University of Nevada, Reno

The roles of divergence and hybridization in shaping patterns of genetic and phenotypic variation across the evolutionary continuum in *Juniperus* and *Piper*

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Ecology, Evolution, and Conservation Biology

by

Kathryn A. Uckele

Dr. Thomas L. Parchman/Dissertation Advisor

August, 2022

THE GRADUATE SCHOOL

We recommend that the dissertation
prepared under our supervision by

KATHRYN A. UCKELE

entitled

The roles of divergence and hybridization in shaping patterns of genetic and phenotypic variation across the evolutionary continuum in Juniperus and Piper.

be accepted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

Thomas L. Parchman, Ph.D.
Advisor

Lee A. Dyer, Ph.D.
Committee Member

Matthew L. Forister, Ph.D.
Committee Member

Elizabeth G. Pringle, Ph.D.
Committee Member

Marjorie D. Matocq, Ph.D.
Graduate School Representative

David W. Zeh, Ph.D., Dean
Graduate School

August, 2022

Abstract

Genetic and phenotypic variation across populations, species, and radiations mediates the form and outcome of biotic and abiotic interactions and represents a major axis of biodiversity. Resolving patterns of variation across shallow and deep evolutionary divergences can provide key insights into the processes that generate and maintain this variation over micro- and macroevolutionary timescales. Additionally, variation in functional traits that interface with the biotic and abiotic environments plays an important role in adaptive evolution, and can shed light on the drivers of differentiation and diversification. Here, I analyzed genome-scale variation spanning individuals, populations, and species to 1) resolve complex diversification histories, 2) characterize landscape patterns of hybrid admixture and plant secondary chemistry, and 3) characterize macroevolutionary patterns of plant secondary chemistry. First, I reconstructed the evolutionary history of the serrate juniper clade of North America (*Juniperus*) as it diversified into arid habitats of the western United States and Mexico. Second, I examined how admixture across the species boundary influences patterns of genetic and phytochemical variation following secondary contact among three serrate juniper species. Finally, I resolve the timing and tempo of diversification in the Radula clade of *Piper* to understand how secondary chemistry evolves within a diverse tropical plant radiation. My work demonstrates the importance of evolutionary processes occurring along the evolutionary continuum for generating contemporary patterns of variation and diversity.

To my parents for cultivating my interests and providing me with every opportunity to succeed as a student and scientist, and to my sister for her friendship and continuous support.

Acknowledgements

I would like to thank my doctoral committee and collaborators, Robert Adams, Lee Dyer, Matthew Forister, Chris Jeffrey, Marjorie Matocq, Casey Philbin, Beth Pringle, Lora Richards, and Eric Tepe, for providing mentorship and constructive feedback that has been foundational to this work. I would like to express my deep gratitude to my major advisor, Thomas Parchman, who provided thoughtful critique, exceptional mentorship, and the opportunity to develop valuable technical skills in coding and bioinformatics.

During my time at UNR, I have benefitted immensely from academic and personal relationships with members of the EECB community. I am very grateful for the support and feedback I received from the Plant-Insect Group and Hitchcock Center for Chemical Ecology. I will always cherish the memories made during climbing and backpacking trips in the Sierras with Valentina Alaasam, Kristin Charles, Trevor Faske, Chris Halsch, Jenn Heppner, Kelli McKeegan, Danny Nielsen, Devon Picklum, Angela Pitera, Sarah Richman, Jessica Riemche, Erin Suenaga, and Anna Tatarko, and am incredibly grateful for the friendships I have shared with Anne Espeset, Sarah Barga, Jacob Francis, Zach Marion, Cynthia Scholl, and Santiago Villamarín. Additionally, I am so thankful for the solidarity and support I experienced as a student in the Parchman Lab, and for this I would like to thank Trevor Faske, Lanie Galland, Joshua Hallas, and Lana Sheta. In particular, I owe a great deal to my friend and mentor, Joshua Jahner, for taking me under his “duck” wing and supporting me every step of the way.

Words cannot express how thankful I am for my partner, Thomas Riecke, and the impact he has made on my life. He is my endless source of adventure, and the one I can always count on for long hikes in the back country, looking for morels or wolf tracks, getting lost without food or

water, and loving every minute of it. I am so grateful to him, and to my dog Luke, for keeping me grounded in the present and moving forward.

Finally, I am indebted to those that assisted me in the lab to make this work possible, including Shelby Burdo, Joy Lapira, Regina Gojar, Brianna Jones, and Matt Paulsen, and am grateful to the organizations that funded my dissertation research, including the National Science Foundation Graduate Research Fellowship, American Genetic Association, the American Society of Plant Taxonomists, and the Hitchcock Center for Chemical Ecology.

Table of Contents

List of tables

List of figures

Introduction

Understanding the processes shaping patterns of genetic and phenotypic variation over micro- and macroevolutionary timescales is an overarching goal of evolutionary biology. Recent advances in DNA sequencing technology have transformed our ability to generate genome-scale information across populations, species, and communities, offering new opportunities to understand how genome variation shapes and is shaped across this continuum of evolutionary processes (Ekblom and Galindo, 2011). For example, recent population genomic analyses have advanced our understanding of the genetics of adaptive divergence (e.g., Barrett et al., 2019; Nosil et al., 2018) and the role of chromosomal rearrangements in adaptation and reproductive isolation (Huang and Rieseberg, 2020; Todesco et al., 2020; Villoutreix et al., 2020). Analyses of hybrid zones have shed light on the genetic basis of hybrid fitness and reproductive isolation (Gompert et al., 2017; Marques et al. 2019) and have documented genome-wide variation in rates of introgression (Marques et al., 2019). Phylogenomic analyses have provided enhanced resolution of complicated evolutionary histories and have led to an increasing appreciation for reticulation during diversification (Payseur and Rieseberg, 2016; Taylor and Larson, 2019; Moran et al., 2021). Finally, high throughput sequencing data have elucidated the molecular genetic underpinnings of functional trait variation and its role in structuring species interactions and ecological communities (Rudman et al. 2018; Barker et al., 2019; Holeski, 2021).

In the last passage of The Origin of Species, Darwin evokes the image of a ‘tangled bank’, a community where diverse species are intricately woven together into a complex network of interactions (Darwin, 2004). Accumulating research has illustrated the role of species interactions in the evolution and diversification of lineages (Thompson 2005; Toju et al. 2017). For example, antagonistic interactions can produce an escalation of co-adaptation in pairs of

species (e.g., Benkman et al. 2010; Reimche et al. 2020), and highly specialized or obligate interactions may result in co-diversification between divergent taxonomic groups (Kato et al., 2003; Herre et al., 2008). Integral to these coevolutionary processes is functional trait variation that influences the strength and outcome of species interactions. An important and extensively studied example is plant secondary chemistry, a multidimensional phenotype that contributes to defense against herbivores and pathogens (Dyer et al. 2018). Variation at this phenotypic interface shapes interactions between plants, insect herbivores, and their natural enemies (Kursar et al., 2009; Richards et al., 2015) and is theorized to play an important role in the origin and maintenance of biodiversity in tropical communities (Ehrlich and Raven, 1964; Janzen, 1970; Connell 1971; Futuyma and Agrawal, 2009). Plant secondary chemistry varies profoundly across species and communities understanding the causes and consequences of this variation is one motivational basis for the field of chemical ecology (Dyer et al., 2018).

Advances in non-targeted metabolomic methods based on nuclear magnetic resonance (NMR) and mass-spectroscopy (MS) have enabled high-throughput chemical profiling and stand to bridge the gap between phytochemical variation, ecological patterns, and evolutionary processes (Dyer et al., 2018). Until recently, characterization of the plant metabolome required isolation and purification of individual compounds and was time and cost prohibitive for analyses of complex chemical mixtures in non-model organisms (Kuhlisch and Pohnert, 2015; Dyer et al., 2018). Consistent improvements in the resolution, sensitivity, and efficiency of MS instruments have transformed the landscape of chemical ecology research, permitting comparisons of phytochemical variation across structurally diverse compound classes and diverged plant lineages (Macel et al., 2010; Sedio, 2017). For example, Sedio et al. (2018) detected over 126,000 compounds across 203 tree species across forest plots in Maryland and

Panama and found that closely related species were more chemically similar at the temperate site, but not the tropical site. Chemical divergence in the tropics may facilitate species coexistence if herbivore resource utilization is predicted by chemical similarity (Wright, 2002; Sedio and Ostling, 2013), and greater chemical variation within tropical communities may underlie host plant divergence by specialist insect herbivores (Becerra, 1997; Dyer et al., 2007). Richards et al., (2015) used ^{1}H NMR to examine the role of phytochemical diversity in structuring plant-insect interactions in 22 species of *Piper*. They found that higher phytochemical diversity was associated with greater numbers of extreme specialists and generalists and higher herbivore diversity overall. Improvements in analytical chemistry also stand to improve our understanding of the evolutionary patterns and consequences of phytochemical variation. Further research is needed to understand the extent to which phytochemistry varies geographically, and how this may generate spatial variation in the form and outcome of species interactions (Thompson, 2005; Zangerl et al., 2008; Dyer et al., 2018). Furthermore, while an accumulating number of studies have shown a lack of statistical dependence among species' chemical profiles due to their phylogenetic relatedness (i.e., lack of phylogenetic signal) (Becerra, 1997; Agrawal and Fishbein, 2006; Kursar et al., 2009; Haak et al., 2014; Salazar et al., 2018), it is unclear whether this pattern holds across evolutionary scales and for all dimensions of secondary chemistry, from broad compound classes to individual molecules. My dissertation research focuses on two plant genera (*Juniperus* and *Piper*) to investigate the relationships between evolutionary history and phytochemical variation across a continuum of evolutionary divergence.

Juniper is the common name for members of *Juniperus* (Cupressaceae), a diverse conifer genus containing approximately 75 species of woody shrubs and trees that inhabit arid and semi-

arid landscapes of the Northern Hemisphere (Farjon, 2005; Adams, 2014). Dispersal and subsequent vicariance characterize the diversification history of this genus, which originated in Asia during the late Eocene, but established lineages in North America via three independent intercontinental dispersals (Mao et al., 2010). One of these lineages diversified into the serrate leaf juniper clade of North America, a group of 21 xerophytic trees and shrubs that range in their distribution from the Pacific Northwest to Guatemala (Adams, 2014). The serrate leaf junipers are distinguished by microscopic serrations on their scale leaf margins and higher drought tolerance than other North American junipers (Willson et al., 2008). Members of this clade are often the dominant trees in arid habitats of the western United States and Mexico (West et al., 1978; Romme et al., 2009), and many are expanding their ranges in North America (Miller and Wigand, 1994; Weisberg et al., 2007; Romme et al., 2009). Junipers produce fleshy female cones that are dispersed by small mammals and birds and support diverse assemblages of arthropods and their natural enemies (Phillips, 1910; Santos et al., 1999; Tonkel et al., 2021). In many serrate junipers, species boundaries are permeable, allowing hybridization to occur in zones of secondary contact (Adams, 1994; Adams, 2014; Adams et al., 2017; Adams et al., 2020). Reticulation and other sources of gene tree incongruence have challenged previous phylogenetic reconstructions (e.g., Adams et al., 2006; Adams and Schwarzbach, 2013) and have limited our understanding of diversification and the role of hybridization in this foundational plant clade.

To investigate geographic and temporal patterns of diversification within the serrate leaf juniper clade, I generated a phylogenomic data set for 68 accessions representing all 22 extant species of serrate leaf junipers, as well as several close and distant relatives. I utilized likelihood, Bayesian, and coalescent-based approaches to produce highly congruent species topologies that provided vastly improved resolution relative to previous analyses based on Sanger sequencing of

nuclear and chloroplast DNA (Uckele et al., 2021a). The relationships among taxa were largely consistent with taxonomic expectations based on geography and morphology, however, several areas of the tree conflicted with relationships inferred with chloroplast DNA, indicating a potential influence of chloroplast introgression or incomplete lineage sorting of cpDNA. With the Bayesian phylogenetic method, I inferred a time-calibrated diversification history for the serrate leaf junipers based on two juniper fossils. My analyses suggest that the serrate leaf juniper lineage arose in North America during the late Oligocene and experienced a period of elevated diversification between 12 and 5 mya (Uckele et al., 2021a). My results improved our understanding of the geographic pattern and tempo of diversification in *Juniperus*, revealed likely instances of reticulation, and highlighted the utility of reduced-representation sequencing for resolving complex evolutionary histories.

To evaluate the consequences of hybridization for genetic and phenotypic variation across the species boundary, I generated and analyzed population genetic data for a group of three closely related serrate leaf juniper species and their hybrids. Here, I documented complex patterns of ancestry and phytochemical variation shaped by admixture among three distinct parental lineages. Hybrid ancestries spanned the range of possible values between parentals, and geographic patterns of ancestry were consistent with secondary contact in western Nevada and asymmetrical gene flow from west to east. Consistent with hybridization across an environmental gradient, hybrid individuals occupied areas with intermediate climates, and climate variables provided accurate predictions of hybrid ancestry across the landscape. My coauthors and I used mass spectroscopy-based methods to characterize terpenoid variation across 163 distinct compounds, 55 of which were matched to structurally annotated terpenoid compounds. The parental species were phytochemically distinct from each other and from

hybrids, and transgressive segregation was an important and distinguishing feature of hybrid terpenoid variation. My results illustrate how history and environment predict ancestry variation across a hybrid zone, and how hybridization can lead to transgressive phytochemical variation which is likely to have extended ecological consequences.

Finally, I resolved patterns of diversification within a large tropical plant radiation (*Piper*) to understand macroevolutionary patterns of phytochemical variation. *Piper* (Piperaceae) is a genus of tropical understory shrubs, vines, and herbs with over 1,300 described species in the New World tropics alone (Martínez et al., 2015). *Piper* is an old lineage (~72 Ma), however, the bulk of its Neotropical diversity arose between 30 and 40 mya following Andean uplift and the emergence of Central America (Smith et al., 2008; Martínez et al., 2015). Radula, the largest *Piper* clade, exemplifies this pattern, as much of its extant diversity (~450 species) arose relatively recently during the Miocene (Martínez et al., 2015). The timing and tempo of diversification in *Piper* is likewise important for understanding the evolution of *Piper*-associated arthropods, many of which are specialist herbivores. Patterns of *Piper* diversification, for example, have influenced diversification in *Eois* (Geometridae: Larentiidae), a diverse genus of lepidopteran *Piper* specialists (Jahner et al., 2017). *Piper* has also become a model genus for understanding the role of plant secondary chemistry in shaping species interactions across trophic levels (Dyer and Palmer, 2004). *Piper* chemistry is impressively diverse, and includes secondary metabolites from over 15 compound classes (Richards et al., 2015). Variation in *Piper* phytochemistry is high, even among closely-related *Piper* species (Salazar et al., 2016), and chemical variation across populations and species shapes the composition of herbivore communities and the outcomes of plant-insect interactions (Richards et al., 2015; Glassmire et al., 2016). A diversity of these interactions have been described in long-term rearing data sets

(Dyer et al., 2007), making this group well-suited to addressing questions about how the evolution of phytochemical defense mediates plant-insect interactions.

To examine evolutionary patterns of secondary chemistry in Radula, I first generated a phylogenomic data set of 9,113 genetic variants across 44 species from the Radula clade of *Piper* and 21 species from seven other *Piper* clades (Uckele et al., 2021b). I utilized a Bayesian phylogenetic analysis to resolve the evolutionary relationships within Radula and among the major *Piper* clades, and to estimate relative divergence times for each node in the phylogeny. To characterize metabolomic variation across the phylogeny, my coauthors and I analyzed the foliar chemistry of each species with proton nuclear magnetic resonance (^{1}H NMR). Sequencing data substantially improved phylogenetic resolution relative to past studies, and spectroscopic characterization revealed the presence of 35 metabolite classes. Evolutionary patterns were variable across the metabolite classes, but overall exhibited higher phylogenetic signal than the crude ^{1}H NMR data. Evolutionary correlations were detected in two pairs of metabolite classes, where the gain or loss of a class was dependent on the other's state. By resolving patterns of phylogenetic and metabolomic variation within Radula, this study provides insight into the evolution of phytochemical diversity in a chemically and taxonomically diverse plant clade (Uckele et al., 2021b).

References

Adams, R.P., 2014. *Junipers of the world: the genus Juniperus*. Trafford Publishing.

Adams, R.P., 1994. Geographic variation and systematics of monospermous *Juniperus* (Cupressaceae) from the Chihuahua Desert based on RAPDs and terpenes. *Biochemical Systematics and Ecology*, *22*(7), pp.699-710.

Adams, R.P., Johnson, S.T. and Schwarzbach, A.E., 2020. Long distance gene flow facilitated by bird-dispersed seeds in wind-pollinated species: A story of hybridization and introgression between *Juniperus ashei* and *J. ovata* told by nrDNA and cpDNA. *Phytologia*, *102*(2), p.55.

Adams, R.P., Gonzalez-Elizondo, M.S., Gonzalez-Elizondo, M., Ramirez Noya, D. and Schwarzbach, A.E., 2017. DNA sequencing and taxonomy of unusual serrate *Juniperus* from Mexico: Chloroplast capture and incomplete lineage sorting in *J. coahuilensis* and allied taxa. *Phytologia*, *99*(1).

Adams, R.P. and Schwarzbach, A.E., 2013. Phylogeny of *Juniperus* using nrDNA and four cpDNA regions. *Phytologia*, *95*(2), pp.179-187.

Adams, R.P., Nguyen, S., Morris, J.A. and Schwarzbach, A.E., 2006. Re-examination of the taxonomy of the one-seeded, serrate leaf *Juniperus* of southwestern United States and northern Mexico (Cupressaceae). *Phytologia*, *88*(3), pp.299-309.

Agrawal, A.A. and Fishbein, M., 2006. Plant defense syndromes. *Ecology*, *87*(sp7), pp.S132-S149.

Ahrens, C.W., Mazanec, R.A., Paap, T., Ruthrof, K.X., Challis, A., Hardy, G., Byrne, M., Tissue, D.T. and Rymer, P.D., 2019. Adaptive variation for growth and resistance to a novel pathogen along climatic gradients in a foundation tree. *Evolutionary Applications*, *12*(6), pp.1178-1190.

Barker, H.L., Riehl, J.F., Bernhardsson, C., Rubert-Nason, K.F., Holeski, L.M., Ingvarsson, P.K. and Lindroth, R.L., 2019. Linking plant genes to insect communities: Identifying the genetic bases of plant traits and community composition. *Molecular Ecology*, *28*(19), pp.4404-4421.

Barrett, R.D., Laurent, S., Mallarino, R., Pfeifer, S.P., Xu, C.C., Foll, M., Wakamatsu, K., Duke-Cohan, J.S., Jensen, J.D. and Hoekstra, H.E., 2019. Linking a mutation to survival in wild mice. *Science*, *363*(6426), pp.499-504.

Becerra, J.X., 1997. Insects on plants: macroevolutionary chemical trends in host use. *Science*, *276*(5310), pp.253-256.

Benkman, C.W., Parchman, T.L. and Mezquida, E.T., 2010. Patterns of coevolution in the adaptive radiation of crossbills. *Annals of the New York Academy of Sciences*, *1206*(1), pp.1-16.

Connell, J.H., 1971. On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. *Dynamics of Populations*, *298*, p.312.

Darwin, C., 2004. *On the Origin of Species, 1859*. Routledge.

Dyer, L.A. and Palmer, A.D. eds., 2004. *Piper: a model genus for studies of phytochemistry, ecology, and evolution* (p. 228). New York: Kluwer academic/Plenum publishers.

Dyer, L.A., Philbin, C.S., Ochsenrider, K.M., Richards, L.A., Massad, T.J., Smilanich, A.M., Forister, M.L., Parchman, T.L., Galland, L.M., Hurtado, P.J. Espeset, A.E., et al., 2018. Modern approaches to study plant–insect interactions in chemical ecology. *Nature Reviews Chemistry*, *2*(6), pp.50-64.

Dyer, L.A., Singer, M.S., Lill, J.T., Stireman, J.O., Gentry, G.L., Marquis, R.J., Ricklefs, R.E., Greeney, H.F., Wagner, D.L., Morais, H.C., Diniz, I.R., et al., 2007. Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, *448*(7154), pp.696-699.

Ehrlich, P.R. and Raven, P.H., 1964. Butterflies and plants: a study in coevolution. *Evolution*, pp.586-608.

Ekblom, R. and Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), pp.1-15.

Farjon, A., 2005. Monograph of Cupressaceae and *Sciadopitys*. Royal Botanic Gardens, Kew, Richmond.

Futuyma, D.J. and Agrawal, A.A., 2009. Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences*, *106*(43), pp.18054-18061.

Glassmire, A.E., Jeffrey, C.S., Forister, M.L., Parchman, T.L., Nice, C.C., Jahner, J.P., Wilson, J.S., Walla, T.R., Richards, L.A., Smilanich, A.M. and Leonard, M.D., 2016. Intraspecific phytochemical variation shapes community and population structure for specialist caterpillars. *New Phytologist*, *212*(1), pp.208-219.

Gompert, Z., Mandeville, E.G. and Buerkle, C.A., 2017. Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*, pp.207-229.

Haak, D.C., Ballenger, B.A. and Moyle, L.C., 2014. No evidence for phylogenetic constraint on natural defense evolution among wild tomatoes. *Ecology*, *95*(6), pp.1633-1641.

Herre, E.A., Jandér, K.C. and Machado, C.A., 2008. Evolutionary ecology of figs and their associates: recent progress and outstanding puzzles. *Annual Review of Ecology, Evolution, and Systematics*, *39*, pp.439-458.

Holeski, L.M., 2021. The Genetic Basis of Plant-Herbivore Interactions. In Plant-Animal Interactions (pp. 59-91). Springer, Cham.

Huang, K. and Rieseberg, L.H., 2020. Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Frontiers in Plant Science*, *11*, p.296.

Janzen, D.H., 1970. Herbivores and the number of tree species in tropical forests. *The American Naturalist*, *104*(940), pp.501-528.

Kato, M., Takimura, A. and Kawakita, A., 2003. An obligate pollination mutualism and reciprocal diversification in the tree genus *Glochidion* (Euphorbiaceae). *Proceedings of the National Academy of Sciences*, *100*(9), pp.5264-5267.

Kuhlisch, C. and Pohnert, G., 2015. Metabolomics in chemical ecology. *Natural Product Reports*, 32(7), pp.937-955.

Kursar, T.A., Dexter, K.G., Lokvam, J., Pennington, R.T., Richardson, J.E., Weber, M.G., Murakami, E.T., Drake, C., McGregor, R. and Coley, P.D., 2009. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus Inga. *Proceedings of the National Academy of Sciences*, *106*(43), pp.18073-18078.

Macel, M., Van Dam, N.M. and Keurentjes, J.J., 2010. Metabolomics: the chemistry between ecology and genetics. *Molecular Ecology Resources*, *10*(4), pp.583-593.

Mao, K., Hao, G., Liu, J., Adams, R.P. and Milne, R.I., 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytologist*, *188*(1), pp.254-272.

Marques, D.A., Lucek, K., Sousa, V.C., Excoffier, L. and Seehausen, O., 2019. Admixture between old lineages facilitated contemporary ecological speciation in Lake Constance stickleback. *Nature Communications*, *10*(1), pp.1-14.

Martinez, C., Carvalho, M.R., Madrinan, S. and Jaramillo, C.A., 2015. A Late Cretaceous *Piper* (Piperaceae) from Colombia and diversification patterns for the genus. *American Journal of Botany*, *102*(2), pp.273-289.

Miller, R.F. and Wigand, P.E., 1994. Holocene changes in semiarid pinyon-juniper woodlands: response to climate, fire, and human activities in the US Great Basin. *BioScience*, *44*(7), pp.465-474.

Moran, B.M., Payne, C., Langdon, Q., Powell, D.L., Brandvain, Y. and Schumer, M., 2021. The genomic consequences of hybridization. *Elife*, *10*, p.e69016.

Nosil, P., Villoutreix, R., de Carvalho, C.F., Farkas, T.E., Soria-Carrasco, V., Feder, J.L., Crespi, B.J. and Gompert, Z., 2018. Natural selection and the predictability of evolution in *Timema* stick insects. *Science*, *359*(6377), pp.765-770.

Paun, O., Turner, B., Trucchi, E., Munzinger, J., Chase, M.W. and Samuel, R., 2016. Processes driving the adaptive radiation of a tropical tree (*Diospyros*, Ebenaceae) in New Caledonia, a biodiversity hotspot. *Systematic Biology*, *65*(2), pp.212-227.

Payseur, B.A. and Rieseberg, L.H., 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*(11), pp.2337-2360.

Phillips, F.J., 1910. The dissemination of junipers by birds. *Forestry Quarterly*, *8*, pp.60-73.

Reimche, J.S., Brodie Jr, E.D., Stokes, A.N., Ely, E.J., Moniz, H.A., Thill, V.L., Hallas, J.M., Pfrender, M.E., Brodie III, E.D. and Feldman, C.R., 2020. The geographic mosaic in parallel: Matching patterns of newt tetrodotoxin levels and snake resistance in multiple predator–prey pairs. *Journal of Animal Ecology*, *89*(7), pp.1645-1657.

Richards, L.A., Dyer, L.A., Forister, M.L., Smilanich, A.M., Dodson, C.D., Leonard, M.D. and Jeffrey, C.S., 2015. Phytochemical diversity drives plant–insect community diversity. *Proceedings of the National Academy of Sciences*, *112*(35), pp.10973-10978.

Romme, W.H., Allen, C.D., Bailey, J.D., Baker, W.L., Bestelmeyer, B.T., Brown, P.M., Eisenhart, K.S., Floyd, M.L., Huffman, D.W., Jacobs, B.F. and Miller, R.F., 2009. Historical and modern disturbance regimes, stand structures, and landscape dynamics in pinon–juniper vegetation of the western United States. *Rangeland Ecology & Management*, *62*(3), pp.203-222.

Rudman, S.M., Barbour, M.A., Csilléry, K., Gienapp, P., Guillaume, F., Hairston Jr, N.G., Hendry, A.P., Lasky, J.R., Rafajlović, M., Räsänen, K. and Schmidt, P.S., 2018. What genomic data can reveal about eco-evolutionary dynamics. *Nature Ecology & Evolution*, *2*(1), pp.9-15.

Salazar, D., Jaramillo, M.A. and Marquis, R.J., 2016. Chemical similarity and local community assembly in the species rich tropical genus Piper. *Ecology*, *97*(11), pp.3176-3183.

Santos, T., Tellería, J.L. and Virgós, E., 1999. Dispersal of Spanish juniper *Juniperus thurifera* by birds and mammals in a fragmented landscape. *Ecography*, *22*(2), pp.193-204.

Sedio, B.E., 2017. Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytologist*, *214*(3), pp.952-958.

Sedio, B.E., Parker, J.D., McMahon, S.M. and Wright, S.J., 2018. Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology*, *99*(12), pp. 2647-2653.

Sedio, B.E. and Ostling, A.M., 2013. How specialised must natural enemies be to facilitate coexistence among plants?. *Ecology Letters*, *16*(8), pp.995-1003.

Smith, J.F., Stevens, A.C., Tepe, E.J. and Davidson, C., 2008. Placing the origin of two species-rich genera in the late cretaceous with later species divergence in the tertiary: a phylogenetic, biogeographic and molecular dating analysis of *Piper* and *Peperomia* (Piperaceae). *Plant Systematics and Evolution*, *275*(1), pp.9-30.

Taylor, S.A. and Larson, E.L., 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, *3*(2), pp.170-177.

Thompson, J.N., 2005. The Geographic Mosaic of Coevolution. In *The Geographic Mosaic of Coevolution*. University of Chicago Press.

Todesco, M., Owens, G.L., Bercovich, N., Légaré, J.S., Soudi, S., Burge, D.O., Huang, K., Ostevik, K.L., Drummond, E., Imerovski, I. and Lande, K., 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), pp.602-607.

Toju, H., Yamamichi, M., Guimarães, P.R., Olesen, J.M., Mougi, A., Yoshida, T. and Thompson, J.N., 2017. Species-rich networks and eco-evolutionary synthesis at the metacommunity level. *Nature Ecology & Evolution*, *1*(2), pp.1-11.

Tonkel, K.C., Dimitri, L.A., Longland, W.S., Kirchoff, V.S. and Rector, B.G., 2021. Parallel paths in a miniature world. *Ecology*, *102*(10), p.e03460.

Uckele, K.A., Adams, R.P., Schwarzbach, A.E. and Parchman, T.L., 2021a. Genome-wide RAD sequencing resolves the evolutionary history of serrate leaf *Juniperus* and reveals discordance with chloroplast phylogeny. *Molecular Phylogenetics and Evolution*, *156*, p.107022.

Uckele, K.A., Jahner, J.P., Tepe, E.J., Richards, L.A., Dyer, L.A., Ochsenrider, K.M., Philbin, C.S., Kato, M.J., Yamaguchi, L.F., Forister, M.L. and Smilanich, A.M., 2021b. Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs. *Scientific Reports*, *11*(1), pp.1-14.

Villoutreix, R., de Carvalho, C.F., Soria-Carrasco, V., Lindtke, D., De-la-Mora, M., Muschick, M., Feder, J.L., Parchman, T.L., Gompert, Z. and Nosil, P., 2020. Large-scale mutation in the evolution of a gene complex for cryptic coloration. *Science*, *369*(6502), pp.460-466.

Weisberg, P.J., Lingua, E. and Pillai, R.B., 2007. Spatial patterns of pinyon–juniper woodland expansion in central Nevada. *Rangeland Ecology & Management*, *60*(2), pp.115-124.

West, N.E., Tausch, R.J., Rea, K.H. and Tueller, P.T., 1978. Phytogeographical variation within juniper-pinyon woodlands of the Great Basin. *Great Basin Naturalist Memoirs*, pp.119-136.

Willson, C.J., Manos, P.S. and Jackson, R.B., 2008. Hydraulic traits are influenced by phylogenetic history in the drought-resistant, invasive genus *Juniperus* (Cupressaceae). *American Journal of Botany*, *95*(3), pp.299-314.

Wright, J.S., 2002. Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, *130*(1), pp.1-14.

Zangerl, A.R., Stanley, M.C. and Berenbaum, M.R., 2008. Selection for chemical trait remixing in an invasive weed after reassociation with a coevolved specialist. *Proceedings of the National Academy of Sciences*, *105*(12), pp.4547-4552.

Chapter 1: Genome-wide RAD sequencing resolves the evolutionary history of serrate leaf *Juniperus* and reveals discordance with chloroplast phylogeny

Kathryn A. Uckele,[1,*] Robert P. Adams,[2] Andrea E. Schwarzbach,[3] and Thomas L. Parchman[1]

[1] Department of Biology, MS 314, University of Nevada, Reno, Max Fleischmann Agriculture Building, 1664 N Virginia St., Reno, NV 89557, USA

[2] Baylor University, Utah Lab, 201 N 5500 W, Hurricane, UT 84790, USA

[3] Department of Health and Biomedical Sciences, University of Texas - Rio Grande Valley, 1 W University Drive, Brownsville, TX 78520, USA

[*]**Address for correspondence:** Kathryn A. Uckele, kathrynuckele@gmail.com

Citation: Uckele, K.A., Adams, R.P., Schwarzbach, A.E. and Parchman, T.L., 2021. Genome-wide RAD sequencing resolves the evolutionary history of serrate leaf Juniperus and reveals discordance with chloroplast phylogeny. *Molecular Phylogenetics and Evolution*, *156*, p.107022.

Abstract

Juniper (*Juniperus*) is an ecologically important conifer genus of the Northern Hemisphere, the members of which are often foundational tree species of arid regions. The serrate leaf margin clade is native to topologically variable regions in North America, where hybridization has likely played a prominent role in their diversification. Here we use a reduced-representation sequencing approach (ddRADseq) to generate a phylogenomic data set for 68 accessions representing all 22 species in the serrate leaf margin clade, as well as a number of close and distant relatives, to improve understanding of diversification in this group. Phylogenetic analyses using three methods (SVDquartets, maximum likelihood, and Bayesian) yielded highly congruent and well-resolved topologies. These phylogenies provided improved resolution relative to past analyses based on Sanger sequencing of nuclear and chloroplast DNA, and were largely consistent with taxonomic expectations based on geography and morphology. Calibration of a Bayesian phylogeny with fossil evidence produced divergence time estimates for the clade consistent with a late Oligocene origin in North America, followed by a period of elevated diversification between 12 and 5 Mya. Comparison of the ddRADseq phylogenies with a phylogeny based on Sanger-sequenced chloroplast DNA revealed five instances of pronounced discordance, illustrating the potential for chloroplast introgression, chloroplast transfer, or incomplete lineage sorting to influence organellar phylogeny. Our results improve understanding of the pattern and tempo of diversification in *Juniperus*, and highlight the utility of reduced-representation sequencing for resolving phylogenetic relationships in non-model organisms with reticulation and recent divergence.

Keywords: diversification, juniper, RADseq, reticulation, western North America

Introduction

The complex geologic and climatic history of western North America played an important role in the diversification of many plant groups throughout the Cenozoic (Axelrod, 1948, 1950). Tectonic uplift, climate change, transcontinental land bridges, and glacial cycles created opportunity for range shifts, geographic barriers to admixture, and allopatric speciation (Hewitt, 1996; Calsbeek et al., 2003; Hewitt, 2004; Weir and Schluter, 2007). Hybridization has also been prominent in the evolutionary history of Nearctic plant taxa, as glacial cycles allowed periods of isolation and subsequent secondary contact (Swenson and Howard, 2005; Hewitt, 2011). The interactions among topography, climate, and reticulation have shaped diversification and challenged phylogenetic analyses for many plant genera in western North America (e.g., Rieseberg et al., 1991; Kuzoff et al., 1999; Bouillé et al., 2011; Xiang et al., 2018; Shao et al., 2019). However, improved genomic sampling enabled by high-throughput sequencing data has recently increased phylogenetic resolution for many young and reticulated groups (e.g., Stephens et al., 2015; Massatti et al., 2016; McVay et al., 2017; Moura et al., 2020) and generally stands to enhance our understanding of diversification for plant taxa in this region.

Junipers (*Juniperus*, Cupressaceae) are ecologically and economically important conifers of arid and semi-arid landscapes throughout the Northern Hemisphere (Farjon, 2005; Adams, 2014). Unlike other genera in Cupressaceae, the juniper lineage evolved a fleshy female cone, functionally resembling a berry, which is an important food source for many birds and small mammals (Phillips, 1910; Santos et al., 1999). The serrate junipers, distinguished by the presence of microscopic serrations on their scale leaf margins, are particularly resistant to water stress compared with other juniper groups (Willson et al., 2008) and often represent the dominant trees in arid habitats of the western United States and Mexico (West et al., 1978; Romme et al., 2009).

A number of species in this clade are expanding their range in North America, and while the main causes of these expansions are unclear for some taxa (Miller and Wigand, 1994; Weisberg et al., 2007; Romme et al., 2009), fire suppression, over-grazing by cattle, and under-browsing by native herbivores appear to be the dominant factors underlying *J. ashei* and *J. pinchotii* range expansion in west Texas (Taylor, 2008). Despite several attempts to resolve phylogenetic relationships in this ecologically important clade (Mao et al., 2010; Adams and Schwarzbach, 2013a,b), its complex evolutionary history including recent divergence, long generation times, and hybridization have likely obfuscated phylogenetic signal in previous molecular data sets.

The juniper lineage likely originated in Eurasia during the Eocene and subsequently split into three major monophyletic sections (Mao et al., 2010; Adams and Schwarzbach, 2013a): sect. Caryocedrus (1 sp., *J. drupacea*, eastern Mediterranean); sect. Juniperus (14 spp., Asia and the Mediterranean except *J. jackii* and *J. communis*); and the largest clade, sect. Sabina (approximately 62 spp., Northern Hemisphere except *J. procera*). Section Sabina contains three main monophyletic clades (Mao et al., 2010; Adams and Schwarzbach, 2013a): the turbinate, single-seeded, entire leaf margin junipers of the Eastern Hemisphere (16 spp.); the multi-seeded, entire leaf margin junipers of both the Eastern and Western Hemispheres (23 spp.); and the serrate leaf margin junipers (serrate junipers hereafter) of western North America (22 spp.), which are the focus of this study. The ancestral serrate juniper lineage likely arrived in North America from Eurasia via the North Atlantic Land Bridge (NALB) or Bering Land Bridge (BLB) (Mao et al., 2010). Extant serrate junipers are largely restricted to North America, inhabiting arid and semi-arid regions of the western United States, Mexico, and the high, dry mountains of Guatemala (*J. standleyi*; Adams, 2014) (Fig. 1).

A previous phylogenetic analysis based on Sanger sequencing data with complete species-level sampling of the serrate juniper clade was highly biased towards chloroplast DNA (cpDNA), utilizing four cpDNA regions and one nuclear DNA (nrDNA) region [full data set representing 4,411 base pairs (bp), referred to as nr-cpDNA hereafter; Adams and Schwarzbach, 2013b]. Hybridization and discordance between cpDNA and nrDNA based phylogenies have been reported across *Juniperus* (Adams, 2016; Adams et al., 2016) and within the serrate junipers in particular (Adams et al., 2017) and may have contributed to unexpected topologies in the previous predominantly cpDNA based phylogeny (Adams and Schwarzbach, 2013b). Incomplete lineage sorting due to long generation times and recent divergence may have also contributed to paraphyletic and unresolved relationships in the nr-cpDNA analyses of Adams and Schwarzbach (2013b). Multi-locus data encompassing larger genealogical variation should reduce topological uncertainty in this clade, while also allowing for insight into nuclear-chloroplast discordance and its potential causes. Mao et al. (2010) estimated divergence times, diversification rates, and geographic origins of all major juniper clades; however, limited sampling of the serrate juniper clade precluded dating for many of its internal nodes. Divergence time estimation for a complete serrate juniper phylogeny stands to elucidate patterns of diversification at more recent time scales which appear to be important for diversification across the genus (Mao et al., 2010).

High-throughput sequencing technologies have rapidly improved our ability to apply genome-wide information to phylogenetic inference (McCormack et al., 2013; Leaché and Oaks, 2017; Bravo et al., 2019). Data from whole genomes (e.g., Kimball et al., 2019; Allio et al., 2020), whole transcriptomes (e.g., Leebens-Mack et al., 2019), targeted capture (e.g., de La Harpe et al., 2019; Liu et al., 2019; Karimi et al., 2020), and genome-skimming approaches (e.g.,

Liu et al., 2020; Nevill et al., 2020) have resolved evolutionary relationships complicated by incomplete lineage sorting and reticulate evolution (Faircloth et al., 2013; Alexander et al., 2017; Carter et al., 2019). Methods using restriction enzyme digest to reduce genome complexity [e.g., restriction site-associated DNA sequencing (RADseq; Miller et al., 2007; Baird et al., 2008)] have been particularly valuable for phylogenetic applications in non-model organisms due to their ability to sample large numbers of informative polymorphisms without requiring prior genomic resources (Takahashi et al., 2014; Leaché and Oaks, 2017; Near et al., 2018; Salas-Lizana and Oono, 2018; Hipp et al., 2020). RADseq data have improved the resolution of many groups that have been recalcitrant to phylogenetic analysis with small numbers of Sanger-sequenced loci due to rapid, recent, or reticulate evolution (Wagner et al., 2013; Massatti et al., 2016; Paetzold et al., 2019; Rancilhac et al., 2019; Léveillé-Bourret et al., 2020). Although allelic dropout (i.e., the nonrandom absence of sequence data at a locus due to restriction site mutations) can result in larger amounts of missing data across more strongly diverged lineages, analyses of empirical and simulated RADseq data have illustrated its effectiveness for resolving even relatively deep divergences (e.g., up to 60 Mya, Rubin et al., 2012; Cariou et al., 2013; Eaton et al., 2017; Lecaudey et al., 2018; Du et al., 2020).

Here we utilized a double-digest RADseq approach (ddRADseq; Parchman et al., 2012; Peterson et al., 2012) to generate a phylogenomic data set for all extant species of serrate junipers (*Juniperus* sect. Sabina) as well as several close and distant relatives. As methods for phylogenetic inference utilizing multi-locus data make different assumptions about genealogical variation among lineages, we inferred phylogenetic trees using three distinct approaches (SVDquartets, maximum likelihood, and Bayesian). Our results produce consistent and highly resolved topologies, reveal discordance with phylogenies inferred with cpDNA alone, and

illustrate variation in diversification rates consistent with the climatic and geologic history of western North America.

Materials & Methods

Taxon sampling and ddRADseq library prep

We sampled leaf material from 68 individuals representing all 22 serrate juniper species and six outgroup species (Table S1). Most serrate juniper taxa and two outgroup taxa (*Hesperocyparis bakeri* and *H. arizonica*, Cupressaceae; Zhu et al., 2018) were either the same individuals or different individuals collected from the same populations as those analyzed previously by Adams and Schwarzbach (2013b). Thus, analyses of the data presented here have 50 samples (73.5%) in common with Adams and Schwarzbach (2013b) and 18 samples (26.5%) which are unique to this study. Five additional outgroup taxa [*Juniperus drupacea* (*Juniperus* sect. Caryocedrus); *J. communis* (*Juniperus* sect. Juniperus); *J. virginiana, J. sabina* var. *sabina*, and *J. sabina* var. *balkanensis* (smooth leaf junipers of sect. Sabina)] were added to better understand evolutionary divergence at deeper time scales in this genus. Two additional *J. poblana* var. *poblana* localities (Nayarit, MX, and Amozoc de Mota, Puebla, MX), one additional *J. poblana* variety (*J. poblana* var. *decurrens*), and an additional *J. durangensis* locality (Sierra de Gamón, Durango, MX) were included to investigate the potential for recent evolutionary divergence in these taxa. Finally, we substituted *J. ashei* samples from Waco, TX, with *J. ashei* samples from nearby Tarrant County, TX, for this study.

DNA was extracted from dried leaf tissue with Qiagen DNeasy Plant Mini Kits and quantified with a Qiagen QIAxpert microfluidic analyzer prior to library preparation (Qiagen

Inc., Valencia, CA, USA). Reduced-representation libraries for Illumina sequencing were constructed using a ddRADseq method (Parchman et al., 2012; Peterson et al., 2012) in which genomic DNA was digested with two restriction enzymes, *Eco*RI and *Mse*I, and custom oligos with Illumina base adaptors and unique barcodes (8, 9 or 10 bases in length) were ligated to the digested fragments. Ligated fragments were PCR amplified with a high-fidelity proofreading polymerase (Iproof polymerase, BioRad Inc., Hercules, CA, USA) and subsequently pooled into a single library. Libraries were size-selected for fragments between 350 and 450 bp in length with the Pippin Prep System (Sage Sciences, Beverly, MA) at the University of Texas Genome Sequencing and Analysis Facility. Two lanes of single-end 100-base sequencing were executed at the University of Wisconsin-Madison Biotechnology Center using an Illumina HiSeq 2500 platform.

Preparation, filtering, and assembly of ddRADseq data

To identify and discard Illumina primer/adapter sequences and potential biological sequence contaminants (e.g., PhiX, *E. coli*), we used the `tapioca` pipeline (https://github.com/ncgr/tapioca), which uses `bowtie2` (v. 2.2.5; Langmead and Salzberg, 2012) to identify reads which align to a database of known contaminant sequences. To ensure that cpDNA did not influence our analyses, we used the same approach to discard all reads which aligned to the *Juniperus squamata* chloroplast genome (GenBank Accession Number MK085509; Xie et al., 2019). To demultiplex reads to individual, we used a custom Perl script that corrects one or two base sequencing errors in barcoded regions, parses reads according to their associated barcode sequence, and trims restriction site-associated bases. Files with the read data for each individual are available at Dryad (https://doi.org/10.5061/dryad.qbzkh18df).

To process the raw data into a matrix of putatively orthologous aligned loci, we utilized `ipyRAD` (v. 0.9.16; Eaton, 2014) which was designed to process reduced-representation data for phylogenetic workflows and allows for indel variation across samples during clustering (Eaton, 2014; Razkin et al., 2016). We largely used default values, as these settings produced multiple alignments of tractable size which led to highly resolved, supported, and consistent topologies across inference methods. First, nucleotide sites with phred quality scores less than 33, which represent base calls with an error probability greater than 0.0005%, were considered missing and replaced with an ambiguous nucleotide base ("N"). Next, sequences were *de novo* clustered within individuals using `vsearch` (v. 2.14.1; Rognes et al., 2016) and aligned with `muscle` (v. 3.8.155; Edgar, 2004) to produce stacks of highly similar reads. A similarity clustering threshold (*clust_threshold*) of 85% was applied during this and a later clustering step because it produced a thorough yet tractable number of loci and a highly supported topology with the `TETRAD` (SVDquartets) inference method. To ensure accurate base calls, all stacks with a read depth less than 6 were discarded. Observed base counts across all sites in all stacks informed the joint estimation of the sequencing error rate and heterozygosity, which informed statistical base calls according to a binomial model. At this step, each stack within each individual was reduced to one consensus sequence with heterozygote bases represented by IUPAC ambiguity codes, and any consensus sequences with more than 5% ambiguous bases (*max_Ns_consens*) or heterozygous sites (*max_Hs_consens*) were discarded to remove poor alignments. The remaining consensus sequences from all individuals were clustered again, this time across individuals, using the same assembly method and similarity threshold as used in the previous within-sample clustering step. The resulting clusters, which represent putative ddRADseq loci shared across individuals, were discarded if they contained more than 8 indels (*max_Indels_locus*) or 20%

variable sites (*max_SNPs_locus*), as an excess of either could indicate poor alignment. To detect potential paralogs, consensus sequences were removed if they contained one or more heterozygous sites shared across more than 50% of all samples (*max_shared_Hs_locus*) or more than 2 haplotypes (Eaton, 2014). We retained all loci that were present in a minimum of four samples (*min_samples_locus*) to prevent over-filtering of missing data, which can negatively affect downstream inference (Rubin et al., 2012; Wagner et al., 2013; Huang and Knowles, 2014; Takahashi et al., 2014). Two sequence alignment formats, `ipyRAD`'s database file and a `phylip` file of concatenated loci, were used as input for SVDquartets (`TETRAD`) and maximum likelihood (`RAxML`) phylogenetic analyses, respectively. The database file contains the clustered sequence data as well as linkage information for each locus. We used a python script (http://github.com/btmartin721/raxml_ascbias/) to remove all invariant sites from the `phylip` sequence alignment prior to analysis with `RAxML`.

To understand the timing and tempo of diversification within the serrate juniper clade, we utilized fossil evidence to inform divergence time estimates in a Bayesian phylogenetic inference framework. For this analysis, we included one sample per serrate juniper species, including three outgroup samples from the closely related smooth leaf juniper clade (*J. virginiana*, *J. sabina* var. *sabina*, and *J. sabina* var. *balkanensis*), with priority given to juniper samples with higher sequencing coverage depth. Sequencing reads for this subset of 25 samples were *de novo* assembled with default `ipyRAD` parameter values except for the *min_samples_locus* parameter, which was increased from 4 to 20, and the *clust_threshold* parameter, which was increased from 85% to 90%. Increasing these parameters effectively reduced both the proportion of missing data and the size of the sequence alignment to ensure tractable computation time with Bayesian inference methods. However, one caveat of excluding missing data in RADseq data sets is that it

can bias the distribution of mutation rates represented across loci and lower the accuracy of downstream phylogenetic inference (Huang and Knowles, 2014). The resulting `nexus` sequence alignment of concatenated loci was utilized as input for Bayesian analysis (`RevBayes`). Complete information on parameter settings for this and the aforementioned assembly, as well as the sequence alignment files, are archived at Dryad (https://doi.org/10.5061/dryad.qbzkh18df).

Phylogenetic analyses

After removing invariant sites, the `phylip` formatted sequence alignment for all taxa, including outgroups, was analyzed with maximum likelihood as implemented by `RAxML` (v. 8.2.12; Stamatakis, 2014) under the GTR+Γ model of nucleotide substitution corrected for ascertainment bias (`-m ASC_GTRGAMMA`). Support was assessed with 100 rapid bootstrap replicates (`-N 100`), followed by a thorough maximum likelihood search for the best-scoring tree (`-f a`). Although `RAxML` is fast and often used for analysis of concatenated RADseq loci (Lemmon and Lemmon, 2013), phylogenetic inference with concatenated data necessarily ignores genealogical variation among loci and is statistically inconsistent as the number of genes increases (Kubatko and Degnan, 2007; Roch and Steel, 2015).

To account for genealogical variation among sampled loci and to incorporate coalescent stochasticity into analyses, we also conducted species tree inference using a site-based approach, SVDquartets (Chifman and Kubatko, 2014), as implemented by `TETRAD` (Eaton et al., 2017). `TETRAD` is included with `ipyRAD` and implements the SVDquartets algorithm, using information on genotype calls and linkage to sample unlinked SNPs. Briefly, SVDquartets uses the multi-species coalescent model to generate a probability distribution on the data patterns at the tips of a species tree which can be used to compute a score on a quartet of taxa and infer the true quartet

relationship (Chifman and Kubatko, 2014, 2015). These quartet relationships can be inferred for all or a subset of all possible quartets, and a quartet amalgamation software (in this case, `QMC` v. 2.10; Snir and Rao, 2012) joins the inferred quartets into the species tree. Here, we used `TETRAD`'s default number of quartets, which is the number of samples to the power of 2.8, which yielded 135,215 quartets (16.6% of total possible). To quantify support for the nodes of the species tree, we implemented a standard nonparametric bootstrapping procedure for 100 replicates. The inferred tree was manually rooted with the clade containing *Hesperocyparis bakeri* and *H. arizonica*.

To enable comparison of topologies produced with ddRADseq and cpDNA Sanger sequencing data, we repeated the methods of Adams and Schwarzbach (2013b) on the same individuals or different individuals collected from the same populations as those analyzed in the ddRADseq analysis for a total of 66 individual samples. Thus, the cpDNA analysis presented here has 59 samples (89.4%) in common with the aforementioned ddRADseq analyses and 7 substitutional samples (10.6%). DNA extractions, PCR amplifications, and Sanger sequencing of the four chloroplast loci (petN-psbM, trnS-trnG, trnD-trnT, and trnL-trnF) were conducted using the methods described in Adams and Schwarzbach (2013b). The GTR+Γ+I nucleotide substitution model provided the best fit to the cpDNA data according to Akaike's information criterion in `Modeltest` (v.3.7; Posada and Crandall, 1998), and analysis was conducted with `Mr.Bayes` (v.3.1; Ronquist and Huelsenbeck, 2003). Two rounds of four chains were run for a total of 10 million generations, sampling every 1000 generations after an initial burn in of 25% of generations.

To understand diversification rate variation and the timing of divergence events across the serrate juniper clade, we inferred a time-calibrated phylogeny for a subset of individuals

representing all serrate juniper taxa and three closely related outgroup samples from the smooth leaf juniper clade (*J. virginiana*, *J. sabina* var. *sabina*, and *J. sabina* var. *balkanensis*) with a Bayesian method (`RevBayes` v. 1.0.12; Höhna et al., 2017). First, we implemented a model-selection procedure to compare the relative fits with Bayes factors of the JC, HKY, GTR, GTR+Γ, and GTR+Γ+I models of nucleotide substitution. Second, the `nexus` sequence alignment of concatenated loci generated with `ipyRAD` was modeled under the best fit substitution model given a topology modeled with a constant-rate birth-death process, which was parameterized with a sampling fraction of 0.39 due to incomplete sampling of the smooth leaf juniper clade (Kendall, 1948; Nee et al., 1994; Höhna, 2015). We relaxed the assumption of a global molecular clock by allowing each branch-rate variable to be drawn from a lognormal distribution. Eight independent MCMC chains were run for 400,000 generations with a burn-in of 10,000 generations and sampled every 10 generations. Chains were visually assessed for convergence with `Tracer` (v. 1.7.1; Rambaut et al., 2018) and quantitatively assessed with effective sample sizes (ESS) and the Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992) using the *gelman.diag* function in R (`CODA` package; Plummer at al., 2006).

Fossil calibration points and node age prior distributions can influence estimates of divergence times (Graur and Martin, 2004; Sauquet et al., 2012; Wang and Mao, 2016). We used three fossil calibration points: one at the root node for the serrate juniper clade (not shown in Fig. 4A) and two at internal nodes (asterisks, Fig. 4A) representing the MRCA (Most Recent Common Ancestor) of all extant serrate leaf junipers and the MRCA of the western U.S. clade (*J. californica*, *J. osteosperma*, *J. occidentalis*, and *J. grandis*). Fossil assignments were based on morphology and coincided with those made by a previous phylogenetic analysis of *Juniperus* (Mao et al., 2010). Justifications for these assignments can be found in Table S2. A fossil

specimen of *J. creedensis* (23 Mya; Axelrod, 1987), representing the first appearance of a serrate juniper in the fossil record, provided the minimum age constraints for both the root node (representing the MRCA of the serrate leaf juniper clade and the smooth leaf juniper outgroup taxa) and the internal node representing the MRCA of the serrate junipers. The maximum age constraint for the root node, specified with a uniform prior distribution, was the estimated age of the crown lineage of Cupressoideae (134 Mya; Mao et al., 2012), a subfamily of Cupressaceae which contains *Thuja*, *Cupressus*, *Juniperus*, and other genera. A fossil specimen of *J. desatoyana* (16 Mya; Axelrod, 1991), representing a stem ancestor of a subclade containing *J. osteosperma*, *J. occidentalis*, and *J. grandis*, provided the minimum age constraint of 16 Mya for the divergence of this subclade from *J. californica* (i.e., the MRCA of the western U.S. clade). For the internal nodes representing the MRCA of the serrate leaf junipers and the MRCA of the western U.S. clade, the ages of the fossil specimens were modelled as exponential distributions with means of 23 Mya + 1 and 16 Mya + 1, respectively, divided by λ, the parameter of the exponential distribution. The maximum clade credibility tree was inferred from the burned distribution of posterior trees, and the smooth leaf juniper outgroup samples were pruned in R with the *drop.tip* function (`ape` package; Paradis and Schliep, 2019) prior to subsequent visualization and analyses.

The inferred Bayesian chronogram was used to generate a lineage through time plot with the *ltt.plot* function in R (`ape` package; Paradis and Schliep, 2019). To determine whether the rate of lineage diversification was constant through time, we used the *diversi.gof* function in R (`ape` package; Paradis and Schliep, 2019) to compute the Cramér-von Mises and Anderson-Darling goodness-of-fit tests (Stephens, 1974; Paradis, 1998).

To estimate the probability of all possible ancestral ranges at each ancestral node, we utilized the `BioGeoBEARS` package (v. 1.1.2; Matzke, 2013a,b) and its dependencies, `rexpokit` (Matzke et al., 2019) and `cladoRcpp` (Matzke, 2018), in R. This package permits statistical selection of six competing historical biogeographical models (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE, and BAYAREALIKE+J) and includes an additional cladogenetic event, founder-event speciation, represented by the +J notation in DEC+J, DIVALIKE+J, and BAYAREALIKE+J models (Matzke, 2014). While these six methods similarly assume that anagenetic dispersal and extinction occur along branches, they allow for different subsets of cladogenetic range-changing processes. The `BioGeoBEARS` supermodel incorporates all of these different processes, treating them as free parameters which can be excluded or estimated from the data.

Five operational geographic areas (A, western U.S.; B, central U.S.; C, eastern U.S.; D, northern/central MX; E, southern MX; Fig. 5) were defined by both geopolitical and ecologically-relevant boundaries (Level I Ecoregions of North America; see https://www.epa.gov/eco-research/ecoregions). To determine the contemporary geographic range of each species, we referenced U.S. tree species range maps when available (Little, 1971) and juniper range maps otherwise (Adams, 2014) (Table S3). This matrix of distribution information for each species, as well as the maximum clade credibility tree inferred with `RevBayes`, was used as input for ancestral range estimation. We used plotting functions provided by `BioGeoBEARS` to visualize estimates of ancestral range for the model with the lowest AIC.

Results

Assembly of ddRADseq data for phylogenetic inference

Two Illumina HiSeq lanes generated approximately 460 million reads, of which 373,596,722 remained after quality and contaminant filtering. `Bowtie2` aligned 4,007,039 reads (1.07%) to the *J. squamata* chloroplast genome, which we subsequently removed prior to read assembly and SNP calling. Three samples were removed prior to assembly due to low read count relative to other samples, providing 68 samples for `ipyRAD` input. The full data set of 68 samples was initially assembled into 307,146 loci, of which 130,581 remained after filtering, providing 929,267 SNPs (344,189 parsimony informative) for phylogenetic inference with `RAxML` and `TETRAD`. Each individual possessed, on average, approximately five million raw reads which were assembled, on average, into 19,417 loci (14.9% of total loci). Similar to other RADseq phylogenetic data sets (Cariou et al., 2013; Eaton et al., 2017), the resulting sequence alignments provided as input for `RAxML` and `TETRAD` exhibited a large proportion of missing data (84.69% and 83.51% of sites contained missing values, respectively). 10,461,968 invariant sites were removed from the `phylip` formatted sequence alignment prior to analysis with `RAxML`. `TETRAD` sampled 124,530 unlinked SNPs for its analysis.

For the Bayesian analysis, increasing the *min_samples_locus* and *clust_threshold* parameters for assembly of the 22 serrate juniper and 3 outgroup samples effectively diminished the effect of allelic dropout and reduced the proportion of missing data at the expense of incorporating fewer loci for phylogenetic inference. An initial set of 479,143 loci were reduced to 2,390 after filtering steps, providing 18,436 SNPs (7,894 parsimony informative) for phylogenetic inference. On average, each individual possessed 5.7 million raw reads which were assembled into 2,078 loci (86.9% of total loci). Only 14.72% of sites contained missing values in the resulting `nexus` sequence alignment.

Phylogenetic analyses

The maximum likelihood and SVDquartets analyses of ddRADseq data (hereafter referred to as the ddRADseq phylogenies) recovered high support (>95%) for most nodes in the phylogeny, with few exceptions (Fig. 2). The maximum likelihood phylogeny identified nine monophyletic clades within the serrate junipers (Fig. 2 left), which are colored accordingly in Figs. 2-4. The SVDquartets phylogeny resolved the same nine clades (Fig. 2 right), although two were less supported: 1) the Cerro Petosí clade (*J. zanonii* and *J. saltillensis*, which are sympatric on Cerro Petosí, MX) and 2) the subalpine-alpine clade (*J. jaliscana*, *J. standleyi*, and *J. monticola*, which are collectively found in subalpine/alpine environments). The ddRADseq phylogenies consistently recovered deeper relationships among three main monophyletic clades: 1) the western U.S. clade (*J. californica*, *J. osteosperma*, *J. occidentalis*, and *J. grandis*); 2) the *ashei* clade (*J. comitana*, *J. ovata*, and *J. ashei*), the *J. deppeana* species complex, the one-seeded serrate junipers (*J. arizonica*, *J. monosperma*, *J. coahuilensis*, *J. pinchotii*, and *J. angosturana*, which largely exhibit 1 seed per cone); and 3) the Cerro Petosí clade, the *J. durangensis* clade (*J. martinezii* and *J. durangensis* subspp.), the subalpine-alpine clade, *J. flaccida*, and the *J. poblana* species complex. The ddRADseq phylogenies were consistent in their relationships among the three high-level clades, including the placement of the western U.S. clade as basal to the other serrate juniper clades (Fig. 2). Although nearly all relationships were strongly supported and consistent across both phylogenies (Fig. 2), three were inconsistently resolved. In the maximum likelihood phylogeny, the outgroup taxa *J. drupacea* and *J. communis* are in distinct clades, whereas they are sister to one another in the SVDquartets phylogeny (Fig. 2). In the maximum likelihood phylogeny, the *ashei* clade is basal to the *J. deppeana* species complex and the one-seeded group with high support; whereas, in the SVDquartets phylogeny,

the *J. deppeana* species complex is basal, with high support (Fig. 2). Finally, although both placements had low support, the maximum likelihood phylogeny placed *J. flaccida* as sister to the *J. poblana* complex, whereas the SVDquartets phylogeny placed *J. flaccida* as basal to the subalpine-alpine clade (Fig. 2).

Aside from the few conflicts above, the topologies inferred across multiple approaches (maximum likelihood, SVDquartets, and Bayesian) were consistent, highly supported, and congruent with established taxonomy based on morphological and chemical characters (Figs. 2, 4A). Whereas Adams and Schwarzbach (2013b) inferred a paraphyletic relationship for *J. sabina* in which *J. virginiana* was sister to *J. sabina* var. *sabina* (Fig. 1 from Adams and Schwarzbach, 2013b), the ddRADseq phylogenies recovered a monophyletic relationship for the two *J. sabina* varieties (Fig. 2). In addition, three of the nine monophyletic clades recovered with generally high support in the ddRADseq phylogenies (Fig. 2) were paraphyletic in the nr-cpDNA phylogeny of Adams and Schwarzbach (2013b): 1) the western U.S. clade; 2) the *J. ashei* clade; and 3) the subalpine-alpine clade. First, the western U.S. clade was paraphyletic in the nr-cpDNA tree of Adams and Schwarzbach (2013b) and is not basal to the other serrate juniper clades, except for *J. californica*. Second, the *J. ashei* clade was paraphyletic in the nr-cpDNA tree, with *J. comitana* basal to the western U.S. clade, *J. ovata* basal to the Cerro Petosí clade, and *J. ashei* sister to *J. deppeana* (Fig. 1 from Adams and Schwarzbach, 2013b). Third, the nr-cpDNA tree of Adams and Schwarzbach (2013b) placed *J. flaccida* and *J. poblana* in the subalpine-alpine clade, causing the subalpine-alpine clade to be paraphyletic.

Sanger-sequenced data spanning four cpDNA regions (petN-psbM, trnS-trnG, trnL-trnF, trnD-trnT), originally generated by Adams and Schwarzbach (2013b), was reanalyzed here with additional samples to produce a phylogeny for detection of cyto-nuclear discordance when

compared with ddRADseq phylogenies (both analyses were largely based on the same sets of individuals, or individuals from the same populations). The cpDNA phylogeny inferred here had less resolution and a distinctly different topology than that of the combined nr-cpDNA analysis of Adams and Schwarzbach (2013b). Figure 3 illustrates five areas of discordance between the maximum likelihood ddRADseq and Bayesian cpDNA phylogenies. First, the cpDNA phylogeny inferred a sister relationship between *J. sabina* var. *balkinensis* and *J. virginiana* (Fig. 3 right), whereas the maximum likelihood phylogeny inferred a sister relationship between *J. sabina* var. *balkinensis* and *J. sabina* var. *sabina* (Fig. 3 left), consistent with taxonomic expectations. Second, the western U.S. clade is paraphyletic in the cpDNA tree, and *J. californica* is sister to *J. comitana* rather than grouped with the other western U.S. serrate junipers (Fig. 3 right). Third, the cpDNA tree placed *J. zanonii* sister to *J. ovata* and nested within a clade with *J. ashei* (Fig. 3 right), rather than sister to *J. saltillensis* as it is in the maximum likelihood tree (Fig. 3). Fourth, the cpDNA tree also included *J. arizonica* in this highly supported clade, making the one-seeded group (*J. arizonica*, *J. monosperma*, *J. coahuilensis*, *J. pinchotii*, and *J. angosturana*) paraphyletic (Fig. 3 right). Finally, in the cpDNA tree, *J. flaccida* is nested within *J. poblana*, which causes this complex to be paraphyletic (Fig. 3 right).

Diversification history of the serrate junipers

The GTR+ Γ model of nucleotide substitution provided the best fit to the sequence alignment generated for the subset of serrate juniper samples, including three outgroup samples (*J. virginiana*, *J. sabina* var. *sabina*, and *J. sabina* var. *balkanensis*). The Bayesian topology was largely consistent with the maximum likelihood and SVDquartets phylogenies, with an exception being the paraphyletic relationship among the one-seeded junipers (Fig. 4A). The other eight of

the nine monophyletic clades and all three high-level clades recovered by the ddRADseq phylogenies (Fig. 2) were likewise recovered by the Bayesian phylogeny (Fig. 4A) with high support (>99% posterior support for all nodes; Figure 4A). Our Bayesian calibration suggests that the serrate juniper clade arose during the late Oligocene (crown age 23.73 Mya, 95% highest posterior density [HPD]: 23 – 25.15 Mya), which is slightly younger but not inconsistent with previous estimates of 25.82 (23.00 – 31.20) and 29.43 Mya (23.25 – 41.72) inferred from cpDNA data with BEAST and MULTIDIVTIME, respectively (Mao et al. 2010). According to our analysis, the western U.S. clade (*J. californica*, *J. osteosperma*, *J. occidentalis*, and *J. grandis*) arose in the early Miocene (crown age 17.20 Mya, HPD: 16.00 – 19.32 Mya), which is slightly younger but not inconsistent with previous estimates of 19.16 (16.00 – 25.44) and 24.39 (15.88 – 36.64) Mya inferred from cpDNA with BEAST and MULTIDIVTIME, respectively (Mao et al., 2010).

The Bayesian phylogenetic model estimated a mean speciation rate for the serrate juniper and closely related smooth leaf juniper clades of 0.14 sp/Ma (HPD: 2.46E-5 – 0.21 sp/Ma), and an extinction rate of 0.03 sp/Ma (HPD: 7.18E-8 – 0.11 sp/Ma), resulting in a mean net diversification rate (speciation rate – extinction rate) of 0.11 sp/Ma (HPD: -0.07 – 0.20 sp/Ma). A lineage through time plot (Fig. 4B) suggests deviations from a constant rate of diversification over time, which was confirmed quantitatively with the Cramér-von Mises and Anderson-Darling goodness-of-fit tests, both of which rejected the null model of constant diversification rate and exponentially distributed branching times (Cramér-von Mises: $W2 = 2.326$, $p < 0.01$; Anderson-Darling GOF: $A2 = 3.189$, $p < 0.01$). Comparing lineage origination over time with a constant rate of diversification reveals a period of notably elevated diversification from ~12-5 Mya (Fig. 4B).

Comparison of AIC and AICc values for each of the six historical biogeographical models with `BioGeoBEARS` suggested that the DIVALIKE model provided the best fit to the data (AICc weight = 0.62). According to this model, the most probable ancestral range for the serrate juniper clade is a combined range of the Western U.S. and northern/central MX (Fig. 5). The ancestral range of the western U.S. serrate junipers was estimated as the western U.S., but the ancestral range of the remaining serrate junipers was estimated as northern/central MX (Fig. 5).

Discussion

Junipers are considered foundational plants throughout arid regions of North America, where they provide habitat and food resources for numerous animal species (Poddar and Lederer, 1982; Gottfried, 1992; Adams, 2014). The serrate juniper clade is endemic and adapted to arid environments of North America, yet lack of phylogenetic resolution has precluded thorough understanding of how geography and climate may have influenced diversification in this relatively young group. Compared with previous work on limited numbers of serrate juniper taxa and Sanger-sequenced cp and nr loci (Mao et al., 2010; Adams and Schwarzbach, 2013b), the phylogenies inferred here with ddRADseq data offer greater resolution and support, and are largely consistent with longstanding taxonomy. Our results provide insight into the evolutionary history of the serrate junipers, including variation in the tempo of diversification, and reveal notable instances of discordance among phylogenies inferred from nuclear and chloroplast variation. Junipers are considered foundational plants throughout arid regions of North America, where they provide habitat and food resources for numerous animal species (Poddar and Lederer, 1982; Gottfried, 1992; Adams, 2014). The serrate juniper clade is endemic and adapted to arid

environments of North America, yet lack of phylogenetic resolution has precluded thorough understanding of how geography and climate may have influenced diversification in this relatively young group. Compared with previous work on limited numbers of serrate juniper taxa and Sanger-sequenced cp and nr loci (Mao et al., 2010; Adams and Schwarzbach, 2013b), the phylogenies inferred here with ddRADseq data offer greater resolution and support, and are largely consistent with longstanding taxonomy. Our results provide insight into the evolutionary history of the serrate junipers, including variation in the tempo of diversification, and reveal notable instances of discordance among phylogenies inferred from nuclear and chloroplast variation.

Diversification history of the serrate leaf margin junipers

Our results are consistent with the hypothesis (Mao et al. 2010) that the ancestral serrate juniper lineage originated during the Oligocene epoch in North America (Fig. 4). During the Eocene-Oligocene transition (~33.9 Mya), decreasing temperatures and increasing seasonality occurred in many regions globally, potentially favoring the expansion of arid-adapted juniper populations (Kennett, 1977; Buchardt, 1978; Wolfe, 1978). As suggested by Mao et al. (2010), the serrate juniper ancestor may have first reached North America via the North Atlantic Land Bridge (NALB) or the Bering Land Bridge (BLB). The NALB, which provided an Atlantic connection through Greenland, was beginning to fragment during the Eocene, but fossil evidence suggests that it continued to facilitate the transatlantic migration of tree species well into the Miocene (Donoghue et al., 2001; Grímsson and Denk, 2005; Denk et al., 2010; Helmstetter et al., 2019). The BLB, which provided a Pacific connection across the Bering Strait, likely facilitated numerous transcontinental migrations during the Cenozoic (Hopkins, 1959, 1967; Donoghue et

al., 2001; Wang and Ran, 2014), including other North American tree genera (e.g., *Fagus* and *Quercus*, Manos and Stanford, 2001; *Hesperocyparis* + *Callitropsis*, Terry et al., 2016; *Pinus*, Badik et al., 2018; *Picea*, Shao et al., 2019).

We inferred a combined ancestral range for the serrate juniper clade which included the western U.S. and northern/central MX (Fig. 5). Two lines of evidence suggest that the common ancestor of the serrate junipers established in the western United States after migrating from Eurasia and potentially before expanding into northern and central Mexico. First, our results generally suggest that the western U.S. clade is basal to all other serrate juniper clades (Figs. 2). Second, the earliest appearances of serrate junipers in the fossil record date to the late Oligocene and early Miocene in the western United States, and feature characteristics similar to extant western U.S. junipers (Axelrod, 1956, 1987, 1991; Wolfe, 1964). During the Oligocene, the western United States was characterized by drier climates, expanding sclerophyll vegetation, and the origin of many contemporary tree species (Axelrod, 1976; Reveal, 1980). Moderate temperatures during this time shifted mixed conifer and subalpine forests coastward (Axelrod, 1976), which, alongside increasingly xeric conditions throughout the region, may have provided ecological opportunity for serrate juniper establishment.

Divergence time estimates suggest that approximately one-third of all divergence events occurred relatively recently in the serrate juniper clade. Elevated diversification rates occurred from approximately 12 to 5 Mya during the late Miocene and early Pliocene (Fig. 4B). Notably, this period coincided with enhanced diversification rates across juniper generally, which was attributed by Mao et al. (2010) to global cooling and uplift of the Qinghai-Tibetan plateau, though the latter is not relevant for North America. In western North America, uplift of the American Cordillera during the late Miocene (~12-5 Mya) induced a rain shadow effect and the

expansion of arid habitats (Axelrod, 1950, 1985; Leopold and Denton, 1987; Wilson and Pitts, 2010), causing population extirpation and the evolution of drought adapted flora (Reveal, 1980). The serrate junipers are particularly tolerant to water stress (Willson et al., 2008) and may have persisted or expanded into newly vacant habitats during this period. Furthermore, increased fire and the expansion of grassland habitat at lower elevations may have restricted junipers to higher elevations, causing range disjunctions between mountain chains and allopatric divergence across altitudinal zones (Retallack, 1997; Wilson and Pitts, 2010). Indeed, some extant sister species exhibit geographical associations with adjacent mountain ranges, with one example being *J. occidentalis* and *J. grandis*, which diverged around the Miocene-Pliocene boundary: *Juniperus occidentalis* inhabits low to intermediate elevations associated with the Cascade range and *J. grandis* occupies mid to high elevation alpine environments associated with the Sierra Nevada range (Terry et al., 2000). Miocene diversification has also been observed in other temperate trees (*Pinus*; Willyard et al., 2007; *Cupressus*; Xu et al., 2010; *Abies*; Aguirre-Planter et al., 2012; *Quercus* section *Lobatae*, series *Agrifoliae*; Hauser et al., 2017) and has been similarly attributed to falling global temperatures and mountain uplift.

Utilizing ddRADseq data to resolve relationships among the serrate junipers

Our analyses were highly consistent across different inference approaches and recapitulated many of the general patterns suggested by previous analyses, including the monophyly of the "one-seeded", "Cerro Petosí", and "*durangensis*" clades (Adams and Schwarzbach, 2013b) and recognition of two *J. deppeana* varieties, var. *gamboana* and var. *deppeana* (Mao et al., 2010; Adams and Schwarzbach, 2013b). However, ddRADseq analyses, based on more extensive genomic sampling, provided enhanced resolution of early divergences

in the serrate juniper clade by consistently recovering three major groups with high support: 1) the western U.S. clade; 2) the *J. ashei* clade, *J. deppeana* species complex, and one-seeded clade [also suggested by Mao et al. (2010)]; and 3) the Cerro Potosí clade, *J. durangensis* clade, the subalpine-alpine clade, *J. flaccida*, and *J. poblana* species complex. Our analyses additionally recovered some relationships which were previously unresolved due to incomplete sampling, predominantly cpDNA-based inference, or analyses being based on limited genomic sampling (e.g., Mao et al., 2010; Adams and Schwarzbach, 2013b). We highlight noteworthy examples of these results below.

Members of the western U.S. clade (*J. occidentalis*, *J. grandis*, *J. osteosperma*, and *J. californica*) are morphologically cohesive (see Vasek, 1966) and occur along a north-south moisture gradient from the montane zone of the eastern Cascade and Sierra Nevada ranges (*J. occidentalis* and *J. grandis*, respectively), through the pinyon-juniper woodlands of the Great Basin and Colorado Plateau (*J. osteosperma*), to the Mojave Desert (*J. californica*). Nonetheless, both Mao et al. (2010) and Adams and Schwarzbach (2013b) inferred paraphyletic placements of *J. californica* relative to other members of the group. In contrast, our analyses inferred *J. californica* as the most basal member of a monophyletic western U.S. clade (Figs. 2, 4A), consistent with previous taxonomic classification. Our analyses additionally resolved relationships among *J. osteosperma*, *J. occidentalis*, and *J. grandis*, which hybridize in western Nevada (Terry et al., 2000; Terry, 2010; Adams, 2013a,b). *Juniperus grandis* and *J. occidentalis* were previously classified as *J. occidentalis* varieties based on morphological similarities which exhibit clinal variation (Vasek, 1966); however, they were not sister to one another in the analysis of Adams and Schwarzbach (2013b). Our analyses assigned them as sister taxa and

placed *J. osteosperma* basal to them (Figs. 2, 4A), consistent with expectations based on morphology and geography.

Juniperus ashei and *J. ovata* (previously *J. ashei* var. *ovata*; Adams and Baker, 2007) hybridize extensively where they occur parapatrically in the trans-Pecos region of Texas, and were considered subspecies until recent phylogenetic analysis merited the recognition of *J. ovata* at the specific level (Adams and Schwarzbach, 2013b). In contrast to Adams and Schwarzbach (2013b), our analyses indicate a sister relationship for *J. ashei* and *J. ovata*, which is supported by morphology and the geographical proximity of these taxa (Figs. 2, 4A). The inference of *J. comitana* as the basal member of this clade (Figs. 2, 4A), however, is not supported by morphology and chemistry (Adams, 2000) and merits additional research.

We included new collections of *J. durangensis* from Sierra Gamon, Durango, in our analyses due to their atypical morphology relative to type localities of *J. durangensis* (Socorro Gonzales, pers., comm.). Our analyses suggest phylogenetic distinctness of *J. durangensis* from Sierra Gamon, despite growing only 150 km northeast of the type locality near El Salto, Durango (Fig. 2). Ongoing morphological and phytochemical analyses may help determine whether *J. durangensis* from Sierra Gamon merits recognition as a new variety. Similarly, new *J. poblana* accessions were analyzed from Nayarit, Oaxaca, and Puebla, as potential cases of intraspecific divergence. The only additional variety suggested by our analyses besides the previously recognized *J. poblana* var. *decurrens* is represented by samples from Oaxaca, which formed a monophyletic group in both the maximum likelihood and SVDquartet analyses (Fig. 2).

In contrast to Adams and Schwarzbach (2013b), the ddRADseq maximum likelihood analysis placed *J. jaliscana*, *J. monticola*, and *J. standleyi* in a highly-supported monophyletic clade, and *J. flaccida* and *J. poblana* in a distinct sister clade with low support (Fig. 2 left). The

SVDquartets and Bayesian trees likewise indicate monophyly of *J. jaliscana*, *J. monticola*, and *J. standleyi*, but with lower support (Figs. 2 right, 4A). We refer to this group as the "subalpine-alpine clade" because they occur at mid-high elevations. *Juniperus monticola* is widespread in Mexico and occupies subalpine and alpine habitats at elevations of 2400-4500 m (Adams, 2014), while *J. jaliscana* occupies pine-oak forests at elevations of 1335-2670 in southern Durango and northwest Jalisco (Zanoni and Adams, 1979). *Juniperus standleyi* is found in extreme southeast Mexico and Guatemala at elevations of 3000-4250 m (Adams, 2014). Phylogenetically adjacent taxa, *J. flaccida* and *J. poblana*, likewise occur in subalpine habitats but are distinguished morphologically from the subalpine-alpine clade by branches which are flaccid at the tips so that their foliage appears to be drooping (Adams, 2014).

The relationship between *J. flaccida* and *J. poblana* (previously *J. flaccida* var. *poblana*) has been taxonomically challenging due to the paucity of distinguishing morphological features and their ability to hybridize (Zanoni and Adams, 1976; Adams et al., 2018c). Our analyses suggest a distinct taxonomic status for *J. poblana*, but disagree on the relationship between *J. poblana* and *J. flaccida*. Consistent with taxonomic expectations, maximum likelihood and Bayesian phylogenies support a sister relationship between *J. flaccida* and *J. poblana* (although poorly supported in the former) (Figs. 2 left, 4A); however, the SVDquartets tree suggests a more distant placement of *J. flaccida* basal to the subalpine-alpine clade (Fig. 2 right). An affinity of *J. flaccida* towards the subalpine-alpine clade was suggested by the Adams and Schwarzbach (2013b) phylogeny, which recovered a sister relationship between *J. flaccida* and *J. standleyi*. A potential explanation for this, and for conflicting phylogenetic signal in the ddRADseq data, could be introgression from *J. standleyi* into *J. flaccida*.

While the maximum likelihood and SVDquartets analyses produced predominantly consistent results, there were three instances of discordance which highlight areas where gene tree variation may have influenced inference (Maddison, 1997; Huang et al., 2010; Tonini et al., 2015). As incomplete lineage sorting (ILS) is a major source of gene tree-species tree discordance, phylogenetic inference under the multi-species coalescent (e.g., SVDquartets) may perform more accurately under high ILS conditions compared with concatenation approaches (e.g., RAxML) (Chou et al., 2015). Shallow divergences may be especially prone to ILS, which may explain the discordance between the *J. ashei* clade, the *J. deppeana* complex, and the one-seeded clade (Figs. 2, 4A). Alternatively, hybridization is widely reported throughout *Juniperus* (e.g., Adams, 1994; Terry et al. 2000; Adams et al., 2020) and may have contributed to topological discordance in areas of low support, e.g., the relationship of *J. flaccida* (Fig. 2). Finally, allelic dropout in reduced-representation data may complicate the resolution of older splits, and may have played a role in the discordance observed among two outgroup samples, *J. communis* and *J. drupacea* (Fig. 2). Overall, differences in model assumptions and conflicting phylogenetic signal likely influenced the few points of discordance observed among our different inference methods.

Discordance between phylogenies inferred with nuclear and chloroplast DNA

Discordance among nr and cpDNA is common, and can arise from processes including incomplete lineage sorting (Degnan and Rosenburg, 2009), hybridization (Rieseberg and Soltis, 1991; Rieseberg et al., 1996), and lateral transfer of organellar genomes (Stegemann et al., 2012). In angiosperms prone to hybridization, discordance among nr and cpDNA gene trees has often been attributed to introgression and chloroplast capture (e.g., Acosta and Premoli, 2010;

Lee-Yaw et al., 2019; Liu et al., 2020). When maternally inherited in angiosperms, cpDNA exhibits more intraspecific population divergence and higher introgression across species boundaries than nrDNA (Petit and Excoffier, 2009; Du et al., 2009). However, conifer cpDNA is usually paternally inherited through pollen (Neale and Sederoff, 1989; Mogensen, 1996), typically exhibits weaker population differentiation than nr or mtDNA, and is expected to move less readily across species boundaries (e.g., Petit et al., 2005; Gerardi et al., 2010; Godbout et al., 2010). Thus, chloroplast introgression should generally be less likely in conifers, although potential examples of chloroplast introgression and capture have been described (e.g., Liston et al., 2007; Gernandt et al., 2018). Interestingly, theoretical work suggests cp capture may be driven by mitochondrial based cytoplasmic male sterility (Frank, 1989) in hybridizing angiosperms with maternal co-inheritance of mt and cp genomes (Tsitrone et al., 2003). This mechanism couldn't operate in most conifers (e.g., *Picea* and *Pinus*) which inherit mt (maternal) and cp (paternal) genomes separately. However, Cupressaceae (including *Juniperus*) have paternal inheritance of both mt and cp genomes (Mogensen, 1996; Adams, 2019), which could increase the probability of chloroplast capture via cytoplasmic interactions (Tsitrone et al., 2003). Alternatively, lateral transfer of chloroplast through natural grafting during periods of sympatry could lead to apparent chloroplast capture in the absence of hybridization (Stegeman et al., 2012).

As in other conifers (Petit and Hampe, 2006), reproductive isolation is often weak among *Juniperus*, and hybridization has been documented among serrate juniper species including *J. occidentalis* and *J. osteosperma* (Terry et al. 2000; Terry 2010), *J. ashei* and *J. ovata* (Adams et al., 2020), and *J. angosturana* and *J. coahuilensis* (Adams, 1994). Potential cases of introgression or horizontal transfer of cpDNA have also been noted in the group (Adams, 2016;

Adams et al., 2016, 2017). For example, *J. occidentalis* and *J. osteosperma* hybridize extensively in northwestern Nevada, and a cpDNA haplotype fixed in *J. occidentalis* appears to have introgressed through the western range of *J. osteosperma* (Terry et al., 2000, Terry, 2010). A potential case of chloroplast capture occurred in the closely related smooth leaf juniper clade (Fig. 3 right) between *J. thurifera* (chloroplast donor, not shown) and *J. sabina* var. *sabina* (chloroplast recipient), giving rise to the allotetrapoloid *J. sabina* var. *balkanensis* (Adams et al., 2016, 2018a,b; Farhat et al., 2019). The cpDNA tree indicates notable discordance consistent with this idea, placing *J. sabina* var. *balkanensis* in a clade with *J. virginiana* (Fig. 3 right), while ddRADseq analyses inferred the expected monophyletic relationship for the *J. sabina* varieties (Figs. 2, 3 left). As the ddRADseq phylogenies are congruent with taxonomic expectations based on morphology and geography, several strong instances of discordance in the cpDNA phylogeny suggest the potential for chloroplast introgression or transfer, although incomplete lineage sorting remains plausible for several of these cases.

Clear instances of discordance involve species from diverged lineages inferred with nuclear data that unexpectedly share cpDNA variation (Fig. 3). ddRADseq data inferred a western U.S. clade containing *J. californica* (Fig. 3 left), as expected based on morphology and geography; however, the cpDNA tree placed *J. californica* in a well-supported clade with *J. comitana*, which is restricted to southern Mexico/northern Guatemala (Fig. 3 right). Introgression or transfer of a *J. comitana*-type chloroplast from an ancestral *J. comitana* lineage into *J. californica* could underly such discordance (Fig. 3 right). Second, cpDNA placed *J. zanonii*, a sub-alpine plant that grows at the 3550 m summit of Cerro Potosí, NL, Mexico, within a clade with *J. ashei* and *J. ovata*, sibling species that grow on limestone in Central Texas (Adams, 2008) (Fig. 3 right). The *ashei* clade is substantially diverged from *J. zanonii* in

ddRADseq analyses, which placed *J. zanonii* with *J. saltillensis* (Fig. 3 left), consistent with *J. zanonii* and *J. saltillensis* exhibiting altitudinal zonation at Cerro Petosí, Mexico. This discordance could have arisen from chloroplast introgression or transfer from an ancestral *J. ovata*/*J. ashei* into ancestral *J. zanonii*, as these lineages likely experienced sympatry during the Pleistocene (Adams and Baker, 2007). Third, *J. arizonica* and *J. coahuilensis* occur parapatrically, but the two taxa are highly similar morphologically and hybridize in the Trans-Pecos, Texas region (Adams, 2014, 2017). ddRADseq analyses placed *J. arizonica* in the one-seeded group with *J. coahuilensis* (Fig. 3 left), as expected, while cpDNA placed *J. arizonica* within the *J. ashei* clade (Fig. 3 right). Chloroplast introgression or transfer from *J. ashei* to *J. arizonica* could underly such discordance (Fig. 3 right), although incomplete lineage sorting is also possible for these closely related clades. These discordances suggest that nr and cpDNA histories can vary prominently in *Juniperus*, and while evidence for chloroplast capture or horizontal transfer is scarce in conifers, these processes may deserve further study in Cupressaceae.

Conclusion

Our analyses of ddRADseq data produced highly resolved and largely consistent phylogenies depicting the evolutionary history of the serrate junipers of western North America. While these phylogenies were strongly consistent with taxonomic expectations based on morphology and ecology, cpDNA phylogenies illustrated several pronounced cases of discordance, suggesting the potential for processes to differentially influence the evolutionary history of the chloroplast genome. An improved understanding of the timing and tempo of diversification, including the age of origin of the serrate juniper clade and its elevated rate of

diversification during the late Miocene, illustrates how the interaction between geologic, geographic, and climatic processes may have influenced patterns of diversification in this group. This study contributes to a growing body of research demonstrating the effectiveness of reduced-representation sequencing data for resolving the phylogenies of non-model organisms (e.g., Eaton and Ree, 2013; Herrera and Shank, 2016; Massatti et al., 2016; Eaton et al., 2017; Near et al., 2018; Paetzold et al., 2019) and the complex evolutionary histories of western North American taxa characterized by reticulate evolution and recent divergence.

Acknowledgements

We would like to acknowledge María Socorro González-Elizondo, Jim Bartel, Lucio Caamaño Onofre, Allen Coombes, and Alex Tashev for field assistance and sample collection. We are very grateful to Lauren Im, who provided technical support in ArcGIS. We would like to acknowledge and thank Michael May for his invaluable support utilizing RevBayes. We also thank Joshua Jahner and Matthew Forister for their valuable comments on the manuscript. This work was funded by Baylor University (Project No. 032512) to R.P.A., a National Science Foundation Graduate Research Fellowship Award to K.A.U. (Award No. 1650114), and a 2018 EECG award from the American Genetics Association to K.A.U. The authors declare no conflicts of interests.

References

Acosta, M.C., Premoli, A.C., 2010. Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). Mol. Phylogenet. Evol. 54 (1), 235-242, https://doi.org/10.1016/j.ympev.2009.08.008

Adams, R.P., 1994. Geographic variation and systematics of monospermous *Juniperus* (Cupressaceae) from the Chihuahua Desert based on RAPDs and terpenes. Biochem. Syst. Ecol. 22 (7), 699-710, https://doi.org/10.1016/0305-1978(94)90056-6

Adams, R.P., 2000. The serrate leaf margined *Juniperus* (Section Sabina) of the western hemisphere: systematics and evolution based on leaf essential oils and Random Amplified Polymorphic DNAs (RAPDs). Biochem. Syst. Ecol. 28 (10), 975-989, https://doi.org/10.1016/S0305-1978(00)00022-3

Adams, R.P., 2008. Distribution of *Juniperus ashei* var. *ashei* and var. *ovata* around New Braunfels, Texas. Phytologia. 90 (1), 97-102.

Adams, R.P., 2013a. Hybridization between *Juniperus grandis*, *J. occidentalis* and *J. osteosperma* in northwest Nevada I: Terpenes, Leviathan mine, Nevada. Phytologia. 95 (1), 58-69.

Adams, R.P., 2013b. Hybridization between *Juniperus grandis*, *J. occidentalis* and *J. osteosperma* in northwest Nevada II: Terpenes, Buffalo Hills, Northwestern Nevada. Phytologia. 95 (1), 107-114.

Adams, R.P., 2014. Junipers of the World: The genus *Juniperus*. Trafford Publishing, Bloomington.

Adams, R.P., 2016. Two new cases of chloroplast capture in incongruent topologies in the *Juniperus excelsa* complex: *J. excelsa* var. *turcomanica* comb. nov. and *J. excelsa* var. *seravschanica* comb. nov. Phytologia. 98 (3), 219-231.

Adams, R.P., 2017. Multiple evidences of past evolution are hidden in nrDNA of *Juniperus arizonica* and *J. coahuilensis* populations in the trans-Pecos, Texas region. Phytologia. 99 (1), 38-47.

Adams, R.P., 2019. Inheritance of chloroplasts and mitochondria in Conifers: A review of paternal, maternal, leakage and facultative inheritance. Phytologia. 101 (2), 134-138.

Adams, R.P., Baker, L.E., 2007. Pleistocene infraspecific evolution in *Juniperus ashei* Buch. Phytologia. 89 (1), 8-23.

Adams, R.P., Johnson, S.T., Schwarzbach, A.E., 2020. Long distance gene flow facilitated by bird-dispersed seeds in wind-pollinated species: A story of hybridization and introgression between *Juniperus ashei* and *J. ovata* told by nrDNA and cpDNA. Phytologia. 102 (2), 55-74.

Adams, R.P., Boratynski, A., Marcysiak, K., Roma-Marzio, F., Peruzzi, L., Bartolucci, F., Conti, F., Mataraci, T., Tashev, A.N., Siljak-Yakovlev, S., 2018a. Discovery of *Juniperus sabina* var. *balkanensis* R. P. Adams & Tashev in Macedonia, Bosnia-Herzegovina, Croatia and southern Italy and relictual polymoal. rphisms found in nrDNA. Phytologia. 100 (2), 117-127.

Adams, R.P., Farhat, P., Shuka, L., Silak-Yakovlev, S., 2018b. Discovery of *Juniperus sabina* var. *balkanensis* R. P. Adams and A. N. Tashev in Albania and relictual polymorphisms found in nrDNA. Phytologia. 100 (3), 187-194.

Adams, R.P., Johnson, S., Coombes, A.J., Caamaño, L., González-Elizondo, M.S., 2018c. Preliminary examination of hybridization and introgression between *Juniperus flaccida* and *J. poblana*: nrDNA and cpDNA sequence data. Phytologia. 100 (2), 145-152.

Adams, R.P., Schwarzbach A.E., 2013a. Phylogeny of *Juniperus* using nrDNA and four cpDNA regions. Phytologia. 95 (2), 179-187.

Adams, R. P., Schwarzbach, A.E., 2013b. Taxonomy of the serrate leaf *Juniperus* of North America: Phylogenetic analyses using nrDNA and four cpDNA regions. Phytologia. 95 (2), 172-178.

Adams, R.P., Schwarzbach, A.E., Tashev, A.N., 2016. Chloroplast capture by a new variety, *Juniperus sabina* var. *balkanensis* R. P. Adams and A. N. Tashev, from the Balkan peninsula: A putative stabilized relictual hybrid between *J. sabina* and ancestral *J. thurifera.* Phytologia. 98 (2), 100-111.

Adams, R.P., Socorro González-Elizondo M., González-Elizondo M., Ramirez Noy D., Schwarzbach A.E., 2017. DNA sequencing and taxonomy of unusual serrate *Juniperus* from Mexico: Chloroplast capture and incomplete lineage sorting in *J. coahuilensis* and allied taxa. Phytologia. 99 (1), 62-73.

Aguirre-Planter, É., Jaramillo-Correa, J.P., Gómez-Acevedo, S., Khasa, D.P., Bousquet, J., Eguiarte, L.E., 2012. Phylogeny, diversification rates and species boundaries of Mesoamerican firs (*Abies*, Pinaceae) in a genus-wide context. Mol. Phylogenet. Evol. 62 (1), 263-274, https://doi.org/10.1016/j.ympev.2011.09.021

Alexander, A.M., Su, Y.C., Oliveros, C.H., Olson, K.V., Travers, S.L., Brown, R.M., 2017. Genomic data reveals potential for hybridization, introgression, and incomplete lineage

sorting to confound phylogenetic relationships in an adaptive radiation of narrow-mouth frogs. Evolution. 71 (2), 475-488, https://doi.org/10.1111/evo.13133

Allio, R., Scornavacca, C., Nabholz, B., Clamens, A.L., Sperling, F.A., Condamine, F.L., 2020. Whole genome shotgun phylogenomics resolves the pattern and timing of swallowtail butterfly evolution. Syst. Biol. 69 (1), 38-60, https://doi.org/10.1093/sysbio/syz030

Axelrod, D.I., 1948. Climate and evolution in western North America during middle Pliocene time. Evolution. 2, 127-144, https://doi.org/10.2307/2405373

Axelrod, D.I., 1950. Evolution of desert vegetation in western North America. Carnegie Inst. Wash. Publ. 590, 215-306.

Axelrod, D.I., 1956. Mio-Pliocene floras from west-central Nevada. Vol. 33. University of California Press.

Axelrod, D.I., 1976. History of the coniferous forests, California and Nevada. University of California Press, Berkeley.

Axelrod, D.I., 1985. Rise of the grassland biome, central North America. Bot. Rev. 51 (2), 163-201.

Axelrod, D.I., 1987. The late Oligocene Creede flora, Colorado. Vol. 130. University of California Press.

Axelrod, D.I., 1991. The Early Miocene buffalo canyon flora of western Nevada. Vol. 135. University of California Press, Berkeley.

Badik, K.J., Jahner, J.P., Wilson, J.S., 2018. A biogeographic perspective on the evolution of fire syndromes in pine trees (*Pinus*: Pinaceae). R. Soc. Open Sci. 5 (3), 172412, http://dx.doi.org/10.1098/rsos.172412

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3 (10), e3376, http://doi.org/10.1371/journal.pone.0003376

Bouillé, M., Senneville, S., Bousquet, J., 2011. Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. Tree Genet. Genomes. 7 (3), 469-484, http://doi.org/10.1007/s11295-010-0349-z

Bravo, G.A., Antonelli, A., Bacon, C.D., Bartoszek, K., Blom, M.P., Huynh, S., Jones, G., Knowles, L.L., Lamichhaney, S., Marcussen, T. Morlon, H., 2019. Embracing heterogeneity: coalescing the Tree of Life and the future of phylogenomics. PeerJ. 7, e6399.

Buchardt, B., 1978. Oxygen isotope palaeotemperatures from the Tertiary period in the North Sea area. Nature. 275 (5676), 121-123.

Calsbeek, R., Thompson, J.N., Richardson, J.E., 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Mol. Ecol. 12 (4), 1021-1029, https://doi.org/10.1046/j.1365-294X.2003.01794.x

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3 (4), 846-852, https://doi.org/10.1002/ece3.512

Carter, K.A., Liston, A., Bassil, N.V., Alice, L.A., Bushakra, J.M., Sutherland, B.L., Mockler, T.C., Bryant, D.W., Hummer, K.E., 2019. Target capture sequencing unravels *Rubus* evolution. Front. Plant Sci. 10, 1615, https://doi.org/10.3389/fpls.2019.01615

Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics. 30 (23), 3317-3324, https://doi.org/10.1093/bioinformatics/btu530

Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. J. Theor. Biol. 374, 35-47, https://doi.org/10.1016/j.jtbi.2015.03.006

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics, 16 (10), 1-11

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24 (6), 332-340, https://doi.org/10.1016/j.tree.2009.01.009

de La Harpe, M., Hess, J., Loiseau, O., Salamin, N., Lexer, C., Paris, M., 2019. A dedicated target capture approach reveals variable genetic markers across micro- and macro-evolutionary time scales in palms. Mol. Ecol. Resour. 19 (1), 221-234, https://doi.org/10.1111/1755-0998.12945

Denk, T., Grímsson F., Zetter, R., 2010. Episodic migration of oaks to Iceland: Evidence for a North Atlantic “land bridge” in the latest Miocene. Am. J. Bot. 97 (2), 276-287, https://doi.org/10.3732/ajb.0900195

Donoghue, M.J., Bell, C.D., Li, J., 2001. Phylogenetic patterns in Northern Hemisphere plant geography. Int. J. Plant Sci. 162 (S6), S41-S52, https://doi.org/10.1086/323278

Du, F.K., Petit, R.J., Liu, J.Q., 2009. More introgression with less gene flow: chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other conifers. Mol. Ecol. 18 (7), 1396-1407, https://doi.org/10.1111/j.1365-294X.2009.04107.x

Du, Z.Y., Harris, A.J., Xiang, Q.Y.J., 2020. Phylogenomics, co-evolution of ecological niche and morphology, and historical biogeography of buckeyes, horsechestnuts, and their relatives (Hippocastaneae, Sapindaceae) and the value of RAD-seq for deep evolutionary inferences back to the Late Cretaceous. Mol. Phylogenet. Evol. 145, 106726, https://doi.org/10.1016/j.ympev.2019.106726

Eaton, D.A.R., 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics. 30 (13), 1844-1849, https://doi.org/10.1093/bioinformatics/btu121

Eaton, D.A.R, Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Bio. 62 (5), 689-706, https://doi.org/10.1093/sysbio/syt032

Eaton, D.A., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Bio. 66 (3), 399-412, https://doi.org/10.1093/sysbio/syw092

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792-1797, https://doi.org/10.1093/nar/gkh340

Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PloS One. 8 (6), e65923, https://dx.doi.org/10.1371%2Fjournal.pone.0065923

Farhat, P., Siljak-Yakovlev, S., Adams, R.P., Robert, T., Dagher-Kharrat, M.B., 2019. Genome size variation and polyploidy in the geographical range of *Juniperus sabina* L. (Cupressaceae). Bot. Lett. (in press), https://doi.org/10.1080/23818107.2019.1613262

Farjon, A., 2005. Monograph of Cupressaceae and *Sciadopitys*. Royal Botanic Gardens, Kew, Richmond.

Frank, S.A., 1989. The evolutionary dynamics of cytoplasmic male sterility. Am. Nat. 133 (3), 345-376, https://doi.org/10.1086/284923

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7 (4), 457-472, https://www-jstor-org.unr.idm.oclc.org/stable/2246093

Gerardi, S., Jaramillo-Correa, J.P., Beaulieu, J. and Bousquet, J., 2010. From glacial refugia to modern populations: new assemblages of organelle genomes generated by differential cytoplasmic gene flow in transcontinental black spruce. Mol. Ecol. 19 (23), 5265-5280, https://doi.org/10.1111/j.1365-294X.2010.04881.x

Gernandt, D.S., Aguirre Dugua, X., Vázquez-Lobo, A., Willyard, A., Moreno Letelier, A., Pérez de la Rosa, J.A., Piñero, D., Liston, A., 2018. Multi-locus phylogenetics, lineage sorting, and reticulation in *Pinus* subsection *Australes*. Am. J. Bot. 105(4), 711-725, https://doi.org/10.1002/ajb2.1052

Gernandt, D.S., López, G.G., García, S.O., Liston, A., 2005. Phylogeny and classification of *Pinus*. Taxon. 54 (1), 29-42, https://doi.org/10.2307/25065300

Godbout, J., Beaulieu, J. and Bousquet, J., 2010. Phylogeographic structure of jack pine (*Pinus banksiana*; Pinaceae) supports the existence of a coastal glacial refugium in northeastern North America. Am. J. Bot. 97 (11), 1903-1912, https://doi.org/10.3732/ajb.1000148

Gottfried, G.J., 1992. Ecology and management of the southwestern pinyon-juniper woodlands. In: Ffolliott, P.F., Gottfried, G.J., Bennett, D.A., Hernandez, V.M., Ortega-Rubio, C.A., Hamre, R.H. [Tech Coords.], Ecology and management of oak and associated woodlands: perspectives in the southwestern United States and northern Mexico. United States Department of Agriculture, Forest Service, Washington, D.C., pp. 78-85.

Graur, D., Martin, W., 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20 (2), 80-86.

Grímsson, F., Denk, T., 2005. *Fagus* from the Miocene of Iceland: systematics and biogeographical considerations. Rev. Palaeobot. Palynol. 134 (1-2), 27-54, https://doi.org/10.1016/j.revpalbo.2004.11.002

Hauser, D.A., Keuter, A., McVay, J.D., Hipp, A.L., Manos, P.S., 2017. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). Am. J. Bot. 104 (10), 1581-1595, https://doi.org/10.3732/ajb.1700291

Helmstetter, A.J., Buggs, R.J., Lucas, S.J., 2019. Repeated long-distance dispersal and convergent evolution in hazel. Sci. Rep. 9 (1), 1-12, https://doi.org/10.1038/s41598-019-52403-2

Herrera, S., Shank, T.M., 2016. RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol. Phylogenet. Evol. 100, 70-79, https://doi.org/10.1007/s10709-011-9547-3

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58 (3), 247-276, https://doi.org/10.1111/j.1095-8312.1996.tb01434.x

Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 359 (1442), 183-195, https://doi.org/10.1098/rstb.2003.1388

Hewitt, G.M., 2011. Quaternary phylogeography: the roots of hybrid zones. Genetica. 139 (5), 617-638.

Hipp, A.L., Manos, P.S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A.A., Deng, M., Denk, T., Fitz-Gibbon, S., Gailing, O., 2020. Genomic landscape of the global oak phylogeny. New Phytol. 226 (4), 1198-1212, https://doi.org/10.1111/nph.16162

Höhna, S., 2015. The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. J. Theor. Biol. 380, 321-331, https://doi.org/10.1016/j.jtbi.2015.06.005

Höhna, S., Landis, M.J., Heath, T.A., 2017. Phylogenetic inference using RevBayes. Curr. Protoc. Bioinform. 57 (6.16), 1-34, https://doi.org/10.1002/cpbi.22

Hopkins, D.M., 1967. The Bering land bridge. Vol. 3. Stanford University Press, Palo Alto.

Hopkins, D.M., 1959. Cenozoic history of the Bering land bridge. Science. 129 (3362), 1519-1528, https://www.jstor.org/stable/1757656

Huang, H., He, Q., Kubatko, L.S., Knowles, L.L., 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different methods. Syst. Biol. 59 (5), 573-583, https://doi.org/10.1093/sysbio/syq047

Huang, H., Knowles L.L., 2014. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst. Biol. 65 (3), 357-365, https://doi.org/10.1093/sysbio/syu046

Karimi, N., Grover, C.E., Gallagher, J.P., Wendel, J.F., Ané, C., Baum, D.A., 2020. Reticulate evolution helps explain apparent homoplasy in floral biology and pollination in baobabs (*Adansonia*; Bombacoideae; Malvaceae). Syst. Biol. 69 (3), 462-478, https://doi.org/10.1093/sysbio/syz073

Kendall, D.G., 1948. On the generalized "birth-and-death" process. Ann. Math. Stat. 19 (1), 1-15, https://www.jstor.org/stable/2236051

Kennett, J.P., 1977. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic Ocean, and their impact on global paleoceanography. J. Geophys. Res. 82 (27), 3843-3860, https://doi.org/10.1029/JC082i027p03843

Kimball, R.T., Oliveros, C.H., Wang, N., White, N.D., Barker, F.K., Field, D.J., Ksepka, D.T., Chesser, R.T., Moyle, R.G., Braun, M.J., Brumfield, R.T., 2019. A phylogenomic supertree of birds. Diversity, 11 (7), 109, https://doi.org/10.3390/d11070109

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst. Biol. 56 (1), 17-24, https://doi.org/10.1080/10635150601146041

Kuzoff, R.K., Soltis, D.E., Hufford, L., Soltis, P.S., 1999. Phylogenetic relationships within *Lithophragma* (Saxifragaceae): hybridization, allopolyploidy, and ovary diversification. Syst. Bot. 24 (4), 598-615, https://doi.org/10.2307/2419645

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods. 9 (4), 357, http://www.nature.com/doifinder/10.1038/nmeth.1923

Leaché, A.D., Oaks, J.R., 2017. The utility of single nucleotide polymorphism (SNP) data in phylogenetics. Annu. Rev. Ecol. Evol. Syst. 48, 69-84, https://doi.org/10.1146/annurev-ecolsys-110316-022645

Lecaudey, L.A., Schliewen, U.K., Osinov, A.G., Taylor, E.B., Bernatchez, L., Weiss, S.J., 2018. Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. Mol. Phylogenet. Evol. 124, 82-99, https://doi.org/10.1016/j.ympev.2018.02.022

Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., et al., 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature. 574, 679-685, https://doi.org/10.1038/s41586-019-1693-2

Lee-Yaw, J.A., Grassa, C.J., Joly, S., Andrew, R.L., Rieseberg, L.H., 2019. An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). New Phytol. 221 (1), 515-526, https://doi.org/10.1111/nph.15386

Lemmon, E.M., Lemmon A.R., 2013. High-throughput genomic data in systematics and phylogenetics. Annu. Rev. Ecol. Evol. Syst. 44, 99-121, https://doi.org/10.1146/annurev-ecolsys-110512-135822

Leopold, E.B., Denton, M.F., 1987. Comparative age of grassland and steppe east and west of the northern Rocky Mountains. Ann. Mo. Bot. Gard. 74, 841-867, https://doi.org/10.2307/2399452

Léveillé-Bourret, É., Chen, B.H., Garon-Labrecque, M.É., Ford, B.A., Starr, J.R., 2020. RAD sequencing resolves the phylogeny, taxonomy and biogeography of Trichophoreae despite a recent rapid radiation (Cyperaceae). Mol. Phylogenet. Evol. 145, 106727, https://doi.org/10.1016/j.ympev.2019.106727

Liston, A., Parker-Defeniks, M., Syring, J.V., Willyard, A. and Cronn, R., 2007. Interspecific phylogenetic analysis enhances intraspecific phylogeographical inference: a case study in *Pinus lambertiana*. Mol. Ecol. 16 (18), 3926-3937, https://doi.org/10.1111/j.1365-294X.2007.03461.x

Little Jr, E.L., 1971. Atlas of United States trees. Volume 1. Conifers and important hardwoods. Miscellaneous publication 1146. US Department of Agriculture, Forest Service, Washington, DC.

Liu, B.B., Campbell, C.S., Hong, D.Y., Wen, J., 2020. Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. Mol. Phylogenet. Evol. 147, 106784, https://doi.org/10.1016/j.ympev.2020.106784

Liu, Y., Johnson, M.G., Cox, C.J., Medina, R., Devos, N., Vanderpoorten, A., Hedenäs, L., Bell, N.E., Shevock, J.R., Aguero, B., Quandt, D., 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. Nat. Commun. 10 (1), 1-11, https://doi.org/10.1038/s41467-019-09454-w

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46 (3), 523-536, https://doi.org/10.1093/sysbio/46.3.523

Manos, P.S., Stanford, A.M., 2001. The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int. J. Plant Sci. 162 (S6), S77-S93, https://doi.org/10.1086/323280

Mao, K., Hao, G., Liu, J., Adams, R.P., Milne, R.I., 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. New Phytol. 188 (1), 254-272, https://doi.org/10.1111/j.1469-8137.2010.03351.x

Mao, K., Milne, R.I., Zhang, L., Peng, Y., Liu, J., Thomas, P., Mill, R.R., Renner, S.S., 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc. Natl. Acad. Sci. U.S.A. 109 (20), 7793-7798.

Massatti, R., Reznicek, A.A., Knowles, L.L., 2016. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. Am. J. Bot. 103 (2), 337-347, https://doi.org/10.3732/ajb.1500315

Matzke, N.J., 2013a. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. University of California, Berkeley, Berkeley, CA.

Matzke, N.J., 2013b. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5 (4), 242-248.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63 (6), 951-970.

Matzke, N.J., 2018. cladoRcpp v0.15.1: C++ Implementations of Phylogenetic Cladogenesis Calculations. University of Auckland, New Zealand.

Matzke, N.J., Sidje, R., Schmidt, D., 2019. rexpokit v0.26.6.6: R wrappers for EXPOKIT; other matrix functions. School of Biological Sciences, University of Auckland, New Zealand.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol. Phylogenet. Evol. 66 (2), 526-538, https://doi.org/10.1016/j.ympev.2011.12.007

McVay, J.D., Hauser, D., Hipp, A.L., Manos, P.S., 2017. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. Genome. 60 (9), 733-742, https://doi.org/10.1139/gen-2016-0206

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17 (2), 240-248, https://doi.org/10.1101/gr.5681207

Miller, R.F., Wigand, P.E., 1994. Holocene changes in semiarid pinyon-juniper woodlands: response to climate, fire, and human activities in the US Great Basin. BioScience. 44 (7), 465-474, https://doi.org/10.2307/1312298

Mogensen, H.L., 1996. Invited special paper: the hows and whys of cytoplasmic inheritance in seed plants. Am. J. Bot. 83 (3), 383-404, https://doi.org/10.1002/j.1537-2197.1996.tb12718.x

Moura, A.E., Shreves, K., Pilot, M., Andrews, K.R., Moore, D.M., Kishida, T., Möller, L., Natoli, A., Gaspari, S., McGowen, M., Gray, H., 2020. Phylogenomics of the genus *Tursiops* and closely related Delphininae reveals extensive reticulation among lineages and provides inference about eco-evolutionary drivers. Mol. Phylogenet. Evol. 146, 106756.

Neale, D.B., Sederoff, R.R., 1989. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. Theor. Appl. Genet. 77 (2), 212-216.

Near, T.J., MacGuigan, D.J., Parker, E., Struthers, C.D., Jones, C.D., Dornburg, A., 2018. Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. Mol. Phylogenet. Evol. 129, 268-279, https://doi.org/10.1016/j.ympev.2018.09.001

Nee, S., May, R.M., Harvey, P.H., 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 344 (1309), 305-311, https://doi.org/10.1098/rstb.1994.0068

Nevill, P.G., Zhong, X., Tonti-Filippini, J., Byrne, M., Hislop, M., Thiele, K., van Leeuwen, S., Boykin, L.M., Small, I., 2020. Large scale genome skimming from herbarium material for accurate plant identification and phylogenomics. Plant Methods. 16 (1), 1-8, https://doi.org/10.1186/s13007-019-0534-5

Paetzold, C., Wood, K.R., Eaton, D., Wagner, W.L., Appelhans, M.S., 2019. Phylogeny of Hawaiian *Melicope* (Rutaceae): RAD-Seq resolves species relationships and reveals ancient introgression. Front. Plant Sci. 10, 1074, https://doi.org/10.3389/fpls.2019.01074

Paradis, E., 1998. Testing for constant diversification rates using molecular phylogenies: a general approach based on statistical tests for goodness of fit. Mol. Biol. Evol. 15 (4), 476-479, https://doi.org/10.1093/oxfordjournals.molbev.a025946

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35 (3), 526-528, https://doi.org/10.1093/bioinformatics/bty633

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W., Buerkle, C.A., 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol. Ecol. 21 (12), 2991-3005, https://doi.org/10.1111/j.1365-294X.2012.05513.x

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PloS One. 7 (5), e37135, https://dx.doi.org/10.1371%2Fjournal.pone.0037135

Petit, R.J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., Vendramin, G.G., 2005. Invited review: comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol. Ecol. 14(3), 689-701, https://doi.org/10.1111/j.1365-294X.2004.02410.x

Petit, R.J., Excoffier, L., 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24 (7), 386-393, https://doi.org/10.1016/j.tree.2009.02.011

Petit, R.J., Hampe, A., 2006. Some evolutionary consequences of being a tree. Annu. Rev. Ecol. Evol. Syst. 37, 187-214, https://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Phillips, F.J., 1910. The dissemination of junipers by birds. J. For. 8 (1), 60-73.

Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. R news, 6 (1), 7-11.

Poddar, S., Lederer, R.J., 1982. Juniper berries as an exclusive winter forage for Townsend's Solitaires. Am. Midl. Nat. 108 (1), 34-40, https://doi.org/10.2307/2425289

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics, 14 (9), 817-818, https://doi.org/10.1093/bioinformatics/14.9.817

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M.R., Elmer, K.R., Vences, M., Steinfarz, S., 2019. Phylogeny and species delimitation of near Eastern *Neurergus* newts (Salamandridae) based on genome-wide RADseq data analysis. Mol. Phylogenet. Evol. 133, 189-197, https://doi.org/10.1016/j.ympev.2019.01.003

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67 (5), 901, https://dx.doi.org/10.1093%2Fsysbio%2Fsyy032

Razkin, O., Sonet, G., Breugelmans, K., Madeira, M.J., Gómez-Moliner, B.J., Backeljau, T., 2016. Species limits, interspecific hybridization and phylogeny in the cryptic land snail complex *Pyramidula*: the power of RADseq data. Mol. Phylogenet. Evol. 101, 267-278, https://doi.org/10.1016/j.ympev.2016.05.002

Retallack, G.J., 1997. Neogene expansion of the North American prairie. Palaios. 12 (4), 380-390, https://doi.org/10.2307/3515337

Reveal, J.L., 1980. Intermountain biogeography—a speculative appraisal. Mentzelia. 4, 1-92.

Rieseberg, L.H., Beckstrom-Sternberg, S.M., Liston, A., Arias, D.M., 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). Syst. Bot. 50-76, https://doi.org/10.2307/2418973

Rieseberg, L.H., Soltis, D. E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants. 5, 65-84.

Rieseberg, L.H., Whitton, J., Linder, C.R., 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Bot. Neerl. 45 (3), 243-262.

Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theor. Popul. Biol. 100, 56-62, https://doi.org/10.1016/j.tpb.2014.12.005

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4, e2584, https://dx.doi.org/10.7717%2Fpeerj.2584

Romme, W.H., Allen, C.D., Bailey, J.D., Baker, W.L., Bestelmeyer, B.T., Brown, P.M., Eisenhart, K.S., Floyd, M.L., Huffman, D.W., Jacobs, B.F., Miller, R.F., Muldavin, E.H., Swetnam, T.W., Tausch, R.J., Weisberg, P.J., 2009. Historical and modern disturbance regimes, stand structures, and landscape dynamics in pinon–juniper vegetation of the western United States. Rangeland Ecol. Manag. 62 (3), 203-222, https://doi.org/10.2111/08-188R1.1

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19 (12), 1572-1574, https://doi.org/10.1093/bioinformatics/btg180

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PloS One. 7 (4), e33394, https://dx.doi.org/10.1371%2Fjournal.pone.0033394

Salas-Lizana, R., Oono, R., 2018. Double-digest RAD seq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles. Ecol. Evol. 8 (13), 6638-6651, https://doi.org/10.1002/ece3.4147

Santos, T., Tellería, J.L., Virgós, E., 1999. Dispersal of Spanish juniper *Juniperus thurifera* by birds and mammals in a fragmented landscape. Ecography. 22 (2), 193-204, https://doi.org/10.1111/j.1600-0587.1999.tb00468.x

Sauquet, H., Ho, S.Y., Gandolfo, M.A., Jordan, G.J., Wilf, P., Cantrill, D.J., Bayly, M.J., Bromham, L., Brown, G.K., Carpenter, R.J. and Lee, D.M., 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst. Biol. 61 (2), 289-313.

Shao, C.C., Shen, T.T., Jin, W.T., Mao, H.J., Ran, J.H., Wang, X.Q., 2019. Phylotranscriptomics resolves interspecific relationships and indicates multiple historical out-of-North America dispersals through the Bering Land Bridge for the genus *Picea* (Pinaceae). Mol. Phylogenet. Evol. 141, 106610, https://doi.org/10.1016/j.ympev.2019.106610

Snir, S., Rao, S., 2012. Quartet MaxCut: a fast algorithm for amalgamating quartet trees. Mol. Phylogenet. Evol. 62 (1), 1-8, https://doi.org/10.1016/j.ympev.2011.06.021

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30 (9), 1312-1313, https://doi.org/10.1093/bioinformatics/btu033

Stegemann, S., Keuthe, M., Greiner, S., Bock, R., 2012. Horizontal transfer of chloroplast genomes between plant species. Proc. Natl. Acad. Sci. U.S.A. 109 (7), 2434-2438, https://doi.org/10.1073/pnas.1114076109

Stephens, J.D., Rogers, W.L., Mason, C.M., Donovan, L.A., Malmberg, R.L., 2015. Species tree estimation of diploid *Helianthus* (Asteraceae) using target enrichment. Am. J. Bot. 102 (6), 910-920, https://doi.org/10.3732/ajb.1500031

Stephens, M.A., 1974. EDF statistics for goodness of fit and some comparisons. J. Am. Stat. Assoc. 69 (347), 730-737, https://doi.org/10.2307/2286009

Swenson, N.G., Howard, D.J., (2005). Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am. Nat. 166 (5), 581-591, https://doi.org/10.1086/491688

Takahashi, T., Nagata, N., Sota, T., 2014. Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. Mol. Phylogenet. Evol. 80, 137-144, https://doi.org/10.1016/j.ympev.2014.07.016

Taylor, C.A., 2008. Ecological consequences of using prescribed fire and herbivory to manage *Juniperus* encroachment. In: Van Auken, O.W. (Ed.), Western North American Juniperus Communities. Springer, New York, pp. 239-252.

Terry, R.G., 2010. Re-evaluation of morphological and chloroplast DNA variation in *Juniperus osteosperma* Hook and *Juniperus occidentalis* Torr. Little (Cupressaceae) and their putative hybrids. Biochem. Syst. Ecol. 38 (3), 349-360, https://doi.org/10.1016/j.bse.2010.03.001

Terry, R.G., Nowak, R.S., Tausch, R.J., 2000. Genetic variation in chloroplast and nuclear ribosomal DNA in Utah juniper (*Juniperus osteosperma*, Cupressaceae): evidence for interspecific gene flow. Am. J. Bot. 87 (2), 250-258, https://doi.org/10.2307/2656913

Terry, R.G., Pyne, M.I., Bartel, J.A., Adams, R.P., 2016. A molecular biogeography of the New World cypresses (*Callitropsis*, *Hesperocyparis*; Cupressaceae). Plant Syst. Evol. 302 (7), 921-942.

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and species tree methods exhibit statistically indistinguishable accuracy under a range of simulated conditions. PLoS Curr. 7.

Tsitrone, A., Kirkpatrick, M., Levin, D.A., 2003. A model for chloroplast capture. Evolution. 57 (8), 1776-1782, https://doi.org/10.1111/j.0014-3820.2003.tb00585.x

Vasek, F.C., 1966. The distribution and taxonomy of three western junipers. Brittonia. 18 (4), 350-372, https://doi.org/10.2307/2805152

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22 (3), 787-798, https://doi.org/10.1111/mec.12023

Wang, Q., Mao, K.S., 2016. Puzzling rocks and complicated clocks: how to optimize molecular dating approaches in historical phytogeography. New Phytol. 209 (4), 1353-1358.

Wang, X.Q., Ran, J.H., 2014. Evolution and biogeography of gymnosperms. Mol. Phylogenet. Evol. 75, 24-40, https://doi.org/10.1016/j.ympev.2014.02.005

Weir, J.T., Schluter, D., 2007. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. Science. 315 (5818), 1574-1576, https://doi.org/10.1126/science.1135590

Weisberg, P.J., Lingua, E., Pillai, R.B., 2007. Spatial patterns of pinyon–juniper woodland expansion in central Nevada. Rangeland Ecol. Manag. 60 (2), 115-124, https://doi.org/10.2111/05-224R2.1

West, N.E., Tausch, R.J., Rea, K.H., Tueller, P.T., 1978. Phytogeographical variation within juniper-pinyon woodlands of the Great Basin. Great Basin Naturalist Memoirs. 8 (2), 119-136, https://www.jstor.org/stable/23376562

Willson, C.J., Manos, P.S., Jackson, R.B., 2008. Hydraulic traits are influenced by phylogenetic history in the drought-resistant, invasive genus *Juniperus* (Cupressaceae). Am. J. Bot. 95 (3), 299-314, https://doi.org/10.3732/ajb.95.3.299

Willyard, A., Syring, J., Gernandt, D.S., Liston, A., Cronn, R., 2007. Fossil calibration of molecular divergence infers a moderate mutation rate and recent radiations for *Pinus*. Mol. Biol. Evol. 24 (1), 90-101, https://doi.org/10.1093/molbev/msl131

Wilson, J.S., Pitts, J.P., 2010. Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: a resource for biologists. Prog. Phys. Geogr. 34 (4), 419-441, https://doi.org/10.1177%2F0309133310363991

Wolfe, J.A., 1964. Miocene floras from Fingerrock wash, southwestern Nevada. US Geological Survey Professional Paper. 454-N, 1-36.

Wolfe, J.A., 1978. A paleobotanical interpretation of Tertiary climates in the Northern Hemisphere: Data from fossil plants make it possible to reconstruct Tertiary climatic changes, which may be correlated with changes in the inclination of the earth's rotational axis. Am. Sci. 66 (6), 694-703, https://www.jstor.org/stable/27848958

Xiang, Q.P., Wei, R., Zhu, Y.M., Harris, A.J., Zhang, X.C., 2018. New infrageneric classification of *Abies* in light of molecular phylogeny and high diversity in western North America. J. Syst. Evol. 56 (6), 562-572, https://doi.org/10.1111/jse.12458

Xie, S., Jialiang, L., Jibin, M., Jingjing, X., Kangshan, M., 2019. The complete chloroplast genome of *Juniperus squamata* (Cupressaceae), a shrubby conifer from Asian Mountains. Mitochondrial DNA Part B. 4 (2), 2137-2139.

Xu, T., Abbott, R.J., Milne, R.I., Mao, K., Du, F.K., Wu, G., Zhaxi, C., Liu, J., 2010. Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. BMC Evol. Biol. 10 (1), 194.

Zanoni, T.A., Adams, R.P., 1976. The genus *Juniperus* in Mexico and Guatemala: Numerical and chemosystematic analysis. Biochem. Syst. 4 (3), 147-158.

Zhu, A., Fan, W., Adams, R.P., Mower, J.P., 2018. Phylogenomic evidence for ancient recombination between plastid genomes of the *Cupressus-Juniperus-Xanthocyparis* complex (Cupressaceae). BMC Evol. Biol. 18 (1), 137.

Figure Legends

Figure 1. The serrate leaf junipers are distributed across arid and semi-arid regions of the western United States, Mexico, and Guatemala. Colors representing sampling localities correspond with those designating serrate juniper clades in the phylogenies of Figures 2-4. Outgroup specimens are not shown in map. Map created with ArcGIS Pro 2.4.0 (http://www.esri.com).

Figure 2. Phylogenetic analyses of ddRADseq data with maximum likelihood (left) and SVDquartets (right) provide largely consistent topologies for the serrate juniper clade and its relatives. Nine monophyletic clades resolved by both methods are indicated by colored boxes. Bootstrap support values are reported for all nodes. Branch lengths are not meaningful for the SVDquartets tree.

Figure 3. Comparison of the maximum likelihood ddRADseq tree (left) to a Bayesian cpDNA tree (right) reveals five clear instances of discordance, indicated by dashed arrows. Nine low-level clades resolved with ddRADseq data (Fig. 2) are indicated by colored boxes.

Figure 4. (A) Maximum clade credibility tree (MCC) from analyses in `RevBayes` of the serrate leaf juniper clade calibrated with fossil evidence. Smooth leaf juniper outgroup taxa were excluded from the figure for clarity. Asterisks identify two of the three calibration nodes (the calibrated root node is not shown because it was pruned prior to visualization; see Methods and Table S2 for details). All nodes received greater than 99% Bayesian posterior support. The nine low-level clades resolved in RAxML and SVDquartets phylogenetic analyses of the full set of

ddRADseq data (Fig. 2) are indicated by colored boxes. (B) Lineage through time plot for the serrate juniper clade generated with the Bayesian MCC tree in panel A. Grey dashed line represents linear diversification rate through time given the estimated crown age of the serrate clade and the extant number of species.

Figure 5. Ancestral ranges for the serrate junipers based on a dated phylogeny produced with `RevBayes` and the DIVALIKE model in `BioGeoBEARS`. The map inset shows the delineation of five operational areas (A, western U.S.; B, central U.S.; C, eastern U.S.; D, northern/central MX; E, southern MX), which, along with information of species distributions, informed the geographic ranges assigned to each species and model-based estimates of ancestral ranges. Pie charts at each node represent the marginal probabilities for each range estimated with maximum likelihood, where the colors of the pie sectors either represent single ancestral ranges indicated within the map inset or a possible combination of two ancestral ranges, in which case a novel color was chosen.

Figure 1


Western U.S. clade
ashei clade
deppeana complex
One-seeded clade
Cerro Potosí clade
durangensis clade
Subalpine-alpine clade
J. flaccida
poblana complex
J. occidentalis
J. osteosperma
J. grandis
J. californica
J. monosperma
J. deppeana var. deppeana
J. ashei
J. pinchotii
J. ovata
J. arizonica
J. durangensis Sierra de Gamón
J. poblana var. decurrens
J. zanonii
J. flaccida
J. durangensis
J. coahuilensis
J. angosturana
J. poblana var. poblana Nayarit
J. martinezii
J. saltillensis
J. jaliscana
J. monticola
J. poblana var. poblana Amozoc
J. poblana var. poblana Oaxaca
J. comitana
J. deppeana var. gamboana
J. standleyi
N

Figure 2

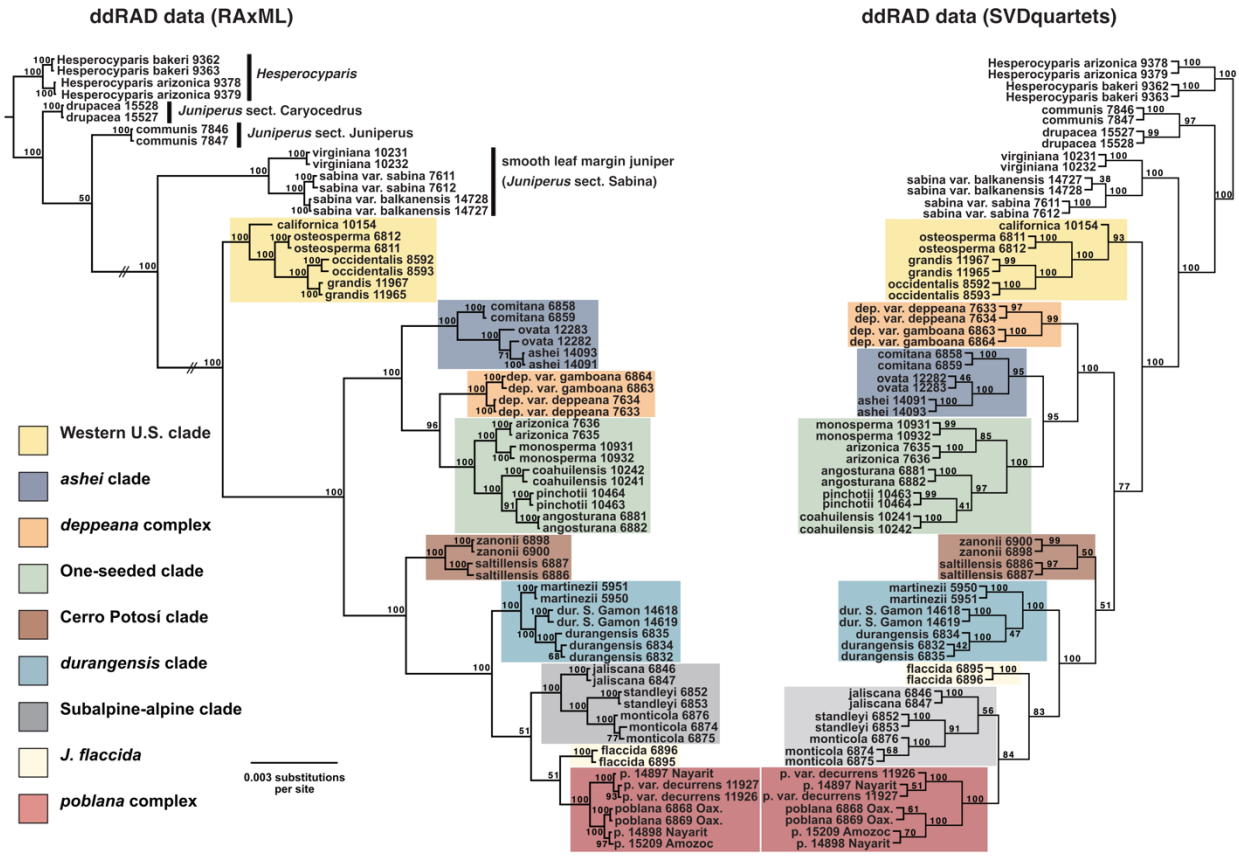

ddRAD data (RAxML)
ddRAD data (SVDquartets)
Hesperocyparis bakeri 9362
Hesperocyparis bakeri 9363
Hesperocyparis arizonica 9378
Hesperocyparis arizonica 9379
Hesperocyparis
drupacea 15528
drupacea 15527
Juniperus sect. Caryocedrus
communis 7846
communis 7847
Juniperus sect. Juniperus
virginiana 10231
virginiana 10232
sabina var. sabina 7611
sabina var. sabina 7612
sabina var. balkanensis 14728
sabina var. balkanensis 14727
smooth leaf margin juniper
(Juniperus sect. Sabina)
californica 10154
osteosperma 6812
osteosperma 6811
occidentalis 8592
occidentalis 8593
grandis 11967
grandis 11965
comitana 6858
comitana 6859
ovata 12283
ovata 12282
ashei 14093
ashei 14091
dep. var. gamboana 6864
dep. var. gamboana 6863
dep. var. deppeana 7634
dep. var. deppeana 7633
arizonica 7636
arizonica 7635
monosperma 10931
monosperma 10932
coahuilensis 10242
coahuilensis 10241
pinchotii 10464
pinchotii 10463
angosturana 6881
angosturana 6882
zanonii 6898
zanonii 6900
saltillensis 6887
saltillensis 6886
martinezii 5951
martinezii 5950
dur. S. Gamon 14618
dur. S. Gamon 14619
durangensis 6835
durangensis 6834
durangensis 6832
jaliscana 6846
jaliscana 6847
standleyi 6852
standleyi 6853
monticola 6876
monticola 6874
monticola 6875
flaccida 6896
flaccida 6895
p. 14897 Nayarit
p. var. decurrens 11927
p. var. decurrens 11926
poblana 6868 Oax.
poblana 6869 Oax.
p. 14898 Nayarit
p. 15209 Amozoc
Western U.S. clade
ashei clade
deppeana complex
One-seeded clade
Cerro Potosí clade
durangensis clade
Subalpine-alpine clade
J. flaccida
poblana complex
0.003 substitutions per site

Figure 3

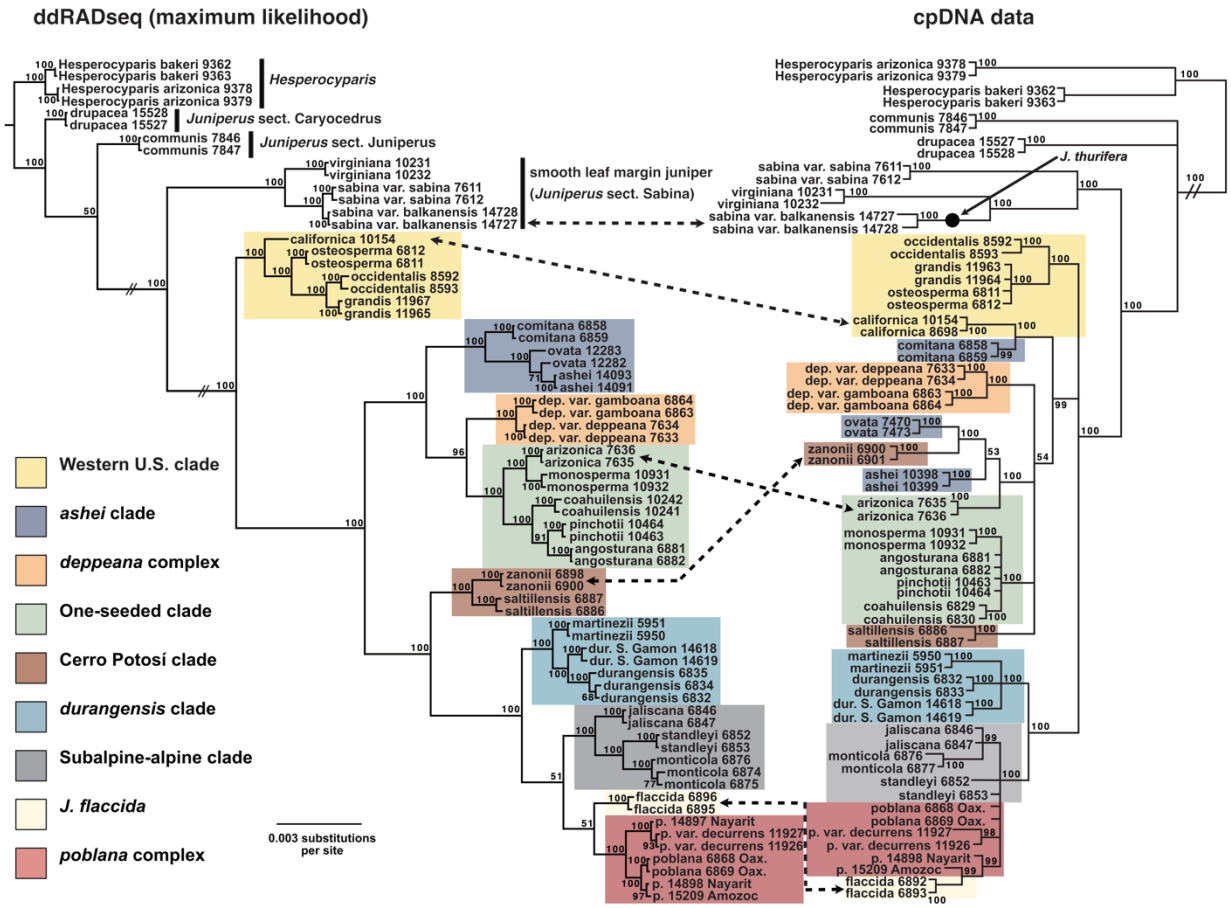

ddRADseq (maximum likelihood)
cpDNA data
Hesperocyparis
Juniperus sect. Caryocedrus
Juniperus sect. Juniperus
smooth leaf margin juniper
(Juniperus sect. Sabina)
J. thurifera
Western U.S. clade
ashei clade
deppeana complex
One-seeded clade
Cerro Potosí clade
durangensis clade
Subalpine-alpine clade
J. flaccida
poblana complex
0.003 substitutions per site

Figure 4

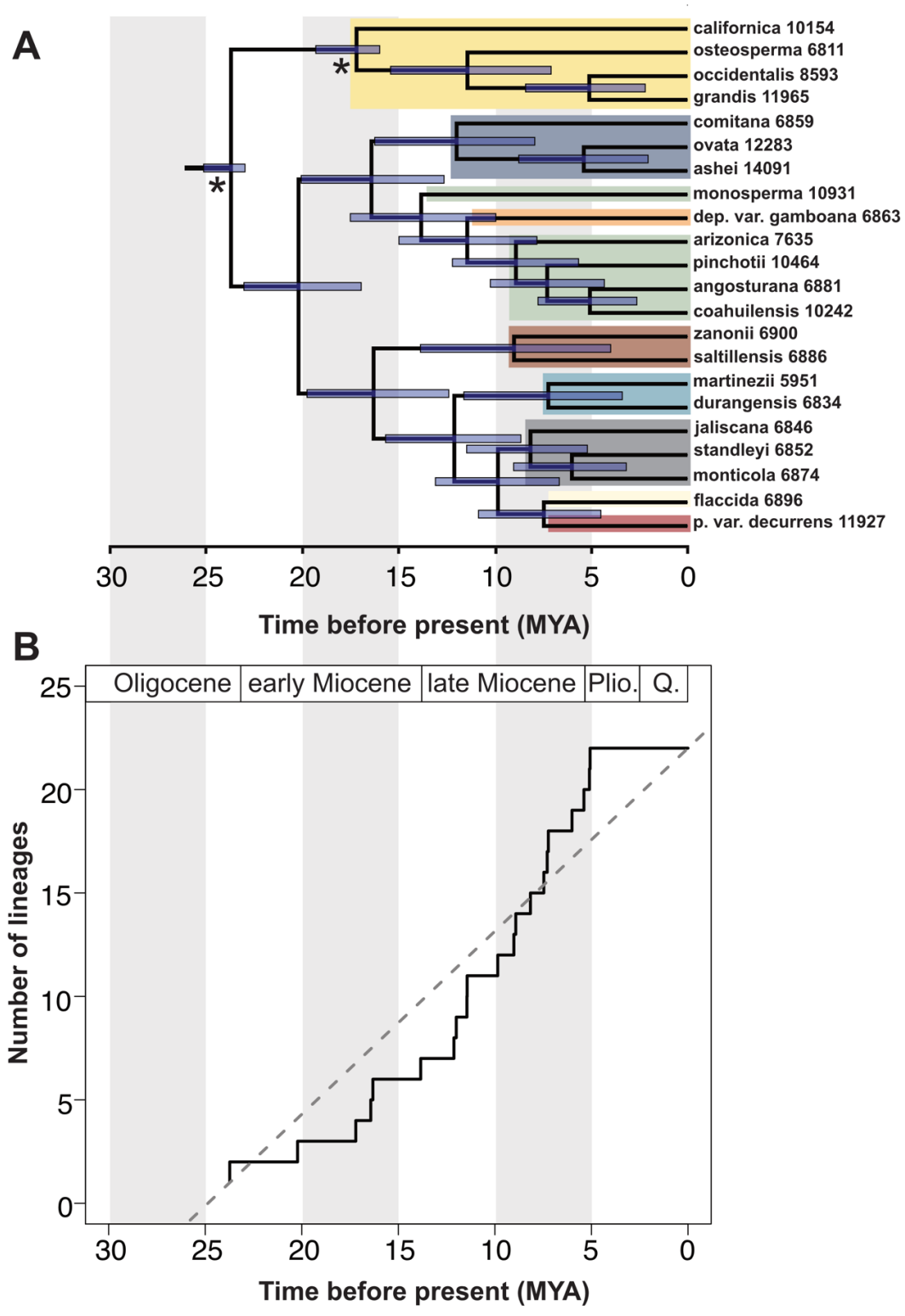
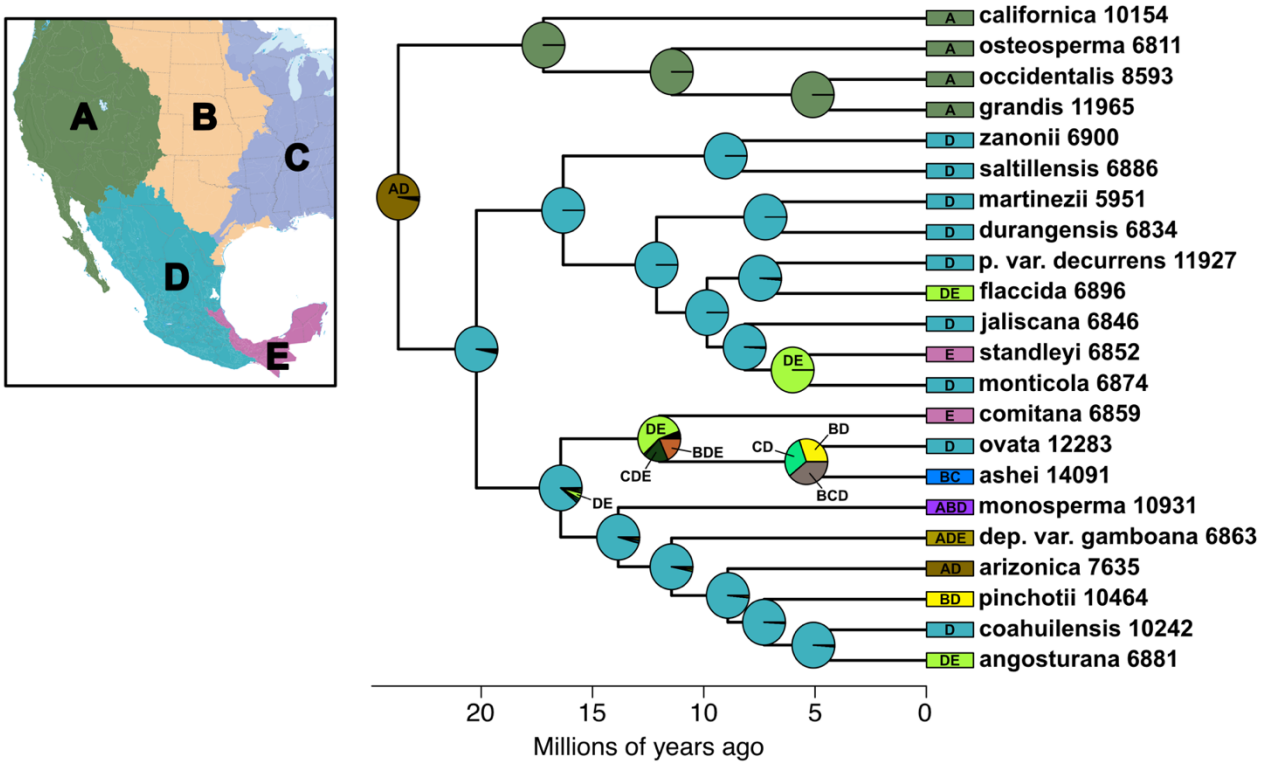

A
californica 10154
osteosperma 6811
occidentalis 8593
grandis 11965
comitana 6859
ovata 12283
ashei 14091
monosperma 10931
dep. var. gamboana 6863
arizonica 7635
pinchotii 10464
angosturana 6881
coahuilensis 10242
zanonii 6900
saltillensis 6886
martinezii 5951
durangensis 6834
jaliscana 6846
standleyi 6852
monticola 6874
flaccida 6896
p. var. decurrens 11927
30 25 20 15 10 5 0
Time before present (MYA)
B
Oligocene
early Miocene
late Miocene
Plio.
Q.
Number of lineages
0 5 10 15 20 25
30 25 20 15 10 5 0
Time before present (MYA)

Figure 5


A
B
C
D
E
AD
DE
BDE
CDE
CD
BD
BCD
californica 10154
osteosperma 6811
occidentalis 8593
grandis 11965
zanonii 6900
saltillensis 6886
martinezii 5951
durangensis 6834
p. var. decurrens 11927
flaccida 6896
jaliscana 6846
standleyi 6852
monticola 6874
comitana 6859
ovata 12283
ashei 14091
monosperma 10931
dep. var. gamboana 6863
arizonica 7635
pinchotii 10464
coahuilensis 10242
angosturana 6881
20
15
10
5
0
Millions of years ago

Supplementary material

Table S1. Clade designation, voucher number, locality description, longitude/latitude, and elevation for each of the samples used in this study. Voucher specimens are deposited at BAYLU herbarium, Baylor University, Waco, TX.

Species	Clade	Accession	Locality	Latitude	Longitude	Elevation (m)
Hesperocyparis arizonica	*Hesperocyparis*	Adams 9378, 9379	ex *J. Bartel 1580A, B*, upper Bear Canyon, Pima Co., AZ	32° 21'47.9" N	110° 42'50.3" W	1590
Hesperocyparis bakeri	*Hesperocyparis*	Adams 9362, 9363	ex *J. Bartel 1572A, B*, Thousand Lake Wilderness, Shasta, CA	40° 45'3.3" N	121° 36'30" W	1590
J. angosturana	Serrate junipers (sect. Sabina)	Adams 6881, 6882	21 km e of Cerritos, San Luis Potosi, MX	22° 17' 0.0" N	100° 08' 59.4" W	1040
J. arizonica	Serrate junipers (sect. Sabina)	Adams 7635, 7636	Rock Hound State Park, Luna Co., NM	30° 11' 12" N	107° 36' 42.4" W	1400
J. ashei	Serrate junipers (sect. Sabina)	Adams 14091, 14093	Benbrook Lake, Tarrant Co., TX	32° 37.515' N	97° 30.154' W	250
J. californica	Serrate junipers (sect. Sabina)	Adams 10154, 10155	Jct. I15 and US 395, San Bernardino Co., Hesperia, CA	34° 24' 19.4" N	117° 23' 38.1" W	1131
J. coahuilensis	Serrate junipers (sect. Sabina)	Adams 10241, 10242	18 km n of Durango, MX, on Mex. 45	24° 09' 04.067" N	104° 42' 27.7" W	1938
J. comitana	Serrate junipers (sect. Sabina)	Adams 6858, 6859	27 km W of Laguna Montebello on Mex. 307, Chiapas, MX	16° 08' 20" N	91° 54' 25" W	1540
J. communis var. *communis*	sect. Juniperus	Adams 7846, 7847	Stockholm, Sweden	59° 19' 53" N	18° 04' 20" E	30
J. deppeana var. *deppeana*	Serrate junipers (sect. Sabina)	Adams 7633, 7634	Sacramento Mtns., 13-14 km E. of Alamogordo, Otero Co., NM	32° 57' 36" N	105° 48' 44" W	2050
J. deppeana var. *gamboana*	Serrate junipers (sect. Sabina)	Adams 6863, 6864	17 km N. of Comitan on Mex. 190., Chiapas, MX	16° 22' 0.7" N	92° 13' 12" W	1900

J. drupacea	sect. Caryocedrus	Adams 15227, 15228	Achladokampos Pass, Arkadia, Greece	37° 31' 25" N	22° 38' 19" E	600
J. durangensis	Serrate junipers (sect. Sabina)	Adams 6832, 6834, 6835	52 km W. El Salto, Durango, MX	23° 42' 31" N	105° 43' 25" W	2735
J. durangensis (Sierra de Gamón)	Serrate junipers (sect. Sabina)	Adams 14618, 14619	*ex. Socorro Gonzalez* 7391 a, b, Sierra de Gamón, Durango, MX	24° 30' N	104° 14' 25" W	2500
J. flaccida	Serrate junipers (sect. Sabina)	Adams 6895, 6896	22 km E. San Roberto Jct., Nuevo, Leon, MX	24° 41' 54" N	100° 04' 59" W	1720
J. grandis	Serrate junipers (sect. Sabina)	Adams 11965, 11967	Meyers, El Dorado Co., CA	38º 51' 5.2" N	120º 01' 14.6" W	1937
J. jaliscana	Serrate junipers (sect. Sabina)	Adams 6846, 6847	19 km E. Mex. Hwy 200, on dirt road to Cuale, Jalisco, MX	20° 31.28' N	105° 9.2' W	940
J. martinezii	Serrate junipers (sect. Sabina)	Adams 5950, 5951	42 km N. Lagos de Moreno, Jalisco, MX	21° 35' N	101° 36' W	2084
J. monosperma	Serrate junipers (sect. Sabina)	Adams 10931, 10932	E. of Reserve, Catron Co., NM	33° 44' 36.7" N	108° 43' 31.7" W	1825
J. monticola f. *monticola*	Serrate junipers (sect. Sabina)	Adams 6874, 6875, 6876	El Chico Natl. Park, Hidalgo, MX	20° 08' 54" N	98 ° 40' 16" W	2750
J. occidentalis	Serrate junipers (sect. Sabina)	Adams 8593, 8594	Sisters, Deschutes Co., OR	44° 17' 28.4" N	121° 33' 11" W	960
J. osteosperma	Serrate junipers (sect. Sabina)	Adams 6811, 6812	Little Cottonwood Canyon, Salt Lake City, UT	40° 34' 20.4" N	111° 46' 21.08" W	1680
J. ovata	Serrate junipers (sect. Sabina)	Adams 12281, 12282, 12283	5 km W. of Ozona on FM2398, Crockett Co., TX	30° 42' 9.8" N	101° 16' 21.6" W	772
J. pinchotii	Serrate junipers (sect. Sabina)	Adams 10463, 10464	12 km N. of Meridian, Bosque Co., TX	32° 00' 48.6" N	97° 42' 51.2" W	289
J. poblana var. *poblana*	Serrate junipers (sect. Sabina)	Adams 6868, 6869	62 km S. Oaxaca, MX	16° 41' 36" N	96° 19' 41" W	1710
J. poblana var. *poblana*	Serrate junipers (sect. Sabina)	Adams 14898	N.E. of El Maguey, Nayarit, Mexico	22° 07' 40" N	104° 47' 47" W	1430
J. poblana var. *poblana*	Serrate junipers (sect. Sabina)	Adams 15209	ex *L Caamano A and Allen Coombes* 10173, Amozoc de Mota, Puebla, MX	19° 01' N	98° 01' W	2300
J. poblana var. *decurrens*	Serrate junipers (sect. Sabina)	Adams 11927, 11926	Durango, MX	25° 14' 11" N	106° 26' 55.7" W	1818
J. sabina var. *sabina*	Smooth leaf junipers (sect. Sabina)	Adams 7611, 7612	2 km S St. Niklaus,	46° 09' 24" N	7° 47' 40" E	1300

			Baltschieder, Switzerland			
J. sabina var. *balkanensis*	Smooth leaf junipers (sect. Sabina)	Adams 14727, 14728	ex *A. Tashev,* ns Tsena 1,2, Mount Tsena near Notia village, Greece	41º 08' 29.4" N	22º 14' 42.2" E	1630
J. saltillensis	Serrate junipers (sect. Sabina)	Adams 6886, 6887	14 km E. San Roberto Jct., Nuevo Leon, MX	20˚ 40' 48" N	100˚ 10' 02" W	2090
J. standleyi	Serrate junipers (sect. Sabina)	Adams 6852, 6853	24 km N.W. Huehuetango, Guatemala	15˚ 25' 34" N	91˚ 28' 17" W	3130
J. virginiana	Smooth leaf junipers (sect. Sabina)	Adams 10231, 10232	32 km east of Knoxville on north side of I40, Sevier Co., TN	35° 59.470' N	83° 36.472' W	324
J. zanonii	Serrate junipers (sect. Sabina)	Adams 6898, 6899	Cerro Potosí, Nuevo Leon, MX	24˚ 52' N	100˚ 13' 40.8" W	3490

Table S2. Molecular node dating was conducted within a Bayesian framework utilizing fossil evidence to inform the ages of the root and two internal nodes.

Calibration node and age prior	**Fossil assignment and stratigraphic period**	**Holotype information**	**Justification**
Root; uniform prior [134, 23]	Stem lineage of the MRCA of *J. californica* and *J. osteosperma*; Late Oligocene (minimum age constraint)	UC Museum of Paleontology (UCMP) 7447, holotype of *Juniperus creedensis* (Axelrod 1987) (minimum age constraint)	The minimum age of the root was constrained as follows: the holotype of *Juniperus creedensis* (UCMP 7447) was collected from the Creede Formation (late Oligocene), Rocky Mts., Colorado, United States (Axelrod 1987) and represents the first appearance of the serrate junipers group. The seed cones and shoots of this fossil resemble extant *J. osteosperma* and *J. californica*. The maximum constraint for the root node prior (134 Ma) represents the age of the crown lineage of Cupressoideae (subfamily within Cupressaceae containing *Thuja*, *Cupressus*, *Juniperus*, and other genera) estimated by Mao et al. (2012).
MRCA of extant serrate leaf junipers; diffuse exponential prior with mean 23 Mya + 1/ λ	Stem lineage of the MRCA of *J. californica* and *J. osteosperma*; Late Oligocene	UC Museum of Paleontology (UCMP) 7447, holotype of *Juniperus creedensis* (Axelrod 1987)	The holotype of *Juniperus creedensis* (UCMP 7447) was collected from the Creede Formation (late Oligocene), Rocky Mts., Colorado, United States (Axelrod 1987) and represents the first appearance of the serrate junipers group. The seed cones and shoots of this fossil resemble extant *J. osteosperma* and *J. californica*.
MRCA of *J. californica*, *J. osteosperma*, *J. occidentalis*, and *J. grandis* (western U.S. juniper clade); diffuse exponential	Stem lineage of the MRCA of *J. osteosperma*, *J. occidentalis*, and *J. grandis*; Early Miocene	UC Museum of Paleontology (UCMP) 9355, holotype of *Juniperus desatoyana* (Axelrod 1991)	The holotype of *Juniperus desatoyana* (UCMP 9355) was collected from Buffalo Canyon, Churchill County, Nevada, United States (Axelrod 1991). Seed cones and twigs are very similar to extant *J. occidentalis*; however, a conservative placement of this fossil is as a stem relative to the group containing *J.*

prior with mean 16 Mya + 1/ λ	*osteosperma*, *J. occidentalis*, and *J. grandis*.

Axelrod, D.I., 1987. The late Oligocene Creede flora, Colorado. Vol. 130. University of California Press.

Axelrod, D.I., 1991. The Early Miocene buffalo canyon flora of western Nevada. Vol. 135. University of California Press, Berkeley.

Mao, K., Milne, R.I., Zhang, L., Peng, Y., Liu, J., Thomas, P., Mill, R.R., Renner, S.S., 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc. Natl. Acad. Sci. U.S.A. 109 (20), 7793-7798.

Table S3. Area assignments used for ancestral range estimation with `BioGeoBears`. Areas were defined by both geopolitical and ecologically-relevant boundaries designated by the Level I Ecoregions of North America (https://www.epa.gov/eco-research/ecoregions). Regions are coded as letters (A-E) in the map inset in Fig. 5 as follows: A, western U.S.; B, central U.S.; C, eastern U.S.; D, northern/central MX; E, southern MX.

Species	**Areas**
J. californica	Western U.S.
J. osteosperma	Western U.S.
J. occidentalis	Western U.S.
J. grandis	Western U.S.
J. deppeana var. *gamboana*	Western U.S., northern/central MX, southern MX
J. monosperma	Western U.S., central U.S., northern/central MX
J. comitana	Southern MX
J. ovata	Northern/central MX
J. ashei	Central U.S., eastern U.S.
J. angosturana	Northern/central MX, southern MX
J. coahuilensis	Northern, central MX
J. pinchotii	Central U.S., northern/central MX
J. arizonica	Western U.S., northern/central MX
J. zanonii	Northern/central MX
J. saltillensis	Northern/central MX
J. martinezii	Northern/central MX
J. durangensis	Northern/central MX
J. jaliscana	Northern/central MX
J. standleyi	Southern MX
J. monticola	Northern/central MX
J. flaccida	Northern/central MX, southern MX

J. poblana var. *decurrens*	Northern/central MX

Chapter 2: History and environment shape population genetic and phytochemical variation across three western *Juniperus* and their hybrids

Kathryn A. Uckele[1,2,3]

Casey Philbin[3]

Lora Richards[1,2,3]

Lee Dyer[1,2,3]

Joshua P. Jahner[4]

Robert P. Adams[5]

Thomas L. Parchman[1,2]

[1]Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV, 89557, USA

[2]Department of Biology, University of Nevada, Reno, NV, 89557, USA

[3]Hitchcock Center for Chemical Ecology, University of Nevada, Reno, NV 89557, USA

[4]Department of Botany, University of Wyoming, Laramie, WY 82071, USA

[5]Biology Department, Baylor University, Waco, TX, 76798, USA

1. Abstract

Hybridization is a common consequence of secondary contact among plant species which can have extended ecological consequences through its effects on phenotype. Junipers (*Juniperus*) are foundational tree species in many semi-arid landscapes of the western United States, and hybridize frequently. *Juniperus grandis*, *J. occidentalis*, and *J. osteosperma* are a closely related group of species known to hybridize in a complex zone of secondary contact in western Nevada, permitting investigation into the drivers of hybridization and its phenotypic consequences. Here, we generated population genomic data (9,125 SNPs; 326 individuals; 25 populations) to quantify patterns of genetic variation across populations and species and to characterize ancestry variation in hybrids and phytochemical consequences. Our analyses indicated that the parental species represent well-differentiated, monophyletic lineages which are phenotypically and ecologically distinct. As is common in other groups of conifers, we detected little evidence for genetic differentiation among populations within each species. Hybrids occupied environmental conditions intermediate of the parental species, exhibited ancestry from all three parental species, and were largely F_1 and advanced generation backcrosses. We then used GCMS to characterize terpenoid variation across parental species and hybrids to better understand how underlying environmental and ancestry variation shape phytochemical variation. Despite high variability among hybrids, the hybrid chemotype was identifiable and distinct from parentals due in part to transgressive variation at a subset of terpenoid compounds. Our results illustrate how history and environment predict ancestry variation across a hybrid zone, and how admixture across the species boundary generates novel variation in an ecologically-important trait.

KEYWORDS: admixture**,** population genetics, *Juniperus*, hybridization, secondary contact, GC-MS, plant secondary chemistry, terpenoid

2. Introduction

Recent applications of genome-scale data at the population and phylogenetic levels have increasingly revealed a prevalence of reticulate evolution and hybridization across disparate taxonomic lineages (Linan et al., 2021; Esquerré et al., 2021; Wang et al., 2021). Hybridization occurs in as many as 34% of plant families (Mallet 2005) where it has likely contributed pervasively to patterns of diversification (Soltis et al., 2015; Tank et al, 2015; Alix et al., 2017) as well as the evolution of domesticated (Arnold, 2004) and invasive forms (Ellstrand and Schierenbeck 2000). When allopatrically diverged lineages hybridize upon secondary contact, the proximate and ultimate outcomes of hybridization depend on early-generation stochasticity, ecological context, and the genetic basis of reproductive isolation (Baack and Rieseberg, 2007; Mandeville et al., 2017; McFarlane et al., 2021). If reproductive isolation is nearly complete, F_1 hybrids may have reduced fertility and F_2 and backcross hybrids may rarely occur (Dobzhansky, 1970; Burke and Arnold, 2001; Burton et al., 2013). In the absence of strong isolating mechanisms, advanced-generation hybrids may form that span the ancestry continuum, generating a bridge between previously isolated lineages. Such hybrid zones can shed light on barriers to hybrid admixture (Kay, 2006; Mandeville et al., 2019) and have utility for mapping the genetic basis of complex phenotypes (Buerkle and Lexer, 2008; Barker et al., 2019). By generating pronounced patterns of phenotypic variation, hybridization in foundational plant species can also have consequences that extend to community and ecosystem levels (e.g., Wimp et al., 2005; Caseys et al., 2012).

Trees and shrubs in the genus *Juniperus* are foundational species in many arid regions of the Northern Hemisphere. The serrate leaf junipers of North America (21 species) diversified into the increasingly arid and mountainous regions of the western United States and Mexico over

the last ca. 23 million years (Mao et al., 2010; Uckele et al., 2021). Hybridization is common in the serrate leaf juniper clade (Adams, 1994; Adams et al., 2017; Adams et al., 2020); a well-documented example involves hybridization among *Juniperus grandis*, *J. occidentalis*, and *J. osteosperma* where they have come into secondary contact in western Nevada (Terry et al., 2000; Terry 2010). *Juniperus grandis* (Sierra juniper) and *J. occidentalis* (western juniper) occur along the drier, eastern aspects of the Sierra and Cascade ranges, respectively, and diverged from one another approximately 5 Mya (Uckele et al., 2021). Ancestral to them is *J. osteosperma* (Utah juniper), whose lineage diverged from that of *J. grandis* and *J. occidentalis* approximately 11 Mya (Uckele et al., 2021), and occurs throughout the Great Basin and Colorado Plateau ecoregions. *Juniperus grandis*, *J. occidentalis*, and *J. osteosperma* outcompete grasses and shrubs to form large, natural single- and double-species stands in otherwise tree-impoverished regions, and commonly associate with members of the pinyon pine group (*Pinus* subsection Cembroides) to form pinyon-juniper woodlands, an important and expanding ecosystem in much of the western United States and northern Mexico (Miller and Wigand, 1994; Weisberg et al., 2007). Pinyon-juniper woodlands exist in warmer climates than other regional tree-dominated ecosystems, and are associated with high diversities of plant, bird, mammal and invertebrate species (Gottfried et al., 1995; Tonkel et al., 2021). In foundational species such as these, the effects of hybridization can be far-reaching, impacting community composition and ecosystem processes via novel variation in functional traits (Evans et al., 2008; Floate et al., 2016).

The geographic ranges of many North American trees, including *Juniperus*, shifted in response to Pleistocene glaciation cycles (Roberts and Hamann, 2015). Hybridization at secondary contact zones among closely related species commonly cluster near western mountain ranges at regions of pronounced environmental transition (Remington 1968; Swenson and

Howard, 2005). At the height of the Wisconsin glacial, *J. grandis*, *J. occidentalis*, and *J. osteosperma* likely persisted in geographically isolated glacial refugia. Analyses of packrat midden macrofossils suggest that *J. grandis* occurred in the mesic western aspect of the central Sierra Nevada range, *J. osteosperma* existed significantly south of its contemporary range in the current-day Sonoran, Mojave, and Chihuahuan deserts, and *J. occidentalis* occurred, at least temporarily, in the northern Great Basin (Betancourt et al., 1990). However, over the last 11 Kya of the Holocene interglacial, *J. osteosperma* migrated north, ostensibly entering into secondary contact with *J. grandis* and *J. occidentalis* in the eastern foothills of the Sierra Nevada range (Betancourt et al., 1990). The resulting hybrid zone in western Nevada spans an ecotone characterized by the transition from high alpine to low basin (Vasek, 1966).

Previous studies based on morphological, phytochemical, and genetic data provided evidence for hybridization in the region (Vasek, 1966; Terry *et al.,* 2000; Terry *et al.,* 2010; Adams, 2013ab), but were based on few characters and provided limited resolution of hybrid admixture. For example, Terry et al., (2000) suggested introgression from *J. occidentalis* into *J. osteosperma*, however, small numbers of chloroplast and nuclear ribosomal loci precluded their ability to distinguish hybrid ancestry arising from *J. grandis* and *J. occidentalis* and to characterize genome-wide patterns of ancestry. Additionally, chemical analyses by Adams (2013a,b) inferred terpenoid variation consistent with backcrossed and stabilized hybrid populations, respectively, but lacked the geographic scope to characterize these patterns across the hybrid zone. More generally, population genetic analyses in North American *Juniperus* are lacking, despite their ecological significance for many arid- and semi-arid regions of the continent.

Understanding how hybridization affects functional traits like plant secondary chemistry is important due to the well-documented role of phytochemical variation in mediating species interactions (Wimp et al., 2005), the diversity of ecological communities (Abrahamson et al., 2003; Poelman et al., 2009), and the flow of resources through ecosystems (Schweitzer et al., 2008). Conifers produce a large number of distinct and structurally diverse terpenoid molecules that shape interactions via their effects on herbivore behavior and physiology (Keeling and Bohlmann, 2006). A well-documented example is myrcene and other monoterpene volatiles, which elicit mass beetle attacks when they synergize with bark beetle pheromones (Seybold et al., 2006). *Juniperus* spp. in particular exhibit impressive terpenoid structural diversity (Otto and Wilde, 2001) that has been linked to Lepidopteran performance (Pardikes et al., 2019), deer diet preference (Schwartz et al., 1980), and selective browsing damage (Markó et al., 2011). Adams (2013a, 2013b) sought to differentiate hybrid and parental individuals within the *J. grandis*, *J. occidentalis*, and *J. osteosperma* hybrid zone based on terpenoid data (32 and 28 terpenoid characters, respectively), but found contrasting patterns of hybrid terpenoid variation, suggesting that accurate assignment of admixed individuals to specific hybrid classes may be required to resolve this discrepancy.

Hybridization produces a spectrum of potential phenotypic outcomes - ranging from intermediacy to transgression - which can have consequential impacts on the composition of ecological communities. Phytochemically intermediate hybrids are often associated with higher species richness and variable rates of herbivory than either of the parentals (Bangert et al., 2004), and may theoretically promote expansions of host range by specialized herbivores (Floate & Whitham, 1993; Araujo et al., 2015; Floate et al., 2016). For example, hybridization in the lodgepole pine x jack pine hybrid zone by the mountain pine beetle may have facilitated range

expansion by the mountain pine beetle from its native host, lodgepole pine, onto its novel host, jack pine, by increasing its geographic proximity to the novel host or by providing a stepping stone from one fitness peak to another (Erbilgin, 2018). Additionally, transgressive segregation is a prominent feature of admixture among some diverged plant lineages (Rieseberg et al., 1999), and can produce compound concentrations in hybrids that surpass the range of the parental species (Cheng et al., 2011). For example, Adams (2013a) found that hybrids of *J. grandis*, *J. occidentalis*, and *J. osteosperma* were transgressive at a majority of major terpenes (18/32), effectively obscuring intermediate variation and complicating the detection of hybrids. By generating novel phytochemical variation, transgressive segregation can enlarge the "working surface" upon which natural selection may operate (Stelkens and Seehausen, 2009), and can promote phenotypic and lineage diversification into underutilized areas of niche space (Seehausen, 2004; Rieseberg et al., 2007). Although most late-generation plant hybrids tend to be more susceptible to insect herbivory than parental and F_1 hybrids, a small proportion of hybrids may be more resistant, suggesting a role for transgressive segregation in promoting the diversification of phytochemical defense (Cheng et al., 2011). Modern approaches for analyzing phytochemical diversity coupled with genome-wide approaches for quantifying ancestry variation stand to shed light on the chemical outcomes of ancestry variation and form a basis for understanding its potential ecological consequences.

Here we analyzed patterns of genetic, phytochemical, and environmental variation across the three parental species and hybrids spanning a broad admixture zone in western Nevada. We generated population genomic data (9,125 SNPs, 326 individuals) for 25 populations to explore the predictors of genetic variation within and among each species and their hybrids. We resolved the pattern of diversification among the three focal species, quantified spatial genetic structure

across the range of each, and characterized ancestry variation of hybrids along with its environmental predictors. To understand the influence of ancestry on phytochemical variation, we quantified variation in composition and concentrations for 163 terpenoids using gas chromatography coupled to mass spectrometry (GC-MS) across the same trees for which we generated genetic data. Our results illustrate that hybrids had ancestry involving all three parental species and followed a geographic and environmental gradient across the zone. Hybrid terpenoid chemistry was distinct from the parentals and was characterized by transgressive and intermediate inheritance at many key terpenoids with documented antifungal properties in other conifers.

3. Methods

3.1 Sampling, library preparation, and DNA sequencing

We sampled leaf material from 388 individuals for genetic and phytochemical analyses. Of those 388, 326 were from throughout the ranges of *J. grandis*, *J. occidentalis*, and *J. osteosperma* (hereafter referred to as parental taxa) and the putative hybrid zone. We sampled three *J. grandis* populations from its northern range in central California, five *J. occidentalis* populations from its southern range in southern Oregon and northern California, seven *J. osteosperma* populations from Nevada, eastern California, and Utah, and ten putative hybrid populations from western and central Nevada (Fig. 1; Supplementary Table 1). The remaining 62 samples were taken from populations of *J. arizonica*, *J. californica*, *J. deppeana*, and *J. scopulorum*, which were utilized as outgroup samples for phylogenetic and population genetic analyses.

Leaf samples were dried and stored in silica gel prior to DNA extraction with Qiagen DNeasy Plant Mini Kits (Qiagen Inc., Valencia, CA, USA). Genomic DNA was quantified and quality checked with a QIAxpert microfluidic analyzer (Qiagen Inc., Valencia, CA, USA), and reduced-representation libraries were constructed using a double digest RAD sequencing approach (ddRADseq; Parchman *et al*., 2012; Peterson *et al*., 2012). Briefly, genomic DNA from each individual was digested with *Eco*RI and *Mse*I restriction enzymes. Custom oligos containing barcodes and Illumina adaptors were annealed to the *Eco*RI cut sites, and oligos containing the alternative Illumina adaptor were annealed to the *Mse*I cut sites. Digests were PCR amplified with a high-fidelity proofreading polymerase (Iproof polymerase, BioRad Inc., Hercules, CA, USA) and pooled for sequencing. Libraries were size-selected for fragments between 350 and 450 bp in length with the Pippin Prep System (Sage Sciences, Beverly, MA) at the University of Texas Genome Sequencing and Analysis Facility (Austin, TX). Four lanes of single-end 100-base sequencing were executed with an Illumina HiSeq 2500 at the University of Wisconsin Madison Biotechnology Center (Madison, WI).

We employed a filtering pipeline (https://github.com/ncgr/tapioca) which uses `bowtie` (Langmead *et al*., 2009) and several databases to align and remove potential sequencing contaminants (e.g., Illumina adaptors and primers, PhiX, *E. coli*). To demultiplex the reads, we used a custom Perl script which matches the barcode sequence from each read to the correct individual and parses the reads accordingly. Parsed reads were split into individual fastq files, which are available at Dryad (DOI).

3.2 Phylogenetic analyses

To clarify patterns of divergence among our parental taxa within the context of other North American *Juniperus* species, we conducted phylogenetic analysis of the putatively pure parental and outgroup samples. We sampled one individual from each of the 15 pure parental populations and 6 outgroup populations, prioritizing samples with the largest numbers of reads that aligned to the *de novo* reference assembly (described in the next section). We used `ipyRAD` (v. 0.7.19; Eaton, 2014) to assemble and filter reads using default parameter settings, as this parameterization provided a highly resolved and supported tree topology. `Ipyrad` parameter specification and alignment statistics can be found in the Supplemental Materials. One SNP was randomly chosen from each stack and analyzed with SVDquartets (Chifman & Kubatko, 2014; 2015), a multi-species coalescent, quartet amalgamation approach, which we implemented with `TETRAD` (Eaton et al., 2017). All possible quartets were inferred using singular value decomposition and amalgamated to produce an unrooted species tree. We conducted 100 bootstrap replicates to generate statistical support at each node and manually rooted the tree with *J. scopulorum* according to Mao *et al.,* (2010).

3.3 Genetic variation across populations, species, and hybrids

Due to the lack of a suitable reference genome for *Juniperus*, we used a *de novo* assembly to generate a reference of genomic regions sampled with the ddRADseq approach. We used `cdhit` (cd-hit-est; Li & Godzik, 2006; Fu *et al*., 2012) to cluster unique reads which were present in at least four samples with a similarity threshold of 80%. This clustered 1,168,219 reads into a *de novo* reference assembly consisting of 160,256 contigs. We used the `aln` and `samse` algorithms in `bwa` (v. 0.7.5a; Li & Durbin 2009) to align reads from each individual to

the *de novo* reference. The resulting BAM format alignments were used to identify variant positions, call single nucleotide polymorphisms (SNPs), and estimate genotype likelihoods with `samtools` and `bcftools` (v. 1.3; Li & Durbin, 2009). We filtered for alignment quality by removing loci with site quality scores (QUAL) lower than 20 and genotype quality scores (GQ) lower than 8, and retained only bi-allelic SNPs. We implemented the `HDplot` pipeline (McKinney *et al*., 2017) in `python` and `R` to filter paralogous loci according to the expected proportion of heterozygous individuals within a population and allelic ratios within heterozygous individuals. Using `vcftools` (v. 0.1.14; Danecek *et al*., 2011), we randomly sampled one SNP per ddRADseq locus to reduce the effects of linkage disequilibrium (--thin 90) and set the minimum minor allele frequency (--maf) to 5%. To understand how genetic structure is partitioned across populations and species, we conducted independent variant calling procedures on 1) all 388 samples (including outgroup samples) and 2) a subset of 325 samples which included parental taxa and hybrids.

To infer ancestry variation across individuals and populations, we used the admixture model implemented with `entropy` (Gompert *et al*., 2014; Shastry et al., 2021). Similar to `STRUCTURE` (Pritchard *et al*., 2000; Falush *et al*., 2003), `entropy` is a hierarchical Bayesian model which utilizes genotype likelihood data and model-based estimates of individual (admixture proportions, q) and population (allele frequency) parameters to inform the prior probability of genotypes at each locus. This framework allows genotype estimates to be probabilistically inferred for sites with low-depth or missing data. As we were interested in the consequences of hybridization among *J. grandis*, *J. occidentalis*, and *J. osteosperma*, we defined the number of source populations, *K*, to be three for the analysis of parental species and hybrid individuals. For the analysis of all samples including outgroups, we set *K* to seven to reflect the

number of different species. To facilitate convergence, we initialized the MCMC chains with cluster membership probabilities generated from linear discriminant analysis and *K*-means clustering of the first five principal components calculated for genotype point estimates obtained from genotype likelihoods (see Jombart et al., 2010). For each subset of samples, we ran four independent chains with a burn-in phase of 30,000 iterations, after which we sampled every tenth iteration for 60,000 iterations. To assess adequate mixing and convergence, we plotted and visualized posterior trace plots in `R` (R Core Team, 2020).

In addition to the admixture model, `entropy` specifies an ancestry complement model, which can be used to distinguish hybrids representing different ancestry classes (F_1, F_2, backcross, and late-generation hybrids). In addition to estimating the proportion of an individual's genome that is derived from one of *K* source populations (q), the ancestry complement model also estimates the proportion of an individual's genome in which allele copies were inherited from the same or different source populations (Q_{12}). For example when hybridization occurs between two parental populations ($K = 2$), Q_{11} and Q_{22} represent the proportions of the genome for which both allele copies were descended from the same parental species, whereas Q_{12} represents the proportion of the genome for which the allele copies were descended from different parental species, i.e., the proportion of interspecific ancestry. One can distinguish first-generation hybrids (F_1) from later-generation hybrids (F_2, F_3, F_N) by posterior estimates of Q_{12}, which are expected to equal one in F_1s and less than one in later-generation hybrids. Additionally, one can distinguish backcrossed hybrids (BC) from other ancestry classes (F_1, F_2, F_N) by values of q, which are expected to equal 0.5 on average in hybrids without backcrossed parentage and < 0.5 or > 0.5 in BCs, depending on the direction of backcrossing. Multiple combinations of interspecific ancestry are possible for $K > 2$, making the interpretation

of ancestry classes increasingly challenging. Given this, here we treated *J. grandis* and *J. occidentalis* as one ancestral population (hereafter referred to as the western parentals) owing to their much smaller divergence from one another relative to their joint divergence from *J. osteosperma* (see Results, Fig. 1). We ran and assessed the `entropy` MCMC following the same procedure as above. We plotted the means of the posterior distributions for q and Q_{12} for each individual to visualize patterns of ancestry classes within and across hybrid populations. To quantify the relationship between q and Longitude across the study area, we employed a Bayesian beta regression using the *brm* function, and the adjusted R^2 was computed with the *loo_R2* function (brms package; Bürkner, 2017) in `R`.

To visualize genetic structure across populations and species, we conducted principal components analysis (PCA) on the genotype probabilities inferred using the admixture model for the two sets of samples (all individuals and parental taxa/hybrids). Because unbalanced sampling can influence PCA, we randomly sampled 15 hybrids and 15 individuals from each parental taxon for the PCA of all individuals. To estimate genetic differentiation among parental and hybrid populations, we calculated Nei's D (Nei, 1972) and Hudson's F_{ST} (Hudson *et al*., 1992) from population allele frequencies using custom `R` scripts. To visualize genetic distance among parental and hybrid populations, Nei's D estimates were used to construct a neighbor-joining tree with the *nj* function (`ape` package; Paradis & Schliep 2018) in `R`. To assess whether population genetic differentiation was associated with geographic distance within the three parental species, a pattern expected under isolation by distance, we performed multiple regression on distance matrices (MRM) using the *MRM* function in `R` (`ecodist` package; Goslee and Urban, 2007). Geographic distances (km) between populations were calculated from geographic coordinates (see Supplementary Table 1) using the *earth.dist* function (`fossil` package; Vavrek 2011) in `R`

and used as predictors of Nei's D in MRM models, where significance was assessed with 1000 permutations.

We calculated two estimates of genome-wide genetic diversity, Θ_π (the average number of polymorphisms found among pairwise sequence comparisons; Tajima 1983) and Θ_w (the number of segregating sites; Watterson 1975), for each parental and hybrid population using methods that account for genotype uncertainty implemented in `ANGSD` (Korneliussen et al., 2014). First, we estimated the site allele frequency likelihood (SAF) from the BAM files and *de novo* reference assembly using the "-doSaf" option. We estimated the folded site frequency spectrum (SFS) from the SAF using the `realSFS` utility program (Nielsen et al., 2012), and used the SAF and SFS to calculate Θ_π and Θ_w for each site with the "saf2theta" option. Finally, to assess whether the populations are evolving neutrally and exhibiting mutation-drift balance, we calculated Tajima's D (Tajima, 1989) for each site with the `thetaStat` utility program and the "doStat" option (Korneliussen et al., 2013). Briefly, Tajima's D compares two method of moments estimates of genetic diversity, Watterson's estimator (Θ_w) and nucleotide diversity (Θ_π), which are expected to be equal in neutrally evolving populations. We calculated the means and 95% confidence intervals for Θ_π, Θ_w, and Tajima's D across all sites in `R`. A negative value for Tajima's D indicates an abundance of rare alleles caused by population expansion, and a value greater than 0 indicates a scarcity of rare alleles caused by population contraction.

3.4 Ecological divergence among parental taxa and hybrids

To examine the extent of ecological divergence among parental taxa and hybrids, we first conducted a discriminant analysis of principal components (DAPC, Jombart et al., 2010) based on climate variation using distribution-wide occurrence data for each of the three parental taxa and the hybrid zone. For each parental species, we randomly sampled 200 geographic coordinates associated with plot observations from the US Forest Service Forest Inventory and Analysis (USFIA) database (http://www.fia.fs.fed.us/), which includes data for over 900 tree species found in 355,000 plots across the United States. We considered all *J. osteosperma* plot observations west of 118° W longitude as hybrid observations, based on our analyses of population genetic structure and ancestry. We extracted 12 climate variables from PRISM (PRISM Group, 2007) and 13 functionally relevant climate variables from the climatic water deficit toolbox for ArcGIS 10.1 (Dilts and Yang, 2015). To conduct the DAPC, we supplied species and hybrid assignments as *a priori* groupings. The first 10 principal components of the scaled data were analyzed with Linear Discriminant Analysis using the *dapc* function, and the first two discriminant functions were visualized with the *scatter.dapc* function in `R` (`adegenet` package; Jombart 2008).

To assess how climate variation predicts ancestry and genetic variation in our study populations, we obtained climate data from the 25 parental and hybrid populations utilizing the same methods as previously, but including five more climate variables for a total of 30 (Table 1). We first assessed the extent to which climatic variation predicts ancestry across our populations using random forest. Parental and hybrid groups (*J. grandis*, *J. occidentalis*, *J. osteosperma*, hybrid) were predicted based on the known group origin for each individual and climatic data. We used the `randomForest` package in `R` (Liaw and Wiener, 2002) to train the random forest

algorithm with default hyperparameters on a random subset of 70% of the individuals and to assess model performance by predicting the group of origin for the remaining 30% of individuals. We conducted an additional random forest analysis using hybrid admixture proportions to examine the extent to which patterns of ancestry are influenced by climate and geography. For this analysis, we plotted a histogram of $K = 2$ admixture proportions for all parental and hybrid individuals, which formed a trimodal distribution, and we kept all hybrid individuals that fell within the range of the intermediate peak. The resulting hybrid admixture proportions ranged from 0.057 to 0.824. Admixture proportions were predicted by the same set of 30 climate variables, as well as geographic distances between each population and the centroid of the parental ranges. The centroids of the parental ranges were calculated from published range maps for *J. occidentalis* and *J. osteosperma* (Little, 1971). Using the `h2o` package in `R`, the model was tuned with a random sample of 70% of individuals across a range of hyperparameter settings (`mtries`, `min_rows`, `nbins`, and `sample_rate`) with the "RandomDiscrete" search strategy in the *h2o.getGrid* function. We evaluated model performance on the remaining 30% of individuals utilizing the *h2o.performance* function (`h2o` package; LeDell et al., 2022) in `R`.

To evaluate the extent to which climatic variables explain overall patterns of genomic variation across parental and hybrid populations, we utilized redundancy analysis (RDA) (Legendre & Makarenkov, 2002; Capblancq and Forester, 2021). Genotype probabilities inferred for 9,125 SNPs comprised the response matrix. To reduce multicollinearity among the climate variables, we removed variables with the largest mean correlations until all remaining correlations were less than 0.7. The remaining seven climate variables were scaled and used as multivariate predictors. We conducted RDA using the *rda* function (vegan package, Oksanen et

al., 2020) in R, and assessed the significance of the model and each constrained axis with the *anova.cca* and *anova* functions (Legendre et al., 2011).

3.5 Phytochemical variation across species and hybrids

We used a gas chromatography-mass spectrometry (GC-MS) approach to characterize secondary chemistry profiles for nearly all parental and hybrid individuals (314/326). Briefly, silica-dried leaf samples were lysed with a Qiagen TissueLyser II (Qiagen Inc., Valencia, CA, USA). To target terpenoid molecules, we transferred 20 mg of lysed tissue to a 2 dram glass screw cap vial and extracted with 1 mL of hexanes containing n-Eicosane internal standard (0.25 mM). After vortexing for 10 seconds, samples were sonicated in ice water for 15 minutes and centrifuged with a Genevac EZ-2 (SP Scientific, Gardiner, NY, USA) at atmospheric pressure with pre-chilled inserts for 10 minutes. After centrifugation, samples were chilled in a -20 °C freezer for 20 minutes to minimize evaporation during transfer into a GC-MS autosampler vial. Extracts (1 µL) were injected onto an Agilent 7890A gas chromatograph coupled to an Agilent 5975C quadrupole mass spectrometer equipped with an Agilent HP-5MS, (5%-Phenyl)-methylpolysiloxane, 30 m x 250 μm x 0.25 μm capillary column (Agilent Technologies Inc., Santa Clara, CA, USA). A full description of the GC-MS protocol and data processing can be found in the Supplementary Materials.

As preliminary PCA of the terpenoid data demonstrated that individuals from the ZA population (*J. grandis* x *J. occidentalis* hybrids) were outliers and dominated the variance explained by PCs 1 and 2. As a result, these 14 individuals were removed, leaving 300 individuals. As the GC-MS data exhibited strong positive skew (mean = 4.79, sd = 2.82), the data was log2 transformed before ordination and statistical comparison of groups. These analyses

resulted in presence and concentration information for 163 compounds in each of the 300 samples. Of these 163 compounds, 56 (34%) were annotated based on a NIST match (score > 60) with the Adams GCMS library (Adams, 2007), which provides spectra, retention times, and structures for 2205 essential oil components. Some compounds could be attributed to multiple peaks based on their match score. In such cases, the peak having the closest retention index (RI) to the database value was annotated as that compound and others were considered structural analogs of that compound. Three additional compounds had no matches (40 < score < 60) in the Adams library, but were broadly classified according to the structure of their highest match score hit (Supplementary Table 4). The remaining peaks were of insufficient concentration to annotate or had no match in the Adams library (score < 40).

We used PCA to summarize the major axes of chemical variation among individuals. PCA on unscaled data was used as it provided the best resolution between the parental and hybrid groups and did not accentuate compounds with low concentrations and/or variances. To examine whether variation in the first PCs was primarily due to variation in presences or concentrations, we compared PCoAs based on binary distances (qualitative variation) with PCoAs based on Euclidean distances (quantitative variation). A description of the methods utilizing this approach can be found in the Supplementary Methods.

To assess whether terpenoid variation could accurately assign individuals to a parental or hybrid group, we conducted random forest classification with terpenoid variation at 163 compounds. As before (Section 3.4), the random forest algorithm was implemented with the *randomForest* command (`randomForest`, Liaw and Wiener, 2002) in `R`. The number of trees and number of variables tried at each split (m_{try}) was tuned with a random subset representing 70% of the individuals using cross-validation performance. To evaluate model accuracy, we used

the *predict* and *confusionMatrix* (`caret` package; Kuhn, 2021) commands to predict the remaining 30% of the data and summarize the results in `R`.

To determine the independent contributions of climate and genetic variation on phytochemical variation across parental and hybrid populations, we utilized a variance partitioning approach with partial redundancy analysis (pRDA). The response matrix comprised the GC-MS data representing the concentrations of 163 compounds for 300 focal and hybrid individuals. We used two climatic, two genetic, and two geographic predictors in the models: 1) the first two components of a PCA of climate variables which explained 73.74% of total variance; 2) the first two components of a PCA of genotype probabilities which explained 50.56% of the total variance; and 3) geographic coordinates (longitude and latitude) for each population. First, a standard RDA was conducted with all predictors to obtain the total constrained and unconstrained variance. Next, we conducted three pRDAs, where two of the three sets of predictors were conditioned at a time to calculate the individual contributions of climate, genetics, and geography in explaining chemical variation. Similar results were obtained using the *varpart* function (`vegan`; Oksanen et al., 2020) in `R`.

To quantify patterns of phenotype expression in hybrids as dominant, intermediate, or transgressive, we used pairwise Mann Whitney U tests (adjusted for multiple comparisons with the Bonferroni method) to compare the distributions of peak areas among ancestry classes (*J. grandis*, n = 36; *J. occidentalis*, n = 51; *J. osteosperma*, n = 90; hybrids, n = 123) for each compound (n = 163) using the *pairwise.wilcox.test* function in `R`. Though we report the results of the non-parametric Mann Whitney U tests here due to positive skew of the data, one-way ANOVA models and post-hoc Tukey's HSD tests were also conducted and yielded similar results (not shown). In cases where significant differences were observed, hybrid phenotypic

expression was classified as dominant, intermediate, or transgressive. Dominant hybrid expression was further classified into *J. grandis*-like, *J. occidentalis*-like, or *J. osteosperma*-like expression, or in cases where hybrids resembled two parentals they were classified appropriately.

4. Results

4.1 Phylogenetic analyses

Assembly and filtering with `ipyrad` produced a set of 66,280 SNPs for phylogenetic inference. On average, 19,858 loci were recovered for each sample and over half of the loci were present in only 4 or 5 samples. The phylogeny inferred with the multispecies coalescent TETRAD model (Fig. 1) provided 100% bootstrap support for each species lineage, consistent with past work (Uckele et al., 2021). The ingroup (*J. grandis*, *J. occidentalis*, and *J. osteosperma*) was resolved as monophyletic with 100% bootstrap support, as was the serrate leaf juniper clade which included the ingroup as well as *J. arizonica*, *J. californica*, and *J. deppeana*. In a previous analysis (Uckele et al., 2021), *J. californica* was resolved as the most basal member of the ingroup. Here, *J. californica* was resolved with *J. arizonica* and *J. deppeana*, though with low bootstrap support (45%). The ingroup relationships inferred here are similar to those inferred by Uckele et al., (2021), where *J. osteosperma* is basal to *J. grandis* and *J. occidentalis* (Fig. 1).

4.2 Genetic variation across populations, species, and hybrids

Four lanes of sequencing generated 930.7 million raw reads, which were subsequently reduced to 738.6 million reads after contaminant filtering and barcode matching. The mean number of assembled reads per individual was 359,245. After variant calling and stringent

filtering with `HDplot` and `vcftools`, we retained 8,319 and 9,125 SNPs for the data set of all samples and the subset of parental and hybrid individuals, respectively.

The PCA of individuals from all sampled species illustrated subtle differentiation among the ingroup taxa (*J. grandis*, *J. occidentalis*, *J. osteosperma*) and pronounced differentiation between ingroup and outgroup taxa along PC1, and differentiation between three outgroup lineages along PC2 : 1) *J. scopulorum*, 2) *J. californica*, and 3) *J. arizonica* and *J. deppeana* (Fig. 2a). The PCA of parental and hybrid individuals separated *J. osteosperma* from both *J. grandis* and *J. occidentalis* along PC1 (50.611% explained variance), and separated *J. grandis* from *J. occidentalis* along PC2 (3.441% explained variance) (Fig. 2b). Hybrids were not tightly clustered and formed a continuous bridge between parental species clusters (Fig. 2b). The PCAs of each parental taxa overall explained much less variance and exhibited less clustering (Fig. 2c-e), however, some evidence for genetic structure was observed in *J. grandis* (Fig. 2d), though inference is limited by the lack of population sampling in this parental species. Consistent with a lack of genetic structuring with PCA, MRM models of genetic distance (Nei's D) by geographic distance were insignificant for each parental taxa (Fig. 2f-h).

Consistent with the maintenance of species boundaries and low levels of population genetic structure, interspecific values of F_{ST} were 4.55 times greater on average (mean = 0.091, s = 0.022) than intraspecific values (mean = 0.02, s = 0.004) (Fig. 3). Genetic differentiation was more pronounced among *J. osteosperma* and both *J. occidentalis* (mean = 0.111, s = 0.003) and *J. grandis* (mean = 0.087, s = 0.002) than between *J. grandis* and *J. occidentalis* (mean = 0.056, s = 0.004). Similar to the phylogeny and patterns illustrated with PCA, the neighbor-joining tree constructed with Nei's D genetic distances infers a close relationship among populations of *J.*

grandis and *J. occidentalis*, and most hybrid populations form an evolutionary grade between *J. osteosperma* and the *J. grandis* - *J. occidentalis* clade (Fig. 4a).

Genetic diversity measures (Θ_π and Θ_w) were highly correlated (r = 0.89) and high across populations (Supplementary Table 2). Genetic diversity was approximately 4% higher within hybrids (Θ_π = 0.0117, SD = 0.0003; Θ_w = 0.0134, SD = 0.0011) than in *J. osteosperma* (Θ_π = 0.0112, SD = 0.0002; Θ_w = 0.0128, SD = 0.0008), 9% higher in hybrids than in *J. occidentalis* (Θ_π = 0.0107, SD = 0.0001; Θ_w = 0.0107, SD = 0.0003), and 13% higher in hybrids than in *J. grandis* (Θ_π = 0.0104, SD = 0.0004; Θ_w = 0.0105, SD = 0.0005) (Supplementary Table 2). Mean per-site Tajima's D was negative for every population except UA (*J. occidentalis*) (Supplementary Table 3). Tajima's D was lowest in *J. osteosperma* (-0.4415, SD = 0.1104) and hybrids (-0.4339, SD = 0.1239), and approximately 2.7 and 3.7 times lower than in *J. grandis* (-0.166, SD = 0.0588) and *J. occidentalis* (-0.1189, SD = 0.0979), respectively (Supplementary Table 3).

The $K = 3$ `entropy` model assigned the three parental taxa as ancestral populations. Under this model, most hybrids exhibited ancestry from all three parental lineages; however, individuals from the most western hybrid population (population ZA) exhibited ancestry predominantly from *J. grandis* and *J. occidentalis* (Fig. 4a). Population mean ancestries varied geographically, with western hybrid populations exhibiting nearly equal proportions of ancestry from the three parental species, and eastern hybrid populations exhibiting a greater proportion of *J. osteosperma* ancestry (Fig. 4b). Similarly, northern hybrid populations possessed a higher proportion of *J. occidentalis* ancestry, whereas southern hybrid populations possessed a higher proportion of *J. grandis* ancestry (Fig. 4b). We estimated interspecific ancestry (Q) using a K = 2

model because interpretation of such estimates is difficult for $K = 3$ models, and because *J. occidentalis* and *J. grandis* exhibit much less divergence from one another than from *J. osteosperma*. In other words, the $K = 2$ model recognized *J. grandis*/*J. occidentalis* as one ancestry and *J. osteosperma* as another, while still representing the salient features of divergence and admixture. Plotting individual admixture proportions (q) by estimates of interspecific ancestry (Q) under a $K = 2$ entropy model revealed a lack of late-generation hybrids and a preponderance of F_1 and backcrossed hybrids (Fig. 4c). Approximately 37% of hybrids were backcrossed to *J. osteosperma*, 25% were backcrossed to the western parental lineage, 24% were F_1 hybrids, 11% were *J. grandis* x *J. occidentalis* hybrids, and only 3% were advanced generation hybrids. We detected a longitudinal cline in estimates of q and Q with a sharp transition from western parental to hybrid ancestry from approximately 123 to 119.5° W, and a more gradual transition to *J. osteosperma* ancestry from approximately 119.5 to 109° W (Fig. 4d). The highest variation in hybrid ancestries (q and Q) was observed at the point of inflection (~119.5° W), and featured F_1, late-generation, and BC hybrids to the western parentals. Alternatively, BC hybrids to *J. osteosperma* were observed east of this point of inflection, at a larger range of longitudes (Fig. 4d).

4.3 Ecological divergence among parental taxa and hybrids

DAPC on the range-wide climatic data (200 plot occurrences for each species) assuming *a priori* taxon groups resolved distinct environmental variation characterizing the parental groups and hybrids occupying intermediate environmental space (Fig. 5a). The first discriminant function separated the *J. grandis* and *J. osteosperma* clusters, with the hybrid and *J. occidentalis* clusters as intermediate. The second discriminant function separated *J. occidentalis* from the

other two parental clusters, but there remained substantial overlap between the *J. occidentalis* and hybrid clusters (Fig. 5a). Monsoonality, minimum summer temperature, seasonality of precipitation, and the fraction of AET from monthly precipitation were most predictive of species and hybrid range-wide distributions in the DAPC analysis.

We then used complementary analyses to quantify environmental predictors of genetic variation among the populations sampled for this study. First we conducted a random forest classification model of ancestry class predicted by climatic variables. This model (RF_a, Table 1) exhibited nearly perfect classification accuracy (95% CI: 96.48 - 100), suggesting that climate variables are highly predictive of ancestry class. The five most important variables with the highest mean decrease in accuracy (MDA) were the fraction of actual evapotranspiration from month's precipitation rather than soil water, summer precipitation, cumulative potential evapotranspiration, cumulative climatic water deficit, and monsoonality (Table 1). Next, we conducted random forest regression (RF_b, Table 1) predicting the admixture proportions of hybrid individuals as a function of distances to each parental range and climatic variables. The best model explained 86% of the variance in the test data, and the most important variables were winter precipitation and seasonality of precipitation, representing 39.82% and 35.38% of the total variable importance respectively (Table 1).

To complement the random forest classification model, we conducted an RDA of parental and hybrid genetic variation explained by climatic variation, with the genotype matrix as the response. The first two RDA axes explained a majority of the constrained variance and produced distinct parental species clusters with hybrids intermediate between them (Fig. 5b). Both *J. grandis* and *J. occidentalis* were associated with higher values of cumulative water supply and seasonality of precipitation, whereas *J. osteosperma* was associated with higher

values of pdaetswb, the fraction of AET from the month's precipitation rather than from soil water, which is highly correlated with summer precipitation (r = 0.71) and monsoonality (r = 0.79). Hybrids were associated with intermediate values of these climatic variables. The third RDA axis depicted genetic variation within hybrids associated with higher values of unmet demand for water, maximum winter temperature, and minimum annual temperature (Fig. 5c).

4.4 Phytochemical variation across species and hybrids

Integration and correspondence of GC-MS data yielded 163 peak bins (putative terpenoid compounds). Of these, 55 were matched to known molecules and included one aliphatic, one phenylpropanoid, 15 monoterpenes, 19 diterpenes (2 bicyclic, 17 tricyclic), and 19 sesquiterpenes (6 monocyclic, 11 bicyclic, and 2 tricyclic) (Supplementary Table 4). Of the 163 compounds, 109 (67%) exhibited differentiation in presence or concentration across species and hybrids. Of those, 49 (45%) were dominant towards one or two parental taxa, 30 (27%) were transgressively high, 27 (25%) were intermediate, and 3 (3%) were transgressively low in hybrids. The first PC axis of terpenoid chemistry (12.73% explained variance) differentiated *J. grandis* and *J. occidentalis*, and captured transgressive variation in hybrids, i.e., concentrations in hybrids which exceeded those of the parental taxa (Fig 6a,b). The second PC axis (8.21% explained variance) represented terpenoid variation that was intermediate in hybrids and distinct in *J. osteosperma* individuals (Fig. 6a,e). Principal coordinate analyses of Euclidean and binary chemical distances indicated that PC1 is driven more by variation in terpenoid concentration, while PC2 is driven more by variation in terpenoid composition (presences and absences) (Supplementary Figure 1). The hybrid group exhibited the highest variation in compound concentrations, followed by *J. osteosperma* (Supplementary Results). Of the compounds that

loaded most significantly onto PCs 1 and 2, the majority were tricyclic diterpenes (Supplementary Table 4). The compounds associated with PC1 were transgressive or dominant towards *J. osteosperma* in hybrids, whereas the compounds associated with PC2 were predominantly intermediate in hybrids (Fig. 6).

The overall classification accuracy of random forest was 92.55% (95% CI: 85.26 - 96.95). Model sensitivity was lowest for *J. osteosperma* (81.48%) and *J. grandis* (90%) and highest for *J. occidentalis* (100%) and hybrids (97.62%). Model specificity was lowest for hybrids (88.46%), high for *J. osteosperma* (98.51%), and perfect for *J. grandis* and *J. occidentalis*. When applied to the test data subset, random forest misclassified one *J. grandis* as hybrid, one hybrid as *J. osteosperma*, and five *J. osteosperma* as hybrids. 15 of the 18 terpenoid compounds with higher than average loadings on PCs 1 and 2 were also rated as important variables by random forest (Supplementary Table 4).

We used RDA-based variance partitioning to disentangle the effects of climate and genetic variation on terpenoid variation across hybrids and species. The explanatory variables together (genetic, geographic, and climate) explained 11.92% of the variance in the GC-MS data (Table 2). Variance partitioning resulted in the largest proportion of variance (3.28%) explained by genetics, the second largest by geography (2.77%) and the smallest by climate (1.28%). The remaining 4.59% of the explained variance was confounded, meaning the individual effects of genetics, geography, and environment could not be disentangled (Table 2).

5. Discussion

5.1 Genetic variation across populations, species, and hybrids

Hybridization is increasingly recognized as an important evolutionary process that has influenced diversification and challenged phylogenetic inference across diverse plant lineages (Rieseberg and Soltis, 1991; Folk et al., 2017). Early phylogenetic analyses of the western North American serrate leaf juniper clade based on small numbers of nuclear and chloroplast loci (Adams et al., 2006) suggested several paradoxical relationships that may have arisen due to hybridization and ancient chloroplast capture (Uckele et al., 2021). For example, previous analyses based in large part on chloroplast DNA (Adams et al., 2006; Adams and Schwarzbach, 2013) resolved *J. grandis* and *J. osteosperma* as sister taxa despite extensive morphological evidence suggesting instead that *J. grandis* and *J. occidentalis* are sister taxa. Our analysis, similar to Uckele et al., (2021), was based on thousands of nuclear loci and inferred a highly-supported sister relationship between *J. grandis* and *J. occidentalis* (Fig. 1). Relationships among the parental taxa were resolved with 100% bootstrap support (Fig. 1), suggesting strong phylogenetic concordance among nuclear loci despite ongoing contemporary hybridization.

Despite often spanning large geographic areas, forest trees commonly exhibit low levels of population genetic differentiation (Petit and Hampe 2006; Neale 2007). Consistent with high gene flow across populations, the juniper parental species exhibited low levels of population genetic differentiation that were not associated with geographic distance. Evidence for isolation by distance was notably lacking (Fig. 2f-h), even among *J. osteosperma* populations spanning nearly one thousand kilometers (Fig. 2f). In contrast, species boundaries were well-defined despite gene flow among them, and substantive differentiation between *J. grandis* and *J. occidentalis* supported their evolutionary independence (Fig. 3). Similar to other North American trees, levels of genetic diversity were high within parental and hybrid populations (Menon et al., 2018; Acosta et al., 2019; Haselhorst et al., 2019) (Supplementary Table 2).

Genetic diversity in the form of standing genetic variation is increasingly recognized as an important predictor of adaptive potential to environmental change and may be particularly important in forest trees, which are sessile, long-lived, and geographically widespread (De Carvalho et al., 2010). For temperate plants in particular, high levels of genetic diversity may also facilitate persistence on the landscape in response to long- and short-term climatic variations that characterized the late Neogene (Barnosky, 1987). Genetic diversity was highest in hybrids and *J. osteosperma* (Supplementary Table 2), signaling the importance of admixture, large population size, and broad environmental tolerance in generating and maintaining genetic diversity. *Juniperus osteosperma* occupies the largest geographic area and elevational range of the three parentals, and fossil evidence suggests that while elevational shifts were prominent, the geographic extent of *J. osteosperma* during the last glacial maximum remained constant in the Great Basin Desert and even expanded south into the Mojave Desert, southern Sierras, southern Central Valley, and Sonoran Desert (Thompson 1990; Nowak et al., 1994). In contrast, *J. grandis* and *J. occidentalis* did not inhabit their contemporary geographic ranges until the early Holocene, and likely faced greater geographic displacement and range contraction in response to ice-age oscillations (Cole, 1983; Nowak et al., 1994). The Holocene brought warmer, arid conditions which caused widespread decline of juniper woodlands until the late Holocene. Cooler, wetter conditions of the Neoglaciation period marked the end of woodland decline and prompted vigorous juniper growth and downslope expansion (Wigand et al., 1995). This growth is reflected in population values of Tajima's D, which were negative for every population except UA, and significantly more negative for *J. osteosperma* (Table 3).

5.2 Patterns of admixture in secondary contact zone

Among junipers, species boundaries are often permeable (Adams, 1994; Adams 2016; Adams et al., 2016; Adams et al., 2017; Adams et al., 2018; Adams et al., 2020; Farhat et al., 2020), suggesting an important role of hybridization in the evolution of the genus. Secondary contact among the parental species was likely established by the early Holocene, when the western parentals occupied their contemporary distributions and milder climates allowed for downslope expansion (Cole, 1983; Nowak et al., 1994). Previous genetic analyses provided preliminary evidence for hybridization and introgression into the *J. osteosperma* background (Terry et al., 2000), but were limited by small numbers of genetic markers. Utilizing a ddRADseq approach, we obtained a substantial improvement in genomic sampling and resolution of hybrid ancestry. Hybrids possessed ancestry from three distinct parental lineages, and admixture proportions spanned the range of possible values from 0 to 1. Only one population that we sampled (ZA) resulted from hybridization between *J. grandis* and *J. occidentalis*, however, the parental *J. grandis* and *J. occidentalis* populations appear to be partially admixed as well, suggesting that the transition from the Cascade to the Sierra Nevada range may represent another contact zone where interspecific gene flow occurs (Fig. 4a). The transition from western parental ancestry to *J. osteosperma* ancestry forms a longitudinal cline from west to east, and the inflection of this cline occurs within the ecotone marking the transition from the Cascade-Sierra mountains to the Great Basin Desert (Fig. 4d). Similar to other studies (Adams, 2013a,b), we found that parental individuals were rare within the hybrid zone (Fig. 4a). The majority of hybrids were F_1s (27%) or BCs (70%) and only four individuals were F_2 or later-generation hybrids (Fig. 4c). This deficit of F_N hybrids despite high densities of presumably fertile F_1 hybrids is surprising, and potentially suggests a reduction in recombinant fitness in the F_2 generation (Burke and Arnold, 2001) or highly asymmetrical gene flow from

outside to inside the hybrid zone (Barton and Hewitt, 1985). Backcrossing occurred in both directions, however, gene flow into the *J. grandis*/*J. occidentalis* background was primarily restricted to the zone of secondary contact in western Nevada, whereas gene flow towards *J. osteosperma* extended well into central Nevada (Fig. 4b,d). This pattern is consistent with previous analyses (Terry et al., 2000) and with prevailing wind patterns in western Nevada, which are strongly asymmetrical favoring west to east. Wind strength and direction can shape landscape genetic patterns in trees, and have been shown to be most important for wind-pollinated trees like conifers (Kling and Ackerly, 2021).

Recent studies have uncovered strong genetic-environmental associations as putative evidence for environmental selection against hybrids (Menon *et al*., 2018; Huang et al., 2019; Wang et al., 2020). We found that climatic variation perfectly predicted whether each population belonged to a parental or hybrid group. The most important predictor in this analysis was the positive difference between actual evapotranspiration and soil water balance (Table 1), which is highest in areas occupied by *J. osteosperma* where precipitation, rather than soil water, is an important source of moisture during the growing season. This variable is also highly correlated with monsoonality and summer precipitation (Supplementary Figure 3), both of which are higher in regions occupied by *J. osteosperma* that receive a greater percentage of rainfall during the summer months from southerly, subtropical sources. Hybrid climatic variation was consistently intermediate to that of *J. grandis* and *J. osteosperma* (Fig. 5b), where the precipitation regime of the former is broadly characterized as wet and winter-dominated and the latter as dry and monsoonal. One exception was cumulative water supply, which was lower for hybrid populations east of the contact zone that are most severely impacted by the Sierra Nevada rain shadow (Fig. 5c). Furthermore, we found that climatic variation and the

geographical proximity to parental ranges provided highly accurate predictions of admixture within the hybrid zone, and winter precipitation and seasonality of precipitation were most important to this prediction (Table 1). Winter precipitation and seasonality of precipitation are both highest in areas inhabited by *J. grandis* where the greatest input of annual precipitation occurs during the winter as snow. Other important variables quantified the amount of excess moisture available for soil water storage, runoff, or deep percolation (Table 1), thus reinforcing that variation in the timing and amount of precipitation is highly influential in shaping patterns of admixture within the hybrid zone.

5.3 Phytochemical variation across species and hybrids

Conifers produce an astounding number of terpenoids with key roles in defense and communication (Keeling and Bohlmann, 2006). Terpenoid variation is shaped by both biotic and abiotic factors, including herbivores, pathogens, nutrients, light, and water availability, as well as having a substantial genetic component (Moore et al., 2014). The contributions of genetic, climatic, and geographic variation were highly confounded but together explained 12% of the variation in terpenoid composition and concentrations across species and hybrids (Table 2). Genetic variation had the largest individual effect on terpenoid variation, and climatic variation had the smallest, consistent with a strong genetic basis of terpenoid biosynthesis (Table 2). Principal components analysis of terpenoid variation resolved discrete parental and hybrid groups (Fig. 6a), and prediction of these groups with random forest performed well, suggesting that ancestry is a significant predictor of parental and hybrid chemistry. Similar to others (Moore et al., 2014), our analyses were unable to explain the majority of terpenoid variation, which may be attributed to sampling error, unmeasured biotic and abiotic variation, or biosynthetic noise.

Large gene families, namely terpenoid synthases and cytochrome P450-dependent monooxygenases, are responsible for the impressive chemical diversity found in conifer oleoresin. As many terpenoid synthases and P450 enzymes are multi-product and multi-substrate, respectively, small changes in expression can have large effects on phenotype (Pichersky and Raguso, 2018). As a result, hybridization can disrupt biosynthetic networks such that compounds are expressed in novel concentrations or tissues and exert unique selective pressures on herbivore communities. Hybrid terpenoid chemistry was differentiated from parental chemistry along the first two principal component axes, with each axis representing a different aspect of novel hybrid variation: transgressive concentrations on the first, and intermediate concentrations on the second (Fig. 6). Transgressive segregation appears more frequently when parental species are phenotypically similar (Stelkens and Seehausen, 2009), and is believed to occur because of the complementary action of additive alleles, specifically, the fixation of alternative antagonistic alleles in independent lineages experiencing stabilizing selection (Kim and Rieseberg, 1999). Consistent with this idea, the terpenoid variation of *J. grandis* and *J. osteosperma* overlapped considerably along the first principal component axis that was also characterized by transgressive inheritance (Fig. 6b). In contrast, the parental species are phenotypically dissimilar along the second axis, where intermediate inheritance in hybrids prevails (Fig. 6e). The most important compounds for predicting parental and hybrid groups with random forest also loaded strongly onto the first two principal component axes and were largely represented by tricyclic diterpenes (Supplementary Table 4), of which a substantial number were biosynthetically related to diterpene acids. Diterpene acids are actively involved in herbivore resistance (Tomlin et al., 1996) and have negative, dose-dependent effects on herbivore feeding (Powell and Raffa, 1999) and growth rates of bark beetle fungal associates (Mason et al., 2015),

yet relatively little is known about variation in their concentrations and sources of this variation. We found that the concentrations of these diterpenes were distinct among the parental species, and exhibited novel intermediate and transgressive concentrations in hybrids (e.g., Fig. 6c,d,f,). Diterpenes and sesquiterpenes were overrepresented in our analyses because many of the more volatile monoterpenes were lost during tissue storage and extraction. This aspect precluded thorough investigation of monoterpenoid variation, however, numerous studies (Adams 2013a, Adams 2013b, more) have investigated monoterpenoid content and variation in these parental species and hybrids, while few have addressed di- and sesquiterpenoid variation despite it being an important component of conifer oleoresin.

6. Conclusion

Our reduced representation approach generated the largest SNP data set applied in *Juniperus* to date, provided improved resolution of evolutionary history and hybrid admixture, and facilitated an understanding of phytochemical diversity spanning a full gradient of ancestry. Despite hybridization, our data fully resolved the evolutionary relationships among the parental species and was consistent with previous studies indicating that *J. grandis* and *J. occidentalis* are evolutionarily independent taxa (Adams et al., 2006; Adams and Schwarzbach, 2013; Uckele et al., 2021). Our population genetic analyses further demonstrated clear divergence between *J. grandis, J. occidentalis*, and *J. osteosperma* and permitted novel insight into admixture among three evolutionarily distinct lineages. Most hybrids exhibited ancestry from all three parental species, and variation in ancestry was predicted by the geographical proximity to parental ranges and environmental variation across the landscape. This study contributes to an accumulating body of work documenting interfertility among members of *Juniperus* (Palma-

Otal et al., 1983; Adams and Kistler, 1991; Terry et al., 2000; Adams, 2015; Adams et al., 2017), which collectively suggest that hybridization has been a prominent process in the evolution of the genus, as it has for other tree genera (e.g., Bouillé et al., 2010; Grattapaglia et al., 2012; Cannon and Petit, 2020). Our results also demonstrate that hybrid admixture has a significant effect on terpenoid variation, which was highly variable in hybrids and distinct from the parentals. We observed transgressive variation at a number of terpenoids with well-documented antiherbivore and antifungal effects in other conifers. In many tree systems, hybrid genetic and phytochemical variation exhibits extended effects on ecological community structure and ecosystem processes (Driebe and Whitham, 2000; Dungey et al., 2000; Wimp et al., 2007; Jarvis et al., 2017). Our study provides preliminary evidence that hybridization and its effect on functional trait variation may be consequential for species interactions and diversity within a juniper hybrid zone.

7. Acknowledgements

K.A.U. was supported by the National Science Foundation Graduate Research Award (Award No. 1650114), J.P.J. by the Modelscape Consortium with funding from NSF (OIA-2019528), and R.P.A by Baylor University (Project No. 032512). The authors declare no conflict of interest.

8. References

Abrahamson, W.G., Hunter, M.D., Melika, G. and Price, P.W., 2003. Cynipid gall-wasp communities correlate with oak chemistry. *Journal of Chemical Ecology*, *29*(1), pp.209-223.

Acosta, J.J., Fahrenkrog, A.M., Neves, L.G., Resende, M.F., Dervinis, C., Davis, J.M., Holliday, J.A. and Kirst, M., 2019. Exome resequencing reveals evolutionary history, genomic diversity, and targets of selection in the conifers *Pinus taeda* and *Pinus elliottii*. *Genome biology and evolution*, *11*(2), pp.508-520.

Adams, R.P., Schwarzbach, A.E. and Tashev, A.N., 2016. Chloroplast capture in *Juniperus sabina* var. *balkanensis* RP Adams and AN Tashev, from the Balkan peninsula: A new variety with a history of hybridization with *J. thurifera*. *Phytologia*, *98*(2) pp.219-231.

Adams, R.P., 2016. Two new cases of chloroplast capture in incongruent topologies in the *Juniperus excelsa* complex: *J. excelsa* var. *turcomanica* comb. nov. and *J. excelsa* var. *seravschanica* comb. nov. *Phytologia*, *98*(3), pp.219-231.

Adams R.P., Schwarzbach A.E., 2013. Taxonomy of the serrate leaf *Juniperus* of North America: Phylogenetic analyses using nrDNA and four cpDNA regions. *Phytologia*, *95*(2), pp.172–178.

Adams R.P., Nguyen S., Morris J.A., Schwarzbach A.E., 2006. Re-examination of the taxonomy of the one-seeded, serrate leaf *Juniperus* of southwestern United States and northern Mexico (Cupressaceae). *Phytologia*, *88*(3), pp.299–308.

Adams, R.P. and Kistler, J.R., 1991. Hybridization between *Juniperus erythrocarpa* Cory and *Juniperus pinchotii* Sudworth in the Chisos mountains, Texas. *The Southwestern Naturalist*, *36*(3), pp.295-301.

Adams, R.P., Socorro González-Elizondo M., González-Elizondo M., Ramirez Noy D., Schwarzbach A.E., 2017. DNA sequencing and taxonomy of unusual serrate *Juniperus* from Mexico: Chloroplast capture and incomplete lineage sorting in *J. coahuilensis* and allied taxa. *Phytologia*, *99*(1), pp.62-73.

Adams, R.P., Johnson, S., Coombes, A.J., Caamaño, L., González-Elizondo, M.S., 2018. Preliminary examination of hybridization and introgression between *Juniperus flaccida* and *J. poblana*: nrDNA and cpDNA sequence data. *Phytologia*, *100*(2), pp.145-152.

Adams, R.P., Johnson, S.T., Schwarzbach, A.E., 2020. Long distance gene flow facilitated by bird-dispersed seeds in wind-pollinated species: A story of hybridization and introgression between *Juniperus ashei* and *J. ovata* told by nrDNA and cpDNA. *Phytologia*, *102*(2), pp.55-74.

Adams, R.P., 1994. Geographic variation and systematics of monospermous *Juniperus* (Cupressaceae) from the Chihuahua Desert based on RAPDs and terpenes. *Biochemical Systematics and Ecology*, *22*(7), pp.699-710.

Adams, R.P., 2007. Identification of essential oil components by gas chromatography/mass spectrometry (Vol. 456, pp. 544-545). Carol Stream: Allured publishing corporation.

Adams R.P., 2013a. Hybridization between *Juniperus grandis*, *J. occidentalis* and *J. osteosperma* in northwest Nevada I: Terpenes, Leviathan Mine, Nevada. *Phytologia*, *95*(1), pp.58-69.

Adams R.P., 2013b. Hybridization between *Juniperus grandis*, *J. occidentalis* and *J. osteosperma* in northwest Nevada II: Terpenes, Buffalo Hills, northwestern Nevada. *Phytologia*, *95*(1), pp.107-114.

Adams, R.P., 2015. Allopatric hybridization and introgression between *Juniperus maritima* RP Adams and *J. scopulorum* Sarg.: evidence from nuclear and cpDNA and leaf terpenoids. *Phytologia*, *97*(1), pp.55-66.

Alix, K., Gérard, P.R., Schwarzacher, T. and Heslop-Harrison, J.S., 2017. Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Annals of Botany*, *120*(2), pp.183-194.

Araujo, S.B., Braga, M.P., Brooks, D.R., Agosta, S.J., Hoberg, E.P., von Hartenthal, F.W. and Boeger, W.A., 2015. Understanding host-switching by ecological fitting. *PLoS One*, *10*(10), p.e0139225.

Arnold, M.L., 2004. Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology*, *13*(5), pp.997-1007.

Baack, E.J. and Rieseberg, L.H., 2007. A genomic view of introgression and hybrid speciation. *Current Opinion in Genetics & Development*, *17*(6), pp.513-518.

Bangert, R.K., Turek, R.J., Martinsen, G.D., Wimp, G.M., Bailey, J.K. and Whitham, T.G., 2005. Benefits of conservation of plant genetic diversity to arthropod diversity. *Conservation Biology*, *19*(2), pp.379-390.

Barker, H.L., Riehl, J.F., Bernhardsson, C., Rubert-Nason, K.F., Holeski, L.M., Ingvarsson, P.K. and Lindroth, R.L., 2019. Linking plant genes to insect communities: Identifying the genetic bases of plant traits and community composition. *Molecular Ecology*, *28*(19), pp.4404-4421.

Barnosky, C.W., 1987. Response of vegetation to climatic changes of different duration in the late Neogene. *Trends in Ecology & Evolution*, *2*(8), pp.247-250.

Barton N.H., Hewitt G.M., 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), pp.113-148.

Betancourt, J.L., Van Devender, T.R. and Martin, P.S. eds., 1990. *Packrat middens: the last 40,000 years of biotic change*. University of Arizona Press.

Bouillé, M., Senneville, S. and Bousquet, J., 2011. Discordant mtDNA and cpDNA phylogenies indicate geographic speciation and reticulation as driving factors for the diversification of the genus *Picea*. *Tree Genetics & Genomes*, *7*(3), pp.469-484.

Buerkle C.A., Lexer C., 2008. Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution*, *23*(12), pp.686-694.

Burke, J.M. and Arnold, M.L., 2001. Genetics and the fitness of hybrids. *Annual review of genetics*, *35*(1), pp.31-52.

Bürkner, P.C., 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software*, *80*, pp.1-28.

Burton, R.S., Pereira, R.J. and Barreto, F.S., 2013. Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, *44*, pp.281-302.

Cannon, C.H. and Petit, R.J., 2020. The oak syngameon: more than the sum of its parts. *New Phytologist*, *226*(4), pp.978-983.

Capblancq, T. and Forester, B.R., 2021. Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), pp.2298-2309.

Caseys, C., Glauser, G., Stölting, K.N., Christe, C., Albrectsen, B.R. and Lexer, C., 2012. Effects of interspecific recombination on functional traits in trees revealed by metabolomics and genotyping-by-resequencing. *Plant Ecology & Diversity*, *5*(4), pp.457-471.

Cheng, D., Vrieling, K. and Klinkhamer, P.G., 2011. The effect of hybridization on secondary metabolites and herbivore resistance: implications for the evolution of chemical diversity in plants. *Phytochemistry Reviews*, *10*(1), pp.107-117.

Chifman, J. and Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), pp.3317-3324.

Chifman, J. and Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of theoretical biology*, *374*, pp.35-47.

Cole, K., 1983. Late Pleistocene vegetation of Kings Canyon, Sierra Nevada, California. *Quaternary research*, *19*(1), pp.117-129.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T. and McVean, G., 2011. The variant call format and VCFtools. *Bioinformatics*, *27*(15), pp.2156-2158.

De Carvalho, D., Ingvarsson, P.K., Joseph, J., Suter, L., Sedivy, C., Macaya-Sanz, D., Cottrell, J., Heinze, B., Schanzer, I. and Lexer, C., 2010. Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Molecular Ecology*, *19*(8), pp.1638-1650.

Dilts, T.E. and Yang, J., 2015. Climatic water deficit toolbox for ArcGIS 10.1. *University of Nevada, Reno, NV, USA*.

Dobzhansky, T. and Dobzhansky, T.G., 1971. *Genetics of the evolutionary process* (Vol. 139). Columbia University Press.

Driebe, E.M. and Whitham, T.G., 2000. Cottonwood hybridization affects tannin and nitrogen content of leaf litter and alters decomposition. *Oecologia*, *123*(1), pp.99-107.

Dungey, H.S., Potts, B.M., Whitham, T.G. and Li, H.F., 2000. Plant genetics affects arthropod community richness and composition: evidence from a synthetic eucalypt hybrid population. *Evolution*, *54*(6), pp.1938-1946.

Eaton, D. A., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic biology*, *66*(3), 399-412.

Ellstrand, N.C. and Schierenbeck, K.A., 2000. Hybridization as a stimulus for the evolution of invasiveness in plants?. *Proceedings of the National Academy of Sciences*, *97*(13), pp.7043-7050.

Erbilgin, N., 2019. Phytochemicals as mediators for host range expansion of a native invasive forest insect herbivore. *New Phytologist*, *221*(3), pp.1268-1278.

Esquerré, D., Keogh, J.S., Demangel, D., Morando, M., Avila, L.J., Sites Jr, J.W., Ferri-Yáñez, F. and Leaché, A.D., 2022. Rapid radiation and rampant reticulation: Phylogenomics of South American Liolaemus lizards. *Systematic Biology*, *71*(2), pp.286-300.

Evans, L.M., Allan, G.J., Shuster, S.M., Woolbright, S.A. and Whitham, T.G., 2008. Tree hybridization and genotypic variation drive cryptic speciation of a specialist mite herbivore. *Evolution: International Journal of Organic Evolution*, *62*(12), pp.3027-3040.

Falush, D., Stephens, M. and Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), pp.1567-1587.

Farhat, P., Takvorian, N., Avramidou, M., Garraud, L., Adams, R.P., Siljak-Yakovlev, S., Kharrat, M.B.D. and Robert, T., 2020. First evidence for allotriploid hybrids between *Juniperus thurifera* and *J. sabina* in a sympatric area in the French Alps. *Annals of Forest Science*, *77*(4), pp.1-17.

Floate, K.D. and Whitham, T.G., 1993. The" hybrid bridge" hypothesis: host shifting via plant hybrid swarms. *The American Naturalist*, *141*(4), pp.651-662.

Floate, K.D., Godbout, J., Lau, M.K., Isabel, N. and Whitham, T.G., 2016. Plant–herbivore interactions in a trispecific hybrid swarm of *Populus*: assessing support for hypotheses of hybrid bridges, evolutionary novelty and genetic similarity. *New Phytologist*, *209*(2), pp.832-844.

Folk, R.A., Mandel, J.R. and Freudenstein, J.V., 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic biology*, *66*(3), pp.320-337.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), pp.3150-3152.

García, C., Guichoux, E. and Hampe, A., 2018. A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. *turbinata*). *Tree Genetics & Genomes*, *14*(6), pp.1-9.

Gompert, Z., Lucas, L.K., Buerkle, C.A., Forister, M.L., Fordyce, J.A. and Nice, C.C., 2014. Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular ecology*, *23*(18), pp.4555-4573.

Goslee, S.C. and Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, pp.1-19.

Gottfried, G.J., Swetnam, T.W., Allen, C.D., Betancourt, J.L. and Chung-MacCoubrey, A.L., 1995. Pinyon-juniper woodlands. *United States Department of Agriculture Forest Service General Technical Report RM*, pp.95-132.

Grattapaglia, D., Vaillancourt, R.E., Shepherd, M., Thumma, B.R., Foley, W., Külheim, C., Potts, B.M. and Myburg, A.A., 2012. Progress in Myrtaceae genetics and genomics: Eucalyptus as the pivotal genus. *Tree Genetics & Genomes*, *8*(3), pp.463-508.

Haselhorst, M.S., Parchman, T.L. and Buerkle, C.A., 2019. Genetic evidence for species cohesion, substructure and hybrids in spruce. *Molecular ecology*, *28*(8), pp.2029-2045.

Huang, K., Andrew, R.L., Owens, G.L., Ostevik, K.L. and Rieseberg, L.H., 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, *29*(14), pp.2535-2549.

Hudson, R.R., Slatkin, M. and Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics*, *132*(2), pp.583-589.

Jarvis, K.J., Allan, G.J., Craig, A.J., Beresic-Perrins, R.K., Wimp, G., Gehring, C.A. and Whitham, T.G., 2017. Arthropod communities on hybrid and parental cottonwoods are phylogenetically structured by tree type: Implications for conservation of biodiversity in plant hybrid zones. *Ecology and evolution*, *7*(15), pp.5909-5921.

Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), pp.1403-1405.

Jombart, T., Devillard, S. and Balloux, F., 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), pp.1-15.

Kay, K.M., 2006. Reproductive isolation between two closely related hummingbird pollinated neotropical gingers. *Evolution*, *60*(3), pp.538-552.

Keeling, C.I. and Bohlmann, J., 2006. Genes, enzymes and chemicals of terpenoid diversity in the constitutive and induced defence of conifers against insects and pathogens. *New Phytologist*, *170*(4), pp.657-675.

Kim, S.C. and Rieseberg, L.H., 1999. Genetic architecture of species differences in annual sunflowers: implications for adaptive trait introgression. *Genetics*, *153*(2), pp.965-977.

Kling, M.M. and Ackerly, D.D., 2021. Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Sciences*, *118*(17).

Korneliussen, T.S., Moltke, I., Albrechtsen, A. and Nielsen, R., 2013. Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC bioinformatics*, *14*(1), pp.1-14.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, *15*(1), pp.1-13.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R.C. and Benesty, M., 2021. Caret: Classification and Regression Training [Internet]. *R package version*.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L., 2009. Bowtie: An ultrafast memory-efficient short read aligner. *Genome Biol*, *10*(3), p.R25.

LeDell, E. and Poirier, S., 2020, July. H2o automl: Scalable automatic machine learning. In *Proceedings of the AutoML Workshop at ICML* (Vol. 2020).

Legendre, P. and Makarenkov, V., 2002. Reconstruction of biogeographic and evolutionary networks using reticulograms. *Systematic Biology*, *51*(2), pp.199-216.

Legendre, P., Oksanen, J. and ter Braak, C.J., 2011. Testing the significance of canonical axes in redundancy analysis. *Methods in Ecology and Evolution*, *2*(3), pp.269-277.

Li, H. and Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), pp.1754-1760.

Li, W. and Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), pp.1658-1659.

Liaw, A. and Wiener, M., 2002. Classification and regression by randomForest. *R news*, *2*(3), pp.18-22.

Linan, A.G., Lowry, P.P., Miller, A.J., Schatz, G.E., Sevathian, J.C. and Edwards, C.E., 2021. RAD-sequencing reveals patterns of diversification and hybridization, and the accumulation of reproductive isolation in a clade of partially sympatric, tropical island trees. *Molecular Ecology*, *30*(18), pp.4520-4537.

EL Little, J., 1971. Atlas of United States trees, Volume 1, conifers and important hardwoods. *US Department of Agriculture. Miscellaneous Publication*, *1146*.

Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends in ecology & evolution*, *20*(5), pp.229-237.

Mandeville, E.G., Parchman, T.L., Thompson, K.G., Compton, R.I., Gelwicks, K.R., Song, S.J. and Buerkle, C.A., 2017. Inconsistent reproductive isolation revealed by interactions between *Catostomus* fish species. *Evolution letters*, *1*(5), pp.255-268.

Mandeville, E.G., Walters, A.W., Nordberg, B.J., Higgins, K.H., Burckhardt, J.C. and Wagner, C.E., 2019. Variable hybridization outcomes in trout are predicted by historical fish stocking and environmental context. *Molecular ecology*, *28*(16), pp.3738-3755.

Mao, K., Hao, G., Liu, J., Adams, R.P. and Milne, R.I., 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytologist*, *188*(1), pp.254-272.

Markó, G., Novák, I., Bernáth, J. and Altbäcker, V., 2011. Both gas chromatography and an electronic nose reflect chemical polymorphism of juniper shrubs browsed or avoided by sheep. *Journal of chemical ecology*, *37*(7), pp.705-713.

Mason, C.J., Klepzig, K.D., Kopper, B.J., Kersten, P.J., Illman, B.L. and Raffa, K.F., 2015. Contrasting patterns of diterpene acid induction by red pine and white spruce to simulated bark beetle attack, and interspecific differences in sensitivity among fungal associates. *Journal of Chemical Ecology*, *41*(6), pp.524-532.

McFarlane, S.E., Senn, H.V., Smith, S.L. and Pemberton, J.M., 2021. Locus-specific introgression in young hybrid swarms: drift may dominate selection. *Molecular Ecology*, *30*(9), pp.2104-2115.

McKinney, G.J., Waples, R.K., Seeb, L.W. and Seeb, J.E., 2017. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources*, *17*(4), pp.656-669.

Menon, M., Bagley, J.C., Friedline, C.J., Whipple, A.V., Schoettle, A.W., Leal-Sàenz, A., Wehenkel, C., Molina-Freaner, F., Flores-Rentería, L., Gonzalez-Elizondo, M.S. and

Sniezko, R.A., 2018. The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Molecular Ecology*, *27*(5), pp.1245-1260.

Miller, R.F. and Wigand, P.E., 1994. Holocene changes in semiarid pinyon-juniper woodlands: response to climate, fire, and human activities in the US Great Basin. *BioScience*, *44*(7), pp.465-474.

Moore, B.D., Andrew, R.L., Külheim, C. and Foley, W.J., 2014. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *New Phytologist*, *201*(3), pp.733-750.

Neale, D.B., 2007. Genomics to tree breeding and forest health. *Current Opinion in Genetics & Development*, *17*(6), pp.539-544.

Nei, M., 1972. Genetic distance between populations. *The American Naturalist*, *106*(949), pp.283-292.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y. and Wang, J., 2012. SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PloS one*, *7*(7), p.e37558.

Nowak, C.L., Nowak, R.S., Tausch, R.J. and Wigand, P.E., 1994. Tree and shrub dynamics in northwestern Great Basin woodland and shrub steppe during the Late-Pleistocene and Holocene. *American Journal of Botany*, *81*(3), pp.265-277.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P. and Stevens, M.H.H., 2020. vegan: Community Ecology Package. R package version 2.5-6. 2019.

Otto, A. and Wilde, V., 2001. Sesqui-, di-, and triterpenoids as chemosystematic markers in extant conifers—a review. *The Botanical Review*, *67*(2), pp.141-238.

Palma-Otal, M., Moore, W.S., Adams, R.P. and Joswiak, G.R., 1983. Morphological, chemical, and biogeographical analyses of a hybrid zone involving *Juniperus virginiana* and *J. horizontalis* in Wisconsin. *Canadian Journal of Botany*, *61*(10), pp.2733-2746.

Paradis, E. and Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), pp.526-528.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. and Buerkle, C.A., 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular ecology*, *21*(12), pp.2991-3005.

Pardikes, N.A., Forister, M.L. and Dyer, L.A., 2019. Preference and performance of Lepidoptera varies with tree age in juniper woodlands. *Ecological Entomology*, *44*(1), pp.140-150.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one*, *7*(5), p.e37135.

Petit, R.J. and Hampe, A., 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, *37*, pp.187-214.

Pichersky, E. and Raguso, R.A., 2018. Why do plants produce so many terpenoid compounds?. *New Phytologist*, *220*(3), pp.692-702.

Poelman, E.H., Dam, N.M., Loon, J.J.A., Vet, L.E. and Dicke, M., 2009. Chemical diversity in *Brassica oleracea* affects biodiversity of insect herbivores. *Ecology*, *90*(7), pp.1863-1877.

Powell, J.S. and Raffa, K.F., 1999. Effects of selected *Larix laricina* terpenoids on *Lymantria dispar* (Lepidoptera: Lymantriidae) development and behavior. *Environmental Entomology*, *28*(2), pp.148-154.

PRISM CLIMATE GROUP. 2014. PRISM climate data. Oregon State Univ. Available online at prism.oregonstate.edu; last accessed Apr. 11, 2018.

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

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Remington, C.L., 1968. Suture-zones of hybrid interaction between recently joined biotas. In *Evolutionary biology* (pp. 321-428). Springer, Boston, MA.

Rieseberg, L.H. and Soltis, D.E., 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants*, *5*(1), pp.65-84.

Rieseberg, L.H., Archer, M.A. and Wayne, R.K., 1999. Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), pp.363-372.

Rieseberg, L.H., Kim, S.C., Randell, R.A., Whitney, K.D., Gross, B.L., Lexer, C. and Clay, K., 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, *129*(2), pp.149-165.

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804), 20142903.

Schwartz, C.C., Regelin, W.L. and Nagy, J.G., 1980. Deer preference for juniper forage and volatile oil treated foods. *The Journal of Wildlife Management*, pp.114-120.

Schweitzer, J.A., Madritch, M.D., Bailey, J.K., LeRoy, C.J., Fischer, D.G., Rehill, B.J., Lindroth, R.L., Hagerman, A.E., Wooley, S.C., Hart, S.C. and Whitham, T.G., 2008. From genes to ecosystems: the genetic basis of condensed tannins and their role in nutrient regulation in a *Populus* model system. *Ecosystems*, *11*(6), pp.1005-1020.

Seehausen, O., 2004. Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*(4), pp.198-207.

Seybold, S.J., Huber, D.P., Lee, J.C., Graves, A.D. and Bohlmann, J., 2006. Pine monoterpenes and pine bark beetles: a marriage of convenience for defense and chemical communication. *Phytochemistry Reviews*, *5*(1), pp.143-178.

Shastry, V., Adams, P.E., Lindtke, D., Mandeville, E.G., Parchman, T.L., Gompert, Z. and Buerkle, C.A., 2021. Model-based genotype and ancestry estimation for potential hybrids with mixed-ploidy. *Molecular Ecology Resources*, *21*(5), pp.1434-1451.

Soltis, P.S., Marchant, D.B., Van de Peer, Y. and Soltis, D.E., 2015. Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, *35*, pp.119-125.

Stelkens, R. and Seehausen, O., 2009. Genetic distance between species predicts novel trait expression in their hybrids. *Evolution: International Journal of Organic Evolution*, *63*(4), pp.884-897.

Swenson, N.G. and Howard, D.J., 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist*, *166*(5), pp.581-591.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), pp.437-460.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), pp.585-595.

Tank, D.C., Eastman, J.M., Pennell, M.W., Soltis, P.S., Soltis, D.E., Hinchliff, C.E., Brown, J.W., Sessa, E.B. and Harmon, L.J., 2015. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. *New Phytologist*, *207*(2), pp.454-467.

Terry, R.G., 2010. Re-evaluation of morphological and chloroplast DNA variation in *Juniperus osteosperma* Hook and *Juniperus occidentalis* Torr. Little (Cupressaceae) and their putative hybrids. *Biochemical Systematics and Ecology*, *38*(3), pp.349-360.

Terry, R.G., Nowak, R.S. and Tausch, R.J., 2000. Genetic variation in chloroplast and nuclear ribosomal DNA in Utah juniper (*Juniperus osteosperma*, Cupressaceae): evidence for interspecific gene flow. *American Journal of Botany*, *87*(2), pp.250-258.

Thompson, R.S., 1990. Late Quaternary Vegetation and Climate. *Packrat middens: the last 40,000 years of biotic change*, p.200.

Tomlin, E.S., Borden, J.H. and Pierce Jr, H.D., 1996. Relationship between cortical resin acids and resistance of Sitka spruce to the white pine weevil. *Canadian Journal of Botany*, *74*(4), pp.599-606.

Tonkel, K.C., Dimitri, L.A., Longland, W.S., Kirchoff, V.S. and Rector, B.G., 2021. Parallel paths in a miniature world. *Ecology*, *102*(10), p.e03460.

Uckele, K.A., Adams, R.P., Schwarzbach, A.E. and Parchman, T.L., 2021. Genome-wide RAD sequencing resolves the evolutionary history of serrate leaf Juniperus and reveals discordance with chloroplast phylogeny. *Molecular Phylogenetics and Evolution*, *156*, p.107022.

Vasek, F.C., 1966. The distribution and taxonomy of three western junipers. *Brittonia*, *18*(4), pp.350-372.

Vavrek, M.J., 2011. Fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, *14*(1), p.16.

Wang, H., Yin, H., Jiao, C., Fang, X., Wang, G., Li, G., Ni, F., Li, P., Su, P., Ge, W. and Lyu, Z., 2020. Sympatric speciation of wild emmer wheat driven by ecology and chromosomal rearrangements. *Proceedings of the National Academy of Sciences*, *117*(11), pp.5955-5963.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), pp.256-276.

Weisberg, P.J., Lingua, E. and Pillai, R.B., 2007. Spatial patterns of pinyon–juniper woodland expansion in central Nevada. *Rangeland Ecology & Management*, *60*(2), pp.115-124.

Wigand, P.E., Hemphill, M.L. and Sharpe, S.E., 1995. *Great Basin semi-arid woodland dynamics during the late Quaternary* (No. CONF-9409325-). Rust Geotech, Inc., Grand Junction, CO (United States).

Wimp, G.M., Martinsen, G.D., Floate, K.D., Bangert, R.K. and Whitham, T.G., 2005. Plant genetic determinants of arthropod community structure and diversity. *Evolution*, *59*(1), pp.61-69.

Wimp, G.M., Wooley, S., Bangert, R.K., Young, W.P., Martinsen, G.D., Keim, P., Rehill, B., Lindroth, R.L. and Whitham, T.G., 2007. Plant genetics predicts intra-annual variation in phytochemistry and arthropod community structure. *Molecular Ecology*, *16*(23), pp.5057-5069.

Table 1. Variable importance is provided for DAPC and both random forest models (RF_a and RF_b). The five most important variables for each model are bolded. DAPC was conducted on range-wide climatic data extracted from 200 populations from throughout the hybrid zone and parental ranges. In contrast, RDA and random forest were conducted on data from the parental and hybrid localities (Fig. 1 and supplementary figure X). Asterisks in the first column indicate variables that were used as predictors in the RDA model (Fig. 5). RF_a utilized a random forest classification approach to predict ancestry class, whereas RF_b utilized a random forest regression approach to predict hybrid admixture proportions. Where variables were absent in models, they are indicated with NA. Abbreviations are as follows: AET = actual evapotranspiration, PET = potential evapotranspiration, SWB = soil water balance, WS = water supply, CWD = climatic water deficit.

Variable	Definition	DAPC	RF_a	RF_b
cumlPET	Cumulative PET; amount of water that would be evapo-transpired if enough water were available	0.0075	**5.6**	0.55
cumlAET	Cumulative AET; actual amount of water that is evapo-transpired; proxy for productivity	0.0195	2.29	0.14
cumlSWB	Cumulative SWB; quantity of water stored in the soil from one month to the next	0.0043	2.54	0.15
cumlWS*	Cumulative WS	0.0023	3.82	0.01
cumlCWD	Cumulative CWD; proxy for drought stress	0.0003	**4.07**	0.03
SWB:AET	Ratio SWB to AET; values > 1 indicate more stored soil water than used in AET	0.0011	2.29	**9.32**
WS:AET	Ratio WS to AET; values > 1 indicate more water for soil water storage, runoff, or deep percolation than used in AET	NA	4.07	**10.88**
AET:CWD	Ratio of AET to CWD; values > 1 indicate mesic climate, values < 1 indicate xeric climate	0.0042	4.83	0.13
PET:AET*	Ratio of PET to AET; relative drought indicator; values > 1 indicate an unmet demand for water	0.0031	3.56	0.01

pdAET-SWB*	Positive difference between AET and SWB; fraction of AET from month's precipitation, not soil water	**0.1246**	**7.12**	0.09
pdWS-AETorSWB	Positive difference between WS and the greater of AET or SWB; cumulative water available for runoff or deep percolation	0.0134	1.27	**12.54**
WS:AETorSWB	Spring ratio of WS and the greater of AET or SWB; values > 1 indicate spring water available for runoff or deep percolation	0.0031	3.56	0.34
monsoon	Monsoonality; pattern of pronounced precipitation during summer months	**0.2903**	**4.83**	6.48
pcseas	Seasonality of precipitation	**0.1317**	3.05	**12.02**
prcpann	Annual precipitation	0.0009	2.04	0.02
prcpspr	Spring precipitation	0.0029	2.04	1.47
prcpsum	Summer precipitation	NA	**7.38**	2.85
prcpfal	Fall precipitation	0.0053	1.27	0.13
prcpwin	Winter precipitation	NA	2.04	**19.64**
mintemp	Minimum temperature	**0.0635**	6.87	0.04
maxtemp	Maximum temperature	0.0048	3.82	0.15
temprang	Temperature range	0.0539	2.8	1.33
mntmpspr	Minimum spring temperature	0.0185	2.54	0.05
mxtmpspr	Maximum spring temperature	NA	1.53	3.13
mntmpsum	Minimum summer temperature	**0.1283**	1.78	0.03
mxtmpsum	Maximum summer temperature	NA	2.04	0.63
mntmpfal	Minimum fall temperature	0.0355	3.56	0.24
mxtmpfal	Maximum fall temperature	0.0034	2.04	3.14
mntmpwin*	Minimum winter temperature	0.0435	3.31	0.15
mxtmpwin*	Maximum winter temperature	0.0338	2.54	2.89

latitude	NA	NA	3.2
longitude	NA	NA	8.2

Table 2. Results of variance partitioning approach using redundancy analysis (RDA) and partial redundancy analysis (pRDA). For all models, the terpenoid data, which includes variation in concentration and presence/absence across 163 terpenoid compounds, was the response matrix. For the full RDA model, terpenoid chemistry was predicted by the first two components of climate and genetic PCAs. For the pure climate model, the variance explained by the genetic PCs was partialled out, isolating the individual contribution of climate. Alternatively, for the pure genetic model, the variance explained by the climate PCs was partialled out, isolating the effect of genetic variation.

dbRDA models	**Inertia**	**R^2**	**p (>F)**	**Proportion of explainable variance**	**Proportion of total variance**
Full model: *chem* (Euclidean) ~ *clim.* + *genet.* + *geo.*	19.430	0.119	0.001	1	0.119
Pure climate: *chem* (Euclidean) ~ *clim.* | *(genet.* + *geo.*)	2.086	0.013	0.001	0.107	0.013
Pure genetic: *chem* (Euclidean) ~ *genet.* | *(clim.* + *geo.*)	5.350	0.033	0.001	0.275	0.033
Pure geographic: *chem* (Euclidean) ~ *geo.* | (*clim.* + *genet*.)	4.519	0.028	0.001	0.233	0.028
Confounded climate/genetic/geography	7.476			0.385	0.046

Figure legends

Figure 1. Phylogenetic tree estimated with SVDquartets, left, shows evolutionary relationships among the parental taxa. Outgroup taxa are represented by black tips, whereas branches and tips of the parental lineages are colored according to the map of sampling localities to the right. Representative photos of the parental taxa are provided to the right of the phylogeny: A) prostate *Juniperus osteosperma* at Fisher Towers, UT (photo credit: Floris van Bruegel); B) *J. grandis* at Donner Lake, CA; C) *J. occidentalis* woodland near Smith Rock State Park, OR. The geographic locations for each photo are labeled with their associated letters on the map to the right. This map provides a geographic view of the sampling localities, which are overlaid onto polygons representing the geographic species ranges for the parental species: *J. occidentalis* (green), *J. grandis* (gold), and *J. osteosperma* (blue). Hybrid populations are gray, and each locality is labeled with a two-letter population name. Additional locality information can be found in Supplementary Table 1.

Figure 2. Genetic variation is structured strongly at the level of species and hybrids, but weakly across populations. Genotype probabilities estimated with `entropy` were analyzed with PCA for ingroup and outgroup individuals (A) and parental species and hybrids (B), separately, and points were colored by species. PCA was also conducted on the genotype probabilities for each parental taxa separately (C-E), and points were shaded by population (see Figure 1). Finally, relationships between geographic and genetic distance were insignificant for all parental taxa (F-H), though limited sampling of *J. grandis* and *J. occidentalis* likely precluded inference to some degree.

Figure 3. Genome-wide estimates of F_{ST} are visualized as a bubble plot, where the size of the bubble corresponds to the F_{ST} value. All pairwise estimates among parental populations were calculated, and additional information on population names can be found in Figure 1 or Supplementary Table 1. Unfilled bubbles represent interspecific estimates of F_{ST}, whereas filled bubbles indicate intraspecific measures of F_{ST}.

Figure 4. The admixture model of `entropy`, panel A, was parameterized with three sources of ancestry ($K = 3$) and produced parental populations with complete or nearly complete ancestry from one of the three parental lineages, which are denoted by color, and hybrid populations with mixed ancestry from two or three parental lineages. Individual ancestries (q) are provided in the upper barplot, while corresponding population mean ancestries are provided in the lower barplot and as pie charts in the right hand map, panel B. Relationships among populations are provided by a neighbor-joining tree based on Nei's D genetic distances. The ancestry complement model of `entropy`, panel C, was parameterized with two sources of ancestry ($K = 2$) for ease of interpretation (see Methods for details) by representing sister taxa, *J. grandis* and *J. occidentalis*, as a single ancestral source population which is referred to as the western parentals. The ancestry complement model estimates interspecific ancestry, Q, the proportion of an individual's genome in which allele copies were inherited from the same or different source populations. Panel C is a plot of interspecific ancestries (Q) as a function of individual ancestries (q) for each parental and hybrid individual, and shows the distribution of hybrid classes across the hybrid zone. Panel D plots individual ancestries (q) as a function of longitude. A color ramp provides approximate values for the associated interspecific ancestries (Q).

Figure 5. Bioclimatic variation across the hybrid zone is distinct across ancestry classes, and hybrid populations are located in a region which is predominantly intermediate to the parentals. (A) 200 geographic coordinates randomly sampled from throughout the geographic ranges of each parental species and hybrids. All *J. osteosperma* observations west of -118° longitude were treated as hybrid observations, based on results of population genetic analyses (see Figure 4). (B) Bioclimatic data was extracted for each locality in panel A analyzed with discriminant analysis of principal components using the first ten principal components. (C and D) Redundancy analyses (RDA) were used to visualize how bioclimatic variables explain patterns of genetic variation across the hybrid zone. In contrast to the discriminant analysis in panel B, these analyses were based on genotype probabilities and bioclimate data associated with the 25 populations sampled in this study (map insets, lower left). The proximity of points to one another in the RDA biplot indicates similar responses of genetic variation to bioclimatic predictors, and right-angle projections of points onto vectors approximates the variable values for any given point, so that smaller projections indicate larger values for a given variable, and vice versa. RDA axes 1 and 2 are displayed in panel B, and axes 1 and 3 in panel C.

Figure 6. PCA based on variation in both composition and concentration of 163 terpenoid compounds (A). Violin plots of the transformed scores from PC1 and PC2 are shown in panels B and E, respectively. Two examples of compounds which loaded strongly onto PC1 are shown in panels C and D. Both compounds are transgressive in hybrids. Two examples of compounds which loaded strongly onto PC2 are shown in panels F and G. One compound (dehydro-abietal) is intermediate in hybrids, and the other (abieta-dien-one) is dominant towards the *J. osteosperma* phenotype. Abbreviations for groups are as follows: GR (*J. grandis*), HY (hybrids),

OC (*J. occidentalis*), OS (*J. osteosperma*). Asterisk superscripts on chemical names indicate the PC axis for which that compound loaded significantly onto (* = PC 1, ** = PC 2). Terpenoid variation in each parental group was compared to the hybrid group with t-test, and significance for those tests are denoted as follows: ns ($p > 0.05$), * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 0.001$), **** ($p \leq 0.0001$).

Figures

Figure 1

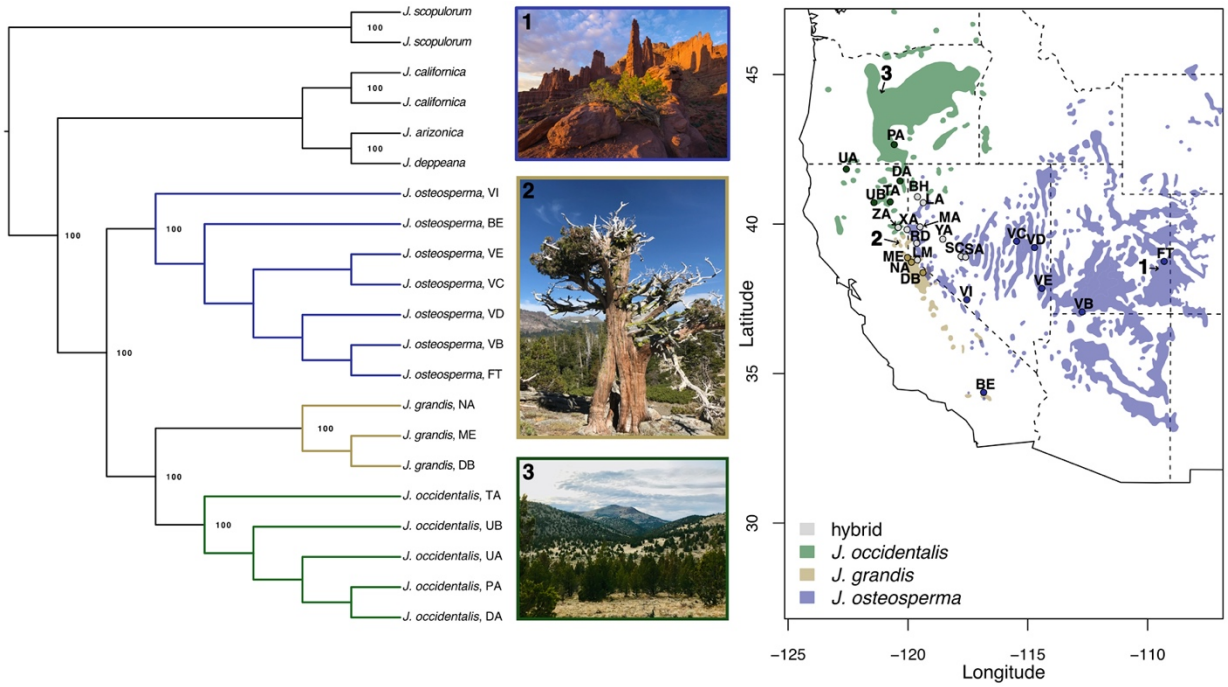


Figure 2

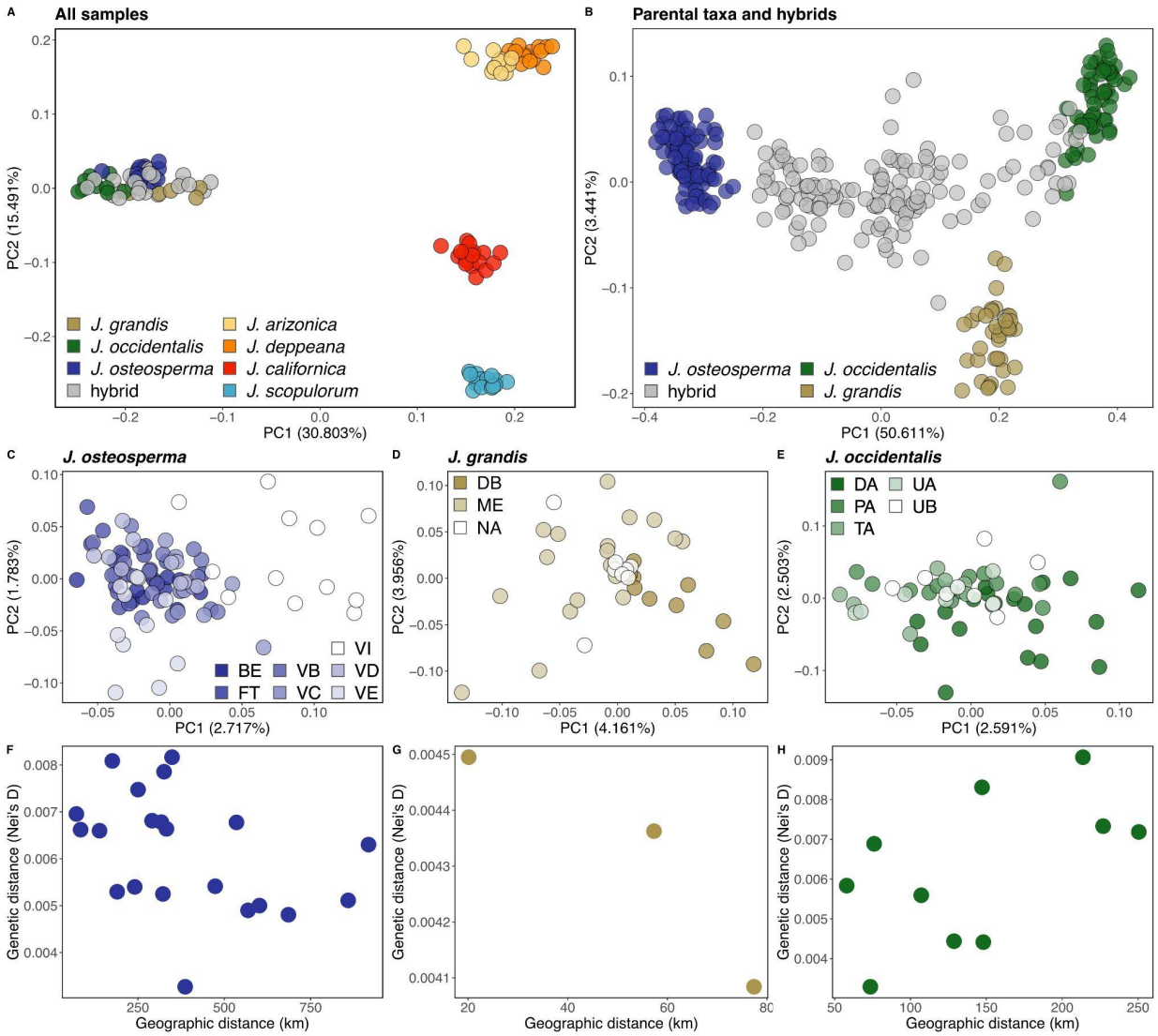

A
All samples
PC2 (15.491%)
PC1 (30.803%)
J. grandis
J. occidentalis
J. osteosperma
hybrid
J. arizonica
J. deppeana
J. californica
J. scopulorum
B
Parental taxa and hybrids
PC2 (3.441%)
PC1 (50.611%)
J. osteosperma
hybrid
J. occidentalis
J. grandis
C
J. osteosperma
PC2 (1.783%)
PC1 (2.717%)
BE
FT
VB
VC
VI
VD
VE
D
J. grandis
PC2 (3.956%)
PC1 (4.161%)
DB
ME
NA
E
J. occidentalis
PC2 (2.503%)
PC1 (2.591%)
DA
PA
TA
UA
UB
F
Genetic distance (Nei's D)
Geographic distance (km)
G
Genetic distance (Nei's D)
Geographic distance (km)
H
Genetic distance (Nei's D)
Geographic distance (km)

Figure 3

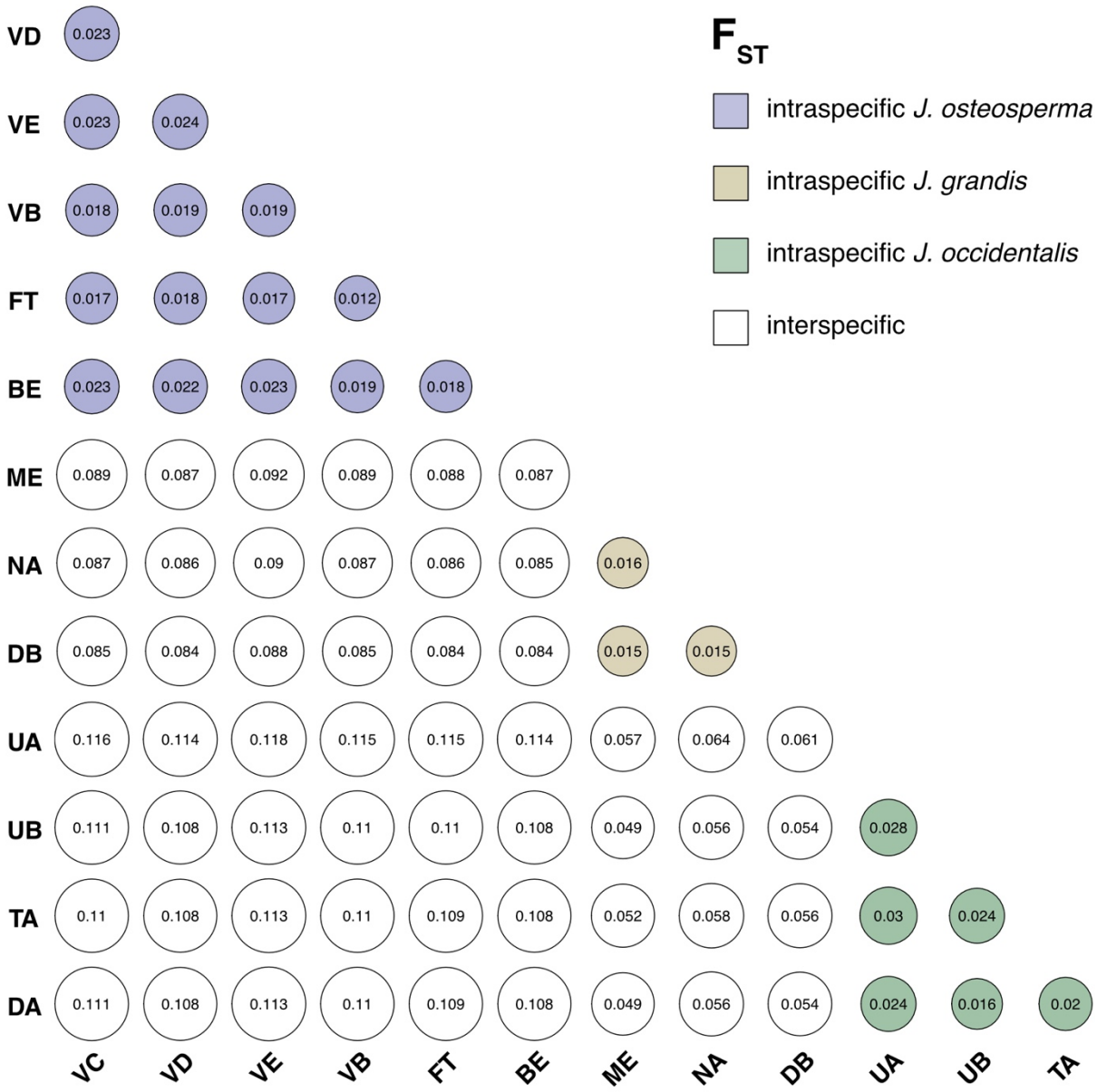

F ST
intraspecific J. osteosperma
intraspecific J. grandis
intraspecific J. occidentalis
interspecific
VD 0.023
VE 0.023 0.024
VB 0.018 0.019 0.019
FT 0.017 0.018 0.017 0.012
BE 0.023 0.022 0.023 0.019 0.018
ME 0.089 0.087 0.092 0.089 0.088 0.087
NA 0.087 0.086 0.09 0.087 0.086 0.085 0.016
DB 0.085 0.084 0.088 0.085 0.084 0.084 0.015 0.015
UA 0.116 0.114 0.118 0.115 0.115 0.114 0.057 0.064 0.061
UB 0.111 0.108 0.113 0.11 0.11 0.108 0.049 0.056 0.054 0.028
TA 0.11 0.108 0.113 0.11 0.109 0.108 0.052 0.058 0.056 0.03 0.024
DA 0.111 0.108 0.113 0.11 0.109 0.108 0.049 0.056 0.054 0.024 0.016 0.02
VC VD VE VB FT BE ME NA DB UA UB TA

Figure 4

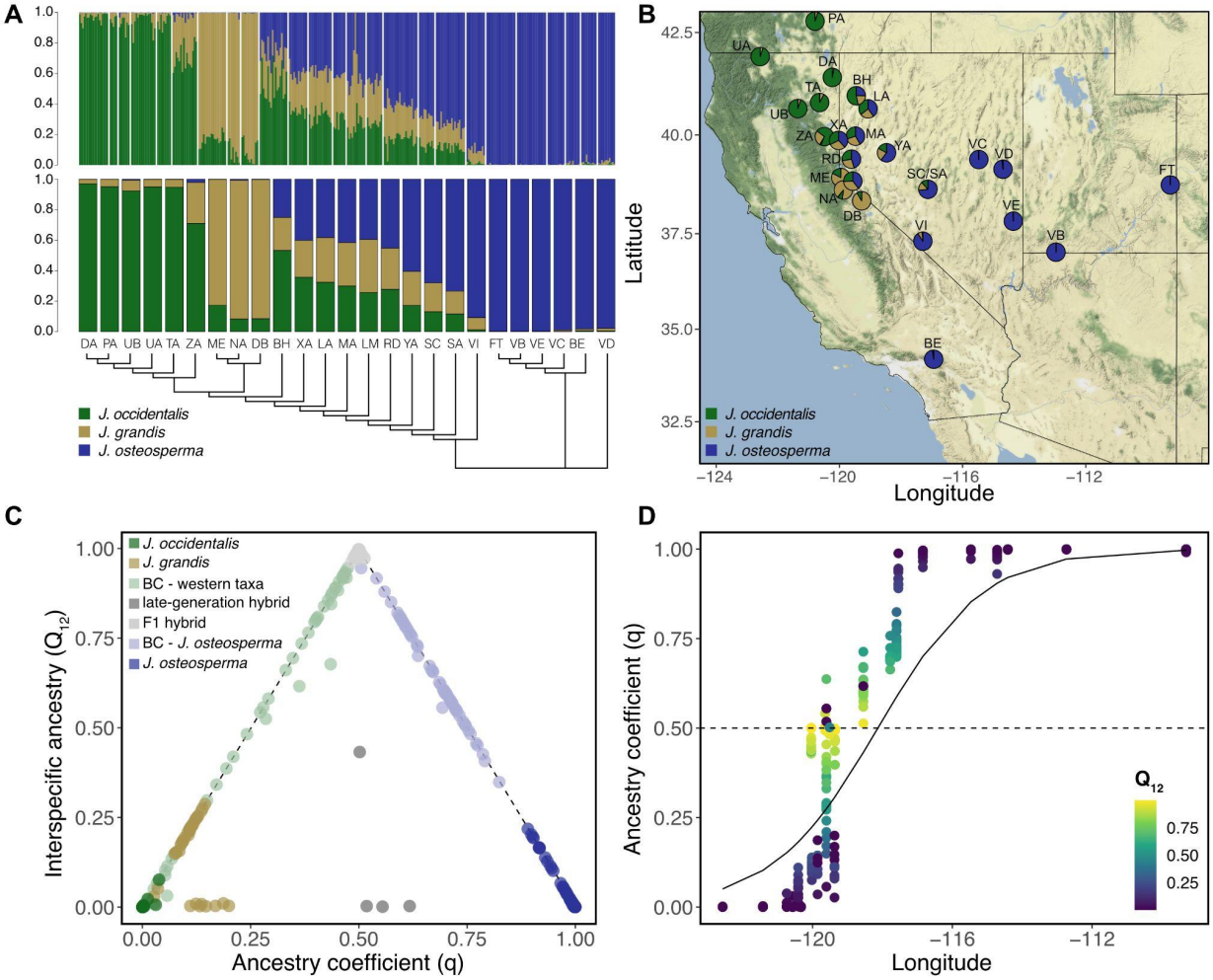

A
1.0
0.8
0.6
0.4
0.2
0.0
DA PA UB UA TA ZA ME NA DB BH XA LA MA LM RD YA SC SA VI FT VB VE VC BE VD
J. occidentalis
J. grandis
J. osteosperma
B
PA
UA
DA
BH
TA
LA
UB
XA
ZA
MA
YA
VC
VD
RD
SC/SA
FT
ME
NA
VE
DB
VI
VB
BE
Latitude
42.5
40.0
37.5
35.0
32.5
−124
−120
−116
−112
Longitude
C
J. occidentalis
J. grandis
BC - western taxa
late-generation hybrid
F1 hybrid
BC - J. osteosperma
J. osteosperma
Interspecific ancestry (Q12)
Ancestry coefficient (q)
0.00
0.25
0.50
0.75
1.00
D
Ancestry coefficient (q)
Q12
0.75
0.50
0.25
Longitude
−120
−116
−112

Figure 5

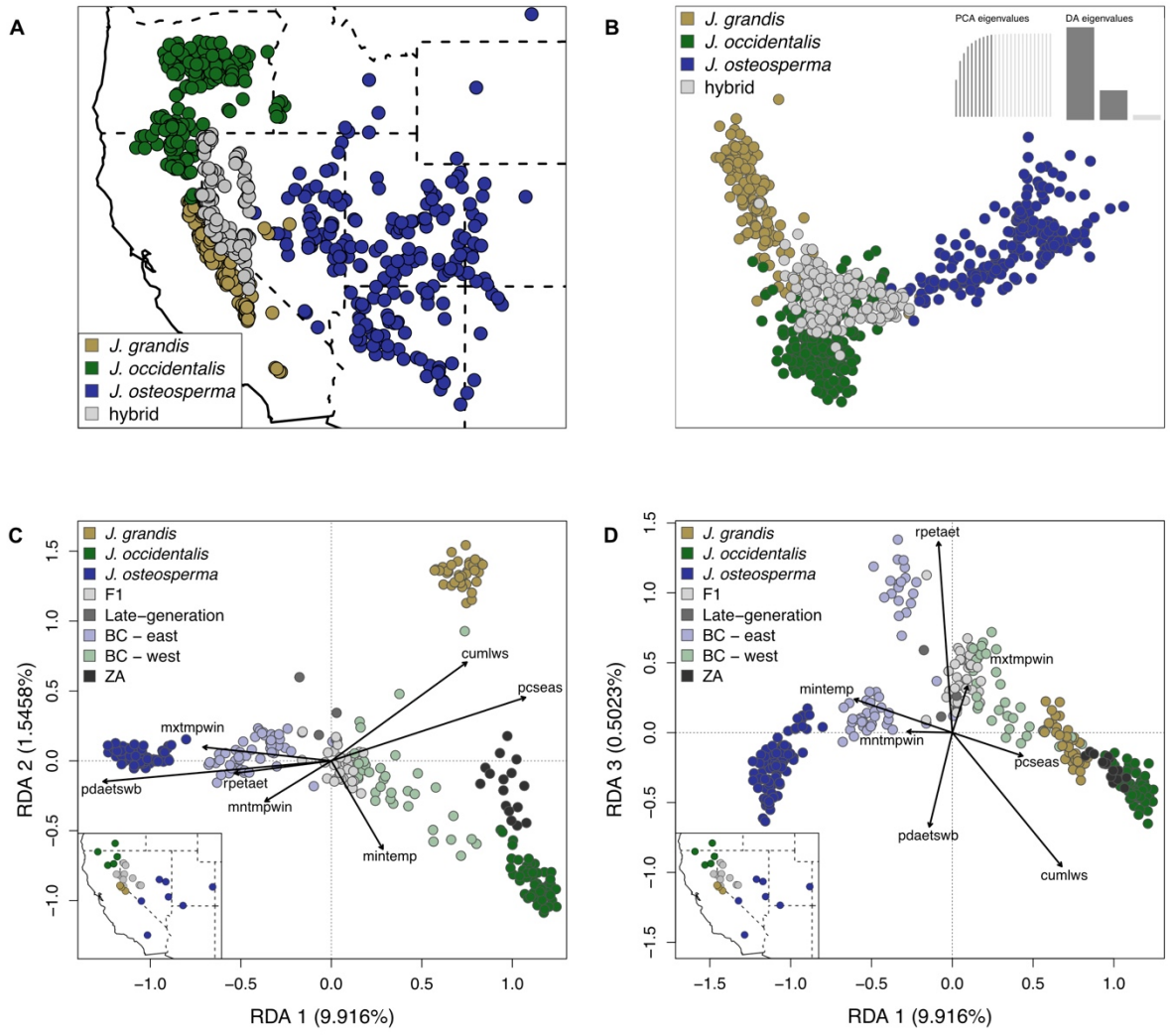
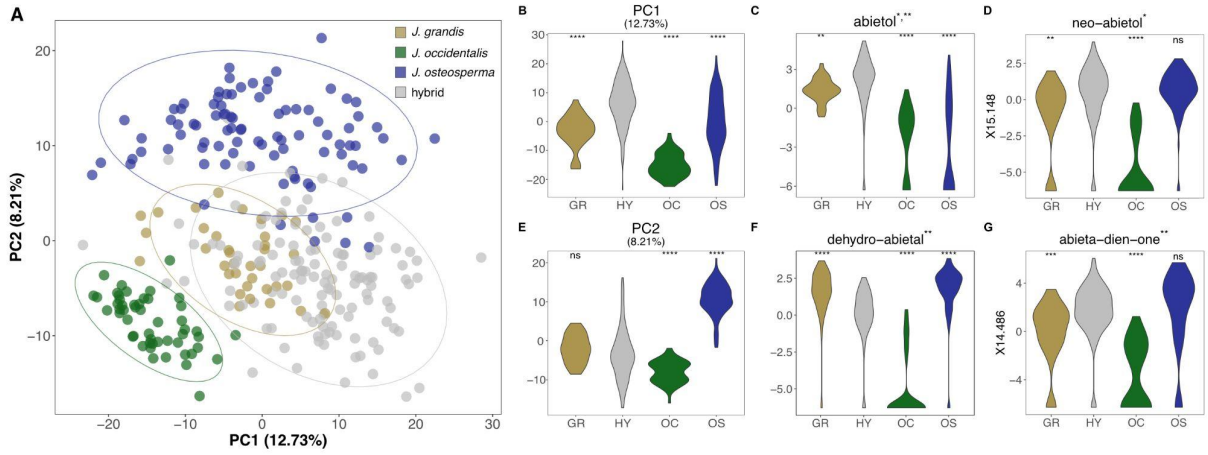

A
J. grandis
J. occidentalis
J. osteosperma
hybrid
B
J. grandis
J. occidentalis
J. osteosperma
hybrid
PCA eigenvalues
DA eigenvalues
C
J. grandis
J. occidentalis
J. osteosperma
F1
Late−generation
BC − east
BC − west
ZA
cumlws
pcseas
mxtmpwin
pdaetswb
rpetaet
mntmpwin
mintemp
RDA 2 (1.5458%)
RDA 1 (9.916%)
D
J. grandis
J. occidentalis
J. osteosperma
F1
Late−generation
BC − east
BC − west
ZA
rpetaet
mxtmpwin
mintemp
mntmpwin
pcseas
pdaetswb
cumlws
RDA 3 (0.5023%)
RDA 1 (9.916%)

Figure 6


A
J. grandis
J. occidentalis
J. osteosperma
hybrid
PC2 (8.21%)
PC1 (12.73%)
B
PC1
(12.73%)
C
abietol
D
neo−abietol
X15.148
E
PC2
(8.21%)
F
dehydro−abietal
G
abieta−dien−one
X14.486
GR
HY
OC
OS
ns

Supplementary Information

History and environment shape population genetic and phytochemical variation across three western *Juniperus* and their hybrids

Kathryn A. Uckele, Casey Philbin, Lora Richards, Lee Dyer, Joshua P. Jahner,

Robert P. Adams, Thomas L. Parchman

Supplementary Methods

Ipyrad parameterization

Briefly, reads were *de novo* assembled within individuals using `vsearch` (v. 2.14.1; Rognes et al., 2016) and aligned with `muscle` (v. 3.8.155; Edgar, 2004) to produce stacks of highly similar reads with over 85% sequence similarity (*clust_threshold*). Consensus sequences with more than 5% ambiguous bases (*max_Ns_consens*) or 5% heterozygous sites (*max_Hs_consens*) likely represent poorly aligned regions, and were discarded. The remaining consensus sequences were then clustered across individuals. Any stacks with more than 8 indels (*max_Indels_locus*), 20% variable sites (*max_SNPs_locus*), or one heterozygous site shared across more than 50% of the samples (*max_shared_Hs_locus*) are indicative of poor alignment or paralogy and were discarded. To avoid overfiltering missing data, we retained all stacks that were present in at least four samples (*min_samples_locus*).

Gas chromatography-mass spectroscopy protocol

Extracts (1 µL) were injected onto an Agilent 7890A gas chromatograph coupled to an Agilent 5975C quadrupole mass spectrometer (GC-MS) equipped with an Agilent HP-5MS, (5%-Phenyl)-methylpolysiloxane, 30 m x 250 μm x 0.25 μm capillary column (Agilent Technologies Inc., Santa Clara, CA, USA). Injections were split 1:10 at an inlet temperature of 250 °C, pressure of 7.1 psi, total He flow of 14 mL/min, septum purge flow of 3 mL/min, and a split flow of 10 mL/min. The 18.25 minute run began with an oven temperature of 40 °C with He carrier gas flow of 1 mL/min at 7.1 psi. This temperature was held for two minutes before elevating temperature at 20 °C/min to 325 °C and holding at 325 °C for an additional two minutes. The electron impact source (70 eV) temperature was 230 °C with the MS Quad temperature set to

150 °C. An *n*-alkane standard mix (Restek 31633, C_{10}-C_{40}) was injected at the start of each day for retention index calibration.

GC-MS data processing

Raw `Chemstation` GC-MS chromatograms were converted so they could be opened in `MassHunter Qualitative Analysis` using `GC/MS Translator` (Agilent Technologies Inc., Santa Clara, CA, USA). In `MassHunter Qualitative Analysis`, total ion chromatograms (TIC) were integrated using the `Agile integrator` in the "find compounds by integration" feature. The resulting retention time (rt) and peak area list was tabulated using the `GCalignR` package (Ottensmann et al., 2018) in R using a maximum retention time (rt) difference of 0.05 min from the mean for group inclusion, and a minimum rt difference of 0.1 min between peak groups. Tabulated peak areas were normalized to n-Eicosane internal standard area and plant mass before statistical analysis. To account for analytical limits of detection, which vary from instrument to instrument, we replaced all zeroes in our data with the minimum peak area detected.

Describing patterns of qualitative and quantitative terpenoid variation with PCoA

To assess whether patterns produced with PCA were dominated by differences in chemical concentration or composition, we visually compared PCoAs conducted with binary and Euclidean chemical distances. A PCoA based on binary chemical distances would accentuate compositional (presence/absence) differences among samples, whereas a PCoA based on Euclidean chemical distances should be similar to PCA and depict variation in both concentrations and composition. Distances for PCoA were calculated using the *vegdist* function (`vegan`; Oksanen et al., 2020) in R. To calculate binary chemical distances among individuals,

we first performed presence/absence standardization and calculated distances using Jaccard's dissimilarity index, whereas non-binary (Euclidean) chemical distances were simply calculated without standardization using the Euclidean dissimilarity index. PCoA was conducted in R using the *pco* function (Goslee and Urban, 2007). Next, we tested whether parental and hybrid groups were qualitatively and quantitatively different by applying a method that is commonly used in beta diversity analyses. First, we tested for homogeneity of variance among groups using the *betadisper* function in R (`vegan`; Oksanen et al., 2020). For pairs of groups with similar group variances, we used the *adonis* function in R (`vegan`; Oksanen et al., 2020) to conduct permutational multivariate analysis of variance (MANOVA) using distance matrices to assess whether groups were significantly chemically different. We conducted tests on binary and Euclidean-based PCoAs and distance matrices to discern whether groups were qualitatively and quantitatively different, respectively.

Supplementary Results

Describing patterns of qualitative and quantitative terpenoid variation with PCoA

As expected, the PCoA based on Euclidean chemical distances among samples was identical to the PCA of the covariance matrix produced from the terpenoid data. Alternatively, the PCoA based on binary chemical distances among samples resembled a 90 degree rotation of the Euclidean PCoA and the PCA, where the first binary PCoA axis was similar to the second Euclidean PCoA and PC axes, and vice versa (Supplementary Figure 1). This suggests that the second Euclidean PCoA and PC axes are driven more by terpenoid composition (compound presences/absences) than the first axes, which are driven more by variation in concentrations. Tests of homogeneity of variance among the groups of samples (*J. grandis*, *J. occidentalis*, *J. osteosperma*, and hybrids) with *betadisper* were significant for Euclidean and binary distances, suggesting that variances among groups were significantly different. For the test based on Euclidean distances, each group exhibited distinct variances, precluding tests of compositional similarity with *adonis*. The hybrid group exhibited the most variance, second was *J. osteosperma*, third was *J. grandis*, and *J. occidentalis* exhibited the least group variance. For the test of homogeneity of variance based on binary distances, the *J. osteosperma* and *J. occidentalis* groups exhibited similar group variances, but tests of compositional similarity with *adonis* indicated that they were compositionally distinct. Similarly, *J. grandis* and the hybrid group exhibited similar group variances, but subsequent tests with *adonis* confirmed that they were also compositionally distinct.

Supplementary Table & Figures

Table 1: Information on collection localities for focal and outgroup *Juniperus* taxa from the western United States. See Figure 1 in the main text for a geographic view of the focal localities.

Id	Taxon label	Locality	N	Longitude (°W)	Latitude (°N)	Elevation (meters)
Focal Taxa						
BE	*J. osteosperma*	San Bernardino, CA	11	34.343487	-116.84196	1465
BH	hybrid	Buffalo Hills, NV	20	40.88507	-119.60353	1548
DA	*J. occidentalis*	Parker Creek, CA	20	41.46667	-120.33333	1690
DB	*J. australis*	False Hot Spring Pass, CA	11	38.34694	-119.37139	2297
FT	*J. osteosperma*	Fisher Towers, UT	20	38.72444	-109.30917	1443
LA	hybrid	Gerlach, NV	12	40.68777	-119.36055	1189
LM	hybrid	Leviathan Mine, NV	18	38.77353	-119.60447	1954
MA	hybrid	south of Pyramid Lake, NV	9	39.88	-119.5131	1171
ME	*J. australis*	Meyers, CA	19	38.85144	-120.02065	1946
NA	*J. australis*	Grover Waterfalls, CA	11	38.6975	-119.8567	1821
PA	*J. occidentalis*	Paisley, OR	20	42.68372	-120.57216	1425
RD	hybrid	Virginia Mountains, NV	14	39.38361	-119.63833	1792
SA	hybrid	Humboldt-Toiyabe National Forest, NV	20	38.86889	-117.59611	2098
SC	hybrid	Humboldt-Toiyabe National Forest, NV	10	38.89222	-117.7775	2054
TA	*J. occidentalis*	north of Eagle Lake, CA	11	40.82416	-120.73944	1693
UA	*J. occidentalis*	Hornbrook, CA	7	41.91611	-122.57277	687
UB	*J. occidentalis*	Subway Cave Lava Tubes, CA	10	40.69222	-121.41972	1329
VB	*J. osteosperma*	Coral Pink Sand Dunes, UT	20	37.035	-112.7342	1795
VC	*J. osteosperma*	Little Antelope Summit, NV	10	39.39667	-115.46861	2260
VD	*J. osteosperma*	Cave Lake State Park, NV	10	39.18722	-114.71778	2186
VE	*J. osteosperma*	Cathedral Gorge, NV	11	37.81917	-114.41222	1446
VI	*J. osteosperma*	Lida, NV	12	37.44139	-117.54389	2134
XA	hybrid	Hallelujah Junction, NV	15	39.78833	-120.03778	1561
YA	hybrid	Stillwater, NV	22	39.52167	-118.54722	1187
ZA	hybrid	north of Sierraville, CA	20	39.64246	-120.36759	1495
Outgroup Taxa						
GB	*J. californica*	Culp Valley, CA	5	33.23	-116.460833	1012
HA	*J. californica*	Culp Valley, CA	10	33.23	-116.460833	1012
WB	*J. scopulorum*	Rocky Mountain National Park, CO	10	40.51389	-105.5889	3184
WC	*J. scopulorum*	Estes Park, CO	5	40.368889	-105.538056	2356
ZB	*J. deppeana*	Paradise Cemetery, AZ	16	31.931820	-109.208412	1695
ZC	*J. deppeana*	Paradise Road, AZ	14	31.931667	-109.204722	1695

Table 2: Genetic diversity (θ_π and θ_w) was calculated for each focal locality listed in Table S1.

Locality	Taxon label	θ_π	θ_w	Tajima's D range (mean)
BE	*J. osteosperma*	0.01147	0.01292	-0.43797 — -0.43151 (-0.43474)
BH	*hybrid*	0.01184	0.01323	-0.37989 — -0.37393 (-0.37691)
DA	*J. occidentalis*	0.01073	0.01099	-0.20873 — -0.20157 (-0.20515)
DB	*J. grandis*	0.01007	0.01042	-0.21415 — -0.20793 (-0.21104)
FT	*J. osteosperma*	0.01132	0.01417	-0.62874 — -0.62218 (-0.62546)
LA	*hybrid*	0.01198	0.01394	-0.49811 — -0.49213 (-0.49512)
LM	*hybrid*	0.01167	0.01284	-0.36621 — -0.36061 (-0.36341)
MA	*hybrid*	0.01185	0.01314	-0.40167 — -0.39627 (-0.39897)
ME	*J. grandis*	0.01083	0.01108	-0.19082 — -0.18404 (-0.18743)
NA	*J. grandis*	0.01015	0.01007	-0.10215 — -0.0966 (-0.09938)
PA	*J. occidentalis*	0.01055	0.01087	-0.22399 — -0.21724 (-0.22062)
RD	*hybrid*	0.01198	0.01346	-0.4246 — -0.4188 (-0.4217)
SA	*hybrid*	0.01190	0.01525	-0.67098 — -0.66486 (-0.66792)
SC	*hybrid*	0.01187	0.01365	-0.48431 — -0.47852 (-0.48142)
TA	*J. occidentalis*	0.01080	0.01067	-0.10837 — -0.10254 (-0.10545)
UA	*J. occidentalis*	0.01082	0.01033	0.01613 — 0.02197 (0.01905)
UB	*J. occidentalis*	0.01067	0.01041	-0.0852 — -0.07911 (-0.08216)
VB	*J. osteosperma*	0.01115	0.01283	-0.45228 — -0.44551 (-0.4489)
VC	*J. osteosperma*	0.01114	0.01273	-0.47911 — -0.4729 (-0.47601)
VD	*J. osteosperma*	0.01141	0.01304	-0.48033 — -0.47407 (-0.4772)
VE	*J. osteosperma*	0.01101	0.01203	-0.36433 — -0.35788 (-0.3611)
VI	*J. osteosperma*	0.01113	0.01177	-0.27079 — -0.26408 (-0.26743)
XA	*hybrid*	0.01104	0.01147	-0.22727 — -0.22084 (-0.22406)
YA	*hybrid*	0.01190	0.01459	-0.56472 — -0.55837 (-0.56155)
ZA	*hybrid*	0.01117	0.01224	-0.35071 — -0.34433 (-0.34752)

Table 3: Genetic diversity (θ_π and θ_w) and Tajima's D was calculated for the parental taxa, *Juniperus osteosperma*, *J. occidentalis*, and *J. grandis*. Distributions for these species are shown in Figure 1.

Taxa	θ_π	θ_w	Tajima's D range (mean)
J. osteosperma	0.01150	0.01744	-0.90137 — -0.89458 (-0.89798)
J. occidentalis	0.01090	0.01221	-0.35958 — -0.35187 (-0.35573)
J. grandis	0.01089	0.01230	-0.37383 — -0.36692 (-0.37037)

Table 4: Annotated compounds from GC-MS analysis of hybrid and parental individuals. Of 163 peak bins, 55 were matched to known compounds and annotated. Retention time (RT), retention index (RI), the difference between the retention index of the matched compound and the observed retention index (Delta RI), and match score (Match) are provided for each annotation. The compounds with the largest loadings on PC1 and PC2 are designated with single or double asterisks, respectively.

RT	RI	Delta RI (DB)	Match	Name	Class	Oxidation
5.6321	913	11	71	alpha-Thujene	Monoterpene	Hydrocarbon
6.1126	970	1	93	Sabinene	Monoterpene	Hydrocarbon
6.5286	1020	19	82	delta-2-Carene	Monoterpene	Hydrocarbon
6.8944	1064	62	85	alpha-phellandrene	Monoterpene	Hydrocarbon
7.0099	1077	36	65	trans-sabinene hydrate	Monoterpene	Monohydric
7.2797	1110	26	64	trans-para-menth2-en-1-ol	Monoterpene	Monohydric
7.6792	1156	4	88	Z-isocitral	Monoterpene	Aldehyde
7.708	1161	20	95	camphor	Monoterpene	Ketone
7.7737	1168	70	62	heptenol acetate	Aliphatic	Ester
7.9012	1184	19	86	Borneol	Monoterpene	Monohydric
7.9522	1189	16	94	terpinen-4-ol	Monoterpene	Monohydric
8.1876	1220	16	92	verbenone	Monoterpene	Ketone
8.7161	1296	13	94	isobornyl acetate	Monoterpene	Monohydric
8.9444	1322	0	92	methyl geranate	Monoterpene	Ester
9.0267	1339	194	66	para-menth-3-en-8-ol	Monoterpene	Monohydric
9.854	1461	7	94	pinchotene acetate	Monoterpene	Aromatic, dihydroxy
10.197	1516	23	87	trans-muurola-4(14),5-diene	Bicyclic sesquiterpene	Hydrocarbon
10.316	**1534**	**21**	**93**	**gamma-cadinene****	**Bicyclic sesquiterpene**	**Hydrocarbon**
10.382	1545	10	95	elemicin	Phenylpropanoid	
10.501	1563	17	92	hedycaryol	Monocyclic sesquiterpene	Monohydric
10.718	1597	24	90	germacrene D-4-ol	Monocyclic sesquiterpene	Monohydric
10.784	1613	63	91	cis-muurol-5-en-4-beta-ol	Bicyclic sesquiterpene	Monohydric
10.884	1626	19	66	beta-oplopenone	Bicyclic sesquiterpene	Ketone
10.979	1642	24	66	epi-cedrol	Tricyclic sesquiterpene	Monohydric
11.023	1655	33	82	5-epi-7-epi-alpha-eudesmol	Bicyclic sesquiterpene	Monohydric
11.087	**1662**	**18**	**89**	**alpha-muurolol****	**Bicyclic sesquiterpene**	**Monohydric**

11.191	1682	76	92	10-epi-gamma-eudesmol	Bicyclic sesquiterpene	Monohydric
11.235	1683	34	79	beta-eudesmol	Bicyclic sesquiterpene	Monohydric
11.353	**1710**	**36**	**65**	**analog: 8-alpha-11-elemodiol***	**Monocyclic sesquiterpene**	**Dihydroxy**
11.629	1757	11	86	8-alpha-11-elemodiol	Monocyclic sesquiterpene	Dihydroxy
11.636	1760	21	93	oplopanone	Bicyclic sesquiterpene	Monohydric ketone
11.813	1792	0	91	8-alpha-acetoxyelemol	Monocyclic sesquiterpene	Dihydroxy ester
12.007	**1827**	**42**	**80**	**Flourensadiol***	**Tricyclic sesquiterpene**	**Dihydroxy**
12.078	1844	52	74	8-alpha-acetoxyelemol	Monocyclic sesquiterpene	Dihydroxy ester
12.148	1854	79	76	2-alpha-hydroxy-amorpha-4,7(11)-diene	Bicyclic sesquiterpene	Monohydric
12.365	1894	7	64	Oplopanoyl acetate	Bicyclic sesquiterpene	Monohydric ketone
13.064	2036	49	95	Manool oxide	Tricyclic diterpene	Pyran
13.32	2087	28	85	13-epi-manool	Bicyclic diterpene	Monohydric
13.506	**2131**	**44**	**88**	**Abietadiene***	**Tricyclic diterpene**	**Hydrocarbon**
13.943	**2227**			**Unknown***	**Tricyclic diterpene**	
14.06	2256	34	86	sclareol	Bicyclic diterpene	Dihydroxy
14.165	2276	36	66	analog: abieta-7,13-dien-3-one	Tricyclic diterpene	Ketone
14.196	2286	27	81	analog: abieta-7,13-dien-3-one	Tricyclic diterpene	Ketone
14.274	**2298**	**32**	**71**	**dehydro-abietal****	**Tricyclic diterpene**	**Aldehyde**
14.295	2305	140	67	7-alpha-hydroxy-trans-totarol	Tricyclic diterpene	Dihydroxy
14.46	2346	48	94	4-epi-abietal	Tricyclic diterpene	Aldehyde
14.486	**2358**	**46**	**99**	**abieta-7,13-dien-3-one***	**Tricyclic diterpene**	**Ketone**
14.687	2395	52	73	analog: abietol	Tricyclic diterpene	Monohydric
14.668	2396	5	71	analog: abietol	Tricyclic diterpene	Monohydric
14.777	**2426**	**113**	**85**	**analog: abieta-7,13-dien-3-one***	**Tricyclic diterpene**	**Ketone**
14.888	**2447**	**46**	**93**	**abietol*,****	**Tricyclic diterpene**	**Monohydric**

14.966	**2467**			**Unknown****	**Tricyclic diterpene**	
15.016	**2486**	**0**	**61**	**3-alpha-14,15-dihydro-manool oxide***	**Tricyclic diterpene**	**Monoxy**
15.057	2487	190	65	analog: abieta-7,13-dien-3-one	Tricyclic diterpene	Ketone
15.148	**2515**	**47**	**77**	**neo-abietol***	**Tricyclic diterpene**	**Monohydric**
15.187	**2523**	**34**	**62**	**Hinokienone****	**Tricyclic diterpene**	**Monohydric ketone**
15.406	**2580**	**38**	**80**	**Totarolone***	**Tricyclic diterpene**	**Monohydric ketone**
15.499	**2599**			**Unknown****	**Tricyclic diterpene**	**Monohydric ketone**
15.738	**2661**			**Unknown****	**Tricyclic diterpene**	**Monohydric ketone**
15.92	**2712**			**Unknown****	**Tricyclic diterpene**	

Figure S1. Principal coordinates analysis of terpenoid Euclidean (A) and binary (B) distances across parental and hybrid individuals.

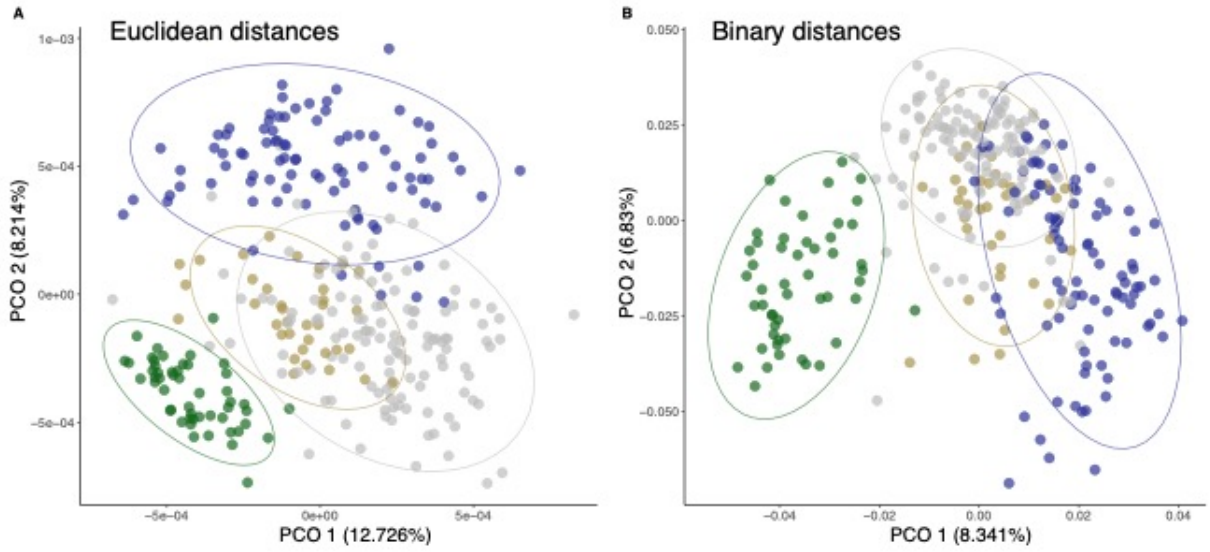


Figure S2. Correlations among moisture-related variables. Descriptions of each variable can be found in the main text (Table 1).

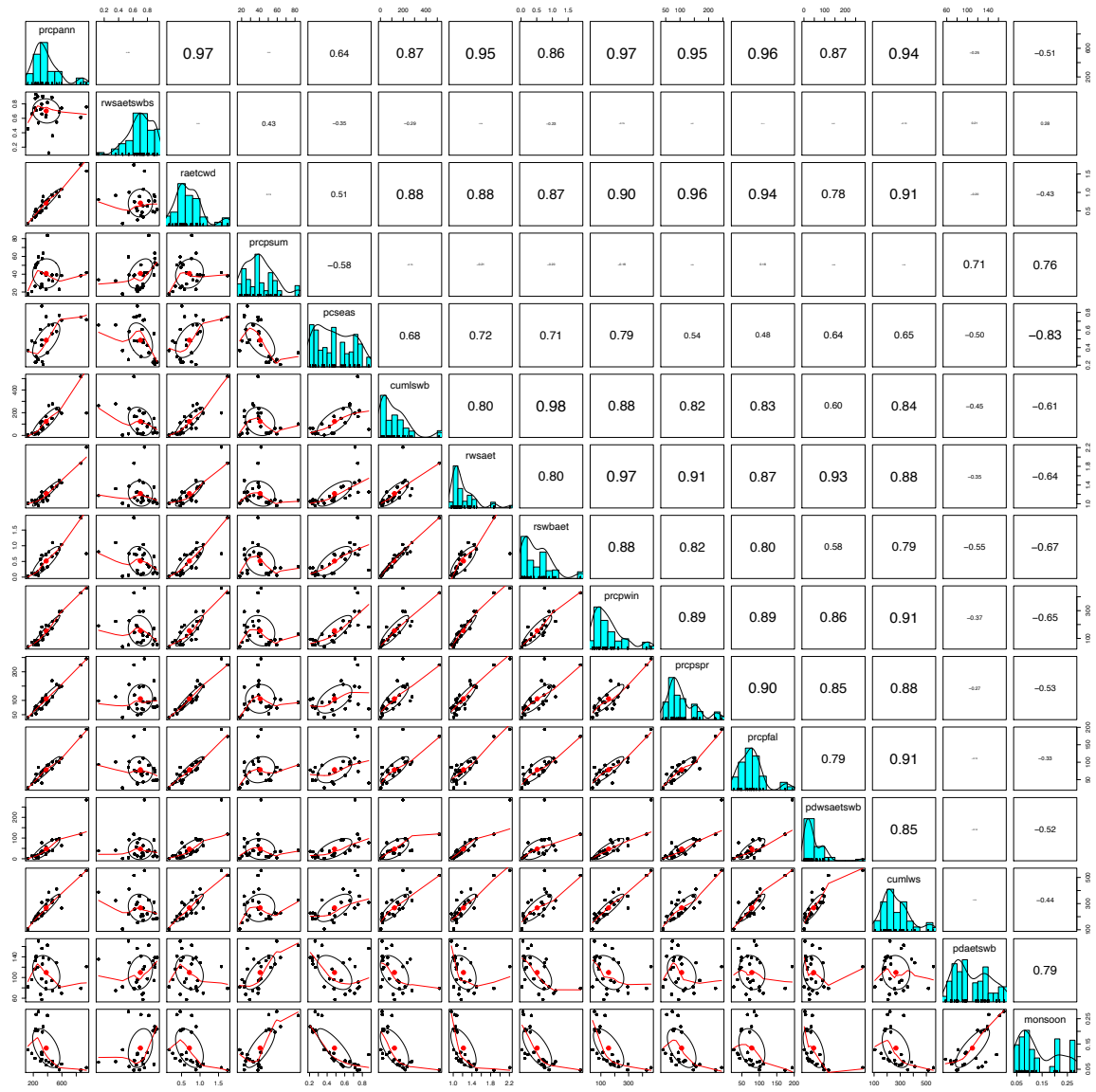


Figure S3. Correlations among temperature-related variables. Descriptions of each variable can be found in the main text (Table 1).

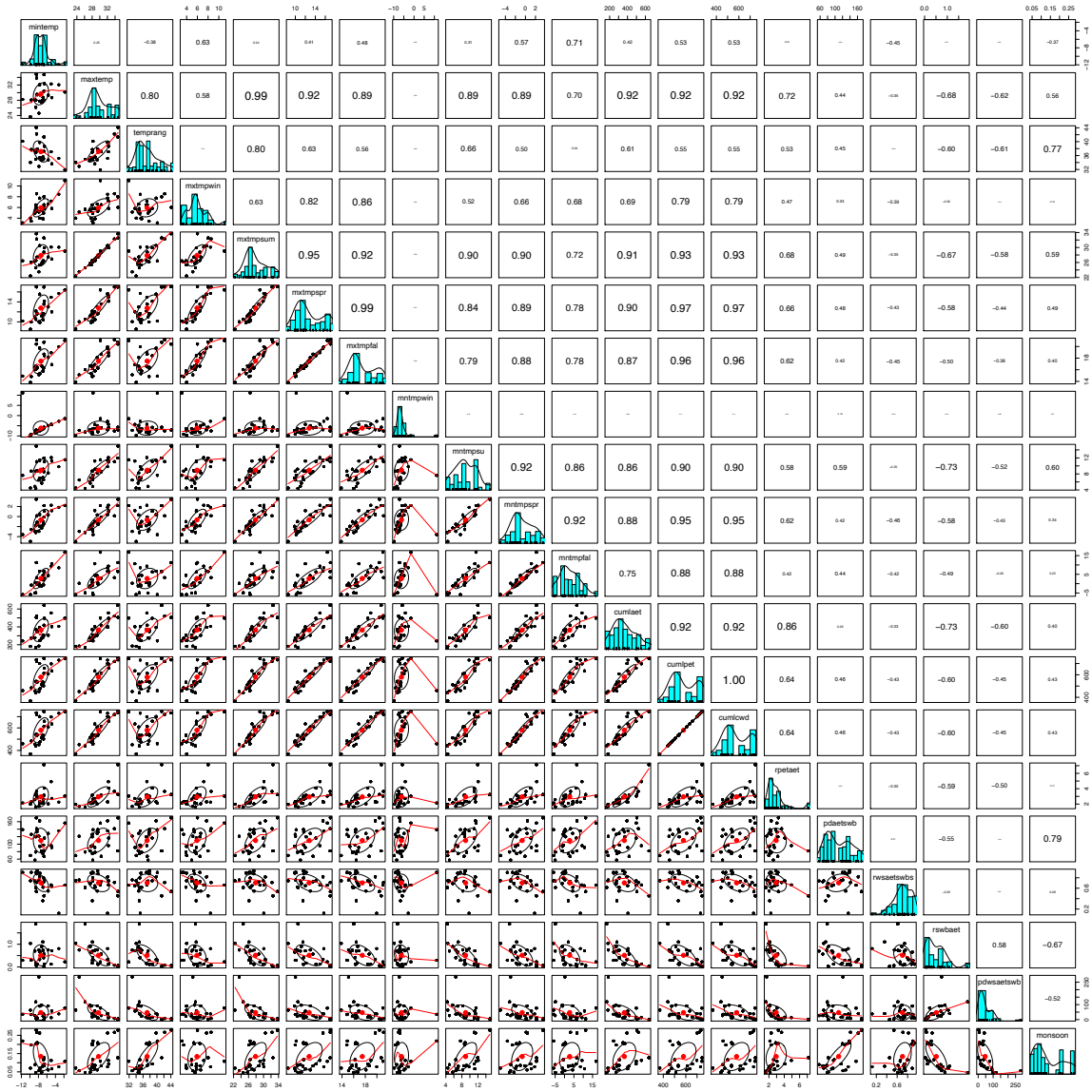
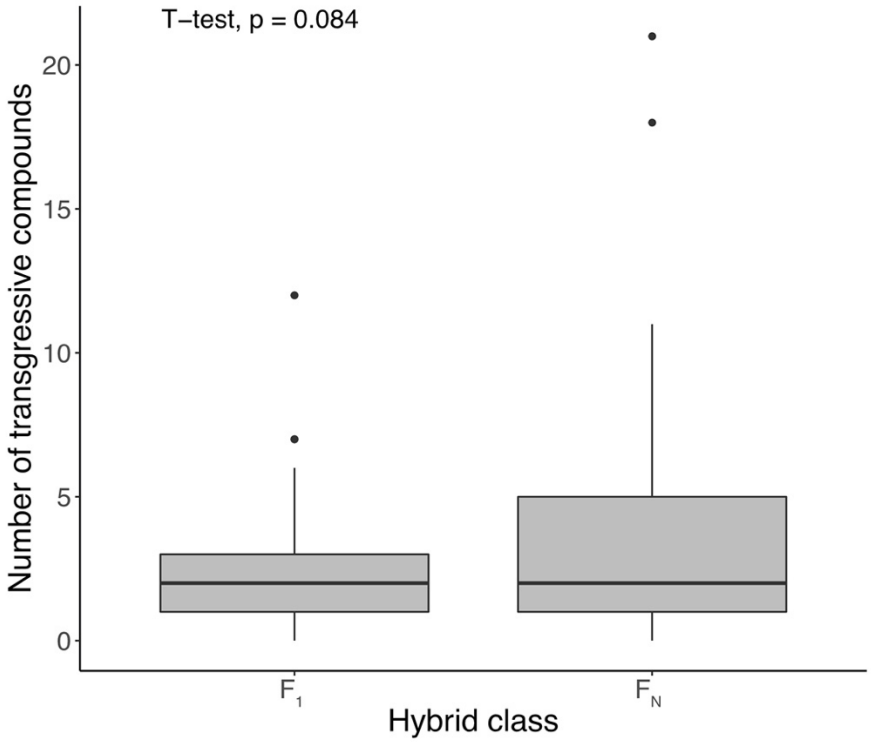


Figure S4. Transgressive compounds are more prevalent in advanced generation hybrids (F_N) than F_1 hybrids.



Chapter 3: Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs

Kathryn A. Uckele,[1,2,3]* Joshua P. Jahner,[1,2]* Eric J. Tepe,[4] Lora A. Richards,[1,2,3] Lee A. Dyer,[1,2,3,5] Kaitlin M. Ochsenrider,[6] Casey S. Philbin,[3] Massuo J. Kato,[7] Lydia F. Yamaguchi,[7] Matthew L. Forister,[1,2,3] Angela M. Smilanich,[1,2] Craig D. Dodson,[6] Christopher S. Jeffrey,[1,3,6] Thomas L. Parchman[1,2]

**These authors contributed equally*

[1]Program in Ecology, Evolution, and Conservation Biology, University of Nevada, Reno, NV 89557, USA

[2]Department of Biology, University of Nevada, Reno, NV 89557, USA

[3]Hitchcock Center for Chemical Ecology, University of Nevada, Reno, NV 89557, USA;

[4]Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221, USA;

[5]Sección Invertebrados, Museo Ecuatoriano de Ciencias Naturales, Quito, Ecuador

[6]Department of Chemistry, University of Nevada, Reno, NV 89557, USA

[7]Department of Fundamental Chemistry, Institute of Chemistry, University of São Paulo, São Paulo, Brazil

Author for correspondence: Kathryn A. Uckele, Email: kathrynuckele@gmail.com

Citation: Uckele, K.A., Jahner, J.P., Tepe, E.J., Richards, L.A., Dyer, L.A., Ochsenrider, K.M., Philbin, C.S., Kato, M.J., Yamaguchi, L.F., Forister, M.L. and Smilanich, A.M., 2021. Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs. *Scientific reports*, *11*(1), pp.1-14.

Abstract

Foundational hypotheses addressing plant-insect co-diversification and plant defense theory typically assume a macroevolutionary pattern whereby closely related plants have similar chemical profiles. However, numerous studies have documented variation in the degree of phytochemical trait lability, thus raising the possibility that phytochemical evolution is more nuanced than initially assumed. We utilize proton nuclear magnetic resonance (^{1}H NMR) data, chemical classification, and double digest restriction site associated DNA sequencing (ddRADseq) to resolve evolutionary relationships and characterize the evolution of secondary chemistry in the Neotropical plant clade Radula (*Piper*; Piperaceae). Sequencing data substantially improved phylogenetic resolution relative to past studies, and spectroscopic characterization revealed the presence of 35 metabolite classes. Metabolite classes displayed phylogenetic signal, whereas the crude ^{1}H NMR spectra featured little evidence of phylogenetic signal in multivariate tests of chemical resonances. Evolutionary correlations were detected in two pairs of compound classes (flavonoids with chalcones; *p*-alkenyl phenols with kavalactones), where the gain or loss of a class was dependent on the other's state. Overall, the evolution of secondary chemistry in Radula is characterized by strong phylogenetic signal of traditional compound classes and weak phylogenetic signal of specialized chemical motifs, consistent with both classic evolutionary hypotheses and recent examinations of phytochemical evolution in young lineages.

Keywords: ddRADseq, nuclear magnetic resonance (^{1}H NMR), phylogenetic comparative analyses, phylogenetic signal, phytochemistry, *Piper*, Radula

Introduction

Plant secondary chemistry affects plant-herbivore interactions at various stages throughout an insect's lifespan: mixtures of compounds can shape adult oviposition preferences[1], specific chemical compounds can stimulate larval feeding[2], specific chemotypes can deter insect herbivores via toxicity or physiological disruptions[3], and sequestered metabolites can alter immune function against natural enemies[4]. Plants capable of developing novel chemical defenses are hypothesized to accrue higher fitness due to enemy release[5], potentially resulting in the diversification of plant lineages with conserved chemical phenotypes (the escape and radiate hypothesis[6]). Coevolutionary hypotheses and plant defense theory have yielded clear predictions that herbivory, additional trophic interactions, and resource availability shape the evolution of plant defenses, including secondary metabolites[7, 8]. However, an evolutionary response to these biotic and abiotic pressures could be complex and highly context-dependent.

Due in part to the enzymatic complexity of metabolic biosynthesis, phylogenetic conservatism is the null hypothesis for the evolution of plant secondary chemistry[9, 10]. Indeed, expectations of phylogenetic conservatism appear to hold at deep evolutionary scales; for example, the family Solanaceae is characterized by the presence of tropane alkaloids[11], though they are consistently present in only 3 of 19 tribes (Datureae, Hyoscyameae, Mandragoreae) and sporadically found elsewhere[12]. Further, recent work suggests that classes of secondary metabolites are more likely to be phylogenetically conserved in large seed plant clades (e.g., eudicots and superasterids) than at lower taxonomic scales (e.g., orders and families)[13]. However, at shallower scales, numerous studies provide evidence for evolutionary lability in chemical traits within genera[7, 14-16], suggesting that surveys of phytochemical variation within young plant lineages might yield variable perspectives on the evolution of secondary chemistry.

Adding further complexity, many studies have found evidence for strong evolutionary associations among chemical classes[16, 17]. For example, Johnson et al.[18] found a strong positive correlation between flavonoids and phenolic diversity and a strong negative correlation between ellagitannins and flavonoids across a phylogeny of 26 evening primroses (*Oenethera*: Onagraceae). Such associations are relevant because they may reflect evolutionary constraints, and their causes may be varied. For example, positive associations may be associated with chemical defense syndromes[9, 19] or synergistic effects of multiple classes on herbivore deterrence[20]. Alternatively, negative associations might be consistent with evolutionary tradeoffs or at least different optima in defense space[18, 19]. By leveraging advances in organic chemistry and genomics, we stand to increase metabolomic and phylogenetic resolution to provide novel insight into the evolution of phytochemistry.

Recent advances in chemical ecology have improved perspectives on phytochemical diversity across a broad range of taxonomic groups and metabolite classes[21, 22]. High throughput processing of plant tissue, rapid advances in spectroscopy, and improved ordination and network analyses have enabled characterization of metabolomic variation across plant communities[10, 15, 22-24] and stand to enhance our understanding of phytochemical evolution across taxonomic scales[21]. Additionally, structural spectroscopic approaches like ^{1}H NMR can provide improved resolution of structural variation across a wide range of metabolite classes. Selection on the plant metabolome is inherently multivariate, arising from diverse herbivore communities and environmental conditions[10, 25], and even relatively small structural changes can impart disproportionate shifts in bioactivity. Thus, approaches that capture a larger proportion of the structural variation underlying phytochemical phenotypes could be well suited to addressing hypotheses concerning evolutionary patterns.

Next-generation sequencing data has reinvigorated phylogenetic analyses of traditionally challenging groups characterized by recent or rapid diversification[26]. Reduced representation DNA sequencing approaches [e.g., ddRADseq; genotyping-by-sequencing (GBS)] have been increasingly utilized in phylogenetic studies due to their ability to effectively sample large numbers of orthologous loci throughout the genomes of non-model organisms without the need for prior genomic resources[27]. Nearly all such studies have reported increased topological accuracy and support compared with past phylogenetic inference based on smaller numbers of Sanger-sequenced loci[28, 29], especially when applied to diverse radiations[30, 31]. While reduced representation approaches have clear phylogenetic utility at relatively shallow time scales, they have also performed well for moderately deep divergence[29, 32].

Piper (Piperaceae) is a highly diverse, pantropical genus of nearly 2,600 accepted species[33], with the highest diversity occurring in the Neotropics[34]. Chemically, *Piper* is impressively diverse[35-37]: chemical profiling in a modest number of taxa has yielded 667 different compounds from 11 distinct structural classes thus far[35, 36, 38, 39]. This phytochemical diversity has likely contributed to the diversification of several herbivorous insect lineages that specialize on *Piper*, including the geometrid moth genus *Eois*[40] (Larentiinae). Furthermore, phytochemical diversity in *Piper* communities has been shown to shape tri-trophic interactions and the structure of tropical communities[36, 39, 41]. As a species-rich genus with abundant and ecologically consequential phytochemical diversity, *Piper* represents a valuable system for understanding how complex diversification histories underlie the evolution of phytochemical diversity.

Piper is an old lineage (~72 Ma), yet most of its diversification occurred in the Neotropics during the last 30-40 My following Andean uplift and the emergence of Central

America[34, 42]. The largest clade of *Piper*, Radula, exemplifies this pattern, as much of its extant diversity (~450 species) arose relatively recently during the Miocene[34]. Such bouts of rapid and recent diversification have limited the efficacy of traditional Sanger sequencing methods to resolve the timing and tempo of diversification in *Piper*[42, 43]. Past phylogenetic analyses utilizing Sanger-sequenced nuclear and chloroplast regions have consistently inferred eleven major clades within *Piper*; however, phylogenetic resolution within these clades has been elusive[42-45]. Phylogenetic inference based on genome-wide data spanning a range of genealogical histories should facilitate an understanding of evolutionary patterns of phytochemical diversity in *Piper* and their consequences for plant-insect codiversification.

We leveraged complementary phylogenomic, metabolite classification, and ^{1}H NMR data sets to generate a *Piper* phylogeny and explore the evolution of secondary chemistry within the largest *Piper* clade (Radula). We used reduced representation sequencing (ddRADseq) to generate genome-wide data for 71 individuals, spanning eight *Piper* clades but focusing on Radula, for phylogenetic analyses. Due to its ability to characterize subtle structural variation across a wide range of compound classes, we used nuclear magnetic resonance (^{1}H NMR) spectroscopy to quantify phytochemical diversity in the same individuals. Our goals were to: 1) resolve the evolutionary relationships within the Radula clade of *Piper* included in this study; 2) characterize metabolomic variation across the genus and within Radula in particular; and 3) quantify the strength of phylogenetic signal and test for evolutionary associations in Radula secondary chemistry. Because secondary chemistry is an emergent composite phenotype of many traits that can evolve semi-independently, we expected to detect mixed strengths of phylogenetic signal and strong associations among a subset of traits over evolutionary time.

Results

Phylogenetic analyses

After contaminant filtering and demultiplexing, we retained ~313 million Illumina reads for phylogenetic analyses. Initial clustering, variant calling, and filtering assembled reads into 362,169 ddRADseq loci. There was a high proportion of missing data, presumably due to allelic dropout increasing with high levels of divergence among *Piper* clades. For Bayesian phylogenetic inference, we mitigated the influence of missing data by removing loci absent in >30% of samples. The final dataset for phylogenetic analysis consisted of 641 ddRADseq loci (~86 bp in length each) that housed 9,113 genetic variants (51% parsimony informative). Aligned loci were concatenated into a nexus alignment with missing data at 18.9% of sites.

Bayesian phylogenetic analysis of ddRADseq data resolved eight major Neotropical *Piper* clades with high posterior support (Fig. 1). While past phylogenetic studies supported the monophyly of seven of these eight clades (Macrostachys, Radula, Peltobryon, Pothomorphe, Hemipodion, Isophyllon, and Schilleria)[34, 43], our analysis resolved an additional clade, Churumayu. Notably, Isophyllon and Churumayu were highly supported, monophyletic clades and not nested within Radula, as was inferred in previous analyses[43]. Contrary to previous phylogenetic hypotheses of *Piper*[34, 43], our analysis might suggest Churumayu is the most basal clade, but we caution that this node had very low posterior support (51%). Intrageneric relationships below the clade level were highly resolved, with nearly all nodes exhibiting greater than 95% posterior support, including within the diverse Radula clade (Fig. 1). Our phylogenetic hypothesis for Radula indicates three species (*P. hispidum*, *P. colonense*, *P. lucigaudens*) may be paraphyletic.

Phytochemical diversity in *Piper*

All but four individuals included in the inferred *Piper* tree were successfully chemically extracted and profiled. Nearly all common compound classes that have been previously reported in *Piper*[46] were observed from our compound characterization analysis (see Table S2). This analysis revealed the presence of broad metabolite classes that are ubiquitous across plant families (e.g., lignans, flavonoids/chalcones, etc.) as well as classes that are specifically common in *Piper* (e.g., amides) (Fig. 2, Table S2). Specific compound characterization revealed genus specific compounds and compound classes (piplartine, cenocladamide, crassinervic acid, kava lactones), as well as metabolites that are more rarely reported in plants (putrescine diamides, nerolidyl catechol, alkenyl phenols, anuramide peptides) (Fig. 2, Table S2). Alternative methods, such as sampling across a species' ontogeny, sampling reproductive parts or roots, and storing freshly collected tissue in methanol rather than air drying would add to a more comprehensive picture of variation in phytochemical diversity across and within species, but our sampling was standardized to allow for initial comparisons across species, some of which were collected in remote regions.

Metabolite phylogenetic signal and evolutionary associations

We recovered 35 metabolite classes, of which only eight were sufficiently present across our taxa to afford tests of phylogenetic signal and correlated evolution. For all eight metabolite classes, estimates of D were low and did not deviate from a null distribution generated under a scenario of Brownian motion (Table 1), consistent with phylogenetic signal. Two of the eight traits, phenolic glycosides and lignans, exhibited strong phylogenetic signal ($D < 0$), while the remaining six traits exhibited weak phylogenetic signal ($0 < D < 1$). Further, all metabolite

classes had observed values of D that differed from a null distribution generated under a phylogenetic randomness scenario (Table 1). The mean of the observed D estimates for the metabolite classes was 0.04, with the largest D statistic observed for the flavonoid class (d_{obs} = 0.49) and the smallest observed for the phenolic glycosides (d_{obs} = -1.18) (Table 1).

Of the 28 pairwise tests of correlated evolution, only two were significant based on a significance level of 0.05. Evidence for correlated evolution was detected in two pairs of metabolite classes: 1) flavonoids and chalcones; and 2) *p*-alkenyl phenols and kavalactones/butenolides. For the first pair of traits, a model of contingency in which changes in chalcones depend on the state of flavonoids provided the best fit to the data (Table 2). In this model, when flavonoids are present, chalcone gains are almost two times more probable than chalcone losses; however, when flavonoids are absent, chalcone losses are much more probable than chalcone gains (Fig. 3). The alternative contingency model for this pair of traits (i.e., changes in flavonoids depend on the state of chalcone) was also a good fit to the data (Table 2). According to this model, when chalcones are present, flavonoid transitions are extremely probable, with flavonoid gains being approximately eight times more probable than flavonoid losses. Alternatively, when chalcones are absent, flavonoid losses are approximately five times more probable than flavonoid gains (Fig. 3). For the second pair of traits, *p*-alkenyl phenols and kavalactones/butenolides, the best fit model was one of interdependent correlated evolution in which changes in *p*-alkenyl phenol depend on the state of kavalactones/butenolides, and vice versa (Table 2). When kavalactones/butenolides are present, *p*-alkenyl phenol transitions are more probable than when they are absent, with the loss of *p*-alkenyl phenols being much more probable than the gain of *p*-alkenyl phenols under both scenarios. Alternatively, when *p*-alkenyl phenols are present, the loss of kavalactones/butenolides is extremely probable relative to the

gain of kavalactones/butenolides, which is rarely observed. When *p*-alkenyl phenols are absent, kavalactones/butenolides are rarely gained or lost (Fig. 3).

Phylogenetic signal in high-dimensional metabolomic data

While the eight metabolite classes uniformly exhibited at least moderate levels of phylogenetic signal, evidence for phylogenetic signal in multivariate analyses of the crude ^{1}H NMR data was largely absent. PCo axes 1 & 2 and 3 & 4 explained 32.8% and 16.0% of variance in the ^{1}H NMR data, respectively, but showed little clustering by clade (Fig. 4a). Permutational multivariate analyses of variance were not significant for combinations of either PCo 1 & 2 ($P = 0.407$) nor 3 & 4 ($P = 0.142$), suggesting that different clades do not form distinct clusters in chemospace based on their ^{1}H NMR spectra.

According to the MRM models, phylogenetic distance significantly predicts phytochemical distance within Radula ($\beta = 4.503$, $P = 0.013$) but not across all clades ($\beta = 1.775$, $P = 0.46$) (Fig 4b). It is important to note that the proportion of variance explained by the significant MRM model is low ($R^2 = 0.039$), suggesting that the majority of variation in NMR data cannot be explained by phylogenetic distance.

Analyses with the generalized *K* statistic (K_{mult}) indicated lower levels of phylogenetic signal in the metabolomic data than expected under a Brownian motion model of evolution for *Piper* generally ($K_{mult} = 0.1606$, $P = 0.001$) and for Radula specifically ($K_{mult} = 0.1803$, $P = 0.001$). Still, the observed K_{mult} was higher than all K_{mult} values obtained with permutations of the ^{1}H NMR dataset (Fig. S1). Additionally, few K_{mult} tests of the permuted data yielded significant *P*-values (4.4% of permutations), indicating that the estimate we observed, though subtle and

lower than Brownian motion expectations, was real and not a statistical artifact of zero-inflation in the data.

Discussion

Piper is a hyper-diverse lineage in which phytochemical diversity has influenced evolutionary and ecological processes and shaped complex tropical communities[15, 39]. However, limitations in both the degree of phylogenetic resolution and the understanding of phytochemical diversity in this group have precluded analyses of phylogenetic signal and correlated evolution of phytochemistry. Phylogenies inferred here with ddRADseq data substantially improved resolution and support compared to past studies of *Piper*, which were limited by interspecific variation in small numbers of Sanger-sequenced loci[34, 42, 43]. Although the data set did not include members from all previously recognized groups, analyses resolved eight monophyletic Neotropical *Piper* clades, six of which have been inferred in previous analyses of the genus based on chloroplast psbJ-petA and ITS[34, 43]. Two of the eight clades, Churumayu and Isophyllon, had been previously nested within Radula[43]; however, our results suggest that they are independent monophyletic lineages (Fig. 1). Despite low support for several deep divergences, the phylogeny inferred here had strong resolution and support for recent relationships, including within Radula (Fig. 1), consistent with other recent reduced representation sequencing studies that have generated high quality phylogenies at shallow time scales[28, 31, 32]. However, a potential limitation of such sequencing designs may include the recovery of fewer loci shared by more distantly related samples due to allelic dropout[47]. It is possible that allelic dropout, potentially exacerbated by strict filtering based on missing data, led to weak support values for deep splits in the phylogeny, many of which occurred early in the

history of the Neotropical *Piper* lineage[34]. Nonetheless, the resulting subset of data (641 loci; 9,113 SNPs) was sufficient for inferring a largely resolved phylogeny, highlighting the potential promise of reduced representation sequencing for resolving evolutionary histories even in groups spanning moderately deep divergence. Although our sampling was limited to 44 of 450 estimated species within Radula, the extent of sampling is a substantial improvement over past phylogenetic analyses for the group[42, 43].

Comparative studies have taken diverse approaches to analyzing metabolomic data, each providing a unique perspective on the evolution of specialized metabolites[10, 24]. Here, we first characterized the presence/absence of 35 metabolite classes commonly used to categorize plant secondary compounds that are hierarchically nested into three levels of structural resolution. Specific categories at the lowest level of the hierarchy, representing specialized structural motifs or specific molecules, were rare across species and precluded tests of phylogenetic signal and correlated evolution at our level of taxonomic sampling (Fig. 2). Despite not being able to test for phylogenetic signal, clustering is evident for more specific categories, such as crassinervic acid and prenylated flavonoids, which are only present in small subclades but include particularly effective defenses[36, 46]. Alternatively, broader metabolite classes at intermediate and high positions in the hierarchy that are directly tied to fundamental secondary metabolite biosynthetic pathways were more abundant across species and exhibited moderately high levels of phylogenetic signal across Radula (Table 1, Fig. 2). This pattern may be expected if initial biosynthetic steps are conserved over longer evolutionary scales, permitting the abundance of broad chemical classes, yet later stage modifications of these core structures are more evolutionarily labile, causing structural similarity to be low even among related species. Flavonoids are a good example of this pattern, with pathways that form the flavonoid scaffold

being very conserved, as they are catalyzed by modified enzymes from ubiquitous metabolic pathways, but then subsequent biosynthetic steps (e.g., those catalyzed by p450 enzymes) modify these scaffolds[48], yielding unique molecules towards the tips of evolutionary trees (Fig. 3E). For example, late-stage modification of common flavonoid scaffolds can result in the production of non-aromatic protoflavonoids. These compounds rarely occur across the plant kingdom and have only recently been found in one species of *Piper*[49], but this type of subtle structural modification that leaves most of the flavonoid scaffold intact dramatically enhances the cytotoxic properties compared to that of the parent flavonoid[50, 51].

One key prediction from the escape and radiate hypothesis is that adaptive defensive traits should be phylogenetically conserved within the lineage they evolved, but this prediction has mostly been evaluated with broad classes of secondary metabolites at high taxonomic scales[6, 13, 48] rather than specific compounds in recent diversifications[7, 10, 16]. A growing number of studies conducted at shallow evolutionary scales suggest low phylogenetic signal in many chemical traits[14, 15, 18]. While evidence for low phylogenetic signal is often attributed to high evolutionary rates (i.e., evolutionary lability), simulations under various evolutionary processes and conditions indicate that the relationship between phylogenetic signal and rate of trait evolution is not necessarily straightforward, and evidence for low phylogenetic signal is not an indication of any single evolutionary process[52]. Nonetheless, understanding how phylogenetic signal responds to variation in phylogenetic scale is informative in a comparative sense, especially among different traits or classes of traits generated with different levels of analytical resolution. Phylogenetic signal is also a useful starting point for developing insights into the drivers of herbivorous insect radiations, as codiversification in many of these lineages is structured in part by chemical defense and biotic interactions[40, 53]. Our results are generally

consistent with the predictions of moderately strong signal for broad classes of compounds, as well as the lack of signal for specific structures captured by ^{1}H NMR data.

The ^{1}H NMR data address a different set of hypotheses than data from categorization of individual molecules – peaks represent resonances associated with particular molecular structures rather than individual compounds, and the chemical shift (frequency), shape, and abundance of these resonances are extremely sensitive to subtle structural changes. ^{1}H NMR spectroscopy easily detects a great range and subtle differences in compositional and structural complexity, including increasing size, asymmetry and oxidation states, that might be predicted to evolve in response to divergent selection across plant populations responding to different suites of enemies[22]. Low levels of phylogenetic signal in the ^{1}H NMR data is also likely due to the fact that many molecular features of small defensive molecules have potentially evolved in a convergent manner across *Piper*, such as the kavalactones, *p*-alkenyl phenols, piplartine, oxidized prenylated benzoic acids, chromenes, anuramide peptides, and phenethyl amides.

There are numerous limitations that could affect estimates of phylogenetic signal in comparative studies[54] that are relevant to the analyses presented here. First, incomplete taxon sampling likely influenced our results to some degree, but sampling was conducted randomly, and the probability that a particular species was sampled was unlikely related to any aspect of its chemical phenotype[55]. Low sampling proportion in clades other than Radula may have reduced our power to detect phylogenetic signal across all our sampled clades[55] (Fig. 4c). However, despite only sampling approximately 10% of the Radula clade of *Piper*, our sample size should provide sufficient power to infer phylogenetic signal in this clade if present[56, 57] (Fig. 4b). Second, while topological errors and small sample size may have reduced our power to detect phylogenetic signal at deeper time scales[58], more comprehensive genomic sampling produced

enhanced phylogenetic resolution of the Radula clade, where we focused the majority of phylogenetic comparative methods. In addition, we were unable to quantify the measurement error associated with the chemical traits within species, which can decrease the statistical power for detecting phylogenetic signal[56, 59, 60]. It is also possible that environmental effects on our chemical traits could bias estimates of phylogenetic signal and correlations[59].

The causes of correlated evolution, including linkage, epistasis, and selection, are difficult to detect without careful approaches in quantitative genetics and population genomics. Nevertheless, one advantage of examining the presence/absence of multiple classes of defensive compounds in a phylogenetic context is that it is possible to test for expected patterns of correlated evolution due to shared metabolic pathways (e.g., flavonoids and cardenolides[7]) or due to adaptive advantages of specific mixtures. Recent studies detecting evolutionary associations among chemical traits[17, 18] have posited that the branching structure of metabolic pathways could potentially drive this pattern. If metabolite classes share a common precursor, one might expect evolutionary tradeoffs and negative covariation. Alternatively, if metabolite classes lie along the same metabolic pathway, an increase in one class may be concomitant with increases in another (or vice versa), causing positive covariation among the classes. There are also numerous empirical examples supporting the hypotheses that positive correlations may be driven by functional redundancy[61] or selection for synergistic effects on herbivores[20] rather than the structural constraints of metabolism. Suites of covarying defensive traits, or defense syndromes, have been detected in several plant genera[9, 53] and plant communities[62], and have been predominantly used to describe covariation among mechanical and chemical defenses. It is interesting to note the correlated evolution of the flavones/chalcones and the *p*-alkenyl phenols/kavalactones could be due to metabolic constraints, as well as possible adaptations via

synergistic (e.g., kavalactones in *P. methysticum*) or other mixture-associated defensive attributes[22]. Flavonoids and chalcones are directly linked biosynthetically, such that the inherent reactivity of the chalcone moiety permits the enzymatic processes that result in cyclization to the flavonoid scaffold (Fig. 3E). This strong biosynthetic tie yields a clear prediction that the presence of one would depend on the other, and indeed our structural analysis found many cases where both metabolite classes co-occurred in the same sample. Revealing the relationship between the kavalactones and *p*-alkenyl phenols is more tenuous because both classes are less prevalent across our samples. Kavalactones and *p*-alkenyl phenols are dramatically different compounds that diverge at a much earlier branch point from a common cinnamic/coumaric acid precursor. Whereas one polyacetate chain extension pathway leads to the long-chain lipophilic substituent, characteristic of the *p*-alkenyl phenols, the other chain extension pathway conserves oxidation states through the chain extension process to produce the lactones (kavalactones or butenolides) through cyclization reactions (Fig. 3E). The overall outcome is different than the chalcone-flavonoid relationship; in this case, two dramatically different compounds are produced by divergence from a common early-stage biosynthetic precursor in contrast to the immediate biosynthetic precursor relationship between chalcones and flavonoids. Broader sampling across *Piper* and Radula will be necessary to confirm this unexpected relationship between kavalactones and *p*-alkenyl phenols.

Conclusion

Here we sought to advance understanding of phylogenetic relationships within *Piper* while simultaneously investigating the mode and manner of phytochemical evolution in this group. In addition to generating a well-resolved phylogeny, our results support theoretical

expectations that broad classes of compounds display higher degrees of phylogenetic signal than molecular features revealed by ^{1}H NMR data. In addition, trait associations observed in Radula can be used to pose functional hypotheses about genetic constraints or biases on phytochemical evolution and how these factors structure plant-animal interactions. Such investigations are one of the emerging frontiers in terrestrial ecology, and we hope that our study provides one example of how collaborative and multi-disciplinary research can progress in this area.

Methods

Study system and sample collection

For phylogenetic and chemical analyses, we collected leaf material from 71 individuals representing 65 Neotropical *Piper* species from the following clades: Churumayu (N = 3), Hemipodium (N = 1), Isophyllon (N = 5), Macrostachys (N = 4), Peltobryon (N = 2), Pothomorphe (N = 1), Radula (N = 44), and Schilleria (N = 5). This study complied with all local and national regulations/guidelines, and vouchers for all collections were deposited in herbaria in the country of origin as stipulated in the permit documents (Table S1). Brazilian collections were made under permit No. 15780-6 from the Sistema de Autorização e Informação em Biodiversidade (SISBIO). Costa Rican collections were made under the permits R-054-2018-OT-CONAGEBIO and R-055-2018-OT-CONAGEBIO from the Ministerio del Ambiente y Energía (MINAE). Collections from Ecuador were conducted under the permit 03-IC-FAU/FLO-DNP/MA granted by the Ministerio del Ambiente. Collections from Panamá were covered by the permit SE/AP-15-13 from the Autoridad Nacional del Ambiente (ANAM). Finally, Peruvian collections were covered by the permit 288-2015-SERFOR-DGGSPFFS granted by the Servicio Nacional Forestal de Fauna Silvestre (SERFOR). All collections were identified by E.J.T. in the

field, and confirmed with vouchers in the herbarium using regional keys, where available, comparison with type specimens, and experience with the genus. For chemical profiling and DNA sequencing, we collected the youngest, fully expanded leaves and dried them immediately with silica gel. While drying on silica gel may not inhibit enzymatic activity and could limit our analyses to relatively stable molecules, this is not an issue for the phylogenetic analyses described below. Collections were only made from mature individuals in the field. Vouchers were pressed, dried, and deposited in one or more herbaria for future reference and species verification (Table S1). To investigate the evolution of phytochemistry at a relatively shallow evolutionary scale, we conducted the majority of our sampling within Radula[34].

Phylogenetic analyses

Genome-wide polymorphism data was generated for 71 individuals for phylogenetic analyses. Either the same accession sampled for chemical analysis, or an individual from the same population as the one sampled, were sequenced with a genotyping-by-sequencing approach[63] that is analogous to ddRADseq[64]. Briefly, genomic DNA was digested with two restriction enzymes, *Eco*RI and *Mse*I. Sample-specific barcoded oligos containing Illumina adaptors were annealed to the *Eco*RI cut sites, and oligos containing the alternative Illumina adaptor were annealed to the *Mse*I cut sites. Fragments were PCR amplified and pooled for sequencing. The library was size-selected for fragments between 350 - 450 base pairs (bp) with the Pippin Prep System (Sage Sciences, Beverly, MA), and sequenced on two lanes of an Illumina HiSeq 2500 at the University of Texas Genome Sequencing and Analysis Facility (Austin, TX). Single-end, 100 bp, raw sequence data were filtered for contaminants (*E. coli*, *Phi*X, Illumina adaptors or primers) and low quality reads using `bowtie2_db`[65] and a pipeline of

bash and perl scripts (https://github.com/ncgr/tapioca). We used custom perl scripts to demultiplex our reads by individual and trim barcodes and restriction site-associated bases.

Assembly and initial filtering was conducted with `ipyRAD` v.0.7.30[66]. `ipyRAD` was specifically designed to assemble ddRADseq data for phylogenetic applications, permits customization of clustering and filtering, and allows for indel variation among samples[66]. Because a suitable *Piper* genome was not available at the time of analysis, we generated a *de novo* consensus reference of sampled genomic regions with `ipyRAD`. Briefly, nucleotide sites with phred quality scores lower than 33 were treated as missing data. Sequences were clustered within individuals according to an 85% similarity threshold with `vsearch`[67] and aligned with `muscle`[68] to produce stacks of highly similar ddRADseq reads (hereafter, ddRADseq loci). The sequencing error rate and heterozygosity were jointly estimated for all ddRADseq loci with a depth >6, and these parameters informed statistical base calls according to a binomial model. Consensus sequences for each individual in the assembly were clustered once more, this time across individuals, and discarded if possessing >8 indels (max_Indels_locus), >50% heterozygous sites (max_shared_Hs_locus), or >20% variable sites (max_SNPs_locus). To reduce the amount of missing data in our alignment matrix, ddRADseq loci were retained if they were present in at least 50 of 71 samples. The nexus file of concatenated consensus sequences for each individual, including invariant sites, were used as input for the Bayesian phylogenetic methods described below. The nexus alignment as well as complete information on additional parameter settings for this analysis are archived at Dryad (https://doi.org/10.5061/dryad.j6q573nc7).

To resolve patterns of diversification and to provide a foundation for investigating variation in patterns of phytochemical evolution, we estimated a rooted, calibrated tree according

to a relaxed clock model in `RevBayes` v.1.0.12[69], which provides the ability to specify custom phylogenetic models for improved flexibility compared with other Bayesian approaches. The prior distribution on node ages was defined by a birth-death process in which the hyper priors on speciation and extinction rates were exponentially distributed with $\lambda = 10$. We relaxed the assumption of a global molecular clock by allowing each branch-rate variable to be drawn from a lognormal distribution. After comparing the relative fits of JC, HKY, GTR, and GTR+Gamma nucleotide substitution models with Bayes factors, we modeled DNA sequence evolution according to the best-fit HKY model. Eight independent MCMC chains were run for 100,000 generations with a burn-in of 1,000 generations and sampled every 10 generations. Chains were visually assessed for convergence with `Tracer` v.1.7.1[70] and numerically assessed with effective sample sizes (ESS), the Gelman−Rubin convergence diagnostic[71], and by comparing the posterior probabilities of clades sampled between MCMC chains. The maximum clade credibility (MCC) tree provided the ultrametric fixed tree topology and relative node ages for phylogenetic comparative methods described below.

Chemical profiling

Crude proton nuclear magnetic resonance (^{1}H NMR) spectroscopy was chosen for chemotype mapping due to its ability to characterize subtle structural variation across a wide range of compound classes in a single, reproducible, non-destructive analysis[39]. Briefly, after leaf samples were ground to fine powder, approximately 100.0-2000.0 mg of leaf material were ground and transferred to a glass screw cap test tube with 10 ml of methanol, sonicated for 10 minutes, and filtered. This step was repeated and both filtrates were combined in a pre-weighed 20 ml scintillation vial. The solvent was removed *in vacuo* and dissolved in 0.6 ml methanol-d_4

for ^{1}H NMR analysis. Crude ^{1}H NMR solutions were standardized to 13.1±3.8 mg/mL when possible and analyzed on a Varian 400 MHz solution state NMR spectrometer with autosampler. Data were processed using MestReNova software (Mestrelab Research, Santiago de Compostela, Spain). Spectra from the crude extracts were aligned with the solvent peak (CD_3, δ = 3.31 ppm), baseline corrected, and phase corrected. Solvent and water peaks were removed and the binned spectra were normalized to a total area of 100. This data set is referred to as "crude ^{1}H NMR". In addition to crude ^{1}H NMR spectral chemotyping, we further annotated samples based upon the presence or absence of compound classes. To further gain structural resolution across the crude extracts that were sampled, aliquots of the ^{1}H NMR extracts were diluted and subjected to GC-MS and LC-MS analysis (see Supplementary Information for additional details). Crude extracts were classified using chemotaxonomic classifications outlined in Parmar's comprehensive review of *Piper* phytochemistry[35], and our rationale for assigning chemical classes is outlined for each species in Table S2. Briefly, phenolic compounds were identified from high-resolution matches to the METLIN mass spectrometry database[72]. Database hits were then confirmed by agreement of crude ^{1}H NMR chemical shifts with literature values for phenolics known to be found in *Piper*, but not always Radula species. Many compounds identified by LC-MS as flavonoids and chalcones had multiple possible METLIN matches, which confounded NMR confirmation. In these cases, we were still able to differentiate flavonoids from chalcones by characteristic UV spectra (l_{max} ~ 350 nm). Phenylpropanoids and *p*-alkenyl phenols were identified based on characteristic GC-MS fragmentation for these compound classes known to be found in *Piper*. *Piper* amides were characterized in a similar fashion, starting from high-resolution mass spectrometric matches and confirming with known ^{1}H NMR data from the literature. In some cases, crude 2D-NMR analysis (COSY, HSQC) was

used to confirm structural classifications. COrrelated SpectroscopY (COSY) was used to identify ^{1}H NMR that were contained within the same molecule, while Heteronuclear Single Quantum Coherence (HSQC) spectroscopy was used to identify the carbon (^{13}C) resonances associated with certain proton (^{1}H) signals to verify the presence of specific functional groups[73]. Only the most abundant and spectroscopically apparent compounds were classified due to the low sensitivity of NMR. 35 total classes were identified at three levels of structural resolution. At the coarsest level of resolution, we identified compounds as phenolics, nitrogen-containing, or sesquiterpenes. Within the phenolics, we identified 9 intermediate and 17 high-resolution subclasses. Within the nitrogen-containing compounds, we identified three intermediate and three high-resolution subclasses. Finer resolution was not characterized for the sesquiterpene class. This hierarchical set of 35 traits is referred to as "metabolite classes" (Fig. 2). Additional details on chemical profiling can be found in the Supplementary Information.

Phylogenetic signal and evolution of metabolite classes

To assess whether metabolite classes were phylogenetically conserved across Radula, we quantified phylogenetic signal in these binary traits using the D statistic[57]. The D statistic calculates the sum of sister-clade differences, Σd_{obs} for an observed tree and binary trait, and scales this value with the distributions of sums expected under two disparate evolutionary models, random and Brownian motion (Σd_r and Σd_b, respectively), using the following equation:

$$D = \frac{[\Sigma d_{obs} - mean(\Sigma d_b)]}{[mean(\Sigma d_r) - mean(\Sigma d_b)]}$$

Thus, D is expected to equal 1 when the observed binary trait is distributed randomly, lacking phylogenetic signal, and is expected to equal 0 when it exhibits phylogenetic signal as expected under Brownian motion. As tests of phylogenetic signal with the D statistic are most accurate when the ratio of presences and absences is closer to 1:1[57], we tested for phylogenetic signal in eight of the 35 metabolite classes (outlined in white in Fig. 2) which were present in a sufficient proportion of taxa. We used the *phylo.d* function in the `caper` package[74] in `R` v.4.0.0[75] to calculate the observed D for a subset of binary traits that were sufficiently present across the phylogeny. This value was compared to a distribution of D values simulated under models of phylogenetic randomness (D = 1) and pure Brownian motion (D = 0) to determine whether the observed D differed from either zero or one.

To detect evolutionary associations among pairs of metabolite classes within Radula, we used Pagel's method[76] that models evolutionary changes in two binary traits, X and Y, as continuous-time Markov processes in which the probabilities of state transition at one trait may depend on the state at the other trait. We tested all pairwise associations among the eight metabolite classes that were represented by a sufficient number of Radula taxa to provide accurate tests of evolutionary associations ($N = 28$). Significant tests of correlated evolution were followed by tests of contingency, in which changes at X depend on the state of Y, or vice versa. Model fits, comparisons, and plots were performed with the *fitPagel* function in the `phytools` package[77] in `R`.

Multivariate analyses of phylogenetic signal with crude ^{1}H NMR spectra

While the analyses above based on broad classifications of structurally determined metabolites provide a coarse view of phytochemical evolution, these classifications are anchored

to the foundations of plant secondary metabolite biosynthesis. Using ^{1}H NMR spectra as a raw chemotype should allow a more detailed multivariate perspective on phytochemical diversity. Studies on other plant taxa have typically detected some signal and evolutionary correlations for broad classes of compounds but not necessarily for specific compounds or biologically active moieties, both of which can be inferred from ^{1}H NMR data. Multivariate approaches to phylogenetic comparative methods have provided insight into covarying suites of related traits, while simultaneously increasing the statistical power to detect phylogenetic signal[78] and differences in trait means among taxa[79]. Indeed, these multivariate approaches might be particularly useful when exploring the evolution of complex phenotypes, like the plant metabolome, which exhibit trait covariances due to metabolomic or functional associations[20]. Here we utilize three multivariate methods to detect patterns of phylogenetic signal for 263 resonances found in the crude ^{1}H NMR data representing all 35 metabolite classes: 1) principal coordinate analyses (PCoA); 2) multiple regression on distance matrices (MRM); and 3) multivariate estimation of phylogenetic signal.

To visualize patterns of chemotypic variation across all sampled species from all clades, we first analyzed the ^{1}H NMR data with PCoA. First we calculated the Manhattan distances between all pairwise species with the *dist* function in `R`, and then conducted PCoA on the distance matrix using the *pcoa* function in `R`. If the major axes of metabolomic variation are phylogenetically conserved, the plotted species scores should be clustered by clade in a rotated principal coordinate (PCo) space. Alternatively, if metabolomic variation is randomly distributed across the phylogeny, there should be little to no clustering by clade[80]. The degree to which plant clade predicted chemical similarity was assessed using permutational multivariate analysis of

variance (permanova)[81] in the `vegan` package[82] in `R` based on Euclidean distances of the first four PCo axes.

Mantel tests have been frequently used to assess the degree of phylogenetic signal in multivariate data[10, 83, 84] by estimating the relationship between phylogenetic and phenotypic distances. Simulations under scenarios of measurement error have found instances where Mantel tests outperform traditional univariate methods in detecting phylogenetic signal, especially as the number of traits increases[60]. Because we were unable to account for measurement error in our study, we utilized Multiple Regression on distance Matrices (MRM)[85] to examine the relationship between metabolomic and phylogenetic distance at two evolutionary scales (within Radula and across all clades). Euclidean distances were calculated from the crude ^{1}H NMR spectra using the *dist* function in `R`, and phylogenetic distances for Radula only and all clades were calculated using the *cophenetic* function in the `ape` package[89] in `R`. MRM analyses were implemented using the *MRM* function with 1000 permutations in the `ecodist` package[90] in `R`.

Since Blomberg's K[56] statistic exhibits higher statistical power to detect phylogenetic signal relative to Mantel tests[91], we quantified phylogenetic signal of the crude ^{1}H NMR at both evolutionary scales using a multivariate generalization of the K statistic (K_{mult})[92] with the *physignal* function in the `geomorph` package[93] in `R`. Similar to the aforementioned D statistic, the K statistic compares the observed variation to that expected under Brownian motion, but the K statistic does not scale this comparison by the variation exhibited under a completely random evolutionary model[56, 92]. Values of K greater than 1 indicate phylogenetic signal greater than expected under Brownian motion, whereas values between 0 and 1 indicate less signal than expected under Brownian motion. Significance for the generalized K statistic was assessed by permuting the ^{1}H NMR peak data among the tips of the phylogeny for 999 iterations. To

determine whether the zero-inflated nature of the ^{1}H NMR data influenced the detection of phylogenetic signal, we permuted our ^{1}H NMR data set over 1000 iterations by randomly indexing our original ^{1}H NMR data matrix. This permutation method preserves the original proportion of zeros in the matrix while obfuscating any observed phylogenetic signal. The generalized *K* statistic test was calculated for each permutation, and our observed generalized *K* statistic was compared to the null distribution of permuted values.

References

1. Thompson, J. N. & Pellmyr, O. Evolution of oviposition behavior and host preference in Lepidoptera. *Annu. Rev. Entomol.* **36**, 65–89 (1991).
2. Bowers, M. D. Iridoid glycosides and host-plant specificity in larvae of the buckeye butterfly, *Junonia coenia* (Nymphalidae). *J. Chem. Ecol.* **10**, 1567–1577 (1984).
3. Zagrobelny, M. *et al.* Cyanogenic glucosides and plant–insect interactions. *Phytochemistry* **65**, 293–306 (2004).
4. Richards, L. A. *et al.* Synergistic effects of iridoid glycosides on the survival, development and immune response of a specialist caterpillar, *Junonia coenia* (Nymphalidae). *J. Chem. Ecol.* **38**, 1276–1284 (2012).
5. Berenbaum, M. Toxicity of a furanocoumarin to armyworms: a case of biosynthetic escape from insect herbivores. *Science* **201**, 532–534 (1978).
6. Ehrlich, P. R. & Raven, P. H. Butterflies and plants: a study in coevolution. *Evolution* **18**, 586–608 (1964).
7. Agrawal, A. A., Salminen, J. P. & Fishbein, M. Phylogenetic trends in phenolic metabolism of milkweeds (*Asclepias*): evidence for escalation. *Evolution* **63**, 663–673 (2009).

8. Maron, J. L., Agrawal, A. A. & Schemske, D. W. Plant-herbivore coevolution and plant speciation. *Ecology* **100**, e02704 (2019).
9. Agrawal, A. A. & Fishbein, M. Plant defense syndromes. *Ecology* **87**, S132–S149 (2006).
10. Salazar, D. *et al.* Origin and maintenance of chemical diversity in a species-rich tropical tree lineage. *Nat. Ecol. Evol.* **2**, 983 (2018).
11. Griffin, W. J & Lin, G. D. Chemotaxonomy and geographical distribution of tropane alkaloids. *Phytochemistry* **53**, 623–637 (2000).
12. Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3–19 (2003).
13. Zhang, Y. *et al.* Phylogenetic patterns suggest frequent multiple origins of secondary metabolites across the seed plant "tree of life". *Natl. Sci. Rev.* **7**, 964–977 (2020).
14. Kursar, T. A. *et al.* The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proc. Natl. Acad. Sci. USA* **106**, 18073–18078 (2009).
15. Salazar, D., Jaramillo, M. A. & Marquis, R. J. Chemical similarity and local community assembly in the species rich tropical genus *Piper*. *Ecology* **97**, 3176–3183 (2016).
16. Allevato, D. M., Groppo, M., Kiyota, E., Mazzafera, P. & Nixon, K. C. Evolution of phytochemical diversity in *Pilocarpus* (Rutaceae). *Phytochemistry* **163**, 132-146 (2019).
17. Boachon, B. *et al*. Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol. Plant* **1**, 1084–1096 (2018).
18. Johnson, M. T., Ives, A. R., Ahern, J. & Salminen, J. P. Macroevolution of plant defenses against herbivores in the evening primroses. *New Phytol.* **203**: 267–279 (2014).

19. Agrawal, A. A. Macroevolution of plant defense strategies. *Trends Ecol. Evol.* **22**, 103–109 (2007).

20. Richards, L. A., Dyer, L. A., Smilanich, A. M. & Dodson, C. D. Synergistic effects of amides from two *Piper* species on generalist and specialist herbivores. *J. Chem. Ecol.* **36**, 1105–1113 (2010).

21. Sedio, B. E. Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytol.* **214**, 952–958 (2017).

22. Dyer, L. A. *et al*. Modern approaches to study plant–insect interactions in chemical ecology. *Nat. Rev. Chem.* **2**, 50-64 (2018).

23. Richards, L. A. *et al.* Phytochemical diversity and synergistic effects on herbivores. *Phytochem. Rev.* **15**, 1153–1166 (2016).

24. Sedio, B. E., Parker, J. D., McMahon, S. M. & Wright, S. J. Comparative foliar metabolomics of a tropical and a temperate forest community. *Ecology* **99**, 2647–2653 (2018).

25. Fine, P. V. A. *et al.* The growth–defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* **87**, S150–S162 (2006).

26. Léveillé-Bourret, É., Chen, B. H., Garon-Labrecque, M. É., Ford, B. A. & Starr, J. R. RAD sequencing resolves the phylogeny, taxonomy and biogeography of Trichophoreae despite a recent rapid radiation (Cyperaceae). *Mol. Phylogenet. Evol.* **145**, 106727 (2020).

27. Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M. & Eckert, A. J. RADseq approaches and applications for forest tree genetics. *Tree Genet. Genomes* **14**, 39 (2018).

28. Massatti, R., Reznicek, A. A. & Knowles, L. L. Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: A case study in *Carex* sect. *Racemosae*. *Am. J. Bot.* **103**, 337–347 (2016).

29. Du, Z. Y., Harris, A. J. & Xiang, Q. Y. J. Phylogenomics, co-evolution of ecological niche and morphology, and historical biogeography of buckeyes, horsechestnuts, and their relatives (Hippocastaneae, Sapindaceae) and the value of RAD-seq for deep evolutionary inferences back to the Late Cretaceous. *Mol. Phylogenet. Evol.* **145**, 106726 (2020).

30. Fernández-Mazuecos, M. *et al.* Resolving recent plant radiations: power and robustness of genotyping-by-sequencing. *Syst. Biol.* **67**, 250-268 (2017).

31. Paetzold, C., Wood, K. R., Eaton, D., Wagner, W. L. & Appelhans, M. S. Phylogeny of Hawaiian *Melicope* (Rutaceae): RAD-Seq resolves species relationships and reveals ancient introgression. *Front. Plant Sci.* **10**, 1074 (2019).

32. Eaton, D. A., Spriggs, E. L, Park, B. & Donoghue, M. J. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Syst. Biol.* **66**, 399–412 (2017).

33. Callejas-Posada, R. Piperaceae. in: *Flora Mesoamericana* vol. 2, pt. 2 (eds. Davidse, G., Ulloa Ulloa, C., Hernández, H. M. & Knapp, S.) 1-618 (Missouri Botanical Garden Press, 2020).

34. Martínez, C., Carvalho, M. R., Madriñán, S. & Jaramillo, C. A. A late Cretaceous *Piper* (Piperaceae) from Colombia and diversification patterns for the genus. *Am. J. Bot.* **102**, 273–289 (2015).

35. Parmar, V. S. *et al.* Phytochemistry of the genus *Piper*. *Phytochemistry* **46**, 597–673 (1997).

36. Dyer, L. A. & Palmer, A. D. N. *Piper: a model genus for studies of phytochemistry, ecology, and evolution*. (Kluwer Academic/Plenum Publishers, 2004).

37. Richards, L. A. *et al.* Phytochemical diversity drives plant–insect community diversity. *Proc. Natl. Acad. Sci. USA* **112**, 10973–10978 (2015).

38. Kato, M. J. & Furlan, M. Chemistry and evolution of the Piperaceae. *Pure Appl. Chem.* **79**, 529–538 (2007).

39. Richards, L. A., Oliveira, C. & Dyer, L. A. Shedding light on chemically mediated tri-trophic interactions: A ^{1}H-NMR network approach to identify compound structural features and associated biological activity. *Front. Plant Sci.* **9**, 1155 (2018).

40. Jahner, J. P. *et al.* Host conservatism, geography, and elevation in the evolution of a Neotropical moth radiation. *Evolution* **71**, 2885–2900 (2017).

41. Glassmire, A. E. *et al.* Intraspecific phytochemical variation shapes community and population structure for specialist caterpillars. *New Phytol.* **212**, 208–219 (2016).

42. Smith, J. F., Stevens, A. C., Tepe, E. J. & Davidson, C. Placing the origin of two species-rich genera in the late cretaceous with later species divergence in the tertiary: a phylogenetic, biogeographic and molecular dating analysis of *Piper* and *Peperomia* (Piperaceae). *Plant Syst. Evol.* **275**, 9 (2008).

43. Jaramillo, M. A. *et al.* A phylogeny of the tropical genus *Piper* using ITS and the chloroplast intron psbJ–petA. *Syst. Bot.* **33**, 647–660 (2008).

44. Molina-Henao, Y. F., Guerrero-Chacón, A. L. & Jaramillo, M. A. Ecological and geographic dimensions of diversification in *Piper* subgenus *Ottonia*: A lineage of Neotropical rainforest shrubs. *Syst. Bot.* **41**, 253–262 (2016).

45. Asmarayani, R. Phylogenetic relationships in Malesian–Pacific *Piper* (Piperaceae) and their implications for systematics. *Taxon* **67**, 693-724 (2018).

46. Salehi, B. *et al. Piper* species: a comprehensive review on their phytochemistry, biological activities and applications. *Molecules* **24**, 1364 (2019).

47. Cariou, M., Duret, L. & Charlat, S. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecol. Evol*. **3**, 846–852 (2013).

48. Yonekura-Sakakibara, K., Higashi, Y. & Nakabayashi, R. The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* **10**, 943 (2019).

49. Freitas, G. C. *et al.* Cytotoxic non-aromatic B-ring flavanones from *Piper carniconnectivum* C. DC. *Phytochemistry* **97**, 81–87 (2014).

50. Hunyadi, A., Martins, A., Danko, B., Chang, F. R. & Wu, Y. C. Protoflavones: A class of unusual flavonoids as promising novel anticancer agents. *Phytochem. Rev.* **13**, 69–77 (2014).

51. Latif, A. D. *et al.* Protoflavone-chalcone hybrids exhibit enhanced antitumor action through modulating redox balance, depolarizing the mitochondrial membrane, and inhibiting ATR-dependent signaling. *Antioxidants* **9**, 1–18 (2020).

52. Revell, L. J., Harmon, L. J. & Collar, D. C. Phylogenetic signal, evolutionary process, and rate. *Syst. Biol.* **57**, 591–601 (2008).

53. Endara, M. J. *et al.* Coevolutionary arms race versus host defense chase in a tropical herbivore-plant system. *Proc. Natl. Acad. Sci. USA* **114**, E7499-E7505 (2017).

54. Kamilar, J. M. & Cooper, N. Phylogenetic signal in primate behaviour, ecology and life history. *Philos. Trans. R. Soc. B* **368**, 20120341 (2013).

55. Garamszegi, L. Z. & Møller, A. P. Nonrandom variation in within-species sample size and missing data in phylogenetic comparative studies. *Syst. Biol.* **60**, 876-880 (2011).

56. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).

57. Fritz, S. A & Purvis, A. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* 24, 1042–1051 (2010).

58. Sakamoto, M. & Venditti, C. Phylogenetic non-independence in rates of trait evolution. *Biology Letters* **14**, 20180502 (2018).

59. Ives, A. R., Midford, P. E. & Garland, T. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* **56**, 252–270 (2007).

60. Hardy, O. J. & Pavoine, S. Assessing phylogenetic signal with measurement error: a comparison of Mantel tests, Blomberg et al.'s *K*, and phylogenetic distograms. *Evolution* **66**, 2614–2621 (2012).

61. Romeo, J. T., Saunders, J. A. & Barbosa, P. *Phytochemical diversity and redundancy in ecological interactions*, vol. 30. (Springer Science & Business Media, 2013).

62. Kursar, T. A. & Coley, P. D. Convergence in defense syndromes of young leaves in tropical rainforests. *Biochem. Syst. Ecol.* **31**, 929–949 (2003).

63. Parchman, T. L. *et al.* Genome-wide association genetics of an adaptive trait in lodgepole pine. *Mol. Ecol.* **21**, 2991–3005 (2012).

64. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135 (2012).

65. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359 (2012).

66. Eaton, D. A. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics* **30**, 1844–1849 (2014).

67. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016).

68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

69. Höhna, S. *et al.* RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).

70. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).

71. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7**, 457-472 (1992).

72. Guijas, C. *et al.* METLIN: a technology platform for identifying knowns and unknowns. *Anal. Chem.* **90**, 3156-3164 (2018).

73. Crews, P., Rodríguez, J., & Jaspars, M. *Organic structure analysis*. (Oxford University Press, 2010)

74. Orme, D. *et al.* caper: comparative analyses of phylogenetics and evolution in R. R package version 1.0.1. https://CRAN.R-project.org/package=caper (2018)

75. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, AT. https://www.R-project.org/ (2020).

76. Pagel, M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. B* **255**, 37–45 (1994).

77. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

78. Zheng, L. *et al.* New multivariate tests for phylogenetic signal and trait correlations applied to ecophysiological phenotypes of nine *Manglietia* species. *Funct. Ecol.* **23**, 1059–1069 (2009).

79. Clavel, J., Escarguel, G. & Merceron, G. mvmorph: an R package for fitting multivariate evolutionary models to morphometric data. *Methods Ecol. Evol.* **6**, 1311–1319 (2015).

80. Klingenberg, C. P. & Gidaszewski, N. A. Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Syst. Biol.* **59**, 245-261 (2010).

81. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32-46 (2001).

82. Oksanen, J. *et al.* vegan: Community Ecology Package, R package version 2.5-7. https://CRAN.R-project.org/package=vegan (2020)

83. Cardini, A. & Elton, S. Does the skull carry a phylogenetic signal? Evolution and modularity in the guenons. *Biol. J. Linn. Soc.* **93**, 813–834 (2008).

84. Easson, C. G. & Thacker, R. W. Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front. Microbiol.* **5**, 532 (2014).

85. Lichstein, J. W. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecology* **188**, 117-131 (2007).

86. Abouheif, E. A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* **1**, 895–909 (1999).

87. Pavoine, S., Ollier, S., Pontier, D. & Chessel, D. Testing for phylogenetic signal in phenotypic traits: new matrices of phylogenetic proximities. *Theor. Popul. Biol.* **73**, 79–91 (2008).

88. Jombart, T. & Dray, S. adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics* 26, 1907–1909 (2010).

89. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

90. Goslee, S. C. & Urban, D. L. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* **22**, 1–19 (2007).

91. Harmon, L. J. & Glor, R. E. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* **64**, 2173–2178 (2010).

92. Adams, D. C. A generalized *K* statistic for estimating phylogenetic signal from shape and other high-dimensional multivariate data. *Syst. Biol.* **63**, 685–697 (2014).

93. Adams, D. C. & Otárola-Castillo, E. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393-399 (2013).

Acknowledgements

This research was funded by the National Science Foundation (DEB-1145609 and DEB-1442103) to C.S.J., L.A.D., L.A.R., M.L.F., T.L.P., and A.M.S., by the National Science Foundation Graduate Research Award (Award No. 1650114) to K.A.U., and by FAPESP (Award No 2014/50316-7) to M.J.K. Fellowship support for K.A.U., K.M.O., and C.S.P. and funding for chemical instrumentation and analysis was provided by the Hitchcock Center for Chemical Ecology at the University of Nevada, Reno. We thank Jennifer L. McCracken for her assistance

with the collection of GC-MS data for the categorical chemical characterization, and we thank Chris Feldman, Beth Leger, and Steve Vander Wall for their guidance during the earliest stages of this project.

Author contributions

M.L.F., L.A.D., A.M.S., C.S.J., L.A.R., and T.L.P. developed the original idea for the research and secured funding. E.J.T., M.J.K., and L.F.Y. collected specimens. E.J.T. extracted DNA from plant specimens. K.A.U. and T.L.P. generated ddRADseq libraries. K.A.U. and J.P.J. analyzed the genetic data. K.M.O. and L.A.R. performed chemical extractions and analyses. C.S.J., C.S.P., and C.D.D. executed chemical annotation and structure determination. K.A.U. and J.P.J. wrote the first draft of the manuscript, and all authors contributed to subsequent revisions.

Competing interests

The authors declare no competing interests.

Table 1. Estimates of phylogenetic signal (D)[57] for a subset of metabolite classes (see Methods for explanation of subset). To ask whether traits evolved under scenarios of Brownian motion (D = 0) or phylogenetic randomness (D = 1), observed values of D were compared to null distributions of D modeled under each scenario.

			Randomness (H_0: D=1)		Brownian (H_0: D=0)	
Metabolite class	Observed D	Σd_{obs}	$mean(\Sigma d_r)$	*P*	$mean(\Sigma d_b)$	*P*
Flavonoids	0.49	14.18	17.56	**0.012**	11.01	0.093
Chalcones	0.39	9.77	12.18	**0.019**	8.24	0.235
Phenolic glycosides	-1.18	3.11	7.01	**0.000**	5.19	0.95
Lignans	-0.02	4.16	5.47	**0.036**	4.19	0.564
PBA	0.22	12.40	17.51	**0.001**	10.96	0.293
p-alkenyl phenols	0.33	9.47	12.30	**0.010**	8.19	0.265
Kavalactones/butenolides	0.02	5.17	6.99	**0.027**	5.18	0.504
Piper amides	0.1	5.37	7.00	**0.033**	5.18	0.482

Table 2. Correlated evolution was detected in two pairs of metabolite classes with Pagel's method[76]: 1) chalcones and flavonoids; and 2) kavalactones/butenolides and *p*-alkenyl phenols. A model comparison framework was employed to evaluate four potential models of trait evolution using AIC: correlated evolution (transition rate in one trait depends on state at another, and vice versa); contingent change (transition rate in one trait depends on state at another, but not the converse); and independent evolution.

Comparison	Model	AIC	Δ AIC	AIC weight
Chalcones, flavonoids	Chalcones contingent on flavonoids	87.40	0	0.55
	Flavonoids contingent on chalcones	88.41	1.01	0.33
	Correlated evolution	90.54	3.14	0.11
	Independent evolution	95.32	7.92	0.01
kavalactones/butenolides, *p*-alkenyl phenols	Correlated evolution	62.35	0	0.95
	p-alkenyl phenols contingent on kavalactones/butenolides	69.65	7.29	0.03
	Kavalactones/butenolides contingent on *p*-alkenyl phenols	70.61	8.26	0.02
	Independent evolution	71.57	9.22	0.01

Figure legends

Figure 1. Maximum clade credibility tree of 48 samples from the Radula clade of *Piper* and 23 outgroup species inferred with a Bayesian analysis of 641 concatenated ddRADseq loci (55,298 base pairs) comprising 9,113 genetic variants (of which 4,674 are parsimony informative). The outgroup taxa were sampled across multiple *Piper* clades: Churumayu, Isophyllon, Hemipodium, Macrostachys, Peltobryon, Pothomorphe, and Schilleria. All nodes are supported by at least 95% posterior support except where noted with circles or labels. Blue circles indicate support values between 85-95%. Red circles indicate support values between 75-85%. Three nodes with less than 75% posterior support were given numerical support values. Blue bars at each node denote the 95% highest posterior density interval on relative node ages. The photos to the right of the tree showcase a sample of *Piper* diversity, including a few of the species which were included in this study: **A.** *Piper hillianum* (Macrostachys), **B.** *P. acutifolium* (Peltobryon), **C.** *P. umbellatum* (Pothomorphe), **D.** *P. pseudofuligineum* (Radula), **E.** *P. concepcionis* (Radula), **F.** *P. disparipes* (Radula), **G.** *P. friedrichsthalii* (Radula), **H.** *P. dilatatum* (Radula), **I.** *P. bredemeyeri* (Radula), **J.** *P. immutatum* (Radula), **K.** *P. erubescentispicum* (Radula), and **L.** the widespread and often weedy *P. aduncum* (Radula). (Photo credits: E. J. Tepe)

Figure 2. Patterns of chemical variation are displayed for individuals in this study. Taxa comprise the columns of the matrix and are ordered according to their inferred phylogenetic relationships. Groups of columns are colored according to their designated *Piper* clade. Black circles within the phylogenetic tree designate nodes with posterior support values greater than 85%. Each row of the matrix represents a metabolite class that was detected from ^{1}H NMR and MS-based methods, and dark grey cells indicating the presence of that class in that taxa. Classes

are hierarchically nested; capitalized font signifies the three classes at the highest level (and coarsest resolution), italicized font signifies the intermediate level, and black font signifies the lowest level (and highest resolution). Rows outlined in white indicate traits that were analyzed for phylogenetic signal in Radula. To the left of the matrix are representative compounds for a subset of metabolite classes that were detected in our samples.

Figure 3. Evolutionary associations were detected in two pairs of traits according to Pagel's test[76] of correlated evolution: 1) flavonoids and chalcones and 2) *p*-alkenyl phenols and kavalactones/butenolides. Filled shapes indicate presences and unfilled shapes indicate absences of flavonoids (circles), chalcones (squares), *p*-alkenyl phenols (diamonds), and kavalactones/butenolides (triangles), respectively. The shapes used in the phylogenetic plots (**A** and **C**) are repeated below (**B** and **D**) to depict four states comprising all combinations of presences and absences in the pair of traits. Arrows represent transition rates between states. **B.** As both models of contingent change provided good fits to the flavonoid and chalcone data, both sets of transition rates are displayed, with the first set of values (bolded) corresponding to the best supported model (chalcone evolution contingent on flavonoid state) and the second set of values corresponding to the alternative contingency model (flavonoid evolution contingent on chalcone state). **D.** The best fit model to the *p*-alkenyl phenol and kavalactone/butenolide data was one of dependent evolution, where *p*-alkenyl phenol evolution is dependent on the state at the kavalactone/butenolide trait, and vice versa. Panel **E** illustrates the enzymatic processes and branch points along biosynthetic pathways that give rise to the four classes of metabolites. Chalcones are immediate biosynthetic precursors of flavonoids, where the inherent reactivity of the chalcone moiety permits cyclization to the flavonoid scaffold. Subtle structural changes to

the flavonoid scaffold caused by late-stage oxidation can produce protoflavonoids, a rare class of metabolite with potent cytotoxic activity. In contrast, the pathways of *p*-alkenyl phenols and kavalactones diverge much earlier and embark on distinct chain elongation pathways that lead to long-chain lipophilic substituent characteristic of the *p*-alkenyl phenols in one case, and lactones (kavalactones and butenolides) in the other case.

Figure 4. **A.** Chemospace of all 67 *Piper* samples constructed with the crude ^{1}H NMR data across 277 peaks. Point colors were chosen according to clade designation as portrayed in the phylogenetic tree in Figure 1. A legend is provided in the rightmost PCoA plot. Error bars depict estimates of ordination mean +/- 3 standard errors for each clade. **B.** MRM analyses recovered a significant positive relationship between phylogenetic and chemical distances calculated among samples from the Radula clade (left), but did not recover a significant relationship when calculated among all samples from all clades (right).

Figure 1

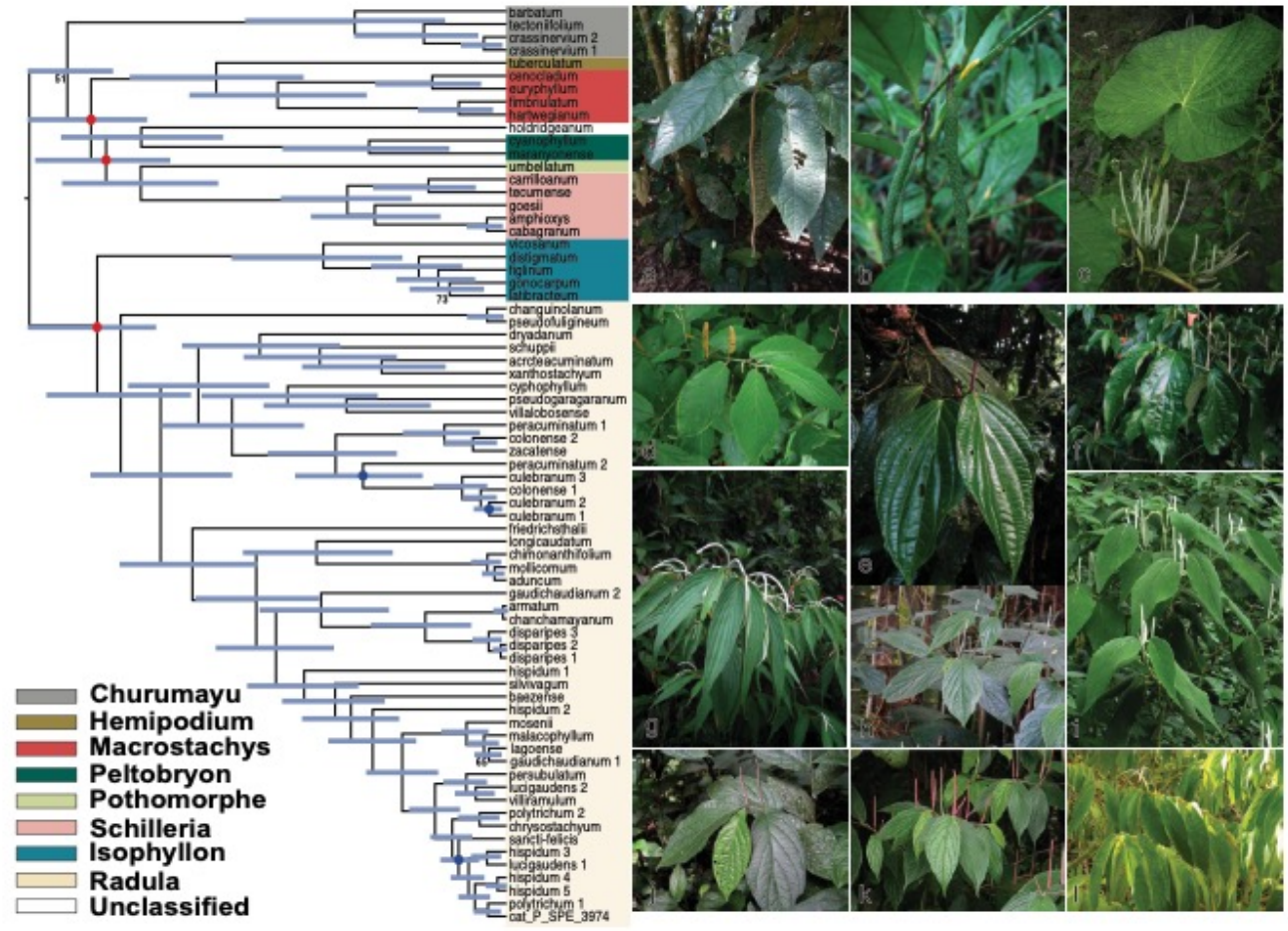

barbatum
tectonifolium
crassinervium 2
crassinervium 1
tuberculatum
cenocladum
euryphyllum
fimbriulatum
hartwegianum
holdridgeanum
cyanophyllum
steyermarkense
umbellatum
carrilloanum
tecumense
goesii
amphioxys
cabagranum
vicosanum
distigmatum
riglinum
gonocarpum
latibracteum
changuinolanum
pseudofuligineum
dryadanum
schuppii
acrocteacuminatum
xanthostachyum
cyphophyllum
pseudogaragaranum
villalobosense
peracuminatum 1
colonense 2
zacatense
peracuminatum 2
culebranum 3
colonense 1
culebranum 2
culebranum 1
friedrichsthalii
longicaudatum
chimonanthifolium
mollicomum
aduncum
gaudichaudianum 2
armatum
chanchamayanum
disparipes 3
disparipes 2
disparipes 1
hispidum 1
silvivagum
baezense
hispidum 2
mosenii
malacophyllum
lagoense
gaudichaudianum 1
persubulatum
lucigaudens 2
villamulum
polytrichum 2
chryscstachyum
sancti-felicis
hispidum 3
lucigaudens 1
hispidum 4
hispidum 5
polytrichum 1
cat_P_SPE_3974
Churumayu
Hemipodium
Macrostachys
Peltobryon
Pothomorphe
Schilleria
Isophyllon
Radula
Unclassified

Figure 2

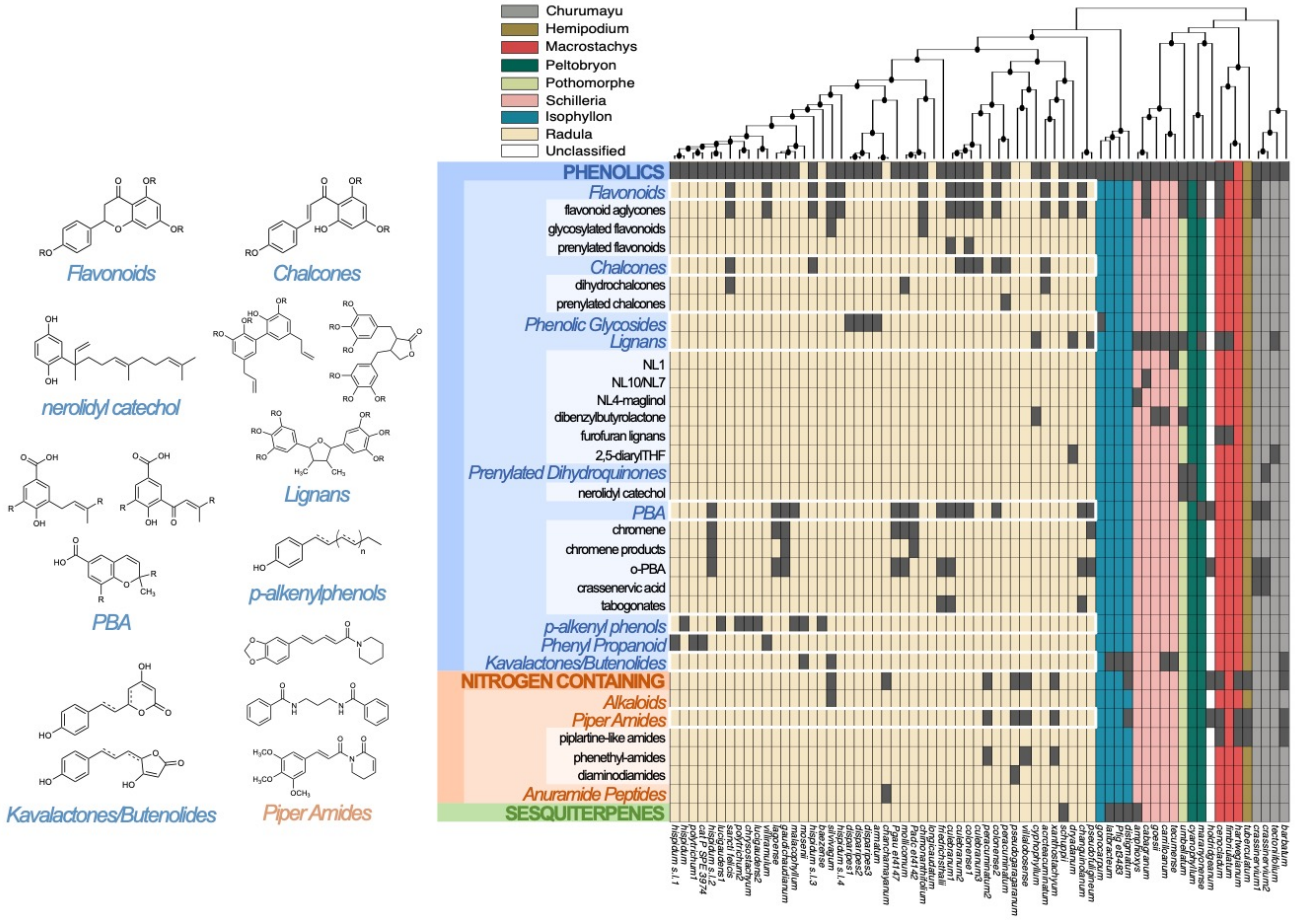

Churumayu
Hemipodium
Macrostachys
Peltobryon
Pothomorphe
Schilleria
Isophyllon
Radula
Unclassified
PHENOLICS
Flavonoids
flavonoid aglycones
glycosylated flavonoids
prenylated flavonoids
Chalcones
dihydrochalcones
prenylated chalcones
Phenolic Glycosides
Lignans
NL1
NL10/NL7
NL4-maglinol
dibenzylbutyrolactone
furofuran lignans
2,5-diarylTHF
Prenylated Dihydroquinones
nerolidyl catechol
PBA
chromene
chromene products
o-PBA
crassenervic acid
tabogonates
p-alkenyl phenols
Phenyl Propanoid
Kavalactones/Butenolides
NITROGEN CONTAINING
Alkaloids
Piper Amides
piplartine-like amides
phenethyl-amides
diaminodiamides
Anuramide Peptides
SESQUITERPENES
Flavonoids
Chalcones
nerolidyl catechol
Lignans
p-alkenylphenols
PBA
Kavalactones/Butenolides
Piper Amides

Figure 3

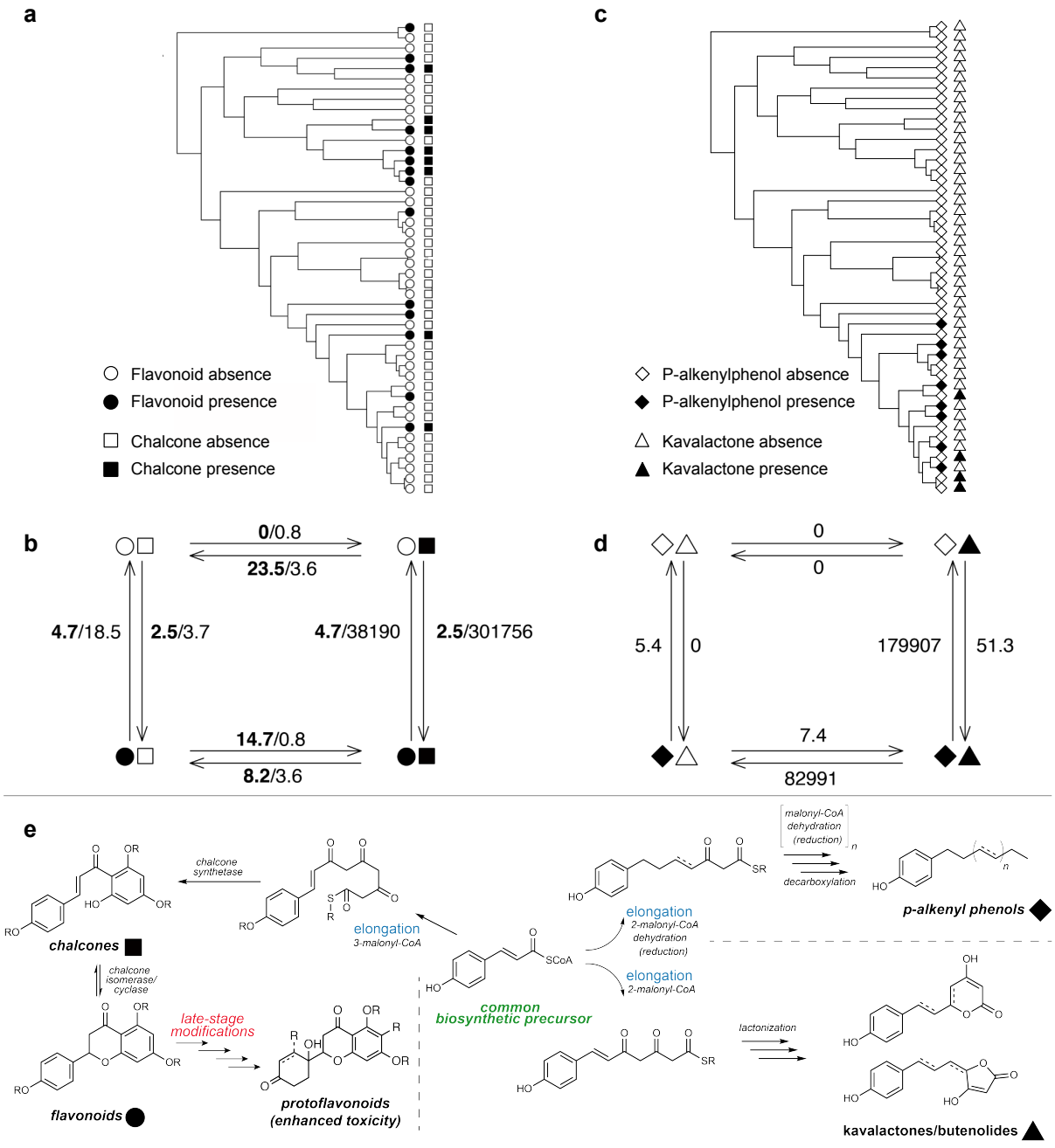
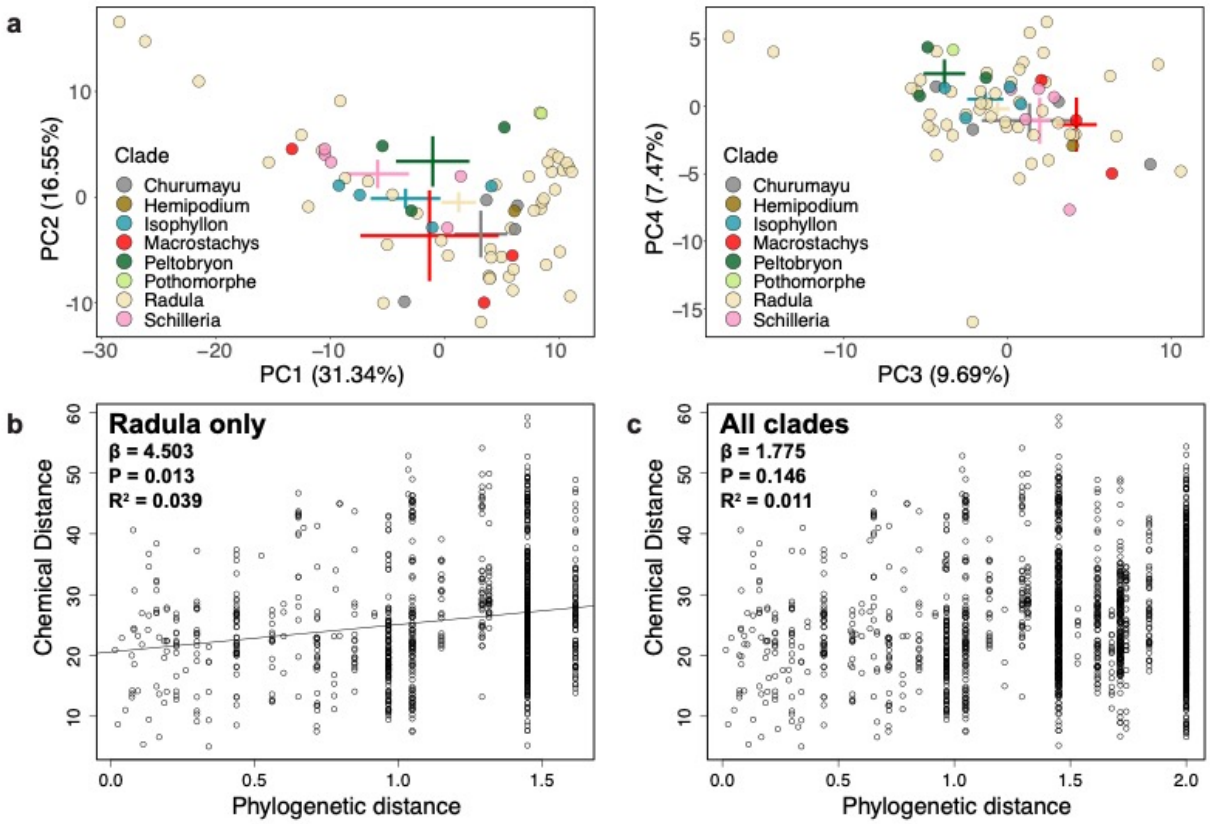

a
Flavonoid absence
Flavonoid presence
Chalcone absence
Chalcone presence
c
P-alkenylphenol absence
P-alkenylphenol presence
Kavalactone absence
Kavalactone presence
b
0/0.8
23.5/3.6
4.7/18.5
2.5/3.7
4.7/38190
2.5/301756
14.7/0.8
8.2/3.6
d
0
0
5.4
0
179907
51.3
7.4
82991
e
chalcone synthetase
chalcones
elongation
3-malonyl-CoA
chalcone isomerase/ cyclase
late-stage modifications
flavonoids
protoflavonoids (enhanced toxicity)
common biosynthetic precursor
elongation
2-malonyl-CoA dehydration (reduction)
elongation
2-malonyl-CoA
malonyl-CoA dehydration (reduction)
decarboxylation
p-alkenyl phenols
lactonization
kavalactones/butenolides

Figure 4


a
PC2 (16.55%)
PC1 (31.34%)
PC4 (7.47%)
PC3 (9.69%)
Clade
Churumayu
Hemipodium
Isophyllon
Macrostachys
Peltobryon
Pothomorphe
Radula
Schilleria
b
Radula only
β = 4.503
P = 0.013
R² = 0.039
c
All clades
β = 1.775
P = 0.146
R² = 0.011
Chemical Distance
Phylogenetic distance

Supplementary Information

Phytochemistry reflects different evolutionary history in traditional classes versus specialized structural motifs

Kathryn A. Uckele, Joshua P. Jahner, Eric J. Tepe, Lora A. Richards, Lee A. Dyer, Kaitlin M. Ochsenrider, Casey S. Philbin, Massuo J. Kato, Lydia F. Yamaguchi, Matthew L. Forister, Angela M. Smilanich, Craig D. Dodson, Christopher S. Jeffrey, Thomas L. Parchman

Supplementary Methods

Nuclear magnetic resonance spectroscopy

Foliar tissue (100.0-2000.0 mg) from each sample was cooled with liquid nitrogen and ground to a fine powder using a mortar and pestle and was combined with 10 mL of methanol (HPLC grade, Fisher Scientific, Pittsburgh, PA). Samples were vortexed for 30 seconds, sonicated for 10 minutes before the supernatant was decanted and syringe filtered (0.45 μm). This process was repeated once on the retained plant material and filtered supernatants were combined and concentrated to dryness using a Genevac centrifugal evaporator. Samples were placed on a hi-vac overnight before determining extract mass. All samples underwent a deuterium exchange (HDX) to minimize proton peaks in ^{1}H NMR spectra by thrice reconstituting in methanol-d1 (Cambridge Isotope Laboratories, Tewksbury, MA) and re-drying. In cases where extract mass was above 20 mg, extracts were aliquoted to standardize extract mass to 13.1±3.8 mg/mL, and in some cases extract mass was below this mean (2.0-17.9 mg). Dried deuterated extracts were reconstituted in 600 μL methanol-d4 with 0.01% TMS and syringe filtered into NMR tubes for ^{1}H NMR analysis. FIDs were collected for each sample on a Varian 400 MHz NMR spectrometer (128 scans) before processing using MestReNova software (Mestrelab Research, Spain). Spectra were aligned to the residual methanol solvent peak (C$\underline{D}_3$OD, 3.31 ppm), Global phase-corrected, Whittaker baseline-corrected, and peak picked before being used for structural annotation. Further isolation was completed on crude extracts of *P. holdridgeanum, P. cabagranum,* and *P. peracuminatum* using a combination of flash column chromatography and reverse phase C18 preparatory medium pressure chromatography. Structural elucidation was completed on these isolated compounds using ^{1}H, ^{13}C, COSY, HSQC, NOESY, and HMBC NMR spectroscopy.

Liquid and gas chromatography-mass spectrometry

During the first step of HDX, 100 uL of ^{1}H NMR stock was was diluted 1:10 with protonated methanol for LC-MS analysis. Extracts were injected (1 µL) onto an Agilent (Santa Clara, CA) 1200 analytical HPLC equipped with a binary pump, autosampler, column compartment and diode array UV detector and eluted at 0.500 mL/min through a Kinetex EVO C18 column (Phenomenex, 2.1 x 100 mm, 2.6 µ, 100 Å; Torrance, CA) at 40 °C. The linear binary gradient was comprised of buffers A (Optima-grade water containing 0.1 % formic acid, Fisher Scientific) and B (Optima-grade acetonitrile containing 0.1 % formic acid, Fisher Scientific) changing over 20 minutes accordingly: 0-1 min 20% B, ramp to 50% B at 6 min, ramp to 100% B at 12 min, 12-16 min hold at 100% B, 16-17 min ramp to 20% B, 17-20 min hold at 20% B. Liquid chromatography was coupled to an Agilent 6230 Time-of-Flight mass spectrometer via an electrospray ionization source (ESI-TOF; gas temperature: 325 °C, flow: 10 L/m; nebulizer pressure: 35 psig; VCap: 3500 V; fragmentor: 165 V; skimmer: 65 V; octopole: 750 V). Raw data were processed and analysed in Agilent MassHunter.

A portion of the crude extract (~ 1-3 mg) was dissolved in CH_2Cl_2 (2 mL) and analyzed in the UNR Chemistry Department Shared Instrument Lab using an HP Agilent 7890A GC System coupled with an Agilent 5975C MSD (Agilent Technologies, Santa Clara, CA, USA), equipped with a DB-Ultra Inert capillary column (30 m x 250 µm x 0.25 µm; Agilent J&W GC Columns, Santa Clara, CA, USA). The carrier gas used was ultra-pure He set at a flow rate of 1.0 mL/min with a pressure of XX psi with an injection port temperature of 250 °C. Initial oven temperature 80 °C, with an initial hold time of 1.5 min, then ramping at 17°C/min to 300 °C, with a final hold time of 15 min. The resulting GC-MS data were recorded and processed using

MassHunter Workstation Quantitative Analysis software. Comparisons to literature values and NIST database searches was used to complete the categorical analysis, which is detailed in the supplementary table S2.

Table S1. Sampling information for all taxa. Herbarium acronyms are as follows: CINC=Margaret H. Fulford Herbarium, University of Cincinnati; CR=Herbario Nacional at the Museo Nacional de Costa Rica, San José, Costa Rica; PMA=Herbarium at the Universidad de Panamá, Panamá; QCNE=Herbario Nacional del Ecuador at the Museo Ecuatoriano de Ciencias Naturales del Instituto Nacional de Biodiversidad, Quito, Ecuador; SPF=Herbarium of the Universidade de São Paulo, São Paulo, Brazil; USM=Herbarium of the Universidad Nacional de San Marcos, Lima, Perú.

Species	Country	Collector	Voucher (herbarium)	Clade
P. aduncum var. *cordulatum* (C.DC.) Yunck.	Brazil	M. Kato	*M. Kato K-1978 (SPF)*	Radula
P. arcteacuminatum Trel.	Costa Rica	E.J. Tepe	*E.J. Tepe 3534 (CR)*	Radula
P. amphioxys Trel.	Panamá	E.J. Tepe	*E.J. Tepe 4044 (PMA)*	Schilleria
P. armatum Trel. & Yunck.	Peru	E.J. Tepe	*E.J. Tepe 4394 (USM)*	Radula
P. baezense Trel.	Ecuador	A.E. Glassmire	*A.E. Glassmire YY1 (CINC)*	Radula
P. barbatum Kunth	Ecuador	E.J. Tepe	*E.J. Tepe 3005 (QCNE)*	Churumayu
P. cabagranum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3531 (CR)*	Schilleria
P. carrilloanum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 4069 (PMA)*	Schilleria
P. cenocladum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 3999 (PMA)*	Macrostachys
P. chanchamayanum Trel.	Peru	E.J. Tepe	*E.J. Tepe 4393 (USM)*	Radula
P. changuinolanum Trel.	Panamá	E.J. Tepe	*E.J. Tepe 3961 (PMA)*	Radula
P. chimonanthifolium Kunth	Brazil	M. Kato	*M. Kato K-1960 (SPF)*	Radula
P. chrysostachyum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3482 (CR)*	Radula
P. colonense C.DC. (1)	Panamá	E.J. Tepe	*E.J. Tepe 4032 (PMA)*	Radula
P. colonense C.DC. (2)	Costa Rica	E.J. Tepe	*E.J. Tepe 3502 (CR)*	Radula
P. crassinervium Kunth (1)	Costa Rica	E.J. Tepe	*E.J. Tepe 3463 (CR)*	Churumayu
P. crassinervium Kunth (2)	Brazil	M. Kato	*M. Kato K-1954 (SPF)*	Churumayu
P. culebranum C.DC. (1)	Costa Rica	E.J. Tepe	*E.J. Tepe 3413 (CR)*	Radula
P. culebranum C.DC. (2)	Costa Rica	E.J. Tepe	*E.J. Tepe 3527 (CR)*	Radula
P. culebranum C.DC. (3)	Panamá	E.J. Tepe	*E.J. Tepe 4005 (PMA)*	Radula
P. cyanophyllum Trel.	Costa Rica	E.J. Tepe	*E.J. Tepe 3427 (CR)*	Peltobryon
P. cyphophyllum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3541 (CR)*	Radula
P. disparipes Trel. (1)	Costa Rica	E.J. Tepe	*E.J. Tepe 3426 (CR)*	Radula
P. disparipes Trel. (2)	Costa Rica	E.J. Tepe	*E.J. Tepe 3543 (CR)*	Radula
P. disparipes Trel. (3)	Costa Rica	E.J. Tepe	*E.J. Tepe 3486 (CR)*	Radula
P. distigmatum Yunck.	Panamá	E.J. Tepe	*E.J. Tepe 3972 (PMA)*	Isophyllon
P. dryadanum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 3963 (PMA)*	Radula
P. euryphyllum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 4000 (PMA)*	Macrostachys
P. figlinum Trel.	Costa Rica	E.J. Tepe	*E.J. Tepe 3483 (CR)*	Isophyllon
P. fimbriulatum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 3956 (PMA)*	Macrostachys
P. friedrichsthalii C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 131 (CR)*	Radula
P. gaudichaudianum Kunth (1)	Brazil	M. Kato	*M. Kato K-1949 (SPF)*	Radula
P. gaudichaudianum Kunth (2)	Brazil	M. Kato	*M. Kato K-1983 (SPF)*	Radula
P. goesii Yunck.	Brazil	M. Kato	*M. Kato K-1964 (SPF)*	Schilleria
P. gonocarpum Trel.	Panamá	E.J. Tepe	*E.J. Tepe 3959 (PMA)*	Isophyllon
P. hartwegianum (Benth.) C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 3966 (PMA)*	Macrostachys
P. hispidum Sw. (1)	Costa Rica	E.J. Tepe	*E.J. Tepe 3430 (CR)*	Radula

P. hispidum s.l. (2)	Costa Rica	E.J. Tepe	*E.J. Tepe 3476 (CR)*	Radula
P. hispidum Sw. (3)	Costa Rica	E.J. Tepe	*E.J. Tepe 3496 (CR)*	Radule
P. hispidum Sw. (4)	Costa Rica	E.J. Tepe	*E.J. Tepe 3509 (CR)*	Radula
P. hispidum Sw. (5)	Costa Rica	E.J. Tepe	*E.J. Tepe 3537 (CR)*	Radula
P . holdridgeanum W.C.Burger	Costa Rica	E.J. Tepe	*E.J. Tepe 4196 (CR)*	Unclassified
P. lagoense C.DC.	Brazil	M. Kato	*M. Kato K-1944 (SPF)*	Radula
P. latibracteum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 4050 (PMA)*	Isophyllon
P. longicaudatum Trel. & Yunck.	Ecuador	E.J. Tepe	*E.J. Tepe 3036 (QCNE)*	Radula
P. lucigaudens C.DC. (1)	Panamá	E.J. Tepe	*E.J. Tepe 3993 (PMA)*	Radula
P. lucigaudens C.DC. (2)	Panamá	E.J. Tepe	*E.J. Tepe 4028 (PMA)*	Radula
P. malacophyllum (C.Presl.) C.DC.	Brazil	M.Kato	*M. Kato K-1945 (SPF)*	Radula
P. maranyonense Trel.	Ecuador	E.J. Tepe	*E.J. Tepe 3034 (QCNE)*	Peltobryon
P. mollicomum Kunth	Brazil	M. Kato	*M. Kato K-1942 (SPF)*	Radula
P. mosenii C.DC.	Brazil	M. Kato	*M. Kato 1948 (SPF)*	Radula
P. peracuminatum C.DC. (1)	Costa Rica	E.J. Tepe	*E.J. Tepe 3433 (CR)*	Radula
P. peracuminatum C.DC. (2)	Panamá	E.J. Tepe	*E.J. Tepe 4062 (PMA)*	Radula
P. persubulatum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 4068 (PMA)*	Radula
P. polytrichum C.DC. (1)	Panamá	E.J. Tepe	*E.J. Tepe 3965 (CR)*	Radula
P. polytrichum C.DC. (2)	Costa Rica	E.J. Tepe	*E.J. Tepe 3470 (CR)*	Radula
P. pseudofuligineum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3499 (CR)*	Radula
P. pseudogaragaranum Trel.	Panamá	E.J. Tepe	*E.J. Tepe 4041 (PMA)*	Radula
P. sancti-felicis Trel.	Costa Rica	E.J. Tepe	*E.J. Tepe 3415 (CR)*	Radula
P. schuppii A.H. Gentry	Ecuador	E.J. Tepe	*E.J. Tepe 1562 (QCNE)*	Radula
P. silvivagum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3523 (CR)*	Radula
P. sp	Panamá	E.J. Tepe	*E.J. Tepe 3974 (PMA)*	Radula
P. tectoniifolium Kunth	Brazil	M. Kato	*M. Kato K-1958 (SPF)*	Churumayu
P. tecumense Trel.	Panamá	E.J Tepe	*E.J. Tepe 4008 (PMA)*	Schilleria
P. tuberculatum Jacq.	Panamá	E.J. Tepe	*E.J. Tepe 4039 (PMA)*	Hemipodion
P. umbellatum L.	Panamá	E.J. Tepe	*E.J. Tepe 3967 (PMA)*	Pothomorphe
P. vicosanum Yunck.	Brazil	M. Kato	*M. Kato K-1966 (SPF)*	Isophyllon
P. villalobosense Yunck.	Ecuador	E.J. Tepe	*E.J. Tepe 2952 (QCNE)*	Radula
P. villiramulum C.DC.	Panamá	E.J. Tepe	*E.J. Tepe 4031 (PMA)*	Radula
P. xanthostachyum C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3542 (CR)*	Radula
P. zacatense C.DC.	Costa Rica	E.J. Tepe	*E.J. Tepe 3438 (CR)*	Radula

Table S2. Combined evidence from ^{1}H NMR, GC-MS, LC-MS, LC-UV, and in few cases, 2D-NMR was used to assign metabolite classes to each species. This table includes relevant high-resolution masses, nominal and fragment masses and NMR chemical shifts that were used to classify the compounds present in each individual. In cases where published literature values provided structural confirmation, those reference(s) are listed.

Voucher (herbarium)	Species	Categorical observations	References
E.J. Tepe 3005 (QCNE)	*P. barbatum*	Crude ^{1}H NMR analysis supports the presence of a dihydropyridone ring of piplartine and a minor analog. The major component indicated by ^{1}H NMR is 3,4,5-trimethoxycinnamic acid methyl ester. GC and LC-MS conforms with the ^{1}H NMR analysis, with the major component as the methyl ester along with the carboxylic acid. Additionally, this analysis supports the presence of dihydropiplartine.	[1]
M. Kato K-1958 (SPF)	*P. tectoniifolium*	Crude ^{1}H NMR analysis is consistent with the two *meso*-stereoisomers of grandisin as the major components, with the all syn isomer as the major of the two stereoisomers. Chemical shift analysis is consistent with Kato, et al and other reported values for grandisin. GC-MS analysis confirms the presence of the grandisin stereoisomers.	[2]
E.J. Tepe 3463 (CR)	*P. crassinervium*	Crude ^{1}H NMR analysis is consistent with data reported by Kato et al in previous studies of the natural products chemistry of *P. crassinervium*. The mixture is dominated by the known dihydroquinone, with minor amounts of crassinervic acid, its analogs, and other oxidized prenylated benzoic acid derivatives.	[3-6]
M. Kato K-1954 (SPF)	*P. crassinervium*	The mixture is dominated by crassinervic acid, which was identified by crude ^{1}H NMR analysis and overlapped with the NMR data reported in the literature. Additional analysis by GC-MS confirmed the presence of crassinervic acid as the major component along with other oxidized PBAs that have been previously isolated from *P. crassinervium*. GC-MS analysis confirmed the presence of the flavone sakuranetin, whose fragmentation and molecular ion were consistent with that reported in the NIST database.	[3-6]
E.J. Tepe 4039 (PMA)	*P. tuberculatum*	Characteristic resonances in the crude ^{1}H NMR spectrum indicated the presence of piplartine, which is consistent with literature data for this compound and previous natural product studies of *P. tuberculatum*. LC-MS analysis confirmed the presence of piplartine as well as other minor piper amides (piperolyene and dihydropiperolyene), with other minor long-chain isobutyl amides. GC-MS analysis confirms the presence of piplartine (major), 4,5-Dihydropiperlonguminine, desmethoxypiplartine, and another piperonal derived amide (m/z = 287 with a 135 bp).	[7,8]
E.J. Tepe 3999 (PMA)	*P. cenocladum*	Piplartine, cenocladamide, and sintenpyridone all confirmed by GC-MS and supported by characteristic resonances in the ^{1}H NMR spectrum that overlap with values reported in previous studies of the natural product chemistry of *P. cenocladum*. GC-MS analysis matches that originally reported for this species.	[9]
E.J. Tepe 3956 (PMA)	*P. fimbriulatum*	Crude LC-MS analysis reveals 3 major peaks that are consistent with the bis-furan neolignans diayangambin, sesartemin, and an isomeric analog. ^{1}H-NMR resonances in the crude extract are consistent with the presence of these types of neolignans.	[10-14]

E.J. Tepe 3966 (PMA)	*P. hartwegianum*	^{1}H NMR shows resonances characteristic of piplartine-like amides. LC-MS confirms piplartine as major component of the mixture along with flavonoids/flavones that are supported by ^{1}H NMR analysis. LC-MS analysis also indicates the presence of minor piplartine analogs and long chain piperonal type amides.	
E.J. Tepe 4196 (CR)	*P. holdridgeanum*	Holdrigeanic acid (a prenylated benzoic acid derivative) was found as the major compound that was isolated and characterized using 2D-NMR analysis. LC-MS analysis of the crude extract indicated a variety of other long-chain amides along with the holdrigeanic acid.	Jeffrey and Oliveira, unpubl. data
E.J. Tepe 3034 (QCNE)	*P. maranyonense*	Resonances of 4-nerolidylcatechol type phenol are clearly apparent in the NMR and indicates this as the major component with minor amounts of related compounds. LC-MS analysis supports the presence of a methylated 4-nerolidylcatechol along with other related components.	[15-19]
E.J. Tepe 4069 (PMA)	*P. carrilloanum*	Crude ^{1}H NMR analysis indicates the presence of eupomatenoid-like lignans with evidence of terminal double bonds, many oxygenated, CHOR, Methoxy, methyl doublets. LC-MS is consistent with the presences of an array of these lignans.	[20-26]
E.J. Tepe 4008 (PMA)	*P. tecumense*	Very complex ^{1}H NMR spectrum indicative of a mixture of (neo)lignans, which is supported by a variety of peaks with the molecular ion of m/z = 416 in the LC-MS and the GC-MS fragmentation patterns. The presence of methyl doublets and a highly complex aromatic region is supportive of this conclusion. Sodium adduct is more apparent in the LC-MS, which is a common observation in the ESI-MS analysis of neolignans that are clusin/cubeba like. Masses related to compounds isolated from *P. clusii* and *P. cubeba.*	[27-29]
M. Kato K-1964 (SPF)	*P. goesii*	Crude ^{1}H NMR analysis indicates characteristic lignan resonances with LC-MS and GC-MS indicating (-)-5"-Methoxyhinokinin (MW = 384) isolated from *P. cubeba* and *P. trichostachyon.*	[30]
E.J. Tepe 4044 (PMA)	*P. amphioxys*	Crude ^{1}H NMR analysis indicates characteristic lignan resonances with LC-MS and GC-MS indicating similar patterns to other lignan rich *Piper* species with the observation of peaks with m/z = 296, 372, 326, 356, and 374.	
E.J. Tepe 3961 (PMA)	*P. changuinolanum*	^{1}H NMR analysis suggests the presence of 2-4 sets of tri-substituted aromatic groups with prenyl alkene and methyl resonances suggests a mixture of prenylated benzoic acid derivatives. Additional downfield resonances at 8.5 ppm are characteristic of the oxidized type of PBA similar to resonances found *ortho* to the phenol in methyl tabogonate. LCMS confirms a complex mixture of prenylated phenols along with their oxidized derivatives and GC- and LC-MS indicate the presence of minor flavonoids.	[4, 31-33]
E.J. Tepe 3499 (CR)	*P. pseudofuligineum*	^{1}H NMR indicates flavones, with two trisubstituted aromatics, a cyclized methyl tabagonate, and its carboxylic acid derivative (LC-MS, m/z = 220.22 and 234.25). A minor prenylated flavanone based upon ^{1}H NMR evidence and mass defect in the HRMS is also present.	[4, 31-33]
E.J. Tepe 3963 (PMA)	*P. dryadanum*	^{1}H NMR analysis and GC-MS shows almost exclusively grandisin, with the *anti*-C2 symmetric stereochemistry that is consistent with the NMR resonances reported in the literature.	[2]

E.J. Tepe 1562 (QCNE)	*P. schuppii*	GC-MS suggests sesquiterpenes and a flavanone by comparisons to other GC-MS spectra and the NIST database.	
E.J. Tepe 3534 (CR)	*P. arcteacuminatum*	GC-MS and LC-MS indicate phenethylamine or dihydrochalcone type piper amides (m/z = 288). ^{1}H NMR is consistent with this analysis, clearly showing two sets of triplets of ethylene system and aromatic resonances corresponding to a benzamide or cinnamate type of mono-substitution. These compounds are similar to the altamide/tembamide amides that have been previously isolated from *Piper*.	[34]
E.J. Tepe 3542 (CR)	*P. xanthostachyum*	^{1}H NMR analysis is consistent with the presence of flavonoids, and this is supported by GC-MS and LC-MS analysis, which indicates 4 major flavonoid components. Clear *para*-hydroxybenzyl fragmentation is indicated in the GC-MS of the flavonoids. Some indication of minor amides in the LC-MS.	
E.J. Tepe 3541 (CR)	*P. cyphophyllum*	^{1}H NMR analysis is consistent for the presence of a mixture of lignans with methyl doublets and complexity in the oxygenated and aromatic regions of the crude ^{1}H NMR spectrum. LC-MS is consistent with lignans and/or isoprenylated flavanoids, however NMR is consistent with lignans.	
E.J. Tepe 4041 (PMA)	*P. pseudogaragaranum*	LC-MS and crude ^{1}H NMR is consistent with benzoylated tyramine (phenethylamides) with dihydrochalcones. Dihyrdoxymethoxydihydrochalcones (2) and tyramine with a methoxybenzoyl group.	[34]
E.J. Tepe 2952 (QCNE)	*P. villalobosense*	^{1}H NMR cleanly shows two major compounds with the LC-MS demonstrating two major components [m/z = 311.1785 (M+H) and 297.1626 (M+H)]. ^{1}H-NMR MS are consistent for two related bis amides *N,N′*-dibenzoylputricene and *N,N′*-dibenzoylcadavine, which have been isolated previously from *Haplophyllum*.	[35-38]
E.J. Tepe 3433 (CR)	*P. peracuminatum*	^{1}H NMR clearly shows a prenylated and diprenylated dihydrochalcone that were confirmed by isolation and full 2D spectroscopic characterization. LC-MS analysis confirms the presence of these dihydrochalcone derivatives.	Jeffrey and Oliveira, unpubl. data
E.J. Tepe 3502 (CR)	*P. colonense*	Two chalcones (one major: one minor) are clearly indicated in the ^{1}H NMR spectrum the LC- and GC-MS with their corresponding flavanones. Sakuranetin is present in the GC- and LC-MS analysis and confirmed by the NIST database. ^{1}H NMR analysis also indicates one prenyl group that is consistent with the presence of methyl 3-prenyl-4-hydroxybenzoate in the LC/GC-MS. Clear diastereotopic ABX pattern indicative of a dihydroflavone as one of the major components.	
E.J. Tepe 4062 (PMA)	*P. peracuminatum*	^{1}H NMR analysis clearly shows resonances that are characteristic of *trans*-double bonds (J ~ 15 Hz). Additional complexity across the downfield and upfield regions of the spectrum with a clear CH_2CH_2 phenethyl structure indicated by two triplets. Cinnamoyl fragmentation in the GC- and LC-MS suggests a mixture of cinnamylphenethylamides related to those that have been isolated from other species of *Piper*.	
E.J. Tepe 3413 (CR)	*P. culebranum*	LC- and GC-MS provide evidence of two chalcones that are identical to those found in *P. colonense* (m/z = 314 and 270) along with the corresponding flavanones, and sakuranetin. Neutral loss of water from m/z = 284 indicating hydroxyflavone of m/z = 256.	

E.J. Tepe 4032 (PMA)	*P. colonense*	2 chalcones. Clear diastereotopic ABX pattern indicative of a dihydroflavone as one of the major components. Prenyl alkene resonance is consistent with the presence of methyl 3-prenyl-4-hydroxybenzoate. 2 chalcones (one major: one minor) with their corresponding flavanones, including sakuranetin.	
E.J. Tepe 3527 (CR)	*P. culebranum*	Methyl 3-prenyl-4-hydroxybenzoate is clearly present in the LC-MS and is supported in the ^{1}H NMR with one prenylated alkene resonance. Other peaks correspond to the presence of two chalcones. Clear diastereotopic ABX pattern indicative of a dihydroflavone as one of the major components. Prenyl checks out with methyl 3-prenyl-4-hydroxybenzoate. 2 chalcones (one major: one minor) with their corresponding flavanones. Sakuranetin.	
E.J. Tepe 4005 (PMA)	*P. culebranum*	^{1}H NMR analysis supports the presence of methyl tabagonate and its methyl ether, which is further supported in the LC/GC-MS with m/z = 314, 270, 284. Additional chalcones and flavanones present.	[4, 31-33]
E.J. Tepe 131 (CR)	*P. friedrichsthalii*	Spectroscopic analysis consistent with a cyclized (chromanone) of methyl tabagonate and an oxidized tabagonate with a third related compound.	[4, 31-33]
E.J. Tepe 3036 (QCNE)	*P. longicaudatum*	LC-MS yielded a peak that corresponded to the mass of anofinic acid methyl ester ($[M+H]^{+}$ = 235.0997) that could not be confirmed by NMR due to low signal. Anofinic acid occurs naturally in mushrooms and is similar in structure to gaudichaudianic acid Stigmasterol. Sitosterol are all apparent in the GC-MS with 2-aminosteroids/triterpenes (C30+NH2). The aminosteroid mass was confirmed by LC-MS: $[M+H]^{+}$ = 420.3503.	[39]
M. Kato K-1960 (SPF)	*P. chimonanthifolium*	LC-MS is indicative of the presence of glycosylated. There are some interesting protons in the 6.2-6.9 range that could suggest chromenes or a styrenyl alkene. GC-MS indicates m/z = 284, 256, which are supportive of the presence of 5,7-dihydroxyflavanone and wogonin.	
M. Kato K-1942 (SPF)	*P. mollicomum*	NMR is low signal, however LC-MS shows the acid aduncum chromene, dihydroxyp-methoxy dihydrochalcone, piperaduncin A, and piperaduncin B.	[40]
M. Kato K-1978 (SPF)	*P. aduncum* var. *cordulatum*	LC-MS flavanoid glycosides, which is consistent with the ^{1}H NMR data. Dihydrochalcones similar to what has been previously isolated from *Piper aduncum* are indicated. Additional, chromene type structures that are related to aduncum chromenes and their oxidized derivatives.	[41-44]
M. Kato K-1983 (SPF)	*P. gaudichaudianum*	Spectroscopic analysis consistent with compounds previously reported from *Piper gaudichaudianum.*	[4, 31, 45]
E.J. Tepe 4394 (USM)	*P. armatum*	The NMR and GC-MS suggests a benzoylamide similar to the phenylalanine peptides that have been isolated previously from *Piper aurantiamide* A and B. Significant ^{1}H NMR resonance overlap with these reports. LC-MS indicates a m/z = 253, sometimes 254 could be a fragment.	[46]
E.J. Tepe 4393 (USM)	*P. chanchamayanum*	The phytochemical profile was surprisingly similar to that of *P. disparipes*, having all three of phenolic glycosides found there.	

E.J. Tepe 3426 (CR) *E.J. Tepe 3486 (CR)* *E.J. Tepe 3543 (CR)*	*P. dispariipes*	An early eluting peak (0.6 min) with a mass defect suggesting a phenolic glycoside ($[M+H]^+$ = 543.1347), a peak at 0.75 min that had a mass suggesting a trihydroxy flavanone glycoside (METLIN match, $[M+H]^+$ = 565.1627), and a peak at 0.9 min that had a mass suggesting a tetrahydroxy flavanone (such as kaempferol) glycoside (METLIN match, $[M+H]^+$ = 579.1777). These predominant flavonoid glycosides were found in all three *P. dispariipes* individuals, which all had very similar chemistry. These proposed classes could not be confirmed by ^{1}H NMR.	
E.J. Tepe 3430 (CR)	*P. hispidum*	7+ flavonoids are clearly indicated in the GC-MS and is additionally supported by the ^{1}H NMR spectrum.	
E.J. Tepe 3523 (CR)	*P. silvivagum*	We found a peak that was attributed to anofinic acid methyl ester in *P. longicaudatum.* Both the retention time (5.8 min) and *m/z* matched. We also found a mass that corresponded to trihydroxy biphenyl structure (METLIN match; $[M+H]^+$ = 203.0729), which was categorized as a phenyl propanoid, and two aliphatic amides (METLIN match; $[M+H]^+$ = 404.3599, 446.3683).	
A.E. Glassmire YY1 (CINC)	*P. baezense*	GC-MS clearly indicates a series of 4-alkenyl phenols: 2 dienyl-C12; 1-C12 monoalkenyl, and 1-C14 alkenyl along with polyunsaturated fatty acids. ^{1}H NMR is characteristic for the presence of alkenyl phenols.	[47]
E.J. Tepe 3476 (CR)	*P. hispidum*	Two major flavonoids indicated by GC-MS, and this is clearly supported in the crude ^{1}H NMR spectrum.	
M. Kato 1948 (SPF)	*P. mosenii*	^{1}H NMR and MS were inconclusive, although tabogenates have been found in the literature, they were not found in our analysis.	[33]
M. Kato K-1945 (SPF)	*P. malacophyllum*	^{1}H NMR and MS are consistent with *p*-alkenyl phenols that have been previously isolated from this species. LC-MS and NMR confirm the presence of gibbilimbol.	[48]
M. Kato K-1944 (SPF)	*P. lagoense*	^{1}H NMR analysis indicates chromene type peaks similar to those isolated from *Piper kelleyii.* LC-MS indicates isobaric compounds to the *kelleyii* chromene/guachaudianic acid, oxidation products. A farnesylated hydroxy benzoic acid derivative is also indicated. Dimeric compounds and oxidized dimeric compounds are indicated in the LC-MS.	
M. Kato K-1949 (SPF)	*P. gaudichaudianum*	Gaudichaudianic acid and related prenylated compounds are clearly indicated by LC-MS and ^{1}H NMR analysis.	[4]
E.J. Tepe 3993 (PMA)	*P. lucigaudens*	Low signal in the NMR, however, there are peaks that align with pyrone kava lactones of 5,6-dehydrokavain or desmethoxyyangonin and its dimer (perhaps [2+2]). Additional flavonoids are indicated in the GC/LC-MS.	[49]
E.J. Tepe 4031 (PMA)	*P. villiramulum*	^{1}H NMR and GC-MS analysis for the presence of *p*-alkenylphenols that have been previously isolated from this species.	[50]
E.J. Tepe 3965 (CR)	*P. polytrichum*	^{1}H NMR and GC-MS indicates the presence of three major alkenyl phenols, m/z = 344, 314 (diene) along with fatty acids related to those previously isolated from *Piper* species.	[48]
E.J. Tepe 3482 (CR)	*P. chrysostachyum*	^{1}H NMR and GC-MS indicates the presence of a series of alkenyl phenols.	[47]
E.J. Tepe 3415 (CR)	*P. sancti-felicis*	GC-MS and LC-MS are both consistent with a mixture chalcones, dihydrochalcones, and flavonoids, with some mix of methylation. ^{1}H NMR is consistent with this analysis with distinct peaks for the chalcone alkene and a mono-substituted aromatic.	
E.J. Tepe 3496 (CR)	*P. hispidum*	GC-MS confirms the presence of *p*-alkenylphenols with additional unsaturated fatty acids.	

E.J. Tepe 4028 (PMA)	*P. lucigaudens*	LC-MS analysis indicates chromenes and related dimers that are isobaric to the *P. kelleyi* system. ^{1}H NMR is consistent with the presence of these type of compounds.	[49]
E.J. Tepe 3537 (CR)	*P. hispidum*	Major compound (m/z =275, M+H) and ^{1}H NMR are nearly identical to 3509 and 3974. Resonances match for a butenolide or kava-type lactone that has been previously isolated from *Piper sanctum.*	[51]
E.J. Tepe 3509 (CR)	*P. hispidum*	Nearly identical to *P. hispidum E.J. Tepe 3537 (CR).*	[51]
E.J. Tepe 3965 (CR)	*P. polytrichum*	^{1}H NMR and GC-MS analysis clearly indicate the presence of a series of *p*-alkenylphenols that are related to those previously isolated from *Piper* species.	[47]
E.J. Tepe 3974 (PMA)	*P.* sp.	Spectra are nearly identical to those of *P. hispidum E.J. Tepe 3537 (CR)* and *P. hispidum E.J. Tepe 3509 (CR).*	[51]

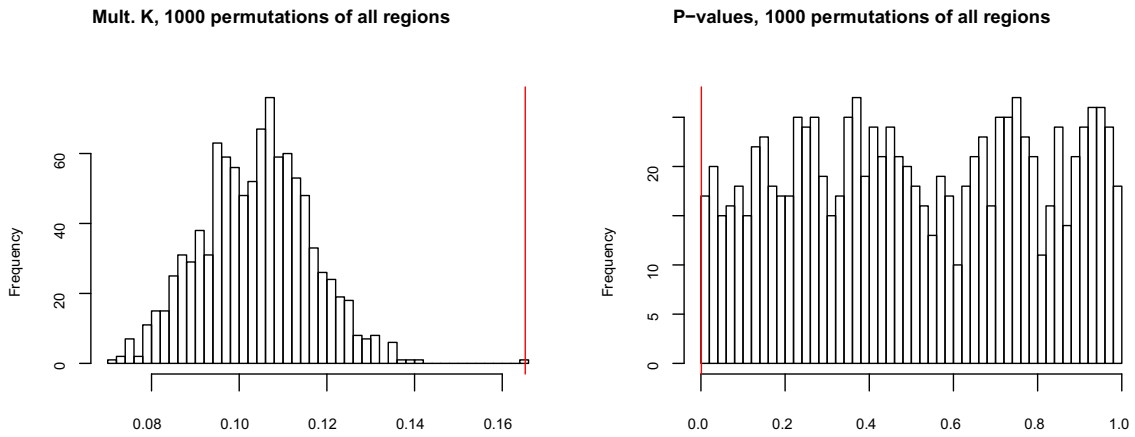


Figure S1. Results of the multivariate K test on 1000 permutations of all chemical regions indicate that our observed, significant phylogenetic signal is not an artifact of zero inflation exhibited by the ^{1}H NMR data. Vertical red lines represent our observed values for the multivariate K statistic and its associate *P*-value.

Supplementary References

1. Duh, C. Y., Wu, Y. C. & Wang, S. K. Cytotoxic pyridine alkaloids from *Piper aborescens*. *Phytochemistry* **29**, 2689-2691 (1990).
2. Ramos, C. S. *et al.* Configuration and stability of naturally occurring all-*cis*-tetrahydrofuran lignans from *Piper solmsianum*. *RSC Adv.* **7**, 46932-46937 (2017).
3. Danelutte, A. P., Lago, J. H. G., Young, M. C. M., & Kato, M. J. Antifungal flavanones and prenylated hydroquinones from *Piper crassinervium* Kunth. *Phytochemistry* **64**, 555-559 (2003).
4. Lago, J. H. G. *et al.* Benzoic acid derivatives from *Piper* species and their fungitoxic activity against *Cladosporium cladosporioides* and *C. sphaerospermum*. *J. Nat. Prod.* **67**, 1783-1788 (2004).
5. Lopes, A. A. *et al. In vitro* activity of compounds isolated from *Piper crassinervium* against *Trypanosoma cruzi*. *Nat. Prod. Res.* **22**, 1040-1046 (2008).
6. Yamaguchi, L. F., Lago, J. H. G., Tanizaki, T. M., Di Mascio, P., & Kato, M. J. Antioxidant activity of prenylated hydroquinone and benzoic acid derivatives from *Piper crassinervium* Kunth. *Phytochemistry* **67**, 1838-1843 (2006).
7. Chaves, M. C. O., Júnior, A. G. F., & Santos, B. V. O. Amides from *Piper tuberculatum* fruits. *Fitoterapia* **74**, 181-183 (2003).
8. da Silva, R. V. *et al.* Antifungal amides from *Piper arboretum* and *Piper tuberculatum*. *Phytochemistry* **59**, 521-527 (2002).
9. Dodson, C. D., Dyer, L. A., Searcy, J., Wright, Z., & Letourneau, D. K. Cenocladamide, a dihydropyridone alkaloid from *Piper cenocladamide*. *Phytochemistry* **53**, 51-54 (2000).
10. Ahmed, A. A. *et al.* Two highly oxygenated eudesmanes and 10 lignans from *Achillea holosericea*. *Phytochemistry* **59**, 851-856 (2002).
11. Mota, J. d. S. *et al. In vitro* trypanocidal activity of phenolic derivatives from *Peperomia obtusifolia*. *Planta Med.* **75**, 620-623 (2009).
12. Rios, M. Y., Ocampo-Acuña, Y. D., Ramírez-Cisneros, M. A., & Salazar-Rios, M. E. Furofuranone lignans from *Leucophyllum ambiguum. J. Nat. Prod.* **83**, 1424-1431 (2020).
13. Shao, S.-Y., Yang, Y.-N, Feng, Z.-M., Jiang, J.-S., & Zhang, P.-C. An efficient method for determining the relative configuration of furofuran lignans by ^{1}H NMR spectroscopy. *J. Nat. Prod.* **81**, 1023-1028 (2018).
14. Solís, P. N. *et al.* A new larvicidal lignan from *Piper fimbriulatum*. *Pharm. Biol.* **43**, 378-381 (2005).
15. Bagatela, B. S. *et al.* Evalution of antimicrobial and antimalarial activities of crude extract, fractions and 4-nerolidylcathecol from the aerial parts of *Piper umbellata* L. (Piperaceae). *Nat. Prod. Res.* **27**, 2202-2209 (2013).
16. Baldoqui, D. C., Bolzani, V. d. S., Furlan, M., Kato, M. J., & Marques, M. O. Flavones, lignans, and terpene from *Piper umbellata* (Piperaceae). *Quím. Nova* **32**, 1107-1109 (2009).
17. Kijjoa, A., Giesbrecht, A. M., Akisue, M. K., Gottlieb, O. R., & Gottlieb, H. E. 4-nerolidylcatechol from *Potomorphe umbellata*. *Planta Med.* **39**, 85-87 (1980).
18. da Silva Pinto, A. C. *et al. Piper peltatum*: biomass and 4-nerolidylcatechol production. *Planta Med.* **76**, 1473-1476 (2010).
19. Tabopda, T. K. *et al.* Bioactive aristolactams from *Piper umbellatum*. *Phytochemistry*, **69**, 1726-1731 (2008).

20. Carinin, M., Aldini, G., Orioli, M., & Facino, R. M. Antioxidant and photoprotective activity of a lipophilic extract containing neolignans from *Krameria triandra* roots. *Planta Med.* **68**, 193-197 (2002).
21. Freixa, B., Vila, R., Ferro, E. A., Adzet, T., & Cañigueral, S. Antifungal principles from *Piper fulvescens*. *Planta Med.* **67**, 873-875 (2001).
22. Johann, S. *et al.* Antifungal activities of compounds isolated from *Piper abutiloides* Kunth. *Mycoses* **52**, 499-506 (2009).
23. Macedo, A. L. *et al.* Isolation of a larvicidal compound from *Piper solmsianum* C.DC. (Piperaceae). *Nat. Prod. Res.* **32**, 2701-2704 (2018).
24. Marçal, F. J. B., Cortez, D. A. G., Ueda-Nakamura, T., Nakamura, C. V., & Dias Filho, B. P. Activity of the extracts and neolignans from *Piper regnellii* against methicillin-resistant *Staphylococcus aureus* (MRSA). *Molecules* **15**, 2060-2069 (2010).
25. Pessini, G. L., Dias Filho, B. P., Nakamura, C. V., Ferreira, A. G., & Cortez, D. A. G. Neolignans and the analysis of the essential oil of *Piper regnellii* (Miq.) C. DC. var. *pallescens* (C. DC.) Yunck leaves. *Rev. Bras. Farmacogn.* **15**, 199-204 (2005).
26. Scodro, R. B. L. *et al.* Anti-tuberculosis neolignans from *Piper regnellii*. *Phytomedicine*, **20**, 600-604 (2013).
27. Badheka, L. P., Prabhu, B. R., & Mulchandani, N. B. Dibenzylbutyrolactone lignans from *Piper cubeba*. *Phytochemistry*, **25**, 487-489 (1986).
28. Badheka, L. P., Prabhu, B. R., & Mulchandani, N. B. Lignans of *Piper cubeba*. *Phytochemistry* **26**, 2033-2036 (1987).
29. Prabhu, B. R., & Mulchandani, N. B. Lignans from *Piper cubeba*. *Phytochemistry* **24**, 329-331 (1985).
30. Koul, S. K., Taneja, S. C., Pushpangadan, P., & Dhar, K. L. Lignans of *Piper trichostachyon*. *Phytochemistry* **27**, 1479-1482 (1988).
31. Gaia, A. M., Yamaguchi, L. F., Jeffrey, C. S., & Kato, M. J. Age-dependent changes from allylphenol to prenylated benzoic acid production in *Piper gaudichaudianum* Kunth. *Phytochemistry* **106**, 86-93 (2014).
32. Terreaux, C., Gupta, M. P., & Hostettmann, K. Antifungal benzoic acid derivatives from *Piper dilatatum*. *Phytochemistry* **49**, 461-464 (1998).
33. Zermiani, T. *et al.* Morphological and phytochemical characterization of *Piper mosenii. Nat. Prod. Commun.* **14**, 1934578X1901400118 (2019).
34. Maxwell, A., & Rampersad, D. *β*-phenylethylamine-derived amides from *Piper guayranum. J. Nat. Prod.* **52**, 411-414 (1989).
35. Nesmelova, E. F., Bessonova, I. A., & Yunusov, S. Y. Haplamidine – a new alkaloid from *Haplophyllum latifolium*. *Khim. Prir. Soedin.* **3**, 427 (1977).
36. Nesmelova, E. F., Bessonova, I. A., & Yunusov, S. Y. The structure of haplamide and the synthesis of haplobucharine. *Khim. Prir. Soedin.* **2**, 289 (1977).
37. Nesmelova, E. F., Bessonova, I. A., & Yunusov, S. Y. Alkaloids of *Haplophyllum latifolium*. Structure of haplatine. *Khim. Prir. Soedin.* **6**, 758-764 (1978).
38. Nesmelova, E. F., Bessonova, I. A., & Yunusov, S. Y. Alkaloids of *Haplophyllum latifolium*. Structure and synthesis of haplamide and haplamidine. *Khim. Prir. Soedin.* **6**, 749-752 (1978).
39. Tan, R. X., Wolfender, J.-L., Ma, W. G., Zhang, L. X., & Hostettmann, K. Secoiridoids and antifungal aromatic acids from *Gentiana algida*. *Phytochemistry* **41**, 111-116 (1996).
40. Orjala, J., Wright, A. D., Behrends, H., Folkers, G., & Sticher, O. Cytotoxic and antibacterial dihydrochalcones from *Piper aduncum*. *J. Nat. Prod.* **57**, 18-26 (1994).

41. Baldoqui, D. C. *et al.* A chromene and prenylated benzoic acid from *Piper aduncum*. *Phytochemistry* **51**, 899-902 (1999).
42. Lago, J. H. G. *et al.* Prenylated benzoic acid derivatives from *Piper aduncum* L. and *P. hostmannianum* C. DC. (Piperaceae). *Phytochem. Lett.* **2**, 96-98 (2009).
43. Moreira, D. d. L., Guimarães, E. F., & Kaplan, M. A. C. A chromene from *Piper aduncum*. *Phytochemistry* **48**, 1075-1077 (1998).
44. Orjala, J., Erdelmeier, C. A. J., Wright, A. D., Rali, T., & Sticher, O. Two chromenes and a prenylated benzoic acid derivative from *Piper aduncum*. *Phytochemistry*, **34**, 813-818 (1993).
45. Batista, J. M. *et al.* Absolute configuration and selective trypanocidal activity of gaudichaudianic acid enantiomers. *J. Nat. Prod.* **74**, 1154-1160 (2011).
46. Banerji A., & Ray, R. Aurantiamides: a new class of modified dipeptides from *Piper aurantiacum. Phytochemistry* **20**, 2217-2220 (1981).
47. Maynard, L. D. *et al.* Secondary metabolites in a neotropical shrub: spatiotemporal allocation and role in fruit defense and dispersal. *Ecology* **101**, e03192 (2020).
48. Yoshida, N. C., *et al.* Alkenylphenols from *Piper dilatatum* and *P. diospyrifolium. Phytochem. Lett.* **25**, 136-140 (2018).
49. Jeffrey, C. S. *et al.* Antiherbivore prenylated benzoic acid derivatives from *Piper kelleyi*. *J. Nat. Prod.* **77**, 148-153 (2014).
50. Galinis, D. L., & Wiemer, D. F. Villiramulins A and B: new phenol derivatives from *Piper villiramulum. J. Org. Chem.* **58**, 7804-7807 (1993).
51. Pelter, A., & Hansel, R. Epoxipiperolid from *Piper sanctum*. *Z. Naturforsch B* **27**, 1186-1190 (1972).