RESEARCH ARTICLE

Short Title: Marx et al.—Sawtooth Community Phylogenetics

Increasing phylogenetic stochasticity at high elevations on summits across a remote North American wilderness

Hannah E. Marx[1,2,3,4], Melissa Richards[1], Grahm M. Johnson[1], and David C. Tank[1,2]

Manuscript received 25 October 2018; revision accepted 24 April 2019

[1] Department of Biological Sciences, University of Idaho, 875 Perimeter Dr. MS 3051, Moscow, Idaho 83844-3051, USA

[2] Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Dr. MS 3051, Moscow, Idaho 83844-3051, USA

[3] Current address: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048, USA

[4] Author for correspondence (e-mail: marxh@umich.edu); ORCID iD 0000-0002-3526-7878

Citation: Hannah E. Marx, Melissa Richards, Grahm M. Johnson, and David C. Tank. 2019. Increasing phylogenetic stochasticity at high elevations on summits across a remote North American wilderness. *American Journal of Botany* 106(7): XXXX.

DOI: XXXX

PREMISE: At the intersection of ecology and evolutionary biology, community phylogenetics can provide insights into overarching biodiversity patterns, particularly in remote and understudied ecosystems. To understand community assembly of the high alpine flora in the Sawtooth National Forest, USA, we analyzed phylogenetic structure within and between nine summit communities.

METHODS: We used high-throughput sequencing to supplement existing data and infer a nearly completely sampled community phylogeny of the alpine vascular flora. We calculated mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) to quantify

phylogenetic divergence within summits, and assessed whether maximum elevation explains phylogenetic structure. To evaluate similarities between summits, we quantified phylogenetic turnover, taking into consideration microhabitats (talus vs. meadows).

RESULTS: We found different patterns of community phylogenetic structure within the six most species-rich orders, but across all vascular plants phylogenetic structure was largely not different from random. There was a significant negative correlation between elevation and tree-wide phylogenetic diversity (MPD) within summits: overdispersion degraded as elevation increased. Between summits, we found high phylogenetic turnover driven by greater niche heterogeneity on summits with alpine meadows.

CONCLUSIONS: Our results provide further evidence that stochastic processes may also play an important role in the assembly of vascular plant communities in high alpine habitats at regional scales. However, order-specific patterns suggest that adaptations are still important for assembly of specific sectors of the plant tree of life. Further studies quantifying functional diversity will be important in disentangling the interplay of eco-evolutionary processes that likely shape broad community phylogenetic patterns in extreme environments.

KEY WORDS: alpine; community phylogenetics; elevation; high-throughput sequencing; Idaho; mean nearest taxon distance; mean pairwise distance; mega-phylogeny; vascular plants

In an ecological context, evolutionary history provides a useful tool for quantifying overall diversity (Pavoine and Bonsall, 2010; Winter et al., 2013; Jarzyna and Jetz, 2016) and a framework to address potential eco-evolutionary drivers of diversity patterns (Webb et al., 2002). On time-scaled phylogenies, branch lengths quantify the evolutionary time that separates species; more closely related species are expected to share ecologically relevant functional traits, assuming that such traits and niches are phylogenetically conserved (Webb et al., 2002; Cavender-Bares et al., 2009). Generally, this community phylogenetic approach is used to assess the importance of environmental filtering ("clustering" of closely related species within communities in a species pool) or competition defined by limiting similarity ("overdispersion" of distantly related species assemblages) for community assembly (Webb, 2000; Webb et al., 2002).

Alternatively, communities could be shaped by assembly processes that are, at least to some degree, species-neutral, such as colonization and local extinction (MacArthur and Wilson, 1967; Hubbell, 2001).

Importantly, many complex ecological and evolutionary processes influence community assembly (Vellend, 2010), requiring careful consideration of system-specific a priori hypotheses (Gerhold et al., 2015) and cautious interpretations of the resulting community phylogenetic patterns (Mayfield and Levine, 2010). The assumption of phylogenetic niche conservation (PNC) has been debated (for a review, see Munkemüller et al., 2015), and even with PNC, coexistence theory predicts that competition can produce clustering if differences in interspecific competitive hierarchy fitness dominate the assembly process (Mayfield and Levine, 2010; HilleRisLambers et al., 2012). Ideally, to interpret processes governing species' coexistence, additional information about the species' functional traits would be analyzed in conjunction with phylogenetic relationships (Cavender-Bares and Wilczek, 2003; Cavender-Bares et al., 2009; Cadotte et al., 2013), especially since different traits may have different levels of conservatism or convergence depending on the community (Cavender-Bares et al., 2006). Detailed, environmentally defined regional species surveys can be used to delineate environmental filtering in relation to dispersal limitation or competitive exclusion (Kraft et al., 2015), and explicitly test specific environmental, historical, biotic, and neutral hypotheses to explain coexistence (Gerhold et al., 2015). In remote and understudied ecosystems, such as the high alpine, the community phylogenetic approach can be particularly useful for providing insights into macroecological and evolutionary processes driving diversity (Marx et al., 2017).

With steep environmental gradients over increasing elevation, mountains provide ideal "natural experiments" for understanding general patterns of biodiversity (Körner, 2000; Graham et al., 2014) and adaptive evolution (Körner, 2007; Körner et al., 2011). Alpine regions are the only terrestrial biome with a global distribution (Körner, 2003), yet they represent some of the largest gaps in floristic knowledge (Kier et al., 2005). This is especially concerning because ranges of alpine plants are anticipated to shift with a changing climate (Körner, 2000; Dullinger et al., 2012; Pauli et al., 2012; Morueta-Holme et al., 2015), so documenting the present floristic diversity in alpine regions is a priority. Previous studies have therefore used the "phylogenetic-patterns-as-a-proxy" for ecological similarity framework (Gerhold et al., 2015) to test the hypothesis that physiologically harsh high alpine environments should filter for closely related

species sharing similar traits adapted to abiotic pressures, including low temperatures, extended periods of drought, and extreme ultraviolet radiation (Körner, 1995, 2011). Phylogenetic clustering has been identified within high-elevation sites (Li et al., 2014; Jin et al., 2015; Marx et al., 2017), and decreases in pairwise divergence have been positively correlated with temperature and precipitation (Li et al., 2014). However, recent community phylogenetic studies in high alpine habitats are challenging the ubiquity of abiotic constraints and environmental filtering for shaping communities. Random phylogenetic diversity has been found across taxonomic and spatial scales within summits (Le Bagousse-Pinguet et al., 2017; Marx et al., 2017), and neutral models of colonization and extinction have been able to explain community phylogenetic structure of dominant plant orders (Marx et al., 2017). One study also found a tendency toward overdispersion with increasing elevation, contrary to predictions of environmental filtering (Le Bagousse-Pinguet et al., 2017).

These contrasting patterns of community phylogenetic structure likely emerge from complex ecological and evolutionary processes that shape biodiversity in high alpine ecosystems (Graham et al., 2014), and we are far from having a general characterization of elevational diversity patterns across mountain ranges for plants (but for birds, see Quintero and Jetz, 2018). Mountain summits are often inhabited by globally rare or locally endemic lineages (Smith and Cleef, 1988; Kier et al., 2009), so inferring phylogenetic relationships among species is challenging because taxa are either not represented in supertrees or molecular sequence data are not readily available in repositories such as GenBank for mega-phylogenetic approaches. Targeted PCR enrichment (Cronn et al., 2012), combined with high-throughput sequencing technologies, provides a solution to retrieving genetic sequence data for entire community assemblages (reviewed in Godden et al., 2012; Grover et al., 2012). These methods are proving useful for resolving diversity patterns within specific lineages (e.g., Uribe-Convers et al., 2016) but are not yet being applied in macroecological contexts. Importantly, high-throughput sequencing technologies could potentially capture intraspecific variation between communities, which has been largely unexplored in previous studies of alpine community assembly using either supertree (Li et al., 2014) or mega-phylogenetic (Jin et al., 2015; Le Bagousse-Pinguet et al., 2017; Marx et al., 2017) approaches.

The present study seeks to fill a gap in our knowledge of high alpine community assembly by describing the phylogenetic structure of the flora across summits within the

Sawtooth National Forest (SNF), a remote North American wilderness located in south-central Idaho, USA (Fig. 1). We present the first detailed floristic survey of nine high alpine summits across three mountain ranges. From these collections, we used targeted high-throughput sequencing to supplement publicly available sequences mined from GenBank and compiled a detailed molecular dataset. We used the combined dataset to infer relationships among all species with a mega-phylogenetic approach (Smith et al., 2009; Roquet et al., 2012; Marx et al., 2016) and quantified community phylogenetic structure within and between alpine summits. To test the hypothesis that intense environmental conditions filter for closely related species that are physiologically able to survive in high alpine habitats in the SNF, we correlated patterns of community phylogenetic structure with maximum elevation on each summit. Many climatic and geologic factors constitute the local environment, but elevation (a.s.l.) has been used as a proxy for increasing environmental severity of temperature and precipitation in alpine habitats in general (Korner, 2007) and has been examined in previous studies of high alpine community phylogenetic structure (Machac et al., 2011; Jin et al., 2015). In central Idaho, corresponding gradients of temperature and precipitation over elevation have been shown to delimit ranges of certain endemic species (Steele et al., 1981; Ertter and Mosely, 1992).

If environmental filtering is structuring alpine communities, we expect a negative relationship between phylogenetic distance and maximum elevation: closely related species should occur together more often than by chance (low pairwise phylogenetic distance) at higher elevations (assuming trait conservation for physiologically relevant traits). On the other hand, if traits are convergent or if competition is strong, we expect increasing elevation to promote phylogenetic overdispersion (assuming niche conservation). Alternatively, diversity patterns might instead be explained by dispersal limitation in this island-like system (MacArthur and Wilson, 1967), in which case phylogenetic structure within summits is expected to be no different from random, while turnover between summits should be correlated with geographic proximity (Marx et al., 2017). To address how microhabitats impact community phylogenetic structure above tree line, we separated species collected in alpine meadows from those occurring only on talus slopes. Finally, taxonomic scale is known to impact community phylogenetic structure (Cavender-Bares et al., 2006; Graham et al. 2016), and distinct clades have experienced adaptive radiations into alpine ecosystems (reviewed in Hughes and Atchison, 2015). To assess

how clade-specific strategies may drive community diversity, we investigated patterns across all vascular plants as well as within the six most species-rich taxonomic orders separately.

<h1>MATERIALS AND METHODS

<h2>Study area and species collections

We sampled nine alpine summits in the Sawtooth, White Cloud, and Pioneer mountain ranges within the SNF (Fig. 1). The SNF is known for its immense mountainous terrain (Reid, 1963) and encompasses over 200,000 acres of federally designated wilderness area. Lying within the Rocky Mountain chain, this region was formed by the tectonic uplift of the Idaho and Sawtooth batholith (Kiilsgaard et al., 1970). Recent geologic episodes, including the Laramide orogeny in the late Mesozoic and extensive glaciations in the quaternary, resulted in the sharp topography and surface rock formation we currently observe (Borgert et al., 1999), giving the area its name (Kiilsgaard et al., 1970). The mountain ranges within the forest boundary include some of the most remote alpine biomes in the contiguous United States, and its alpine flora has been drastically understudied. Besides management-focused efforts (Schlatterer, 1972; Harper et al., 1978), no systematic surveys of this region have been conducted.

Our collections were focused on sampling alpine species, here defined as plants occurring in areas above tree line (Billings and Mooney, 1968; Körner, 2003), because they represent a major shift in climate (Richardson and Friedland, 2009). The collection approach followed other floristic studies based out of the Rocky Mountain Herbarium (e.g., Hartman, 1992; Lukas et al., 2012) and the Stillinger Herbarium (G. M. Johnson et al., unpublished data). Starting at the highest point on each summit, all aspects were traversed by spiraling down to tree line (as terrain allowed), and an individual of each species was collected to represent the diversity, which ranged from lycophytes through angiosperms and included herbaceous plants, shrubs, and small trees. We sampled in the months of July and August to capture peak phenology. Phenology in high alpine communities is regulated by environmental cues such as snow melt (Winkler et al., 2018), and for the most part is coordinated across species due to the short growing season. Specimens were pressed in the field, and leaf tissues were preserved in silica for molecular analyses. Imaging and processing of the collections were conducted at the University of Idaho Stillinger Herbarium (ID), where all voucher specimens were deposited (Appendix S1). Identifications were made using Hitchcock and Cronquist (1973), with nomenclature following the updated

taxonomy in the Consortium of Pacific Northwest Herbaria data portal (http://www.pnwherbaria.org; Consortium of Pacific Northwest Herbaria, 2013). The combined list of identified species that were collected constitutes the "alpine species pool" considered. Spatial Euclidean distances between summits were calculated from GPS coordinates.

<h2>Molecular sequence data

Total genomic DNA was extracted from silica-dried leaf tissue for all collections following a modified 2x-CTAB extraction protocol (Doyle and Doyle, 1987). Six gene regions with varying rates of molecular evolution that are frequently employed to resolve both recent and distant phylogenetic relationships (Soltis et al., 2011) were chosen for the present study and included representatives of the nuclear (ITS) and chloroplast (*atpB*, *matK*, *ndhF*, *rbcL*, and *trnTLF*) genomes. For all vascular plants that were collected on each alpine summit, we used targeted polymerase chain reaction (PCR) to amplify the six gene regions (PCR details are presented in Appendices S2 and S3). Following PCR, the resulting amplicons were pooled for high-throughput sequencing on an Illumina MiSeq platform using 300 bp paired-end reads and 1% of a sequencing lane.

Pooled reads from the Illumina MiSeq runs were demultiplexed using the dbcAmplicons pipeline, and consensus sequences were generated using the R script "reduce_amplicons.R" (https://github.com/msettles/dbcAmplicons) following the workflow detailed in Uribe-Covers et al. (2016). Briefly, for each sample, read-pairs were identified, sample-specific dual barcodes and target-specific primers were identified and removed (allowing the default matching error of four bases), and each read was annotated to include the species name and read number for the gene region. To eliminate fungal contamination that may have been amplified with ITS, and nonspecific amplification of poor PCR products for all gene regions, each read was screened against a user-defined reference file of annotated sequences retrieved from GenBank (using the "-screen" option in dbcAmplicons). Reads that mapped with default sensitivity settings were kept. Each read was reduced to the most frequent length variant, paired reads that overlapped by ≥10 bp (default) were merged into a single continuous sequence, and a consensus sequence without ambiguities was produced ("-p consensus" in "reduce_amplicons.R"). Paired reads that did not overlap were concatenated using Phyutility version 2.2.4 (Smith and Dunn, 2008), and any merged segments were added to the concatenated reads.

Processed MiSeq reads for each gene region were aligned using MAFFT version 7.273 (Katoh and Standley, 2013) with default settings, and segments that were divided for PCR amplification were aligned separately. All alignments were loaded into Geneious version 7.1.9 (http://www.geneious.com; Kearse et al., 2012), where visual inspection in addition to a batch blast to the NCBI nucleotide database helped identify incorrect sequences that escaped our primary screening (e.g., resulting from fungal contamination, nonspecific amplification, or contaminated samples). Incorrect sequences (those whose BLAST hit did not match with the species and/or gene region identification) were discarded and the segment was realigned. Each gene segment was then concatenated using Phyutility, resulting in a final alignment for each gene region.

In some cases, gene regions were not amplified for every species collected using the targeted high-throughput sequencing approach described above. Therefore, we used the PHLAWD pipeline (Smith et al., 2009) to retrieve published sequences from GenBank and supplement our newly produced sequence data. The PHLAWD pipeline incorporates GenBank taxonomy to sequentially profile-align increasingly higher taxonomic groups together with MAFFT, and outputs a single alignment file for each query. Using the combined list of species that were identified across all summits, we searched for the same six gene regions that were amplified with PCR. Infraspecific taxa were collapsed to the species level to avoid pseudoreplication, and if there was more than one sequence for a species available in GenBank the longest was kept.

Alignments from the high-throughput sequencing and GenBank output from PHLAWD were combined for each gene region, gaps were removed, sequences were realigned using MAFFT, and alignments were cleaned using Phyutility to remove sites that were missing ≥50% of data. To initially assess taxonomic concordance, gene trees were estimated for each region under maximum likelihood (ML) criterion using the GTR-CAT model of nucleotide substitution and 1000 bootstrap replicates in RAxML version 8.2.8 (Stamatakis, 2006). After visual inspection of gene trees, taxonomic conflicts with phylogenetic expectations following the Angiosperm Phylogeny Group IV classification (Angiosperm Phylogeny Group, 2016) were removed from each gene region, the alignments and cleaning described above were repeated, and the longest sequence from either data source (GenBank or our sequencing results) was retained

to represent each species. All gene regions were then concatenated into a final "total dataset" alignment using Phyutility.

<h2>Community phylogenetic inference

The total dataset's sequence alignment was used to infer a ML estimate for the species-level community phylogeny of all vascular alpine plants collected in the SNF in RAxML with a GTR-CAT model partitioned by gene region and using the auto MRE bootstrap convergence option to determine the number of bootstrap replicates for stable support values (Pattengale et al., 2009). All analyses were run on the CIPRES cyberinfrastructure for phylogenetic research (Miller et al., 2010; last accessed July 29, 2017). Following Marx et al. (2016), we used the "congruification" approach (Eastman et al., 2013) in the R package "geiger" version 2.0 (Pennell et al., 2014) to place node calibrations from the detailed time-tree estimate of Zanne et al. (2014) on congruent nodes in the SNF community phylogeny. Penalized likelihood was then used to scale molecular branch lengths to time as implemented in "treePL" version 1.0 (Smith and O'Meara, 2012).

<h2>Community phylogenetic structure

Evolutionary relationships estimated from the SNF community phylogeny were used to summarize phylogenetic patterns within (α-diversity) the nine alpine summit communities sampled for all vascular plants (Tracheophyta), as well as the six most species-rich orders. We calculated the mean nearest taxon distance (MNTD) and the mean pairwise distance (MPD; Webb, 2000) to quantify divergence at fine and broad phylogenetic scales, respectively (Mazel et al., 2016; Tucker et al., 2016). We assessed whether observed phylogenetic patterns were different from a random expectation by randomly resampling each community phylogeny 10,000 times (random draw null model) and calculating the standardized effect sizes (SES) for each metric in the R package "picante" version 1.6-2 (Kembel et al., 2010). Significance of SES was assessed from ranks of each observed metric compared with the null model distribution using two-tailed *P*-values ($\alpha = 0.05$). Positive SES values indicate greater-than-expected observed phylogenetic divergence from the species pool of the alpine flora within the SNF (phylogenetic overdispersion), while negative values indicate that observed divergence is less than expected (phylogenetic clustering).

Changes in community phylogenetic structure between summits (β-diversity) were summarized with two different metrics. First, we calculated the unique branch-length contribution in relation to the total branch lengths shared between each community with the UniFrac index (Lozupone and Knight, 2005), which has been used to quantify turnover in other studies of alpine phylogenetic structure (Jin et al., 2015). This broad measure of phylogenetic divergence between sites (Baselga, 2009) does not discern between richness gradients of species-poor communities nested within species-rich communities (Wright and Reeves, 1992) or spatial turnover, whereby environmental filtering or historical processes cause distinct lineages to replace others between sites (Qian et al., 2005). Following Leprieur et al. (2012), we decomposed UniFrac to separate the phylogenetic divergence between summits attributed to accumulation of species richness (UniFrac PD) from divergences that represent true gain or loss of species due to replacement (UniFrac Turn). SES of UniFrac indices were quantified by comparing observed values to a null distribution of indices from tips shuffled across the community phylogeny (n = 999) using the R code provided in Leprieur et al. (2012). Second, we identified nodes within the community phylogeny where species or clades of species were contributing to turnover with the metric Π_{ST} (Hardy and Senterre, 2007), which measures changes in mean phylogenetic distances between sites compared to within sites. We used a randomization that shuffles species across the community phylogeny ("1s"; n = 999) to test the significance of the phylogenetic structure (Hardy, 2008) with the R package "spacodiR" version 0.13.0115 (Eastman et al., 2011). $\Pi_{ST} > 0$ indicates spatial phylogenetic clustering (i.e., species within plots are more closely related than between plots), while $\Pi_{ST} < 0$ indicates spatial phylogenetic overdispersion (i.e., species within plots are less closely related than between plots; Hardy and Senterre, 2007).

<h2>Environmental drivers of diversity patterns

To test for environmental filtering of closely related species, we assessed the relationship between patterns of phylogenetic divergence and elevation. Within summits, we used simple linear regression with SES MNTD and SES MPD as the dependent variables, and maximum elevation as the independent variable. In addition, we used a linear mixed model with elevation as a fixed factor and mountain range as a random factor. Regression assumptions were verified with diagnostic plots (residual vs. fitted values, Q-Q plots). For turnover between sites, elevation

and geographic coordinates were expressed as pairwise Euclidean distances between sites using the "vegdist" function R package "vegan" version 2.4-3 (Oksanen et al., 2017). Each compositional pairwise β-diversity matrix was correlated with maximum elevation and spatial distance using multiple regression on matrices (MRM; nperm = 999; Lichstein, 2007) as implemented in the R package "ecodist" version 1.2.9 (Goslee and Urban, 2007). On four summits (Thompson Peak, D.O. Lee Peak, Salzburger Spitzl, and Hyndman Peak) we encountered one additional habitat type besides talus slopes: high alpine meadows. To assess whether niche heterogeneity was driving patterns of phylogenetic divergence on these summits, we removed species collected in meadows and compared SES metrics for species collected only on talus slopes ("Talus") with all species collected above tree line together ("All Alpine"). Inspection of Q-Q plots indicated normal distribution of SES MNTD and SES MPD for each category and sample variances were homogeneous, so we used a paired *t*-test with a 95% confidence interval to test the difference in means. Statistical analyses were conducted in R version 3.2.3 (R Core Team, 2015).

<h1>RESULTS

<h2>Species collections

A total of 476 specimens (155 unique species) were collected, and between 28 (Braxon Peak) and 78 (Hyndmann Peak) species were sampled on each summit. Four summits (Thompson Peak, D.O. Lee Peak, Salzburger Spitzl, and Hyndman Peak) had alpine meadows. The six plant orders with the greatest species richness were the Asterales ($N = 37$), Poales ($N = 19$), Caryophyllales ($N = 19$), Lamiales ($N = 12$), Brassicales ($N = 9$), and Ericales ($N = 8$) (Fig. 2). Vouchers and images can be viewed online at the Consortium of the Pacific Northwest Herbaria data portal (http://www.pnwherbaria.org; Consortium of Pacific Northwest Herbaria, 2013). Voucher information for each collection and the community matrix showing the presence of species across summits are provided in Appendices S1 and S4.

<h2>Molecular sequence data and community phylogeny

Because we used universal primers that were designed primarily for angiosperms and/or seed plant systematics to generate molecular sequence data, there was taxonomic variation (and biases) in the efficacy of amplification for each gene region, and none of the ferns or lycophytes

amplified (Appendix S4). Amplification of certain gene regions (and segments) was more successful in certain clades than in others. Rates of amplification in graminoids were particularly low, especially for *matK*. The segments *ndhF2* and *ndhF3* worked better for graminoids than *ndhF1*, and *trn_cf* worked better than *trn_ab* for graminoids and gymnosperms. The *atpB* primers amplified well for graminoids and gymnosperms (especially *atpB1*). ITS amplified well across a broad range of taxonomic lineages, but there was significant nonspecific amplification or fungal contamination, which had to be removed prior to compiling this dataset. The *rbcL* primers amplified well overall across all taxonomic groups, in one entire segment, and had few reads with nonspecific amplification to remove (for summary statistics from Illumina read processing, including the number of raw reads and the number of reads remaining after screening and reduction, see Appendix S5).

After MiSeq reads were processed, screened, and reduced, the high-throughput approach generated novel sequence data for 419 individuals (88% of samples collected on different summits) and 145 unique species (94% of all species collected). By supplementing missing gene regions with available data from GenBank, the total dataset included 152 species (98% of species collected). Ambiguous species (i.e., those identified only to genus) were excluded from the analyses of phylogenetic structure, resulting in a community phylogenetic dataset representing 149 taxa (for MiSeq and GenBank accessions for each gene region that was used in the total dataset for each species, see Appendix S4). The cleaned and concatenated total sequence alignment had 3193 bp and 37.11% gaps (Appendix S5). The ML estimate of alpine community phylogenetic relationships across the total dataset was consistent with the Angiosperm Phylogeny Group IV classification (Angiosperm Phylogeny Group 2016; Fig. 2), and the majority of deep and shallow nodes showed bootstrap support >75 (Appendix S6).

<h2>Community phylogenetic structure and statistical analyses

Within summits, observed MNTD was no different than expected from a random assemblage of alpine flora across vascular plants (Fig. 3, top panel). However, there was statistically significant overdispersion of MPD on Horstmann Peak and clustering on D.O. Lee Peak (Fig. 3, bottom panel). Within specific clades, phylogenetic structure was also largely not different from random. The few summits that did have statistically significant order-level phylogenetic structure were mostly clustered, except for Poales on Thompson Peak, which was significantly overdispersed.

Summits with alpine meadows did not have a higher (or lower) phylogenetic divergence than those without (Appendix S7). With increasing maximum elevation, there was a slight but nonsignificant increase in phylogenetic distance between closely related species (MNTD; Fig. 4A) and a significant decrease in pairwise phylogenetic divergence across Tracheophyta (MPD; Fig. 4B) (adjusted $R^2 = 0.3668$, $P = 0.04933$; complete results for phylogenetic α-diversity are included in Appendix S8).

The decomposed UniFrac index revealed higher-than-expected true turnover of distinct plant lineages between four of the 36 pairwise summit comparisons, and none were significantly lower than expected (Fig. 5A, above diagonal). When species collected in high alpine meadows were removed (from Thompson Peak, D.O. Lee Peak, Salzburger Spitzl, and Hyndman Peak), turnover between summits was no different from random overall (Fig. 5A, below diagonal). However, neither maximum elevation ($R^2 = 0.0147$, $P = 0.4497$) nor spatial distance ($R^2 = 0.0228$, $P = 0.4190$) explained phylogenetic β-diversity. The only clades with species less widespread among summits than expected by chance (higher than expected Π_{ST}) were the order Lamiales and the Rosid clade (Fig. 5B). Otherwise, turnover among lineages was random. A table summarizing β-diversity results can be found in Appendix S9.

<h1>DISCUSSION

High alpine ecosystems across the remote Sawtooth National Forest comprise a diverse array of vascular plants (Fig. 2), dominated by species in the orders Asterales, Poales, Caryophyllales, Lamiales, Brassicales, and Ericales. Significant patterns in community phylogenetic structure across vascular plants were only found on two summits: tree-wide overdispersion on Horstmann Peak and clustering on D.O. Lee Peak (MPD; Fig. 3). Otherwise, tip-wise phylogenetic structure was not different from random across all vascular plants (MNTD; Fig. 3). The influence of spatial and taxonomic scale on patterns of community phylogenetic structure has been well documented in the literature (Emerson and Gillespie, 2008; Vamosi et al., 2009; Park et al., 2018), and when source pools are defined more broadly, communities tend to be more phylogenetically clustered than expected (Cavender-Bares et al., 2006). This pattern was confirmed here, as significant phylogenetic clustering was mostly found within specific clades on a few summits when the source pool for each summit community was reduced from the entire

Tracheophyta to orders (Fig. 3). Still overall, significant order-specific phylogenetic structure was idiosyncratic and sparse, suggesting clade-specific community assembly mechanisms.

To test the hypothesis that extreme environments filter for closely related species in high alpine communities, we investigated the relationship between these phylogenetic patterns and maximum elevation. We predicted that the physiologically extreme environment is filtering for closely related species driving observed diversity patterns (Billings and Mooney, 1968), as has been found in previous studies of high alpine community phylogenetic structure (Li et al., 2014; Jin et al., 2015). In the Hengduan Mountains Region of China, Li et al. (2014) investigated the community phylogenetic structure of alpine flora across 27 elevation belts ranging from 3000 to 5700 m. Within sites, they found phylogenetic overdispersion at lower elevations and phylogenetic clustering at higher elevations (more significant overdispersion with higher temperatures and precipitation; Li et al., 2014). However, at the highest elevations (>5500 m) phylogenetic structure became random (Li et al., 2014), possibly indicating relaxed environmental filtering between tree line and summit belts. In the Rocky Mountain National Park, Colorado, USA, Jin et al. (2015) assessed phylogenetic turnover between 569 plots ranging in elevation from 2195 to 3872 m, and found that plant species were more closely related than expected (high phylogenetic clustering) overall within plots, and had a higher-than-expected turnover within than among plant clades between plots (Jin et al., 2015). Abiotic environment defined by the elevation of individual plots explained turnover across alpine communities more than spatial distance between sites, implying a regional environmental filter and niche conservatism within clades, which was particularly strong for communities sampled east of the Continental Divide. Our results did show that elevation was significantly correlated with MPD across vascular plants (Fig. 4). While the summits at lower elevations consisted of plant assemblages that were more distantly related than expected, phylogenetic structure of summits at higher elevations was not significantly different than a random sample of the alpine species pool (Fig. 4B). This significant negative relationship between MPD and elevation suggests that the environment may be shaping community-wide assembly in high alpine areas of the SNF, but not toward significant clustering of close relatives with shared derived traits adapted to extreme alpine conditions, as expected. Instead, phylogenetic structure shifted from significantly overdispersed on summits with lower maximum elevation to each species having an equal

probability of co-occurring at higher maximum elevations (Fig. 3). While not significant, this trend also held for the order Poales.

On the other hand, a positive trend between maximum elevation and MNTD was found overall (across vascular plants and within most orders) but was not significant (Fig. 4A). As elevation increased, tip-wise phylogenetic distances also increased. If traits are conserved, overdispersion of distantly related species is sometimes interpreted to result from competition in the community phylogenetic framework (Webb, 2000). However, if ecologically relevant traits are convergent, habitat filtering is instead expected to produce phylogenetic overdispersion. Significant phylogenetic overdispersion was also found within the order Poales on Thompson Peak for MPD (Fig. 3), and MPD was positively related to maximum elevation for the Caryophyllales and Ericales, though not significantly (Fig. 4B). Globally these orders, and the Caryophyllales in particular, are known to contain many species with a cushion life form, suggesting frequent evolutionary convergence toward this trait (Boucher et al., 2016), which could explain the overdispersion found here. Niche or habitat heterogeneity could also allow distantly related species to fill space (Stein et al., 2014), promoting phylogenetic overdispersion. Despite higher species richness on summits with alpine meadows, habitat heterogeneity did not drive patterns of community phylogenetic structure within summits (Appendix S7). Spatially, high turnover of phylogenetic diversity does appear to be attributable to plants found in high alpine meadows (Fig. 5A)—at least between Snowyside Peak and the three summits D.O. Lee Peak, Hyndman Peak, and Thompson, and between Thompson and Hyndman. Species in the order Lamiales and the Rosid clade were found to be more closely related within than between summits (spatial clustering; Fig. 5B), which might indicate specific niche preferences in these lineages. Finally, facilitation has been shown to increase phylogenetic diversity of plant communities generally (Valiente-Banuet and Verdú, 2007) and within alpine communities in particular (Choler et al., 2001). Taken together, patterns of phylogenetic structure within and between summit communities suggest that functional trait convergence and niche differentiation promote the co-occurrence of distant lineages at high elevations in the SNF, but further work detailing traits and environmental conditions will be necessary to support this.

Besides adaptation, species-neutral processes are expected to shape biodiversity in island-like systems, such as high alpine summits (MacArthur and Wilson, 1967; reviewed in Marx et al., 2017). A recent study simulated communities under different assembly processes, and the

results revealed how overdispersion can also be caused by stochastic processes, such as local extinction or limited dispersal following allopatric speciation (Pigot and Etienne, 2015). In fact, given that allopatric speciation has been shown to drive communities toward overdispersion, clustering should be difficult to detect at all under the random-draw null model (implemented here), unless (1) rates of extinction are high enough to decouple from allopatric speciation, (2) the source pool itself was completely formed by colonization, or (3) the source pool was poorly sampled (Pigot and Etienne, 2015). The lack of significant phylogenetic clustering across vascular plants coupled with largely random phylogenetic structure at the highest maximum elevations (Fig. 3) could signify the importance of such stochastic assembly processes in central Idaho.

Farther south, in the Rocky Mountains of Colorado, phylogenetic clustering of closely related species within alpine summits was found (for MPD), and environment explained high turnover within clades (Jin et al., 2015). While this is part of the same greater mountain range, it is possible that scale effects explain the differences in phylogenetic patterns (Cavender-Bares et al., 2006). In the present study, we defined communities at the summit level (everything occurring above tree line), while in Colorado, communities were defined at the plot level (~400 m^2 in area). At a similar spatial scale (summit level) in the Écrins National Park, France, however, few high alpine summit communities were significantly clustered (for either MNTD or MPD), and phylogenetic structure was not explained by a series of environmental variables that were tested (including elevation; Marx et al., 2017). Instead, models explicitly accounting for species-neutral assembly processes such as colonization and local extinction were able to explain phylogenetic patterns, providing further indication that these processes play an important role in shaping diversity at this regional scale. But clade-specific patterns differ between Idaho and France: phylogenetic patterns within the Poales mirrored the negative relationship found for MPD across vascular plants in the SNF (Fig. 4B), while environmental conditions were mostly found to drive clustering within the Caryophyllales in the French Alps. The architecture of these alpine ranges is incredibly complex (Körner at al., 2011; Elsen and Tingley, 2015), and factors such as the age of mountain orogeny, bioclimatic belts, or the extent of dynamic glacial histories should be considered in greater detail to compare cross-continental community phylogenetic relationships and more rigorously test hypotheses explaining how processes like historic

biogeography shape the evolution and ecology of alpine biodiversity globally (Graham et al., 2014).

Our results and those of other studies (e.g., Li et al., 2014; Jin et al., 2015; Le Bagousse-Pingueta et al., 2017; Marx et al., 2017) illustrate the potential for patterns of phylogenetic diversity to elucidate dominant processes driving species co-occurrence in extreme regions; however, many assumptions about functional trait evolution and community assembly processes have the potential to be violated when evolutionary relationships are used as a proxy for ecological niche similarity (reviewed in Gerhold et al., 2015). Rather than viewing phylogenetic patterns as a proxy for ecological similarity and accepting the myriad of underlying assumptions, community phylogenetic diversity has strong potential to inform how macroevolutionary processes shape the diversity of multispecies assemblages we observe across space (Gerhold et al., 2015). Because alpine ecosystems are found on every continent, patterns of phylogenetic community structure can be compared globally to assess how rates of diversification constrain (or promote) alpine diversity. Still, a central challenge for moving toward investigating macroevolutionary drivers of community phylogenetic patterns (the "phylogenetic-patterns-as-a-cause" approach; Gerhold et al., 2015) is that more studies across lineage-pools are necessary to compare across alpine regions.

We have demonstrated here how combining novel and available molecular sequence data efficiently resolved phylogenetic community structure across remote summits. We leveraged a targeted high-throughput sequencing approach recently developed for plant systematics (Cronn et al., 2012; Godden et al., 2012; Grover et al., 2012; Uribe-Convers et al., 2016) to directly sample community-wide genetic diversity. These novel molecular sequences captured phylogenetic relationships for 88% of alpine plant individuals collected across the nine summits. However, the taxonomic specificity of the primer pairs used for amplification was biased toward seed-producing vascular plants (i.e., excluded ferns and lycophytes). Including primers that are optimized for these groups would be more effective for documenting the complete flora. We were able to supplement taxonomic gaps in the high-throughput dataset with publicly available molecular sequence data from GenBank, resulting in a nearly complete (98%) species-level phylogenetic representation of the alpine flora.

This total data approach could be tractable for other high-elevation ecosystems facing a similar deficit in molecular sequence data for community phylogeny inference. The ability to

effectively sequence multiple gene regions from hundreds of plant species at a time also presents an opportunity to capture intraspecific genetic variation of multispecies assemblages across regions, which is not possible when a single sample is used to represent species diversity. Investigating signatures of selection at the population level could provide deeper insights into the mechanistic basis underlying patterns of community phylogenetic structure, such as the evolution of key traits or life forms that are important for survival at these extremes (e.g., Boucher et al., 2016). Additionally, the targeted high-throughput sequencing approach presented here is extendable across taxonomic lineages, presenting exciting avenues for community phylogenetic networks of plants with associated pollinators (Pellissier et al., 2013) or microbes (Bryant et al., 2008). Furthermore, the power to detect environmental filtering increases with the size of the species pool (Kraft et al., 2007). Supplementing available sequence data with high-throughput sequencing to sample larger source pools could be used to more explicitly test stochastic models of species-neutral colonization and local extinction (Pigot and Etienne, 2015) and the relative importance of adaptive and species-neutral processes for generating and maintaining biodiversity in high alpine habitats.

<h1>CONCLUSIONS

Mountains are ideal for testing how ecological and evolutionary mechanisms shape diversity patterns we observe across space (Graham et al., 2014), but the extreme environmental conditions that define high alpine areas also pose a challenge to research efforts, so comparisons among regions remain limited. Collections from the first detailed floristic survey of nine summits across the SNF in central Idaho contribute to our global synthesis of montane biodiversity, and community phylogenetic relationships from combined novel and publicly available molecular sequence data show patterns of increasing phylogenetic stochasticity over an elevation gradient. While we interpret these results as an indication that the environment may not be a broad selective force across the vascular plant community as a whole at high elevations, we recognize that elevation gradients comprise complex geographic effects (Körner, 2007), and regional distinctions of specific climatic and topological properties will be important for global comparisons in the future. Clade-specific signatures of phylogenetic clustering indicate that environmental filtering may be more important for certain branches of the tree of life than others, and trends toward phylogenetic overdispersion over increasing elevation suggest that traits

important for functioning in high alpine habitats may have converged in different lineages. Aggregating functional and phylogenetic distances (Cadotte et al., 2013) will be useful in future studies to assess convergence (Cavender-Bares et al., 2006) and differentiate between the complex drivers of diversity across taxonomic levels.

<h1>AUTHOR CONTRIBUTIONS

H.E.M. and D.C.T. conceptualized this research. H.E.M. organized collecting expeditions to the Sawtooth National Forest. M.R. conducted laboratory work (total DNA extractions and PCR). H.E.M. performed analyses and wrote the manuscript, with input from G.M.J. and D.C.T.

<h1>ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation through Graduate Research Fellowship grant no. DGE-1144254 and grants no. EF-1550838 and FESD 1338694. Collections were made possible by support from an expedition fund of the University of Idaho Stillinger Trust, and molecular work was supported by a Graduate Student Research Award from the Society of Systematic Biologists and by a Rosemary Grant Research Award from the Society for the Study of Evolution. The Stillinger Herbarium processed and stores this collection. The Sawtooth National Forest is thanked for granting collection permits. Comments from the editor and two anonymous reviewers greatly improved the manuscript. This floristic survey would not have been achievable without the aid of the Sawtooth Mountain Guides (including Drew Daly, Ryan Jung, Matt Scrivner, and Chris Lundy), the SNF Ranger Station in Stanley, Idaho, and field assistants Basya R. Clevenger, Chloe M. Stenkamp-Strahm, and Shauna K. Thomas. I also thank Bill Rember for his geological perspective and for welcome breaks reminiscing about the Sawtooths and Stanley, Idaho—a reminder of the beautiful diversity of this place. This work is dedicated to the memory of Samuel W. Marx, whose encouragement inspired this expedition.

<h1>DATA ACCESSIBILITY

A detailed workflow—including read processing scripts, known sequences used for MiSeq read screening, reference sequence files for PHLAWD searches, and R scripts—is available on GitHub (https://github.com/hmarx/Sawtooth-Alpine-PD). Raw MiSeq reads are deposited in the Sequence Read Archive (BioProject: PRJNA530061), and assembled gene regions are available

on GenBank (for accession numbers corresponding to each sample, see Appendix S10): ITS (MK802340–MK802514), *atpB* (MK800171–MK800415), *ndhF* (MK800615–MK800980), *matK* (MK800416–MK800614), *rbcL* (MK749304–MK749388), and *trnTLF* (MK800981–MK801096). The alpine community matrix, gene region sequence alignments before cleaning, concatenated sequence alignment post-cleaning, and the community phylogeny are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.d06j48t (Marx et al., 2019).

<h1>SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Appendix S1. Voucher numbers and collection information for each sample.

Appendix S2. Supplemental Methods.

Appendix S3. Table with target-specific primer pair sequences and citations.

Appendix S4. Accession number for the high-throughput (miseq) or GenBank (ncbi) sequence used for each gene region and community matrix.

Appendix S5. Table summarizing high-throughput sequencing read statistics and aligned sequence length for each gene region.

Appendix S6. Maximum likelihood phylograms estimated from concatenated gene regions.

Appendix S7. Comparison of phylogenetic community structure between microhabitats within each summit.

Appendix S8. Table summarizing phylogenetic α-diversity.

Appendix S9. Table summarizing phylogenetic β-diversity.

Appendix S10. GenBank accession numbers for each gene region.

<h1>LITERATURE CITED

Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Baselga, A. 2009. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* 19: 134–143.

Billings, W. D., and H. A. Mooney. 1968. The Ecology of Arctic and Alpine Plants. *Biological Reviews* 43: 481–529.

Borgert, J. A., K. A. Lundeen, and G. D. Thackray. 1999. Glacial Geology of the Southeastern Sawtooth Mountains. In: Hughes SS, Thackray GD, eds. *Guidebook to the Geology of Eastern Idaho. Idaho Museum of Natural History*. Pocatello, USA: Idaho Museum of Natural History, 205–217.

Bryant, J. A., C. Lamanna, H. Morlon, A. J. Kerkhoff, B. J. Enquist, and J. L. Green. 2008. Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences, USA* 105: 11505–11511.

Boucher, F. C., S. Lavergne, M. Basile, P. Choler, and S. Aubert. 2016. Evolution and biogeography of the cushion life form in angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 20: 22–31.

Cadotte, M., C. H. Albert, and S. C. Walker. 2013. The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecology Letters* 16: 1234–1244.

Cavender-Bares, J., A. Keen, and B. Miles. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* 87: S109–S122.

Cavender-Bares, J., K. H. Kozak, P. V. A. Fine, and S. W. Kembel. 2009. The merging of community ecology and phylogenetic biology. *Ecology Letters* 12: 693–715.

Cavender-Bares, J., and A. Wilczek. 2003. Integrating micro- and macroevolutionary processes in community ecology. *Ecology* 84: 592–597.

Choler, P., Michalet, R., and R. M. Callaway. (2001). Facilitation and competition on gradients in alpine plant communities. *Ecology* 82: 3295–3308.

Consortium of Pacific Northwest Herbaria. 2007-2018 (continuously updated) Website http://www.pnwherbaria.org/ [accessed 05 March 2013].

Cronn, R., B. J. Knaus, A. Liston, P. J. Maughan, M. Parks, J. V. Syring, and J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99: 291–311.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Dullinger, S., A. Gattringer, W. Thuiller, D. Moser, N. E. Zimmermann, A. Guisan, W. Willner, et al. 2012. Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change* 2: 619–622.

Eastman, J. M., L. J. Harmon, and D. C. Tank. 2013. Congruification: support for time scaling large phylogenetic trees. *Methods in Ecology and Evolution* 4: 688–691.

Eastman, J. M., C. E. T. Paine, and O. J. Hardy. 2011. spacodiR: structuring of phylogenetic diversity in ecological communities. *Bioinformatics* 27: 2437–2438.

Elsen, P. R., and M. W. Tingley. 2015. Global mountain topography and the fate of montane species under climate change. *Nature Climate Change* 5: 772–776.

Emerson, B. C., and R. G. Gillespie. 2008. Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution* 23: 619–630.

Ertter, B., and B. Moseley. 1992. Floristic regions of Idaho. *Journal of the Idaho Academy of Science* 28.2: 57-70.

Gerhold, P., J. F. Cahill Jr., M. Winter, I. V. Bartish, and A. Prinzing. 2015. Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology* 29, 600–614.

Godden, G. T., I. E. Jordon-Thaden, S. Chamala, A. A. Crowl, N. García, C. C. Germain-Aubrey, et al. 2012. Making next-generation sequencing work for you: approaches and practical considerations for marker development and phylogenetics. *Plant Ecology and Diversity* 5: 427–450.

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

Graham, C. H., A. C. Carnaval, C. D. Cadena, K. R. Zamudio, T. E. Roberts, J. L. Parra, C. M. McCain, et al. 2014. The origin and maintenance of montane diversity: Integrating evolutionary and ecological processes. *Ecography* 37: 711-719.

Graham, C. H., D. Storch, and A. Machac. 2018. Phylogenetic scale in ecology and evolution. *Global Ecolology and Biogeography* 27: 175–187.

Grover, C. E., A. Salmon, and J. F. Wendel. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *American Journal of Botany* 99: 312-319.

Hardy, O. J. 2008. Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology* 96: 914–926.

Hardy, O. J., and B. Senterre. 2007. Characterizing the Phylogenetic Structure of Communities by an Additive Partitioning of Phylogenetic Diversity. *Journal of Ecology* 95: 493–506.

Harper, K. T., C. D. Freeman, K. W. Ostler, and L. G. Klikoff. 1978. *The flora of the great basin mountain ranges: Diversity, sources, and dispersal ecology.* Provo, USA: Intermountain biogeography: a symposium.

Hartman, R.L. 1992. The Rocky Mountain Herbarium, associated floristic inventory, and the flora of the Rocky Mountains project. *Journal of the Idaho Academy of Science* 28: 22–43.

HilleRisLambers, J., P. B. Adler, W. S. Harpole, J. M. Levine, and M. M. Mayfield. 2012. Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics* 43: 227–248.

Hitchcock, C. L., and A. Cronquist. 1973. *Flora of the Pacific Northwest: An Illustrated Manual*. Seattle, WA, USA: University of Washington Press.

Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography.* Princeton, NJ, USA: Princeton University Press.

Hughes, C. E., and G. W. Atchison. 2015. The ubiquity of alpine plant radiations: from the Andes to the Hengduan Mountains. *New Phytologist* 207: 275–282.

Jarzyna, M. A, and W. Jetz. 2016. Detecting the Multiple Facets of Biodiversity. *Trends in Ecology & Evolution* 31: 527–538.

Jin, L. S., M. W. Cadotte, and M.-J. Fortin. 2015. Phylogenetic turnover patterns consistent with nicheconservatism in montane plant species. *Journal of Ecology* 103: 742–749.

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

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647–1649.

Kembel, S. W., P.D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Kier, G., H. Kreft, T. M. Lee, W. Jetz, P. L. Ibisch, C. Nowicki, J. Mutke , et al. 2009. A Global Assessment of Endemism and Species Richness across Island and Mainland Regions. *Proceedings of the National Academy of Sciences, USA* 106: 9322–9327.

Kier, G., J. Mutke, E. Dinerstein, T. H. Ricketts, W. Küper, H. Kreft, and W. Barthlott. 2005. Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography* 32: 1107–1116.

Kiilsgaard, T. H., V. L. Freeman, and J. S. Coffman. 1970. Mineral resources of the Sawtooth Primitive Area, Idaho. In: *Studies related to wilderness–primitive areas; Geological Survey Bulletin: 1319-D.* Washington, USA: United States Government Printing Office.

Körner, C. 1995. Alpine plant diversity: a global survey and functional interpretations. *Ecological Studies* 113: 45–62.

Körner, C. 2000. Why are there global gradients in species richness? Mountains might hold the answer. *Trends in Ecology & Evolution* 15: 513–514.

Körner, C. 2003. *Alpine Plant Life: Functional Plant Ecology of High Mountain Ecosystems.* Springer, Berlin, Germany & New York.

Körner, C. 2007. The use of "altitude" in ecological research. *Trends in Ecology & Evolution* 22: 569–574.

Körner, C. 2011. Coldest places on earth with angiosperm plant life. *Alpine Botany* 121: 11–22.

Körner, C., J. Paulsen, and E. M. Spehn. 2011. A definition of mountains and their bioclimatic belts for global comparisons of biodiversity data. *Alpine Botany* 121: 73–78.

Kraft, N. J. B., P. B. Adler, O. Godoy, E. C. James, S. Fuller and J. M. Levine. 2015. Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology* 29: 592–599.

Kraft, N. J. B., W. K. Cornwell, C. O. Webb, and D. D. Ackerly. 2007. Trait Evolution, Community Assembly, and the Phylogenetic Structure of Ecological Communities. *The American Naturalist* 170: 271–283.

Le Bagousse-Pinguet, Y., P. Liancourt, L. Götzenberger, F. de Bello, J. Altman, V. Brozova, Z. Chlumska, et al. 2018. A multi-scale approach reveals random phylogenetic patterns at the edge of vascular plant life. *Perspectives in Plant Ecology, Evolution and Systematics* 30: 22–30.

Leprieur, F., C. Albouy, J. De Bortoli, P. F. Cowman, D. R. Bellwood, and D. Mouillot. 2012. Quantifying Phylogenetic Beta Diversity: Distinguishing between "True" Turnover of Lineages and Phylogenetic Diversity Gradients. *PLoS ONE* 7: e42760.

Li, X. H., X. X. Zhu, Y. Niu, and H. Sun. 2013. Phylogenetic clustering and overdispersion for alpine plants along elevational gradient in the Hengduan Mountains Region, southwest China. *Journal of Systematics and Evolution* 52: 280–288.

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

Lomolino, M. V. 2001. Elevation gradients of species-density: Historical and prospective views. *Global Ecology and Biogeography* 10: 3–13.

Lozupone, C., and R. Knight. 2005. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology* 71: 8228–8235.

Lukas, L.E., B. E. Nelson, and R.L. Hartman. 2012. A floristic inventory of vascular plants of the Medicine Bow National Forest and vicinity, southeastern Wyoming, USA. *Journal of the Botanical Research Institute of Texas* 6: 759-787.

MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, USA.

Machac, A., Janda, R., Dunn, R. R., and N. J. Sanders. 2011. Elevational gradients in phylogeneticstructure of ant communities reveal the interplay of biotic and abiotic constraints on diversity *Ecography* 34: 364–371.

Marx, H. E., C. Dentant, J. Renaud, R. Delunel, D. C. Tank, and S. Lavergne. 2017. Riders in the sky (islands): Using a mega-phylogenetic approach to understand plant species distribution and coexistence at the altitudinal limits of angiosperm plant life. *Journal of Biogeography* 44: 2618-2630.

Marx, H. E., D. E. Giblin, P. W. Dunwiddie, and D. C. Tank. 2016. Deconstructing Darwin's Naturalization Conundrum in the San Juan Islands using community phylogenetics and functional traits. *Diversity and Distributions* 22: 318–331.

Marx, H. E., M. Richards, G. M. Johnson, and D. C. Tank. 2019. Data from: Increasing phylogenetic stochasticity at high elevations on summits across a remote North American wilderness. Dryad Digital Repository. doi:10.5061/dryad.d06j48t

Mayfield, M. M., and J. M. Levine. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters* 13 1085–1093.

Mazel, F., T. J. Davies, L. Gallien, J. Renaud, M. Groussin, T. Münkemüller, and W. Thuiller. 2016. Influence of tree shape and evolutionary time-scale on phylogenetic diversity metrics. *Ecography* 39: 913-920.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees: Presented at the Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp. 1–8.

Morueta-Holme, N., K. Engemann, P. Sandoval-Acuña, J. D. Jonas, R. M. Segnitz, and J.-C. Svenning. 2015. Strong upslope shifts in Chimborazo's vegetation over two centuries since Humboldt. *Proceedings of the National Academy of Sciences, USA* 112: 12741–12745.

Munkemüller, T., F. C. Boucher, W. Thuiller, and S. Lavergne. 2015. Phylogenetic niche conservatism—common pitfalls and ways forward. *Functional Ecology* 29: 627–639.

Oksanen, J. F., G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. and Minchin, et al. 2017. vegan: Community Ecology Package. R package version 2.4-3. https://CRAN.R-project.org/package=vegan

Park, D. S., S. Worthington, and Z. Xi. 2018. Taxon sampling effects on the quantification and comparison of community phylogenetic diversity. *Molecular Ecology* 27: 1296-1308

Pattengale, N. D., M. Alipour, O. R. P. Bininda-Emonds, B. M. E. Moret, and A. Stamatakis. 2009. How Many Bootstrap Replicates Are Necessary? In: Batzoglou S, ed. *Research in Computational Molecular Biology: 13th Annual International Conference, RECOMB 2009*. Tucson, AZ, USA. May 18-21, 2009. Berlin, Germany & Heidelberg, Germany: Springer Berlin Heidelberg, pp. 184–200.

Pauli, H., M. Gottfried, S. Dullinger, O. Abdaladze, M. Akhalkatsi, J. L. B. Alonso, G. Coldea, et al. 2012. Recent Plant Diversity Changes on Europe's Mountain Summits. *Science* 336: 353–355.

Pavoine, S., M. B. Bonsall. 2010. Measuring biodiversity to explain community assembly: a unified approach. *Biological Reviews* 86: 792–812.

Pellissier, L., N. Alvarez, A. Espíndola, J. Pottier, A. Dubuis, J.-N. Pradervand, and A. Guisan. 2013. Phylogenetic alpha and beta diversities of butterfly communities correlate with climate in the western Swiss Alps. *Ecography* 36: 541–550.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, M. E. Alfaro, and L. J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30: 2216–2218.

Pigot, A.L., and R. S. Etienne. 2015. A new dynamic null model for phylogenetic community structure. *Ecology Letters* 18: 153–163.

Qian, H., R. E. Ricklefs, and P. S. White. 2005. Beta diversity of angiosperms in temperate floras of eastern Asia and eastern North America. *Ecology Letters* 8: 15–22.

Quintero, I., and W. Jetz. 2018. Global elevational diversity and diversification of birds. *Nature* 555: 246–250

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website https://www.R-project.org/.

Reid, R. R. 1963. Reconnaissance geology of the Sawtooth Range. *Idaho Bureau of Mines and Geology* 1–61.

Richardson, A. D., and A. J. Friedland. 2009. A Review of the Theories to Explain Arctic and Alpine Treelines Around the World. *Journal of Sustainable Forestry* 28: 218–242.

Roquet, C., W. Thuiller, and S. Lavergne. 2012. Building megaphylogenies for macroecology: taking up the challenge. *Ecography* 35: 1–14.

Schlatterer, E. F. 1972. A preliminary description of plant communities found on the Sawtooth, White Cloud, Boulder, and Pioneer Mountains. *United States Forest Service Report – Intermountain Region.*

Smith, J. M. B., and A. M. Cleef. 1988. Composition and Origins of the World's Tropicalpine Floras. *Journal of Biogeography* 15: 631–645.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 1–12.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24: 715–716.

Smith, S. A., and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690.

Steele, R. W. and the Intermountain Forest Range Experiment Station. 1981. Forest habitat types of central Idaho. U.S. Dept. of Agriculture, Forest Service Intermountain Forest and Range Experiment Station.

Stein, A., K. Gerstner, and H. Kreft. 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters* 17: 866–880.

Tucker, C. M., M. W. Cadotte, S. B. Carvalho, T. J. Davies, S. Ferrier, S. A. Fritz, R. Grenyer, et al. 2016. A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews* 92: 698–715.

Uribe-Convers, S., M. L. Settles, and D. C. Tank. 2016. A Phylogenomic Approach Based on PCR Target Enrichment and High Throughput Sequencing: Resolving the Diversity within the South American Species of Bartsia L. (Orobanchaceae). *PLoS ONE* 11: e0148203.

Valiente-Banuet, A., and M. Verdú. 2007. Facilitation can increase the phylogenetic diversity of plant communities. *Ecology Letters* 10: 1029–1036.

Vamosi, S. M., S. B. Heard, J. C. Vamosi, and C. O. Webb. 2009. Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology* 18: 572–592.

Vellend, M. 2010. Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology* 85: 183-206.

Webb, C. O. 2000. Exploring the Phylogenetic Structure of Ecological Communities: An Example for Rain Forest Trees. *American Naturalist* 156: 145–155.

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33: 475–505.

Winkler, D. E., R. J. Butz, M. J. Germino, K. Reinhardt, and L. M. Kueppers. 2018. Snowmelt Timing Regulates Community Composition, Phenology, and Physiological Performance of Alpine Plants. *Frontiers in Plant Science* 9: 1–13.

Winter, M., V. Devictor, and O. Schweiger. 2013. Phylogenetic diversity and nature conservation: where are we? *Trends in Ecology & Evolution* 28: 199–204.

Wright, D. H., and J. H. Reeves. 1992. On the Meaning and Measurement of Nestedness of Species Assemblages. *Oecologia* 92: 416–428.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Figure 1. Map of the Sawtooth National Forest (gray area on map inlay) in Idaho, USA, showing the locations of the nine high alpine summits sampled. Triangle colors correspond to different summits, and triangle size is proportional to maximum elevation: Horstmann Peak (3155 m), Braxon Peak (3156 m), Thompson (3203 m), Snowyside Peak (3246 m), Mount Cramer (3266 m), D.O. Lee Peak (3457 m), Salzburger Spitzl (3536 m), Castle Peak (3601 m), and Hyndman Peak (3660 m).

Figure 2. Community phylogeny of the alpine flora of the Sawtooth National Forest, Idaho, USA. Color bars on tips match colors of summits on the map and indicate the presence of each species on each summit. Gray bars closest to tip names indicate species that were collected from alpine meadows. Asterisks mark species with molecular sequence data available in GenBank, and diamonds indicate species with sequence data generated from high-throughput sequencing. Nodes that were congruent with the reference timetree ("congruified") are indicated by open black circles. Nodes with a light gray dot have bootstrap support (BS) between 75% and 95%, and those with a black dot have BS support ≥95%. Summits with high alpine meadows include Thompson Peak, D.O. Lee Peak, Salzburger Spitzl, and Hyndman Peak. Representative species within the six most species rich orders are highlighted: (A) *Carex* sp. on the summit of Snowyside Peak; (B) *Draba oligosperma* on Thompson Peak; (C) *Eriogonum ovalifolium* clinging to Snowyside Peak; (D) *Phyllodoce glanduliflora* at the base of Thompson Peak summit; (E) *Castilleja miniata* covering a high alpine meadow on Hyndman Peak; and (F) *Hulsea algida* scattered across a talus slope on Salzburger.

Figure 3. Phylogenetic α-diversity of alpine flora on summits across the Sawtooth National Forest, Idaho, USA: standardized effect size (SES) for mean nearest taxon distance (MNTD) and mean pairwise distance (MPD) estimated from the alpine community phylogeny. Tile rows show

community phylogenetic structure within each summit (ordered by increasing elevation across the *x*-axis) for all vascular plants (Tracheophyta) and each of the six most species-rich orders (Asterales, Poales, Caryophyllales, Lamiales, Brassicales, and Ericales). Warm tones (positive SES values) indicate phylogenetic overdispersion (high phylogenetic divergence), and cool tones (negative SES) indicate phylogenetic clustering (low phylogenetic divergence). Tiles with dots denote higher (or lower) observed divergence than expected by chance (from random resampling the community phylogeny of all alpine plants; $P < 0.05$). Cells filled with an "x" had too few species for comparison (only one species was present).

Figure 4. Statistical analysis of the relationship between environment and phylogenetic community structure in the Sawtooth National Forest, Idaho, USA. Linear regression of maximum elevation (independent variable) on standardized effect sizes (SES) for (A) mean nearest taxon phylogenetic distance (MNTD) and (B) mean pairwise phylogenetic distance (MPD). Separate models were performed for all vascular plants (Tracheophyta) and each of the six most species-rich orders (Asterales, Poales, Caryophyllales, Lamiales, Brassicales, and Ericales).

Figure 5. Phylogenetic β-diversity of alpine flora on summits across the Sawtooth National Forest, Idaho, USA. (A) Pairwise matrices showing species turnover between summits, measured by standardized effect sizes (SES) of UniFrac distances, decomposed into the portion corresponding to true turnover (UniFrac Turn) for all alpine species (top half) and those collected only on talus slopes (excluding species collected from alpine meadows; bottom half). Tiles with warm tones indicate high turnover between summit pairs (i.e., summits have unique species); cool tones indicate low turnover between summit pairs (i.e., summits share the same species). Tiles with asterisks show summit pairs with higher or lower turnover than expected (from random resampling of the phylogeny; $P < 0.05$). (B) Phylogenetic turnover between clades on the community phylogeny of summit species measured by Π_{ST} for all alpine species. Species subtending nodes with red dots have a higher-than-expected turnover between summits (i.e., appear only on certain summits), species subtending nodes with blue dots have a lower-than-expected turnover between summits (i.e., appear across all summits), and species subtending nodes with open circles have turnover no different than random.