



Resolving neutral and deterministic contributions to genomic structure in *Syntrichia ruralis* (Bryophyta, Pottiaceae) informs propagule sourcing for dryland restoration

Rob Massatti¹ · Kyle D. Doherty² · Troy E. Wood³

Received: 3 July 2017 / Accepted: 4 November 2017 / Published online: 14 November 2017
© Springer Science+Business Media B.V., part of Springer Nature (outside the USA) 2017

Abstract

Syntrichia ruralis is a cosmopolitan moss that occupies steep environmental gradients. In arid to semi-arid regions of the world it is a key component of biological soil crusts, which are fundamental to healthy dryland ecosystem processes. As such, *S. ruralis* has attracted the attention of conservationists seeking to restore degraded biological soil crust communities and their associated vascular flora. Here, we generate genomic data for *S. ruralis* populations that span climatic gradients across the Colorado Plateau of the southwestern USA to investigate the contributions of neutral and deterministic processes to the partitioning of genomic structure. Although *S. ruralis* appears to be highly dispersible, geographic proximity significantly predicts genomic similarity. In addition, even when taking into account apparently high migration rates among populations and spatial autocorrelation of allele frequencies, some genomic variation is explained by environmental gradients correlated with elevation and latitude. Consequently, efforts to restore dryland ecosystems by establishing *S. ruralis* as a foundation should include strategies to ensure that propagule sources of this moss are environmentally stratified and targeted to the current/future climates of restoration sites.

Keywords Phylogeography · North America · RADseq · Procrustes · Biological soil crust · Colorado Plateau

Introduction

The advent of molecular techniques that can efficiently yield abundant, genome-wide data (e.g., Ekblom and Galindo 2011; McCormack et al. 2013) has broadened researchers' abilities to infer ecological and evolutionary processes. For example, genomic data has been used to uncover adaptation to both novel and parallel environments (Colosimo et al. 2005; Nosil 2007), the outcomes of secondary contact between divergent evolutionary lineages (Gompert et al.

2010), and the effects of climate diversity on differentiation and diversification within species (Papadopoulou and Knowles 2015a, b). Such genomic study is especially relevant to restoration and conservation efforts, where accurate inference may have substantial impacts on success and thus the shaping and implementation of management and policy. Genomic analysis is particularly useful for less studied taxa, where evolutionary or ecological considerations are not evident *a priori*. Conservation-minded questions often focus on taxonomic identity, population connectedness and effective population sizes, the distribution of genetic variation, and inferring adaptation to environmental gradients; all of these topics have been successfully addressed using genomic data (e.g., Eckert et al. 2010; Escudero et al. 2014; Kool et al. 2013; Hipp et al. 2014; Merwe et al. 2014). Given that any genetic analysis used to inform conservation decisions should contain sufficient power to discriminate between deterministic and neutral processes with reasonable confidence, researchers are fortunate that practical, cost-effective methods for generating large genomic data sets exist (e.g., see Narum et al. 2013 and references therein).

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10592-017-1026-7>) contains supplementary material, which is available to authorized users.

✉ Rob Massatti
rmassatti@usgs.gov

¹ Southwest Biological Science Center, U.S. Geological Survey, Flagstaff, AZ 86001, USA

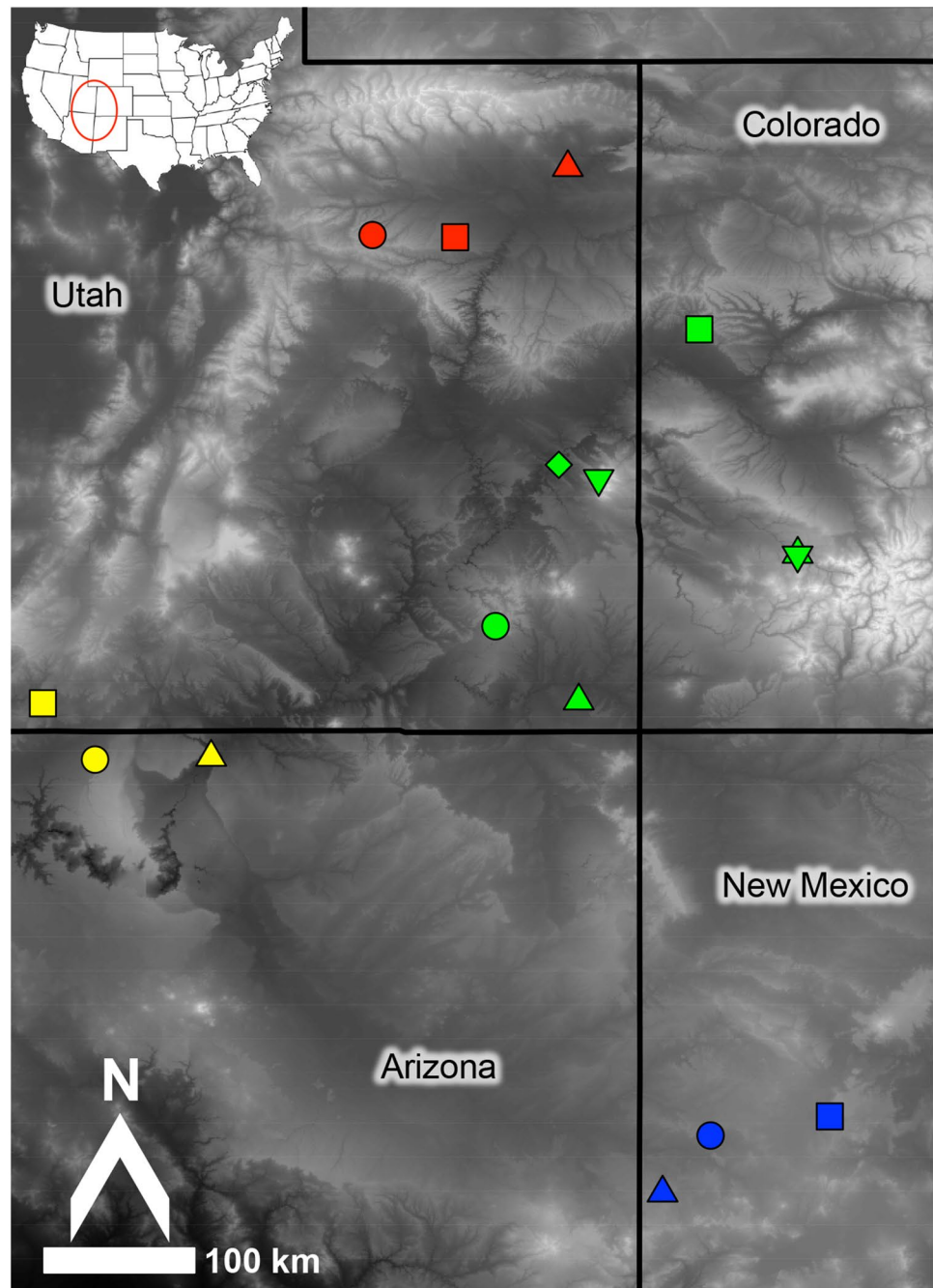
² School of Forestry, Northern Arizona University, 200 East Pine Knoll Drive, Flagstaff, AZ 86011, USA

³ 507 E. Charles Road, Flagstaff, AZ 86001, USA

Here, we assess phylogeographic structure in the widely distributed moss species, *Syntrichia ruralis* (Hedw.) F. Weber & D. Mohr (synonyms include *Tortula ruralis*, *Barbula ruralis*), across the Colorado Plateau of southwestern North America (Fig. 1). Regionally, *S. ruralis* is a common component of biological soil crusts (Rosentreter et al. 2007), which occupy arid landscapes worldwide. Biological soil crusts provide important, foundational ecosystem services in drylands, yet they are easily degraded by many types of land use activities and tend to be slow to recover (Weber et al. 2014). Consequently, significant resources are being devoted

to investigate biological soil crust reestablishment methods, including using *S. ruralis* reintroduction as a catalyst for biological soil crust regeneration (e.g., Antoninka et al. 2016). Prior to implementing a restoration protocol using *S. ruralis*, propagules must be generated for deployment, but data are lacking on the distribution of genetic diversity within the species and whether it corresponds with adaptive differentiation across environments. While recent experimental studies of *S. ruralis* suggest that populations exhibit ecological differences (Condon and Pyke 2016; Doherty et al. 2017),

Fig. 1 *Syntrichia ruralis* sampling localities. Colored shapes match those used in Figs. 2 and 3; see Fig. 2 for population names and regional affiliations (denoted by colors). The background greyscale depicts elevation, with whiteness positively correlated with elevation (range of background: 330–4228 m). Coordinates of localities, as well as associated environmental data, are provided in Supplementary Material Table 1. (Color figure online)



the utility and value of developing one versus multiple propagule sources for *S. ruralis* is largely unknown.

Biological characteristics unique to mosses lead to conflicting expectations regarding the distribution of genetic diversity and opportunity for local adaptation within and among *S. ruralis* populations on the Colorado Plateau. First, the dispersibility of *S. ruralis* may be high enough to overwhelm any partitioning of genetic variation at this regional scale, and populations may harbor similar levels of genetic diversity compared to the species' region-wide extent. *Syntrichia ruralis* occurs in a diverse range of environments, including dry to moist soil and rock from low to high elevations, on all continents except Antarctica and Australia (Mishler 2007). Within North America, it has been recorded in northern and southern states from Maine to California, in several Canadian provinces, and in Greenland. Cosmopolitan distributions of organisms often suggest high dispersibility, which is supported in many moss species by molecular and experimental studies (Shaw 2000; Vanderpoorten et al. 2008; Lonnell et al. 2012; Lewis et al. 2014; Paasch et al. 2015). For example, several moss species with amphi-Atlantic distributions apparently function as metapopulation networks via high rates of migration (Desauniers et al. 2016). The geographic scale in our study (i.e., the Colorado Plateau) represents less than 400,000 km² (i.e., less than 720 km of latitudinal distance and 550 km of longitudinal distance; Fig. 1). In the Great Basin region west of the Colorado Plateau, Paasch et al. (2015) estimated an average of 32 migrants per generation in a closely related species, *S. caninervis*, between sites separated by approximately 270 km. Finally, known locations of *S. ruralis* across the study area indicate that the species is limited by environmental rather than dispersal limitation (i.e., it is always found when environmental conditions are appropriate; personal communication, J. Belnap and M. Bowker). Cumulatively, these data suggest that gene flow may impede the formation of phylogeographic structure and adaptive differentiation in *S. ruralis* at small to moderate spatial scales. If true, sourcing propagules from a single population could be acceptable for restoration purposes.

Alternatively, life history characteristics related to *S. ruralis*' reproduction may promote population level differentiation across the Colorado Plateau. In all mosses, the haploid, gametophytic generation is dominant, and sexual reproduction depends upon sperm being able to pass through a water film to archegonia for the production of sporophytes and spores. In arid climates with little moisture, such as those prevalent across the Colorado Plateau, opportunities to reproduce sexually appear to occur infrequently (Paasch et al. 2015; personal observation, K. Doherty), limiting recombination among haplotypes and the overall abundance of sporophytes (Mishler 1990). Lower rates of sexual reproduction are expected to result in fewer opportunities for gene

flow and more differentiation among populations, although asexual reproduction and dispersal via gametophytic fragmentation is known to occur (e.g., Miles and Longton 1990). In addition, *S. ruralis* is a monoicous species, allowing for both intergametophytic and intragametophytic selfing, which leads to elevated levels of heterozygote deficiency (Eppley et al. 2007); in fact, intragametophytic selfing can result in complete homozygosity in a single generation (Hendrick 1987). *Syntrichia ruralis*' reproductive mode, along with other factors that promote differentiation (e.g., topographic complexity associated with steep environmental gradients; Sexton et al. 2014), would suggest that creating multiple propagule sources will preserve the highest levels of genetic diversity and/or local adaptation.

We develop a genomic dataset for 15 *S. ruralis* populations sampled across the Colorado Plateau to investigate: (1) If genomic structure exists at the scale of the Colorado Plateau; and (2) If genomic variation correlates with regional environmental gradients. We find evidence of population genomic structure across these 15 *S. ruralis* populations, and our analyses suggest that some of this structure is explained by environmental gradients associated with latitude and elevation. In light of these results, we provide basic guidance for propagule sourcing across the Colorado Plateau. In addition, we discuss refining sampling schemes for future projects to account for tight correlations between geography and environmental variation, which will bolster inference into environmental adaptation. We hope that our work directly informs ongoing international efforts to develop plant materials sources (Plant Conservation Alliance 2015) that are adapted to the environmental conditions where they are deployed and, consequently, that maximize restoration success and reduce overall costs (Oldfield and Olwell 2015).

Materials and methods

Field sampling and DNA extraction

Individuals from 15 *S. ruralis* populations stratified across the monsoonal precipitation gradient on the Colorado Plateau were sampled during fall 2012 and spring 2013 (Fig. 1; Supplementary Material Table 1). Populations were separated by at least 20 km. Seven to eighteen *S. ruralis* cushions separated by at least 10 m were sampled from each population. Field identification of *S. ruralis* was based upon costa color (red) and strongly recurved laminae (Rosentreter et al. 2007). Dry tissue samples were placed in coin envelopes and stored in the dark at 20 °C. Field identification of the samples was complemented by evaluation of prepared slides of laminae; laminae that were one cell thick, strongly recurved along their entire lengths, and that lacked branched papillae were considered diagnostic for *S. ruralis* (Weber and

Wittman 2007). Single gametophytes determined to be *S. ruralis* were isolated from each cushion for DNA extraction. Gametophyte tissues were disrupted using a GenoGrinder 2000 (SPEX SamplePrep, Metuchen, New Jersey, USA), and DNA was extracted using a modified sorbitol protocol (Storchova et al. 2000).

Next-generation sequencing

We selected 107 specimens for RADseq genotyping (Supplementary Material Table 2). Extracted genomic DNAs were processed into a reduced complexity library using a restriction fragment based procedure (for details, see Peterson et al. 2012). Briefly, each DNA sample was doubly digested with EcoRI and SphI, followed by the ligation of Illumina adaptor sequences and unique 5 bp barcodes. Ligation products were pooled and the fragments were amplified by 12 cycles of PCR. A Pippin Prep (Sage Science, Beverly, Massachusetts, USA) was used to size select fragments between 400 and 500 bp. The resultant pool of fragments was sequenced at the University of Texas Genome Sequencing and Analysis Facility in two lanes on a HiSeq 2500 (Illumina, San Diego, California, USA) to generate 100 bp, single-end reads.

Illumina data processing

Raw data were demultiplexed, filtered, and assembled *de novo* in pyRAD version 3.0.66 (Eaton 2014). Only reads with fewer than four low quality sites, no adaptor contamination, and unambiguous barcodes and restriction cut sites were retained. Following initial processing, sequences of each individual were clustered into highly similar stacks (i.e., sets of sequences inferred to be from a single locus) using VSEARCH (Rognes et al. 2016), followed by the estimation of rates of sequencing error (Lynch 2008). Because the gametophytes from which tissues were sampled are haploid, heterozygosity (H) was set to 0. Individuals' stacks of raw reads were independently summarized into consensus sequences, which were subsequently clustered among individuals to generate RADseq loci (i.e., sets of sequences inferred to be homologous). Because of variation in sequencing coverage, mutations in the restriction site defining the RAD loci, and other factors, not every individual had a sequence for every RADseq locus (Eaton 2014). Parameter values used for clustering were based on suggestions from the literature (Eaton and Ree 2013; Eaton 2014; Takahashi et al. 2014; Massatti et al. 2016) but modified to account for the haploid genome of *S. ruralis* gametophytes and the intraspecific scope of our study. For example, the percentage similarity required to cluster sequences into a stack was set at the relatively stringent level of 0.95 because heterozygosity could be excluded *a priori* and all samples came from a single species across a relatively small geographic area. In

addition, the minimum stack depth for a cluster was set to 3 because, when it is known that $H=0$, the binomial probability of correctly calling a site is higher at low depths. Full parameter details are archived with associated data sets on Dryad (<https://doi.org/10.5061/dryad.18bv1>).

Dataset preparation

Following data processing, we used a custom script to filter the variant call file (vcf) created by pyRAD and to write a STRUCTURE formatted file (script available at <https://doi.org/10.5061/dryad.18bv1>). First, SNPs and individuals that contained high proportions of missing data were removed to prevent their disproportionate effects on analyses (e.g., see Wang et al. 2012—large amounts of missing data can skew biological patterns). Specifically, we excluded all SNPs with > 70% missing data across individuals and individuals missing > 90% of loci (a threshold which set off a clear outlier group of individuals). Next, we iteratively pruned the SNP dataset in 10% intervals (from 70 to 10% missing data per SNP) and assessed the information content and the impact of missing data in the resultant dataset with principal component analysis (PCA) using the `dudi.pca` function in R (R Core Team 2016; center = T, scale = T with missing data replaced by the mean frequency of the corresponding allele, as recommended by Jombart et al. 2008) and by assessing Θ genetic diversity statistics and F_{ST} values calculated in ARLEQUIN version 3.5 (Excoffier and Lischer 2010). The goal of this second step was to maximize the information content in the dataset while minimizing missing data (see Massatti and Knowles 2014). Finally, we randomly selected one SNP per RADseq locus in order to comply with the assumptions of the analytical methods we employed. Thus, the final dataset (available at <https://doi.org/10.5061/dryad.18bv1>) represents a balance between minimizing potentially negative impacts of missing data on inference and minimizing the loss of information content (signal).

Genomic differentiation and structure

To visualize the major axes of population genetic variation contained in the SNP dataset, we used two complementary approaches: (1) Bayesian clustering implemented in STRUCTURE version 2.3.4 (Pritchard et al. 2000; Falush et al. 2003) and (2) a multivariate ordination method that accounts for spatial patterns, spatial principal component analysis (sPCA), implemented in the ADEGENET package (Jombart 2008) in R. We used both methods to allow comparison of results across approaches with different sets of assumptions. For example, STRUCTURE assumes that loci are in equilibrium and unlinked while sPCA does not. Furthermore, Bayesian clustering may be inappropriate when populations are structured across a cline (Jombart et al. 2008) because it

may overestimate genetic structure, while a spatially explicit multivariate method can identify genetic structure, including clines, and accounts for spatial autocorrelation (Frantz et al. 2009). sPCA achieves this by incorporating Moran's I (Moran 1948, 1950), which allows the detection of structure at local (e.g., differentiation between neighboring individuals) and global (e.g., genetically distinguishable groups and clines; see Jombart et al. 2008 for details) levels.

STRUCTURE was run across K -values ranging from 1 to 10 without assigning population membership *a priori*. Ten independent runs per K were conducted, each with 150,000 burn-in and 500,000 MCMC iterations, using the “Admixture Model” and “Correlated Allele Frequency Model” with default settings; our application of STRUCTURE to *S. ruralis* is similar to that of other studies on haploid organisms (e.g., Falush et al. 2003). STRUCTURE HARVESTER (Earl and vonHoldt 2012) and DISTRUCT (Rosenberg 2004) were used to visualize results, and the most probable K was chosen based on ΔK (Evanno et al. 2005). For sPCA analysis, geographic locations of individuals were created by jittering population locations (factor = 3), and a Delauney triangulation graph was used to create the connection network required by the sPCA function.

At the individual level, the fit of genomic distances to environmental and/or geographic distances was quantified using Procrustes analyses. Within a genetic context, Procrustes analyses are interpreted under the assumption that individuals from populations closer in geographic proximity will be more closely related (i.e., isolation by distance), and deviations from this pattern bring to light interesting biological and/or historical phenomena (see Wang et al. 2012; Knowles et al. 2016 for details and examples). We apply this technique to test whether geography + environment, as implicit within the environmental PCA distance matrix (see below) due to tight correlations between latitude and environmental gradients (see Results), explains more genomic variation than does geographic distance alone. These analyses complement Mantel tests, partial Mantel tests, and linear models (see below) because they rely upon individual level genomic differentiation (i.e., distances among individuals in PC space), yet they represent a testing framework for the same phenomenon (i.e., if environmental variation explains genomic variation). Analyses were performed using the ‘protest’ function of VEGAN in R. The significance of the association statistic between the genetic PC values and geography or environmental space was evaluated based on 10,000 permutations. Note that we used PCA values (PC1 and PC2) for Procrustes analyses instead of sPCA values; because sPCA analyses incorporate geographic information into genetic patterns, testing the fit of sPCA values to geography would result in a spuriously high correlation.

At the level of populations, the impact of environmental factors and geographic distance on genetic structure was

tested in R with Mantel tests, partial Mantel tests, and linear models. Distance matrices for these analyses included: (1) genetic distance between populations as represented by pairwise F_{ST} -values; (2) Euclidean environmental distance between populations calculated from principal component 1 (PC1) and PC2 of an environmental PCA; and (3) Euclidean geographic distance between populations as calculated in ARC GIS 10 (Esri, Redlands, California, USA). The environmental PCA was performed on 19 bioclimatically informative raster layers (Hijmans et al. 2005) in ARC GIS 10; PCA was performed directly on the climate layers instead of values extracted from sampling points to capture the variation in the environment while avoiding any bias in the sampling points. Each of the potential predictors of genetic distance (geographic distance and environmental distance) was tested separately with Mantel tests; significance was assessed by 10,000 permutations. Similarly, a partial Mantel test assessed the correlation between environmental variation and genetic distance while partialling out geographic distance. The distance matrices were also used to create sequential general linear models in R using the `lm()` function; the residuals from an initial linear model between geographic and genetic PC distances were used in a linear model with environmental distance to determine the contribution of environmental distance on genetic patterns, while controlling for geographic distance.

Results

NGS data quality, processing, and dataset construction

DNA extraction was successful for 107 of the approximately 150 individuals sampled. Two lanes of Illumina sequencing produced nearly 150 million reads across these 107 individuals (average of $1,550,667 \pm 957,581$), of which more than 132 million passed filtering and quality control procedures (Supplementary Material Table 2). Based on low coverage or low quality of reads, 34 individuals were excluded from further analyses (Supplementary Material Table 2). Filtering the resulting dataset to include only unlinked SNPs with fewer than 60% missing data across individuals resulted in 1678 SNPs. Alternative datasets containing fewer missing data did not noticeably affect downstream analyses (e.g., patterns of differentiation as assessed by F_{ST} and PCA were similar when missing data was $\leq 60\%$) and are not discussed.

Genomic differentiation and structure

STRUCTURE analyses across multiple K -values indicated that $K=3$ is the most likely number of genetic clusters (the K value following the highest ΔK , hereafter referred to as

the most likely K). We present illustrations of both $K=2$ and $K=3$, as both are helpful in assessing genetic structure within *S. ruralis* (Fig. 2). Higher K values decomposed genetic variation into groupings that are difficult to interpret, and thus are not presented. At $K=2$, genetic variation within *S. ruralis* essentially broke down into northern (red) and southern (green) genomic axes (Fig. 2). However, within the North and South regions, there are admixed populations (i.e., populations containing individuals wholly or partially with the less dominant genomic axis), and most populations from the East and West regions contain a complement of both genomic axes. At $K=3$, the green genomic axis breaks down into two components—a green and a blue genomic axis (Fig. 2). The blue genomic axis is largely absent in the North region and high elevation populations of the East region.

The first two sPCA eigenvalues associated with global structure were used to characterize genetic variation among *S. ruralis* populations (Supplementary Material Figs. 1 and 2); sPC1 and sPC2 account for 35.5 and 14.8% of the

variation in the SNP dataset, respectively. Using color to represent genetic similarity, individuals in populations from the North region are easily distinguished from individuals in South region populations, while East and West populations are intermediate, containing either a mix of North and South identities or containing individuals of intermediate identities (Fig. 3a). These patterns are concordant with those resolved by STRUCTURE at $K=2$, as well as patterns in the regular PCAs that were used when assessing information content across screened datasets generated at variable levels of missing data (Supplementary Material Fig. 3).

When populations are graphed in environmental space (Fig. 3b), genetic similarity resolved by the sPCA is similarly situated along PC2 compared to latitude in Fig. 3a, but it is better sorted along PC1 compared to longitude in Fig. 3a (i.e., the colors form a more continuous gradient from yellow to red). Loading values of the variation in the bioclimatic variables onto PC1 and PC2 indicate that PC1 is positively correlated with environmental variables primarily attributed to higher elevations (e.g., lower mean annual

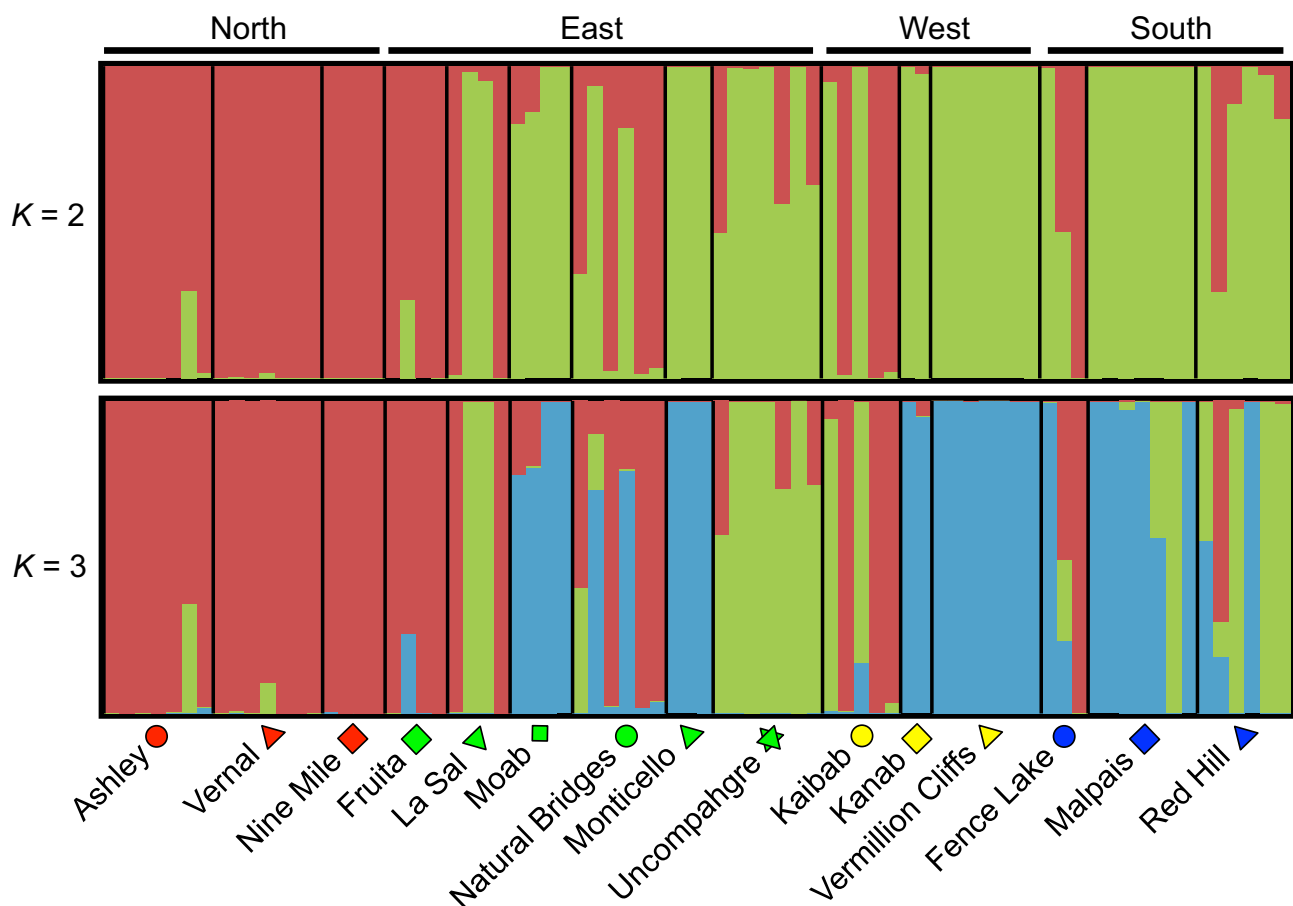


Fig. 2 Plots of posterior probabilities for individuals, which could be assigned to 2 ($K=2$) or 3 ($K=3$) groups by the STRUCTURE algorithm. The K groups within an analysis are differentiated by color, and thinner vertical black lines delimit populations, whose names are listed

across the bottom. Symbols following population names correspond to those used in Figs. 1 and 3. Regional groupings of populations referred to in the text are noted along the top, and match the colors of the symbols. (Color figure online)

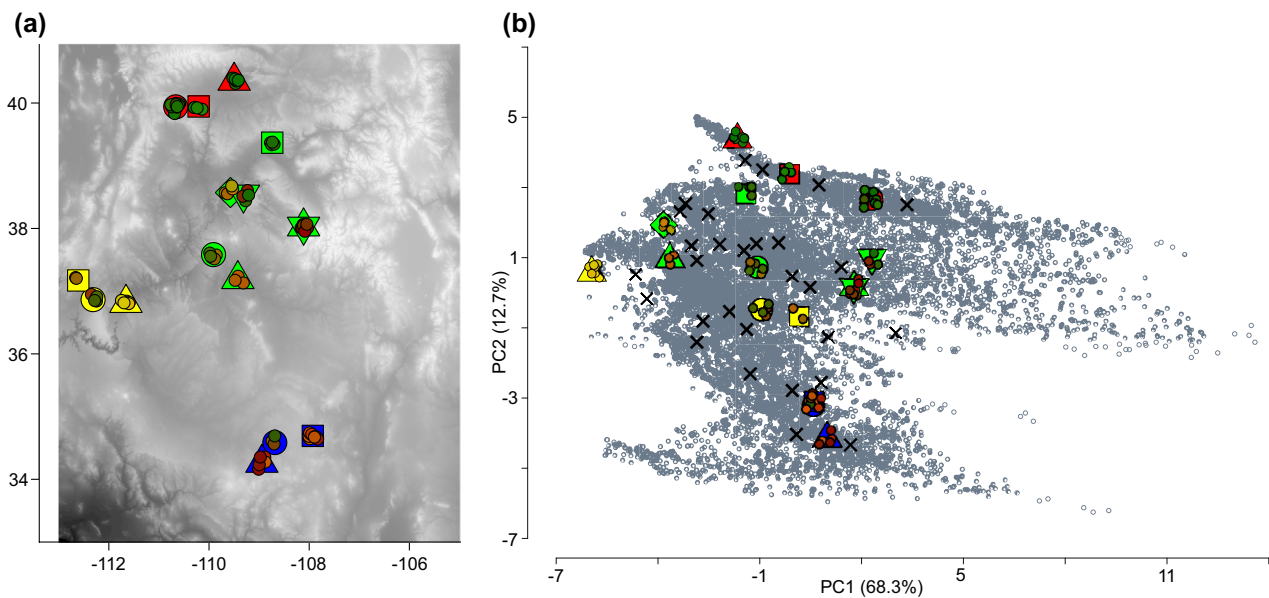


Fig. 3 Populations mapped in **a** geographic space, and **b** environmental space. Symbols match those of Figs. 1 and 2. Within the environmental PCA (**b**), small grey circles represent the environmental space of the Colorado Plateau, and X's mark other known *S. ruralis* populations (i.e., they help depict *S. ruralis*' overall environmental tolerance across the Colorado Plateau). PC2 is highly correlated with geographic distance (see “Results”). Populations' individuals are depicted by overlain, smaller colored circles, where the colors among individuals represent genetic similarity, as calculated from sPC1 and sPC2. Individuals are artificially jittered so they do not

completely overlap within a population. The gradation of individuals' colors along the x and y axes in (**a**) and (**b**) illustrates why the Procrustes correlation is higher when rotating genetic similarity to environmental space compared to geographic space; compared to (**a**), genetic similarity in (**b**) forms a more continuous gradient from yellow to red along the x-axis and from red to green along the y-axis. In other words, accounting for environmental distance and geographic distance together (**b**) explains genetic structure better than considering geographic distance in isolation (**a**). (Color figure online)

temperature, higher annual precipitation), while PC2 is positively correlated with environmental variables attributed to higher latitudes on the Colorado Plateau (e.g., colder winter temperatures, warmer temperatures during drier summers) (Supplementary Material Table 5). Both PC1 and PC2 are significantly correlated with geography ($p < 0.0001$ for each), but the correlation is weak for PC1 ($r = 0.05$) compared to PC2 ($r = 0.69$), as resolved by Mantel tests between geographic and PC1 or PC2 distances calculated for 1000 randomly sampled points across the study area. The pattern of genetic similarity when graphed in environmental space (Fig. 3b) is similar to the $K = 3$ STRUCTURE result, with lower elevation sites having lower PC1 values (i.e., corresponding roughly to the blue genomic axis in Fig. 2) and higher elevation sites having higher PC1 values (i.e., the green genomic axis in Fig. 2). The Procrustes analysis between genetic similarity and geographic distance (i.e., the fit of the two datasets in Fig. 3a) resolved a significant correlation ($t = 0.58$, $p = 0.001$), while the analysis between genetic similarity and environmental space (i.e., Fig. 3b) had a higher correlation ($t = 0.72$, $p = 0.001$), indicating that considering geographic and environmental distance together accounts for more genetic variation than does geographic distance alone.

Genomic diversity varied among populations, with θ_S ranging from 25.6 to 161.2 (average = 100.1 ± 37.4) and θ_π ranging from 15.7 to 94.2 (average = 63.5 ± 21.3) (Supplementary Material Table 3). Pairwise genetic differentiation (F_{ST}) ranged from ~ 0 to ~ 0.19 (Supplementary Material Table 4). Mantel correlations were significant for both F_{ST} by environmental distance ($r = 0.31$, $p = 0.0097$) and F_{ST} by geographic distance ($r = 0.22$, $p = 0.044$) (see Supplementary Material Table 4 for distance matrices). Furthermore, environmental distance remained a significant predictor of genetic distance when geographic distance was accounted for as a covariate ($r = 0.24$, $p = 0.044$). While Mantel tests have been critiqued when used for population genetic inference (e.g., see Raufaste and Rousset 2001; Guillot and Rousset 2013), our focus here is on the relative importance of environmental versus spatial prediction of genetic structure. Furthermore, we do not solely depend on the Mantel framework to evaluate determinants of genetic structure, but present them as complementary to other analytic approaches. In congruence with the results from the Mantel tests, an initial general linear model resulted in geographic distance explaining 7.8% of the genetic PC dataset variation ($p < 0.0001$). A subsequent linear model resulted in the environmental

PC data explaining 11% of the variation within the residual dataset ($p < 0.0001$).

Discussion

This study highlights the power of next-generation sequencing data to resolve genomic structure within a putatively highly dispersible species at a relatively small spatial scale. Results of complementary analyses suggest *S. ruralis*' genome displays a significant pattern of isolation by distance (IBD) in addition to adaptation to regional environmental gradients, namely those associated with elevation and latitude. Below we discuss these patterns in relation to neutral and deterministic processes, as well as to *S. ruralis*' unique biological features. We also discuss future avenues of research that would help untangle the contributions of geographic and environmental distance to patterns of genomic variation.

Population structure of *Syntrichia ruralis*

Syntrichia ruralis is perceived to be highly dispersible due to its expansive range over multiple continents and dispersal patterns resolved in other moss species, including one that is very closely related (Paasch et al. 2015). Our results support a high dispersal capacity for this species. At both $K=2$ and $K=3$ within STRUCTURE analyses (Fig. 2), populations in all regions contain admixed individuals and/or non-admixed individuals representing alternative genomic axes (i.e., colors). This result is reinforced in the sPCA (Fig. 3) and the PCA (Supplementary Material Fig. 3) by the overlap/mixing of individuals across populations in color or in PC space, respectively. Similar mixing of genomic axes within populations has been resolved in other species inferred to be highly dispersible (e.g., Pinheiro et al. 2014). Despite the inferred high dispersibility of *S. ruralis*, we resolved a significant pattern of isolation by distance (IBD) using both Mantel and Procrustes analyses, with the latter manifesting as a latitudinal trend of differentiation (Figs. 2, 3 and Supplementary Material Fig. 3). Practically, these results imply that creating a restoration propagule source from one or a few geographically proximate populations will be insufficient to represent the genetic diversity of the species across the Colorado Plateau.

Population level sampling across the Colorado Plateau for phylogeographic analyses is largely absent for other species, which makes setting the genomic patterns of *S. ruralis* into a comparative context difficult. The most analogous study may be Galbreath et al. (2010), in which three distinct mitochondrial lineages for the American pika were resolved. Higher differentiation in pikas compared to mosses is unsurprising due to their dispersal method and limited climatic

niche (Smith 1974), but it does illustrate that environmentally specialized species with low dispersibility will show definite spatial genetic structure across the Colorado Plateau. Another habitat specialist in the adjacent Great Basin region, the pallid kangaroo mouse, displays east–west differentiation into distinct clades, with further division into subclades in the eastern portion of its range (Hafner et al. 2008). Phylogeographic research involving plants in the western United States generally has been conducted at large spatial scales and focused on the Rocky Mountains and/or Pacific Northwest; even in instances where population sampling occurred across smaller areas, genetic variation was more finely partitioned compared to *S. ruralis* (e.g., Richardson et al. 2002; DeChaine and Martin 2005; Brunsfeld et al. 2007; Massatti and Knowles 2014).

Signatures of environmental adaptation

The *S. ruralis* data provide evidence that population level genomic variation is structured along environmental gradients, despite potentially high rates of gene flow. For example, the Mantel test was significant when correlating genetic distance (F_{ST}) with distance in environmental PC space (Fig. 3), and this relationship remained significant after controlling for geographic distance. Considering individual level genomic PC data instead of F_{ST} values, Procrustes analyses resolved a higher correlation when fitting genetic data to environmental space compared to geographic space. Note that geography is implicit within environmental space; neither environmental PC1 nor PC2 was independent of geographic distance. Finally, the linear regression between environmental distance and the residuals resulting from a regression between genomic distance and geographic distance was highly significant, explaining approximately 11% of the residual variation. Similar to our suggestion that the IBD pattern prevents a geographically limited propagule source from accurately representing the genomic diversity of *S. ruralis*, the ability of environmental variables to explain genomic variation implies that stratifying propagule sources across relevant environmental gradients would best represent environmentally driven (and potentially adaptive) genomic variation.

Genomic patterns displayed in the STRUCTURE and PCA analyses can also be interpreted in an environmental context. In the $K=3$ STRUCTURE analysis, the populations dominated by the red genomic axis (North region populations + Fruita; Fig. 2) fall north of the average northern extent of summer monsoon moisture (Mitchell 1976). The blue and green genomic axes correspond relatively well with lower and higher elevation sampling sites, respectively, within the region that experiences summer monsoons (Supplementary Material Table 1). Within the genomic PCA that does not account for geography (Supplementary Material Fig. 3), PC1

may be environmentally interpreted as a monsoonal gradient axis (from low intensity to high intensity), while PC2 corresponds with elevation (also from low to high, when discounting populations from the North region). However, we note that due to the complex ways in which environmental variables covary, not to mention how axes of genetic variation interact, direct environmental interpretation of the genomic PCA is difficult. Cumulatively, our data support a pattern of isolation by environment (IBE) (Nosil et al. 2008, 2009) within *S. ruralis*, a result that has emerged from many genetic studies on plants (see Sexton et al. 2014 and references therein). While we are not attempting to resolve individual loci exhibiting signs of selection (see below), isolation by adaptation (IBA) may explain why the dominant axes of genomic variation correspond to the major environmental gradients across the Colorado Plateau (Orsini et al. 2013). For researchers and practitioners whose goals are to create geographically tailored propagule sources for restoration and conservation purposes (e.g., Broadhurst et al. 2008; Havens et al. 2015), resolving the underlying drivers (i.e., loci) of environmental adaptation may not be necessary in instances where IBA shapes genomic structure.

Ecological studies of local adaptation in moss species reveal that populations prefer their respective local environments. Experiments utilizing *S. ruralis* suggest, at a minimum, adaptation to different moisture regimes. For example, when two *S. ruralis* populations were sampled from sites with different environments (cool/mesic and warm/dry) and transferred to a separate warm/dry site, the warm/dry-sourced material produced more biomass during three years of observation (Condon and Pyke 2016). In another study, individuals from five environmentally unique populations of *S. ruralis* were grown under uniform greenhouse conditions and then subjected to a watering protocol designed to induce stress. Populations displayed a range of tolerances to these watering treatments, perhaps due to adaptation to site-specific precipitation regimes (Doherty et al. 2017). Similar adaptive processes are inferred to be at work in other moss species, in which genetic variation is commonly structured across altitudinal gradients that are strongly correlated with temperature and precipitation (Korpelainen et al. 2012; Pisa et al. 2013; Magdy et al. 2016).

Given that we infer IBA likely shaped *S. ruralis*' genomic variation, the processes promoting the maintenance of alternative genomic axes within populations becomes intriguing, especially in the context of creating appropriate propagule sources for restoration projects. Resulting from the putatively limited opportunities for recombination between individuals and low levels of spore generation in arid climates (Bowker et al. 2000; Nath and Asthana 2001), asexual reproduction/dispersal and environmental filtering (i.e., Cavender-Bares et al. 2004) may be the predominant methods by which *S. ruralis* populations are tuned to local

environments. However, landscapes with high topographic complexity, and thus a diversity of microhabitats at a small spatial scale, may promote the persistence of immigrants adapted to different suites of environmental characteristics. Combined with infrequent episodes of sexual reproduction, these processes would explain the mixture of alternative genomic axes within populations (Figs. 2, 3 and Supplementary Material Fig. 3) that underlies the inferred patterns of IBD and IBE. Furthermore, this mixture of more and less adapted genomic axes may be maintained through time in an environmentally transitional landscape by high levels of phenotypic plasticity (e.g., desiccation tolerance; Stark et al. 2012); ultimately, genomic diversity may allow populations the flexibility to respond relatively quickly to changing environmental conditions. We note, however, that the generation time for *S. ruralis* is unknown, though its ability to withstand extreme environmental events (e.g., drought) may suggest an ability to persist for long periods. Modeling alternative hypotheses, such as random assembly versus environmental filtering, under a range of environmental and demographic parameters (e.g., He et al. 2013; Bemmels et al. 2016; Massatti and Knowles 2016; Knowles and Massatti 2017) would help determine if and how the process of local adaptation proceeds and the relative importance of environmental, demographic, and historical processes to genomic structure.

Alternative explanations and future analyses

The resolution of genomic structure across *S. ruralis* populations is a testament to the utility of next-generation sequencing data, given the spatial scale of our analyses and the biology of *S. ruralis*. While our analyses indicate that the environment significantly explains some genomic variation, additional research would help rule out alternative hypotheses. For example, the resolved genomic patterns could be consistent with the recolonization of East and West region populations from northern and southern refugia following the Last Glacial Maximum (sensu Knowles and Massatti 2017). However, if population extirpation and recolonization strictly followed neutral demographic processes, we would not expect major axes of environmental variation to significantly explain a relatively large proportion of genomic variation beyond that which is tightly correlated with geography.

Patterns of *S. ruralis* genomic variation could also indicate that ecological and evolutionary processes occur at a geographic scale larger than the Colorado Plateau. For example, a wider geographic sampling of *S. ruralis* populations may resolve discrete genomic axes associated with distinct environments, such as a low elevation genomic axis (blue in Fig. 3b), a montane genomic axis (green in Fig. 3b), and a genomic axis outside of the main monsoonal influence (red in Fig. 3b). Due to the limited geographic scope of this study and the steep environment gradients across the

Plateau, we could be resolving few populations displaying what may be an “optimal” genomic identity. In other words, the short geographic distances between environmentally distinct locations (for example, the low elevation Moab population was sampled 20 km from the high elevation La Sal; Fig. 1) result in high levels of genomic mixing compared to the species’ overall range, which would suggest that the Colorado Plateau is a place where several different genomic axes intermingle (similar to the pattern resolved in Galbreath et al. 2010, as discussed above). Accordingly, it is important to sample populations over an appropriate geographic scale, depending on the ecology of the study organism and the questions at hand (Anderson et al. 2010; Manel et al. 2010).

Other important considerations for testing adaptation to local environments include constructing sampling schemes that explicitly control for geography and using environmental variables that are relevant to the organism (see Rellstab et al. 2015 and references therein). Due to the correlation between geography and environmental space across the Colorado Plateau, it is difficult to disentangle the contribution of geographic and environmental processes, a problem that haunts many spatial genetic analyses. For example, it may be that the continuity of *S. ruralis* populations across the Colorado Plateau, along with high migration rates, swamp out any signal of IBD, and we may be misinterpreting environmental adaptation as IBD. Only a refined sampling scheme specifically designed to disentangle environmental distance from geographic distance will be able to address this issue. For example, Lotterhos and Whitlock (2015) suggest sampling scattered pairs of closely spaced populations that exhibit substantial environmental differences in order to identify loci adapted to environmental gradients; this approach controls for both geography and for genetic structure arising from neutral demographic processes. In addition, assessing environmental adaptation may be refined by only using environmental variables with low levels of correlation or that are specifically hypothesized to have biological influence (Rellstab et al. 2015). Furthermore, using modeling techniques and/or machine learning algorithms to determine how allele frequencies change according to environmental variables would resolve the variables most important to *S. ruralis* (Lee and Mitchell-Olds 2011; Briec et al. 2015; Shryock et al. 2015).

The quality of the *S. ruralis* genomic dataset imposes analytical limitations on our demographic and environmental adaptation inferences. While biological patterns are robust to analyses relying on different assumptions (e.g., STRUCTURE versus sPCA analyses, and Mantel tests versus Procrustes rotations versus regression), our dataset contains a high proportion of missing data (up to 60% per SNP across 1678 SNPs). Missing data affect our ability to calculate accurate and significant locus-specific F_{ST} values, thus making F_{ST} -based outlier tests for adaptive loci unreliable. In addition,

missing data limit the opportunity to use modeling methodologies (e.g., FASTSIMCOAL2; Excoffier et al. 2013) to calculate demographic parameters within tight confidence intervals; optimally, modeling demographic scenarios with appropriate data would provide the best estimates of migration rates among regions and among populations within regions. Given the limitations discussed thus far, future research would benefit from (1) a region-wide sampling scheme designed to dissociate environmental from geographic distance; (2) representative sampling of populations located in regions adjacent to the Colorado Plateau; (3) including fewer individuals per library to increase overall depth of coverage; (4) uniform (and higher) individual sampling per population (i.e., some of our populations contain as few as 2 to 3 individuals); and (5) potentially modifying aspects of the ddRAD library protocol to further reduce the complexity of the genome (e.g., using different restriction enzymes).

Conservation implications

One motivation for generating genomic data for *S. ruralis* across the Colorado Plateau was to inform propagule sourcing for biological soil crust restoration. An efficient program of *S. ruralis* materials development (e.g., Antoninka et al. 2016) should identify an appropriate set of source materials for restoration projects that can be deployed across a breadth of environments. Our data suggest that *S. ruralis* restoration material may be best tailored to a given restoration site by minimizing environmental distance rather than geographic distance. The genomic patterns identified here imply that sources should include, at a minimum, populations at higher and lower elevations both within and outside of the area affected by summer monsoonal precipitation. Because the monsoonal precipitation gradient is tightly linked to latitude, stratifying by latitude satisfies the requirement to maximize geographic distance to maintain genomic diversity (due to the pattern of IBD). We do acknowledge that some populations were significantly mixed in terms of genomic ancestry; however, it remains to be seen whether individuals not tailored to a specific environment (i.e., that do not follow our expectations of local environmental adaptation) persist for multiple generations, or if they are temporary inhabitants that immigrated into a local environment to which they are ill-suited. Given that we suggest matching *S. ruralis* propagules to restoration sites by environmental distance, careful consideration must be given to not only a site’s current climate, but also that which it will likely encounter in the future (e.g., Havens et al. 2015).

Acknowledgements The authors are grateful for the support and helpful advice of their colleagues, including E. Milano for assistance with next-generation sequencing, M. Bowker for moss expertise, and J. Belnap and two anonymous reviewers for comments that improved the quality of this manuscript. This work was funded by the Bureau of

Land Management's Colorado Plateau Native Plant Program. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

References

- Anderson CD, Epperson BK, Fortin MJ, Holderegger R, James P, Rosenberg MS, Scribner KT, Spear S (2010) Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Mol Ecol* 19:3565–3575
- Antoninka A, Bowker MA, Reed SC, Doherty K (2016) Production of greenhouse-grown biocrust mosses and associated cyanobacteria to rehabilitate dryland soil function. *Restor Ecol* 2:324–335
- Bemmels JB, Ortego J, Knowles LL (2016) Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Mol Ecol* 25:4889–4906
- Bowker MA, Stark LR, McLetchie DN, Mishler BD (2000) Sex expression, skewed sex ratios, and microhabitat distribution in the dioecious desert moss *Syntrichia caninervis* (Pottiaceae). *Am J Bot* 87:517–526
- Brieuc MS, Ono K, Drinan DP, Naish KA (2015) Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Mol Ecol* 24:2729–2746
- Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Veski PA, Yates C (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol Appl* 1:587–597
- Brunsfeld SJ, Miller TR, Carstens BC (2007) Insights into the biogeography of the Pacific Northwest of North America: evidence from the phylogeography of *Salix melanopsis*. *Syst Bot* 32:129–139
- Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA (2004) Phylogenetic overdispersion in Floridian oak communities. *Am Nat* 163:823–843
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* 307:1928–1933
- Condon LA, Pyke DA (2016) Filling the interspace—restoring arid land mosses: source populations, organic matter, and overwintering govern success. *Ecol Evol* 6:7623–7632
- DeChaine EG, Martin AP (2005) Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. *Am J Bot* 92:477–486
- Désamoré A, Patiño J, Mardulyn P, McDaniel SF, Zanatta F, Laenen B, Vanderpoorten A (2016) High migration rates shape the postglacial history of amphi-Atlantic bryophytes. *Mol Ecol* 25:5568–5584
- Doherty KD, Bowker MA, Antoninka AJ, Johnson NC, Wood TE (2017) Biocrust moss populations differ in growth rates, stress response, and microbial associates. *Plant Soil*. <https://doi.org/10.1007/s11104-017-3389-4>
- Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Res* 4:359–361
- Eaton DA (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30:1844–1849
- Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). *Syst Biol* 62:689–706
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB (2010) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185:969–982
- Eklom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* 107:1–5
- Eppley SM, Taylor PJ, Jesson LK (2007) Self-fertilization in mosses: a comparison of heterozygote deficiency between species with combined versus separate sexes. *Heredity* 98:38–44
- Escudero M, Eaton DA, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Mol Phylogenet Evol* 79:359–367
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* 10:564–567
- Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLOS Genet* 9:e1003905
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J Appl Ecol* 46:493–505
- Galbreath KE, Hafner DJ, Zamudio KR, Agnew K (2010) Isolation and introgression in the Intermountain West: contrasting gene genealogies reveal the complex biogeographic history of the American pika (*Ochotona princeps*). *J Biogeogr* 37:344–362
- Gompert Z, Lucas LK, Fordyce JA, Forister ML, Nice CC (2010) Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Mol Ecol* 19:3171–3192
- Guillot G, Rousset F (2013) Dismantling the Mantel tests. *Methods Ecol Evol* 4:336–344
- Hafner JC, Upham NS, Reddington E, Torres CW (2008) Phylogeography of the pallid kangaroo mouse, *Microdipodops pallidus*: a sand-obligate endemic of the Great Basin, western North America. *J Biogeogr* 35:2102–2118
- Havens K, Vitt P, Still S, Kramer AT, Fant JB, Schatz K (2015) Seed sourcing for restoration in an era of climate change. *Nat Area J* 35:122–133
- He Q, Edwards DL, Knowles LL (2013) Integrative testing of how environments from the past to the present shape genetic structure across landscapes. *Evol Int J Org Evol* 67:3386–3402
- Hedrick PW (1987) Genetic load and the mating system in homosporous ferns. *Evol Int J Org Evol* 41:1282–1289
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978
- Hipp AL, Eaton DA, Cavender-Bares J, Fitzek E, Nipper R, Manos PS (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. *PLOS One* 9:e93975
- Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403–1405

- Jombart T, Devillard S, Dufour AB, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* 101:92–103
- Knowles LL, Massatti R (2017) Distributional shifts—not geographic isolation—as a probable driver of montane species divergence. *Ecography*. <https://doi.org/10.1111/ecog.02893>
- Knowles LL, Massatti R, He Q, Olson LE, Lanier HC (2016) Quantifying the similarity between genes and geography across Alaska's alpine small mammals. *J Biogeogr* 43:1464–1476
- Kool JT, Moilanen A, Treml EA (2013) Population connectivity: recent advances and new perspectives. *Landsc Ecol* 28:165–185
- Korpelainen H, Jägerbrand A, von Cräutlein M (2012) Genetic structure of mosses *Pleurozium schreberi* (Willd. ex Brid.) Mitt. and *Racomitrium lanuginosum* (Hedw.) Brid. along altitude gradients in Hokkaido. *Jpn J Bryol* 4:309–312
- Lee CR, Mitchell-Olds TH (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Mol Ecol* 20:4631–4642
- Lewis LR, Rozzi R, Goffinet B (2014) Direct long-distance dispersal shapes a New World amphitropical disjunction in the dispersal-limited dung moss *Tetraplodon* (Bryopsida: Splachnaceae). *J Biogeogr* 41:2385–2395
- Lönnell N, Hylander K, Jonsson BG, Sundberg S (2012) The fate of the missing spores—patterns of realized dispersal beyond the closest vicinity of a sporulating moss. *PLOS One* 7:e41987
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* 24:1031–1046
- Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Mol Biol Evol* 25:2409–2419
- Magdy M, Werner O, McDaniel SF, Goffinet B, Ros RM (2016) Genomic scanning using AFLP to detect loci under selection in the moss *Funaria hygrometrica* along a climate gradient in the Sierra Nevada Mountains, Spain. *Plant Biol* 18:280–288
- Manel S, Poncet BN, Legendre P, Gugerli F, Holderegger R (2010) Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Mol Ecol* 19:3824–3835
- Massatti R, Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of codistributed species: genomic evidence in montane sedges (*Carex* L.) from the Rocky Mountains. *Evol Int J org Evol* 68:2833–2846
- Massatti R, Knowles LL (2016) Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. *Mol Ecol* 25:3974–3986
- Massatti R, Reznicek AA, Knowles LL (2016) Utilizing RADseq data for phylogenetic analysis of challenging taxonomic groups: a case study in *Carex* sect. *Racemosae*. *Am J Bot* 103:337–347
- McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT (2013) Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 66:526–538
- Merwe M, McPherson H, Siow J, Rossetto M (2014) Next-Gen phylogeography of rainforest trees: exploring landscape-level cpDNA variation from whole-genome sequencing. *Mol Ecol Res* 14:199–208
- Miles CJ, Longton RE (1990) The role of spores in reproduction in mosses. *Bot J Linean Soc* 104:149–173
- Mishler BD (1990) Reproductive biology and species distinctions in the moss genus *Tortula*, as represented in Mexico. *Syst Bot* 1:86–97
- Mishler BD (2007) *Syntrichia*. In: *Flora of North America Editorial Committee* (eds) 1993 + *Flora of North America North of Mexico*, 20 + vols. New York and Oxford. vol 27, p 618
- Mitchell VL (1976) The regionalization of climate in the western United States. *J Appl Meteorol* 15:920–927
- Moran PA (1948) The interpretation of statistical maps. *J R Stat Soc Series B Methodol* 10:243–251
- Moran PA (1950) Notes on continuous stochastic phenomena. *Biometrika* 37:17–23
- Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* 22:2841–2847
- Nath V, Asthana AK (2001) Reproductive Biology of Bryophytes. In: *Reproductive Biology of plants*, Springer, Berlin, pp. 148–174
- Nosil P (2007) Divergent host plant adaptation and reproductive isolation between ecotypes of *Timema cristinae* walking sticks. *Am Nat* 169:151–162
- Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: 'isolation by adaptation' and multiple roles for divergent selection. *Evolution Int J org Evolution* 62:316–336
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402
- Oldfield S, Olwell P (2015) The right seed in the right place at the right time. *Bioscience* 65:955–956
- Orsini L, Vanoverbeke J, Swillen I, Mergey J, de Meester L (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol* 22:5983–5999
- Paasch AE, Mishler BD, Nosratinia S, Stark LR, Fisher KM (2015) Decoupling of sexual reproduction and genetic diversity in the female-biased Mojave Desert moss *Syntrichia caninervis* (Pottiaceae). *Int J Plant Sci* 176:751–761
- Papadopoulou A, Knowles LL (2015a) Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution Int J org Evolution* 69:1501–1517
- Papadopoulou A, Knowles LL (2015b) Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance across a Mediterranean Pleistocene Aggregate Island Complex. *Mol Ecol* 24:4252–4268
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS One* 7:e37135
- Pinheiro F, Cozzolino S, Draper D, de Barros F, Felix LP, Fay MF, Palma-Silva C (2014) Rock outcrop orchids reveal the genetic connectivity and diversity of inselbergs of northeastern Brazil. *BMC Evol Biol*:1–15
- Pisa S, Werner O, Vanderpoorten A, Magdy M, Ros RM (2013) Elevational patterns of genetic variation in the cosmopolitan moss *Bryum argenteum* (Bryaceae). *Am J Bot* 100:2000–2008
- Plant Conservation Alliance (2015) National seed strategy for rehabilitation and restoration 2015–2020. US Department of the Interior, Bureau of Land Management. <https://www.blm.gov/programs/natural-resources/native-plant-communities/national-seed-strategy>. Accessed 30 June 2017
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 3 March 2017
- Raufaste N, Rousset F (2001) Are partial mantel tests adequate? *Evolution Int J org Evolution* 55:1703–1705
- Rehstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Mol Ecol* 24:4348–4370
- Richardson BA, Brunsfeld SJ, Klopfenstein NB (2002) DNA from bird dispersed seed and wind disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Mol Ecol* 11:215–227

- Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:e2584
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Res* 4:137–138
- Rosentreter R, Bowker M, Belnap J (2007) A field guide to biological soil crusts of western US drylands: common lichens and bryophytes. US Government Printing Office, Denver. p. 104
- Sexton JP, Hangartner SB, Hoffmann AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution Int J org Evolution* 68:1–15
- Shaw AJ (2000) Molecular phylogeography and cryptic speciation in the mosses. *Mielichhoferia elongata* and *M. mielichhoferiana* (Bryaceae). *Mol Ecol* 9:595–608
- Shryock DF, Havrilla CA, DeFalco LA, Esque TC, Custer NA, Wood TE (2015) Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: a multivariate, spatially-explicit approach to guide ecological restoration. *Conserv Genet* 16:1303–1317
- Smith AT (1974) The distribution and dispersal of pikas: influences of behavior and climate. *Ecology* 55:1368–1376
- Stark LR, Brinda JC, Nicholas McLetchie D, Oliver MJ (2012) Extended periods of hydration do not elicit dehardening to desiccation tolerance in regeneration trials of the moss *Syntrichia caninervis*. *Int J Plant Sci* 173:333–343
- Štorchová H, Hrdličková R, Chrtek J Jr, Tetera M, Fitze D, Fehrer J (2000) An improved method of DNA isolation from plants collected in the field and conserved in saturated NaCl/CTAB solution. *Taxon* 49:79–84
- 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
- Vanderpoorten A, Devos N, Goffinet B, Hardy OJ, Shaw AJ (2008) The barriers to oceanic island radiation in bryophytes: insights from the phylogeography of the moss *Grimmia montana*. *J Biogeogr* 35:654–663
- Wang C, Zöllner S, Rosenberg NA (2012) A quantitative comparison of the similarity between genes and geography in worldwide human populations. *PLOS Genet* 8:e1002886
- Weber WA, Wittman RC (2007). *Bryophytes of Colorado*. Pilgrims Process
- Weber B, Büdel B, Belnap J (2014) *Biological soil crusts: an organizing principle in drylands*. Springer International Publishing, Cham