The historical context of contemporary climatic adaptation: a case study in the climatically dynamic and environmentally complex southwestern United States

Rob Massatti and L. Lacey Knowles

Introduction

While genome-wide (e.g. phylogeographic) and locus-specific (e.g. genome scan) methods are often used to investigate neutral and adaptive genetic variation, respectively, integrating both into a combined framework is less common. However, such a framework is valuable when trying to understand how evolutionary processes, such as adaptation, unfold in dynamic, heterogeneous systems. For example, while the impacts of cycling glacial and interglacial periods during the Pleistocene on species’ distributions across northerly latitudes are clear (Massatti and Knowles 2016, Wächter et al. 2016, Knowles and Massatti 2017, Bemmels et al. 2019), their effects on adaptive processes and the genetic variation available for selection to act upon has received less attention (Sork et al. 2016). Considering adaptive processes is important because mismatches between species’ fundamental and realized niches in glacial versus interglacial
periods are expected (Jackson and Overpeck 2000), particularly in terms of the environmental space occupied (Fitzpatrick and Hargrove 2009). For example, geographic constraints on species’ distributions during glacial periods (e.g., Italian or Iberian peninsular refugia – Hewitt 2000) likely increased the probability that species experienced aberrant climatic conditions (e.g., novel climates sensu Williams and Jackson 2007) or a subset of climatic space compared to their interglacial distributions. Given that the cycling of glacial and interglacial periods was too quick for all of an organism’s adaptively significant genetic variation to be generated by mutation during any individual glacial or interglacial period (Hewitt 2000), standing genetic variation must have been important for selection to act upon (Barrett and Schluter 2008). Understanding the process of adaptation is important because it helps us understand how rapidly species can adapt to future environmental changes (Palumbi 2001).

A framework structured around complementary investigations to elucidate a species’ responses to demographic and selective processes can help determine the role of standing genetic variation relative to dynamic histories. We apply this framework to *Hilaria jamesii* (Torr.) Benth. (Poaceae; common name – James’ galleta grass; synonym – *Pleuraphis jamesii*), a rhizomatous, wind-pollinated, perennial grass restricted to the arid landscapes of the southwestern United States (Fig. 1). As a C_{4} (or ‘warm-season’) species and given its habitat preference, *H. jamesii* would not have persisted across most, if not all, of its current distribution during Pleistocene glacial periods. In contrast to the arid and semi-arid floras that are currently distributed across lower elevations of the Colorado Plateau and Great Basin, floristic patterns reconstructed from pollen and packrat midden fossils (Cole 1990) support mixed coniferous forests during the last glacial period (Thompson and Anderson 2000, Ray and Adams 2001, Grayson 2006, Waltari and Guralnick 2009). Furthermore, the refugia available to *H. jamesii* were constrained by the regional geography and paleoclimate. For example, the Sierra Nevada in California and the Rocky Mountains in Colorado physically prevented westward and eastward migration, respectively. Likewise, cooler climates in the north would have physiologically prevented northward migration. As such, refugial areas were restricted south of presently suitable habitat (Cole et al. 2008, 2013), which increased the likelihood of mismatches between *H. jamesii*’s potential habitat based

Figure 1. Geographic map of the southwestern United States, with the Colorado Plateau (sensu lato) outlined in red. Larger pie charts within the Colorado Plateau overlap with sampling localities and depict the local proportion of the genetic clusters delimited by STRUCTURE across all sampled individuals. Smaller pie charts outside of the Colorado Plateau represent individual plants (sampled from herbarium specimens) and their relative affinities to the genetic clusters. The dotted line approximates the range limit of *H. jamesii*; several outlier localities are present south of this polygon. In the text, the ‘eastern’ population corresponds to the localities that are predominantly orange, the ‘northern’ population to predominantly green and the ‘western’ population to predominantly blue. The black star in southeastern Utah represents the location of Canyonlands National Park.
on its climatic tolerances and the actual climatic conditions it encountered. Moreover, during post-glacial recolonization of the Colorado Plateau and Great Basin, migration across steep climatic gradients, such as those associated with higher latitudes and increasingly continental climates, suggests the species was subject to recent, strong selective pressures.

Here, we test if *H. jamesii*’s putative adaptive genetic variation is linked to its glacial and post-glacial history. Specifically, we employ genome scan methods to detect putative adaptive loci, which are interpreted in the context of 1) the species’ biogeographic history, 2) coalescent model-based evaluations of alternative historical scenarios detailing the timing and order of population splits and the number of ancestral populations (i.e. glacial refugia) and 3) the region’s contemporary climatic gradients, as well as estimates of the paleoclimate since the last glacial period inferred from independent sources. Our data support a single, southern refugium for the species during the last glacial period and subsequent migration north and west concurrent with regional, post-glacial warming. This migration brought the species into contact with steep climatic gradients, which the species successfully adapted across such that it presently occupies a broader environmental space compared to its glacial refugium; the timing of range expansion and complexity of adaptation to climatic gradients support the hypothesis that natural selection acted upon standing genetic variation. This framework can be extended to other regional species to investigate the pervasiveness of standing genetic variation in facilitating adaptive responses to Pleistocene glacial cycles.

**Material and methods**

**Field sampling and data generation**

*Hilaria jamesii* was collected from 27 localities on the Colorado Plateau during the 2017 growing season (Fig. 1, Supplementary material Appendix 1 Table A1). Leaf tissues from 10 to 25 plants each separated by ≥ 30 m were collected from each locality and stored in silica desiccant. DNA was extracted using DNeasy 96 Plant Kits (Qiagen, Germantown, MD, USA). Sampling localities were chosen by stratifying the climatic space occupied by *H. jamesii* across the Colorado Plateau and selecting multiple sites from representative clusters (Fig. 2). This sampling scheme was designed to mitigate the confounding effects of covariation between geographic distance and environmental gradients in genetic/adaptive analyses (e.g. genetic processes, such as isolation by distance, may be difficult to distinguish from adaptive processes, as might be expected along steep environmental gradients – Rellstab et al. 2015). In addition, *H. jamesii*’s regional distribution was represented by herbarium voucher specimens (Fig. 1, Supplementary material Appendix 1 Table A2), and an *H. mutica* (Buckley) Benth. specimen was included as an outgroup for phylogenetic analyses.

Genomic DNAs from 387 individuals (349 from sampling localities + 38 herbarium specimens) were individually barcoded and processed into libraries using a restriction fragment-based procedure (Peterson et al. 2012) modified to utilize separate indexing reads. Raw data were demultiplexed using custom scripts and ‘fastq-mlt’ in ea-utils (Aronesty 2011). The remaining processing was accomplished using stacks v2.2 (Catchen et al. 2013) and followed the r80 protocol detailed in Rochette and Catchen (2017) (Supplementary material Appendix 1).

**Population structure and history**

Two approaches were employed to infer genetic structure, including 1) Bayesian clustering implemented in structure v2.3.4 (Falush et al. 2003) and 2) a multivariate ordination method that accounts for spatial patterns, spatial principal components analysis (sPCA), implemented in the adegenet package (Jombart 2008) in R. These methods complement one another because they rely on different assumptions (Jombart et al. 2008, Frantz et al. 2009; Supplementary material Appendix 1). In addition, a maximum-likelihood phylogenetic analysis in PhyML 3.0 (Guindon et al. 2010) estimated the evolutionary history of *H. jamesii*. Model selection was implemented using smart model selection (Lefort et al. 2017) and nodal support values were calculated from 1000 bootstrap iterations.

Demographic modeling to elucidate the number of refugial *H. jamesii* populations, the history and timing of population splits, and approximations of population sizes was conducted using the allele frequency spectrum method (Gutenkunst et al. 2009) implemented in fastsimcoal2 (v2603; Excoffier et al. 2013). Parameters under two 3-population models (allowing for a single refugium or multiple refugia – Fig. 3) were estimated because population structure analyses may be interpreted as supporting a variable number of refugia (see Results). The sampling localities with significant admixture were excluded given the goal was to estimate regional diversification history (see Supplementary material Appendix 1 Table A1 for population assignments). Loci identified as potentially adaptive by either redundancy analysis or using latent factor mixed models (see below) were excluded and one SNP per locus was randomly chosen to satisfy the fastsimcoal2 assumption that SNPs are in linkage equilibrium. To remove missing data for the calculation of the joint site frequency spectrum and minimize errors with allele frequency estimates, each population was subsampled using easySFS.py (<https://github.com/isaacovercast/easySFS.py>) and only loci found in ≥ 40 individuals from each of the eastern, northern and western populations were retained (see Results for population definitions and Supplementary material Appendix 1).

**Characterizing climatic gradients**

Regional-scale climatic gradients across the Colorado Plateau were characterized using the 19 bioclimatically informative variables from WorldClim v2 (Fick and Hijmans 2017), as well as a data layer representing monsoonal precipitation, which was created using the sum of July–September
average precipitation/the total yearly precipitation (per cell) from WorldClim v2. We reduced the 19 bioclimatic variables + monsoonal precipitation to three principal components (PCs) using ‘princomp’ in R, as implemented through ‘rasterPCA’ in rstoolbox. PC1–3 explained 42, 35 and 11% of the climatic variation, respectively. Correlations between the PC axes, elevation, longitude, latitude and the bioclimatic variables suggest PC1 is a proxy for elevational climatic gradients ($r = -0.81$), PC2 is a proxy for a latitudinal climatic gradients ($r = -0.91$) and PC3 is closely associated with mean diurnal temperature range ($r = -0.83$) (Supplementary material Appendix 1 Table A3). Environmental PCs were subsequently used to identify putative adaptive loci (see below).

**Identifying putative adaptive loci**

Redundancy analysis (RDA), a multivariate environmental association method shown to detect weak, multilocus selection...
and identify adaptation with low rates of false positives and high rates of true positives across a range of demographic histories, sampling designs and sample sizes (Capblancq et al. 2018, Forester et al. 2018), was used to identify putative adaptive loci across climatic gradients. RDA was implemented following the methodology outlined in Forester et al. (2018). Population-level allele frequencies were computed using ‘makefreq’ in adegenet. Putative adaptive loci were identified using a two standard deviation cutoff (i.e. \( z = 2 \)), which is desirable because it allows identification of adaptive loci affected by weak and strong selection. We avoid making assumptions about the strength of the selection, which is implied by tests of selection based on more stringent cutoffs, because adaptation from standing genetic variation may be facilitated by weak and/or strong selective forces (Hermisson and Pennings 2005, Barrett and Schluter 2008).

Given that the demographic history of *H. jamesii* is linked with regional climatic gradients (e.g. those associated with latitude – see Results and Fig. 1), we also used latent factor mixed models (LFMMs) because LFMMs directly include an estimation of population structure when identifying adaptive loci with the goal of reducing false identifications (i.e. outlier loci resulting from demographic, rather than adaptive processes) (Frichot et al. 2013). Multiple LFMMs were generated using \( K \)-values (i.e. the number of latent factors) ranging from 3 to 5, which represent the most likely \( K \) genetic clusters suggested by structure and sPCA, as well as closely related values (Frichot and François 2015). The latent factors were implemented in LFMMs using ‘snmf’ in lea v1.2.0 in \( r \) (Frichot and François 2015) (Supplementary material Appendix 1).

To verify that putative adaptive loci may be biologically meaningful and adaptively significant, we blasted all loci identified by RDA and LFMMs (932 – see Results) to the National Center for Biotechnology Information’s nr database (<https://blast.ncbi.nlm.nih.gov>). We retained putative adaptive loci that matched database sequences of graminoids with an \( E \)-value of \( \leq 10^{-3} \). We then investigated the potential function of these sequences using the UNIPROT database (<www.uniprot.org>). We note that, while determining specific candidate loci influencing adaptation is beyond the scope of this study (Pavlidis et al. 2012) and unfeasible due to the lack of general understanding across a wide range of genes of how sequence variations are adaptive, it remains an informative exercise to confirm the potential function of putative adaptive loci. In addition, such inferences can provide valuable information to researchers assessing the function of protein sequences and how their variations may coincide with exogenous environmental factors.

**Results**

**NGS data quality, processing and dataset construction**

Illumina sequencing produced \( > 6.4 \times 10^6 \) reads across 387 individuals (average of \( 1.35 \times 10^6 \) reads per individual) after removing low-quality reads and adapter contamination (Supplementary material Appendix 1 Table A1, A2). Because of low coverage and/or putative admixture with *H. mutica*, 18 individuals were excluded from further analyses. The dataset for sPCA, structure, RDA and LFMMs included one randomly chosen SNP from every locus that was present in \( \geq 50\% \) of individuals from each sampling locality; this metric was chosen to provide a robust dataset for RDA, which utilized population-level allele frequencies. As a result, the dataset for sPCA and structure included 369 individuals and 9534 unlinked SNPs (average missing data per SNP = \( 17.2 \pm 9.7\% \); average missing data per individual = \( 17.2 \pm 3.8\% \)). The dataset used in LFMMs and RDA contained the same loci with fewer individuals (341 individuals and 9534 unlinked SNPs); individuals represented by herbarium vouchers were excluded because of RDA’s population-level allele frequency requirement. Genetic diversity and corrected AMOVA \( F_{ST} \)-statistics were calculated across 86 867 loci and the maximum-likelihood phylogenetic inference was conducted on 63 209 SNPs; loci for these analyses were determined by the filtering values used in the stacks populations program (see Methods). Data generated during this study are available from the USGS ScienceBase-Catalog (Massatti 2020).

**Population structure and history**

A population genetic structure of \( K=3 \) was inferred by STRUCTURE analyses as the most likely number of genetic clusters in *H. jamesii* (Fig. 1). Similarly, the first three eigenvalues associated with global structure were resolved as significant by sPCA, with sPC axes 1–3 explaining 67% of the total genetic variation (Supplementary material Appendix 1 Fig. A1). The three *H. jamesii* populations correspond to non-overlapping geographic regions, which hereafter are referred to as the ‘eastern’, ‘northern’ and ‘western’ population (Fig. 1). The eastern population is resolved as most similar to the outgroup (*H. mutica*) in the maximum-likelihood tree reconstruction, while the herbarium specimens representing the Great Basin are the most differentiated relative to the outgroup (Fig. 4). Given that the included sampling localities represent the majority of *H. jamesii*’s range, it is unlikely that there are unidentified, major genetic clusters, though substructure within these populations may be expected (Massatti and Knowles 2014).

Comparison of the divergence models (Fig. 3) identified a population history of sequential divergence (i.e. model 2) as most probable; an AIC score of 263 910 as opposed to 263 942 indicates a minimal loss of information with this model, as opposed to one involving divergence from a single refugial population (i.e. model 1). However, there are many similarities in parameter estimations under the two alternative population histories (Table 1), with the primary differences limited to estimated migration rates between the western and northern populations and their times of divergence. Regardless, estimated divergence times postdate the Last Glacial Maximum (LGM) in both models, indicating the recency of observed genetic structure in *H. jamesii* (i.e. there is no genetic evidence that contemporary populations were founded from > 1 refugium that persisted during the last glacial period).
Diversity statistics averaged across sampling localities of the three populations indicate consistent levels of expected heterozygosity and nucleotide diversity and low values of Wright’s inbreeding coefficient (Supplementary material Appendix 1 Table A4). Only a slight deviation is apparent when grouping the transitional sampling localities together (i.e. those in southeastern Utah and western Colorado – Fig. 1), where slightly higher heterozygosity and diversity values consistent with the localities’ admixed identities were estimated.

Figure 4. Maximum likelihood tree showing relationships among sampling localities. Branch support is measured from 1000 bootstrap replicates. Colored circles next to sampling locality names depict the genetic cluster to which each aligns – localities in which > 1 genetic cluster is prominent are colored with the dominant genetic cluster at the center and the secondary genetic cluster as the outer margin (i.e. localities from southeastern Utah and western Colorado, Fig. 1). The inset displays the relationship between *H. jamesii* and the outgroup (*H. mutica*). See Supplementary material Appendix 1 Table A1 for sampling locality information. This analysis also includes the herbarium specimens – Supplementary material Appendix 1 Table A2 details the locality to which each herbarium specimen was assigned.

Combs Ridge

Hidden Mtn

Bear Mtns

Manzanares

East Sangres

Eastern Plains

Black Mesa

Coal Creek

Jesus Mesa

Box Canyon

Rio Grande

Needles

Sevenmile Canyon

San Rafael Swell

WY Basin

Westwater

Poison Spring

Book Cliffs

Vernal

Factory Butte

Devils Playground

Salt Lake

Western Great Basin

North Rim

Kaiparowits Plateau

Alvej Wash

Blue Mesa

Red Mtn

Lees Ferry

Pleuraphis mutica

Box Canyon

Rio Grande

Jesus Mesa

Coal Creek

Black Mesa

Eastern Plains

East Sangres

Manzanares

Bear Mtns

Hidden Mtn

Combs Ridge

Gypsum Gap

Kutz Canyon

Fruita Canyon

Lees Ferry

Red Mtn

Salt Lake

Western Great Basin

North Rim

Kaiparowits Plateau

Alvey Wash

Blue Mesa

Red Mtn

Lees Ferry

Fruita Canyon

Gypsum Gap

Kutz Canyon

Combs Ridge

Hidden Mtn

Bear Mtns

Manzanares

East Sangres

Eastern Plains

Black Mesa

Coal Creek

Jesus Mesa

Box Canyon

Rio Grande

Needles

Sevenmile Canyon

San Rafael Swell

WY Basin

Westwater

Poison Spring

Book Cliffs

Vernal

Factory Butte

Devils Playground

Figure 4. Maximum likelihood tree showing relationships among sampling localities. Branch support is measured from 1000 bootstrap replicates. Colored circles next to sampling locality names depict the genetic cluster with which each aligns – localities in which > 1 genetic cluster is prominent are colored with the dominant genetic cluster at the center and the secondary genetic cluster as the outer margin (i.e. localities from southeastern Utah and western Colorado, Fig. 1). The inset displays the relationship between *H. jamesii* and the outgroup (*H. mutica*). See Supplementary material Appendix 1 Table A1 for sampling locality information. This analysis also includes the herbarium specimens – Supplementary material Appendix 1 Table A2 details the locality to which each herbarium specimen was assigned.

Diversity statistics averaged across sampling localities of the three populations indicate consistent levels of expected heterozygosity and nucleotide diversity and low values of Wright’s inbreeding coefficient (Supplementary material Appendix 1 Table A4). Only a slight deviation is apparent when grouping the transitional sampling localities together (i.e. those in southeastern Utah and western Colorado – Fig. 1), where slightly higher heterozygosity and diversity values consistent with the localities’ admixed identities were estimated. \( F_{ST} \)-values between sampling localities varied in a predictable manner, with localities closer in geographic proximity (and within the same population) having lower values compared to geographically distant sampling localities or those assigned to different populations (Supplementary material Appendix 1 Table A5). *Hilaria jamesii*’s average \( F_{ST} \) was low (0.07), which was beneficial for implementing RDA because this value is similar to the structure in the testing dataset used in Forester et al. (2018) (i.e. \( F_{ST} = 0.05 \)).

Identification of putative adaptive loci

Genomic scan methods resolved a total of 932 putative adaptive loci (9.8% of the total loci assessed), of which 131 were identified by both RDA and LFMMs. Specifically, RDA identified 652 putative adaptive loci; the distribution of \( z \)-scores between 2 and 3 indicates that most loci are weakly selected, although strongly selected loci are present as well (Supplementary material Appendix 1 Fig. A2).
Table 1. Results of parameter estimation with FASTSIMCOAL2 under two alternative models. Composite maximum likelihood estimates of: effective population sizes (N) are presented as the number of individuals per population; divergence times (T) are presented as the number of generations (i.e. number of years ago, as 1 generation = 1 yr); and migration rates (m) are presented as 2 Nm. See Fig. 3 to determine where each parameter maps onto the respective model. 95% confidence intervals were calculated from 100 parametric bootstrap replicates.

| Parameter | Model 1 | 95% CI | Model 2 | 95% CI |
|-----------|---------|--------|---------|--------|
| N_1       | NA      | NA     | 176 299 | 159 909–178 637 |
| N_{East}  | 179 122 | 176 968–178 788 | 143 085 | 136 571–141 797 |
| N_{North} | 128 895 | 124 668–125 929 | 126 756 | 121 636–124 481 |
| N_{Areq}  | 5 982 953 | 5 943 443–6 150 475 | 5 326 009 | 5 141 818–5 330 378 |
| T_1       | 20 050 | 20 661–20 906 | 11 557 | 11 111–11 417 |
| T_2       | 5.6    | 5.5–5.7 | 16.625 | 16 852–17 672 |
| m_{ne}    | 7.8    | 7.9–8.1 | 6.5    | 6.5–6.9 |
| m_{mm}    | 5.9    | 6.1–6.2 | 7.4    | 7.3–7.9 |
| m_{me}    | 3.4    | 3.4–3.5 | 1.3    | 0.9–1.2 |
| m_{mm}    | 3.4    | 3.4–3.5 | 0.7    | 0.5–0.7 |

LFMMs resolved 411 unique putative adaptive loci after intersecting the results across K-values (Supplementary material Appendix 1 Table A6). The number of inferred putative adaptive loci represents a small portion of the total loci; however, the genome size of *Hilaria* likely has a small 1C-value (<1000 Mbp) based on estimates from closely related graminoid species in the Cynodonteae tribe (synonym Chlorideae; e.g. species in *Bouteloua, Chloris, Cynodon* and *Muhlenbergia* – see <https://cvalues.science.kew.org>). We note that it is unknown whether all putative adaptive loci have adaptive function (or are closely linked to an adaptive locus), or if some are neutral or deleterious loci displaying allele frequency gradients resulting from, for example, genetic drift correlated with the demographic expansion of the species (Peischl et al. 2013, Gilbert et al. 2017). However, our results and interpretations herein rely upon cumulative patterns across loci, which were corroborated by independent analyses.

Finally, our blast analysis of putative adaptive loci to NCBI’s nr database revealed that 175 out of 932 unique loci are closely and significantly aligned with known graminoid sequences. Furthermore, 125 out of the 175 graminoid-related loci have known or inferred function according to the UNIPROT database (Supplementary material Appendix 1 Table A7). We reiterate that it is currently unfeasible to assess the adaptive significance of these loci according to the climatic gradients used in this study, as the functional variation in these loci, as well as how that variation would respond to climatic variation, is unknown. In addition, not finding a match between putative adaptive loci and a sequence in NCBI or UNIPROT does not mean that the locus was falsely identified. Instead, unmatched loci may not yet be characterized (especially from a closely related species), or they may be in non-coding regions that are closely linked to genomic regions under selection.

**Correlation of putative adaptive loci with climatic gradients**

The overall RDA model was significant (p < 0.001) and the proportion of variance explained by the three PC axes (i.e. the adjusted $R^2$) was 8.6%, which is not surprising under the assumption that most SNPs are neutral. Furthermore, the first two constrained axes were significant (p < 0.01 and p < 0.02, respectively), and the variance inflation factors for the predictor variables were low (1.04, 1.05 and 1.07 for PC1–3, respectively). Therefore, multicollinearity among these predictors should not influence the interpretation of the results. Most putative adaptive loci were aligned with the latitudinal climatic axis (PC2 – 51.8%), followed by the mean diurnal temperature range axis (PC3 – 41.9%); relatively few loci were aligned with elevation-related climatic gradients (PC1 – 6.1%) (Supplementary material Appendix 1 Fig. A2). LFMMs recovered a more even distribution of SNPs across the PC axes, although similar to RDA, the fewest putative adaptive loci were associated with PC1 and the most were associated with PC2 (Supplementary material Appendix 1 Table A6).

When sampling localities are visualized along the significant RDA axes, it becomes clear that their positions in relation to the climatic predictors show a correspondence with the species’ population genetic structure (Fig. 5). Each population’s sampling localities cluster together and are well-differentiated along the climate predictors; the transitional sampling localities (i.e. those containing > 25% of minor genetic axes – Fig. 1) are also intermediate in their relatedness to the climatic predictors (Fig. 5). Populations are situated from south to north along the latitudinal climate axis (PC2) – the intermediate position of the western population along this axis makes sense because, for example, this geographic area receives less monsoonal precipitation than that covered by the eastern population, even though the western population sampling localities are distributed across roughly similar southerly latitudes (i.e. the intensity of monsoonal precipitation is greatest in the southeast and diminishes to the northwest). In contrast, populations are distributed from west to east along the mean diurnal temperature range axis (PC3), reflective of the general geographic pattern of this climatic gradient (e.g. Supplementary material Appendix 1 Table A1). All populations have sampling localities distributed across a similar range of elevations (Supplementary material Appendix 1 Table A1), and there is correspondingly less differentiation among populations along the elevational
Figure 5. Redundancy analysis triplots for (a) the major RDA axes and (b) the magnification of these axes to highlight SNP loadings. In (a), the dark gray cloud of points at the center of the plot represents the SNPs, and the colored points represent the Colorado Plateau sampling localities coded by their dominant genetic cluster (colors match those used in Fig. 1, but population identity did not inform the test). Localities in which > 1 genetic cluster is evident are colored with the dominant genetic cluster at the center and the secondary genetic cluster as the outer margin (Fig. 1). In (b), candidate loci are shown as colored points, colored by the most highly correlated PC axis. SNPs not identified as candidate loci (i.e. neutral SNPs) are shown in light gray. In both panels, the blue vectors represent climate-derived PC axes (see Supplementary material Appendix 1 Table A3 for detailed correlations).
climate axis (PC1). When considering the putative adaptive loci in ordination space (Fig. 5b), those most strongly correlated with latitude load on axis 1, while loci correlated with mean diurnal temperature range load predominantly on axis 2. These results are not simply an artifact of demographic processes (e.g. migration), as LFMMs, which explicitly control for population structure, resolved similar signals of adaptation (Supplementary material Appendix 1 Table A2).

Discussion

Aspects of the southwestern North American landscape point to multiple factors that could individually influence neutral and/or adaptive variation. These include: 1) climatic shifts associated with Pleistocene glaciations, 2) steep, contemporary climatic gradients and 3) regional topographic complexity that impacted historical and contemporary dispersal. When these three dimensions are assessed simultaneously, it is difficult to develop scenarios where they do not jointly affect an organism’s neutral and adaptive variation. As such, the details of, and interaction among, these factors are important for interpreting selection and the source of adaptive variation. Here, we describe the multifaceted evidence for the interdependency of neutral and adaptive variation in *H. jamesii* and discuss implications for adaptation from standing genetic variation and restoration across the Colorado Plateau.

Codependency of neutral and adaptive variation

*Hilaria jamesii* persisted during the last glacial period in a single, large refugium (Table 1). Although the precise location of this refugium is unknown due to the lack of macro or microfossils for the species, landscape features, pollen data, paleoclimate data and macrofossils from other organisms indicate this grass would not have persisted on the Colorado Plateau or in areas to the north, east or west, due to dispersal barriers and/or climates outside of the species’ tolerance (Thompson and Anderson 2000, Ray and Adams 2001, Grayson 2006). Accordingly, these factors suggest a southern refugium, which is supported by the maximum likelihood phylogenetic reconstruction due to the nesting of a well-supported clade containing sampling localities forming the western + northern populations within a clade that includes the eastern population sampling localities (Fig. 4). Although high differentiation between *H. jamesii* and the outgroup, combined with low differentiation within *H. jamesii*, may affect the reliability of the phylogenetic reconstruction, we note a similar refugium south-east of the Colorado Plateau has been supported for *Bouteloua gracilis*, a C₄ graminoid presently co-distributed with *H. jamesii* across the Colorado Plateau (Brown and Gersmehl 1985). In addition to providing specific expectations for patterns of neutral variation resulting from the post-glacial expansion of the species, the location of its refugium and timing of its expansion have important consequences for non-neutral variation.

An *H. jamesii* refugium south of the Colorado Plateau suggests that the range of climatic conditions *H. jamesii* experienced during the last glacial period differed from the climatic conditions to which it is presently exposed. While *H. jamesii* is currently distributed across steep latitudinal climatic gradients that are strongly correlated with increasing continental climatic conditions (e.g. more northerly latitudes have more significant annual variation in temperature and little to no summer precipitation from the monsoon weather pattern), a southern refugium would likely have been subjected to southerly latitudinal climates, including persistent monsoonal precipitation, since before the LGM (Holmgren et al. 2007, Coats et al. 2008, Cole et al. 2013). As the climate warmed throughout the region after the LGM, *H. jamesii* likely tracked suitable habitat northward by traversing lower elevation habitat along the western slope of the Rocky Mountains and into southeastern Utah (similar to *B. gracilis* – Tso and Allan 2019). Once in Utah, coalescent analyses support two independent migration routes (Table 1), which resulted in the northern and western populations (Fig. 1). This migration not only increased the species’ geographic extent across the Intermountain West, but it also caused *H. jamesii* to encounter the region’s steep latitudinal climatic gradients, which are the gradients associated with *H. jamesii*’s putative adaptive variation.

Climatic gradients related to latitude and mean diurnal temperature range (Supplementary material Appendix 1 Table A3) are the main influences on *H. jamesii*’s putative adaptive alleles (Fig. 5); these gradients may strongly affect fitness and therefore drive patterns of adaptive variation. For example, precipitation in the spring (i.e. prior to May) and, in more southeasterly geographic locations, mid to late summer (i.e. monsoonal precipitation occurring from July through September) can both promote growth and flowering of *H. jamesii* (Schwinning et al. 2005). As such, individuals pre-adapted to utilize the additional summer moisture may have higher fitness compared to those that lack such adaptations because there is a higher probability that they produce at least one, if not two, sets of seeds per year. Furthermore, seeds resulting from spring flowering may themselves flower and produce seed under optimal monsoonal conditions (Massatti unpubl.). Alternatively, monsoonal precipitation may affect individuals through other fitness components, such as the regeneration niche of adapted individuals (e.g. individuals adapted to an extended growing season due to the prolonged maintenance of soil moisture; Poorter 2007). These mechanisms provide a direct link to physiologically important constraints in dryland species. However, the field-based demonstration of monsoonal adaptation is difficult from both the perspective of effectively simulating monsoonal precipitation as well as challenges associated with interpreting comparisons to control plots because of fluctuations in the yearly strength of the monsoon rains (Wertin et al. 2015), which highlights the utility of inferring putative adaptive variation using genomic data.

Additional evidence supporting the correlation of neutral and adaptive processes in generating contemporary genetic patterns is gleaned from considering the timing of *H. jamesii*’s population splits (Table 1) in conjunction with independent evidence of the regional climate dynamics following the
LGM. Based on the changing distributions of pollen and plant macrofossils for plants adapted to monsoonal precipitation, the monsoon weather pattern migrated north following the LGM (Shafer 1989, Coats et al. 2008), before again retreating southward to its contemporary configuration (Cole et al. 2013). Our demographic modeling data suggest the concomitant northward shift of *H. jamesii* with monsoonal precipitation. Under the phylogenetic hypothesis established by the maximum likelihood tree (Fig. 4), the ancestor of the northern + western populations (*N*, in model 2 of Fig. 3) diverged from the eastern population approximately 16 600 yr ago (Table 1). This ancestor would have been located north of the eastern population, given the species was expanding its range due to warming temperatures (i.e. transitioning to its present-day distribution). Interestingly, a location north of the eastern population corresponds well with the northernmost extent of the monsoons (near Canyonlands National Park in Utah – see star in Fig. 1; Shafer 1989) and with the northernmost extent of sampling localities containing a genetic component aligning with the eastern population. The northern and western populations diverged ca 11 500 yr ago, which is roughly coincident with the southward retreat of monsoonal precipitation ca 9000 yr ago (Shafer 1989). Note that the divergence time of 11 500 yr ago is based on one generation per year and may be overestimated if *H. jamesii* individuals adapted to utilize monsoonal precipitation (as were the individuals that migrated up from the south) are able to produce offspring that could reproduce within the same year (see above). The absence of monsoonal precipitation for these northern *H. jamesii* individuals must have created a strong selective regime. Surviving under these climatic conditions (e.g. with no reliable monsoonal precipitation) would have precipitated evolutionary changes and as a consequence opened up a vast amount of habitat, paving the way for *H. jamesii*’s expansion to the north and west (see distributions of the northern and western populations in Fig. 1). In support of this idea, differentiation between the northern and western populations is neither geographical (i.e. restricted to isolation by distance from the location of their shared ancestor) nor strictly related to latitude, as these populations display putative selectively-driven differentiation with respect to mean diurnal temperature range (i.e. PC3; Fig. 5b).

**Standing genetic variation as a source for adaptive variation**

During the climatic oscillations of the Pleistocene, species across the Intermountain West experienced cycles of extinction, confinement to refugia and expansion. Our results indicate *H. jamesii*’s most recent migration from its refugium throughout its contemporary distribution unfolded since the LGM (Table 1). However, the species’ exposure to at least some latitudinal climate gradients occurred even more recently due to the timing of the northern expansion, and subsequent retreat, of the summer monsoons (as recently as 9000 yr ago – see above). Similarly, differentiation across the mean diurnal temperature range climatic gradient occurred as the western and northern populations migrated across the landscape after they diverged from their common ancestor (< 11 000 yr ago, Table 1). Given the complex signatures of environmental adaptation and the recency of migration, standing genetic variation may have been the source of the observed gradients in allele frequencies. Standing genetic variation is supported because: 1) adaptation can proceed faster from standing genetic variation compared to new mutations because beneficial alleles are present and at higher frequencies within the population, 2) beneficial alleles from standing genetic variation may have been pre-tested by selection in past environments (in this case, previous interglacial periods) and 3) alleles are present in both the eastern population (i.e. close to the putative glacial refugium for the species) and the northern and western population sampling localities (see data files – Massatti 2020) (Hermisson and Pennings 2005, Barrett and Schluter 2008). Standing genetic variation may not only be important in population establishment associated with the repeated expansion and contraction of populations, but it also places constraints on the current extent of *H. jamesii*’s distribution. It is also possible that some putative adaptive loci could have arisen more recently (e.g. since the LGM). However, the complexity of adaptation (along multiple environmental axes and across a number of weakly and strongly selected loci) and the regional glacial history (e.g. repeated glacial/interglacial cycles during the Pleistocene – Richmond and Fullerton 1986) make it less likely that new mutations (as opposed to pre-existing mutations) were the main source of the putative adaptive loci identified here. A comparative phylogeographic/adaptive framework would be valuable to determine if species with similar life history characteristics and distributions share similar histories and allele frequency gradients suggestive of adaptation from standing genetic variation (Papadopoulou and Knowles 2016).

**Restoration relevance on the Colorado Plateau**

The consideration of neutral and putative adaptive genetic variation has direct relevance to management and restoration on the Colorado Plateau. Specifically, there is interest in developing *H. jamesii* plant materials for restoration purposes due to its trampling tolerance, ability to prevent erosion in semi-desert sites, and palatability for cattle and horses (St. John et al. 2012). Our data show that strategies to use *H. jamesii* restoration materials should not only depend on a specific combination of local climatic conditions, but also on the history of local populations. Specifically, the data highlight that plants are sensitive to climatic gradients closely aligned with latitude, but that those associated with elevation may be less important (contrary to guidance provided for other restoration species – see Shryock et al. 2017, Massatti et al. 2018a) (Fig. 5a). Likewise, when choosing among seed sources to be used in the development of regionally-specific restoration materials (Bucharova et al. 2019), even geographically proximate populations may differ significantly with respect to population history as a result of the interaction between topography and migration dynamics from glacial
refugia. We note that, while molecular genetics provides an estimate of putative adaptation, results should be confirmed with field-based experiments (e.g. common gardens or reciprocal transplants, as summarized in Kilkenny 2015). Nonetheless, our results may be used to develop restoration materials that consider both adaptive genetic variation and genetic differentiation (i.e. population history) (McKay et al. 2005, Massatti et al. 2018b), such that the materials support natural patterns of genetic variation while maintaining some type of ‘local’ adaptation. Furthermore, strategies to maintain genetic diversity may facilitate the maintenance of adaptive variation important to past, and likely future, environmental changes.

Data availability statement

Data generated during this study are available from the USGS ScienceBase-Catalog: <https://doi.org/10.5066/P9CNFWOX> (Massatti 2020).

Acknowledgements – We would like to thank Ernie Nelson at the Rocky Mountain Herbarium and Michael Piep at the Intermountain Herbarium for helping acquire the herbarium specimens used in this study. Gery Allan, Lela Andrews and Erica Sukovich at Northern Arizona University’s Environmental Genetics and Genomics Laboratory provided expert support on laboratory procedures. Colleagues at the Southwest Biological Science Center, especially John Bradford, engaged in many useful discussions regarding this work. This manuscript has been greatly improved by comments from Daniel Winkler, Andrew Eckert and one anonymous reviewer. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Geological Survey.

Funding – We are grateful for funding provided by the Bureau of Land Management’s Colorado Plateau Native Plant Program.

Author contributions – RM designed this research, performed the analyses and analyzed the results, with contributions from L.L.K. RM wrote the original manuscript and both authors contributed to the final manuscript.

References

Aronesty, E. 2011. ea-utils: command-line tools for processing biological sequencing data. – <https://expressionanalysis.github.io/ea-utils/?>.

Barrett, R. D. and Schluter, D. 2008. Adaptation from standing genetic variation. – Trends Ecol. Evol. 23: 38–44.

Bemmels, J. B. et al. 2019. Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. – Proc. Natl Acad. Sci. USA 116: 8431–8436.

Brown, D. A. and Gersmehl, P. J. 1985. Migration models for grasses in the American midcontinent. – Ann. Assoc. Am. Geogr. 75: 383–394.

Bucharova, A. et al. 2019. Mix and march: regional admixture prov- enancing strikes a balance among different seed-sourcing strategies for ecological restoration. – Conserv. Genet. 20: 7–17.

Capblancq, T. et al. 2018. Evaluation of redundancy analysis to identify signatures of local adaptation. – Mol. Ecol. Resour. 18: 1223–1233.

Catchen, J. et al. 2013. Stacks: an analysis tool set for population genomics. – Mol. Ecol. 22: 3124–3140.

Coats, L. L. et al. 2008. 50,000 years of vegetation and climate history on the Colorado Plateau, Utah and Arizona, USA. – Quat. Res. 70: 322–338.

Cole, K. L. 1990. Reconstruction of past desert vegetation along the Colorado River using packrat middens. – Palaeoecogr. Palaeoclimatol. Palaeoecol. 76: 349–366.

Cole, K. L. et al. 2008. Modeling future plant distributions on the Colorado Plateau: an example using Pinus edulis. – In: van Riper III, C. and Sogge, M. K. (eds), The Colorado Plateau III. Integrating research and resources management for effective conservation. Univ. of Arizona Press, pp. 319–330.

Cole, K. L. et al. 2013. The biogeographic histories of Pinus edulis and Pinus monophylla over the last 50,000 years. – Quat. Int. 310: 96–110.

Excoffier, L. et al. 2013. Robust demographic inference from genomic and SNP data. – PLoS Genet. 9: e1003905.

Falush, D. et al. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. – Genetics 164: 1567–1587.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315.

Fitzpatrick, M. C. and Hargrove, W. W. 2009. The projection of species distribution models and the problem of non-analog climate. – Biodivers. Conserv. 18: 2255.

Forester, B. R. et al. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. – Mol. Ecol. 27: 2215–2233.

Frantz, A. C. et al. 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.

Frichot, E. and François, O. 2015. LEA: an R package for landscape and ecological association studies. – Methods Ecol. Evol. 6: 925–929.

Frichot, E. et al. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. – Mol. Biol. Evol. 30: 1687–1699.

Gilbert, K. J. et al. 2017. Local adaptation interacts with expansion load during range expansion: maladaptation reduces expansion load. – Am. Nat. 4: 368–380.

Grayson, D. K. 2006. The late quaternary biogeographic histories of some Great Basin mammals (western USA). – Quat. Sci. Rev. 25: 2964–2991.

Guindon, S. et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. – Syst. Biol. 59: 307–321.

Gutenkunst, R. N. et al. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. – PLoS Genet. 5: e1000695.

Hermisson, J. and Pennings, P. S. 2005. Soft sweeps: molecular genetic evidence for adaptive evolution in nature. – Genetica 1223–1233.

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

Holmgren, C. A. et al. 2007. Inferences about winter temperatures and summer rains from the late Quaternary record of C4 perennial grasses and C3 desert shrubs in the northern Chihuahuan Desert. – J. Quat. Sci. 22: 141–161.
Appendix 1.

Supplementary material (available online as Appendix ecog-04840 at [www.ecography.org/appendix/ecog-04840]).

Appendix 1.