Glaciation history and geographic barriers shape genetic diversity of Lewis flax across North America

Peter A. Innes*[1], Zach Marcus[1], Brent S. Hulke[2], and Nolan C. Kane[1]

[1]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado, USA
[2]USDA-ARS Sunflower Improvement Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, North Dakota, USA

Abstract

Factors shaping genetic variation at the continental scale include geographic barriers, historical disturbances, and genetic constraints like mating system. Here, we examine how such factors combine to influence population diversification in *Linum lewisii* (Lewis flax, prairie flax), a perennial forb native across much of North America. We used whole genome sequencing, spatial analyses, and demographic inference to investigate how mountains and recent glaciation have shaped the evolution of Lewis flax and to test expectations about the evolution of mating system during range expansions. Our dataset comprises 168 genomes sampled from public seed repositories, herbarium collections, and new field collections from Alaska and the Hudson Bay to Mexico, and from California to the Great Plains. We found that Lewis flax consists of four main genetic clusters that are not reflected in currently accepted intraspecific taxonomy. Demographic inference suggests this structuring occurred within the last 100,000 years, during the Last Glacial Period, with the Rocky Mountains acting as a dispersal barrier. Genomic data also suggest Lewis flax is predominantly outcrossing, despite its well-known mating-system shift from self-incompatible heterostyly to self-compatible homostyly. Furthermore, despite its initial hypothesized colonization of North America via Beringia, we found no evidence of a northern refugium. Our results instead suggest that current northern populations of Lewis flax derive from a range expansion following the Last Glacial Maximum. This expansion involved a reduction in nucleotide diversity, increased inbreeding, and increased linkage disequilibrium across the genome, which might reflect a shift towards increased selfing in northern populations, potentially facilitating colonization in areas with pollinator limitation. The mixed mating system of Lewis flax thus combines with historical processes and geographical barriers to affect genetic variation in ways that would be different in related obligate outcrossers. By revealing distinct lineages and centers of genetic diversity, this range-wide population genomic analysis also provides foundational knowledge for its use in ecological restoration and its ongoing neodomestication as a perennial oilseed food crop.

Keywords

Population genomics, demographic inference, phylogeography, whole genome sequencing, *Linum lewisii*, museum specimens

Introduction

Understanding the forces that structure genetic variation within species is an essential goal of population genetics that unifies both basic and applied research. The study of population genetic diversity helps trace the evolutionary history of a species; it also provides foundational knowledge for conservation, restoration, and agriculture. Species management and breeding decisions are often based on some knowledge of population genetic structure, which can impact the compatibility of a cross or reveal distinct lineages of conservation concern. Intraspecific genetic variation can also serve as a reservoir for crop improvement and wild species' adaptation to environmental change or biotic threats.

Quaternary glacial cycles are a major factor that have shaped global patterns of population genetic diversity through profound alteration of climate, continental landscapes, and demographic trajectories (Fonseca et al. 2023; Hewitt 2004). Most recently, during the Last Glacial Period (115–11.7 kya), many Northern Hemisphere temperate plant and animal species experienced southward range contractions into climatically suitable refugia, followed by postglacial northward expansion. This movement pattern generates some predictions regarding population genetic diversity and divergence (Provan et al. 2008). Because genetic variation tends to accumulate in populations that persist through time, while range expansions typically involve genetic bottlenecks, there is an expectation of higher diversity in southern refugial populations relative to northern populations (Comes et al. 1998). Furthermore, if a species inhabited multiple distinct refugia, genetic divergence between the separated populations is likely to occur due to genetic drift, and genetic structuring may occur along the expansion path as well (Hewitt 1996).

*Corresponding author: peter.innes@colorado.edu

Genetic impacts of Pleistocene glaciation are widely documented in North American and European plants. Many such studies have illustrated southern refugia and northern expansions, although there are also cases of alternative northern refugia (Beatty et al. 2010; Provan et al. 2008). Eastern North America has been a primary focus geographically speaking (Lumibao et al. 2017; Lyman et al. 2022; Soltis, Morris, et al. 2006). In western North America, plant phylogeography research has typically focused on trees (Bagley et al. 2020; Callahan et al. 2013; Evans et al. 2015; Jaramillo-Correa et al. 2009); or species with more limited regional distributions, especially in the Pacific Northwest (Shafer et al. 2010; Soltis, Gitzendanner, et al. 1997; Stone et al. 2021; Tomimatsu et al. 2009). By comparison, our knowledge of population genetic patterns in herbaceous species more widely distributed across western North America remains limited.

Lewis flax (*Linum lewisii* Pursh.) is a perennial, blue-flowered flax with a distribution spanning western North America from Alaska to northern Mexico, with some isolated populations around Hudson Bay in Canada (Fig. S1; Mosquin 1971). This broad distribution makes Lewis flax a useful system for investigating continental-scale phylogeographic patterns. There are two other blue flax species native to North America: *L. pratense* (Norton) Small, which has a smaller range, overlapping with the southeastern extent of *L. lewisii*, and the lesser known *L. rzedowskii* Arreguín, which has a very restricted endemic range in central Mexico and is critically endangered (González-Velasco et al. 2022). All other close relatives of Lewis flax inhabit Eurasia. The ancestor of Lewis flax is thought to have colonized North America via Beringia during its split from Eurasian populations 3.3–3.8 Mya (Harris 1968; McDill et al. 2009). A notable evolutionary shift that coincided with the expansion of blue flax into North America was the loss of self-incompatibility, which, in *Linum*, is tied to loss of heterostyly polymorphism, the likely ancestral state of the genus involving two genetically determined mating types with distinct floral morphology (Maguilla et al. 2021; McDill et al. 2009). Lewis flax is homostylous and self-compatible, although it typically relies on insect pollinators to set seed (Kearns et al. 1994), apparently due to spatial separation of its anthers and stigmas, which is known as herkogamy (Webb et al. 1986). Still, its propensity for selfing versus outcrossing is unclear, and variation in the degree of herkogamy occurs in some parts of its range (Mosquin 1971). This raises the possibility that selfing rates differ across the range, associated with herkogamy variation or differences in pollinator associations.

Lewis flax has spread across a considerable distance and variety of habitats on the North American continent. This broad adaptability has made it a popular species included in seed mixes for restoration of native plant communities, especially on disturbed lands in the Intermountain West (Kitchen 1995; Leger et al. 2015; Winslow et al. 2009). To this end, the United States Department of Agriculture (USDA) selected and released the "Maple Grove" pre-variety germplasm in 2003, which was originally collected from Millard County, Utah in the southwestern United States. Maple Grove Lewis flax seed is commercially grown and widely available from seed companies and garden stores, and is one of only a few native forb species in the western United States for which this is the case (Jones 2019). There is also a recent effort to domesticate Lewis flax as a perennial oil seed crop in part due to its seeds being rich in healthy α-linolenic acid (Innes, Gossweiler, et al. 2022; Pull et al. 2023). For both of these purposes, understanding the genetic variation within Lewis flax is useful. For instance, Maple Grove is widely planted for restoration but might not be suitable in particular regions if genetic differentiation has involved local adaptation, or if maintaining genetically distinct local populations is a priority. Likewise, this knowledge helps inform choices of which accessions to use for breeding and which regions to target for additional collections (Raduski et al. 2021).

Here we investigate population genetic structure and diversity in Lewis flax across its range in western North America, from central Alaska south to Mexico, and from the Great Plains to nearly the Pacific Coast. We infer timing of population expansions, contractions, and splits over the past million years and examine whether demographic changes follow North America's history of glaciation. We investigate the major genetic lineages of Lewis flax and how these coincide with geographical features that delineate species or subspecies boundaries in other taxa. We also seek to better characterize inbreeding patterns across the range of the species, potentially giving insights into forces affecting the evolution of mating systems. We consider how our findings might be applied to the use of Lewis flax in restoration and agricultural settings.

Methods

Plant material

Our initial sampling combined 172 Lewis flax genotypes from seed banks (USDA National Plant Germplasm System and Plant Gene Resources of Canada, $n = 34$), herbaria ($n = 66$), personal seed collections ($n = 70$), and the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA, $n = 2$) (Table S1). Sampling was conducted opportunistically due to limited resources and travel prohibitions during the COVID-19 pandemic. Most geographic locations are represented by a single genotype (Figure S2), with the main exception being that we had dense sampling in Boulder County, Colorado ($n = 32$ across 13 localities).

Seed bank accessions and personal collections were grown from seed either in a greenhouse setting at University of Colorado, Boulder or in field settings at North Dakota State University and University of Minnesota, Twin Cities. We sampled herbarium specimens from the University of Colorado Herbarium (COLO), the Rocky Mountain Herbarium at University of Wyoming (RM), the University of Manitoba Vascular Plant Herbarium (WIN), the Arizona–Sonora Desert Museum (ASDM), and the Desert Botanical Garden (DES). Acronyms follow the Index Herbariorum database. We selected specimens that filled geographic gaps among our seed collections, were recently collected, appeared in good condition, and had enough leaf material to minimize the impact of destructive sampling (we took leaf tissue from fragment packs associated with the specimen whenever possible). Our study aims did not include an explicit temporal component and thus we prioritized specimens we thought could yield high quality DNA without using intensive ancient DNA protocols.

Of the two *L. lewisii* whole genome shotgun (WGS) samples retrieved from the SRA, Llw01 was from a genotype of unknown provenance (You et al. 2018); the other, TROM V 135817, was from an Arctic University Museum of Norway herbarium specimen collected in Alaska (Alsos et al. 2020). Eight samples lacked geographic coordinates or detailed locality information (Table S1) and thus were excluded from spatial analyses and map visualizations.

DNA extraction, library preparation, and sequencing

Extractions and library preparations were performed separately for non-herbarium samples versus herbarium samples. For the former, we extracted DNA from fresh, lyophilized, or silica-dried leaf tissue using QIAGEN DNeasy Plant Mini kits. Libraries for Illumina whole genome sequencing were prepared with the 96-Plex kit from TWIST Biosciences following the manufacturer's protocol and using an even ratio of the Low and High GC Primer A. DNA amounts across samples were equalized for library input by dilution based on Qubit fluorometer readings. Samples were spread across two TWIST 96-Plex libraries. Library quality control and sequencing was performed by Novogene. The pooled Twist libraries were each sequenced on a single lane of a Nova-Seq 6000 S4 flow cell using 150 bp paired-end reads.

Herbarium sample DNA extractions were performed several weeks to months after the non-herbarium sample extractions but in the same lab space at University of Colorado, Boulder. Equipment and surfaces were thoroughly cleaned with 10% bleach in order to safeguard against contamination; we also treated plastic consumables with UV light prior to extractions. We used a modified QIAGEN DNeasy Plant Mini Kit extraction protocol optimized for herbarium specimen leaf tissue (Marinček et al. 2022). We assessed herbarium DNA quantity and quality with flourometer (Qubit), spectrophotometer (Nanodrop), and gel electrophoresis methods. Herbarium DNA samples passing this quality control step had concentrations ranging from 2–34 ng/µL (mean 13 ng/µL), a 260/280 ratio ranging from 1.6–2.8, and fragments mostly 500 bp or larger based on gel results. A preliminary attempt to include herbarium DNA samples in a pooled Twist 96-Plex library alongside non-herbarium samples resulted in very low amounts of sequence data for the herbarium samples, so we took a different approach. Illumina whole genome libraries were prepared for herbarium DNA samples of sufficient quality using the Illumina DNA Prep workflow (formerly Nextera DNA Flex), which includes an initial fragmentation step. The libraries were sequenced on a single lane of an Illumina NovaSeq X Plus 10B flow cell using 150 bp paired-end reads. Herbarium WGS library preparation and sequencing were performed by SNPsaurus.

Among the 104 non-herbarium samples that were sequenced, we excluded four of five samples that were grown from seed from the same mother plant in Boulder County, Colorado, due to their sibling relationship (Table S1). We also excluded eight of 66 herbarium samples from analysis of the nuclear genome due to low depth (four had depth less than 0.25x and another four had depth less than 0.75x).

DNA sequence processing, alignment, variant calling, and filtering

Raw sequence data were quality trimmed using fastp with default settings (Chen et al. 2018) and aligned to the chromosome-scale *L. lewisii* Maple Grove reference genome [GenBank: GCA_034768395.1 (Innes, Smart, et al. 2023)] with bwa-mem2 (Vasimuddin et al. 2019). We marked duplicate reads with Picard MarkDuplicates and used mosdepth (Pedersen et al. 2018) to calculate average sequencing depth for each sample.

We called variants with bcftools mpileup and bcftools call (Danecek et al. 2021). We used GNU parallel (Tange 2023) to process the nine chromosomes of the Lewis flax genome in parallel and generate VCF files for each chromosome that contained both variant and invariant sites. We excluded indels while filtering SNPs and invariant sites with the following site-level thresholds using bcftools: we kept sites with minimum average mapping quality (MQ) 40, maximum missingness (F_MISSING) 25%, and mean depth [MEAN(FMT/DP)] between 2 and 15. For SNPs we additionally kept only bi-allelic sites with quality (QUAL) of at least 40. The maximum mean depth threshold (15) was slightly less than three times the median of per-sample average depth values and was meant as an initial filter for spurious calls in repetitive or highly paralogous regions. We used two additional filters to limit potential biases resulting from paralogy and repetitiveness in the Lewis flax genome, approximately 70% of which is repetitive (Innes, Smart, et al. 2023). First, using mosdepth, we calculated depth averaged within 500 bp windows across the genome; for each sample, windows with depth greater than 1.5x the within-sample median were defined as high depth regions; we then defined a set of shared high depth regions as those high depth regions identified in at least half of all samples. Second, we used GenMap (Pockrandt et al. 2020) to calculate mappability across the Lewis flax reference genome with a k-mer length of 150 bp and up to three errors (parameters: -K 150 -E 3); low mappability regions were defined as those matching three or more places in the genome. We merged these regions with the shared high depth regions using bedtools. The resulting sites, comprising 262,564,216 bp, around 43% of the length of the nine reference genome chromosomes, were further excluded from our VCF. Filtered variant and invariant sites were then combined back into a filtered "all-sites" VCF containing 123,583,335 invariant sites and 16,218,244 biallelic SNPs, of which 5,417,581 were singletons (i.e. had a minor allele count of 1).

Integrity of herbarium specimen DNA

We were cautious about including herbarium specimens alongside contemporary samples in our analyses, due to potential batch effects that could arise from DNA degradation or the use of different library preparation kits. We used mapDamage2 (Jónsson et al. 2013) to quantify damage patterns in our herbarium DNA and examined the frequency of cytosine deamination (C to T transitions) at the ends of reads, a hallmark of post-mortem degradation. We also looked for batch effects in our subsequent analyses of population structure and diversity simply by comparing technical versus geographic groupings.

Population structure

For analyses of population structure, we used a minimum minor allele frequency threshold of 0.025 and also pruned SNPs based on linkage disequilibrium (LD) using PLINK (v1.90b7.1 Chang et al. 2015) with a window size of 10 kb, a step size of 1 SNP, and a maximum r^2 threshold of 0.5 (plink –indep-pairwise 10kb 1 0.5). This resulted in 1,349,461 SNPs. We performed model-based ancestry inference with ADMIXTURE (Alexander et al. 2009) for values of k ranging from 1 to 10 and using 5-fold cross validation to determine the most likely value of k. Principal component analysis (PCA) was performed with EMU (Meisner et al. 2021), to better account for non-random missingness stemming from variation in sequencing depth among samples. Results from both ADMIXTURE and PCA suggested the *L. lewisii* samples included here could be assigned to four main clusters, which we hereafter refer to as the Northern (N), Central (C), Southwestern (SW) and Southeastern (SE) populations, based on the ancestry coefficient with relative majority.

We investigated the impact of spatial bias in our sampling (particularly the dense sampling in Boulder County, Colorado) by excluding 20 of the 28 Boulder County samples in the full cohort and again running ADMIXTURE and PCA as described above. Prior to analysis we re-filtered the sub-sampled VCF with the same missingness, minor allele frequency, and LD thresholds as described previously. We refer to this dataset as the "reduced cohort".

We used FEEMS (Marcus et al. 2021) to visualize patterns of Lewis flax effective migration and gain additional insight into its spatial genetic structure. For this analysis we excluded nine samples without geographic coordinates as well as the sample from West Virginia (Ames 34946) and subsequently applied the same minor allele frequency filter and LD pruning as before. Missing genotypes were imputed with Beagle (v5.4 (Browning et al. 2018)) prior to LD pruning. Migration was estimated across a spatial grid of triangular cells, each with an area of approximately 6,200 square kilometers, which we cropped to the rough outline of our sample area. We used leave-one-out cross-validation to select the best value of the tuning parameter, λ, following the FEEMS documentation. In a second FEEMS run, we excluded the 13 samples in the formerly glaciated portion of North America, in order to gain better resolution of migration patterns for the core portion of the Lewis flax range.

Genetic diversity, differentiation, and inbreeding

We estimated pairwise nucleotide diversity (π) of each population in 10 kb non-overlapping windows using pixy (v1.2.10.beta2 Korunes et al. 2021) and the filtered all-sites VCF without singleton SNPs. Genome-wide average π for each population was calculated as the sum of counts of differences (across all 10 kb windows) divided by the sum of counts of comparisons, as recommended in the documentation for pixy. Separately, we generated continuous maps of π and observed heterozygosity using wingen (cit) and its rarefaction approach to account for differences in sampling effort across the landscape. Genome-wide average F_{ST} for each population pair was calculated across the same SNPs as for estimation of π, using the function average_hudson_fst() from the python library scikit-allel (v1.3.7 Miles et al. 2021).

We used PLINK to estimate the inbreeding coefficient (F) for each sample, based on the same SNPs used in the diversity and F_{ST} analyses, however, we subset our VCF by genetic cluster in order to limit upward bias due to the Wahlund effect (Hemstrom et al. 2024). To look for evidence of increased selfing at northern latitudes, we fit a basic linear model for $F \sim latitude$.

LD decay curves were generated with PopLDdecay (Zhang et al. 2019), using the filtered SNPs and a minimum minor allele frequency threshold of 0.025.

Demographic inference

We inferred population size histories and divergence times of the four genetic clusters with SMC++ (Terhorst et al. 2017). We limited this analysis to samples with at least 5x depth and with minimal admixture ($q \geq 0.9$) because SMC++ assumes that gene flow does not occur after a split. We used SNPs as filtered above for genetic diversity analyses but without a minor allele frequency threshold so as not to alter the site frequency spectrum. We masked the same high depth and low mappability sites as excluded during initial variant filtration. We converted variant data from VCF to SMC++ format using the command smc++ vcf2smc and iterated over five "distinguished individuals" from each population to form a composite likelihood, which we subsequently used for estimation of population size histories (command: smc++ estimate) and split times (command: smc++ split). We used the per-generation mutation rate from Arabidopsis [7e-9 (Ossowski et al. 2010)] and a generation time of two years based on personal observations that Lewis flax typically does not flower until its second growing season. We set beginning and ending time points to 1e3 and 1e6 years ago, respectively. We accounted for uncertainty in population size histories and split times by resampling the input genomic data of each distinguished individual in 5 Mb blocks for 10 bootstrap replicates, following (Coe et al. 2023).

Chloroplast phylogenetics

We assembled a complete chloroplast genome for Maple Grove from short-read Omni-C data originally used for reference genome scaffolding (Innes, Smart, et al. 2023) but which we found to have sufficient plastid genome coverage for assembly. A draft assembly was performed with SPAdes (v3.11.1 Prjibelski et al. 2020) and four resulting plastid scaffolds were manually joined and circularized. The chloroplast genome was included with the nuclear reference genome for alignments described previously. We included *Linum perenne* and *Linum leonii* as outgroups using publicly available whole genome shotgun data [GenBank: PRJNA262458 (You et al. 2018)]. Variants in the chloroplast genome were called using bcftools with ploidy set to 1. We retained only bi-allelic SNPs with QUAL scores of 100 or greater; we also excluded sites within the second copy of the inverted repeat. SNPs in VCF format were converted to phylip format with vcf2phylip.py

(Ortiz 2019). We used iqtree2 for maximum-likelihood phylogenetic inference with ModelFinder and a partition model delineating SNPs in the long single copy, inverted repeat, and short single copy regions (Chernomor et al. 2016; Kalyaanamoorthy et al. 2017; Minh et al. 2020). The model designation was given as -m MFP+ASC+MERGE to account for SNP data ascertainment bias and to allow for potential merging of partitions. Node support was determined with 1000 bootstrap replicates (Hoang et al. 2018). The resulting tree was visualized with ggtree (Yu et al. 2017), and nodes with less than 75 bootstrap support were collapsed.

Results

Successful whole-genome sequencing of herbarium DNA

The 160 samples included in the nuclear genome analysis had an overall mean and median sequencing depth of 6.8x and 5.6x, respectively. Herbarium samples (Illumina libraries) ranged from 0.75x to 24x depth, with mean 8.5x and median 5.9x depth. The non-herbarium samples (Twist libraries) ranged from 1.3x to 12x, with mean 5.8x and median 5.6x depth. Percentage of reads mapping to the Maple Grove reference genome was 95%–99% in the non-herbarium samples and 97%–99% among the herbarium samples, suggesting the relative level of endogenous DNA in our herbarium samples was high and contamination was not an issue.

Signs of degradation among the herbarium DNA samples were minimal. The frequency of C to T transitions at the 5' end of reads was not elevated; nor did it significantly correlate with the age of the specimen (Fig. S3). Sequencing depth was also uncorrelated with specimen age (Fig. S3). We did observe a weak association between herbarium specimen age and DNA library insert size (Fig. S3), but overall, these results support our decision to not treat the herbarium samples separately in our bioinformatic pipelines. The integrity of the herbarium DNA is believable considering we purposefully chose specimens appearing in good condition and which were mostly collected in the past four decades.

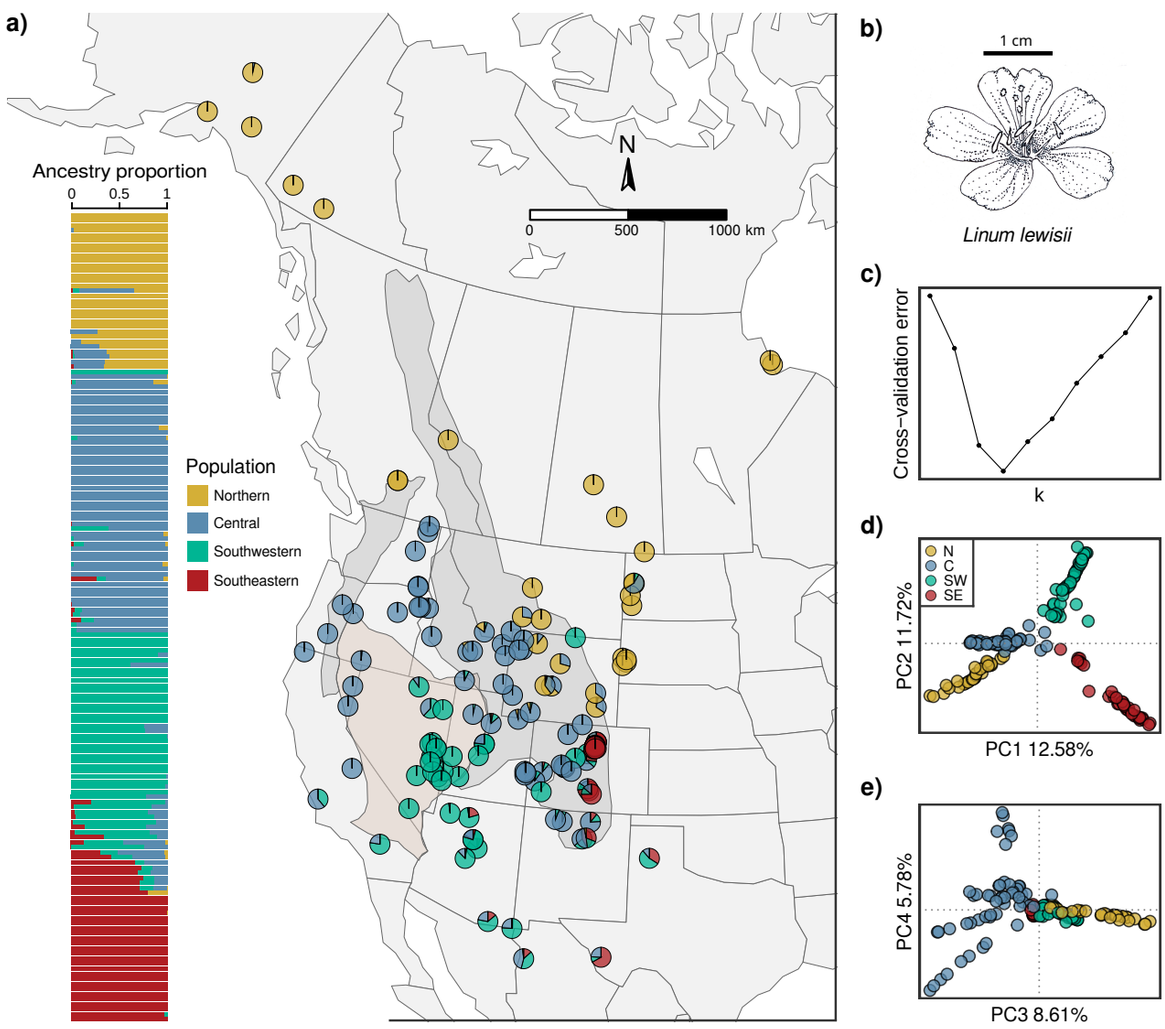

Figure 1: Population genetic structure of Lewis flax across North America inferred from 1,349,461 nuclear SNPs. **a)** ADMIXTURE results and geographic context for $k = 4$. Pie charts represent ancestry coefficients of individual samples. The darker gray shading represents approximate boundaries of the Rocky Mountains and Cascade Range, and the beige shading is the Great Basin, obtained from Natural Earth public map data. For visualization purposes, the Southwestern-type accession Ames 34946 from West Virginia is not shown. **b)** The illustration shows a typical flower of *Linum lewisii*, by Karen Jepson-Innes. **c)** Cross-validation errors of ADMIXTURE runs for $k = 1$ through 10. **d–e)** Principal component analysis of SNP variation. Points are individual samples and are colored based on the relative majority ancestry coefficient at $k = 4$

.

Lewis flax population structure corresponds to four geographic regions

We found the strongest support for $k = 4$ ancestral populations (Figs. 1), which generally correspond to the northern, central, southwestern, and southeastern parts of its range. The Northern population comprised Alaskan and Canadian samples (including the two *L. lewisii* var. *lepagei* samples from Hudson Bay) and reached south—mostly east of the Continental Divide of the Rocky Mountains—into Montana and the Dakotas, as well as Wyoming (Fig. 1). The Central population had the largest number of samples and tracked the Rocky Mountains—primarily west of the Continental Divide—as well as the Cascade Range in Oregon and into California (Fig. 1).

The Southwestern population mainly consisted of samples from the Great Basin in Nevada and Utah, as well as samples from Arizona. Furthermore, the sample from the San Bernardino Mountains in southern California (COLO 539427) and the sample from Sonora, Mexico (ASDM 3338) both had predominantly Southwestern ancestry, while the sample from Chihuahua, Mexico (RM 55736) was more admixed, with a relative majority of Central ancestry. The Maple Grove genotype (Ames 27614), which was used for the reference genome assembly (Innes, Smart, et al. 2023), had an estimated 99.8% Southwestern ancestry. A few samples with 100% Southwestern ancestry were located far outside the core Southwestern region. These included the sample from a naturalized population in West Virginia (Ames 34946; Table S1, Fig. S2), as well as Ames 34266 (northern Wyoming), and three samples from central Colorado (Guanella Pass High, Homestake Reservoir, and Dow Mountain).

The Southeastern population appeared mainly in the Southern Rockies on the eastern side of the Continental Divide. This population mostly included samples from around Boulder County, Colorado, where our sampling was particularly dense, though admixed samples with varying amounts of Southeastern ancestry occur in New Mexico and Texas and, to a lesser extent, Arizona and northern Mexico. When we thinned our sampling in Boulder County from 20 down to 8 samples, several plants were estimated to have a larger amount of Southeastern ancestry (Fig. S4). For instance, the sample from southwestern Texas (COLO 90125) showed 67% SE ancestry in the analysis of the full cohort but 100% SE ancestry in the context of the reduced cohort (Figs. 1, S4).

Further regional population substructure is apparent in the Central population at $k = 5$ and $k = 6$ (Figs. S5, S6). At $k = 7$, the Southwestern population is divided latitudinally (Figs. S5, S6). Smaller geographic clusters in Colorado, Wyoming, and New Mexico appeared in $k = 8$ through $k = 10$ but are difficult to interpret given the low support according to cross-validation error scores (Figs. S5, S6). It is possible that additional sampling, especially towards range margins where our sampling tended to be more sparse, may reveal additional population structure not discovered here.

Principal component analysis of SNP variation also indicated four clusters. The first axis (12.58% variation explained) separated the Northern and Central populations from the Southeastern and Southwestern populations and showed a significant correlation with latitude (Kendall's $\tau = -0.41$; $p < 0.001$) (Fig. S7). The second axis (11.72% variation explained) captured additional structure among all four populations (Fig. 1) and was correlated with longitude (Fig. S7). Visualization of PC3 and PC4 revealed more fine-scale population structure within the Central population (Fig. 1), which is consistent with ADMIXTURE results for $k = 5$ and $k = 6$ (Figs. S5, S6).

We examined sequencing batch and variation in sequencing depth among samples as potential sources of technical bias in our inferences about population structure. We did not observe obvious batch effects between herbarium and non-herbarium samples in the PCA or ADMIXTURE results. Samples clustered by geography rather than by batch (Fig. S8a–b), and there was no correlation between sequencing depth and ADMIXTURE proportions (Fig. S8c). We also did not observe significant effects of sequencing depth on PC1 or PC2, although PC3 and PC4 did show moderate correlations with depth (Fig. S8d). Overall these findings support the interpretation that the clustering patterns are predominantly biological rather than stemming from a technical artifact.

The chloroplast phylogeny also showed clear geographic structuring (Fig. S10). Of the major distinct populations identified in the ADMIXTURE and PCA nuclear SNP analyses, the Northern and Southeastern populations, in particular, are represented in well-supported cpDNA clades. The Southwestern population is found in two cpDNA clades, but these clades also have substantial membership from the Central population. The Central population, as assigned by nuclear SNPs, is found in every major cpDNA lineage, such as the clade containing samples from California and Oregon with only Central membership, which stems from a basal node of the tree. Samples assigned to the Northern population were the most cohesive in their phylogenetic clustering with the notable exception of Mystic 2, from the Black Hills of South Dakota, which fell within the main Southeastern chloroplast clade despite having 100% Northern nuclear ancestry (the three other samples from the Black Hills were in the Northern cpDNA clade). These patterns in the chloroplast, together, are thus largely consistent with the major nuclear lineages, but also suggest localized hybridization and chloroplast capture.

Effective migration surfaces indicate elevated migration in the north

Results from FEEMS further emphasized a latitudinal trend in genetic structure. In the best-fitting model, effective migration was higher than average across the northern half of our sampling area (Fig. 2a). This area of higher effective migration roughly coincided with the distribution of the Northern population (Fig. 1). The southern half of our sampling range generally showed lower than average effective migration (Fig. 2a). FEEMS analysis of only the samples within this southern, unglaciated region revealed that the Southern Rockies had especially low average effective migration rates (Fig. 2b). Visualization of effective migration surfaces across varying levels of the penalization parameter λ further highlighted this pattern (Fig. S9): The Southern Rockies were notably the first feature to appear in the regularization path at high λ (Fig. S9a, indicating strong statistical support for this feature as a migration barrier (Marcus et al. 2021).

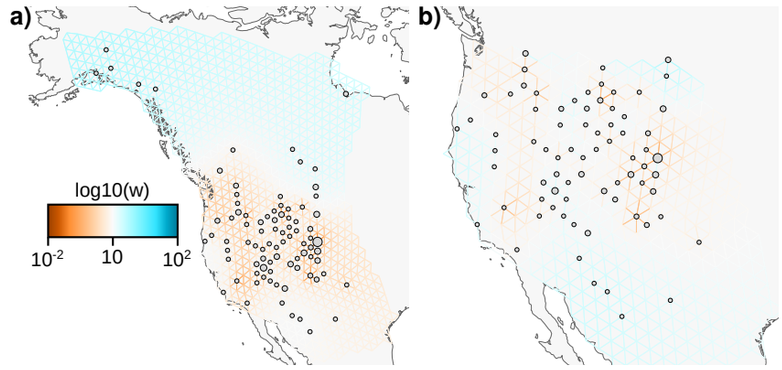

Figure 2: Lewis flax effective migration surfaces estimated with FEEMS for **a)** samples across its full range and **b)** its unglaciated range. Blue shading indicates higher than average effective migration; red shading indicates lower than average effective migration. Circles mark sample locations and circle size varies according to the number of individual plants assigned to the particular node on the grid.

Genetic diversity is lower towards range margins

Nucleotide diversity (π) was highest in the Central population (4.46×10^-3), intermediate in the Southwestern population (3.95×10^-3), and lower in the Northern (3.55×10^-3) and Southeastern (3.48×10^-3) populations (Fig. 3a). Continuous spatial windows of π clearly emphasized low diversity in northern regions (Fig. 3c. The low diversity of the Southeastern population as calculated with pixy was less clearly recapitulated in the wingen results, although windows at the southern edge of our sampling did show reduced diversity (Fig. 3c). F_{ST} was lowest for comparisons with the Central population (ranging from 0.096 to to 0.171) and highest between the Northern and Southeastern population ($F_{ST} = 0.242$) (Fig. 3a). In all four populations, linkage disequilibrium decayed rapidly to background levels within 1 kb but was overall higher in the Northern population compared to the others (Fig. 3b). Inbreeding coefficients were also low, around zero, across all four populations (Fig. 3c). Within the Northern population, inbreeding coefficients showed a significant latitudinal cline, with plants sampled from more northern sites having higher F values (Fig. 3d). This trend was concordant with a reduction in observed heterozygosity among windows encompassing the most Northern samples (Fig. 3e).

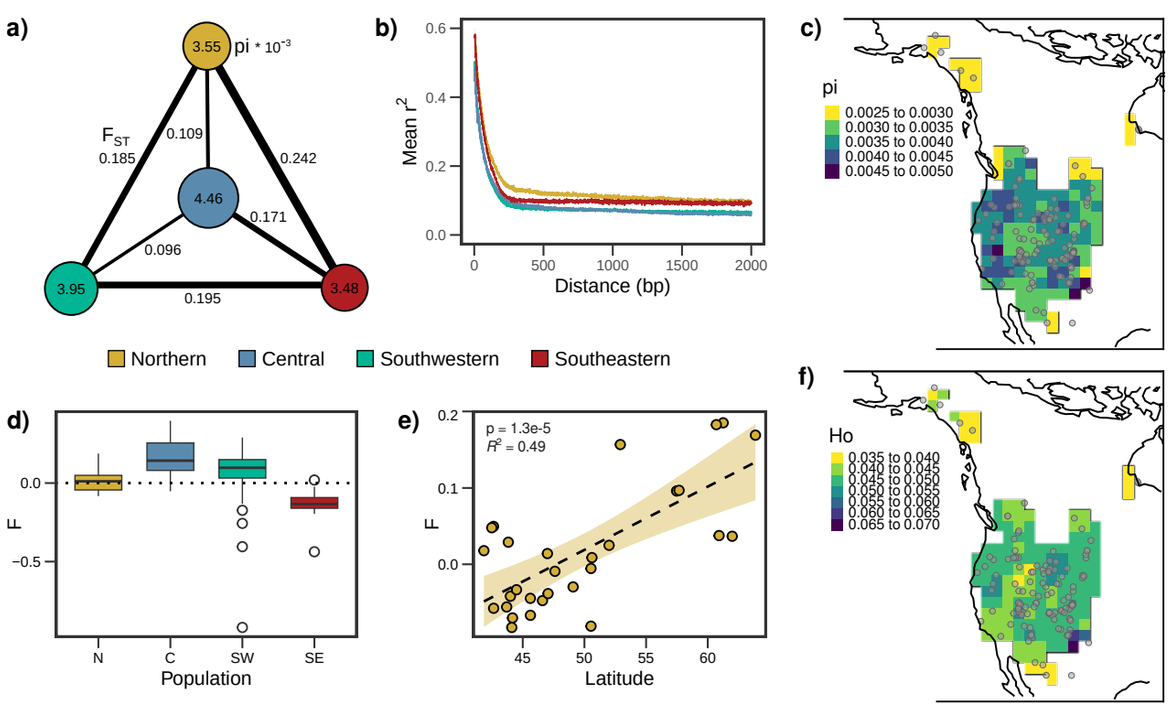

Figure 3: Population diversity, differentiation, linkage disequilibrium, and inbreeding statistics. **a)** Diagram representing population average pairwise nucleotide diversity (π) as circles and average pairwise F_{ST} as edges. Circle size and edge thickness are proportional to the indicated values. **b)** Linkage disequilibrium decay curves in the four populations. **c)** Nucleotide diversity (π) computed for moving windows across the landscape. **d)** Individual inbreeding coefficients (F) summarized at the population level. **e)** Inbreeding coefficients versus latitude within just the Northern population. **f)** Observed heterozygosity (Ho) averaged across individuals within moving spatial windows.

Population size contractions and splits occurred during the Last Glacial Period

Modeling of effective population size (N_e) through time with SMC++ revealed size contractions across all four populations that reached minima during the Last Glacial Period (115–11.7 kya) (Fig. 4). The Northern population, despite showing large historic N_e as late as 1 Mya, showed the greatest contraction and lowest minimum N_e (Fig. 4a). Timing of expansions in both the Central and Northern populations corresponded with the end of the Last Glacial Maximum around 20 kya, however the Northern population showed a much reduced expansion compared to the Central population (Fig. 4a). After the Last Glacial Maximum, the Southwestern population peaked at around 6 kya and subsequently contracted (Fig. 4a).

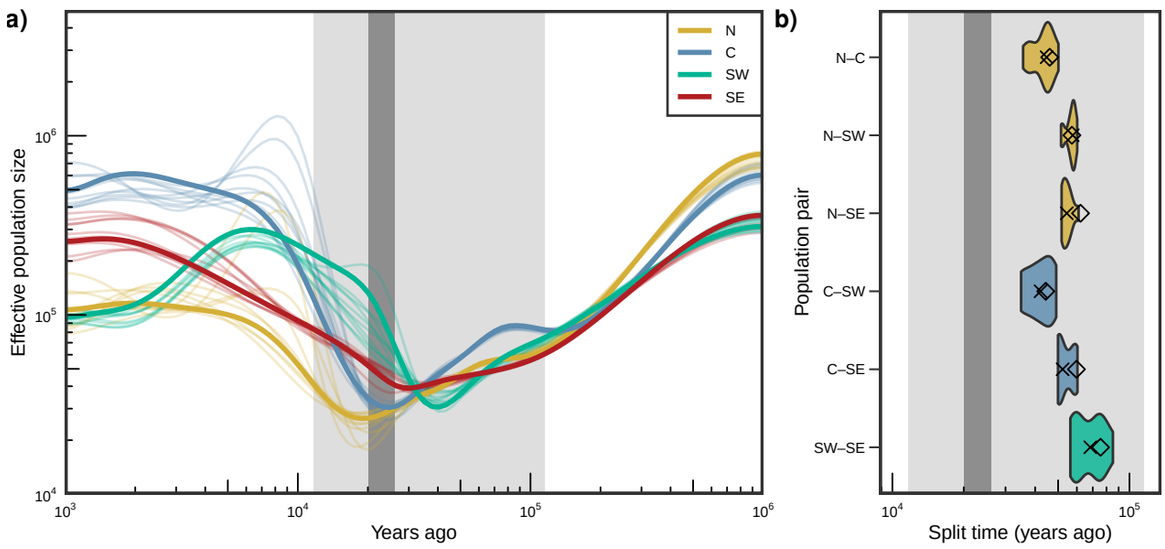

Figure 4: Demographic inference with SMC++ indicates population size contractions and splits during the Last Glacial Period. **a)** Population size histories 1,000–1,000,000 years ago. Thin lines represent 10 bootstrap replicates per population. Light gray and dark gray shading indicate the Last Glacial Period (115–11.7 kya) and Last Glacial Maximum (26–20 kya), respectively. **b)** Estimates of divergence times for each population pair. Diamonds indicate point estimates and violin plots represent bootstrap estimates with the median bootstrap value marked as a cross. Violin plots are colored according to the first population in the pair.

Divergence times for all population pairs were estimated to have occurred during the Last Glacial Period, but before the Last Glacial Maximum (Fig. 4b). The oldest split was between the Southwestern and Southeastern populations (median 68.4 kya). Split times were similar for Northern and Southeastern (median 54.4 kya) and Northern and Southwestern (median 57.6 kya) populations. Splits involving the Central population were estimated to have occurred most recently, ranging from 52.2 kya (C–SE) to 44.7 kya (N–C) and 42.1 kya (C–SW), which is consistent with these same pairs having the lowest F_{ST} (Fig. 3a).

Discussion

Impacts of glaciation and topography on *Linum lewisii* population structure

Our results provide new understanding of major lineages within Lewis flax and point to the historical and contemporary factors that have shaped genetic variation in this widespread species. We found that the population genetic structure of Lewis flax has a clear geographic element best explained by the presence of four ancestral populations, named here as the Northern, Central, Southwestern, and Southeastern populations (Fig. 1). Variation within and among these groups provides insight into the history of Lewis flax on the North American continent. Despite Lewis flax being a relatively cold-tolerant species (Gossweiler et al. 2024) with a hypothesized Beringian origin (McDill et al. 2009), we did not find evidence of a Beringian or other northern refugium. Instead, we found evidence of a genetic bottleneck in the Northern population that is consistent with postglacial range expansion from the south. The Northern population had low nucleotide diversity (π) relative to most other regions (Fig. 3). It also showed the most substantial decline in effective population size and the most recent expansion, at around the end of the Last Glacial Maximum (Fig. 4a). This makes sense considering that during the LGM, the Laurentide ice sheet covered most of the current distribution of the Northern population and would have caused substantial habitat loss and local extirpation. Higher effective migration rates in the north versus lower migration in the south are also consistent with northward range expansion and southern refugia (Fig. 2). Likewise, significant genetic structuring along the latitudinal axis (Fig. S7) supports this scenario and indicates that isolation by distance contributes to Lewis flax population structure. The chloroplast phylogeny (Fig. S10) could indicate that the regions currently inhabited by Central and perhaps Southwestern populations were refugia from which the Northern population is derived, or that the Central population is exposed to gene flow from all other populations. Additional sampling, especially from areas in the southeastern extent of Lewis flax range, would help clarify these hypotheses. Overall, these results align with a trend of north–south genetic structuring in widely distributed western North American plant species, including yellow monkeyflower (*Mimulus guttatus*) (Twyford et al. 2020) and quaking aspen (*Populus tremuloides*) (Callahan et al. 2013), as well as Great Plains species

like common sunflower (*Helianthus annuus*) (Park et al. 2020), and silflower *Silphium integrifolium* (Raduski et al. 2021).

Genetic structuring along the path of postglacial range expansion does not capture the whole story. We also found that the Rocky Mountains mark transitions from Central to Northern ancestry and from Central to Southeastern ancestry (Fig. 1). This pattern, coupled with estimated population split times (Fig. 4), suggest that the Rockies promoted divergence during the Last Glacial Period between populations on either side of the range. Effective migration surfaces also point to the Rocky Mountains—particularly the Southern Rocky Mountains—as a dispersal barrier (Figs. 2, S9). While the Southern Rockies are no longer heavily glaciated, Lewis flax is generally uncommon in the subalpine zone and not found above it, which suggests high mountains remain a significant contemporary barrier. Little is known about Lewis flax seed dispersal except that like other flax species, its seeds form a sticky mucilaginous coating when wet. However, it's unclear whether this provides a long-range dispersal mechanism via adherence to or ingestion by animals, or if it is primarily a means of anchoring to the ground (Kreitschitz et al. 2020; Yang et al. 2012). Among the varied topography of North America, the Rocky Mountains have long been treated as a predominant factor in genetic differentiation and secondary contact of closely related taxa (Jensen et al. 2024; Remington 1968; Swenson et al. 2005). In eastern North America, the Appalachian Mountains have had a similar effects on the population structure of plants like the perennial herbaceous bloodroot (*Sanguinaria canadensis)* (Xia et al. 2024) and American chestnut (*Castanea dentata*) (Sandercock et al. 2022).

Regarding the Southwestern population, we hypothesize that local adaptation to the arid climate of the Great Basin may have contributed to its split from the Central population, which surrounds the western, northern and eastern periphery of the Great Basin (Fig. 1a). Local adaptation is common in Great Basin flora (Baughman et al. 2019) and previous work has identified geographic and environmental trends in Lewis flax trait variation across this region that are consistent with adaptation (Innes, Gossweiler, et al. 2022). The contraction in effective population size of the Southwestern population around 6–5 kya also stood out qualitatively from the trajectories of the other populations and might be related to the middle Holocene periods of high aridity in the Great Basin (Lachniet et al. 2020).

Taxonomy

Our findings regarding Lewis flax population structure do not neatly align with the current accepted infraspecific taxonomic treatment, which delineates three varieties. *Linum lewisii* var. *alpicola* is described as a high-elevation specialist inhabiting mountain ridges (2,000–3,700 m) in California, Idaho, Nevada, and Utah, with reduced stature, flower size, and style length (Flora of North America Editorial Committee 2016). We sampled several plants from described *alpicola* habitat but did not observe any uniquely shared genetic ancestry among these samples (we note, however, that none of our samples were labeled *a priori* as var. *alpicola* and we did not assess phenotypes). Our sampling of Hudson Bay endemic *Linum lewisii* var. *lepagei* was limited to two herbarium specimens from a single area, both of which were grouped within the Northern population. In the chloroplast phylogeny, the *lepagei* samples were placed within a sub-clade of the main Northern clade alongside samples from Saskatchewan and North Dakota. The *lepagei* variety is thought to comprise a distinct lineage given its white flowers and separation from the core range (Flora of North America Editorial Committee 2016; Mosquin 1971), however, this was not apparent in our current sampling. Our analyses of genetic diversity and population demographic history support the idea that Lewis flax recolonized northern latitudes recently, following the Last Glacial Maximum (Figs. 3, 4). Therefore, it's plausible there has not been sufficient time for the *lepagei* variety to clearly differentiate from the rest of the Northern population. Improved sampling of var. *lepagei* as well as var. *alpicola* would be critical to make more conclusive determinations regarding their origins and genetic variation.

The third currently accepted infraspecific, *L. lewisii* var. *lewisii* encompasses the four main populations described here. In our experience, these populations do not show obvious, discrete phenotypic differences, but there is some evidence from previous crossing experiments that reproductive barriers exist between them: hybrids between an Alaskan population (i.e. Northern) and Texan population (putatively Southeastern) showed diminished pollen viability compared to hybrids between the same Texan population and a USDA collection of uncertain origin but "probably from the western United States" (Mosquin 1971). Future work should explore such crossing barriers more extensively and could lend support to updated taxonomic treatments more in line with population genetic structure.

Mating system evolution in North American blue flax

The flax genus is well known for heterostyly, a floral polymorphism that is tied to self-incompatibility and typically comprises two floral morphs with inverse positioning of anthers and stigmas. Mating system is also highly labile in *Linum* and there have been numerous independent losses of heterostyly and self-incompatibility (Maguilla et al. 2021; McDill et al. 2009). More generally, the shift to self-compatibility is one of the most common evolutionary transitions in flowering plants and is often associated with range expansions (Baker 1955). Selfing can be advantageous when pollinators or mates are rare but is expected to have substantial genetic and evolutionary consequences, including increased homozygosity and linkage disequilibrium, the accumulation of deleterious mutations, more pronounced population structure, reduced genetic diversity, and lower effective population size (Wright et al. 2013). Such effects have been observed in a comparison of closely related heterostylous and homostylous yellow flax species (Gutiérrez-Valencia et al. 2024).

Our results indicate that Lewis flax is predominantly outcrossing despite being homostylous and self-compatible. For instance, inbreeding coefficients overall were close to zero and linkage disequilibrium decayed rapidly (Fig. 3). Effective population sizes in recent history were also quite high (Fig. 4). These findings give new context to previous pollination experiments showing Lewis flax generally does not autonomously

self (Kearns et al. 1994). Its outcrossing propensity is likely explained by the fact Lewis flax flowers exhibit herkogamy, with stigmas positioned several millimeters beyond the anthers (i.e. approach herkogamy, Fig. 1). However, this anther–stigma distance has been reported to be significantly reduced in populations towards the northern extent of its range (Mosquin 1971). We accordingly found a substantial latitudinal cline in inbreeding coefficients within the Northern population (Fig. 3e). While our individual-based sampling scheme means we cannot fully eliminate potential bias resulting from local allele frequencies differing from the cluster average (i.e. Wahlund effect), we also observed reduced individual-based heterozygosity rates (H_o) in northern areas (Fig. 3f), indicating that the inbreeding cline is not artifactual. These patterns—in addition to elevated linkage disequilibrium (Fig. 3b and low N_e relative to the other populations, (Fig. 4a)—are consistent with a shift to increased selfing in the Northern population. These results could alternatively be attributed to smaller and more bottlenecked populations, but selfing is a compelling hypothesis given previous observations of reduced herkogamy in northern Lewis flax populations and the fact that strong selfing is linked to diminished herkogamy in Lewis flax's sister taxa, *Linum pratense* ((Mosquin 1971; Uno 1984)). More intensive population-level sampling coupled with measurements of herkogamy will be needed to clarify the potential link between post-glaciation range expansion and increased selfing rates in Lewis flax. This scenario has been documented in other widespread North American forbs, such as American bellflower *Campanula americana* (Koski et al. 2019). In sum, Lewis flax may have a "best of both worlds" reproductive strategy—the ability to self provides reproductive assurance, while predominant outcrossing limits inbreeding depression (Goodwillie et al. 2005). We speculate that this contributed to its colonization of North America as well as its broad prevalence on the continent.

Implications for restoration and agriculture

Human-facilitated migration could explain the disjunct presence of Southwestern-type samples far outside the main geographic region of that population (Fig. 1). This is most assuredly the case for the accession from an isolated population in the Appalachian mountains of West Virginia (Ames 34946, Table S1). It is plausible that humans also moved another accession of Southwestern ancestry, originally collected from a roadside area in northern Wyoming (Ames 34266), as well as three samples of Southwestern ancestry collected in central Colorado (Guanella Pass High, Dow Mountain, and Homestake Reservoir). Commercial production of Lewis flax seed has focused on populations from the Great Basin (particularly the Maple Grove germplasm), due to ecosystem restoration efforts in this region creating high demand for native seed (Innes, Gossweiler, et al. 2022; Pendleton et al. 2008). The aforementioned samples were assigned 100% Southwestern ancestry despite being geographically distant from the southwestern region and were all collected after the 2003 commercial release of Maple Grove. Homestake Reservoir had the same chloroplast haplotype as Maple Grove, while the other disjunct Southwestern-type samples clustered with those from Nevada (Fig. S10). This suggests that Southwestern genotypes, including Maple Grove, from Utah, as well as Nevada genotypes, have been moved far beyond their original provenance. Increasing human-mediated dispersal and prevalence of Southwestern genotypes could have consequences for population structure and evolution of Lewis flax into the future. We encourage land managers to take into account the spatial population structure described here when sourcing Lewis flax for restoration projects, ideally aiming for regionally-sourced, locally-adapted seed.

Ongoing efforts to domesticate lewis flax for edible, oil-rich seed are still in the early stages. The results we present here may help guide collection and crossing efforts. The major genetic lineages we identify here may not be easily intercrossed (Mosquin 1971) and additional data from crossing experiments are needed to better understand these incompatibilities. Such crossing barriers may make trait transfer from one genetic lineage to another via breeding difficult or impossible. Additionally, future seed collection efforts could be targeted to areas where limited seed is available in current repositories, particularly the southern Great Plains, California, Mexico, and Canada, where our sampling mainly came from herbaria if at all. This work will also inform development of diversity panels and novel study populations (such as nested association mapping panels) for future trait genetic mapping.

Conclusion

We sampled whole genomes of *Linum lewisii* across a broad swath of the North American continent. This work relied heavily on herbarium collections, from which we obtained high quality DNA. We found that Lewis flax genetic variation is spatially clustered in four main populations and that glaciation, mountain ranges, and deserts have influenced Lewis flax genetic diversity and structure. The Northern population likely arose from postglacial range expansion back into high latitudes, rather than persistence following the species' hypothesized Beringian origin. The Rocky Mountains are an apparent geographic barrier associated with spatial delineation of the Northern and Central populations, as well as the Southeastern and Central populations. We speculate that the Great Basin may act as an environmental barrier maintaining adaptive divergence of the Southwestern and Central populations. Our results indicate Lewis flax is predominantly outcrossing despite the fact it is homostylous and self-compatible; we did, however, find evidence of elevated inbreeding towards its northern range limit, which could be explained by increased selfing stemming from a reduction in herkogamy. The four clusters identified here were unknown previously and do not closely correspond with current infraspecific taxonomy. Conservation genomic implications of this range-wide phylogeographic analysis include revealing the origin of endangered variety *L. lewisii* var. *lepagei* and the spatial extent to which commercially available Lewis flax seed used in restoration projects matches the genetic composition of wild populations. Additionally, these findings and genomic resources can inform ongoing Lewis flax neodomestication efforts. Overall, this study illustrates the genetic legacy of the last Ice Age in North America and contributes to our understanding of the relationship between continental-scale geography, mating system, and the structuring of population genetic diversity.

Acknowledgments

The authors kindly thank the following herbarium curators and collection managers for obliging our sampling requests: Burrell E. Nelson at Rocky Mountain Herbarium (RM), Diana Sawatzky and Bruce Ford at University of Manitoba Herbarium (WIN), Andrew Salywon at Desert Botanical Garden (DES), Massimo Boscolo at Arizona-Sonora Desert Museum (ASDM), and Dina Clark and Erin Tripp at University of Colorado Herbarium (COLO). The authors thank David Tork and Neil Anderson for providing plant tissue from multiple seed bank accessions, Brian Smart for greenhouse help, Andre Gossweiler for help with field collections, and Dan Klee for helpful discussion of phylogenetic methods. This work was supported with funding from City of Boulder Open Space Mountain Parks, Boulder County Nature Association, and NSF IGERT grant number 1144807.

References

Alexander, D. H., J. Novembre, and K. Lange (2009). "Fast model-based estimation of ancestry in unrelated individuals". In: *Genome Research* 19.9. DOI: 10.1101/gr.094052.109.

Alsos, I. G., S. Lavergne, M. K. F. Merkel, M. Boleda, Y. Lammers, A. Alberti, C. Pouchon, F. Denoeud, I. Pitelkova, et al. (2020). "The Treasure Vault Can be Opened: Large-Scale Genome Skimming Works Well Using Herbarium and Silica Gel Dried Material". In: *Plants* 9.4. DOI: 10.3390/plants9040432.

Bagley, J. C., N. M. Heming, E. E. Gutiérrez, U. K. Devisetty, K. E. Mock, A. J. Eckert, and S. H. Strauss (2020). "Genotyping-by-sequencing and ecological niche modeling illuminate phylogeography, admixture, and Pleistocene range dynamics in quaking aspen (Populus tremuloides)". In: *Ecology and Evolution* 10.11. DOI: 10.1002/ece3.6214.

Baker, H. G. (1955). "Self-Compatibility and Establishment After 'Long-Distance' Dispersal". In: *Evolution* 9.3. DOI: 10.2307/2405656.

Baughman, O. W., A. C. Agneray, M. L. Forister, F. F. Kilkenny, E. K. Espeland, R. Fiegener, M. E. Horning, R. C. Johnson, T. N. Kaye, et al. (2019). "Strong patterns of intraspecific variation and local adaptation in Great Basin plants revealed through a review of 75 years of experiments". In: *Ecology and Evolution* 9.11. DOI: 10.1002/ece3.5200.

Beatty, G. E. and J. Provan (2010). "Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant Orthilia secunda". In: *Molecular Ecology* 19.22. DOI: 10.1111/j.1365-294X.2010.04859.x.

Browning, B. L., Y. Zhou, and S. R. Browning (2018). "A One-Penny Imputed Genome from Next-Generation Reference Panels". In: *The American Journal of Human Genetics* 103.3. DOI: 10.1016/j.ajhg.2018.07.015.

Callahan, C. M., C. A. Rowe, R. J. Ryel, J. D. Shaw, M. D. Madritch, and K. E. Mock (2013). "Continental-scale assessment of genetic diversity and population structure in quaking aspen (Populus tremuloides)". In: *Journal of Biogeography* 40.9. DOI: 10.1111/jbi.12115.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets". In: *GigaScience* 4.1. DOI: 10.1186/s13742-015-0047-8.

Chen, S., Y. Zhou, Y. Chen, and J. Gu (2018). "fastp: an ultra-fast all-in-one FASTQ preprocessor". In: *Bioinformatics* 34.17. DOI: 10.1093/bioinformatics/bty560.

Chernomor, O., A. von Haeseler, and B. Q. Minh (2016). "Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices". In: *Systematic Biology* 65.6. DOI: 10.1093/sysbio/syw037.

Coe, K., H. Bostan, W. Rolling, S. Turner-Hissong, A. Macko-Podgórni, D. Senalik, S. Liu, R. Seth, J. Curaba, et al. (2023). "Population genomics identifies genetic signatures of carrot domestication and improvement and uncovers the origin of high-carotenoid orange carrots". In: *Nature Plants* 9.10. DOI: 10.1038/s41477-023-01526-6.

Comes, H. P. and J. W. Kadereit (1998). "The effect of Quaternary climatic changes on plant distribution and evolution". In: *Trends in Plant Science* 3.11. DOI: 10.1016/S1360-1385(98)01327-2.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, et al. (2021). "Twelve years of SAMtools and BCFtools". In: *GigaScience* 10.2. DOI: 10.1093/gigascience/giab008.

Evans, L. M., G. J. Allan, S. P. DiFazio, G. T. Slavov, J. A. Wilder, K. D. Floate, S. B. Rood, and T. G. Whitham (2015). "Geographical barriers and climate influence demographic history in narrowleaf cottonwoods". In: *Heredity* 114.4. DOI: 10.1038/hdy.2014.115.

Flora of North America Editorial Committee (2016). *Flora of North America North of Mexico*. Vol. 12. 30 vols. New York and Oxford: Oxford University Press.

Fonseca, E. M., T. A. Pelletier, S. K. Decker, D. J. Parsons, and B. C. Carstens (2023). "Pleistocene glaciations caused the latitudinal gradient of within-species genetic diversity". In: *Evolution Letters* 7.5. DOI: 10.1093/evlett/qrad030.

González-Velasco, J., M. Burgos-Hernández, I. G. Galván-Escobedo, and G. Castillo-Campos (2022). "Taxonomic update of the flax family in Mexico". In: *Phytotaxa* 549.2. DOI: 10.11646/phytotaxa.549.2.3.

Goodwillie, C., S. Kalisz, and C. G. Eckert (2005). "The Evolutionary Enigma of Mixed Mating Systems in Plants: Occurrence, Theoretical Explanations, and Empirical Evidence". In: *Annual Review of Ecology, Evolution, and Systematics* 36.1. DOI: 10.1146/annurev.ecolsys.36.091704.175539.

Gossweiler, A. D., B. C. Smart, and B. S. Hulke (2024). "Survival analysis of freezing stress in the North American native perennial flax, Linum lewisii Pursh". In: *AoB PLANTS*. DOI: 10.1093/aobpla/plae022.

Gutiérrez-Valencia, J., P.-I. Zervakis, Z. Postel, M. Fracassetti, A. Losvik, S. Mehrabi, I. Bunikis, L. Soler, P. W. Hughes, et al. (2024). "Genetic Causes and Genomic Consequences of Breakdown of Distyly in Linum trigynum". In: *Molecular Biology and Evolution* 41.5. DOI: 10.1093/molbev/msae087.

Harris, B. D. (1968). "CHROMOSOME NUMBERS AND EVOLUTION IN NORTH AMERICAN SPECIES OF LINUM". In: *American Journal of Botany* 55.10. DOI: 10.1002/j.1537-2197.1968.tb07485.x.

Hemstrom, W., J. A. Grummer, G. Luikart, and M. R. Christie (2024). "Next-generation data filtering in the genomics era". In: *Nature Reviews Genetics*. DOI: 10.1038/s41576-024-00738-6.

Hewitt, G. M. (2004). "Genetic consequences of climatic oscillations in the Quaternary". In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359.1442. DOI: 10.1098/rstb.2003.1388.

Hewitt, G. M. (1996). "Some genetic consequences of ice ages, and their role in divergence and speciation". In: *Biological Journal of the Linnean Society* 58.3. DOI: 10.1111/j.1095-8312.1996.tb01434.x.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh (2018). "UFBoot2: Improving the Ultrafast Bootstrap Approximation". In: *Molecular Biology and Evolution* 35.2. DOI: 10.1093/molbev/msx281.

Innes, P., A. Gossweiler, S. Jensen, D. Tilley, L. St. John, T. Jones, S. Kitchen, and B. S. Hulke (2022). "Assessment of biogeographic variation in traits of Lewis flax (Linum lewisii) for use in restoration and agriculture". In: *AoB PLANTS* 14.2. DOI: 10.1093/aobpla/plac005.

Innes, P. A., B. C. Smart, J. A. M. Barham, B. S. Hulke, and N. C. Kane (2023). *Chromosome-scale Genome Assembly of Lewis Flax (Linum lewisii Pursh.)* DOI: 10.1101/2023.10.10.561607.

Jaramillo-Correa, J. P., J. Beaulieu, D. P. Khasa, and J. Bousquet (2009). "Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes". In: *Canadian Journal of Forest Research* 39.2. DOI: 10.1139/X08-181.

Jensen, A. J., M. V. Cove, B. R. Goldstein, R. Kays, W. McShea, K. Pacifici, B. Rooney, and E. Kierepka (2024). "Geographic barriers but not life history traits shape the phylogeography of North American mammals". In: *Global Ecology and Biogeography* 33.8. DOI: 10.1111/geb.13875.

Jones, T. A. (2019). "Native Seeds in the Marketplace: Meeting Restoration Needs in the Intermountain West, United States". In: *Rangeland Ecology and Management* 72.6. DOI: 10.1016/j.rama.2019.07.009.

Jónsson, H., A. Ginolhac, M. Schubert, P. L. F. Johnson, and L. Orlando (2013). "mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters". In: *Bioinformatics* 29.13. DOI: 10.1093/bioinformatics/btt193.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin (2017). "ModelFinder: fast model selection for accurate phylogenetic estimates". In: *Nature Methods* 14.6. DOI: 10.1038/nmeth.4285.

Kearns, C. A. and D. W. Inouye (1994). "Fly pollination of Linum lewisii (Linaceae)". In: *American Journal of Botany* 81.9. DOI: 10.1002/j.1537-2197.1994.tb15602.x.

Kitchen, S. G. (1995). *Return of the native: a look at select accessions of North American Lewis flax.* General Technical Report INT-GTR-315. USDA Forest Service Intermountain Research Station.

Korunes, K. L. and K. Samuk (2021). "pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data". In: *Molecular Ecology Resources* 21.4. DOI: 10.1111/1755-0998.13326.

Koski, M. H., N. C. Layman, C. J. Prior, J. W. Busch, and L. F. Galloway (2019). "Selfing ability and drift load evolve with range expansion". In: *Evolution Letters* 3.5. DOI: 10.1002/evl3.136.

Kreitschitz, A., E. Haase, and S. N. Gorb (2020). "The role of mucilage envelope in the endozoochory of selected plant taxa". In: *The Science of Nature* 108.1. DOI: 10.1007/s00114-020-01709-7.

Lachniet, M. S., Y. Asmerom, V. Polyak, and R. Denniston (2020). "Great Basin Paleoclimate and Aridity Linked to Arctic Warming and Tropical Pacific Sea Surface Temperatures". In: *Paleoceanography and Paleoclimatology* 35.7. DOI: 10.1029/2019PA003785.

Leger, E. A. and O. W. Baughman (2015). "What Seeds to Plant in the Great Basin? Comparing Traits Prioritized in Native Plant Cultivars and Releases with those that Promote Survival in the Field". In: *Natural Areas Journal* 35.1. DOI: 10.3375/043.035.0108.

Lumibao, C. Y., S. M. Hoban, and J. McLachlan (2017). "Ice ages leave genetic diversity 'hotspots' in Europe but not in Eastern North America". In: *Ecology Letters* 20.11. DOI: 10.1111/ele.12853.

Lyman, R. A. and C. E. Edwards (2022). "Revisiting the comparative phylogeography of unglaciated eastern North America: 15 years of patterns and progress". In: *Ecology and Evolution* 12.4. DOI: 10.1002/ece3.8827.

Maguilla, E., M. Escudero, J. Ruíz-Martín, and J. Arroyo (2021). "Origin and diversification of flax and their relationship with heterostyly across the range". In: *Journal of Biogeography* 48.8. DOI: 10.1111/jbi.14129.

Marcus, J., W. Ha, R. F. Barber, and J. Novembre (2021). "Fast and flexible estimation of effective migration surfaces". In: *eLife* 10. Ed. by G. H. Perry, I. Alves, and W. Tansey. DOI: 10.7554/eLife.61927.

Marinček, P., N. D. Wagner, and S. Tomasello (2022). "Ancient DNA extraction methods for herbarium specimens: When is it worth the effort?" In: *Applications in Plant Sciences* 10.3. DOI: 10.1002/aps3.11477.

McDill, J., M. Repplinger, B. B. Simpson, and J. W. Kadereit (2009). "The Phylogeny of Linum and Linaceae Subfamily Linoideae, with Implications for Their Systematics, Biogeography, and Evolution of Heterostyly". In: *SYSTEMATIC BOTANY* 34.

Meisner, J., S. Liu, M. Huang, and A. Albrechtsen (2021). "Large-scale inference of population structure in presence of missingness using PCA". In: *Bioinformatics* 37.13. DOI: 10.1093/bioinformatics/btab027.

Miles, A., M. R, P. Ralph, N. Harding, R. Pisupati, S. Rae, and T. Millar (2021). *scikit-allel*. Version 1.3.3.
Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear (2020). "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era". In: *Molecular Biology and Evolution* 37.5. DOI: 10.1093/molbev/msaa015.
Mosquin, T. (1971). "Biosystematic studies in the North American species of Linum, section Adenolinum (Linaceae)". In: *Canadian Journal of Botany* 49.
Ortiz, E. M. (2019). *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis*. Version 2.0.
Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch (2010). "The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana". In: *Science* 327.5961. DOI: 10.1126/science.1180677.
Park, B. and J. M. Burke (2020). "Phylogeography and the Evolutionary History of Sunflower (Helianthus annuus L.): Wild Diversity and the Dynamics of Domestication". In: *Genes* 11.3. DOI: 10.3390/genes11030266.
Pedersen, B. S. and A. R. Quinlan (2018). "Mosdepth: quick coverage calculation for genomes and exomes". In: *Bioinformatics* 34.5. DOI: 10.1093/bioinformatics/btx699.
Pendleton, R. L., S. G. Kitchen, E. D. McArthur, and J. E. Mudge (2008). "The 'Appar' flax release: origin, distinguishing characteristics, and use; and a native alternative". In: *Native Plants Journal* 9.1. DOI: 10.2979/NPJ.2008.9.1.18.
Pockrandt, C., M. Alzamel, C. S. Iliopoulos, and K. Reinert (2020). "GenMap: ultra-fast computation of genome mappability". In: *Bioinformatics* 36.12. DOI: 10.1093/bioinformatics/btaa222.
Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov (2020). "Using SPAdes De Novo Assembler". In: *Current Protocols in Bioinformatics* 70.1. DOI: 10.1002/cpbi.102.
Provan, J. and K. D. Bennett (2008). "Phylogeographic insights into cryptic glacial refugia". In: *Trends in Ecology & Evolution* 23.10. DOI: 10.1016/j.tree.2008.06.010.
Pull, Z. A., G. Gramig, B. S. Hulke, A. Gossweiler, and B. Johnson (2023). "First steps toward developing Lewis flax (Linum lewisii Pursh) as an agronomic crop". In: *Renewable Agriculture and Food Systems* 38. DOI: 10.1017/S1742170523000340.
Raduski, A. R., A. Herman, C. Pogoda, K. M. Dorn, D. L. V. Tassel, N. Kane, and Y. Brandvain (2021). "Patterns of genetic variation in a prairie wildflower, Silphium integrifolium, suggest a non-prairie origin and locally adaptive variation". In: *American Journal of Botany* 108.1. DOI: https://doi.org/10.1002/ajb2.1603.
Remington, C. L. (1968). "Suture-Zones of Hybrid Interaction Between Recently Joined Biotas". In: *Evolutionary Biology: Volume 2*. Ed. by T. Dobzhansky, M. K. Hecht, and W. C. Steere. Boston, MA: Springer US. DOI: 10.1007/978-1-4684-8094-8_8.
Sandercock, A. M., J. W. Westbrook, Q. Zhang, H. A. Johnson, T. M. Saielli, J. A. Scrivani, S. F. Fitzsimmons, K. Collins, M. T. Perkins, et al. (2022). "Frozen in time: Rangewide genomic diversity, structure, and demographic history of relict American chestnut populations". In: *Molecular Ecology* 31.18. DOI: 10.1111/mec.16629.
Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman (2010). "Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America". In: *Molecular Ecology* 19.21. DOI: 10.1111/j.1365-294X.2010.04828.x.
Soltis, D. E., M. A. Gitzendanner, D. D. Strenge, and P. S. Soltis (1997). "Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America". In: *Plant Systematics and Evolution* 206.1. DOI: 10.1007/BF00987957.
Soltis, D. E., A. B. Morris, J. S. McLACHLAN, P. S. Manos, and P. S. Soltis (2006). "Comparative phylogeography of unglaciated eastern North America". In: *Molecular Ecology* 15.14. DOI: 10.1111/j.1365-294X.2006.03061.x.
Stone, B. W. and A. D. Wolfe (2021). "Phylogeographic analysis of shrubby beardtongues reveals range expansions during the Last Glacial Maximum and implicates the Klamath Mountains as a hotspot for hybridization". In: *Molecular Ecology* 30.15. DOI: 10.1111/mec.15992.
Swenson, N. G. and D. J. Howard (2005). "Clustering of Contact Zones, Hybrid Zones, and Phylogeographic Breaks in North America." In: *The American Naturalist* 166.5. DOI: 10.1086/491688.
Tange, O. (2023). *GNU Parallel 20231022 ('Al-Aqsa Deluge') [stable]*. DOI: 10.5281/zenodo.10035562.
Terhorst, J., J. A. Kamm, and Y. S. Song (2017). "Robust and scalable inference of population history from hundreds of unphased whole genomes". In: *Nature Genetics* 49.2. DOI: 10.1038/ng.3748.
Tomimatsu, H., S. R. Kephart, and M. Vellend (2009). "Phylogeography of Camassia quamash in western North America: postglacial colonization and transport by indigenous peoples". In: *Molecular Ecology* 18.18. DOI: 10.1111/j.1365-294X.2009.04341.x.
Twyford, A. D., E. L. Y. Wong, and J. Friedman (2020). "Multi-level patterns of genetic structure and isolation by distance in the widespread plant Mimulus guttatus". In: *Heredity* 125.4. DOI: 10.1038/s41437-020-0335-7.
Uno, G. (1984). "The Role of Persistent Sepals in the Reproductive Biology of Linum pratense (Linaceae)". In: *The Southwestern Naturalist* 29.4.
Vasimuddin, M., S. Misra, H. Li, and S. Aluru (2019). "Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems". In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). DOI: 10.1109/IPDPS.2019.00041.
Webb, C. J. and D. G. Lloyd (1986). "The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy". In: *New Zealand Journal of Botany* 24.1. DOI: 10.1080/0028825X.1986.10409726.

Winslow, S., K. Clause, J. Jacobs, and R. Hybner (2009). "REVEGETATION TRIALS IN THE PINEDALE ANTICLINE PROJECT AREA". In: *Journal American Society of Mining and Reclamation* 1. DOI: 10.21000/jasmr09011627.

Wright, S. I., S. Kalisz, and T. Slotte (2013). "Evolutionary consequences of self-fertilization in plants". In: *Proceedings of the Royal Society B: Biological Sciences* 280.1760. DOI: 10.1098/rspb.2013.0133.

Xia, M.-Q., Y.-X. Luo, Y. Suyama, A. Matsuo, S. Sakaguchi, Y.-G. Wang, and P. Li (2024). "Genetic divergence and ecological adaptation of an eastern North American spring ephemeral Sanguinaria canadensis". In: *Diversity and Distributions* 30.4. DOI: 10.1111/ddi.13813.

Yang, X., J. M. Baskin, C. C. Baskin, and Z. Huang (2012). "More than just a coating: Ecological importance, taxonomic occurrence and phylogenetic relationships of seed coat mucilage". In: *Perspectives in Plant Ecology, Evolution and Systematics* 14.6. DOI: 10.1016/j.ppees.2012.09.002.

You, F. M., J. Xiao, P. Li, Z. Yao, G. Jia, L. He, T. Zhu, M.-C. Luo, X. Wang, et al. (2018). "Chromosome-scale pseudomolecules refined by optical, physical and genetic maps in flax". In: *The Plant Journal* 95.2. DOI: 10.1111/tpj.13944.

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam (2017). "ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data". In: *Methods in Ecology and Evolution* 8.1. DOI: 10.1111/2041-210X.12628.

Zhang, C., S.-S. Dong, J.-Y. Xu, W.-M. He, and T.-L. Yang (2019). "PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files". In: *Bioinformatics* 35.10. DOI: 10.1093/bioinformatics/bty875.