



Genomic analysis reveals population structure and guides conservation actions of the endangered plant, *Pedicularis furbishiae* (Orobanchaceae)

Dawson M. White^{1,2} · Eric Normandeau³ · Manuel Lamothe⁴ · Peter R. Nelson⁵ · Richard H. Ree¹ · Martin Williams⁶

Received: 22 April 2025 / Accepted: 21 October 2025

© Crown and Dawson M. White, Eric Normandeau, Peter R. Nelson, Richard H. Ree 2025

Abstract

Pedicularis furbishiae is a globally rare, hemiparasitic perennial restricted to dynamic riverbank habitats along the Wolas-toq (Saint John) River in Maine, USA, and New Brunswick, Canada. The species now faces severe threats from habitat fragmentation, altered hydrology, and demographic collapse—particularly in New Brunswick, where several subpopulations have been recently extirpated and where recovery activities are being undertaken to protect the species from further decline. In this study, we assembled the first reference genome for *P. furbishiae* and conducted reduced-representation genotyping by sequencing (GBS) across 69 individuals sampled from all major subpopulations. Despite recovering a modest number of SNPs (152), we identified four genetically distinct populations: one in Maine with some subpopulation structure and three isolated subpopulations in New Brunswick. Genetic diversity was low overall, with elevated inbreeding coefficients in Grand Falls and Medford populations. All subpopulations contain private alleles, indicating the importance of conserving even the most demographically vulnerable sites. These findings support an integrated conservation strategy combining seed banking, ex situ propagation, and genetic rescue through controlled outcrossing among subpopulations, and provide a critical genomic baseline to inform the recovery of this imperiled species through population reinforcement and reintroduction across its range.

Keywords Hemiparasite · Ex situ conservation · Population enrichment · Population fragmentation · Inbreeding · Genetic rescue · Endemism · Furbish’s lousewort

Introduction

The maintenance of genetic diversity is critical to the long-term survival of rare species experiencing population decline (Frankham 2005). Reduced genetic diversity can limit

adaptive potential, elevate extinction risk, and increase the expression of deleterious alleles through inbreeding depression (Soulé 1985), but gene flow between populations can counteract these effects. The concept of genetic rescue—an increase in population fitness following the introduction

✉ Dawson M. White
dawson.white@gmail.com

✉ Martin Williams
martin.williams@nrca-nrcan.gc.ca

¹ Department of Collections, Conservation, and Research, Field Museum, 1400 S. DuSable Lake Shore Dr, Chicago, IL 60605–2496, USA

² Harvard University Herbaria, 22 Divinity Ave, Cambridge, MA 01238, USA

³ Plateforme de bio-informatique de l’IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval, Québec, QC G1V 0A6, Canada

⁴ Canadian Forest Service, Natural Resources Canada, Laurentian Forestry Centre, 1055 rue du P.E.P.S, Québec, QC G1V 4C7, Canada

⁵ School of Forest Resources, University of Maine, 5755 Nutting Hall, Orono, ME 04469-57556, USA

⁶ Canadian Forest Service, Natural Resources Canada, Atlantic Forestry Center, 1350 Regent street, Fredericton, NB E3C 2G6, Canada

of new genetic variation via gene flow—has emerged as a viable conservation strategy for fragmented and declining species (Thrall et al. 1998; Whiteley et al. 2015). A successful genetic rescue plan begins with understanding the spatial distribution of genetic variation and the degree of population connectivity.

Furbish's lousewort (*Pedicularis furbishiae* S. Watson) is a globally rare, hemiparasitic, herbaceous, perennial plant known from about 20 isolated, scattered subpopulations along a 225-kilometer stretch of the Wolastoq (Saint John River) spanning northern Maine (U.S.) and western New Brunswick (Canada; Fig. 1). It is an early successional species that critically depends on the availability and connectivity of its dynamic and narrow habitat: recently ice-scoured, north-facing riverbanks on the Wolastoq, which are constantly changing due to seasonal flooding, ice-scour, and erosion (Macior 1978; Gawler et al. 1987; Menges 1990). The long-term persistence of *P. furbishiae* is characterized by a dynamic blink-in, blink-out process (Marsh and Trenham 2001) in which colonization of recently scoured riverbanks is balanced by local extirpation due to successional advancement or disturbance from hydrological events (Gawler et al. 1987). It is thus likely that seed dispersal, colonization, establishment, and extirpation along its narrow riverbank habitat yielded a functional metapopulation for *P. furbishiae* (see Halsey et al. 2015) whose survival depends on the consistent and continuous supply of colonizable habitat (Hutchinson 1951).

The hydrological dynamics that have heretofore sustained *P. furbishiae* appear to be changing, raising concerns and questions about its future survival. Increases in winter shocks, winter flows and midwinter rainfall events are possibly causing more disturbances and erosion (Winton et al. 2024). Unfortunately, due to the lack of historical survey data on plant numbers and on habitat loss across the NB landscape, it is difficult to address whether this will be beneficial or problematic for the species due to the lack of subpopulation redundancy, the presence of physical barriers such as dams that could prevent gene flow and minimal seed production due to low plant numbers. By subpopulations, we mean discrete riverbank groupings smaller than 1 km and separated by more than 1 km long holding colonies/patches of *P. furbishiae* individuals (Winton et al. 2024). Even single plants are infrequently encountered outside of these subpopulations, especially in NB.

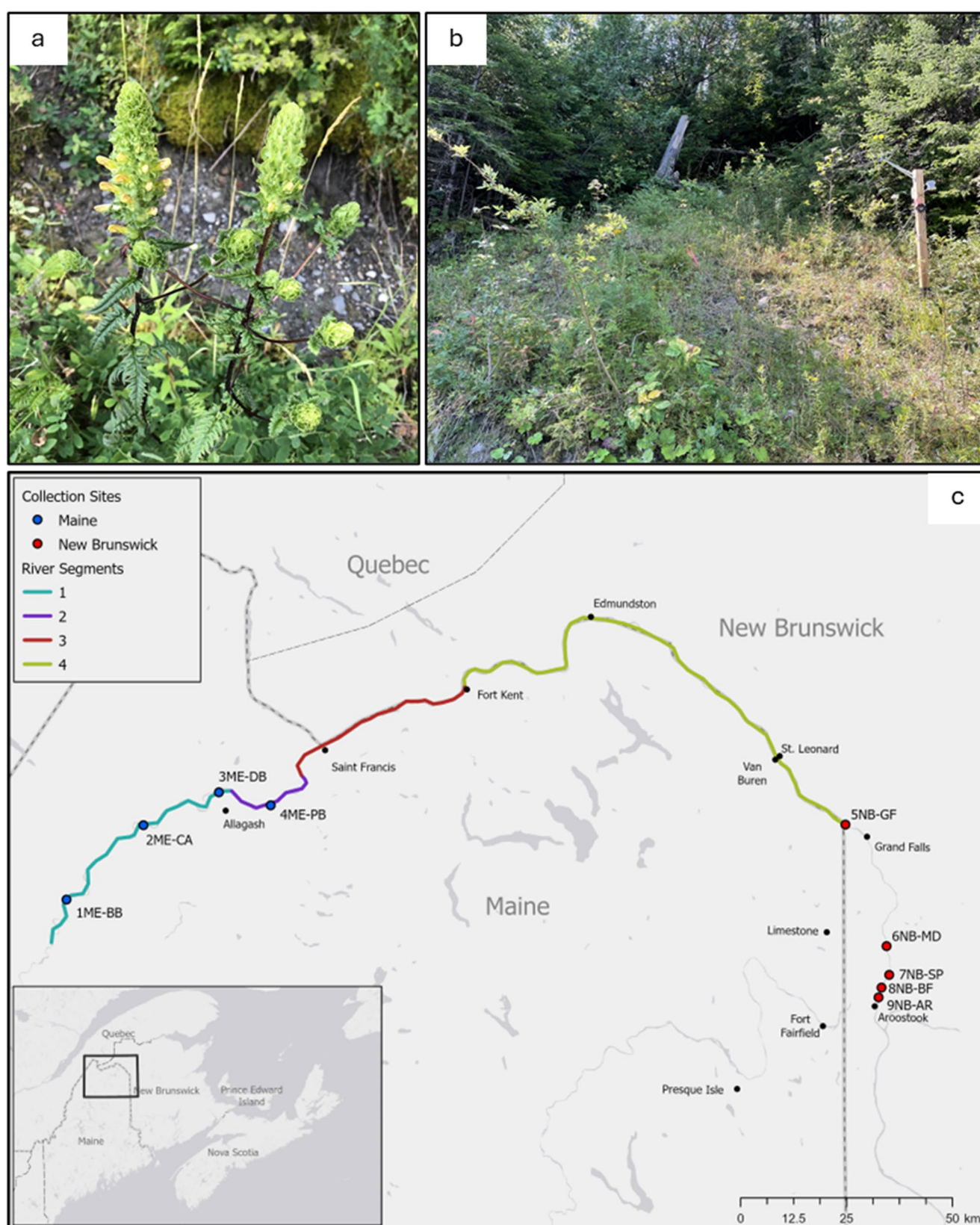
Early studies of the reproductive biology of *P. furbishiae* found it to be pollinated exclusively by the half-black bumble bee (*Bombus vagans*) and reliant on outcrossing for seed set (Macior 1978). It is not known whether it has any system of genetic self-incompatibility. Autonomous self-fertilization appears to be rare but possible; we have observed the

Fig. 1 Furbish's lousewort (*Pedicularis furbishiae* Watson) and geographic distributions. Flowering caudices and riverbank habitat (panel a and b). The riverbank shown in b illustrates an example of typical viable habitat found on the river. A weather station that monitors photosynthetically active radiation, soil moisture and temperature as well as air temperature and relative humidity can be seen on the left side of the image. The river forms the boundary between NB and Maine in segment 3 and 4 of the river. The river has been divided in 4 segments and are described in the Service's Recovery Plan (USFWS 1991). Locations of sampled subpopulations on Wolastoq (Saint John) River. Points for sampled sites are blue (Maine) or red (New Brunswick)

production of viable seeds by a single isolated individual in a controlled field bank with no other congeners present (M. Williams, unpublished). The fruits of *P. furbishiae* are capsules containing small seeds with no known animal dispersal vectors, and floating scapes are likely the primary means by which seeds are dispersed to downstream habitat (Menges 1990).

As louseworts (*Pedicularis* L.) are generally hemiparasitic, another dimension of the biology of *P. furbishiae* is the reliance on host plants. Experiments have shown that seedlings are host-generalist obligate root parasites (Macior 1980), but in the field, adult plants do not appear to maintain haustorial connections to hosts (Macior 1978). In greenhouse tests, effective hosts included alder (*Alnus incana*), wild licorice (*Glycyrrhiza lepidota*), and showy tick trefoil (*Desmodium canadense*; R. Fournier 2017, unpublished; M. Williams 2023–24, unpublished).

Following its botanical discovery by Kate Furbish in 1880 in Van Buren, Maine (Kennedy 2025), *P. furbishiae* is among the few northeastern North American plant taxa named in honor of a female botanist. Historical accounts corroborate that the species was more widespread, particularly in New Brunswick (Fowler 1885), prior to habitat alterations in the 20th century, including the construction of the Grand Falls Dam in 1928. Following decades without confirmed observations, *P. furbishiae* was declared probably extinct (Secretary, Smithsonian Institution 1975). It was then 'rediscovered' in 1976 during environmental surveys for the proposed Dickey-Lincoln hydroelectric project (Richards 1976), and concerns over its protection became a significant factor in the eventual cancellation of the project. This outcome elevated the species as a symbol of conservation action and led some to call it a Lazarus taxon. Although there are no reports highlighting the ecological or cultural benefits of this species, possibly due to its extreme rarity and limited study in the last century or so, it is well-known and a source of pride among many local communities of the Wolastoq watershed as the most threatened and iconic species among the numerous rare plants occurring across the US-Canada border. Because of the level of biodiversity within the watershed, various protection efforts to conserve this rare ecosystem and the species therein have been



implemented (*Wolastoq Priority Place - Wolastoq/Saint John River Priority Place*).

Pedicularis furbishiae was one of the first eleven plant taxa listed as federally endangered under the United States Endangered species act of 1973 (U.S. Fish & Wildlife Service 1978). Surveys conducted in 2016–2017 reported that most of the subpopulations above Dickey, Maine (Fig. 1, near 3ME-DB) appeared stable, but the nine downriver subpopulations in river segments 2–4 have declined since surveys began in the 1980s, having lost resilience to withstand stochastic ice scour and flooding events (U.S. Fish & Wildlife Service 2018). In 2023, the species was downlisted from Endangered to Threatened under the U.S. Endangered Species Act (U.S. Fish & Wildlife Service 2023), acknowledging improved protections at some sites. However, the species remains listed as State Endangered in Maine, and recent observations indicate that subpopulations are in general decline (E. Doucette, pers. comm.). In Canada, *P. furbishiae* has been designated as Endangered since 1980 (Stirret 1980) and was listed as such under Schedule 1 of the Species at Risk Act in 2003. Although a recovery strategy has been in place in Canada since 2010, the species has experienced significant demographic declines (Table 1). A 2025 census revealed that two New Brunswick subpopulations are now reduced to a few reproductive individuals (Medford, $n=1$ and Stirret Preserve, $n=3$). Two more subpopulations, Aroostook and Big Flat, which had two individuals in 2018 (Aroostook) and 24 individuals in 2019 (Big Flat) were extirpated in 2020 and 2025 respectively (M. Williams, unpublished).

The life history traits of *P. furbishiae* combined with changing disturbance regimes thus appear to have caused a breakdown in historical metapopulation dynamics and the active collapse of subpopulations (U.S. Fish & Wildlife Service 2018). These trends underscore the need to understand patterns of genetic diversity and population structure in order to maintain the species' evolutionary potential and guide effective recovery planning (Frankham 2005).

Previous assessments of genetic variation using isozyme markers found no polymorphism among 28 individuals (Waller et al. 1987), and thus no evidence of population structure. However, advances in sequencing technology offer greater power for a re-evaluation of genetic diversity and population differentiation across the species' range. Such insights are essential for informing in situ and ex situ conservation efforts, including seed banking, propagation, and enrichment planting—particularly as the species faces continued habitat loss and the collapse of natural recolonization dynamics.

In this study, we assembled a draft genome of *P. furbishiae* and conducted reduced-representation

Table 1 Geographic information and population trends of *P. furbishiae* subpopulations used in this study

Subpopu- lation ID	Site name	Province/ State	Status – N° indi- viduals in 2025 ^a	Coordinates ^b
1ME-BB	Big Black	Maine	Good – unknown	46.96*°/–69.43*°
2ME-CA	Castonia	Maine	Good – unknown	47.06*°/–69.27*°
3ME-DB	Dicky Bridge	Maine	Poor – unknown	47.11*°/–69.12*°
4ME-PB	Pelletier Brook	Maine	Fair – unknown	47.09*°/–69.01*°
5NB-GF	Grand Falls	New Brunswick	Fair- Increasing – 258	47.06*°/–67.78*°
6NB-MD	Medford	New Brunswick	Poor-Near extirpation – 1	46.88*°/–67.69*°
7NB-SP	Stirret Preserve	New Brunswick	Poor-Near extirpation – 3	46.84*°/–67.69*°
8NB-BF	Big Flat	New Brunswick	Extirpated – 0	46.82*°/–67.71*°
9NB-AR	Aroostook	New Brunswick	Extirpated – 0	46.81*°/–67.71*°

^aMaine site status from U.S. Fish and Wildlife Service (2018). New Brunswick status and census counts from 2025 surveys by M. Williams

^b Due to the endangered status of the species, coordinates have been approximated to avoid identifying the exact site locations

genotyping-by-sequencing (GBS) across individuals sampled throughout the species' range. Our objectives were to: (1) assess levels of genetic diversity and inbreeding across subpopulations; (2) characterize patterns of genetic structure and connectivity between New Brunswick and Maine; and (3) evaluate the implications of these patterns for conservation management. We specifically consider the potential for genetic rescue through outcrossing and the development of ex situ seed sources for enrichment planting. By establishing genomic resources and baseline data for evolutionary conservation biology (Ferrière et al. 2004), we aim to support evidence-based decision-making in the ongoing recovery of this endangered species.

Methods

Genome assembly

Leaf tissue was sampled from a 2-year-old greenhouse grown seedling out planted in a secure field bank. The seed used to produce the plant originated from a seed lot collected at the Grand Falls site in New Brunswick (5NB-GF).

The leaf tissue was collected fresh and flash frozen directly in liquid nitrogen and stored at -80°C upon return from the field.

High molecular weight DNA was extracted from ground leaves using the Circulomics Plant Kit (Pacific Biosciences). Sequencing libraries were then made using the ligation sequencing kit SQK-LSK110 (Oxford Nanopore). Libraries were sequenced with eight sequencing runs using five different PromethION flowcells model FLO-PRO002 with R9.4.1 chemistry. Short read libraries were constructed using the NxSeq AmpFREE low DNA library kit (VWR: CA10837-144) and sequenced on an Illumina NovaSeq S4 flowcell, 2×150 bp. The long reads were assembled using the flye assembler (Kolmogorov et al. 2019) followed by four rounds of polishing with default parameters on the Illumina short reads using Pilon (Walker et al. 2014). We then ran `purge_dups` (Guan et al. 2020) using the default cutoffs to remove duplicated contigs.

We then searched and filtered the assembly for contigs from contaminant organisms. FCS-GX (Astashyn et al. 2024) removed 933 contigs with the EXCLUDE flag (mainly proteobacteria and CFP group bacteria) and 627 contigs with the REVIEW flag (mainly insects and alveolates). We then ran Kraken2 v2.1.3 (Wood et al. 2019) using its complete database on the remaining contigs and dropped an additional 3 contaminated contigs. Finally, genome completeness was assessed using BUSCO with eudicots_odb10 (Manni et al. 2021).

Subpopulation sampling and genotyping-by-sequencing

We sampled leaf tissue from a total of 123 individuals from four sites in Maine, USA (1ME-BB, 2ME-CA, 3ME-PB, 4ME-DB) and five sites in New Brunswick, Canada (5NB-GF, 6NB-MD, 7NB-SP, 8NB-BF, 9NB-SP) in 2016, 2018, and 2019, prior to the current decline described in Table 1. All tissue sampling was done with species-at-risk permits obtained from the province of New Brunswick (NB SAR18-027 and SAR19-020) and permissions from local landowners in Maine and the Maine Natural Areas Program (pers. comm. D. Cameron to PRN). No plants were sampled from United States federal lands. DNA was isolated from silica-dried or frozen leaves with the Plant DNeasy kit (QIAGEN, Hilden, Germany) and 175–1,800 ng of DNA per sample was sent to the University of Wisconsin Biotechnology Center DNA Sequencing Facility for genotyping-by-sequencing (GBS) library preparation following the Elshire et al. (2011), and sequencing. Samples were digested with the ApeKI restriction enzyme, ligated with single-indexed adapters, and sequenced (paired-end 150 bp) on the Illumina NovaSeq 6000.

Sequence data preparation, genotyping and filtration were done using scripts from STACKS v2.62 (Rochette et al. 2019) and `stacks_workflow` v2.62 (https://github.com/e-normandeau/stacks_workflow). Briefly, `cutadapt` v1.18 (`-e 0.2 -m 50`) was used to trim the paired-end reads for quality. Samples were demultiplexed with `process_radtags` v2.62 (`-c -r -t 120 -q -s 0 --barcode_dist 1 2 -E phred33 --renz_1 pstI --renz_2 mspI`). We then aligned the raw reads to the genome assembly using `bwa` v0.7.17-r1188 (`-k 19 -c 500 -O 0,0 -E 2,2 -T 0`; Li and Durbin 2009) and `samtools` v1.8 (`-Sb -q 1 -F 4 -F 256 -F 2048`; Li et al. 2009) and then ran `gstacks` (`--max-clipped`) and `population` (`-p 1 -r 0.5 --ordered-export --fasta-loci --vcf`) to call SNPs.

After these STACKS steps, we used `stacks_workflow` to filter the SNPs as follows: We first filtered the SNPs (`05_filter_vcf_fast.py`, `params: 3 80 0 3`) so that all the genotypes had a minimum coverage of 3 and kept only SNPs for which all the samples, considered as one group, had at most 20% of missing data and for which at least 3 samples had the minor allele. This dataset was used to get the proportion of missing data per sample. Samples with more than 30% of missing data (18 samples) were identified and removed, and the `vcf` was filtered again in the same manner (`05_filter_vcf_fast.py`, `params: 3 70 0 2`), so that all the genotypes had a minimum coverage of 3 and we kept only SNPs for which all the samples had at most 30% of missing data and for which at least 2 samples had a copy of the minor allele. Then, we evaluated pairwise relatedness (`vcftools --relatedness`) and flagged one pair with a high value of 1.18—and flagged the one with the most missing data for removal. We assessed outliers for heterozygosity (`vcftools --het`) found no aberrant individuals. With 19 individuals identified for removal, we backtracked and excluded these from the original VCF and applied one clean round of SNP filtering (`05_filter_vcf_fast.py`, `params: 3 70 0 2`).

Next, we used a modified HD plot approach from Mackinney et al. (2017), to remove SNPs displaying signs of paralogy or over-merging (scripts 08, 09 and 10 from `stacks_workflow`) and retaining only canonical SNPs, referred to as singletons. In this workflow, we first gathered the following statistics from the SNPs: median allelic ratio of heterozygotes, proportion of heterozygotes, proportion of homozygotes with the minor allele, global *Fis*, median genotype coverage of homozygotes, and median genotype coverage of heterozygotes. All these parameters were plotted in a pairwise data in R and the procedure from Mackinney et al. (2017) was used to identify paralogs. SNPs with an *Fis* value lower than -0.4 were deemed as duplicated and those with an *Fis* value lower than -0.8 as diverged, meaning that they are duplicated but that the two separate copies each have two identical alleles. SNPs with median genotype coverages above 20 for homozygotes and 30 for heterozygotes

were also removed as potential duplicates. SNPs with an *Fis* value above 0.9 were deemed of low confidence since they displayed a very high deficit in heterozygosity. SNPs with medium allelic ratios above 0.8 or below 0.2 were also filtered as low-confidence SNPs. After this, we had 4400 SNPs and 104 individuals. However, this SNP dataset showed several artifacts, most notably a negative relationship between observed heterozygosity and depth of coverage, which we then addressed through additional filtering.

To reduce biases in heterozygosity caused by depth of coverage, we iteratively tested SNP filtering parameters using both our STACKS dataset and an independently assembled PyRAD dataset (Eaton 2014; Eaton and Overcast 2020). We evaluated minimum depth thresholds (six or eight), a minimum allele count of two to four ($\approx$ MAF 0.05), and per-individual missing data thresholds ($>50\%$), assessing performance with ADMIXTURE likelihoods, heterozygosity versus depth of coverage, and estimates of *Fis*.

Our final filters—MAC ≥ 4 , minimum depth ≥ 6 , and exclusion of individuals and SNPs with $>60\%$ missing data—substantially reduced heterozygosity bias with respect to coverage (Fig. S2) and produced consistent population structure across independent assemblies. The MAC threshold had the largest impact on ADMIXTURE, with low-frequency alleles introducing clustering instability; however, the final dataset retained numerous private alleles within subpopulations, preserving fine-scale structure. After filtering, we obtained 152 SNPs and 69 individuals (Fig. S3).

Population structure

Principal coordinates analysis (PCoA) was performed on the full dataset using the dartR package (`gl.pcoa()` function; Gruber et al. 2018) and plotted 90th percentile ellipses around each subpopulation (`gl.pcoa.plot()` function). PCoA distances were calculated using loci shared between individuals, with missing genotypes excluded pairwise. We used ADMIXTURE (Alexander et al. 2009) to investigate the number of ancestral populations and admixture proportions across sampled subpopulations. Here, we first used PLINK (`--indep-pairwise`; Purcell et al. 2007) to prune SNPs in high linkage disequilibrium within a sliding window of 100 kbp, step size of 5 kbp, and r^2 threshold of 0.2. This step removed only one SNP. The number of ancestral populations were determined via the ADMIXTURE cross-validation program with major convergence criterion $10e^{-6}$ and minor convergence criterion of $10e^{-13}$. The different ADMIXTURE tests for $K=1-10$ were run with GNU Parallel (Tange 2018). All population statistics, structure, and isolation-by-distance scripts are available at <https://github.com/Erythroxyllum/furibishiae>.

We inferred a distance-based phylogenetic network using SplitsTree CE 6.0.0 (Huson and Bryant 2006). The P Distance method (Hamming 1950) was used (default options) so as to obtain a 69×69 editDistance matrix. The Neighbor Net method (Bryant and Moulton 2004; Bryant and Huson 2023) was used (default options) so as to obtain 156 splits, cyclic. The Show Splits method was used (default options) so as to obtain a Split Network visualization.

Isolation-by-distance analysis

To assess isolation-by-distance (IBD), we conducted Mantel tests comparing pairwise Slatkin's linearized *Fst* values [*Fst*/(1-*Fst*)] with geographic distances among subpopulations. We calculated pairwise *Fst* among the seven subpopulations using the hierfstat R package (Goudet 2005) and the full SNP dataset and calculated pairwise geographic distances among subpopulations as great-circle distances with the Haversine method (geosphere R package; Hijmans 2024). The Mantel statistic reported is Pearson's product-moment correlation coefficient (r), which quantifies the linear association between genetic and geographic distance matrices. We conducted tests using both raw and natural log-transformed geographic distances to evaluate the sensitivity of results to spatial scale. To assess regional patterns of IBD, we performed separate Mantel tests for all subpopulations, for the four Maine (ME) subpopulations, and for the three New Brunswick (NB) subpopulations. We also visualized IBD relationships using linear regression of pairwise genetic vs. geographic distances across all subpopulations and also within and across regions (ME–ME, ME–NB, NB–NB, and All). Figures were created with ggplot2 (Wickham 2016) and combined using patchwork v1.3.2 (Pedersen 2025).

Genetic diversity and differentiation

We used the R package hierfstat (Goudet 2005) with the full SNP dataset to calculate observed heterozygosity within inferred populations and to assess within-population gene diversity (following Nei 1987). To estimate nucleotide diversity (π), we analyzed the assembled GBS loci sequences retained in the final SNP dataset. Ambiguous heterozygous sites were resolved through a random phasing approach, in which alleles were randomly assigned to two pseudo-haplotypes per individual. This process was repeated across 1,000 iterations, with nucleotide diversity calculated for the full set of phased haplotypes in each replicate. We calculated the inbreeding coefficient (*Fis*; Nei 1987) and pairwise *Fst* across 1000 bootstraps using hierfstat to generate confidence intervals for these indices. We calculated the number of private alleles in each population using vcf-contrast (Danecek et al. 2011).

Whole genome demographic inference and heterozygosity

We used the Illumina paired-end whole-genome sequencing data together with our genome assembly to reconstruct historical demography with the Pairwise Sequentially Markovian Coalescent (PSMC). Raw Illumina reads were adapter- and quality-trimmed with fastp v0.23.4 (`-qualified_quality_phred 20, -cut_mean_quality 20, -trim_poly_g, -trim_poly_x, -length_required 35, -dedup`; Chen et al. 2018). Cleaned reads were mapped to the reference genome with bwa-mem2 v2.2.1 (Vasimuddin et al. 2019) using default parameters, and alignments were sorted and marked for duplicates using samtools v1.20 (Li et al. 2009). We filtered BAMs to retain only primary, mapped, non-duplicate reads (samtools view `-F 3844`). Depth distributions were calculated with samtools depth v1.20, and a BED mask of outlier regions (coverage < 10 or > 60) was generated using bedtools v2.31.1 (Quinlan and Hall 2010). We used an average effective coverage depth of $31.2\times$ for snp calling.

Variants were called with bcftools v1.20 (`mpileup -q 20 -Q 20 -C 50 | bcftools call -mv`), and an IUPAC-coded diploid consensus sequence was built with bcftools consensus v1.20 using the coverage mask (Li et al. 2009; Danecek et al. 2021). The consensus FASTA was converted to FASTQ with seqtk v1.4 (Li 2025; `seqtk seq -F I`) and FASTQ scaffolds less than 20,000 bp in length were removed to improve model performance. Diploid FASTQs were converted to PSMC input format with fq2psmcfa (part of psmc v0.6.5; McVean and Cardin 2005; Li and Durbin 2011) using `-q20 -g10000`. The PSMC analysis was run with the standard parameter set (`psmc -N25 -t15 -r5 -p "4+25*2+4+6"`), and with other parameter sets (conservative `"18+16*2+4+4"`, recent-focused `"10*1+14*2+4*4+10"`, and aggressive-recent `"15*1+10*2+4*4+13"`) to test for estimation artifacts in most recent time windows (Hilgers et al. 2025). Results were plotted using our estimated seed-to-seed generation time of three years and the mutation rate of 7×10^{-9} based on rates for *Arabidopsis thaliana* (Weng et al. 2019). Bootstrap replicates were generated with the same parameters. Plots were created with psmc_plot.pl. Scripts for this demographic inference pipeline are available at <https://github.com/Erythroxylum/psmc-pipeline>. In addition to demographic inference, we estimated genome-wide heterozygosity by masking non-callable regions of the assembly (BEDTools v2.31.1; Quinlan and Hall 2010) and restricting variant calls to callable bases. Biallelic SNPs were extracted from the whole-genome VCF using BCFtools v1.20 (Danecek et al. 2021), and heterozygous and homozygous-ALT genotypes were tallied with bcftools query. Heterozygosity was then expressed as the proportion of callable sites

that were heterozygous, scaled per site, using custom Bash and Python scripts (available at <https://github.com/Erythroxylum/genome-wide-Ho>).

Results

Reference genome assembly and genotyping individuals

Nanopore sequencing produced 65Gbp with 8 M reads (average read length: 7Kb) and short read sequencing produced 750 M read pairs. With the contaminant contigs removed, the final polished assembly reached a total size of 3.2Gbp, with an NG50 of 447Kbp in 1653 contigs. The deduplicated assembly reached 2.7 Gbp. The BUSCO result shows 88% complete single copy orthologs and 95% complete single or duplicated orthologs (C:2213 [S:2056, D:157], F:16, M:97, n:2326).

Genotyping-by-sequencing (GBS) yielded an average of 5.1 M reads per individual, ranging from 2.0 to 7.4 M reads. The average proportion of reads successfully mapped to the reference genome was 0.40, ranging from 0.21 to 0.67. After quality controls, the GBS permitted confident genotyping of 69 individuals for 152 SNPs, 151 of which were determined to be unlinked. This dataset had 23% missing data. The raw sequence reads, genome assembly, and GBS data are available on NCBI via BioProject PRJNA1322473.

Population structure

Cross-validation errors across ADMIXTURE models of $K=1-10$ are lowest for $K=4$, followed by $K=7$ (Figs. S4, S5). The genetic clusters at $K=4$ are structured geographically, with one in Maine (comprised of subpopulations 1ME-BB, 2ME-CA, 3ME-PB, and 4ME-DB) and the other three representing the distinct subpopulations in New Brunswick: 5NB-GF, 6NB-MD, and 8NB-BF (Fig. 2a). Individuals in Maine show small ($< 14\%$) admixture proportions with the 8NB-BF ancestral genotype and the 6NB-MD ancestral genotype. Five of the individuals in the 5NB-GF population show admixture proportions up to 48% with up to all three of the other ancestral genotypes (Fig. 2a). Three individuals in the 6NB-MD population show small admixture proportions ($< 15\%$) with 5NB-GF and 8NB-BF, and the 8NB-BF population shows no admixture with the other populations. The $K=7$ model delimits each subpopulation as an ancestral genetic group with two exceptions: individuals in 2ME-CA are modeled as hybrids between 1ME-BB and 3ME-PB, and 8NB-BF has two ancestral groups (Fig. 2a).

The first and second axes (Fig. 2b) of the principal coordinates analysis represent over 35% of the total genetic

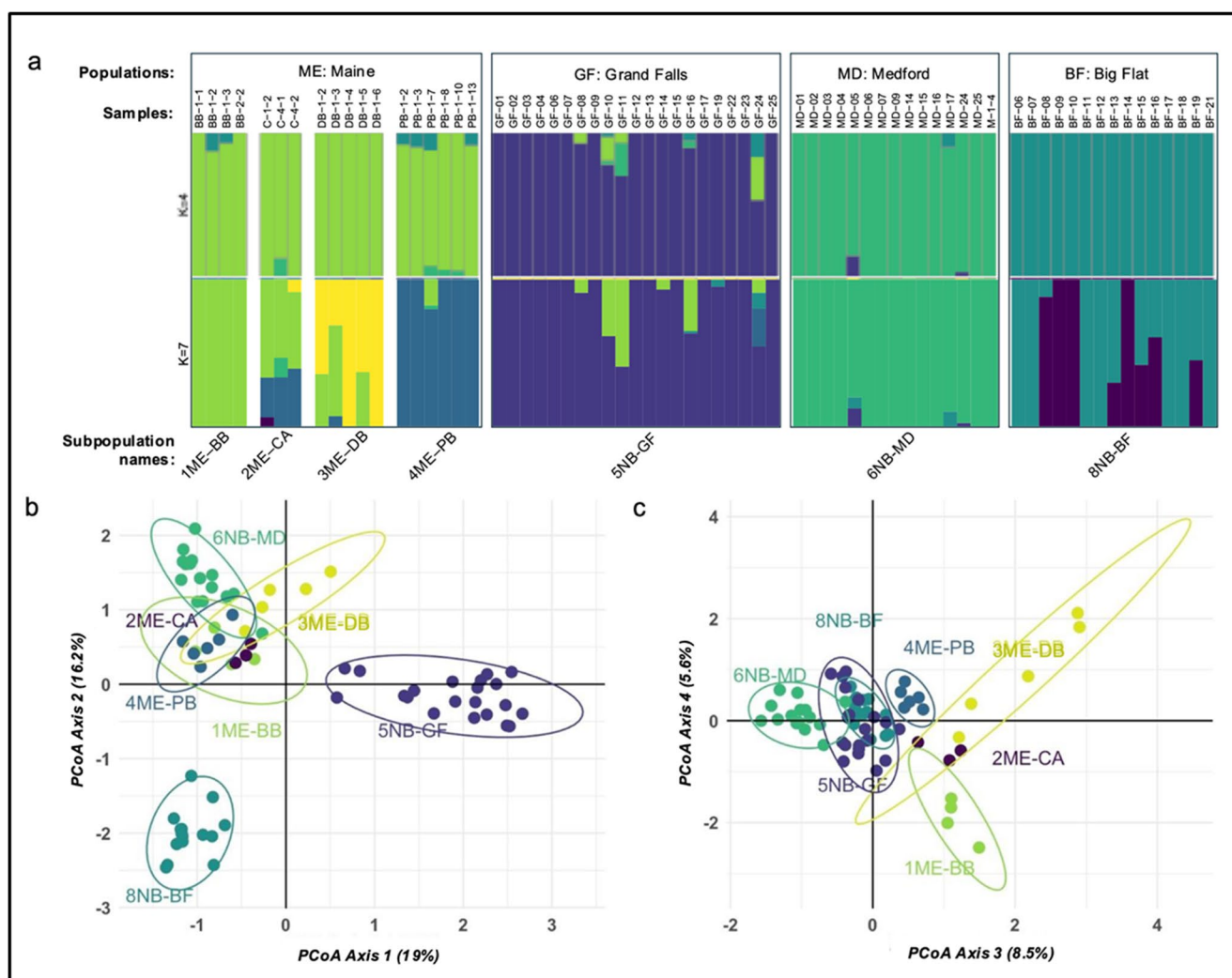


Fig. 2 The population structure of *P. furbishiae*. The ancestral genetic clusters inferred at $K=4$ group the Maine subpopulations together into a single genetic population and the three New Brunswick subpopulations into three genetic populations (panel a). The first and second axes

(panel b) and third and fourth axes (panel c) of the principal coordinates analysis also reveal differentiation and genetic clustering corresponding with geographic subpopulations. Plots showing additional axes are available in Fig. S6

Table 2 Population differentiation. Confidence intervals (95%) of F_{st} values from 1000 bootstrap pseudoreplicates

Population	8NB-BF (Big Flat)	5NB-GF (Grand Falls)	6NB-MD (Medford)
5NB-GF (Grand Falls)	0.11–0.20	-	-
6NB-MD (Medford)	0.22–0.36	0.04–0.09	-
Maine (all)	0.23–0.37	0.18–0.30	0.24–0.35

variance and separate the three New Brunswick subpopulations. The third and fourth principal coordinate axes (Fig. 2c) together explain 14.1% of the genetic variance and separate three of the Maine subpopulations and 6NB-MD. The 2ME-CA subpopulation is not distinct along these multivariate axes and is nested or similar to all other ME subpopulations.

Pairwise genetic differentiation (F_{st}) is moderate to high across populations (Table 2). The greatest differentiation is observed between Maine and New Brunswick subpopulations ($F_{st}=0.18–0.37$) and between 8NB-BF and 6NB-MD ($F_{st}=0.22–0.36$), while the lowest differentiation is between 5NB-GF and 6NB-MD ($F_{st}=0.04–0.09$), suggesting more recent gene flow or divergence among the New Brunswick populations. Estimation of F_{st} on the subpopulation level also reveals high values: the lowest being $F_{st}=0.15–0.37$ for 2ME-CA and 4ME-PB, and the highest being $F_{st}=0.43–0.62$ for 1ME-BB and 8NB-BF (Table S1). These patterns are consistent with the geographic structuring observed in the ADMIXTURE and PCoA analyses (Figs. 2; S6).

The NeighborNet network inferred with SplitsTree explains 97.8% of the variance in the pairwise distance

matrix ($Fit=97.8$), indicating a close correspondence between observed genetic distances and the inferred splits (Fig. 3). Samples are clearly clustered into their respective subpopulations and further grouped by genetic population structure. One new pattern observed is the larger P distance among 3ME-DB individuals indicated by long branches.

Genetic diversity

The genetic diversity of *P. furbishiae* is very low, with pairwise nucleotide diversity (π) estimated at 0.030 and genome-wide heterozygosity from the resequenced individual estimated at 4.47×10^{-4} . Yet, all four focal populations retain private alleles, revealing the persistence of distinct

genetic variants despite overall low diversity. The Maine population (ME) exhibits the highest number of private alleles ($PA=40$, $N=18$), followed by 5NB-GF ($PA=18$, $N=21$), 6NB-MD ($PA=12$, $N=15$), and 8NB-BF ($PA=8$, $N=15$; Table 3).

Observed heterozygosity (H_o) ranges narrowly from 0.12 (6NB-MD) to 0.15 (8NB-BF), whereas gene diversity (H_s) is highest in 5NB-GF ($H_s=0.22$) and lower in the remaining populations ($H_s=0.18$ in ME, 6NB-MD, and 8NB-BF). These H_e and H_o estimates are based on the reduced SNP dataset after stringent filtering and are therefore best interpreted in a comparative rather than absolute sense. The subpopulations in Maine show lower gene diversity, with H_s ranging from 0.11 to 0.18 (Table 3). Inbreeding coefficients

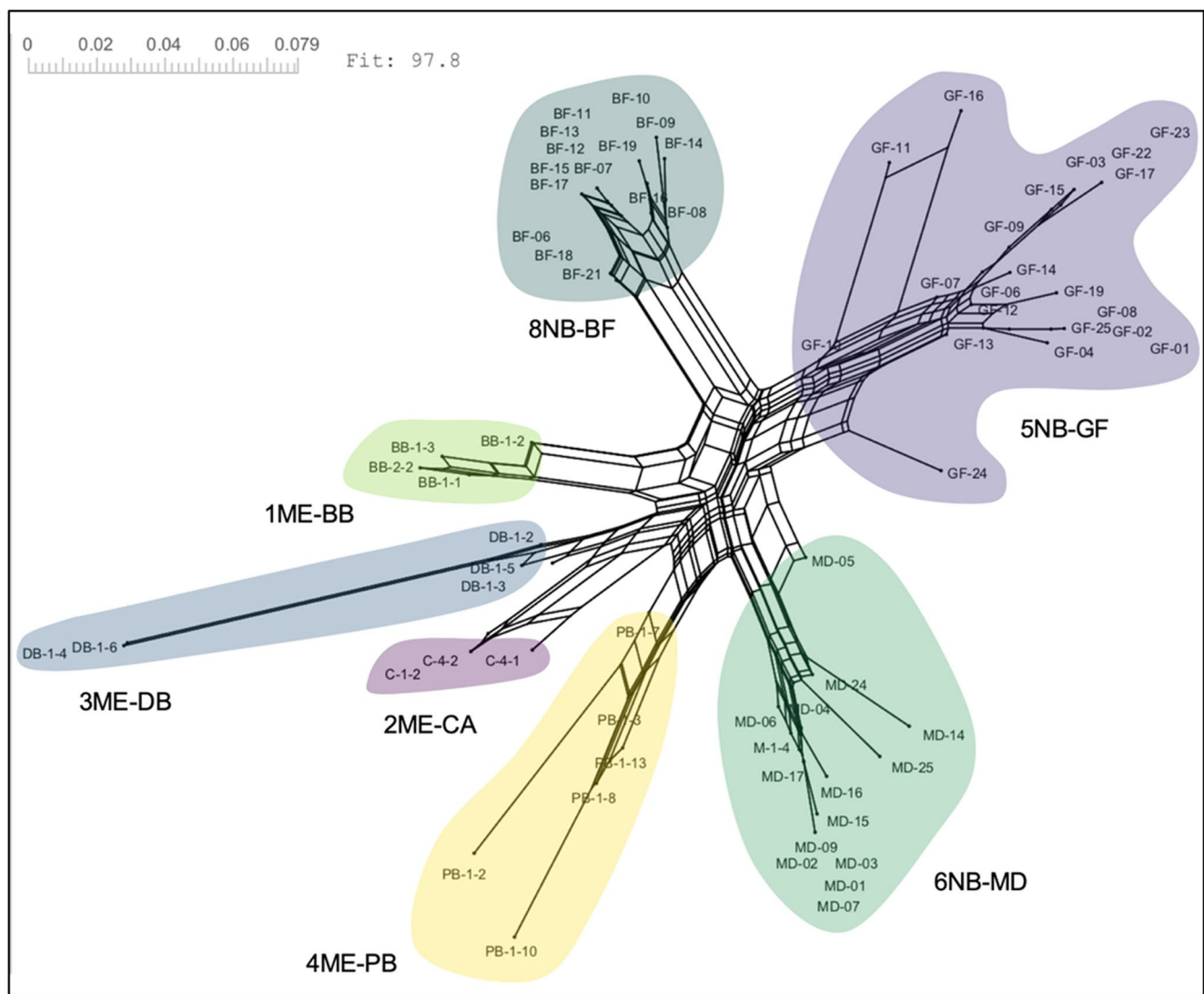


Fig. 3 SplitsTree phylogenetic network inference of *P. furbishiae* individuals colored by subpopulation. The scale bar represents uncorrected p -distance (proportion of nucleotide differences per site) between samples. The Fit value (97.8) indicates the percentage of variance in the

distance matrix explained by the network; higher values reflect a closer correspondence between the observed pairwise distances and the splits network

Table 3 Comparative genetic diversity statistics among *P. furbishiae* populations and subpopulations (N=number of individuals; PA=private alleles; H_o =observed heterozygosity; H_s =gene diversity; F_{is} =inbreeding coefficient 95% confidence intervals from 1000 bootstrap pseudoreplicates)

Population	N	PA	H_o	H_s	F_{is}
Maine (all)	18	40	0.13	0.18	0.14–0.38
1ME-BB	4	3	0.09	0.11	−0.2–0.21
2ME-CA	3	0	0.10	0.12	−0.36–0.37
3ME-PB	6	10	0.18	0.18	−0.32–0.14
4ME-DB	5	12	0.13	0.14	−0.42–0.8
5NB-GF	21	18	0.13	0.22	0.32–0.49
6NB-MD	15	12	0.12	0.18	0.26–0.44
8NB-BF	15	8	0.15	0.18	0.00–0.31

Table 4 Mantel test of significant effect of isolation-by-distance

Comparison	n (pairs)	Mantel r	p-value	log-transformed Mantel r	log p-value
All subpopulations	21	−0.03	0.56	−0.01	0.52
ME subpopulations	6	0.68	0.17	0.64	0.21
NB subpopulations	3	−0.93	1.00	−0.87	1.00

(F_{is}) vary among populations; the lowest is in 8NB-BF (F_{is} =0.00–0.31), suggesting minimal to moderate inbreeding. By contrast, 5NB-GF and 6NB-MD show substantial heterozygote deficiency, with F_{is} ranging from 0.32 to 0.49 and 0.26–0.44, respectively. The ME population exhibits intermediate F_{is} values (0.14–0.38), also consistent with low to moderate inbreeding.

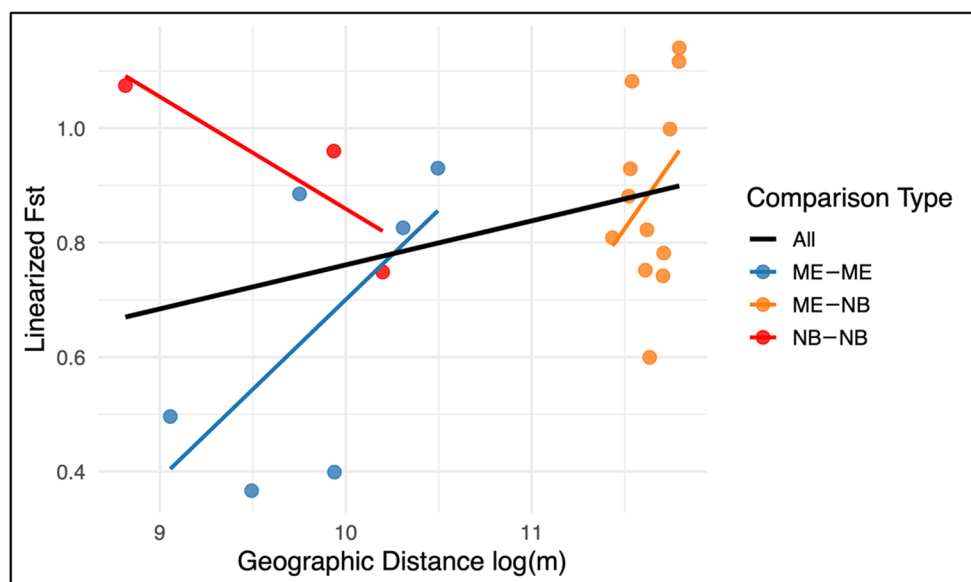
The $K=7$ ADMIXTURE model and the PCoA show subpopulation structure within Maine (Fig. 2a). Population genetics statistics within Maine reveal variability in genetic diversity and inbreeding among sites. Observed heterozygosity (H_o) ranges from 0.09 in 1ME-BB to 0.18

in 3ME-PB, whereas gene diversity (H_s) is generally low, ranging from 0.11 to 0.18 across sites. The highest number of private alleles is found in 4ME-DB (PA=12) and 3ME-PB (PA=10), suggesting some genetic distinctiveness at these sites despite their small sample sizes. Inbreeding coefficients (F_{is}) vary widely among Maine subpopulations, with negative to near-zero values at 1ME-BB (−0.20 to 0.21) and 3ME-PB (−0.32 to 0.14), indicating Hardy-Weinberg equilibrium or a slight excess heterozygosity, while more variable and positive estimates are found at 2ME-CA (−0.36 to 0.37) and especially 4ME-DB (−0.42 to 0.80), which show the widest range, potentially reflecting a mix of related individuals or hidden structure at that site.

Isolation by distance

We found no significant evidence of isolation by distance when analyzing all seven subpopulations (Mantel $r = -0.03$, $p=0.56$; log-transformed $r = -0.01$, $p=0.52$; Table 4; Fig. 4). However, when analyzed separately, the four ME subpopulations show a moderate positive Mantel correlation ($r=0.68$), although not statistically significant ($p=0.17$), suggesting a possible IBD pattern at this scale. By contrast, the three NB subpopulations show a negative correlation ($r = -0.93$) that lacks statistical significance ($p=1.0$; Table 4). These findings indicate that geographic distance is not a strong predictor of genetic differentiation across the full range. Linear regression lines are shown in Fig. 4 to illustrate pairwise trends, but these are not equivalent to the Mantel r statistics, which measure matrix-wide correlations and are evaluated via permutation tests. These patterns are consistent with the Maine subpopulations functioning as a metapopulation.

Fig. 4 Pairwise isolation by distance (IBD) patterns among *Pedicularis furbishiae* subpopulations. Pairwise linearized F_{st} values are plotted against geographic distances (in meters) between subpopulations. Lines show linear regressions for each regional comparison group: ME–ME (blue), ME–NB (orange), NB–NB (red), and all pairwise comparisons combined (black). Note that the linear regression lines shown here are for visualization only and are not equivalent to the Mantel r statistics in Table 4, which are based on matrix correlations using permutation tests



Demographic inference

The PSMC model provided a robust estimate of demography between ca. 1,100–200,000 YBP (Fig. 5), in which an effective population size (N_e) of ca. 17,000 was relatively stable during a period of 100,000–200,000 YBP, followed by a decline during the period synchronic with the Wisconsin glaciation (75,000–11,000 YBP), when N_e dropped to ca. 2,800 by ca. 13,000 YBP. The model then infers a slight increase in N_e to ca. 4,000 at ca. 6,000 YBP followed by a drop to $N_e=2,800$ at ca. 2,200 YBP. Bootstrap pseudoreplicates ($n=100$) closely match the primary inference. Models across the four different parameters were remarkably consistent.

Discussion

Historical and evolutionary context

Our genomic analysis provides new insights into the past, present, and precarious future of *Pedicularis furbishiae*. Prior phylogenetic studies showed it to be most closely related to western North American alpine taxa (*P. bracteosa* and *P. rainierensis*, the latter another narrow endemic) and the boreal species *P. capitata*, in a lineage that migrated from Asia across Beringia (Robart et al. 2015). It thus appears that *P. furbishiae* has arctic-alpine ancestry and an evolutionary history intimately linked to the glacial cycles

of the Quaternary, during which it may have evolved its particular and very specific niche requirements. Our inferences of historical demography indicate that its population size declined and has remained low since the last glacial maximum (Fig. 5), perhaps due to reductions in the availability of ice-scoured riverbanks driven by forest succession and hydrological changes in the current interglacial period. The cumulative evidence suggests that *P. furbishiae* is a glacial relict, surviving tenuously in an ecologically fragile habitat.

Contemporary genetic structure

Our population genomic results reveal how this historical backdrop and the species ecological characteristics have shaped contemporary genetic structure. The species maintains low genetic diversity across its range that is structured into four genetically isolated populations: one in Maine and three in New Brunswick (5NB-GF, 6NB-MD, 8NB-BF). We also detect substructure among Maine sites. These findings build upon an earlier genetic study based on isozymes that did not detect any polymorphism (Waller et al. 1987) and underscore the value of genomic approaches. Low genetic diversity, however, does not by itself equate to reduced fitness. While it constrains adaptive potential, persistently small effective population sizes can facilitate purging of deleterious alleles, reducing genetic load (Robinson et al. 2022). Thus, low diversity in *P. furbishiae* likely reflects its long-term demographic history rather than imminent genetic collapse.

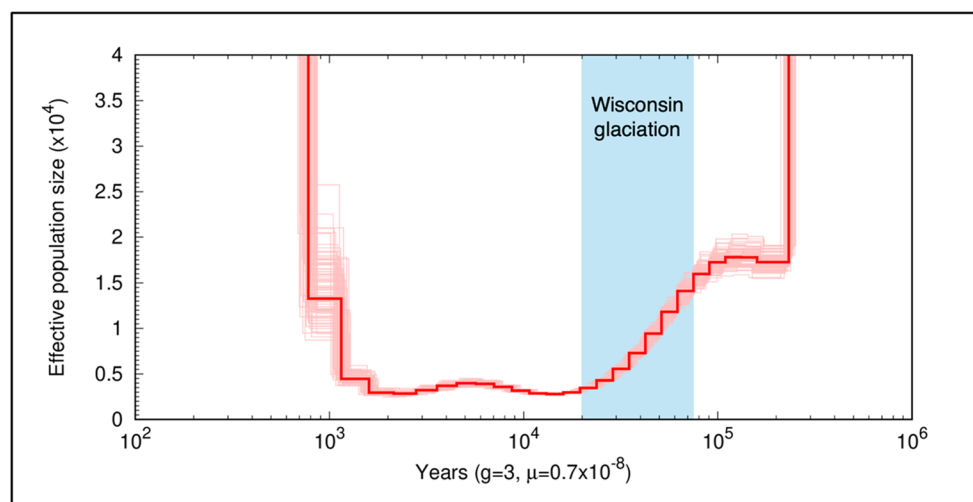


Fig. 5 Demographic history of *Pedicularis furbishiae* inferred with the Pairwise Sequentially Markovian Coalescent (PSMC) under the standard parameter set ($-p\ "4+25*2+4+6"$). Analysis was performed using Illumina whole-genome resequencing data mapped to the *P. furbishiae* reference genome. Time is scaled with a generation rate of three years and a mutation rate of 7×10^{-9} . The bold red line indi-

cates the main PSMC run and the light red lines plot the 100 bootstrap runs. The period of the Wisconsin glaciation is indicated in light blue, with the Wolastoq becoming ice-free 1,100–1,400 YBP (Dalton et al. 2020). Spikes in effective population sizes (N_e) prior to 1,100 YBP and beyond 150,000 YBP reflect methodological limits rather than robust demographic inferences

Maine metapopulation dynamics versus New Brunswick breakdown

Our results support a model in which this early-successional specialist has historically persisted as a dynamic metapopulation, with ephemeral subpopulations ‘blinking in’ to connect others before ‘blinking out’ as they were replaced by ecological succession or stochastic hydrological events (Marsh and Trenham 2001). In Maine, several lines of evidence point to functioning metapopulation dynamics. The four subpopulations show relatively low genetic diversity (average $H_s=0.18$; Table 3), yet they cluster together as a single genetic group at $K=4$ in ADMIXTURE, group tightly along the first two axes of the PCoA (Fig. 2) and share reduced branch lengths in the splits network (Fig. 3). Substructure detected at $K=7$ highlights localized differentiation, with 2ME-CA appearing as a hybrid between 1ME-BB and 4ME-PB, suggesting that recent or ongoing gene flow continues to link sites. Moderate F_{st} values among Maine subpopulations (Table S1) together with a positive correlation between genetic and geographic distance further support partial differentiation with some genetic connectivity (Fig. 4). Although the Mantel test did not recover a statistically significant isolation-by-distance pattern (Table 4), likely due to the small number of sites, the overall evidence suggests that Maine subpopulations retain sufficient connectivity to function as a metapopulation.

In contrast, New Brunswick subpopulations show no evidence of such connectivity. Genetic clusters at both $K=4$ and $K=7$ analyses, clear separation in the PCoA (Fig. 2), and long branches in the splits network (Fig. 3) all point to pronounced isolation. Pairwise F_{st} values are high (up to 0.37; Table 2), and the linearized F_{st} and geographic distance relationships and Mantel test showed a negative, non-significant correlation among New Brunswick sites, indicating no predictable relationship between genetic and geographic distance (Table 4; Fig. 4). These findings suggest that natural metapopulation dynamics have effectively broken down in New Brunswick, leaving its few remaining subpopulations genetically isolated and demographically vulnerable.

Gene flow and demographic constraints

Pairwise F_{st} values confirm strong differentiation between Maine and New Brunswick subpopulations (0.18–0.37; Table 2), consistent with limited historical connectivity across the border. ADMIXTURE analyses detected small proportions of shared ancestry in a few individuals (Fig. 2a), but these signals appear unstructured and potentially incidental. Rather than indicating ongoing seed dispersal or colonization, such weak admixture underscores the absence

of effective gene flow and does not support the potential for natural establishment of new sites or replenishment of existing ones.

Once-abundant subpopulations like Stirrett Preserve and Aroostook, which held over 100 and 300 individuals per subpopulation two decades ago, are now either functionally extirpated or contain very few individuals, reflecting the boom-and-bust dynamics typical of species reliant on stochastic disturbance regimes (New Brunswick Department of Natural Resources and Energy Development 2023; Gyllström 2021). New Brunswick subpopulations have experienced further decline and the species is closer to extirpation in the province. The 2025 census revealed that the Grand Falls subpopulation now harbors the majority (98% or 258 plants) of all known New Brunswick individuals, with only 1 individual found in the Medford subpopulation and three other plants are known from the Stirrett Preserve.

We did have one individual from each of the 7NB-SP and 9NB-AR sites pass our filtering steps, but they were excluded from our final analyses due to concerns that human transplanting activities may have caused the observed genetic affinities with Maine genetic groups (Fig. S7). Specifically, this observation was linked to the hypothesis that gravel and seeds from Maine were brought to Aroostook as fill for the railroad, and that the Stirret Preserve subpopulations might have been bolstered by previous, unrecorded enrichment planting with seed sources from Maine.

Natural recolonization of New Brunswick sites now appears highly unlikely. Over the last 20 years or so, no new subpopulations have been found in the province. Seed dispersal is largely limited to floating scapes severed by rabbits or other herbivores, but ADMIXTURE shows little evidence of downstream dispersal (Fig. 2a). The Grand Falls dam further isolates upstream and downstream habitats. Causes of recent NB crashes include ice scour at Big Flat—where the subpopulation site was scoured down to bare rock. At Aroostook and at Medford we have observed succession and alder encroachment which have lead to debris accumulation and plant loss. At Stirret Preserve, soil erosion and alder encroachment has kept the subpopulation at low levels with no signs of expansion even though the remaining plants are mature and have flowered over the years.

Data limitations and strengths

Our current dataset was sufficient to resolve population structure and reveal conservation-relevant patterns, but its scope was limited for analyzing finer-scale landscape genetic processes, hydrological dynamics, and the genomic basis of adaptation. The relatively small number of SNPs retained (152) reflects the stringent filtering needed to obtain a reliable core of loci with confident genotype calls. While much

of the raw dataset was removed due to low coverage, contamination (particularly in some Maine samples collected later in the season), and the species' overall low genetic diversity, the curated dataset nonetheless revealed clear population structure and meaningful within-site variation.

Conservation actions and ecological considerations

Together, these findings call for proactive conservation strategies. The population structure we documented (F_{st} up to 0.37) supports the inclusion of individuals from all subpopulations in field banks to maximize diversity and the potential benefits of subpopulation mixing. Unlike a homogeneous system where seed could be bulked and planted at random, the presence of structure means that field banks should be designed to track flowering and productivity within and across subpopulation cohorts, ensuring opportunities for open pollination and monitoring potential fitness effects. This aligns with established genetic rescue principles, in which mixing distinct but not overly divergent lineages can increase heterozygosity and enhance adaptive potential (Thrall et al. 1998; Frankham et al. 2011). However, as we explain toward the end of this section, we have not actually observed any signs of inbreeding depression. Additionally, while disruption of local adaptation is not an initial concern, seedlot mixing across subpopulations should buffer against outbreeding depression or total population failure. F_{st} results could also guide the proportion of populations to use in a mix (Bucharova et al. 2019).

Our demographic analysis provides further guidance for restoration. Our PSMC analysis suggests that an effective population size of ~3,000 individuals, which was inferred to be a relatively stable population from 11,000 to 1,100 YBP, may represent a demographic target for sustaining metapopulation dynamics in New Brunswick. Although this estimate is based on a single Grand Falls genome, resequencing of Maine individuals would allow comparison and refinement of demographic goals.

Ultimately, the persistence of *P. furbishiae* in the face of climate-driven habitat volatility will depend on reestablishing gene flow and demographic redundancy by increasing habitat connectivity across its range. In the absence of natural recolonization dynamics, this will require conservationists to simulate metapopulation processes through strategic sourcing, propagation, habitat restoration, and enrichment planting. In line with this need, current recovery plans in New Brunswick call for reintroduction at extirpated sites, augmentation of existing stands, and the creation of novel populations along the Wolastoq and potentially other tributaries (Winton et al. 2024).

Fortunately, a conservation program was initiated by Natural Resources Canada in 2018, in collaboration with New

Brunswick partners, to safeguard and restore *P. furbishiae* before the most recent demographic collapse. Seed sources from three Canadian subpopulations (5NB-GF, 6NB-MD, and 8NB-BF), along with material from Maine, were successfully collected and preserved through cryogenic storage. In parallel, secure field banks were established to maintain living collections representing each New Brunswick subpopulation, supplemented with individuals from Maine to ensure redundancy. These field banks now provide a crucial reservoir of viable seed and plants for propagation and future restoration efforts (Fig. 6).

Observations on seed viability and fitness results from this project are encouraging as they do not seem to indicate any inbreeding depression. Seeds collected from the field have proven viable and were the initial plants transferred to the field banks. Conversely, open-pollinated field bank crosses show no evidence of outbreeding depression to date (M. Williams, unpublished) as F1 generation seed sources have successfully germinated and produced viable plants that have themselves matured and produced viable F2 plants, though continued monitoring is essential. These results indicate that ex situ propagation can effectively combat demographic collapse if integrated with in situ restoration.

Over time, these genetically enriched seed sources will serve as the foundation for two key restoration efforts: (1) reinforcing existing subpopulations in viable habitat and (2) establishing new populations in suitable but currently unoccupied sites. Together, these efforts aim to increase subpopulation redundancy along the Wolastoq (Saint John River) and gradually re-establish a self-sustaining metapopulation with increased connectivity. Adaptive management approaches, including planting trials and longitudinal monitoring, will be essential to track performance, detect unintended outcomes, and iteratively refine strategies.

Future directions

At the landscape scale, the persistence of *P. furbishiae* will depend on reestablishing metapopulation dynamics. The genomic resources generated here provide a critical baseline for understanding genetic diversity, ecological processes, and recovery planning. To improve our characterization of metapopulation dynamics and environmental factors such as ice-scour, future work should focus on resequencing individuals from Maine to refine demographic inference, developing environmental and hydrological models affecting *P. furbishiae*'s goldilocks zone between the river's edge and the tree line, and developing chromosome-scale assemblies for functional genomic research. Crossing with closely related *Pedicularis* species could be considered if inbreeding depression emerges.



Fig. 6 Ex situ enrichment program at Natural Resources Canada in Fredericton, New Brunswick. Seeds and capsules from a single mother plant prior to cryogenic preservation (panel a). Propagation of *P. furbishiae* plants with characteristic dissected leaves grown from seed

with its chosen host showy tick trefoil (*Desmodium canadense*; panel b). After 5 weeks, seedlings are transferred from their individual cells into 4 inch square pots (panel c), before being transferred to the field bank (panel d)

Because natural recolonization is constrained by limited dispersal and the Grand Falls dam, reintroduction efforts in New Brunswick will need to simulate natural connectivity through targeted planting and seed sowing. Because the Grand Falls dam represents a major barrier to downstream dispersal, replanting seeds just below the dam could help re-establish natural propagule flow by allowing seeds to float to downstream New Brunswick sites, simulating the species' historical dispersal pathway. Time-series maps of riverbank disturbance and scouring intensity would greatly aid this process by identifying patches of recently created habitat unlikely to be scoured again in the near term. Further

landscape genetic analyses could help quantify typical dispersal distances and identify barriers or corridors to gene flow, thereby informing the optimal placement of future restoration and reintroduction sites. In addition, host plant availability is likely to influence colonization success. While propagation experiments show that *P. furbishiae* can survive on a variety of hosts and even without one, field associations remain poorly understood. Restoration efforts may benefit from pairing transplants with compatible hosts that are also native to the habitat such as *Desmodium canadense*, which could improve establishment and long-term survival in disturbed habitats.

Conclusion

Our study shows that genetic variability in the species is largely structured by geography, unsurprising given the clustered nature of individuals into subpopulations along the Wolastoq. The ex-situ conservation measures thus far also indicate that demographic rather than genetic collapse is the more pressing concern in the recovery of *P. furbishiae*. Taken together, this suggests that human intervention could reverse the demographic collapse observed in the lower sections of the Wolastoq through targeted population enrichment and could prevent the extirpation or extinction of this narrowly distributed relic species.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10592-025-01729-8>.

Acknowledgements We thank the New Brunswick Department of Energy and Resources Development for providing Species at Risk permits for all collection activities conducted in New Brunswick. Fieldwork assistance was provided by undergraduate students from the University of Maine at Fort Kent enrolled in PRN's Taxonomy of Vascular Plants course. We are especially grateful to Kathleen Forbes and Drs. Tannis Beardmore and Maureen Toner for their help with sampling *Pedicularis furbishiae* leaf tissue from extant sites in New Brunswick, and to Ian DeMerchant for generating the site map used in Fig. 1c. We also thank Haig Djambazian and Shu-Huang Chen of the McGill University Genome Centre for sequencing and bioinformatics support. Finally, we thank two anonymous reviewers for their insightful comments that have improved this paper.

Author contributions PRN, RHR, and MW designed the study. PRN, RHR, and MW contributed to material preparation and data collection. DMW, EN, and ML performed analyses. DMW prepared the manuscript and all authors commented on previous versions. All authors read and approved the final manuscript.

Funding Open access funding provided by Natural Resources Canada.

Data availability Raw genotyping-by-sequencing (GBS) data and whole genome shotgun sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) and are available via BioProject PRJNA1322473. The genome assembly has been submitted to GenBank under accession number [accession pending]. Analysis scripts are available in the github repositories at [https://github.com/enormandeau/stacks_workflow] (https://github.com/enormandeau/stacks_workflow), [<https://github.com/Erythroxyllum/furbishiae>] (<https://github.com/Erythroxyllum/furbishiae>), [<https://github.com/Erythroxyllum/furbishiae>] (<https://github.com/Erythroxyllum/furbishiae>), [<https://github.com/Erythroxyllum/psmc-pipeline>] (<https://github.com/Erythroxyllum/psmc-pipeline>), and [<https://github.com/Erythroxyllum/genome-wide-Ho>].

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19:1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Astashyn A, Tvedte ES, Sweeney D et al (2024) Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol* 25:60. <https://doi.org/10.1186/s13059-024-03198-7>
- Bryant D, Huson DH (2023) NeighborNet: improved algorithms and implementation. *Front Bioinform.* <https://doi.org/10.3389/fbinf.2023.1178600>
- Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21:255–265. <https://doi.org/10.1093/molbev/msh018>
- Bucharova A, Bossdorf O, Hölzel N et al (2019) Mix and match: regional admixture provenancing strikes a balance among different seed-sourcing strategies for ecological restoration. *Conserv Genet* 20:7–17. <https://doi.org/10.1007/s10592-018-1067-6>
- Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Dalton AS, Margold M, Stokes CR et al (2020) An updated radiocarbon-based ice margin chronology for the last deglaciation of the North American ice sheet complex. *Quat Sci Rev* 234:106223. <https://doi.org/10.1016/j.quascirev.2020.106223>
- Danecek P, Auton A, Abecasis G et al (2011) The variant call format and vcfTools. *Bioinformatics* 27:2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Danecek P, Bonfield JK, Liddle J et al (2021) Twelve years of samtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Eaton DAR (2014) PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* 30:1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Eaton DAR, Overcast I (2020) Ipyrad: interactive assembly and analysis of RADseq datasets. *Bioinformatics* 36:2592–2594. <https://doi.org/10.1093/bioinformatics/btz966>
- Elshire RJ, Glaubitz JC, Sun Q et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Ferrière R, Dieckmann U, Couvet D (2004) Evolutionary conservation biology. Cambridge University Press
- Fowler J (1885) A preliminary list of the plants of new Brunswick. Ellis, Robertson & Company
- Frankham R (2005) Genetics and extinction. *Biol Conserv* 126:131–140. <https://doi.org/10.1016/j.biocon.2005.05.002>
- Frankham R, Ballou JD, Eldridge MDB et al (2011) Predicting the probability of outbreeding depression. *Conserv Biol* 25:465–475. <https://doi.org/10.1111/j.1523-1739.2011.01662.x>
- Gawler SC, Waller DM, Menges ES (1987) Environmental factors affecting establishment and growth of *pedicularis furbishiae*, a rare endemic of the St. John river valley, Maine. *Bull Torrey Bot Club* 114:280–292. <https://doi.org/10.2307/2996466>

- Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* 5:184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Gruber B, Unmack PJ, Berry OF, Georges A (2018) Dartr: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* 18:691–699. <https://doi.org/10.1111/1755-0998.12745>
- Guan D, McCarthy SA, Wood J et al (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898. <https://doi.org/10.1093/bioinformatics/bta/a025>
- Gyllström M (2021) Ex Situ conservation of Furbish's Lousewort in New Brunswick –Potential for aligning future work with IUCN guidelines
- Halsey SJ, Bell TJ, McEachern K, Pavlovic NB (2015) Comparison of reintroduction and enhancement effects on metapopulation viability. *Restor Ecol* 23:375–384. <https://doi.org/10.1111/rec.12191>
- Hamming RW (1950) Error detecting and error correcting codes. *Bell Syst Tech J* 29:147–160. <https://doi.org/10.1002/j.1538-7305.1950.tb00463.x>
- Hijmans RJ (2024) Geosphere: spherical trigonometry. R package version 1.5–20. <https://github.com/rspatial/geosphere>
- Hilgers L, Liu S, Jensen A et al (2025) Avoidable false PSMC population size peaks occur across numerous studies. *Curr Biol* 35:927–930e3. <https://doi.org/10.1016/j.cub.2024.09.028>
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267. <https://doi.org/10.1093/molbev/msj030>
- Hutchinson GE (1951) Copepodology for the onthologist. *Ecology* 32:571–577. <https://doi.org/10.2307/1931745>
- Kennedy J (2025) Harvard University Herbaria: all records. <https://gbif.org/occurrence/1998990999>. Accessed 1 Apr 2025
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
- Li H (2025) SeqTK, Toolkit for processing sequences in FASTA/Q formats. <https://github.com/lh3/seqtk>. Accessed 21 Aug 2025
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496. <https://doi.org/10.1038/nature10231>
- Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Macior LW (1978) The pollination ecology and endemic adaptation of *Pedicularis furbishiae* S. Wats. *Bull Torrey Bot Club* 105:268–277. <https://doi.org/10.2307/2484919>
- Macior LW (1980) Population ecology of the furbish lousewort, *Pedicularis furbishiae* S. Wats. *Rhodora* 82:105–111
- Manni M, Berkeley MR, Seppy M, Zdobnov EM (2021) BUSCO: assessing genomic data quality and beyond. *Curr Protoc* 1(12):e323. <https://doi.org/10.1002/cpz1.323>
- Marsh DM, Trenham PC (2001) Metapopulation dynamics and amphibian conservation. *Conserv Biol* 15:40–49. <https://doi.org/10.1111/j.1523-1739.2001.00129.x>
- McKinney GJ, Waples RK, Seeb LW, Seeb JE (2017) Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Mol Ecol Resour* 17:656–669. <https://doi.org/10.1111/1755-0998.12613>
- McVean GAT, Cardin NJ (2005) Approximating the coalescent with recombination. *Philos Trans R Soc Lond B Biol Sci* 360:1387–1393. <https://doi.org/10.1098/rstb.2005.1673>
- Menges ES (1990) Population viability analysis for an endangered plant. *Conserv Biol* 4:52–62. <https://doi.org/10.1111/j.1523-1739.1990.tb00267.x>
- Nei M (1987) Molecular evolutionary genetics. Columbia University, New York
- New Brunswick Department of Natural Resources and Energy Development (2023) Furbish's lousewort *Pedicularis furbishiae*, in new brunswick: status Report, NBDNRD, Fredericton, NB. pp. + apps, p 25
- Pedersen, T. (2025). Patchwork: the composer of plots (Version 1.3.2) [R package]. <https://patchwork.data-imaginist.com>
- Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575. <https://doi.org/10.1086/519795>
- Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Richards CD (1976) Report on rare and unusual plant species within the Dickey-Lincoln School lakes project area. United States Army Corps of Engineers
- Robert BW, Gladys C, Frank T, Kilpatrick S (2015) Phylogeny and biogeography of North American and Asian *Pedicularis* (Orobanchaceae). *Syst Bot* 40:229–258. <https://doi.org/10.1600/036364415X686549>
- Robinson JA, Kyriazis CC, Nigenda-Morales SF et al (2022) The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science* 376:635–639. <https://doi.org/10.1126/science.abm1742>
- Rochette NC, Rivera-Colón AG, Catchen JM (2019) Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* 28:4737–4754. <https://doi.org/10.1111/mec.15253>
- Secretary, Smithsonian Institution (1975) Report on endangered and threatened plant species of the united States. U.S. Government Printing Office, Washington, D.C
- Soulé ME (1985) What is conservation biology? A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems. *Bioscience* 35:727–734. <https://doi.org/10.2307/1310054>
- Stirret GM (1980) COSEWIC status report on the furbish's lousewort *Pedicularis furbishiae* in Canada. Committee on the Status of Endangered Wildlife in Canada
- Tange O (2018) GNU Parallel 2018
- Thrall PH, Richards CM, McCauley DE, Antonovics J (1998) Metapopulation collapse: the consequences of limited gene-flow in spatially structured populations'. In: Bascompte J, Solé RV (eds) Modeling Spatiotemporal dynamics in ecology. Springer, pp 83–104
- U.S. Fish & Wildlife Service (1978) Final determination that eleven plant taxa are endangered and two plant taxa are threatened species
- USFWS 1991. Furbish lousewort (*Pedicularis furbishiae*) revised recovery plan. Prepared by Susanna L. von Oettingen, New England Field Office, U. S. Fish and Wildlife Service, Concord, New Hampshire.
- U.S. Fish & Wildlife Service (2018) Species status assessment for the Furbish's lousewort (*Pedicularis furbishiae*). Version 1.1
- U.S. Fish & Wildlife Service (2023) Endangered and threatened wildlife and plants; Reclassifying Furbish's Lousewort (*Pedicularis furbishiae*) from endangered to threatened status with a section 4(d) Rule
- Vasimuddin Md, Misra S, Li H, Aluru S (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems, pp 314–324

- Walker BJ, Abeel T, Shea T et al (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Waller DM, O'Malley DM, Gawler SC (1987) Genetic variation in the extreme endemic *Pedicularis furbishiae* (Scrophulariaceae). Conserv Biol 1:335–340
- Weng M-L, Becker C, Hildebrandt J et al (2019) Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. Genetics 211:703–714. <https://doi.org/10.1534/genetics.118.301721>
- Wickham H (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. Trends Ecol Evol 30:42–49. <https://doi.org/10.1016/j.tree.2014.10.009>
- Winton S, Chabot A, Zajac K et al (2024) Furbish's lousewort (*Pedicularis furbishiae*) ex situ conservation planning workshop: final report. IUCN SSC Conservation Planning Specialist Group - Regional Resource Center Canada
- Wood DE, Lu J, Langmead B (2019) Improved metagenomic analysis with kraken 2. Genome Biol 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.