Genetic Diversity Evaluation of Rare Endemic Plant Species *Braya longii* and *Braya fernaldii*

by

© **Nathan MacNeil**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science.

Department of Biology

Memorial University of Newfoundland

September 2024

St. John's, Newfoundland and Labrador, Canada

Abstract

Braya longii and *B. fernaldii* are two endangered vascular plants endemic to Newfoundland's (Canada) Northern Limestone Barrens. Currently, there is no firm understanding of the genetics of either species. To date, these two species are managed together under the same *Braya* Action and Recovery Plans due to their shared habitat and close taxonomic relationship. However, an improved understanding of the genetics of both *B. longii* and *B. fernaldii* would improve the future management and survival of these two species, as the use of genetically appropriate source populations in restoration, is a high-priority recovery measure. Genomic analysis is a valuable tool for delineating conservation units and can be performed by next-generation sequencing methods such as genotype-by-sequencing. This research aims first to identify the genetic diversity and structure of the *B. longii* and *B. fernaldii* populations, and second, to provide insight into the relationship between the two species' population genetics and the implications for management/conservation units. I leverage phylogenetic results using SplitsTree4, population structure analysis using STRUCTURE and discriminant analysis of principal components, and SNP analysis using the R package Hierfstat to gain insights into the relationships between each species and within populations. The results support the separation of each species. I propose six divergent lineages as evolutionarily significant units, three for *Braya longii* and three for *B. fernaldii*. I also propose eight management units, with each population managed separately, while Sandy Cove 1 and Savage Cove share a single management unit. These management units can be referenced for making restoration decisions, such as when sourcing individuals. While this research provides new insight into the population genetics of *B. longii* and *B. fernaldii* for the conservation and management of these species, future research could use this dataset to estimate past and current gene flow, effective population sizes and test population expansion or contraction.

Lay summary

Long's Braya and Fernald's Braya are two rare endemic plants found in the Limestone Barrens of the Great Northern Peninsula of Newfoundland (Canada). There is currently no strong understanding of the genetic relationship between these plants. These two plants look very similar, have a similar life history, live in the same ecosystem, and are managed together using the same *Braya* Recovery Plan. Despite having a very similar appearance, a genetic understanding is needed to know how different Long's Braya is from Fernald's Braya, and how different their populations are from each other to properly manage these species to support their long-term conservation. In this study, I processed and analyzed plant samples to gain knowledge about their genetics to suggest groupings for managing and protecting them. This will help inform conservation professionals and researchers of the next steps to protect these species. I propose eight management units representing each population, with the exception of Sandy Cove 1 and Savage Cove, which could be managed together.

Acknowledgements

I acknowledge Tyra Custance for performing the DNA extractions and Patrick Lauriault for helping with fieldwork. I also want to acknowledge the Canadian NSERC-Discovery grant (RGPIN-2014-03976) awarded to Dr. Julissa Roncal. The fieldwork and DNA sequencing were funded by an ECCC-Community Nominated Priority Places on Western Newfoundland Biodiversity grant to Dr. Luise Hermanutz, and the Wildlife Division within the Department of Fisheries, Forestry and Agriculture of Newfoundland and Labrador facilitated collecting permits. Data analysis was run using the Compute Canada Database (CCDB) as a sponsoree of Dr. Lourdes Peña-Castillo's group, enabled in part by support provided by ACENET (ace-net.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

I would like to thank Dr. Julissa Roncal, Dr. Lourdes Peña-Castillo, and Dr. Luise Hermanutz for the amazing supervision, advice, and support provided throughout this process.

Table of contents

List of tables

List of figures

List of abbreviations

AP	Anchor Point
BB	Bellburns
CN	Cape Norman
COSEWIC	Committee on the Status of Endangered Wildlife in Canada
DU	Designatable Units
ESU	Evolutionary Significant Unit
GBS	Genotyping-by-Sequencing
GBS-SNP-CROP	GBS-SNP-Calling Reference Optional Pipeline
GIB	Green Island Brook
ISSR	Inter Simple Sequence Repeats
K	Number of Genetic Clusters
LB	Limestone Barrens
LBSARRT	Limestone Barrens Species at Risk Recovery Team
MU	Management Unit
SARA	Species at Risk Act
SC	Sandy Cove (Inclusive of SC1 and SC2)
SC1	Sandy Cove 1
SC2	Sandy Cove 2
SNP	Single Nucleotide Polymorphism
Selfing	Self-pollination or Inbreeding
VC	Savage Cove
WB	Wild Bight
YP	Yankee Point
bp	Base Pair

Chapter 1

Introduction

Biodiversity is valuable for ecosystems, local communities, eco-tourism industries, medical practice, and agriculture [1, 2]. Over the last 500 years, anthropogenic-driven extinction has accelerated to rates similar to the worldwide mass extinction events from the previous 500 million years, with the current rate being 1000 times higher than natural background extinction rates [3, 4]. Given the high extinction rates, conservation biology and climate change have become major topics of research. Conservation actions that slow the rate of biodiversity loss can also slow anthropogenic-driven climate change [5, 6].

Understanding genetic diversity and structure is essential for preserving biodiversity, as genetic information allows for identifying extinction risk accurately and implementing appropriate recovery strategies [7]. For example, in the case of more diverse populations, by implementing re-introductions using local population diversity levels as a benchmark, populations will have the capacity for long-term adaptation with a lower risk of inbreeding depression [8]. For rare species with low genetic diversity, re-introductions using different sources may result in more diverse populations with a greater chance of long-term adaptation [9]. Genetic variation and geographic population distribution are positively correlated, with reduced population sizes and reduced population ranges having less genetic diversity [10]. Genetic diversity allows for generational adaptation, and genetically diverse populations are more resilient to environmental and climate change [11].

Wildlife management is defined by [12] as "the management of wildlife populations in the context of the ecosystem." For this text, this concept will be referred to as

population management. Conservation professionals may act to conserve populations with low fitness or dwindling population numbers by employing reintroductions [13–17]. If a population decreases year over year, reintroducing individuals from an outside population may be a suitable conservation action. For example, Alpine ibex (*Capra ibex*) goat populations have grown from reintroductions from <100 up to around 50,000. However, the reintroduction efforts and strategies resulted in reduced diversity than the original source population with high levels of inbreeding. Despite this being a successful reintroduction from a population numbers perspective, the high levels of inbreeding may have negative long-term effects on the fitness of these goats [18].

Braya longii and *B. fernaldii* in the family Brassicaceae are two herbaceous, vascular, octoploid plant species with allopolyploid origins [19]. Both species are very morphologically similar (Figure 1.1). Both species primarily self-fertilize and only disperse seeds a short distance (within 50 cm of the adult plant) [20, 21]. They are listed as endangered under the Canadian Species at Risk Act (SARA) and the Newfoundland and Labrador Provincial Endangered Species Act. They are endemic to Newfoundland's northern Limestone Barrens (Figure 2.1). Despite being separate species, these plants are managed under the same *Braya* Recovery and Action Plans [22]. They were designated as "endangered" under COSEWIC due to inhabiting a limited range on the Limestone Barrens of the Great Northern Peninsula of Newfoundland, as well as experiencing habitat loss and degradation, invasive pests, and pathogens [23, 24].

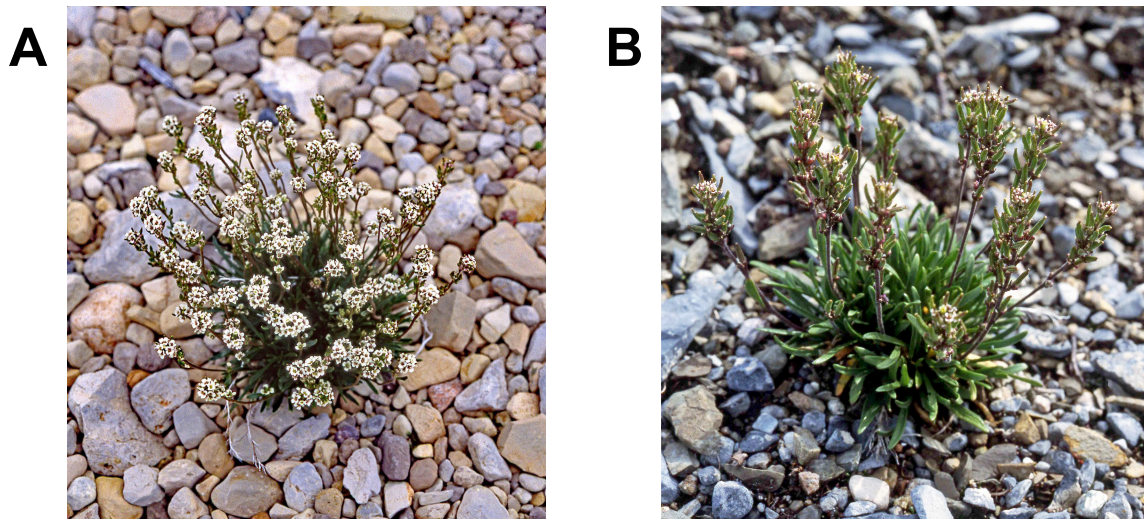

Figure 1.1: A photograph of *B. longii* [25] (A) and a photograph of *B. fernaldii* [26] (B)

Outcrossing is the transfer of gametes for reproduction to other unrelated individuals, while selfing is the fertilization from pollen originating from the same plant. Outcrossing species usually have more genetic diversity and less genetic differentiation between populations than selfing species [27]. Rare, endemic species frequently have low levels of genetic diversity and may experience more differentiation among populations from inbreeding [28]. Island plants have been assumed to have lower levels of genetic diversity than mainland plant species due to several factors, such as genetic bottlenecks from colonization and smaller overall populations [28]. Even though these factors allow us to hypothesize the expected population genetics of the endemic Newfoundland plant species Long's Braya (*Braya longii*) and Fernald's Braya (*Braya fernaldii*), not all studies exhibit island plant populations having lower levels of genetic diversity than mainland populations [29]. Based on the literature review I conducted, there seem to be few studies on the population genetics of rare polyploids like *B.longii* and *B.fernaldii* [19]. Polyploid data can be more challenging to analyze than diploid data due to allele dosage considerations [30]. Allopolyploids are polyploids originating from a hybridization event, and this chromosomal complexity can lead to bias in the estimation of genetic diversity [31]. While there has been no published study on the genetics of Long's and Fernald's Braya, there has been preliminary research in an Honours thesis [32] which used inter-simple sequence repeats (ISSRs) from 21 *B. longii* samples and 28 *B. fernaldii* samples to suggest management units

for creating genetically appropriate seed banks. This ISSR work only analyzed 19 loci, recommended that both species should be considered a single management unit, and suggested that future genetic work should be done on these species.

Conservation professionals should use genomic data for species and population management. Professionals use various genetic groupings, such as COSEWIC's Designatable Units (DU) [33] to manage endangered species by classifying them for population management. Units that can be suggested using genomic data include management units (MUs) and evolutionary significant units (ESUs) [34–36]. MUs are separately managed population groups or subgroups because they are genetically distinct, while ESUs represent evolutionary lineages, and ESUs may contain one or many MUs [34]. ESUs are valuable as they can capture different evolutionary trajectories, which are worth conserving separately as they are highly genetically distant [37].

Single nucleotide polymorphisms (SNPs) are a single nucleotide variation at a specific location in a DNA sequence (Figure 1.2) [38]. One can identify a specific locus with a polymorphism by comparing two DNA sequences and aligning them by loci. SNPs are a measure of genetic diversity, and genotype-by-sequencing (GBS) [39] is a method used to identify SNPs across populations to characterize both gene flow and adaptive variation [40]. Gene flow is the movement of genetic material between populations over time, and adaptive variation is the genetic variation as a result of adaptation to a local environment [41, 42]. Understanding these concepts helps with the designation of MUs to inform re-introductions. Conservation professionals can use MUs and ESUs to avoid genetic inbreeding/outbreeding, which is associated with the risk of losing a parental species from offspring dominating a shared niche [43]. As the environment changes, a historical understanding of population genetics will be a snapshot to inform future long-term management [44].

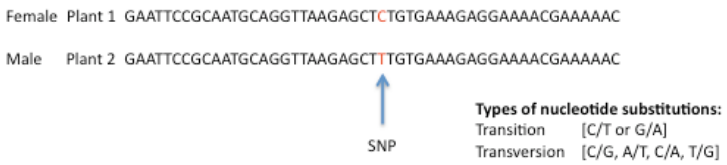

Figure 1.2: An example of a SNP [45]

A high-priority recovery measure in the 2020 amended Action Plan for *B. longii* and *B. fernaldii* [46] is "Using genetically appropriate source populations, reintroduce

Long's Braya and Fernald's Braya by seed, where required, to restore species distribution within its historic range." This research aims to elucidate the population genetics of *B. longii* and *B. fernaldii* to close the knowledge gap for this high-priority recovery measure. To do this, extensive DNA sequencing (GBS) was conducted of both species from across their distribution ranges, and bioinformatics tools and pipelines were used to process the data, resulting in the identification of SNPs.

For my thesis, I had the following specific objectives:

1. Estimate levels of population genetic diversity and structure for both *Braya* species.
2. Using the genetic results from objective 1, recommend management units (MUs) and evolutionary significant units (ESUs) for *Braya longii* and *Braya fernaldii*.
3. Using the MUs and ESUs from objective 2, recommend strategies for long-term *in-situ* and *ex-situ* conservation that will help with future management decisions.

Due to *Braya*'s life history traits of selfing, low-range seed dispersal [20], and small population sizes, I expect that *B. longii* and *B. fernaldii* should exhibit low intra-population genetic diversity and high inter-population genetic differentiation [47, 48]. Another expectation is that every population will be genetically distinct, and therefore each population will constitute a separate MU. The following chapters cover a description of the methods ranging from sample collection in the field to DNA data analysis, the results containing population diversity metrics and genetic structure, a phylogenetic analysis, and interpretations of the results.

Chapter 2

Methods

2.1 Sample Collection and DNA Extraction

In 2017, Patrick Lauriault and Luise Hermanutz collected leaf samples under a provincial permit from 39 *B. longii* and 46 *B. fernaldii* individuals across eight populations (Table 2.1) which span across both species' distributions in the Limestone Barrens of the Great Northern Peninsula in Newfoundland, Canada, for a total of 85 *Braya* samples (Tables 3.1 and 2.1, Figure 2.1). Leaves were selected from healthy-looking (i.e., no obvious signs of pest or pathogen damage) adult individuals from which no more than 20% of the plant's biomass was extracted. Leaf samples were dried and stored in silica gel.

The low number of sampled *Braya* populations and individuals reflects the low total number of populations left in the wild. Both *Braya* species are small and affected by pests and pathogens, limiting the overall potential sampling pool. The *B. fernaldii* population of Port au Choix was the only main population not sampled. It is located along the coast between Anchor Point and Bellburns. This population was not sampled since the population is very small, and the plants are smaller than other *B. fernaldii* plants; hence to secure sufficient leaf tissue for DNA extraction would severely compromise the health of the plant.

Table 2.1: Sampling locality sites for each *Braya* population in the Limestone Barrens of Newfoundland (Figure 2.1).

Sampling localities per species	Population Acronym	Latitude and Longitude	Number of samples sequenced
Braya longii			
Sandy Cove	SC1/SC2	51.36378765, -56.64007902	10/5
Savage Cove	VC	51.35332494, -56.66219518	10
Yankee Point	YP	51.32488489, -56.71294928	14
Braya fernaldii			
Cape Norman	CN	51.61458492, -55.91508865	11
Wild Bight	WB	51.60587311, -55.8820653	6
Green Island Brook	GIB	51.39515877, -56.5315032	15
Anchor Point	AP	51.22969866, -56.76906109	8
Bellburns	BB	50.40533781, -57.50853539	6

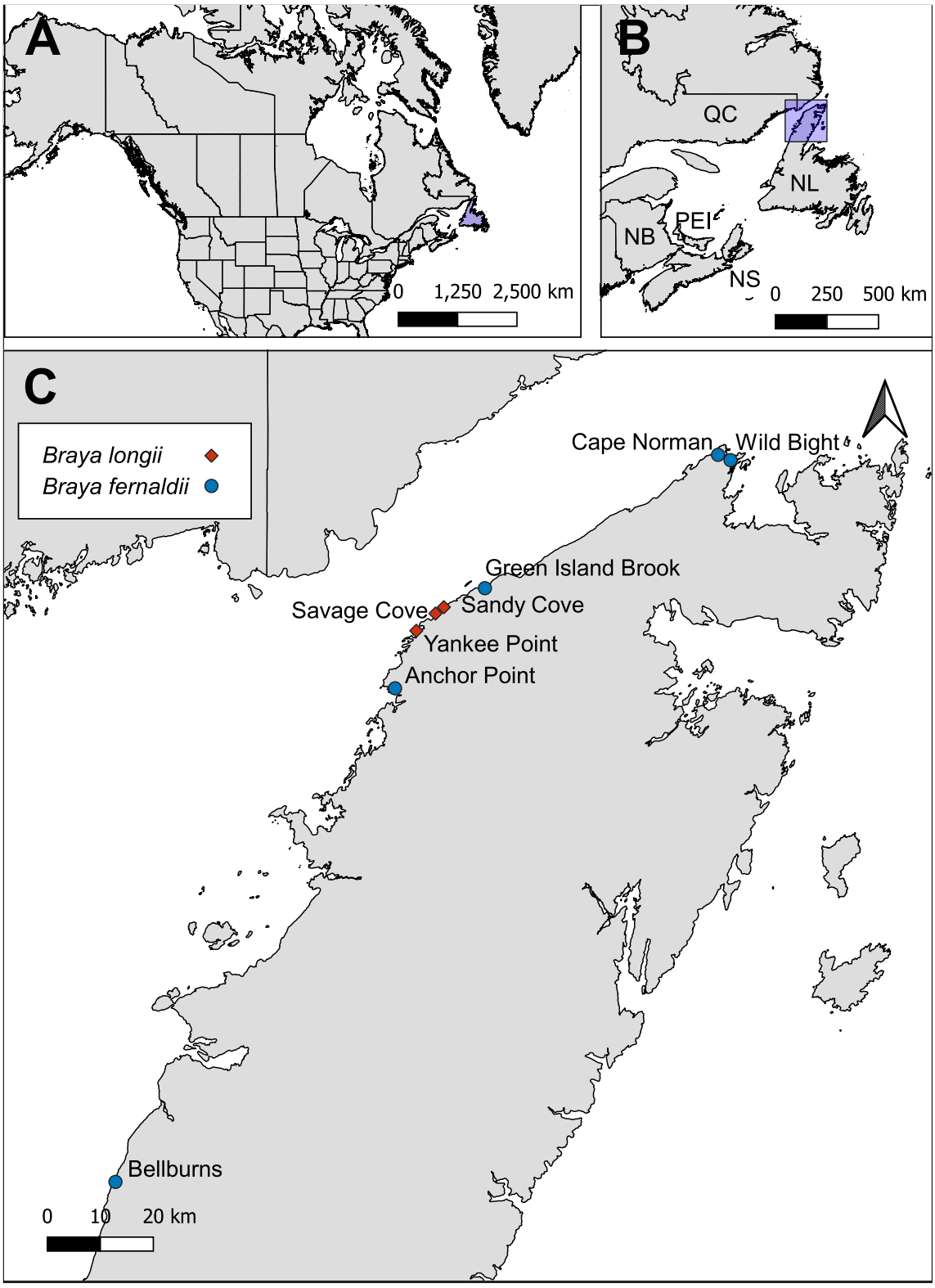

Figure 2.1: A) Map of North America highlighting the position of the island of Newfoundland. B) Map of Canadian Atlantic provinces (NL, QC, PEI, NB, NS) highlighting the study site location. C) Map of *Braya fernaldii* and *B. longii* sampled populations in the Limestone Barrens ecosystem at risk.

I obtained DNA extractions from a previous study conducted by Tyra Custance [32]. Here, I present a brief summary of the DNA extraction protocol. At least 20mg of finely cut leaf tissue was placed inside a 2 mL microcentrifuge tube with two sterilized 3 mm steel balls to help the plant tissue disruption. The microcentrifuge tube was then placed in a QIAGEN TissueLyser which was run for at least 3 min at 50 Hz. In many cases, samples were run for an additional 3 min cycles to ensure complete tissue disruption [32]. The DNeasy Plant Mini Kit from QIAGEN was used to extract the DNA from the disrupted leaf samples following the manufacturer's protocol with a few modifications to improve DNA recovery. Next, 600 μl of AP1 buffer and 4 μl of RNase were used for each sample. The incubation time for cell lysis was 15 mins at 65 °C. After incubation, 195μl of P3 buffer was added [32]. A Qubit fluorometer was used to quantify the total genomic DNA and diluted DNA extractions in EB buffer (Qiagen) to 20 ng/μl.

2.2 Genotyping-by-sequencing and SNP discovery

The Institut de Biologie Intégrative et de Systèmes (IBIS) at Université Laval performed the genotyping-by-sequencing (GBS). GBS is a cost-effective sequencing method that constructs reduced representations of genotypes. Digestion enzymes are used to reduce the complexity of the DNA. IBIS prepared genomic libraries for the 85 DNA samples using the Illumina TruSeq HT adaptor and unique barcodes 10-12 bp long to facilitate demultiplexing. To reduce genome complexity, a rare cutting enzyme, PstI and a common cutting enzyme, MspI, were used, as proposed in Poland et al. [49]. Library preparation and sequencing followed the protocol of Abed et al. [50]. Genome Quebec performed the DNA sequencing on one lane of the Illumina HiSeq 4000 PE100. I inspected data quality with FastQC 0.11.9 [51]. After quality control, I removed five individuals with less than 1 million reads from further analyses, leaving 80 *Braya* individuals analyzed. Low amounts of reads in these samples could arise from low DNA concentration, or issues related to library preparation and sequencing.

Data were processed using the GBS-SNP-Calling Reference Optional Pipeline (GBS-SNP-CROP) v4.1 [52]. I chose this data pipeline because it can handle polyploid individuals by providing ploidy-dependant filtering parameters and a Z-score metric [53]. The Z-score is a metric used to filter homologous variants. It allows

filtering variants from various non-diploid scenarios by identifying the proportion of observed allelic counts for each variant and comparing them to the expected allelic counts. The Z-score is the deviation from this expected value. GBS-SNP-CROP uses Trimmomatic v0.39 [54] to remove low-quality reads and adapter sequences and PEAR v0.9.11 [55] to merge the paired-end reads into single reads. I ran GBS-SNP-CROP using the parameters found in Table 2.2.

In these and subsequent analyses I systematically tried different parameter values and selected the parameter values that provided the "best" results. "Best" results, in this case, were the parameter values that yielded the most interpretable results that made ecological and biological sense. For example, results that aligned with the geographical location or were possible based on the biology of *Braya*. Another important consideration was to be able to retain enough data for the downstream analyses. For instance, when filtering SNPs, I must ensure enough SNPs are left for the population analysis.

Table 2.2: GBS-SNP-CROP steps and corresponding commands used for this analysis

Step	***Command***
Step 1: Parse the raw reads	perl GBS-SNP-CROP-1.pl -d PE -b BarcodeID.txt -fq Braya -s 1 -e 1 -enz1 TGCA -enz2 CGG -t 16
Step 2: Trim based on quality and adaptors	perl GBS-SNP-CROP-2.pl -tm trimmomatic.jar -d PE -fq Braya_001 -t 16 -ph 33 -ad 0
Step 3: Demultiplex	perl GBS-SNP-CROP-3.pl -d PE -b BarcodeID.txt -fq Braya_001
Step 4: Cluster reads and assemble the Mock Reference	perl GBS-SNP-CROP-4.pl -d PE -b BarcodeID.txt -rl 100
Step 5: Align with BWA-mem and process with SAMtools	perl GBS-SNP-CROP-5.pl -bw bwa -d PE -b BarcodeID.txt -ref GSC.MR.Genome.fa -t 16
Step 6: Parse mpileup outputs and produce the variant discovery matrix	perl GBS-SNP-CROP-6.pl -b BarcodeID.txt -out GSC.MasterMatrix.txt -t 24
Step 7: Filter variants and call genotypes	perl GBS-SNP-CROP-7.pl -in GSC.Summary.txt -out GSC.GenoMatrix.txt -mnHoDepth0 5 -mnHoDepth1 20 -mnHetDepth 3 -altStrength 0.962 -mnAlleleRatio 0.25 -mnCall 0.75 -mnAvgDepth 3 -mxAvgDepth 200
Filter based on z-score	awk '$12<=7&&$12>=-7 {print}' GSC.GenoMatrix.txt > GSC.FilteredMatrix.txt
Script 8: Create input files for downstream analyses	perl GBS-SNP-CROP-8.pl -b BarcodeID.txt -in GSC.FilteredMatrix.txt

The sample with the highest number of reads (*B.fernaldii*-BB-617) was selected to create a reference genome using VSEARCH v2.15.1 [56]. This strategy was used because it had been shown to produce the highest number of SNPs when compared to other strategies [57]. Reads were aligned using BWA aligner v0.7.12 [58], and sorted and indexed using SAMTools v1.7 [59]. Unless otherwise stated, parameters for all tools were left at their default values. The parameters used for the initial SNP filtering (Figure 2.2) using the provided GBS-SNP-CROP Step 7 Script can be found in Table 2.3 [60].

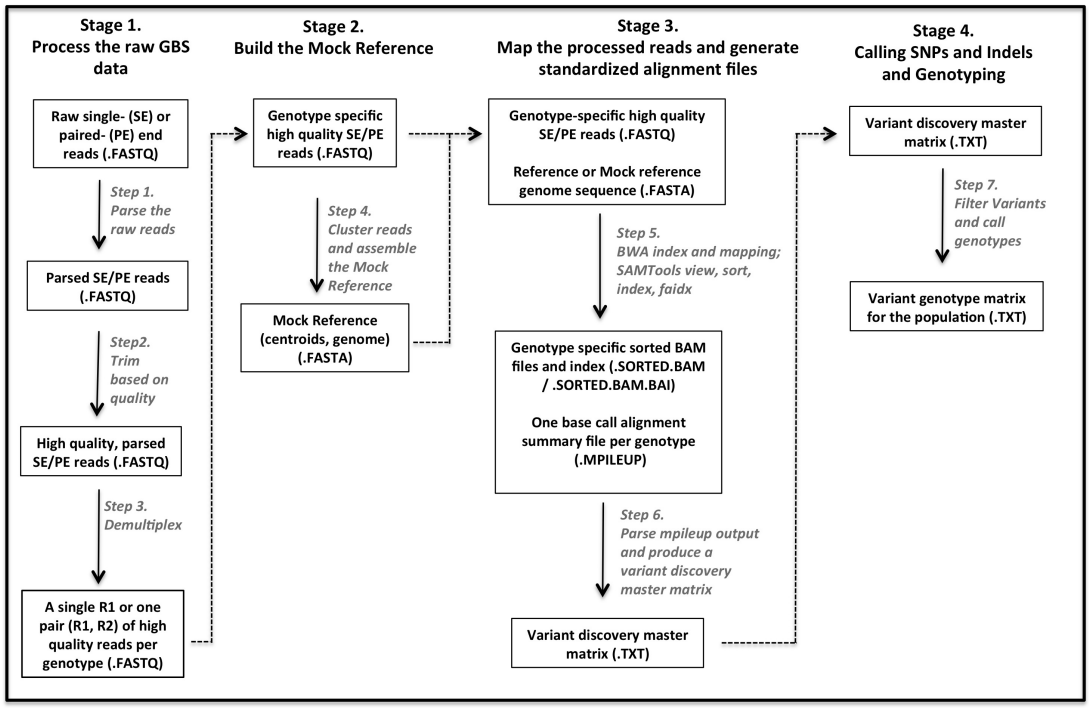

Figure 2.2: GBS-SNP-CROP workflow diagram [61]

Table 2.3: GBS-SNP-CROP Step 7 parameters

Parameter Definition	***Parameter***	***Value***
Minimum depth when secondary allele count is zero	mnHoDepth0	5
Minimum depth when secondary allele count is one	mnHoDepth1	20
Minimum allele depth for calling heterozygotes	mnHetDepth	3
Alternative allele strength	altStrength	0.962
Minimum ratio of more frequent allele depth to less frequent allele depth	mnAlleleRatio	0.25
Minimum percentage of individuals genotyped to keep a variant	mnCall	0.75
Minimum average depth	mnAvgDepth	3
Maximum average depth	mxAvgDepth	200

Version 4.0 of the step 8 script was used instead of 4.1, as the one for 4.1 caused downstream errors in subsequent scripts due to version incompatibility. The dataset was filtered as though it was diploid according to the suggestions in the GBS-SNP-CROP wiki [57] and then filtered for paralogs using the Z-score as also outlined on the wiki. We used the diploid settings to maintain enough data for downstream analysis. The Z-score filtering was done to account for the polyploid nature of *B. longii* and *B. fernaldii*. These parameters were established after several runs of the pipeline, which yielded different SNP sets. Separating sequences deriving from polyploidization into their respective loci when they are non-allelic and highly related is a challenge of processing polyploid data and is mitigated by filtering in GBS-SNP-CROP [52]. Without accounting for polyploidization, the results may be inaccurate by over-counting SNPs resulting from duplication. After the initial filtering, GBS-SNP-CROP provided a Z-score parameter, which I used to address the polyploidization in *Braya* [52]. Further filtering used the Z-score column of the genotype matrix produced by GBS-SNP-CROP to filter out homologous variants using the threshold suggested by [53] of $|Zi| < 7$. I used a variant calling file (VCF) containing the remaining SNPs after initial and Z-score filtering for all downstream analyses.

2.3 Genetic diversity and structure of *Braya* populations in the Limestone Barrens

I obtained genetic diversity statistics using the R package Hierfstat v0.5-11 [62]. I used vcfR v1.12.0 to read the VCF file into R [63]. The basic.stats function in Hierfstat provided general statistics at the population level, such as the observed heterozygosity (HO), observed gene diversity (HS), gene diversity among populations (DST), fixation index (FST), and inbreeding coefficient (FIS). The pairwise.neifst function provided pairwise population FST statistics using Nei's minimum genetic distance [64].

To investigate the genetic structure among *Braya* individuals within and between species, I used a model-based clustering method within a Bayesian framework as implemented in STRUCTURE v.2.3.4 [65]. I used Plink v1.90b6.21 [66] to convert the VCF file to a file that could be used in STRUCTURE. I determined the number of genetic clusters (K) and used default parameter settings unless otherwise stated. I ran analyses with the admixture model, without the linkage model, and without a priori population information. As described in the STRUCTURE documentation [67], Admixture modelling assumes that each individual draws some genetic information from each of the K populations, and linkage modelling is similar to admixture modelling, but linked loci are more likely to originate from the same population. The admixture model was chosen due to the flexibility of the method and the complexity of real-world populations [67]. I ran STRUCTURE using 400,000 Markov Chain Monte Carlo (MCMC) generations after a burn-in of 400,000 generations. Each of the 20 iterations ran for both species together and both species individually. I allowed K to vary from one (no population structure) to 10 genetic clusters. After each of the 20 runs, the STRUCTURE file was uploaded to STRUCTURE HARVESTER v0.6.94 [68], which returned a suggested number of genetic clusters for each run using the ΔK test [69]. The appropriate individual and population files for the suggested K values were used in CLUMPP v1.1.2 [70] to align the different genetic clusters across STRUCTURE runs. The results from CLUMPP on the population and individual files were then processed using DISTRUCT v1.1 [71] to produce bar plots representing the membership coefficient of each individual to genetic clusters.

I conducted a Discriminant Analysis of Principal Components (DAPC) (Figure

3.4) using the R package adegenet v2.1.5 [72]. Two axes were retained in the Discriminant Analysis (DA) step, and four were retained in the Principal Component Analysis (PCA) step. These axes were retained due to the agreement with both the STRUCTURE bar plots and the phylogenetic analysis generated in this analysis. The figure was not as interpretable when following techniques for the suggested number of PCs to retain, and the clusters seemed to be more randomly distributed, so the exact parameters used for the DAPC were decided instead based on agreement with other results. However, a secondary DAPC (Figure 3.5) is provided, which uses the cross-validation (xvalDapc()) function from adegenet [72] to infer the number of PCs to retain. Cross-validation is an optimization procedure used to provide the correct number of PCs to retain [73]. During the cross-validation step, 1000 replicates were carried out at each level of PC retention; the suggested number of PCs was 13, and the suggested number of DAs retained was 8. Lastly, I conducted an analysis of molecular variance (AMOVA) for all sampling locations with SNP data using Arlequin v3.5.2.2 with a missing data threshold set to 0.05, which left 1046 loci available for distance computation [74]. Three stratifications were used: among species, among populations, and within populations.

2.4 Phylogenetic analysis

A python script called vcf2phylip v2.0 [75] was used to convert a VCF file to a nexus file so that SplitsTree4 v4.17.1 [76] could open the diversity matrix. A neighbor-joining unrooted dendrogram was reconstructed using the JukesCantor distance function to create an equal-angle tree. Bootstrap values were generated using 1000 replicates.

Chapter 3

Results

3.1 Sequencing, filtering, and SNP discovery

I obtained two FASTQ files from sequencing, one containing forward reads (19.9 GB compressed) and the other reverse reads (20.4 GB compressed). A Phred quality score is a common metric representing the accuracy of collecting bases during sequencing. In this study, the Phred [77] sequence accuracy is very high (99%), reflecting a Phred quality score of 39 [Fig 3.1].

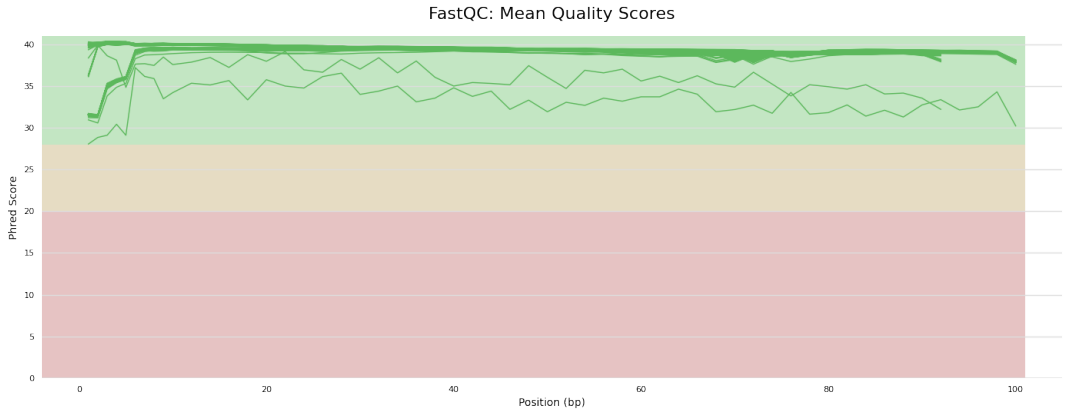

Figure 3.1: Line plots from MultiQC [78] report showing the mean Phred Score per position (bp).

The data consisted of 366,762,326 reads between 85-97 bp long. After sequencing, the data was sorted into the original samples by using barcode identifiers in a process known as demultiplexing. Demultiplexing gave us an average of 2.4 million reads per

sample, ranging from 0 (no reads) to 15 million reads per sample [Fig. 3.2]. Due to a lack of data, I discarded five *B. fernaldii* samples from Cape Norman with fewer than 1 million sequences, leaving 39 *B. longii* and 41 *B. fernaldii* individuals to be analyzed.

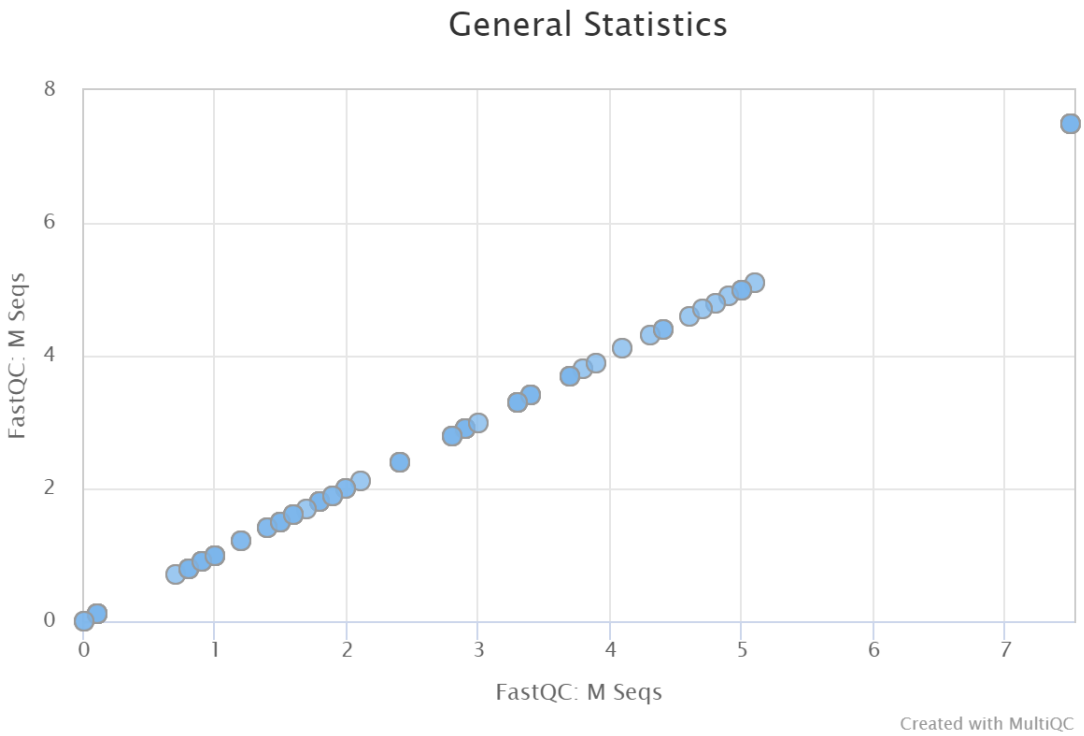

Figure 3.2: Simple figure with Total Sequences (Millions) on both the X and Y axis. Five samples with fewer than one million reads were removed. Figure created using MultiQC [78].

Without filtering, GBS-SNP-CROP yielded 92,671 SNPs. After filtering for high-confidence variants, I retrieved 2,387 SNPs when homologous variants were not filtered using the Z-score method (Figure 2.2, Table 2.2) and 1,449 SNPs when removing homologous variants. All downstream analyses used these 1,449 SNPs. Pipeline parameters were chosen by using suggestions from GBS-SNP-CROP [57] and running the pipeline several times until enough high-quality SNPs were retained from filtering.

3.2 Genetic diversity

The *B. fernaldii* population of Cape Norman displayed the highest observed heterozygosity (0.5821) and observed gene diversity (0.3466), followed by the most southerly population collected from Belburns (Table 2.1). One of the most northern *B. fernaldii* populations, Wild Bight, had the lowest observed heterozygosity (0.5399) and observed gene diversity (0.2724). Cape Norman and Wild Bight are geographically close, being the most northern populations, despite having both the highest (CN) and lowest (WB) observed heterozygosity. For *B. longii*, Yankee Point had the highest observed heterozygosity (0.5424), and Sandy Cove 1 had the highest observed gene diversity (0.2912).

All inbreeding coefficients (FIS) were negative (Table 3.1), suggesting an excess in heterozygotes in all *Braya* populations in relation to what was expected under random mating. However, this result will be discussed in the next chapter in the context of *B. longii* and *B. fernaldii* being allopolyploids [19] and their likely inheritance mode. Allopolyploids are polyploids originating from a hybridization event. Outbreeding is maintained in these populations. Therefore, loss of genetic diversity due to inbreeding is not a threat for *Braya*. Genetic differentiation (FST) between *B. longii* populations was lower than the differentiation between *B. fernaldii* populations. The lower the FST value, the lower the genetic differentiation between individuals/populations. Within *B. longii*, the highest pairwise FST values were for Yankee Point and Savage Cove and the lowest between Sandy Cove and Savage Cove (Table 3.2). Within *B. fernaldii*, the highest pairwise FST values were for Anchor Point and Bellburns and the lowest between Anchor Point and Cape Norman (Table 3.2). Despite the high pairwise FST values, Anchor Point is the geographically closest population to Bellburns, while Bellburns is the most geographically distant from all other populations. Within *B. longii*, the FST values suggest that Yankee Point and Savage Cove are the least differentiated, while Sandy Cove and Savage Cove are the most differentiated. For *B. fernaldii*, Anchor Point and Belburns are the most differentiated, and Anchor Point and Cape Norman are the least differentiated.

The Mantel test did not show a significant correlation between population genetic differentiation and geographic distances (R = -0.015, p-value = 0.533), suggesting a lack of isolation by geographic distance. Unexpectedly, populations closer geographically are not more closely related genetically (Table 3.2).

Table 3.1: Genetic diversity statistics for each *Braya* population in the Limestone Barrens of Newfoundland generated from Hierfstat's basic.stats function. HO = observed heterozygosity, HS = observed gene diversity, DST = gene diversity among populations, and FIS = inbreeding coefficient averaged over loci [79]

Population per species (acronym)	#samples	HO	HS	DST	FIS
Braya longii					
Sandy Cove 1 (SC1)	10	0.5234	0.2912	0.0397	-0.7974
Sandy Cove 2 (SC2)	5	0.5368	0.2714	0.0595	-0.9778
Yankee Point (YP)	14	0.5424	0.2729	0.0580	-0.9872
Savage Cove (VC)	10	0.5415	0.2731	0.0578	-0.9828
Braya fernaldii					
Green Island Brook (GIB)	15	0.5649	0.3031	0.0278	-0.8639
Anchor Point (AP)	8	0.5554	0.2969	0.0340	-0.8707
Wild Bight (WB)	6	0.5399	0.2724	0.0585	-0.9818
Cape Norman (CN)	6	0.5821	0.3466	-0.0157	-0.6796
Bellburns (BB)	6	0.5817	0.3278	0.0031	-0.7746
Total	80	0.5515	0.2944	0.0366	-0.8736

Table 3.2: Above the diagonal are pairwise FST (fixation index) values amongst eight *Braya* populations in the Limestone Barrens. Below the diagonal are distances in kilometres amongst the *Braya* populations obtained from Geographic Distance Matrix Generator v1.2.3 [80]. Population acronyms as in Table 3.1. Bold FST values correspond to the highest and lowest estimates within species, as discussed in the text.

		B. longii			*B. fernaldii*				
		SC	YP	VC	GIB	AP	WB	CN	BB
B. longii	SC	-	0.0403	**0.0146**	0.0600	0.0668	0.0584	0.0708	0.0638
	YP	6.67	-	**0.0483**	0.0877	0.0784	0.0858	0.0994	0.0981
	VC	1.93	4.74	-	0.0799	0.0832	0.0803	0.0992	0.0872
B. fernaldii	GIB	8.31	14.84	10.21	-	0.0754	0.0531	0.0728	0.0628
	AP	17.42	11.29	15.64	24.75	-	0.0818	**0.0454**	**0.0945**
	WB	59.05	65.56	60.96	50.75	74.47	-	0.0873	0.0677
	CN	57.49	64.04	59.40	49.20	73.14	2.48	-	0.0699
	BB	122.89	116.63	121.12	129.79	105.48	175.61	174.85	-

3.3 Population structure

Since both species are a) morphologically similar, b) are sister species in a phylogenetic study of the genus *Braya* [19], and c) are managed under the same action plan, I explored their population structure together and separately. The analysis in STRUCTURE for the two *Braya* species combined showed that the suggested number of genetic clusters (K) was eight, following the Evanno method [69]. I observed the following in the clusters:

- Savage Cove and Sandy Cove 1 of *B. longii* grouped in a single cluster. These are geographically close populations.
- Wild Bight and Belburns populations of *B. fernaldii* formed another genetic cluster. These are geographically distant populations.

- Half of Cape Norman individuals grouped with Anchor Point. These are geographically distant populations.
- The remaining populations formed their own genetic clusters (Figure 3.3A)

When STRUCTURE was run separately for each species, the genetic clusters observed were identical to the combined analysis, reinforcing the confidence of both analyses, with individuals from each species forming four genetic clusters, respectively (Figure 3.3B, 3.3C). Admixture was observed in every collection site, indicating evidence of past or current gene flow between populations and species.

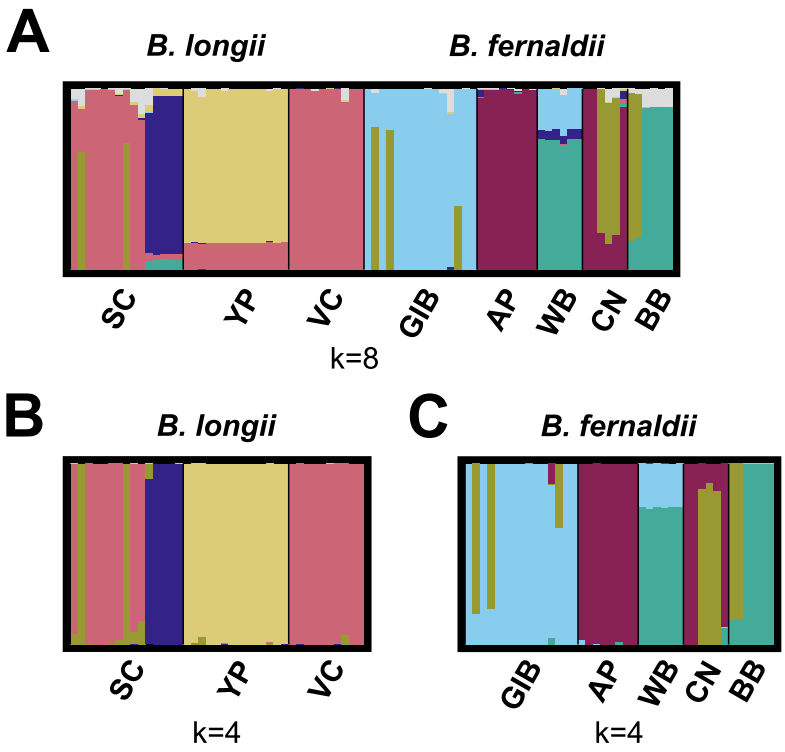

Figure 3.3: Bar plots showing the genetic structure of A) both *Braya* species combined for K of 8, B) *B. longii* populations only for K of 4, and C) *B. fernaldii* populations only for K of 4. Sampled localities are indicated at the bottom of the bar plots. (Table 2.1)

The first two discriminant functions of the DAPC analysis explained a total of 78% of genetic variation (50% and 28% respectively; Figure 3.4). The first axis separated mainly the Yankee Point (YP) and Sandy Cove 2 (SC2) populations of *B.*

longii from the rest. The second axis distinguished mainly the Savage Cove (VC) and Sandy Cove 1 (SC1) populations of *B. longii* from the rest, and these two populations overlap genetically. *Braya fernaldii* individuals appeared in the top right quadrant where all populations were discriminated except for Cape Norman (CN) and Anchor Point (AP). The Sandy Cove 2 population appeared between *B. longii* and *B. fernaldii* populations. In the second DAPC (Figure 3.5), which used cross-validation to inform eigenvalues, the first two discriminant functions explain 56% of genetic variation (31% and 25% respectively). In this DAPC, the two species exhibit more overlap with the *Braya fernaldii* population of Green Island Brook appearing among the *Braya longii* populations of Sandy Cove 1 and Savage Cove. The non-cross-validation DAPC (Figure 3.4) rendered results concordant with the STRUCTURE analysis where no differentiation was achieved between Savage Cove and Sandy Cove 1, and three individuals from Cape Norman appeared in the same genetic group as Anchor Point. The only difference between the DAPC and STRUCTURE results was that the *B. fernaldii* populations of Wild Bight and Bellburns were part of the same genetic cluster (in green) in STRUCTURE but were differentiated in the DAPC. Due to this consistency in results, the non-cross-validation DAPC (Figure 3.4) will be referenced as the main DAPC in this study.

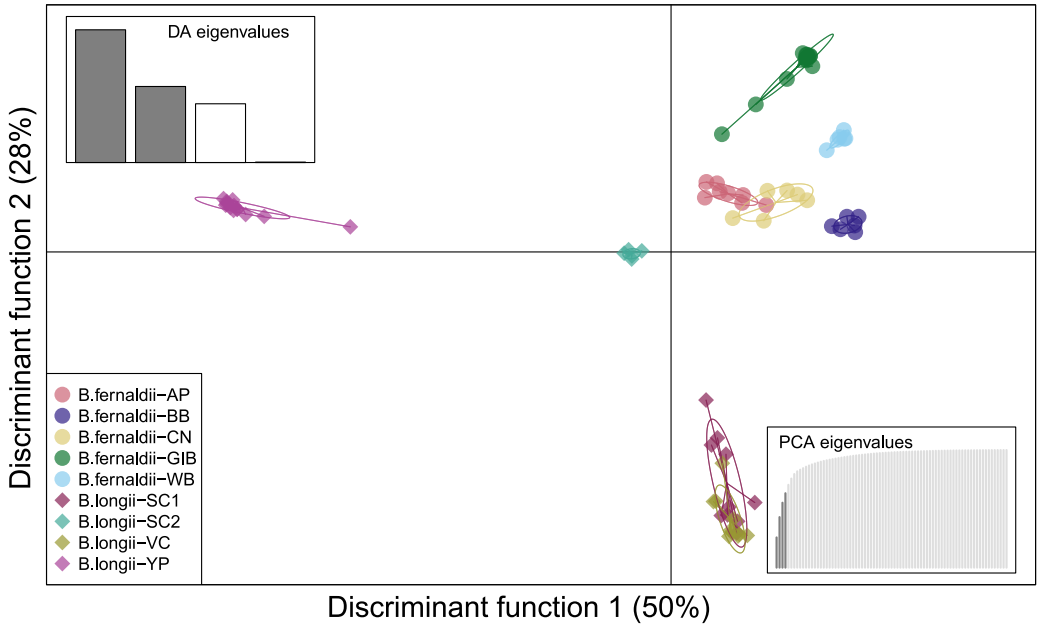

Figure 3.4: Visualization of the discriminant analysis of principal components (DAPC) using genotyping by sequencing data. Inset figures show DA (2) and PCA (4) eigenvalues. Sampling localities appear in different colours. Diamond markers correspond to *B. longii* individuals, and circles correspond to *B. fernaldii* individuals. AP = Anchor Point, BB = Bellburns, CN = Cape Norman, GIB = Green Island Brook, WB = Wild Bight, SC1 = Sandy Cove 1, SC2 = Sandy Cove 2, VC = Savage Cove, YP = Yankee Point

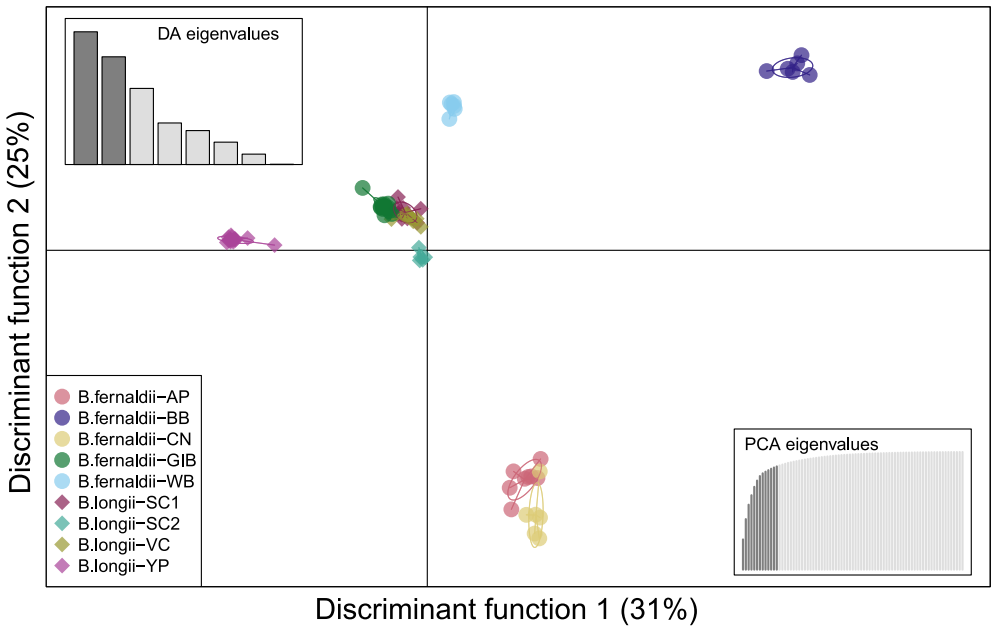

Figure 3.5: Visualization of the discriminant analysis of principal components (DAPC) using the suggested number of PCA eigenvalues (13) from the cross-validation function in adegenet (xvalDapc). Inset figures show DA (8) and PCA (13) eigenvalues. Sampling localities appear in different colours. Diamond markers correspond to *B. longii* individuals, and circles correspond to *B. fernaldii* individuals. AP = Anchor Point, BB = Bellburns, CN = Cape Norman, GIB = Green Island Brook, WB = Wild Bight, SC1 = Sandy Cove 1, SC2 = Sandy Cove 2, VC = Savage Cove, YP = Yankee Point

The AMOVA (Table 3.3) results of GBS data supported the significant population structure revealed in the first two methods. Most genetic variation (50.31%) was found among populations within *Braya* species, 21.89% of the total variation was among *Braya* species, and 27.81% was within populations (Table 3.3). At first view, the amount of variation (27.81%) within populations might seem high; however, as the STRUCTURE analysis indicates, some populations, such as Cape Norman and Wild Bight, have within-population genetic variation. To test whether these populations are driving the within-population variation up, I performed a second AMOVA, including only the five most homogeneous populations from the STRUCTURE analysis (Table 3.3): SC2, YP, VC, AP and BB. In this AMOVA including only homogenous populations (Table 3.4), 72% of the total variation was found among populations within species, 18.47% of the total variation was among *Braya* species, and only 9.59% was found within populations. Arlequin does not report confidence limits on variance

components.

Table 3.3: Analysis of molecular variance as estimated with SNPs using Arlequin [74].

Source of variation	Sum of squares	Variance components	Percentage variation
Among *Braya* species	1990.98	32.65	21.89
Among populations within species	4389.69	75.05	50.31
Within populations	2701.27	41.49	27.81
Total	9081.94	149.1941	100

Table 3.4: Analysis of molecular variance as estimated with SNPs using Arlequin [74] with only data from populations: *B.longii* (SC2, YP, VC), *B. fernaldii* (AP, and WB)

Source of variation	Sum of squares	Variance components	Percentage variation
Among *Braya* species	1002.63	20.44	18.47
Among populations within species	1870.50	79.62	71.94
Within populations	368.50	10.62	9.59
Total	3241.63	110.68	100

3.4 Phylogenetic analysis

The phylogenetic tree showed an evolutionary split between the two *Braya* species (Figure 3.6), supporting taxonomically different species. Each *B. fernaldii* population formed a distinct clade with the exception of Cape Norman, and each clade had bootstrap values of 20 to 90%. Cape Norman individuals are clustered together with Green Island Brook and Anchor Point clades. This pattern can also be observed in the STRUCTURE analysis where Cape Norman individuals share colours with individuals in both Green Island Brook and Anchor Point (Figure 3.3). Bellburns and Wild Bight formed two distinct clades with bootstrap support of 59.3 and 90%, respectively. Likewise, *B. longii* populations did not all form distinct clades. Savage Cove and Sandy Cove 1 formed one lineage (bootstrap of 68%) corroborating the DAPC and STRUCTURE results. Yankee Point and Sandy Cove 2 formed two distinct clades with bootstrap support of 91 and 100%, respectively. As in the DAPC, the Sandy Cove 2 branch is found in between the two species of *Braya*: *B. longii* and *B. fernaldii*.

In this chapter, the genomic structure and phylogenetic analyses supported the taxonomic distinction of *B. longii* and *B. fernaldii*, and the genetic differentiation amongst most collecting sites. Based on these results, I propose MUs and ESUs for conserving these species, as discussed in the next chapter.

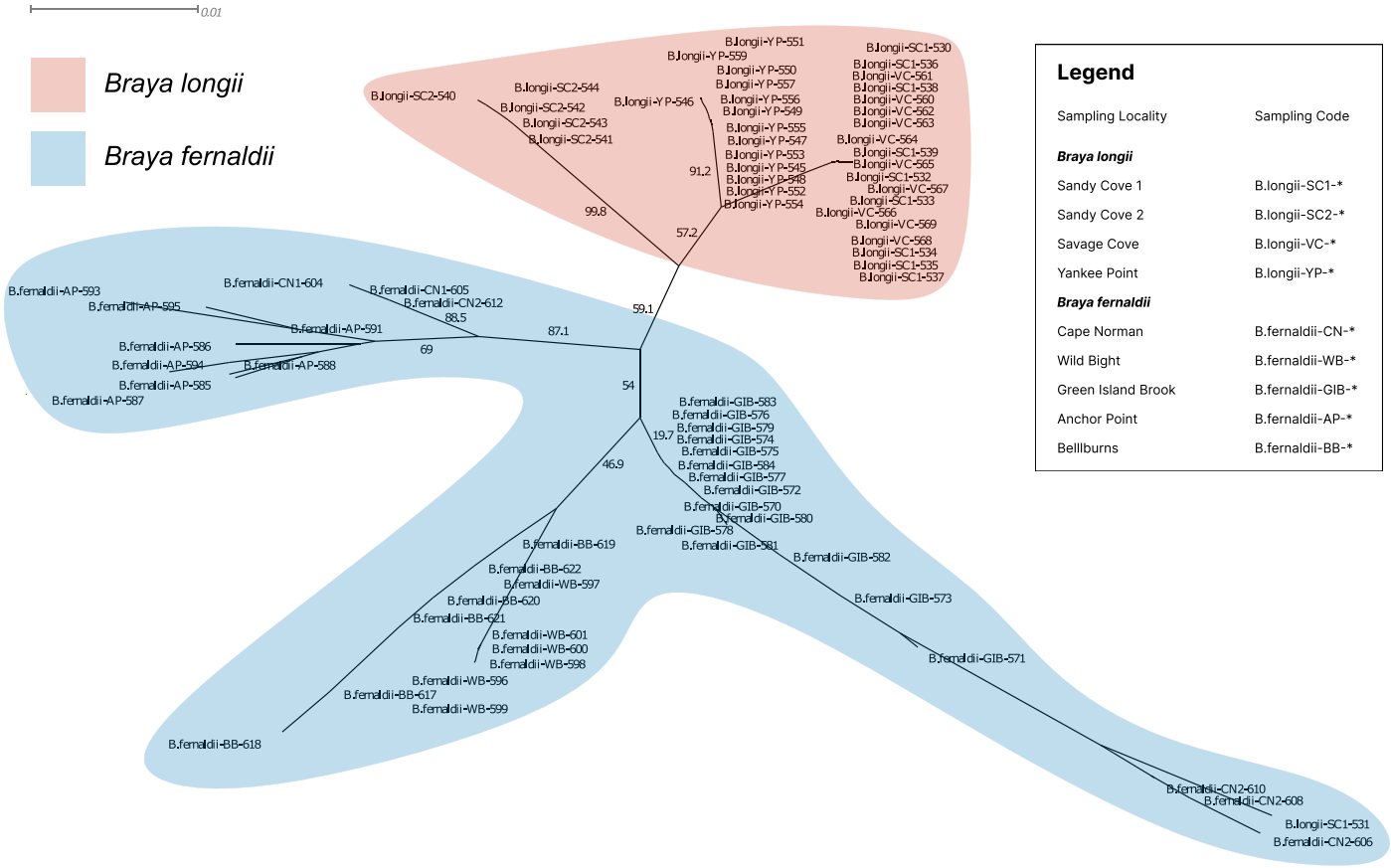

Figure 3.6: Neighbour-joining tree of the 80 *B. longii* and *B. fernaldii* individuals sampled in the Limestone Barrens of Newfoundland. The SNP dataset (1,449 SNPs) was filtered to account for polyploidization. Numbers along branches are bootstrap support values.

Chapter 4

Discussion

Using patterns observed from the genetic diversity and structure results, I propose seven MUs to conserve these two *Braya* species. Based on the phylogenetic results, I recommend six ESUs. I recommend a diverse collection representing each MU and ESU for ex-situ collections. For conservation translocations, careful considerations must be made when deciding which plants or seeds should be introduced to a given area. They should be from a location with similar ecology, genetically diverse, and reproductively healthy [81, Chapter 8]. Due to the complex nature of polyploidism and the various bioinformatic considerations, future genetic research on *Braya longii* and *Braya fernaldii* can still provide novel and increased insight into the genetic relationships between these species.

4.1 Genetic diversity and structure in *Braya longii* and *Braya fernaldii*

Despite the need for a genetic understanding when managing populations [7], there are few plant-focused conservation genetic studies in Canada that discuss MUs or ESUs. However, there are other Canadian conservation plant studies that do leverage molecular markers such as Hills Thistle, Few-flowered Club-rush, Cucumber tree, Eastern Prairie White Fringed Orchid, Pitch Pine, Sugar Maple, Canada Yew and Whitebark Pine [82–88].

Canada Frostweed (*Helianthemum canadense*) and Pitchers Thistle (*Cirsium pitcheri*)

(a) Canada Frostweed (*Helianthemum canadense*) [91]

(b) Pitchers Thistle (*Cirsium pitcheri*) [92]

Figure 4.1: Flowers of Canada Frostweed and Pitchers Thistle

(Figure 4.1) [35, 36] are two Canadian conservation genetic studies which use conservation units. While neither study uses GBS data, they are both evaluating population genetics for the purpose of management. The study on Canada Frostweed used Amplified Fragment Length Polymorphism (AFLP) data [89], and the study on Pitchers Thistle used nuclear microsatellite data. An AMOVA and a STRUCTURE bar plot were analyzed for Canada Frostweed, while Pitchers Thistle leveraged diversity and differentiation statistics. The genetic analysis on *B. longii* and *B. fernaldii* includes evaluation of population structure with STRUCTURE bar plots and DAPC, diversity statistics, and phylogenetic analysis. While these methods are frequently used for agrobiodiversity, such as in watermelon [90], it is less common to find these genetics methods applied to plant conservation studies in Canada.

Canada Frostweed is found in Atlantic Canada and has a genetic evaluation with the outcome of proposed conservation units [35]. Due to declining numbers, individuals from Nova Scotia, Maine, New Hampshire, and Quebec were examined to gain insight into the population genetics for conservation. Similar to what I observed in *Braya*, there is a very geographically separated genetic grouping of Frostweed populations (between Queens County, Nova Scotia and Quebec), which does not align with the rest of the Frostweed studies' genetic groupings corresponding to collection sites. Based on the Frostweed studies' hypothesis that a common source may have introduced both of these populations [35] and the *Braya* phylogenetic analysis (Fig. 3.6),

I recommend that Wild Bight and Bellburns be considered a single ESU despite the sizeable geographic separation between these populations. The reasons behind these two geographically distant populations being genetically close are unclear. Potential explanations could include the transportation of material during road development and the movement of seeds via animals, such as ducks. However, further research is needed to explain this observation.

However, unlike the Frostweed study, which estimated the majority (72%) of variation to be within populations, 9% of variation among populations, and 18% of variation among regions, the majority (50%) of variation for *B. longii* and *B.fernaldii* is found between populations, with 27.81% of variation being within populations, and 21.89% of the total variation being among *Braya* species (Table 3.3). For *B. longii* and *B. fernaldii*, I expect the higher levels of population differentiation to result from the selfing breeding system and small-range seed dispersal. While the 21% variation among species may seem low, *B. longii* and *B. fernaldii* are very morphologically similar and can not be differentiated on one feature alone.

B. longii and *B. fernaldii* primarily self-fertilize, with *B. longii* having a higher outcrossing potential than *B. fernaldii* [20, 93]. In greenhouse tests, Harris [93] found *B. longii* to fully self-fertilize while Parsons and Hermanutz [20] provided evidence of outcrossing potential using hand-pollination experiments. The exact natural selfing and outcrossing rates are unknown. The three separate genetic structure analyses supported the taxonomic distinction of *B. longii* and *B. fernaldii*, and generally align with genetic patterns of self-pollination [47] by having the majority of (50%) of variation (Table 3.3) for *B. longii* and *B.fernaldii* be found between populations [20]. The population of Sandy Cove 2 appeared between both the *B. fernaldii* and *B. longii* populations on the DAPC (Fig. 3.4) and phylogenetic tree (Fig. 3.6). Sandy Cove is bisected by a highway, with Sandy Cove 1 and Sandy Cove 2 located on each side of the road. Road development may seriously impact the distribution and relocation of *Braya* seeds. I hypothesize that the separation of individuals from the Sandy Cove site in two lineages is because of the amount of substrate movement among quarries in the *B. longii*'s distribution. This clustering suggests that the substrate may have been moved between populations at some time. Anecdotal information from quarrying operators supports this hypothesis (Copp and Hermanutz, pers. comm.). Due to the fragmented habitat, SC2 might be a genetic intermediate form/hybrid between *B. fernaldii* and *B.longii*. In a previous study that assessed the hybridization potential

of *B. longii* and *B. fernaldii*, by using floral measurements and hand-pollination experiments, hybridization potential was identified [20]. In wild populations, observed individuals exhibiting intermediate morphology between these species further provided evidence for this hybridization potential. I recommend that future genetic work on *B. fernaldii* and *B. longii* further explores Sandy Cove's nuances.

The diversity statistics of *B. longii* and *B. fernaldii* mirrors patterns seen in a study on *Senna glutinosa* subsp. *glutinosa* (Figure 4.2). This study uses chloroplast regions and nuclear microsatellite markers to gain insight into 480 plants from 20 populations across the Pilbara bioregion [94].

Figure 4.2: Photo of *Senna glutinosa* ssp. *glutinosa* [95] under the following creative commons license: https://creativecommons.org/licenses/by-nc/4.0/

Senna glutinosa is also capable of selfing but it is primarily apomictic. Apomixis is asexual seed development. In this study, similar to what I observed in *Braya*, a high estimated heterozygosity is shown along with negative inbreeding coefficients. This is likely because *Senna glutinosa* subsp. *glutinosa* are allopolyploids [94].

4.2 Polyploidism in plant genetic studies

Challenges with the genetic analysis of polyploids often come from inferring allele frequencies and assumptions about inheritance [96]. Within a genome, inheritance patterns may vary within individuals, resulting in polysomic or disomic inheritance at different loci. Polysomic inheritance is when all variants of the same chromosome can pair in meiosis, while disomic inheritance is the preferential pairing between the chromosomes derived from the same ancestral species. Polysomic inheritance is seen in autopolyploids, while disomic inheritance is seen in allopolyploids [96]. Many genetic analytical frameworks make assumptions regarding inheritance and the dosage of alleles at individual loci. Resolving dosages of alleles is a prerequisite to calculating

observed and expected allele frequencies. The evaluation of inbreeding coefficients depends on observed and expected heterozygosity, which is a problem due to the uncertainty regarding allele dosage. This may result in inaccurate conclusions regarding genetic structure and diversity.

The FIS values are negative in Table 3.1. A disomic inheritance leads to fixed heterozygosity, which would cause negative FIS values [31]. A study [31] illustrated (Figure 4.3) how assumptions of tetrasomy when a species is entirely disomic affect diversity statistics. In this tetrasomy case study, HS is higher than expected, FST is much lower, and FIS values are negative. There is evidence that *Braya longii* and *Braya fernaldii* are octoploid allopolyploids [19]. Allopolyploids tend to have disomic inheritance [96] because chromosomes from the same ancestral genome exhibit preferential pairing. The patterns seen under the tetrasomy/disomic scenario studied in [31] mirror what I have found in *Braya*: negative FIS values, HS higher than expected and FST lower than expected. The octoploid nature of *Braya* implies that this over/underestimation of the diversity statistics would be further exaggerated.

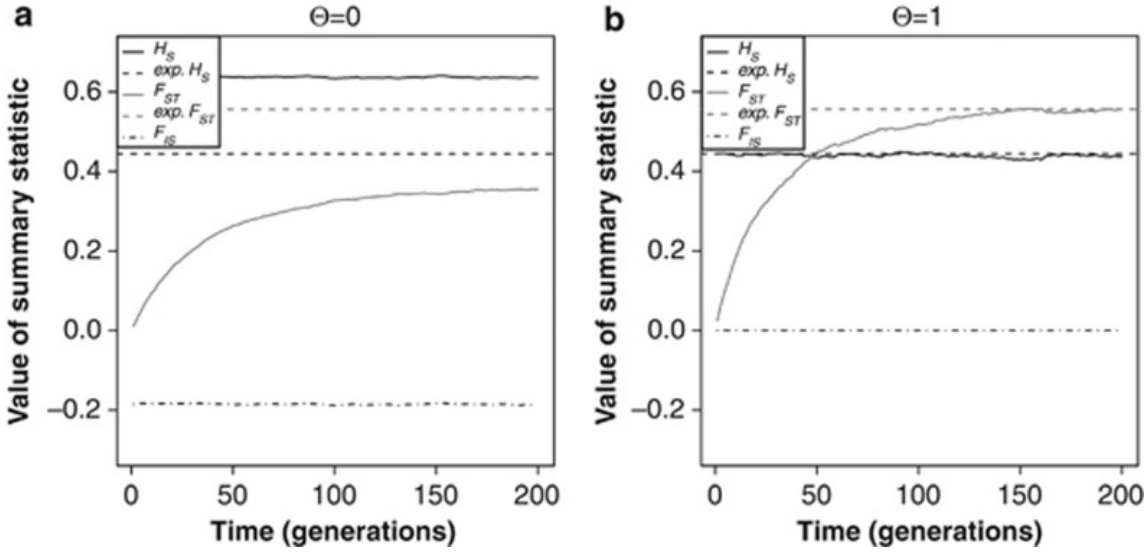

Figure 4.3: The expected heterozygosity HS, the inbreeding coefficients FIS and FST as a function of the number of generations since population divergence under (**a**) full disomy and (**b**) full tetrasomy. The dotted lines show the theoretical expectations for HS and FST. Figure from [31] reproduced with permission from Springer Nature.

There are many challenges that polyploidism introduces to population genetics analyses. One of the major problems is that GBS methods do not handle genotyping duplicated loci, and biologists must filter duplicated loci, which results in less

insightful and comprehensive outcomes and interpretations [97]. Many bioinformatics tools such as OneMap or JoinMap [98–100] only handle haploid or diploid data due to mathematic assumptions. While several data pipelines for GBS data exist, such as UGbS-Flex and Stacks [101, 102], I chose GBS-SNP-CROP [52] for both ease-of-use and for the ability to filter out paralogs. By using the -altStrength parameter when performing the filtering step (7th script) of GBS-SNP-CROP, only strongly bi-allelic SNPs are retained. This parameter filters on a population-level allele frequency [57]. A study on the *Centaurea aspera* L. complex, which includes an autotetraploid subspecies [103], also leveraged GBS-SNP-CROP to obtain SNPs for downstream analysis and used an alternate allele strength value of 0.9 in comparison to the value 0.962 used in this *Braya* study. This higher value was chosen for stricter levels of error control in comparison to the suggested levels for an octoploid of 0.820 [57]. Due to the polyploid nature of these plants and the challenges with inferring allele frequencies, the absolute values of the results should not be viewed as absolute. Instead, all the results together and their corresponding patterns provide insight. The selfing rate is important to consider when interpreting our results because of the high calculated heterozygosity and negative inbreeding values. Such strongly negative FIS values would be remarkable for species with outcrossing potential, and for species that primarily self-fertilize, they would be very unlikely. Our hypothesis is that the FIS values are actually reflecting fixed heterozygosity which is common in allopolyploids.

4.3 Bioinformatic considerations for SNP discovery

It has been shown that variations in bioinformatics pipelines affect the results obtained from these pipelines; for example, in genetic variant identification and in metagenomics analyses [104–107]. With respect to SNP discovery using GBS data, many different software can be used at each step of the pipeline, for example, for filtering, demultiplexing, aligning, and sequencing data. Depending on the software and parameters used, results might differ. For example, the GBS-SNP-CROP Mock Reference workflow identified 18 times more SNPs than TASSEL-UNEAK in a pipeline performance study when using paired-end data from 48 accessions of *Actinidia arguta* [57].

Additionally, differences in the SNPs identified will influence all downstream analyses. While intentionally manipulated for methods validation, Table 3.3 and Table 3.4 illustrated how alterations in data affect final results. The analysis software for this *B. fernaldii* and *B. longii* study was chosen based on benchmarks, accessibility and ease of use. Parameters that provided the most high-quality SNPs were chosen based on the literature and my own trials, leading to various results (Tables 3.1, 3.2, 3.3) that supported one another, were interpretable, and were supported by ecological theory. For example, parameters that lead to results that align with the knowledge that genetics of rare, endemic species often exhibit low levels of genetic diversity and high levels of population differentiation from inbreeding [28] were favoured over parameters that lead to uninterpretable, noisy results.

4.4 Proposed conservation units

While there are many different ways to organize population subgroups for management, Canada's Species at Risk Act recognizes designatable units (DUs). The Committee on the Status of Endangered Wildlife in Canada (COSEWIC) recognizes a DU if it has attributes which are discrete and evolutionarily significant [33]. *Braya longii* and *Braya fernaldii* are each a separate designatable unit under COSEWIC [108, 109]. Management units (MUs) are separately managed population groups or subgroups because they are genetically distinct, while ESUs represent evolutionary lineages. ESUs may contain one or many MUs. While people have provided strategies for designating conservation units, there is no standard approach. Hoelzel [110] suggests leveraging the FST statistic to inform the designation of evolutionarily significant units. However, in this study, FST values cannot be directly used due to the polyploid nature of these plants. Instead of this method, I am using patterns of genetic structure to conserve based on evolutionary trajectory.

Based on the genetic diversity and structure results, I propose seven MUs for the conservation of these two *Braya* species. These are the following for *B. longii*: 1) Sandy Cove 1 and Savage Cove (SC1 & VC), 2) Yankee Point (YP), and 3) Sandy Cove 2 (SC2). For *B. fernaldii*: 4) Anchor Point (AP) and Cape Norman (CN), 5) Wild Bight (WB), 6) Belburns (BB), and 7) Green Island Brook (GIB).

Based on the phylogenetic results, I recommend six ESUs. For *B. longii*: 1) Sandy

Cove 1 and Savage Cove (SC1 & VC), 2) Yankee Point (YP), 3) Sandy Cove 2 (SC2). For *B. fernaldii* : 4) Anchor Point and Cape Norman (AP & CN), 5) Wild Bight and Belburns (WB & BB), and 6) Green Island Brook (GIB). Gravel movement might partially explain some of the observed genetic patterns. However, further research is needed to fully explain the genetic patterns observed.

Based on my results, if there were an interest in reducing management units, the only separate populations that might be combined and treated as a single management unit would be SC1 and VC. However, treating each population as a separate management unit is a more conservative strategy.

4.5 Recommended strategies for long-term *in-situ* and *ex-situ* conservation

Braya longii and *B. fernaldii* have a vegetation height of less than 10cm and inhabit restricted ranges to the naturally occurring Limestone Barrens habitat on the Northern Limestone Barren of Newfoundland [46]. With such few populations, and with such minimal adequate potential habitat, maintaining an ex-situ collection and translocating individuals to areas of location extirpation or new suitable ranges may be a good conservation strategy. The Hawaiian archipelago has several case studies from the Hawaiian Rare Plant Program which outlines conservation efforts on species such as *Cyanea pinnatifda* and *Cyanea grimesiana* subsp. *grimesiana* [111]. I recommend a diverse collection representing each MU and ESU for ex-situ collections. For conservation translocations, careful considerations must be made when deciding which plants or seeds should be introduced to a given area. They should be from a location with similar ecology, genetically diverse, and reproductively healthy [81, Chapter 8]. Using the MUs defined as an outcome of this research, I recommend sourcing individuals from the relevant MU for the given reintroduction.

This study is a stepping stone in satisfying the high-priority species-at-risk-act requirement of reintroducing genetically appropriate individuals to source populations, populations with reduced sizes, or extirpated populations [46] by proposing management units to inform conservation experts. The results of this study also facilitate the planning of future population samplings and the curating of ex-situ collections.

4.6 Conclusions and future research

Using a large dataset of DNA sequences, I propose different management and evolutionary significant units to maintain an ex-situ collection of *Braya*'s genetic diversity to be used in future reintroductions when necessary. While my analyses have shown population differentiation and allowed me to suggest ESUs and MUs, future research can also leverage this high-quality SNP dataset. Future research should estimate the divergence times of each *Braya* lineage or ESU to gain insight into the temporal scale of evolution of these populations in the Limestone Barrens. My SNP dataset could also be used to estimate past and current gene flow, effective population sizes, and test population expansion or contraction, which will complement the management of these endangered species. Each of these additional analyses would provide significant insight and context into the population genetics of *B. longii* and *B. fernaldii*, and they would greatly improve the current understanding of their populations.

Bibliography

1. Stronza, A. L., Hunt, C. A. & Fitzgerald, L. A. Ecotourism for Conservation?, 27 (2019).
2. Sakurai, R., Ota, T. & Uehara, T. Sense of place and attitudes towards future generations for conservation of coastal areas in the Satoumi of Japan. *Biological Conservation* **209,** 332–340. ISSN: 00063207 (May 2017).
3. Gaitán-Espitia, J. D. & Hobday, A. J. Evolutionary principles and genetic considerations for guiding conservation interventions under climate change. *Global Change Biology* **27,** 475–488. ISSN: 1354-1013, 1365-2486 (Feb. 2021).
4. De Vos, J. M., Joppa, L. N., Gittleman, J. L., Stephens, P. R. & Pimm, S. L. Estimating the normal background rate of species extinction. *Conservation Biology* **29,** 452–462. ISSN: 0888-8892, 1523-1739 (Apr. 2015).
5. Shin, Y. *et al.* Actions to halt biodiversity loss generally benefit the climate. *Global Change Biology* **28,** 2846–2874. ISSN: 1354-1013, 1365-2486 (May 2022).
6. IUCN. *Global Species Action Plan: Supporting implementation of the Kunming-Montreal Global Biodiversity Framework* (IUCN, Gland, Switzerland, 2023).
7. Frankham, R. Genetics and extinction. *Biological Conservation* **126,** 131–140. ISSN: 00063207 (Nov. 2005).
8. Di Sacco, A. *et al.* Ten golden rules for reforestation to optimize carbon sequestration, biodiversity recovery and livelihood benefits. *Global Change Biology* **27,** 1328–1348. ISSN: 1354-1013, 1365-2486 (Apr. 2021).
9. Lutz, M. L. *et al.* Using multiple sources during reintroduction of a locally extinct population benefits survival and reproduction of an endangered freshwater fish. *Evolutionary Applications* **14,** 950–964. ISSN: 1752-4571, 1752-4571 (Apr. 2021).

10. Turchetto, C. *et al.* High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. *AoB Plants* **8,** plw002. ISSN: 2041-2851 (2016).

11. Ureta, C. *et al.* Maize yield in Mexico under climate change. *Agricultural Systems* **177,** 102697. ISSN: 0308521X (Jan. 2020).

12. Sinclair, A. R. E., Fryxell, J. M., Caughley, G. & Caughley, G. *Wildlife ecology, conservation, and management* 2nd ed. OCLC: ocm58526307. 469 pp. ISBN: 978-1-4051-0737-2 978-1-4051-3806-2 (Blackwell Pub, Malden, MA ; Oxford, 2006).

13. *Restoring diversity: strategies for reintroduction of endangered plants* (eds Falk, D. A., Millar, C. I. & Olwell, M.) (Island Press, Washington, D.C, 1996). 505 pp. ISBN: 978-1-55963-296-6 978-1-55963-297-3.

14. *Plant Reintroduction in a Changing Climate: Promises and Perils* (eds Maschinski, J., Haskins, K. E. & Raven, P. H.) (Island Press/Center for Resource Economics, Washington, DC, 2012). ISBN: 978-1-59726-343-6 978-1-61091-183-2.

15. Maschinski, J. & Albrecht, M. A. Center for Plant Conservation's Best Practice Guidelines for the reintroduction of rare plants. *Plant Diversity* **39,** 390–395. ISSN: 24682659 (Dec. 2017).

16. *Conservation and Reintroduction of Rare and Endangered Plants in China* (ed Ren, H.) (Springer Singapore, Singapore, 2020). ISBN: 9789811553004 9789811553011.

17. Roncal, J., Maschinski, J., Schaffer, B., Gutierrez, S. M. & Walters, D. Testing appropriate habitat outside of historic range: The case of Amorpha herbacea var. crenulata (Fabaceae). *Journal for Nature Conservation* **20,** 109–116. ISSN: 16171381 (Mar. 2012).

18. Grossen, C., Biebach, I., Angelone-Alasaad, S., Keller, L. F. & Croll, D. Population genomics analyses of European ibex species show lower diversity and higher inbreeding in reintroduced populations. *Evolutionary Applications* **11,** 123–139. ISSN: 17524571 (Feb. 2018).

19. Warwick, S. I., Al-Shehbaz, I. A., Sauder, C., Harris, J. G. & Koch, M. Phylogeny of Braya and Neotorularia (Brassicaceae) based on nuclear ribosomal internal transcribed spacer and chloroplast trnL intron sequences. *Canadian Journal of Botany* **82,** 376–392. ISSN: 0008-4026 (Mar. 1, 2004).

20. Parsons, K. & Hermanutz, L. Conservation of rare, endemic braya species (Brassicaceae): Breeding system variation, potential hybridization and human disturbance. *Biological Conservation* **128,** 201–214. ISSN: 00063207 (Mar. 2006).

21. Tilley, S. *The Factors Governing the Distribution of the Rare Plants* Honours (Memorial University of Newfoundland, Apr. 2003).

22. Canada, E. C̃. C. *Long's Braya (Braya longii) and the Fernald's Braya (Braya fernaldii): amended action plan 2020* aem. Last Modified: 2020-10-27. https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/action-plans/longs-fernalds-braya-2020.html (2020).

23. *Species Profile (Long's Braya) - Species at Risk Public Registry* https://wildlife-species.canada.ca/species-risk-registry/.

24. *Species Profile (Fernald's Braya) - Species at Risk Public Registry* https://wildlife-species.canada.ca/species-risk-registry/.

25. Meades, S. & Meades, W. J. *Gallery - Braya longii (Long's braya) - Flora of Newfoundland and Labrador* Flora of Newfoundland and Labrador. https://newfoundland-labradorflora.ca/gallery/index.cfm?ParentID=105&alpha=b (2020).

26. Meades, S. & Meades, W. J. *Gallery - Braya fernaldii (Fernald's braya) - Flora of Newfoundland and Labrador* Flora of Newfoundland and Labrador. https://newfoundland-labradorflora.ca/gallery/index.cfm?ParentID=103&alpha=b (2023).

27. Hamrick, J. L. & Godt, M. J. W. in *Plant population genetics, breeding, and genetic resources* (eds Brown, A. H. D., Clegg, M. T., Kahler, A. L. & Weir, B. S.) Meeting Name: International Symposium on Population Genetics and Germplasm Resources in Crop Improvement, 43–63 (Sinauer Associates, Sunderland, Mass, 1990). ISBN: 978-0-87893-116-3 978-0-87893-117-0.

28. Pez-Pujol, J. L., Bosch, M. & Simon, J. Population genetics and conservation priorities for the critically endangered island endemic Delphinium pentagynum subsp. formenteranum (Ranunculaceae), 16.

29. García-Verdugo, C. *et al.* Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology* **24,** 726–741. ISSN: 0962-1083, 1365-294X (Feb. 2015).

30. Meirmans, P. G. genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources* **20,** 1126–1131. ISSN: 1755-098X, 1755-0998 (July 2020).

31. Meirmans, P. G. & Van Tienderen, P. H. The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity* **110,** 131–137. ISSN: 0018-067X, 1365-2540 (Feb. 2013).

32. Custance, T. L. *Conservation genetics of the endangered plants, Braya longii and B. fernaldii (Brassicaceae)* Honours (Memorial University of Newfoundland, St. John's, Newfoundland, Canada, 2019).

33. *COSEWIC guidelines for recognizing designatable units* Committee on the Status of Endangered Wildlife in Canada. `https://cosewic.ca/index.php/en/reports/preparing-status-reports/guidelines-recognizing-designatable-units` (2020).

34. Funk, W. C., McKay, J. K., Hohenlohe, P. A. & Allendorf, F. W. Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution* **27,** 489–496. ISSN: 01695347 (Sept. 2012).

35. Yorke, A. F., Mockford, S. & Evans, R. C. Canada frostweed (*Helianthemum canadense* (L.) Michx.; Cistaceae) at the northeastern limit of its range: implications for conservation. *Botany* **89,** 83–89. ISSN: 1916-2790, 1916-2804 (Feb. 2011).

36. Gauthier, M. *et al.* Conservation genetics of Pitcher's thistle (*Cirsium pitcheri*), an endangered Great Lakes endemic. *Botany* **88,** 250–257. ISSN: 1916-2790, 1916-2804 (Mar. 2010).

37. Willi, Y. *et al.* Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proceedings of the National Academy of Sciences* **119,** e2105076119. ISSN: 0027-8424, 1091-6490 (Jan. 5, 2022).

38. Lu, J. *et al.* New insights of CYP1A in endogenous metabolism: a focus on single nucleotide polymorphisms and diseases. *Acta Pharmaceutica Sinica B* **10,** 91–104. ISSN: 22113835 (Jan. 2020).

39. Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12,** 499–510. ISSN: 1471-0056, 1471-0064 (July 2011).

40. Silliman, K. Population structure, genetic connectivity, and adaptation in the Olympia oyster (*Ostrea lurida*) along the west coast of North America. *Evolutionary Applications* **12,** 923–939. ISSN: 1752-4571, 1752-4571 (June 2019).

41. Kling, M. M. & Ackerly, D. D. Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Sciences* **118,** e2017317118. ISSN: 0027-8424, 1091-6490 (Apr. 27, 2021).

42. Charlesworth, D., Barton, N. H. & Charlesworth, B. The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences* **284,** 20162864. ISSN: 0962-8452, 1471-2954 (May 31, 2017).

43. Edmands, S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management: RELATIVE RISKS OF INBREEDING AND OUTBREEDING. *Molecular Ecology* **16,** 463–475. ISSN: 09621083, 1365294X (Nov. 15, 2006).

44. Pamilo, P. & Savolainen, O. Post-Glacial Colonization, Drift, Local Selection and Conservation Value of Populations: A Northern Perspective. *Hereditas* **130,** 229–238. ISSN: 00180661 (May 6, 2004).

45. Warwick Crop Centre, The University of Warwick. *Genetic markers: SNPs* warwick.ac.uk. `https://warwick.ac.uk/fac/sci/lifesci/research/vegin/geneticimprovement/geneticmarker/snp/` (2023).

46. Canada & Environment and Climate Change Canada. *Amended action plan for the Long's braya (Braya longii) and the Fernald's braya (Braya fernaldii) in Canada.* OCLC: 1240697475. ISBN: 978-0-660-35733-1 (2020).

47. J. L. Hamrick & M. J. W. Godt. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **351,** 1291–1298. ISSN: 0962-8436, 1471-2970 (Sept. 30, 1996).

48. Reutemann, A. V. *et al.* Comparative analysis of molecular and morphological diversity in two diploid Paspalum species (Poaceae) with contrasting mating systems. *Plant Reproduction* **37,** 15–32. ISSN: 2194-7953, 2194-7961 (Mar. 2024).

49. Poland, J. A., Brown, P. J., Sorrells, M. E. & Jannink, J.-L. Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE* **7** (ed Yin, T.) e32253. ISSN: 1932-6203 (Feb. 28, 2012).

50. Abed, A. *et al.* in *Barley* (ed Harwood, W. A.) Series Title: Methods in Molecular Biology, 233–252 (Springer New York, New York, NY, 2019). ISBN: 978-1-4939-8942-3 978-1-4939-8944-7.

51. *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data* `https://www.bioinformatics.babraham.ac.uk/projects/fastqc/` (2024).

52. Melo, A. T. O. & Hale, I. Expanded functionality, increased accuracy, and enhanced speed in the de novo genotyping-by-sequencing pipeline GBS-SNP-CROP. *Bioinformatics* **35** (ed Stegle, O.) 1783–1785. ISSN: 1367-4803, 1367-4811 (May 15, 2019).

53. McKinney, G. J., Waples, R. K., Seeb, L. W. & Seeb, J. E. Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations. *Molecular Ecology Resources* **17,** 656–669. ISSN: 1755098X (July 2017).

54. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120. ISSN: 1460-2059, 1367-4803 (Aug. 1, 2014).

55. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620. ISSN: 1367-4803, 1460-2059 (Mar. 1, 2014).

56. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4,** e2584. ISSN: 2167-8359 (Oct. 18, 2016).

57. Melo, A. T. O., Bartaula, R. & Hale, I. GBS-SNP-CROP: a reference-optional pipeline for SNP discovery and plant germplasm characterization using variable length, paired-end genotyping-by-sequencing data. *BMC Bioinformatics* **17,** 29. ISSN: 1471-2105 (Dec. 2016).

58. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760. ISSN: 1367-4803, 1460-2059 (July 15, 2009).

59. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079. ISSN: 1367-4803, 1460-2059 (Aug. 15, 2009).

60. Karim, M. M. & Yu, F. Identification of QTLs for resistance to 10 pathotypes of Plasmodiophora brassicae in Brassica oleracea cultivar ECD11 through genotyping-by-sequencing. *Theoretical and Applied Genetics* **136,** 249. ISSN: 0040-5752, 1432-2242 (Dec. 2023).

61. Lab, H. *GBS-SNP-CROP* original-date: 2015-09-04T20:32:03Z. `https://github.com/halelab/GBS-SNP-CROP` (2023).

62. Goudet, J. hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes* **5,** 184–186. ISSN: 1471-8278, 1471-8286 (Mar. 2005).

63. Knaus, B. J. & Grünwald, N. J. vcf : a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources* **17,** 44–53. ISSN: 1755-098X, 1755-0998 (Jan. 2017).

64. Nei, M. *Molecular evolutionary genetics* (Columbia university press, 1987).

65. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155,** 945–959. ISSN: 1943-2631 (June 1, 2000).

66. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4,** 7. ISSN: 2047-217X (Dec. 2015).

67. Pritchard, J. K., Wen, X. & Falush, D. Documentation for structure software: Version 2.3, 39.

68. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4,** 359–361. ISSN: 1877-7252, 1877-7260 (June 2012).

69. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14,** 2611–2620. ISSN: 0962-1083, 1365-294X (July 2005).

70. Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23,** 1801–1806. ISSN: 1367-4803, 1460-2059 (July 15, 2007).

71. Rosenberg, N. A. distruct: a program for the graphical display of population structure: PROGRAM NOTE. *Molecular Ecology Notes* **4,** 137–138. ISSN: 14718278, 14718286 (Dec. 10, 2003).

72. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24,** 1403–1405. ISSN: 1460-2059, 1367-4803 (June 1, 2008).

73. Jombart, T. & Collins, C. A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0.

74. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources,* 564–567 (Apr. 2010).

75. Ortiz, E. M. *vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis.* Jan. 15, 2019.

76. Huson, D. H. & Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23,** 254–267. ISSN: 1537-1719, 0737-4038 (Feb. 1, 2006).

77. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-Calling of Automated Sequencer Traces Using *Phred.* I. Accuracy Assessment. *Genome Research* **8,** 175–185. ISSN: 1088-9051, 1549-5469 (Mar. 1, 1998).

78. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32,** 3047–3048. ISSN: 1367-4811, 1367-4803 (Oct. 1, 2016).

79. *basic.stats: Basic diversity and differentiation statistics in hierfstat: Estimation and Tests of Hierarchical F-Statistics* R Package Documentation. `https://rdrr.io/cran/hierfstat/man/basic.stats.html` (2024).

80. Ersts, P. *Geographic Distance Matrix Generator (version 1.2.3)* American Museum of Natural History, Center for Biodiversity and Conservation. `https://biodiversityinformatics.amnh.org/open_source/gdmg/` (2023).

81. *Ecological Restoration: Moving Forward Using Lessons Learned* (eds Florentine, S., Gibson-Roy, P., Dixon, K. W. & Broadhurst, L.) (Springer International Publishing, Cham, 2023). ISBN: 978-3-031-25411-6 978-3-031-25412-3.

82. Freeland, J. R., Gillespie, J., Ciotir, C. & Dorken, M. E. Conservation genetics of Hill's thistle (*Cirsium hillii*). *Botany* **88,** 1073–1080. ISSN: 1916-2790, 1916-2804 (Dec. 2010).

83. Nowell, V. *Conservation genetics and habitat characteristics of Few-flowered Club-rush (Trichophorum planifolium (Spreng.) Palla), an endangered woodland sedge in Canada* Master of Science (Carleton University, Ottawa, Ontario, 2015).

84. Paul, J., Budd, C. & Freeland, J. R. Conservation genetics of an endangered orchid in eastern Canada. *Conservation Genetics* **14,** 195–204. ISSN: 1566-0621, 1572-9737 (Feb. 2013).

85. Mosseler, A., Rajora, O., Major, J. & Kim, K. Reproductive and genetic characteristics of rare, disjunct pitch pine populations at the northern limits of its range in Canada. *Conservation Genetics* **5,** 571–583. ISSN: 1566-0621, 1572-9737 (2004).

86. Young, A., Warwick, S. & Merriam, H. Genetic variation and structure at three spatial scales for *Acer saccharum* (sugar maple) in Canada and the implications for conservation. *Canadian Journal of Forest Research* **23,** 2568–2578. ISSN: 0045-5067, 1208-6037 (Dec. 1, 1993).

87. Senneville, S., Beaulieu, J., Daoust, G., Deslauriers, M. & Bousquet, J. Evidence for low genetic diversity and metapopulation structure in Canada yew (*Taxus canadensis*): considerations for conservation. *Canadian Journal of Forest Research* **31,** 110–116. ISSN: 0045-5067, 1208-6037 (Jan. 1, 2001).

88. Krakowski, J., Aitken, S. & El-Kassaby, Y. Inbreeding and conservation genetics in whitebark pine. *Conservation Genetics* **4,** 581–593. ISSN: 15660621 (2003).

89. Vos, P. *et al.* AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23,** 4407–4414. ISSN: 0305-1048, 1362-4962 (1995).

90. Lee, K. J. *et al.* Genetic Diversity Assessed by Genotyping by Sequencing (GBS) in Watermelon Germplasm. *Genes* **10,** 822. ISSN: 2073-4425 (Oct. 18, 2019).

91. *Helianthemum canadense (Longbranch frostweed) — NPIN* `https://www.wildflower.org/gallery/result.php?id_image=3516` (2024).

92. *Cirsium pitcheri (Sand dune thistle) — NPIN* `https://www.wildflower.org/gallery/result.php?id_image=32202` (2024).

93. Harris, J. G. *A revision of the genus Braya (Cruciferae) in North America* Ph. D (Department of Botany, The University of Alberta, Edmonton, Alberta, 1985).

94. Delnevo, N. *et al.* Apomixis goes a long way: Genetic evidence of persistence and long-distance seed dispersal in an ancient landscape. *Journal of Biogeography* **51,** 694–709. ISSN: 0305-0270, 1365-2699 (Apr. 2024).

95. McMaster, I. *Senna glutinosa ssp. glutinosa* iNaturalist NZ. `https://inaturalist.nz/photos/32040530?size=large` (2024).

96. Dufresne, F., Stift, M., Vergilino, R. & Mable, B. K. Recent progress and challenges in population genetics of polyploid organisms: an overview of current state-of-the-art molecular and statistical tools. *Molecular Ecology* **23,** 40–69. ISSN: 0962-1083, 1365-294X (Jan. 2014).

97. Limborg, M. T., Seeb, L. W. & Seeb, J. E. Sorting duplicated loci disentangles complexities of polyploid genomes masked by genotyping by sequencing. *Molecular Ecology* **25,** 2117–2129. ISSN: 0962-1083, 1365-294X (May 2016).

98. Margarido, G. R. A., Souza, A. P. & Garcia, A. A. F. OneMap: software for genetic mapping in outcrossing species: OneMap. *Hereditas* **144,** 78–79. ISSN: 00180661 (June 4, 2007).

99. Ooijen, V. *JoinMap®5: Software for calculating genetic linkage maps in experimental populations of diploid species* 2018.

100. Pereira, G. S., Garcia, A. A. F. & Margarido, G. R. A. A fully automated pipeline for quantitative genotype calling from next generation sequencing data in autopolyploids. *BMC Bioinformatics* **19,** 398. ISSN: 1471-2105 (Dec. 2018).

101. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3 Genes—Genomes—Genetics* **1,** 171–182. ISSN: 2160-1836 (Aug. 1, 2011).

102. Qi, P. *et al.* UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in polyploids without a reference genome: finger millet as a case study. *BMC Plant Biology* **18,** 117. ISSN: 1471-2229 (Dec. 2018).

103. Merle, H., Garmendia, A. & Ferriol, M. Genotyping-by-Sequencing and Morphology Revealed the Role of Polyploidization and Hybridization in the Diversification of the Centaurea aspera L. Complex of Section Seridia (Juss.) DC. (Asteraceae). *Plants* **11,** 1919. ISSN: 2223-7747 (July 25, 2022).

104. O'Rawe, J. *et al.* Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine* **5,** 28. ISSN: 1756-994X (2013).

105. Barbitoff, Y. A., Abasov, R., Tvorogova, V. E., Glotov, A. S. & Predeus, A. V. Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery. *BMC Genomics* **23,** 155. ISSN: 1471-2164 (Dec. 2022).

106. Yao, Z. *et al.* Evaluation of variant calling tools for large plant genome re-sequencing. *BMC Bioinformatics* **21,** 360. ISSN: 1471-2105 (Dec. 2020).

107. Edwin, N. R., Fitzpatrick, A. H., Brennan, F., Abram, F. & O'Sullivan, O. An in-depth evaluation of metagenomic classifiers for soil microbiomes. *Environmental Microbiome* **19,** 19. ISSN: 2524-6372 (Mar. 28, 2024).

108. Hermanutz, L. A. *COSEWIC status appraisal summary on the Long's braya, Braya longii in Canada* OCLC: 781535829. ISBN: 978-1-100-18725-9 (COSEWIC Secretariat, Ottawa, 2011).

109. COSEWIC. *COSEWIC assessment and status report on the Fernald's braya, Braya fernaldii, in Canada.* xi + 36 pp. OCLC: 864394532 (Committee on the Status of Endangered Wildlife in Canada, Ottawa, 2012).

110. Hoelzel, A. R. Where to now with the evolutionarily significant unit? *Trends in Ecology & Evolution* **38,** 1134–1142. ISSN: 01695347 (Dec. 2023).

111. Werden, L. K. *et al.* Ex situ conservation of threatened plant species in island biodiversity hotspots: A case study from Hawai'i. *Biological Conservation* **243,** 108435. ISSN: 00063207 (Mar. 2020).