

RETHINKING REFUGIA: TREE TOPOLOGY, DIVERGENCE DATES, AND DEMOGRAPHIC HISTORY TRACE THE DISTRIBUTION OF THE ENDANGERED PLYMOUTH GENTIAN (*SABATIA KENNEDYANA*) FROM THE PLEISTOCENE GLACIATION TO PRESENT DAY¹

ADRIANA SUAREZ-GONZALEZ^{2,3}, JOLENE T. SUTTON^{4,6}, ANDREW J. TRANT^{5,6}, ELENA ZAMLYNNY^{6,7}, AND SARA V. GOOD^{2,3,8}

²Department of Biology, The University of Winnipeg, 599 Portage Avenue, Winnipeg, Manitoba, R3B 2G3, Canada;

³Department of Botany, University of British Columbia, 3529-6270 University Blvd., Vancouver, British Columbia, V6T 1Z4, Canada; ⁴Department of Biology, University of Hawai'i at Mānoa, 2538 McCarthy Mall, Honolulu, Hawaii 96822 USA; ⁵School of Environmental Studies, P. O. Box 3060, STN CSC, University of Victoria, Victoria, British Columbia, V8W 3R4, Canada;

⁶Department of Biology, Acadia University, 33 Westwood Avenue, Wolfville, Nova Scotia, B4P 2R6, Canada; and ⁷Department of Chemistry, Acadia University, 6 University Avenue, Wolfville, Nova Scotia, B4P 2R6, Canada

- **Premise of study:** Molecular population genetics is a powerful tool to infer how species responded to past environmental change. In the northern hemisphere, interest is increasing in how species responded to changes in ice coverage and temperature during the last glaciation maximum (LGM, between 18 000–21 000 yr ago) with a common assumption that glacial refugia were located at the southern edge of a species range.
- **Methods:** We reconstructed the glacial and postglacial phylogeography of *Sabatia kennedyana*, a member of the Atlantic Coastal Plains Flora with a current distribution from Nova Scotia (NS) to South Carolina, using both cpDNA and nuclear markers. We also examined clinal variation in morphological traits, in particular relative investment in asexual vs sexual growth.
- **Key results:** We find strong evidence that the species did not reside in southern glacial refugia, but rather in primary glacial refugia off the exposed continental shelf extending from Cape Cod and that this area was responsible for the founding of modern populations across the range from Nova Scotia (NS) to the United States. Additionally, based on the finding of higher cpDNA diversity and older cpDNA lineages in NS, we propose that multiple founder events occurred in NS, while only a single lineage gave rise to current populations in the United States.
- **Conclusions:** By understanding how *S. kennedyana* responded to past shifts in climate and by identifying areas of high genetic diversity in the northern range edge, we discuss the potential response of the species to future climate change scenarios.

Key words: Atlantic Coastal Plain; Bayesian analyses; clinal variation; cpDNA; extended Bayesian skyline plot; flora; Gentianaceae; ISSR; northern refugia; Pleistocene; sexual vs asexual investment.

During the Last Glacial Maximum (LGM) 18 000–21 000 yr ago, the landscapes and climates of the northern hemisphere were dramatically altered due to the advance of continental ice sheets that extended over much of the northern portion of the continent (Overpeck et al., 1992). Palaeontological, biogeographical,

and phylogenetic evidence suggests that during glacial maxima, temperate species survived or relocated to refugia at the limits of their range or beyond those limits, leading to isolation, massive migrations, and reductions in population sizes (Hewitt, 2000). As the ice sheets retreated at the end of the LGM, historically glaciated areas were recolonized from refugia, and populations expanded as the climate warmed (Hewitt, 2000). Genetic studies using organellar DNA markers (i.e., chloroplast and mitochondrial DNA) have sought to identify the location of refugia and the patterns of postglacial expansion of taxa by looking for genetic signatures of isolation and a reduction in geographical ranges during the LGM (Hewitt, 1996; Comes and Kadereit, 1998; Hewitt, 2004).

Genetic variation across a range of temperate plant and animal species together with palaeoecological records have been used to identify three major southern refugia in Europe during the LGM: the Mediterranean peninsulas of Iberia, Italy, and the Balkans (Frenzel, 1973; Hewitt, 1996; Tzedakis et al., 2002). In North America, refugia located in Florida and the Gulf Coast

¹Manuscript received 4 June 2014; revision accepted 6 March 2015.

The authors thank S. Boates and M. Elderkin for logistical and financial support and for arranging the permits to sample *S. kennedyana* in Nova Scotia; B. Sorrie, P. Somers, and C. Helms for obtaining the sampling permits in Massachusetts and North Carolina; S. vander Kloet and R. Newell for discussion; and S. Wood, J. and I. Avila, and N. Buckley for logistical help in the field. This study was funded by the World Wildlife Fund Canada, the NS Habitat Conservation Fund, and NSERC grants to S.V.G. and by an Acadia graduate student fellowship to J.S. It was also funded by a Canadian Foundation for Innovation grant that supported equipment and salary funding for E.Z.

⁸Author for correspondence (e-mail: sgood@uwinnipeg.ca)

have been identified (Swenson and Howard, 2005). Provan and Bennett (2008) suggested that in addition to extensive southern glacial refugia, species also persisted throughout the LGM in smaller, cryptic refugia at higher latitudes (Provan and Bennett, 2008). Some of these putative refugia may have existed in the extensive continental shelf region off the New England coast as far as Sable Island (Barrington and Paris, 2007), in an area currently beneath the Atlantic Ocean (Pielou, 1991) (Fig. 1). One North American complex of plant species that are presumed to have persisted at the southern end of their range during the LGM are the Atlantic Coastal Plain Flora (ACPF).

The Atlantic Coastal Plain Flora is a group of over 90 taxonomically unrelated species that occupy oligotrophic wetland habitats including seasonal freshwater ponds, lakeshores, and kettlehole depressions. Generally slow-growing and poor competitors, these plants are adapted to live in areas with high-disturbance regimes, which include damage from waves and large fluctuations in water levels that result in extensive periods of submersion (Keddy, 1985; Wisheu and Keddy, 1994). *Sabatia kennedyana* Fernald (Gentianaceae), a characteristic species of

the ACPF, is a semiaquatic, herbaceous perennial that occurs mostly along the shores of coastal, glaciated ponds throughout the coastal plain of the eastern United States, with nearly all populations in southeastern Massachusetts and a few others located in Rhode Island, North Carolina, and South Carolina (USDA NRCS, 2014). Disjunct populations at the northern limits of the range occur in southwest Nova Scotia (NS).

The primary objective of this phylogeographic study is to characterize the genetic diversity and structure of populations of *S. kennedyana* at various locations across the species' range to identify possible Pleistocene refugia. Specifically, we examined whether there is evidence for glacial refugia at the southern edge of the range and/or in possible cryptic refugia along the eastern seaboard at the northern end of the range. To this end, we used comparative analyses of cpDNA sequence data with data from intersimple sequence repeat (ISSRs): the former are maternally inherited, while the latter are biparentally inherited and dominant. Additionally, because of observed variation in plant morphology and an apparent increase in investment in sexual vs. asexual growth of ramets in the southern part of the

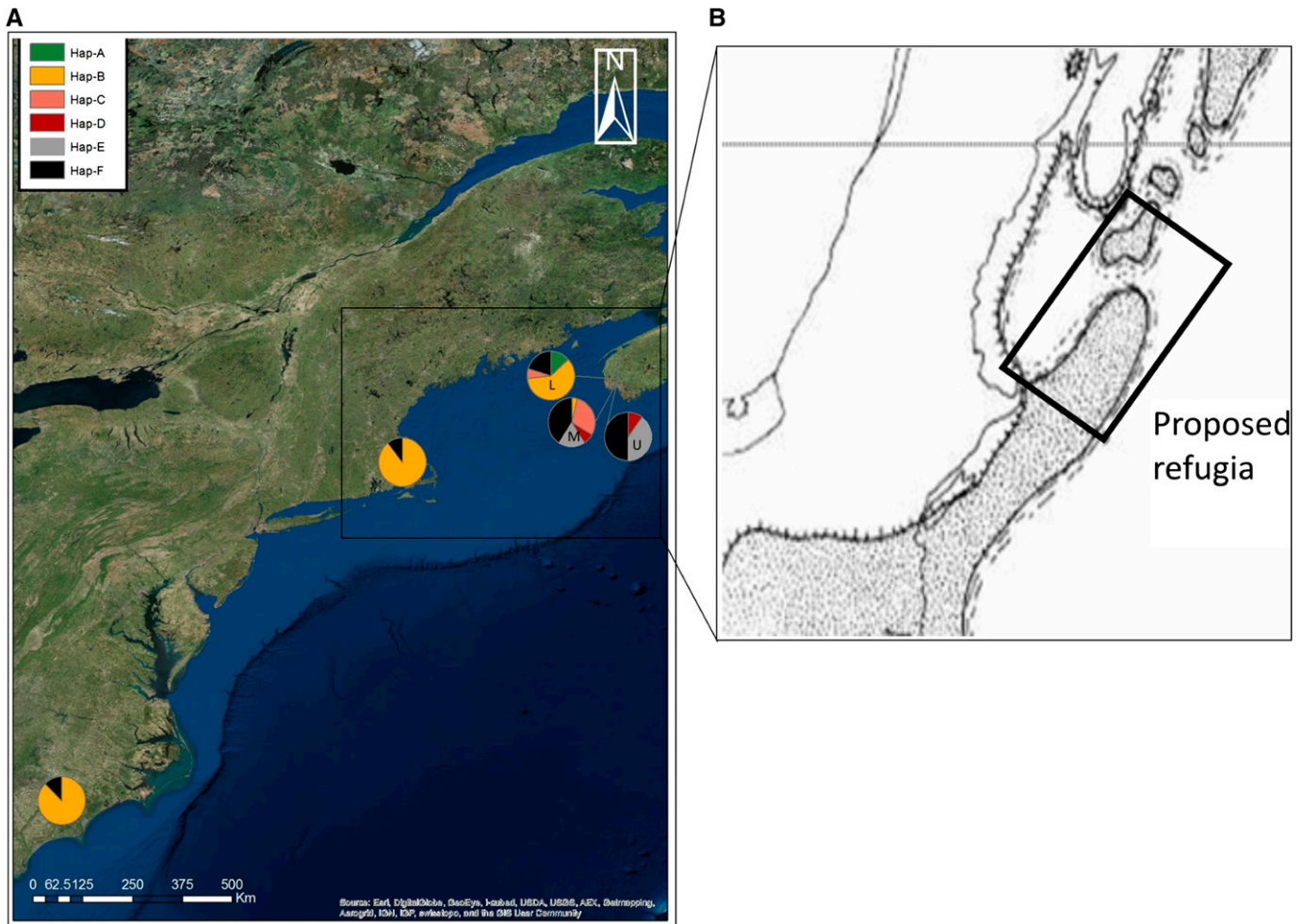


Fig. 1. Sampling sites of extant populations of *Sabatia kennedyana* and inferred historical map of the area during the last glacial maximum (LGM). (A) Map of the frequency distribution of two cpDNA (*trnS-trnG* and *rpl20-rps12* spacers) haplotypes in *Sabatia kennedyana* from five locations representing the northern (Nova Scotia [NS], Canada), central (Massachusetts [MA], USA) and southern (North Carolina [NC], USA) range distribution of the species. Pie charts show haplotypes in NC and MA (USA) and in three region from NS (lower Tusket: L; mid Tusket: M; upper Tusket: U). Sized wedges represent haplotype frequency in each population. Haplotype F in black represents private alleles. (B) Inferred historical map of the area depicted in the black square in (A), as estimated to exist 18 000 yr ago; dark gray areas indicate land masses currently submerged, but hypothesized to be above sea level during the LGM (modified from Pielou, 1991). The location of the proposed refugia is shown by the black rectangle.

range (Sutton, 2007), we examined differences in the patterns of growth and reproductive strategy among populations to compare clinal variation in morphological and genetic characters.

MATERIALS AND METHODS

Study organism—*Sabatia kennedyana* is a semiaquatic, herbaceous, lake-shore perennial herb. It is mainly distributed along the eastern seaboard of the United States, populations naturally occur in Massachusetts (MA), Rhode Island (RI), North Carolina (NC), and South Carolina (SC), and it was introduced into Virginia (USDA NRCS, 2014). At the northern range limits, *S. kennedyana* is restricted to NS where it is considered to be disjunct due to the large geographic separation between NS and populations in Cape Cod (MA) and the rest of the range. Nova Scotian populations are further restricted to 11 lakes in the Tusket River Watershed in the southwest of the province. The species is designated as endangered in Canada under the Species at Risk Act and in Nova Scotia under the Nova Scotia Endangered Species Act due to declines in the species' geographic range (COSEWIC, 2012). The species is also of international concern, with an International Global Rank (G-Rank) of G3, and is listed as critically imperiled in all states except MA, where it is vulnerable (S3) (NatureServe, 2015).

Throughout its range, *S. kennedyana* combines sexual and asexual modes of reproduction. Sexually reproductive flowers are disc-shaped, pale to bright pink, hermaphroditic, and possess a bilocular pistil. Protandry and temporal separation of floral dehiscence on a stem limits self-fertilization, but some self-pollination occurs and substantial seed set can be produced since the species is fully self-compatible (Hill et al., 2006). Sexual reproduction is accomplished through insect pollination, with approximately 87% of insect visitors in Nova Scotia made up of syrphid flies and approximately 5% made up of sweat bees (Trant et al., 2010). However, the diversity, abundance, and proportion of observed pollinator families appear to differ throughout the range (J. T. Sutton and S. V. Good, personal observation).

Basal rosettes are capable of asexual (clonal) growth via stoloniferous runners that produce additional "daughter" rosettes. These daughter rosettes continue to develop through vegetative growth, forming a spreading mat (henceforth referred to as a "clonal cluster"). Eventually, one or more rosettes in a clonal cluster may undergo sexual reproduction by developing an erect flower stalk. Due to the extensive clonal growth exhibited by *S. kennedyana*, it is often very difficult to identify a single genet. Based on the general pattern of clonal growth, for this study, it was determined that "individuals" (i.e., separate plants) would be considered as physically discrete clonal clusters.

Study area—Fourteen populations were sampled from three locations representing the northern (NS, Canada), central (MA, USA), and southern (NC, USA) range of the species (Fig. 1, Table 1). Populations of *S. kennedyana* in SC are on the border with NC; the populations sampled here from NC are on the same river (just north) of the populations of *S. kennedyana* in SC. In NS, samples were collected from seven lakeshores within the lower, mid, and upper Tusket River watershed in Yarmouth County. In NC, two sites were sampled in the Waccamaw River watershed in Columbus County. In MA, eight kettle ponds in and to the north of Cape Cod were sampled which have little or no connectivity via water among them.

Morphology—Eight morphological characters were examined to test for differences in vegetative and flowering investment across the range. Sexual variables included flower stem height, the number of flowers per stem, the number of flower stems per cluster, and the number of flowers per cluster. Additionally, fruits were collected to compare seed set per fruit and fruit mass across the range. Vegetative variables included cluster diameter and the number of rosettes per cluster. With the exception of seed set and fruit mass, all measurements were taken in the field during peak flowering season.

Statistical analysis for morphology—We used generalized linear models (GLMs) with Gaussian error distributions to test for the effect of location on log-transformed data for flower stem height, fruit mass, and cluster diameter. To test the effect of location on the remaining morphological characters (number of flowers per stem, number of stems per cluster, number of flowers per cluster, number of seeds per fruit, and number of rosettes per clusters), we performed GLMs with Poisson error distributions on untransformed data. All statistical analyses were performed with the program R v. 2.15.2 (R Development Core Team, 2011).

Genotyping—Genomic and plastid DNA was extracted from leaf samples (which were collected in the field and immediately frozen in a dryshipper) using GenElute Plant Genomic DNA Kits (Sigma, St. Louis, Missouri, USA), and IDPURE Genomic DNA Isolation Kits (ID Laboratories, London, Ontario, Canada). Three sets of ISSRs loci were amplified using primers 854, 841, and SNB3 (UBC ISSR primer sets) for 179, 42, and 44 samples respectively (Table 1) following a protocol given in Sutton (2007). One of these primers (854) was used in a previous study that examined clonal variation in a subset of the populations (Sutton, 2007), while the additional two markers were scored in fewer individuals but more populations for the purposes of this study. Final PCR products were run on GE Healthcare (Little Chalfont, Buckinghamshire, UK)

TABLE 1. Number of samples obtained for 1161 bp of cpDNA data (*trnS-trnG* + *rpl20-rps12*) and the ISSR data for primers SBN3, 841, and 54 for each of the eight ponds sampled in MA, the two creeks sampled in NC, the two lakes in the southern Tusket watershed, middle Tusket and upper Tusket in Nova Scotia with their corresponding latitude and longitude.

Site	Latitude	Longitude	cpDNA	ISSR SNB3	ISSR 841	ISSR 854
Massachusetts						
Denrock Pond	42.680010	-71.139461	3	3	3	—
Cooks Pond	41.922908	-70.665970	1	1	1	—
Harlow Pond	41.919911	-70.670195	4	3	3	—
Lamson Pond	41.679862	-70.267805	2	1	—	—
Little Cliff Pond	41.755269	-70.013416	2	—	—	—
Marydunn Pond	41.674806	-70.277292	—	—	—	39
North Triangle Pond	41.950579	-70.703530	3	2	3	—
Grassy Pond	42.501868	-71.451938	4	2	3	—
North Carolina						
Seven Creek1	34.277726	-78.539662	13	—	—	35
Seven Creek2	34.056404	-78.656087	3	—	—	—
Lower Tusket						
Wilsons Lake	43.939767	-65.892136	4	3	2	41
Bennetts Lake	43.928375	-65.902836	11	7	7	—
Mid-Tusket						
Lac de l'École Lake	43.935631	-65.813303	9	7	2	—
Kegeshook Lake	43.966303	-65.80065	12	2	7	—
Gillfillan Lake	43.969922	-65.812131	6	4	6	30
Upper Tusket						
Pearl Lake	44.048628	-65.815194	4	3	3	34
Third Lake	43.980653	-65.830364	6	4	4	—
Total			87	42	44	179

polyacrylamide CleanGels. Gels were stained using DNA Silver Staining Kits (GE Healthcare), and lanes scored by eye for the presence (1) or absence of bands (0). For consistency in scoring among gels, a standard of known molecular weight bands (Low ranger, Norgen, Thorold, Ontario, Canada) was run in the first, middle, and last lanes, and two replicate samples were run near each end of all gels. For the chloroplast markers, the *trnS-trnG* intergenic spacer and *rpL20-5'rpS12* intergenic spacer (Hamilton, 1999) were amplified as described by Wood (2006) from 87 samples (Table 1). DNA of purified PCR products was sequenced on the MJ basestation genotyper and reads were called using the licensed genotyping software Cartographer (MJ Research, Waltham, Massachusetts, USA). Sequences from forward and reverse primers were edited and the consensus sequence for each individual extracted using SeqManII (DNA-STAR, Madison, Wisconsin, USA). Both regions were concatenated and a multiple sequence alignment was obtained using Clustal W in the program MEGA5 (Tamura et al., 2011).

Genetic diversity, structure, and haplotype network—The levels of cpDNA sequence polymorphism were evaluated by calculating the number of haplotypes (*h*), haplotype diversity (*H_d*) and nucleotide diversity (π), excluding sites with gaps/missing data, using DNASP v5.0 (Librado and Rozas, 2009). Additionally, the mean pairwise distance between populations was estimated using Kimura's two-parameter model in MEGA5. For the ISSR data, Bayesian estimates of expected within-population heterozygosity (*h_s*) were calculated for the data from two of the three ISSR data sets (SNB3 and 841) using the software Hickory v.1.1 (Holsinger and Lewis, 2003) under the *f*-free model.

To infer population genetic structure among populations and regions and to infer possible refugia during the LGM, we used multiple statistical tests. First, genetic divergence between population pairs was measured using Kimura's two-parameter estimates (Kimura, 1980) (for the cpDNA sequences) and Nei's genetic distance estimate (Nei, 1972) (for the ISSR data) as implemented in MEGA5 and GENALEX (Peakall and Smouse, 2006), respectively. The resulting matrices of genetic distances were used to construct neighbor-joining (NJ) population trees using the program NEIGHBOR included in PHYLIP 3.66 (Felsenstein, 2005). Second, an analysis of molecular variance (AMOVA) was employed to identify genetically homogeneous populations. Significance of the observed *F_{CT}* values was tested using a random permutation procedure available in Arlequin v. 3.01 (Excoffier and Lischer, 2010). We used *F_{ST}* and pairwise differences as the distance method for cpDNA and ISSR data, respectively. A principal component analysis (PCA) was performed as a third approach for correlating genetic composition with geographic location of populations as implemented in the MultiVariate Statistical Package (MVSP) v. 3.21 (Kovach, 2007). For this analysis, we used the data from ISSR primer 854, which consisted of data from fewer populations but more samples per population than the other two primers (see Table 1), thus allowing a better visualization of the distribution of individuals after collapsing the multilocus genotype data.

Patterns of gene flow between populations were estimated using a Bayesian approach in the program Migrate 3.6 (Beerli, 2009). To explore different gene flow scenarios, we compared five models based on our genetic diversity and structure analysis. The models included one panmictic population, two populations (USA and Canada), and 3–4 subgroups of the NS populations (Table 5). We calculated the log marginal likelihoods as the Bezier approximation score and estimated the log Bayes factors (LBF) as $\ln(\text{marginal likelihood of } M_0) - \ln(\text{marginal likelihood of } M_n)$ to determine the level of support for each population model. In the search strategy menu, we used 10000 for the number of recorded steps in chain and 10000 as burn-in for each chain. To compare the models, we used the thermodynamic integration method by turning on static heating with four chains sampled at every 10th interval and using the default temperature scheme.

To examine the glacial refugia hypothesis with the cpDNA data, we constructed a minimum spanning network with the program Arlequin v. 3.01, visualized in the program HapStar (Teacher and Griffiths, 2011)

Examination of phylogenetic relationship and past demographic changes modeled in a Bayesian framework—To estimate the time of divergence of lineages in the US and Canada lineages, we analyzed the cpDNA sequence data in a Bayesian framework using the software BEAST v. 1.8 (Drummond et al., 2012). The divergence times of clades were estimated by partitioning the data into two cpDNA segments, using the best evolutionary model for the substitution process of each marker as identified using BIC, HKY+G, and HKY + I for *trnS-trnG* and *rpL20-5'rpS12* spacers, respectively, and setting the rate of substitution for *trnS-trnG* to 0.0021 and for *rpL20-5'rpS12* to 0.0019 based on the mean and relative estimates found in other members of the euasterids I (Shaw

et al., 2005; Ku et al., 2013). These estimates are slightly higher than an estimated mean rate of evolution across all noncoding elements in the cpDNA of 0.00152 per site per million years (Yamane et al., 2006). We then employed a "strict clock" model based on the recent timescale of divergence among the samples, but allowed the rate to vary by placing a uniform prior on the substitution rate from 0 to 1, with initial values set to 0.0019 and 0.0021. The effect of different rates of substitution, models was evaluated by using both lower (e.g., using the more rigorously supported value of 0.0013 sites/Myr for *rbcl*, a slower evolving protein-coding gene in the cpDNA), and higher (e.g., 0.0026, twice that of *rbcl*, which is at the high end for cpDNA noncoding sequences [Yamane et al., 2006]), estimates of the substitution rate and then estimating relative ages.

Implementation of the analyses in a Bayesian framework allows simultaneous reconstruction of the tree topology and estimated divergence times of lineages including the associated sampling error and also permits estimation of population demographic parameters conditioned on the estimated tree topology. We estimated the credible set of the number of population size changes and the change in effective population size over time, scaled by the mutation rate, by employing a coalescent prior for the tree with an "extended Bayesian skyline plot" and then plotted the estimated effective population size against time (in million years ago [Ma]) with 95% highest posterior density (HPD) values. The posterior distribution of the parameters was sampled by simulating the model with two MCMC chains which were monitored for convergence by examination of the $\ln L$ plots over time using the program Tracer v. 1.3 (Rambaut et al., 2014). After the first 20000 generations were discarded as burn-in, sampling was performed every 2000 generations for another 1000000 generations when the effective sampling size (ESS) of each parameter was above 100 or preferably 200 and the MCMC chain had reached its stationary phase (convergence). Population size changes are presented as means with associated standard deviation (SD) and standard error (SE).

RESULTS

Morphology—Three sexual investment variables increased in size or number going from north to south from NS to MA to NC (flower stem height: $F_{2,1601} = 3483$, $R^2 = 0.81$, $P < 0.001$; number of flowers per stem: $F_{2,1635} = 38480$, $R^2 = 0.98$, $P < 0.001$; number of flowers per cluster: $F_{2,287} = 2112$, $R^2 = 0.94$, $P < 0.001$; Appendices S1 and S2, see Supplemental Data with the online version of this article). Fruit mass and the number of seeds per fruit were significantly lower in NS compared with either MA or NC ($F_{2,158} = 183.90$, $R^2 = 0.70$, $P < 0.001$ and $F_{2,234} = 3.51 \times 10^7$, $R^2 = 1$, $P < 0.001$). The number of rosettes per cluster was greatest in NS, followed respectively by NC and MA ($F_{2,283} = 241.5$, $R^2 = 0.63$, $P < 0.001$); however, cluster diameter was significantly greater in NC compared with either MA or NS ($F_{2,158} = 263$, $R^2 = 0.40$, $P < 0.001$). The number of flower stems per cluster was significantly greater in MA compared with NC or NS ($F_{2,299} = 14.42$, $R^2 = 0.09$, $P < 0.001$), although location only explained 9% of the variation in these data.

Genetic diversity and structure—The spacers between *trnS* and *trnG* and *rpL20-5'rpS12* were aligned with a consensus length of 1161 bp, which was reduced to 812 bp after removing gaps and missing data (alignment file at www.datadryad.org). Among the 87 samples, 28 cpDNA haplotypes harboring 38 informative sites were identified in 812 bp. Haplotype sequences were deposited in the GenBank database under accession numbers JX003130–JX003186 for the *rpL20-5'rpS12* intergenic spacer and JX003187–JX003245 for the *trnS-trnG* intergenic spacer. The levels of haplotype (*H_d*) and nucleotide diversity (π) ranged from 0.978 (upper Tusket) to 0.298 (MA) and from 0.00505 (lower Tusket) to 0.00077 (MA), respectively (Table 2), revealing higher sequence polymorphism in NS compared with populations in MA and NC.

TABLE 2. Genetic diversity of cpDNA sequences from *trnS-trnG* and *rpl20-rps12* spacers sampled from populations of *Sabatia kennedyana* in the Tusket river watershed in Canada and Massachusetts (MA) and North Carolina (NC) in USA.

Location	Population	Samples (<i>N</i>)	Number of haplotypes (<i>h</i>)	Haplotype diversity (<i>H_d</i>)	Nucleotide diversity (π)
Upper Tusket	PE, TH	10	4	0.978	0.00257
Mid Tusket	GI, KE, LA	26	6	0.972	0.00473
Low Tusket	BE, WI	15	9	0.962	0.00505
MA	DEN, GRA, LAM, LIT, NT	19	4	0.298	0.00077
NC	SC1, SC2	15	6	0.571	0.00083

Notes: PE: Pearl Lake, TH: Third Lake, GI: Gillfillan Lake, KE: Kegeshook Lake, LA: Lac de l'École Lake, BE: Bennetts Lake, WI: Wilsons Lake, MA: Massachusetts, DEN: DenRock Pond, GRA: Grassy Pond, LAM: Lamson Pond, LIT: Little Cliff Pond, NT: North triangle Pond, NC: North Carolina, SC: Seven Creeks.

The ISSR primers SNB3, 841, and 854 generated 23, 30, and 54 fragments with a total of 86.96%, 93.33%, and 88.89% polymorphic loci, respectively. The three data sets displayed similar levels of genetic diversity and revealed higher ISSR diversity in MA compared to NS (Table 3).

The neighbor-joining trees based on pairwise genetic distance matrices using either cpDNA or ISSR data, showed that populations from the lower Tusket River watershed (BE-WI) grouped closer to populations in the United States than to other populations in NS (see Figs. S3A and S3B in online Appendix S3). The unrooted NJ tree based on the cpDNA sequences (Appendix S3A) shows a central group consisting of populations from MA and NC, together with populations from the lower Tusket River in NS (BE-WI). Populations from the middle Tusket watershed (GI-KE-LA) are separated from this central group, and populations from the upper Tusket (PE-TH) cluster separately.

ANOVA using cpDNA sequences and ISSR data revealed the highest level of among group differentiation when populations were divided into two groups: USA and Canada. The second highest F_{CT} value was obtained with three groups: (1) upper Tusket, (2) mid and lower Tusket, and (3) United States (Table 4).

The PCA based on ISSR data (Fig. 2) supported the AMOVA results and showed differentiation between Canadian and US populations based on the first two component axes, which explained 31.3% of the variance. Also, ordination of the populations according to the first axes showed populations from the upper Tusket (PE, TH) as a different group (Fig. 2), as shown by the second best model in AMOVA.

TABLE 3. Estimates of expected within-population heterozygosity (*h_s*) of populations of *Sabatia kennedyana* from the Tusket river watershed in Canada and Massachusetts (MA) and North Carolina (NC) in USA using three different ISSR data sets. See Table 1 for abbreviations of population names.

Location	Population	SNB3 <i>h_s</i> (SD)	841 <i>h_s</i> (SD)	54 loci <i>h_s</i> (SD)
Upper Tusket (Canada)	PE	0.191 (0.035)	0.181 (0.030)	0.110 (0.012)
	TH	0.214 (0.031)	0.148 (0.026)	
Mid Tusket (Canada)	GI	0.263 (0.029)	0.212 (0.021)	0.124 (0.008)
	KE	0.247 (0.039)	0.261 (0.023)	
Lower Tusket (Canada)	LA	0.262 (0.027)	0.264 (0.033)	
	BE	0.255 (0.032)	0.291 (0.021)	
MA (USA)	WI	0.299 (0.025)	0.179 (0.033)	0.093 (0.008)
	LAM-NT	0.258 (0.033)	0.285 (0.029)	0.186 (0.009)
NC (USA)	DEN-GRA	0.337 (0.025)	0.258 (0.032)	
	COO-HAR	0.277 (0.031)	0.297 (0.024)	
	SC1			0.127 (0.007)

The comparison of gene flow scenarios in Migrate supported differentiation between Canadian and US populations (Table 5). The LBFs based on the Bezier approximation score also indicated that the second most likely model is one panmictic population, and the third best model is two groups in NS—separating the upper Tusket—and one group in the United States. Results from all the population models showed higher migration rates from NS to the US populations compared with migration rates from the United States to Nova Scotia (Appendix S4). For example, the most likely model (two groups: Canada and USA) showed that the migration rate from Canada to the United States is about 1.5 times higher (1109.12) than the migration rate from the United States to Canada (724.8).

Phylogenetic tree, haplotype network, and demographic population history

The Bayesian reconstruction of the tree topology and estimated divergence date of lineages revealed that three sequences from the lower or middle Tusket River in NS have the oldest coalescent times, ranging from 10000 to 14600 yr old (Fig. 3). The remaining sequences have an estimated most recent common ancestor (MRCA) dating to ~7200 yr, and these sequences group into three main clades. One of these clades has a MRCA ~5800 yr ago and is composed of haplotypes exclusively from the lower and middle Tusket River NS. In the haplotype network (Fig. 4), this ancestral haplotype is represented by the letter C, and its derivative haplotypes are only found in NS, as also shown by the phylogenetic tree. The other clade is younger, with a MRCA of ~3500 yr ago, and contains haplotypes from the two US populations (MA and NC) and from the lower Tusket watershed. This is the most common haplotype and is represented by the letter B in the haplotype network, and it is also the haplotype with highest frequency in the US (88% in NC and 89% in MA, Fig. 1), where it is the founder of all private haplotypes in the region (Fig. 4). The third clade also has a MRCA dating back ~3500 yr, but it contains haplotypes only from the middle and upper Tusket river watershed. The haplotypes in this clade are derivatives from haplotypes B (e.g., haplotype E) and C (e.g., haplotype D) (Fig. 4). Collectively, the results indicate that the oldest and most diverse lineages are in NS, while the US populations contain a subset of the haplotypes found throughout the range.

Using the pooled data set for the US and NS populations (the second best model in the Migrate analyses), the posterior distribution of the credible set of population size changes, estimates that the mean number of population size changes is 1.6 (SD = 0.715, SE = 0.05) with a median of 1, and zero (constant population size) not being in the credible set. The extended Bayesian skyline plot (Fig. 5) on the pooled populations follows a density-dependent exponential growth pattern, suggesting that there

TABLE 4. Analysis of molecular variance (ANOVA) based on cpDNA and ISSR in *Sabatia kennedyana* from Nova Scotia (Canada) and Massachusetts and North Carolina (USA). See Table 2 for abbreviations of population names.

Data	k	Groups	% Variation among groups	qCT	P values
cpDNA	2	(PE-TH-GI-KE-LA-BE-WI), (DR-GR-LM-NT-COO-HAR-NC)	8.54	0.0854	<0.0001
	3	(PE-TH), (GI-KE-LA-BE-WI), (DR-GR-LM-NT-COO-HAR-NC)	7.11	0.071	<0.0001
ISSR (841)	2	(PE-TH-GI-KE-LA- BE-WI), (-DR-GR-LM-NT-COO-HAR)	24.7	0.247	0.02
ISSR (841)	3	(PE-TH), (GI-KE-LA-BE-WI), (DR-GR-LM-NT-COO-HAR)	23.5	0.235	0.004

was exponential growth until about 1500 yr ago, followed by a density-dependent plateau, with a current effective population size of $\sim N_e = 1200$. Parallel analyses conducted on populations only in the United States or Canada show similar results, except that the estimated mean number of population size changes is lower in NS (mean = 1.16, SD = 0.41, SE = 0.01) than in the United States (mean = 1.42, SD = 0.75, SE = 0.02). The skyline plots based on both the NS and US populations are consistent with a model of approximately constant population size followed by expansion, except that, in the NS, population size has been stable and larger over the last 9000 yr (Appendix S5) with evidence of a gradual expansion, while in the United States, the population history is shorter with smaller effective populations sizes but a higher rate of exponential growth (Appendix S5). These models are consistent with a demographic history of gradual postglacial expansion with a density dependent decline in growth in the last 1000 yr.

DISCUSSION

Both the phylogenetic and genetic structure analyses suggest that populations from the northern and central eastern seaboard in the United States (MA and NC) and the Tusket watershed in Nova Scotia, Canada were founded by individuals from one or two northerly located refugia, plausibly on the shore of exposed and ice-free continental shelf extending east from Cape Cod, an area currently beneath the Atlantic Ocean (Fig. 1) (Sweeney and Ogilvie, 1993). Thus, despite the substantial distance (1200

km) between current populations in MA and NC and the ice-free conditions in NC during the LGM, there is no evidence for a separate refugium at the southern end of the distribution. In contrast, we found that individuals from the most northerly populations in the Tusket River watershed, in particular the upper Tusket, harbor divergent haplotypes. We hypothesize that this could be caused by three nonmutually exclusive scenarios: (1) postcolonization isolation by distance in the upper Tusket, (2) greater retention of ancestral diversity in northern parts of the range, or (3) a second refugium located further to the north. Lower cpDNA polymorphism in MA and NC compared with NS do not support the hypothesis of southern refugia. If the refugia for *S. kennedyana* were at the southern end of the current range, we would expect some descendants of these ancestral lineages today, but we do not observe them. Rather, all of the cpDNA haplotypes in MA and NC are descendants of a single ancestral cpDNA haplotype that is shared with individuals from the lower Tusket in NS. Moreover, based on the calibrated Bayesian tree, the haplotypes from MA and NC that form a clade are younger than the MRCA of the haplotypes found in NS. This provides strong evidence for northern refugia and suggests that NS populations of *S. kennedyana* were colonized by multiple founder haplotypes after glaciation. This postglacial expansion may have occurred from a single refugium, likely on the exposed continental shelf at that time, and the subsequent structure in NS may be due to isolation by distance within NS. It is also possible that a second refugium existed in the system of islands close to modern Sable Island; these islands are thought to have had a rich wetland flora due to the entrapment of enormous amounts of water into during the LGM (Pielou, 1991; Keddy et al., 2007).

Morphology—The analyses of morphological characters in *S. kennedyana* indicate that divergent ecological/climatic factors, such as the warmer and longer growing season at the southern end of the range have enabled environmental adaptation permitting greater investment in sexual reproduction. This is in line with our genetic structure and phylogenetic analyses, which showed high levels of population differentiation and divergence between the US and NS populations. However, it is in contrast to our finding of low levels of genetic differentiation among populations of *S. kennedyana* within the United States (MA vs NC) because we find extensive variation across latitudes in morphological traits associated with investment in sexual reproduction to asexual reproduction.

Three morphological characters associated with sexual reproduction (flower stem height, number of flowers per stem, and number of flowers per clonal cluster) decreased in size or number from south to north. Similarly, fruit mass and seed set

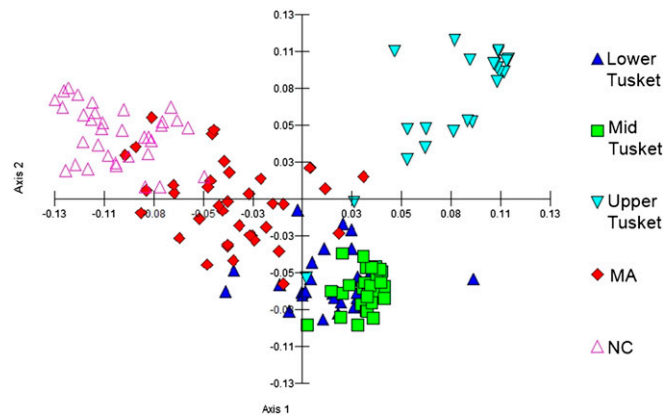


Fig. 2. Plot of the results of the principal component analyses for the ISSR data (primer 841) for populations of *Sabatia kennedyana* in North Carolina (NC), Massachusetts (MA), and the lower Tusket, mid Tusket, and upper Tusket in Nova Scotia.

TABLE 5. Comparison of five different gene flow scenarios based on cpDNA data in *Sabatia kennedyana* from Nova Scotia (Canada) and Massachusetts and North Carolina (USA). See Table 2 for abbreviations of population names.

Model	Bezier approximated score		Model rank
	Lm (mL)	LBF	
2 populations: (Canada), (USA)	–3119.66	0	1
1 population	–3150.94	31.28	2
3 populations: (PE-TH), (GI-KE-LA-BE-WI), (DR-GR-LM-NT-COO-HAR-NC)	–3156.96	37.3	3
3 populations: (PE-TH), (GI-KE-LA), (BE-WI-DR-GR-LM-NT-COO-HAR-NC)	–3288.91	169.25	4
4 populations: (PE-TH), (GI-KE-LA), (BE-WI-), (DR-GR-LM-NT-COO-HAR-NC)	–3310.49	190.83	5

Notes: Ln (mL) = log marginal likelihoods; LBF log Bayes factor = $\ln(\text{marginal likelihood of M0}) - \ln(\text{marginal likelihood of Mn})$. The model rank is in descending order (the preferred model is no. 1).

were reduced in NS compared with either MA or NC. Such reductions in sexual investment at the northern range limits are consistent with other phenotypic plant studies. For example, Pilon and Santamaria (2002) suggested that ecotypes from higher latitudes were generally smaller, but sometimes produced more seeds than their more southern counterparts. Li et al. (1998) reported that in *Arabidopsis thaliana* (Brassicaceae), the size of at least seven morphological characters was reduced from high compared with lower latitudes, although they attributed this clinal variation to genetic differentiation in response to local environmental conditions.

While plants in NS invested less in sexual reproduction compared with plants MA and NC, they also invested more in asexual reproduction, as evidenced by greater numbers of rosettes per cluster. Plants in NC had larger cluster diameters than plants in either MA or NS; however, this was due to larger individual rosettes (J. Sutton, personal observation), rather than increases in the number of rosettes. Thus, larger cluster size in NC is probably a consequence of longer, warmer growing seasons, rather than increased investment to asexual growth. The suppression of sexual components in favor of vegetative (clonal) ones is particularly common at the northern limits of a species' range. Northern populations of *Decodon verticillatus* (L.) Ell. (Lythraceae) produce roughly twice as many branches as more southern populations (Dorken and Eckert, 2001). Biotic and abiotic environmental factors such as temperature, moisture, and nutrient levels and light availability may also affect components of sexual and asexual reproduction (Eckert, 2002). For instance, Van Groenendaal et al. (1996) found an increase in clonality in habitats that were colder, more humid, with lower light availability and/or nutrient poor.

Interestingly, plants had more flower stems and fewer rosettes per cluster in MA compared with NS or NC, suggesting that plants in the middle of the species range may have experienced the largest shift from asexual to sexual allocation. This finding suggests that in addition to changes in growing season associated with latitude, other environmental factors may be driving this phenotypic response. In MA, *S. kennedyana* populations are situated in kettle ponds that frequently undergo prolonged, consecutive years of flooding (Elliston, 2006), while populations in the other two habitats occur on the shorelines of rivers that have more regular seasonal water regimes. Thus, the observation of high levels of sexual investment in MA may be associated with opportunistic sexual reproduction in low water years, resulting in a tradeoff between vegetative and sexual output.

Genetic diversity ISSR vs. cpDNA data—Our analysis of cpDNA genetic diversity revealed generally low levels of

polymorphism across the range, but also higher haplotype and nucleotide diversity in NS compared with populations in the USA. The analyses from the biparentally inherited markers (ISSR) on the other hand, showed similar levels of genetic diversity in populations from MA to those in NS. These apparently contrasting results may be explained by the different effective size and mutation rates of both types of markers; cpDNA has half the effective population size and a much slower mutation rate (1×10^{-9} to 3×10^{-9}) (Wolfe et al., 1987) than biparentally inherited markers such as ISSRs (1×10^{-6} to 1×10^{-2}) (Bhargava and Fuentes, 2010). Thus, the cpDNA genome is not only more sensitive to changes in population size, but levels of diversity should also reflect those in the more distant past since most polymorphism existing in the cpDNA would predate the LGM (Anderson et al., 2006; Maggs et al., 2008). Inferences from ISSR data, which have mutations rates on the order of 1 per 1000 to 100000 per generation per locus, would reflect more recent demographic and ecological events. The finding of relatively high ISSR variation and high levels of investment in sexual reproduction in MA suggests that, despite periodic flooding, which probably results in consecutive years of low seed output (Elliston, 2006), numbers of *S. kennedyana* in MA have probably been historically high. However, there are current concerns about the number of *S. kennedyana* in MA due to recent rapid loss of habitat.

Glacial refugia and postglacial migration—We have multiple lines of evidence supporting a rapid expansion of *S. kennedyana* following the LGM: our calibrated phylogenetic tree, genetic structure, and diversity analyses, as well as the extended Bayesian skyline plot, all suggest that populations have undergone a rapid historical expansion, likely from coastal refugia, into their current habitats on the Eastern Seaboard. Using the cpDNA sequence data, we simultaneously estimated the tree topology, divergence dates, and demographic history of populations across the range by employing a Bayesian analyses (Drummond et al., 2012). In the calibrated tree topology, the three oldest lineages emanate from individuals in the mid or lower Tusket river watershed, while there are three other clades, all 5800 yr old or younger, that show marked phylogeographic clustering. One of these clades harbors haplotypes from the United States and the lower Tusket watershed, while the other two contain haplotypes exclusively found in NS either from the lower/mid or mid/upper Tusket river watershed. Visualization of the timing of the coalescent events on this calibrated phylogeny suggests that populations of *S. kennedyana* expanded exponentially after the last glaciation and then exhibited a density-dependent decline in growth over the last 1000 yr or more. Although the timing of these events is dependent on the

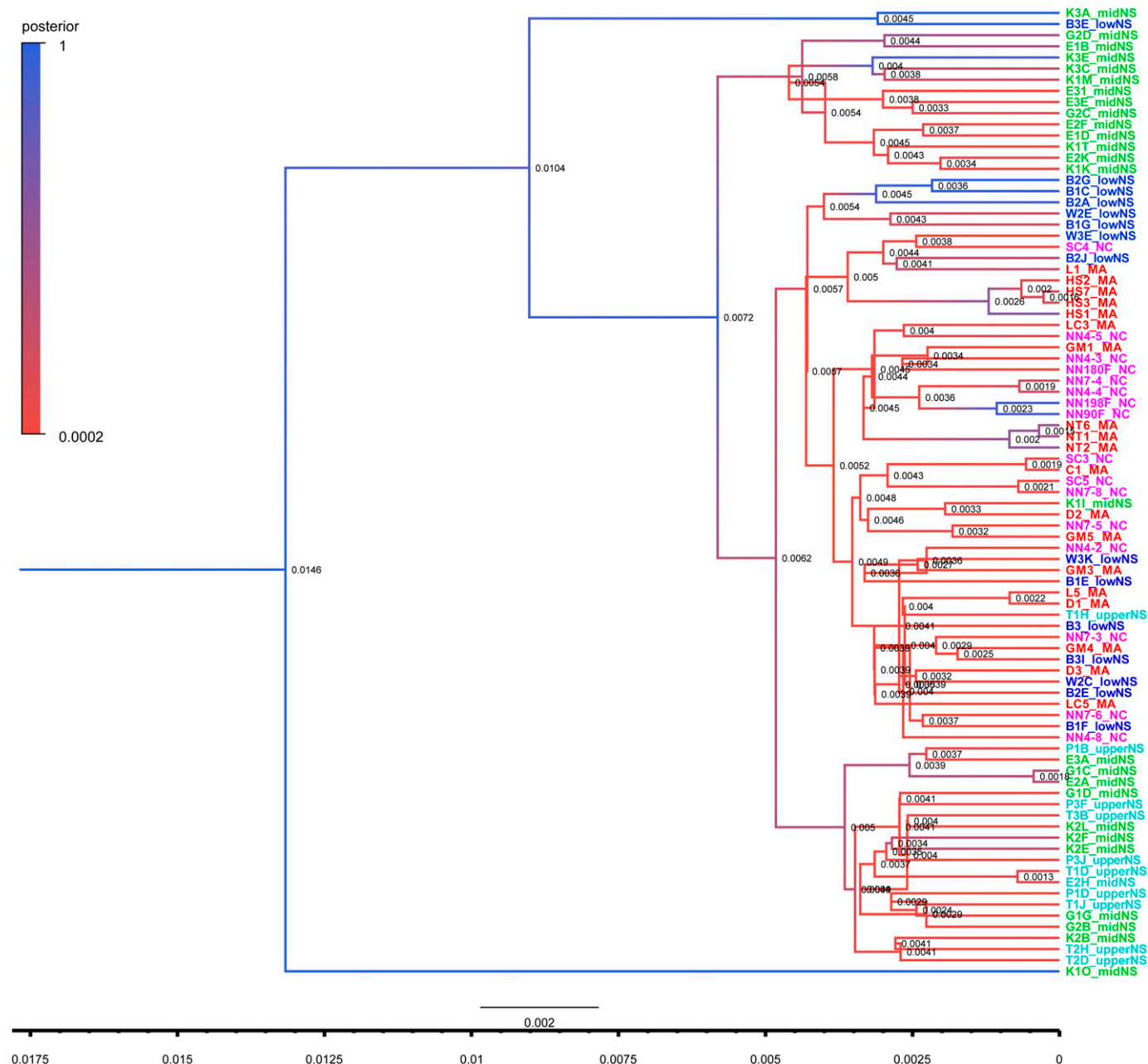


Fig. 3. Phylogenetic tree reconstructed within a Bayesian framework on 87 *rpl20-rpS12* and *trnS-trnG* intergenic spacers from the cpDNA in *Sabatia kennedyana* samples from the upper (cyan), middle (green), and lower (blue) parts of the Tuskett river watershed in NS and from kettleponds in MA (red) and two rivers in NC (purple). Branches are colored by the posterior support for the branch; node labels represent the estimated median age of nodes based on a strict clock framework using the estimate rate of substitution of these intergenic spacers (see text for details).

estimated rate of substitution employed in the Bayesian analyses, experimentation with different substitution rates generated similar results and the three main clades had estimated MRCAs between 3000 and 6000 years ago (not shown). Furthermore, in all models, a demographic model of constant population size was rejected in favor of one or more population size changes that indicated an exponential increase in population size after glaciation and a subsequent (but less dramatic) slow down toward the present.

The Bayesian analyses of the demographic history of populations were performed on the pooled data set, as well as by dividing the data set into populations from the USA and NS.

These models were justified based on the results of the ANOVA and Bayesian migration analyses, which indicate that the best model of population structure separates populations from the United States (MA and NC) from those in NS. However, the second-best migration model groups all populations together. For the Bayesian demographic analyses, pooling all of the data indicates that there has been predominantly exponential growth with a plateau in the last 1000 yr. When the demographic history of NS and MA/NC are reconstructed separately, both areas exhibit predominantly exponential growth post-Pleistocene. However, the estimated population sizes have been higher in NS, probably because it harbors greater haplotype diversity and

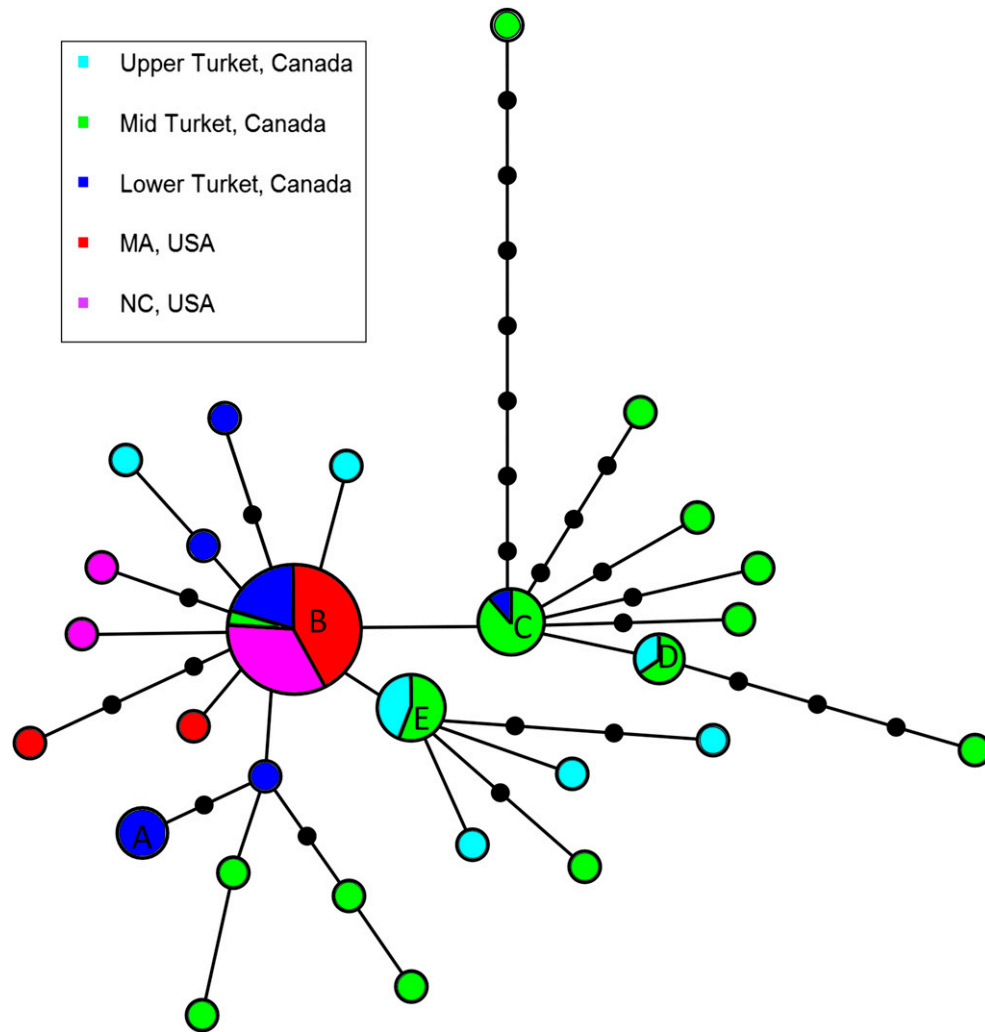


Fig. 4. Haplotype network of the 28 cpDNA haplotypes detected in *Sabatia kennedyana*. Circle size is proportional to haplotype frequency; the color represents groups based on location and genetic structure analyses; upper Tusket (cyan), mid Tusket (green), lower Tusket (blue), Massachusetts (red), and North Carolina (purple). The letters inside the most common haplotypes depict the symbols used in the map from Fig. 1. All other haplotypes are private and correspond to haplotype F in Fig. 1. Haplotypes found in more than one location are colored as pie charts with proportionally sized wedges representing the haplotype frequency in each population. Small black circles represent undetected intermediate haplotypes; the length of the lines connecting the nodes represents the relative genetic distances.

may have experienced less decline in population size in recent history.

The finding of greater genetic diversity in NS is also supported by the haplotype network in which populations from the United States (MA and NC) and the lower Tusket watershed (NS) share many common haplotypes that appear centrally and at high frequency in the haplotype network. Together, these results display a classic pattern of increasing genetic diversity nearing the glacial refugia (Dixon, 2011). This result suggests that all three geographically distant sites were founded by individuals from the same refugium, likely off of the coast of current day Cape Cod. A refugium at this location would have straddled the populations in MA and NS, but would have been more distant from the current populations in NC.

We also find evidence for a possible second refugium farther north. This hypothesis is supported by findings that despite the geographic proximity of populations in NS, located less than 25 km from each other, there is considerably genetic differentiation and higher haplotype diversity in NS than either population in the United States. Although the highest among group variation

was found between the US and NS populations, alternative population models separate individuals from the upper Tusket from those in other NS populations. The NJ trees based on average population pairwise differences also separate the upper Tusket in NS from the mid and lower Tusket populations. These results suggest that there was more than one founder event in NS postglaciation; the divergent NS haplotypes may have emanated from diverse haplotypes in a single refugium, or possibly by the founding of haplotypes from more than one refugium. Alternative explanations for these findings include postcolonization isolation by distance in the upper Tusket and/or greater retention of ancestral diversity in northern parts of the range.

Collectively, our results suggest that unglaciated coastal areas at the edge of the ice sheets may have been important for the survival of this Atlantic Coastal Plain species during the Holocene glaciation. The presence of northern cryptical refugia probably allowed pockets of diversity for other plant and animal species to persist (Stewart and Lister, 2001; Petit et al., 2003). The existence of refugia on the exposed and ice-free northeastern continental shelf in North America was suggested

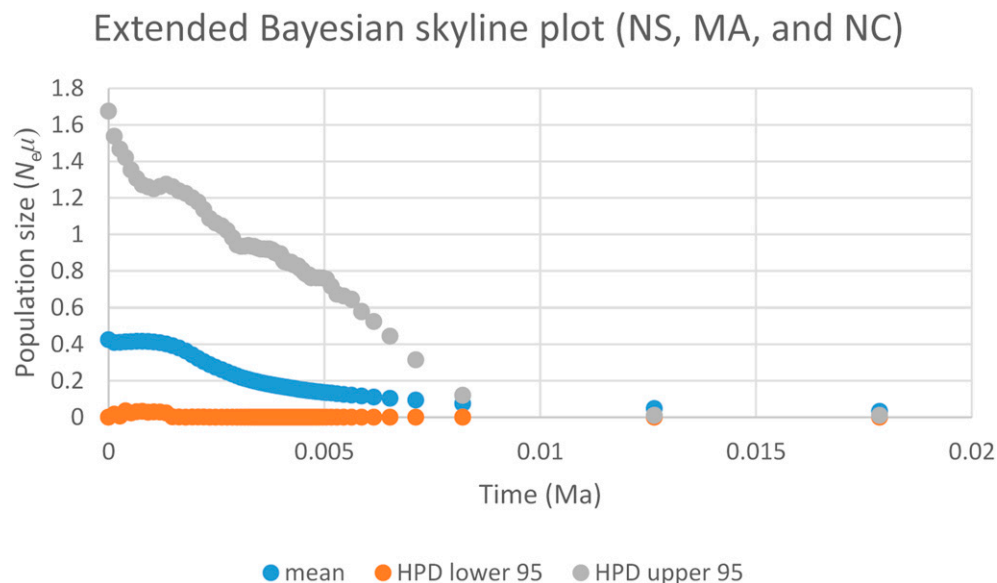


Fig. 5. Extended Bayesian skyline plot on the pooled cpDNA sequence data from the US and NS populations as generated from a Bayesian analyses that estimated historical demographic patterns conditional on the tree topology and time of divergence of lineages. The x-axis represents time in units of million of years ago; y-axis is an estimate of population size scaled by mutation rate. The mean and 95% mean posterior density (HPD) are indicated.

for other plant species in a review by Barrington and Paris (2007). For example, the lack of genetic structure between populations of Peck's avens (*Geum peckii*) located on the alpine summits of the White Mountains in New Hampshire and at sea level in Nova Scotia suggests that coastal bogs in unglaciated regions of the north Atlantic coast may have provided Pleistocene refugia for this alpine species (Paterson and Snyder, 1999). Isozyme data from beachpea (*Lathyrus japonicas*) showed that genetic diversity is higher in populations located in Cape Cod and Newfoundland and that diversity decreases away from these two centers (Schmitz, 2002). Walker (1998) showed that in beachgrass (*Ammophila breviligulata*) populations from the northern Atlantic (Long Island to Nova Scotia) have a greater nucleotide diversity than those in southern Atlantic (North Carolina to New Jersey) and inland (Lake Champlain and the Great Lakes). Collectively, these molecular data suggest the existence of glacial refugia in the North Atlantic in the ice-free exposed area off the coast of Cape Cod for these coastal species. Our study is the first to show that these areas were refugia for members of the Atlantic Coastal Plain, a taxonomic group with a more southern range.

The use of molecular data in relation to the distribution of Pleistocene fossil records is a powerful tool to infer the location of putative northern refugia (e.g., Tremblay and Schoen, 1999). For example, paleoecological records (Pielou, 1991) and cpDNA data (Boys et al., 2005; Walter and Epperson, 2005) suggest the occurrence of northeastern coastal refugia, in addition to fossils that provide evidence for southern Appalachian Mountain refugia, for red pine (*Pinus resinosa*). Additional geological studies (reviewed by Walter and Epperson, 2005) support this hypothesis and suggest that terrestrial ecosystems may have persisted in unglaciated areas off of the present Atlantic coastline due to the drop of sea levels (as much as 100 m) during the LGM (Keddy et al., 2007). Unfortunately, fossil records from the north Atlantic Coastal Plains are limited, since this area is now covered by the Atlantic Ocean. Instead, studies on haplotype distribution and genetic structure on additional

Atlantic Coastal Plain species may further improve our understanding of the survival and postglacial expansion of populations from this region.

Climate change—Understanding the rates by which species responded to past shifts in climate may provide insight into the resilience and vulnerabilities associated with recent climate change. For *S. kennedyana*, range expansion is accomplished via seed dispersal (e.g., floating on water) or by the movement of individuals on floating mats of vegetation dislodged by ice scouring events (A. Trant, personal observation). Estimates of postglacial migration rates are difficult to obtain since little is known about species' distribution at the LGM. For example, trees were originally estimated to have postglacial migration rates of 1500–2000 m/yr, based on pollen records; however, refugia have been identified using cpDNA haplotype distributions, which greatly reduce these estimated rates to less than 100 m/yr (McLachlan et al., 2005; Anderson et al., 2006). Though the exact locations of refugia identified in this study and the length of time *S. kennedyana* has occupied its current locations are unknown, we estimate a minimum migration rate of ~40 m/yr (assuming a refugium was near current day Sable Island and that *S. kennedyana* arrived in NS recently <10 000 yr ago). This migration distance may have been achieved by a stepping stone mosaic of habitat patches, which can be crucial for long-distance dispersal events and maintaining genetic diversity (Saura et al., 2014).

New challenges exist for species trying to track suitable climate given the unprecedented rate of climate change (Cunze et al., 2013) and the high level of landscape fragmentation (Renton et al., 2013). In the case of *S. kennedyana*, the majority of watersheds no longer have natural hydrological regimes, which is important for creating critical habitat for the Atlantic Coastal Plain Flora (ACPF) taxa (Wisheu and Keddy, 1989; Morris et al., 2002). Climate change presents a renewed need for refugia, a place where habitat and climate remain suitable for species to persist (Ashcroft, 2010). Since the distribution of *S. kennedyana*

is restricted by specific habitat and watershed requirements (Wisheu et al., 1994) and not likely by a temperature isocline, the current locations may be appropriate refugia under near future climate change scenarios. A greater conservation importance should be considered for populations at the northern edge of the distribution (NS) since these populations may be able to tolerate the largest increase in temperature and also contain the highest known levels of genetic diversity across the species distribution (Jump and Penuelas, 2005), which may have significant advantages for adapting future change.

LITERATURE CITED

- ANDERSON, L. L., F. S. HU, D. M. NELSON, R. J. PETIT, AND K. N. PAIGE. 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences, USA* 103: 12447–12450.
- ASHCROFT, M. B. 2010. Identifying refugia from climate change. *Journal of Biogeography* 37: 1407–1413.
- BHARGAVA, A., AND F. FUENTES. 2010. Mutational dynamics of microsatellites. *Molecular Biotechnology* 44: 250–266.
- BARRINGTON, D. S., AND C. A. PARIS. 2007. Refugia and migration in the Quaternary history of the New England flora. *Rhodora* 109: 369–386.
- BEERLI, P. 2009. How to use migrate or why are Markov chain Monte Carlo programs difficult to use? In G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi [eds.], *Population genetics for animal conservation*, Conservation Biology vol. 17, 42. Cambridge University Press, Cambridge, UK.
- BOYS, J., M. CHERRY, AND S. DAYANANDAN. 2005. Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *American Journal of Botany* 92: 833–841.
- COMES, H. P., AND J. W. KADEREIT. 1998. The effect of Quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science* 3: 432–438.
- COSEWIC. 2012. COSEWIC assessment and status report on the Plymouth Gentian *Sabatia kennedyana* in Canada. Committee on the Status of Endangered Wildlife in Canada [COSEWIC], Ottawa, Ontario, Canada. Available at <http://www.registrelep-sararegistry.gc.ca/default.asp?lang=En&n=24F7211B-1> [accessed 19 March 2015].
- CUNZE, S., F. HEYDEL, AND O. TACKENBERG. 2013. Are plant species able to keep pace with the rapidly changing climate? *PLoS ONE* 8: e67909.
- DIXON, M. D. 2011. Post-Pleistocene range expansion of the recently imperiled eastern little brown bat (*Myotis lucifugus lucifugus*) from a single southern refugium. *Ecology and Evolution* 1: 191–200.
- DORKEN, M. E., AND C. G. ECKERT. 2001. Severely reduced sexual reproduction in northern populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology* 89: 339–350.
- DRUMMOND, A. J., M. A. SUCHARD, D. XIE, AND A. RAMBAUT. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969–1973.
- ECKERT, C. G. 2002. The loss of sex in clonal plants. *Evolutionary Ecology* 15: 501–520.
- ELLISTON, L. C. O. 2006. Natural history, genetics and population biology of *Sabatia kennedyana* (Plymouth gentian): An endangered plant of Atlantic coastal plain pondshores. MSc dissertation, University of Massachusetts, Boston, Massachusetts, USA.
- EXCOFFIER, L., AND H. E. LISCHER. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.
- FELSENSTEIN, J. 2005. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genome Sciences, University of Washington, Seattle, Washington, USA.
- FRENZEL, B. 1973. Some remarks on the Pleistocene vegetation. *Eiszeitalter und Gegenwart* 23: 281–292.
- HAMILTON, M. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521–523.
- HEWITT, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913.
- HEWITT, G. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 359: 183–195.
- HEWITT, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247–276.
- HILL, N. M., M. T. MYRA, AND M. O. JOHNSTON. 2006. Breeding system and early stage inbreeding depression in a Nova Scotian population of the global rarity, *Sabatia kennedyana* (Gentianaceae). *Rhodora* 108: 307–328.
- HOLSINGER, K. E., AND P. O. LEWIS. 2003. Hickory: A package for analysis of population genetic data V1.0 [computer program]. University of Connecticut, Storrs, Connecticut, USA.
- JUMP, A. S., AND J. PENUELAS. 2005. Running to stand still: Adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010–1020.
- KEDDY, P. A. 1985. Wave disturbance on lakeshores and the within-lake distribution of Ontario's Atlantic coastal plain flora. *Canadian Journal of Botany* 63: 656–660.
- KEDDY, P. A., D. CAMPBELL, T. McFALLS, G. SHAFFER, R. MOREAU, C. DRANGUET, AND R. HELENIAC. 2007. The wetlands of lakes Pontchartrain and Maurepas: Past, present and future. *Environmental Reviews* 15: 43–77.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120.
- KOVACH, W. 2007. Multivariate statistical package, version 3.1 for Windows. Kovach Computing Services, Pentraeth, Wales, UK.
- KU, C., J.-M. HU, AND C.-H. KUO. 2013. Complete plastid genome sequence of the basal asterid *Ardisia polysticta* Miq. and comparative analyses of asterid plastid genomes. *PLoS ONE* 8: e62548.
- LI, B., J.-I. SUZUKI, AND T. HARA. 1998. Latitudinal variation in plant size and relative growth rate in *Arabidopsis thaliana*. *Oecologia* 115: 293–301.
- LIBRADO, P., AND J. ROZAS. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
- MAGGS, C. A., R. CASTILHO, D. FOLTZ, C. HENZLER, M. T. JOLLY, J. KELLY, J. OLSEN, ET AL. 2008. Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. *Ecology* 89: S108–S122.
- McLACHLAN, J. S., J. S. CLARK, AND P. S. MANOS. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology* 86: 2088–2098.
- MORRIS, P. A., N. M. HILL, E. G. REEKIE, AND H. L. HEWLIN. 2002. Lakeshore diversity and rarity relationships along interacting disturbance gradients: Catchment area, wave action and depth. *Biological Conservation* 106: 79–90.
- NEI, M. 1972. Genetic distance between populations. *American Naturalist* 106: 283–292.
- OVERPECK, J. T., R. S. WEBB, AND T. WEBB. 1992. Mapping eastern North American vegetation change of the past 18 ka: No-analogs and the future. *Geology* 20: 1071–1074.
- PATERSON, I. G., AND M. SNYDER. 1999. Genetic evidence supporting the taxonomy of *Geum peckii* (Rosaceae) and *G. radiatum* as separate species. *Rhodora* 101: 325–340.
- PEAKALL, R., AND P. E. SMOUSE. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.
- PETIT, R. J., I. AGUINAGALDE, J.-L. DE BEAULIEU, C. BITTKAU, S. BREWER, R. CHEDDADI, R. ENNOS, ET AL. 2003. Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science* 300: 1563–1565.
- PIELOU, E. 1991. After the ice. University Chicago Press, Chicago, Illinois, USA.
- PILON, J., AND L. SANTAMARÍA. 2002. Clonal variation in the thermal response of the submerged aquatic macrophyte *Potamogeton pectinatus*. *Journal of Ecology* 90: 141–152.
- PROVAN, J., AND K. BENNETT. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution* 23: 564–571.

- R DEVELOPMENT CORE TEAM. 2011. R: A language and environment for statistical computing. Website <http://www.R-project.org/>. R Foundation for Statistical Computing, Vienna, Austria.
- RENTON, M., S. CHILDS, R. STANDISH, AND N. SHACKELFORD. 2013. Plant migration and persistence under climate change in fragmented landscapes: Does it depend on the key point of vulnerability within the lifecycle? *Ecological Modelling* 249: 50–58.
- RAMBAUT, A., M. A. SUCHARD, D. XIE, AND A. J. DRUMMOND. 2014. Tracer v1.6 [computer program]. Available from <http://beast.bio.ed.ac.uk/Tracer>.
- SAURA, S., O. BODIN, AND M.-J. FORTIN. 2014. Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology* 51: 171–182.
- SCHMITZ, S. 2002. The evolutionary and biogeographic history of *Lathyrus japonicus* Willd. North America. Ph.D. dissertation, Univ. Vermont, Burlington, Vermont, USA.
- SHAW, J., E. B. LICKY, J. T. BECK, S. B. FARMER, W. LIU, J. MILLER, K. C. SIRIPUN, ET AL. 2005. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.
- STEWART, J. R., AND A. M. LISTER. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution* 16: 608–613.
- SUTTON, J. T. 2007. Morphology, reproductive strategy, and genetic diversity of the rare atlantic coastal plain flower, *Sabatia kennedyana* (fern): Variation across the species range and in anthropogenically disturbed habitat, MSc dissertation, Acadia University, Wolfville, Nova Scotia, Canada.
- SWEENEY, S., AND R. OGILVIE. 1993. The conservation of coastal plain flora in Nova Scotia. *Maine Naturalist* 1: 131–144.
- SWENSON, N. G., AND D. J. HOWARD. 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *American Naturalist* 166: 581–591.
- TAMURA, K., D. PETERSON, N. PETERSON, G. STECHER, M. NEI, AND S. KUMAR. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731–2739.
- TEACHER, A., AND D. GRIFFITHS. 2011. HapStar: Automated haplotype network layout and visualization. *Molecular Ecology Resources* 11: 151–153.
- TRANT, A. J., T. B. HERMAN, AND S. V. GOOD-AVILA. 2010. Effects of anthropogenic disturbance on the reproductive ecology and pollination service of Plymouth gentian (*Sabatia kennedyana* Fern.), a lakeshore plant species at risk. *Plant Ecology* 210: 241–252.
- TREMBLAY, N., AND D. SCHOEN. 1999. Molecular phylogeography of *Dryas integrifolia*: Glacial refugia and postglacial recolonization. *Molecular Ecology* 8: 1187–1198.
- TZEDAKIS, P., I. LAWSON, M. FROGLEY, G. HEWITT, AND R. PREECE. 2002. Buffered tree population changes in a Quaternary refugium: Evolutionary implications. *Science* 297: 2044–2047.
- USDA NRCS. 2014. The PLANTS Database. U.S. Department of Agriculture, National Resources Conservation Services, Washington, D.C., USA. Website <http://plants.usda.gov> [accessed 24 March 2014].
- VAN GROENENDAEL, J., L. KLIMES, J. KLIMESOVA, AND R. HENDRIKS. 1996. Comparative ecology of clonal plants. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 351: 1331–1339.
- WALKER, P. J. 1998. Biogeography, speciation and concerted evolution in *Ammophila* (Poaceae): Intraspecific variation in the internal transcribed spacer of nuclear ribosomal DNA. Ph.D. dissertation, University of Vermont, Burlington, Vermont, USA.
- WALTER, R., AND B. K. EPPERSON. 2005. Geographic pattern of genetic diversity in *Pinus resinosa*: Contact zone between descendants of glacial refugia. *American Journal of Botany* 92: 92–100.
- WISHEU, I. C., AND P. A. KEDDY. 1989. The conservation and management of a threatened coastal plain plant community in eastern North America (Nova Scotia, Canada). *Biological Conservation* 48: 229–238.
- WISHEU, I. C., AND P. A. KEDDY. 1994. The low competitive ability of Canada's Atlantic coastal plain shoreline flora: Implications for conservation. *Biological Conservation* 68: 247–252.
- WISHEU, I. C., C. J. KEDDY, P. A. KEDDY, AND N. M. HILL. 1994. Disjunct Atlantic coastal plain species in Nova Scotia: Distribution, habitat and conservation priorities. *Biological Conservation* 68: 217–224.
- WOOD, S. C. 2006. Genetic structure of endangered populations of *Coreopsis rosea* in Nova Scotia. Honours thesis, Acadia University, Wolfville, Nova Scotia, Canada.
- WOLFE, K. H., W. H. LI, AND P. M. SHARP. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNA's. *Proceedings of the National Academy of Sciences* 84: 9054–9058.
- YAMANE, K., K. YANO, AND T. KAWAHARA. 2006. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane. *Maize and Rice*. 13: 197–204.