ORIGINAL ARTICLE

Highest genetic diversity at the northern range limit of the rare orchid *Isotria medeoloides*

JL Stone¹, PA Crystal¹, EE Devlin¹, RH LeB Downer² and DS Cameron³

Populations in previously glaciated regions are often genetically depauperate in comparison with populations at lower latitudes, due either to bottlenecks experienced in post-glacial colonization or to contemporary genetic drift in small, peripheral populations. Populations of the rare self-fertilizing North American orchid *Isotria medeoloides* are largest in the previously glaciated region near the northern range limit, allowing us to examine the role of historical versus contemporary processes in determining population genetic diversity and structure. If contemporary processes predominate, genetic diversity should increase with increasing census size. In contrast, if sequential bottlenecks associated with colonization are paramount, diversity should decrease with latitude and be relatively insensitive to census size. We genotyped 299 individuals from 20 populations at four variable microsatellite loci to contrast genetic diversity and structure for populations in previously glaciated regions versus previously unglaciated regions. Populations were highly inbred ($F=0.95$) and highly differentiated ($R_{ST}=0.485$). Across all sampled populations, genetic diversity decreased and genetic differentiation increased with declining population size. Small southern populations were especially differentiated and genetically depauperate. In the glaciated part of the range, genetic diversity increased as populations approached the northern range limit, demonstrating the centrality of contemporary processes for this post-glacial colonist.

Heredity (2012) **109**, 215–221; doi:10.1038/hdy.2012.31; published online 13 June 2012

Keywords: genetic drift; inbreeding; leading edge; migration; orchid; population structure

INTRODUCTION

Species whose ranges span glaciated and unglaciated regions present a valuable opportunity to examine the role of historical processes in shaping the distribution of population genetic diversity. Populations that occupy previously glaciated regions stem from colonization events within the 18 000 years since the last glacial maximum (Pielou, 1991). In contrast, populations at lower latitudes are relatively ancient, although they may have expanded, contracted, or shifted their range during Pleistocene climatic fluctuations. The distribution of genetic diversity in such species will depend upon numerous factors, including whether range expansion was accompanied by demographic expansion, patterns of gene flow during the expansion, the occurrence of physical features constraining colonization routes and the continued suitability of refugial habitats (Excoffier *et al.*, 2009; Knowles and Alvarado-Serrano, 2010). When post-glacial colonization follows a stepping-stone pattern, genetic diversity is expected to steadily decline along the axis of colonization, so that recently colonized populations will show reduced genetic diversity and structure (Austerlitz *et al.*, 1997; Hewitt, 2000; Excoffier *et al.*, 2009). Numerous empirical studies have described declines in genetic diversity accompanying post-glacial colonization from low to high altitudes or latitudes (for example, Hewitt, 2004; Alsos *et al.*, 2007). Interpreting these data can be challenging, however, when expected patterns of genetic diversity and structure owing to contemporary processes coincide with those expected from historical processes (Eckert *et al.*, 2008). In particular, a problem in interpretation arises

because population size tends to decline toward range margins (Sexton *et al.*, 2009). When population size declines toward the periphery, decreased genetic diversity near range limits could be either due to historical population bottlenecks or contemporary genetic drift, or a combination of the two. Studies that have disentangled the effects of current population size and purported colonization history have sometimes but not always found that historic processes have a predominant role in population genetic diversity and structure (reviewed by Eckert *et al.*, 2008).

The geographic distribution of population sizes in the rare North American orchid *Isotria medeoloides* (Pursh) Raf. permits us to examine whether historical colonization processes may overshadow contemporary population genetic processes. In contrast to the case where populations in previously glaciated regions are smaller than those in refugial areas, *I. medeoloides* populations of greater than 100 individuals are all in formerly glaciated New England, with populations in unglaciated areas of the southern Appalachians having fewer than 20 individuals (USFWS, 2007). Thus, we can assess the relative contributions of historical and contemporary processes. If the large populations in New England are genetically depauperate, this would suggest that colonization bottlenecks dictate population genetic diversity, whereas if large populations have relatively high genetic diversity, we can conclude that post-colonization population genetic processes prevail.

In addition to their impact on population genetic diversity, historical colonization processes can leave lasting imprints on

¹Department of Biology, Colby College, Waterville, ME, USA; ²Information Technology, Colby College, Waterville, ME, USA and ³Maine Natural Areas Program, Department of Conservation, Augusta, ME, USA

Correspondence: Dr JL Stone, Department of Biology, Colby College, 5720 Mayflower Hill Drive, Waterville, ME 04901, USA.

E-mail: jstone@colby.edu

Received 3 August 2011; revised 20 March 2012; accepted 16 April 2012; published online 13 June 2012

population genetic structure. If colonization were to proceed according to a stepping-stone or infinite island model across a homogenous landscape, genetic structure would vary somewhat predictably as a function of the migration rate, growth rate and carrying capacity (reviewed by Excoffier *et al.*, 2009). In contrast, leptokurtic dispersal establishes strong genetic structuring that persists for many generations (Ibrahim *et al.*, 1996). *I. medeoloides* should have strongly leptokurtic dispersal, with an excess of both short and long dispersal distances relative to a normal distribution. Tiny, dust-like orchid seeds often land near the parent plant, but rare long-distance events can carry them many kilometers (Arditti and Ghani, 2000; Murren, 2003; Trapnell and Hamrick, 2004). Leptokurtosis of gene flow in *I. medeoloides* is further increased by its mating system. Unlike most orchids, which require insect pollination, gene flow by pollen is negligible in *I. medeoloides* due to its highly selfing mating system (Mehrhoff, 1983; Vitt and Campbell, 1997). Self-fertilizing species generally have strong population genetic structure (Hamrick and Godt, 1996). Therefore, both historical and contemporary processes should generate strong genetic structure throughout the range of this species, although refugial populations may be more distinct to the extent that they are older and have been isolated longer than leading-edge populations (Hampe and Petit, 2005).

The objective of this study was to examine the effects of post-glacial colonization and contemporary processes on population genetic diversity and structure of the rare orchid *I. medeoloides*. We obtained genetic material from more than 10% of known individuals throughout the species range. We used coalescent modeling and frequency-based analyses on genotypes obtained from four variable microsatellite loci to investigate the relationship between geographic location, census population size, and genetic diversity and structure. Specifically, we (1) contrasted diversity and differentiation of populations in glaciated versus unglaciated regions, and (2) asked whether distance to northern range limit or census population size correlated with genetic diversity and differentiation.

MATERIALS AND METHODS

Study species and sampling

I. medeoloides (Vanilloideae) is a non-clonal terrestrial orchid with 150 known populations distributed from Maine to Georgia along the coastal plain and nearby uplands of the eastern United States (USFWS, 2007). It is found in deciduous or mixed forests, often with a fragipan soil (NatureServe, 2011). Most populations contain fewer than 20 emergent stems (USFWS, 2007), with typically only one-third of emergent stems flowering in a given season (Mehrhoff, 1983). Individual plants may remain dormant under the ground for up to 4 years (USFWS, 2007). Emergent stems produce a single whorl of leaves and a solitary terminal flower that persists for 4–7 days (Mehrhoff, 1983). In contrast to its congener, *I. verticillata*, which has brightly colored and fragrant flowers, the flowers of *I. medeoloides* are pale green and scentless (Mehrhoff, 1983). Also in contrast to its congener, which is adapted for cross-pollination, *I. medeoloides* possesses a mechanism for autonomous self-pollination: as the flower develops, the anther drops downward and extrudes mealy pollen until it contacts the stigmatic surface (Mehrhoff, 1983). It appears that *I. medeoloides* routinely self-fertilizes in nature. During 70 h of observation in a large population in southern Maine, no insect visitors were seen (Vitt and Campbell, 1997). A manipulative experiment showed that flowers excluded from pollinator visits produced just as many seeds as control flowers or flowers with supplemental pollen added (Vitt and Campbell, 1997). About 80% of flowers initiate fruit production (Mehrhoff, 1983; Vitt and Campbell, 1997), and the majority of those develop into ripe capsules, producing about 10 000 seeds each (Mehrhoff, 1983).

I. medeoloides was identified as endangered by the US Fish and Wildlife Service (USFWS) in 1982, spurring the quest to locate additional populations and safeguard known populations. Success of these efforts led to

reclassification of the species as threatened in 1994 (USFWS, 2007). Owing to the extensive mapping efforts associated with conservation, the species has a well-documented geographic range. We used the coordinates of the most northeasterly population to define the northern range limit. Demographic population size was estimated by complete census of emergent stems. We used the most recent census data available, which ranged from 2008 to 2010.

Sampling for genetic analysis was concentrated in the northern part of the range, where all large populations are found. Most of the 83 known populations outside of Maine and New Hampshire contain fewer than 20 stems (USFWS, 2007). Exact locations of populations cannot be provided because pressure from collectors is a leading threat to this species (USFWS, 2007). In Maine, we sampled 5 of the 18 known populations and in New Hampshire, we sampled 11 of the 49 known populations. In both cases, sampled populations reflect the statewide geographic distribution of the species. We collected tissue from 1 of the 33 known populations in Virginia and from 3 of the 19 known populations in Georgia. We were not able to collect tissue from any other southern populations, due to extreme rarity in this portion of the species range, or from very small populations in New Hampshire or Maine. In populations having more than 20 stems, tissue was obtained from 10 to 20 stems distributed throughout the population at a spacing ≥ 0.5 m, reaching a sampling intensity of at least 10% of the total population. In populations having fewer than 20 stems, we collected tissue from every individual. A total of 299 individuals were sampled from 20 populations (Figure 1).

Figure 1 Locations of sampled populations of *I. medeoloides*. Inset shows New England populations. Locations of Georgia populations are shown overdispersed, for visibility.

Genotyping

Leaf tissues were dried in silica gel and stored at -80°C . DNA was extracted using the Qiagen DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Seven DNA samples from throughout the species range were sent to C Newton of ATG Genetics (Comox, BC, Canada) for creation and screening of a microsatellite-enriched library. Of the 46 loci screened, 7 polymorphic loci were found; 4 of these could be amplified and scored consistently, and were used in all subsequent analyses (Table 1). We chose to focus on polymorphic loci because we were most interested in the distribution of genetic variation within the species, rather than in comparison with other species.

PCR reactions were carried out in a 25- μl reaction volume containing 1.5 mM MgCl_2 , 0.2 mM of each dNTP, 0.5 μM of each primer, 2 U *Taq* polymerase (New England Biolabs, Ipswich, MA, USA) and 10 ng DNA. In each reaction, the left primer was fluorescently tagged with 6-FAM (Operon, Huntsville, AL, USA). A touchdown PCR program was used. The initial denaturation was at 95°C for 2.5 min, followed by 35 cycles and a final extension at 72°C for 10 min. Each cycle consisted of 95°C for 20 s, annealing for 30 s and extension at 72°C for 30 s. The annealing temperature was decreased from 60 to 50°C by 0.5°C each cycle for the first 20 cycles and kept at 50°C thereafter. PCR products were suspended in deionized formamide with ROX-500 size standard and run under standard conditions on an ABI 3130 genetic analyzer (Applied Biosystems, Carlsbad, CA, USA). Peaks were visualized using Applied Biosystem's GENEMAPPER 4.0 and scored manually.

Population genetic analyses

We used the computer program *INEST* to estimate the frequency of null alleles at each of the four loci (Chybicki and Burczyk, 2009). Unlike other approaches, which assume random mating to estimate the prevalence of null alleles, *INEST* simultaneously estimates inbreeding coefficient and null-allele frequency. Null-allele frequencies were estimated using 10 000 iterations of the Individual Inbreeding Model on genotypes of individuals from New England populations. Observed heterozygosity, expected heterozygosity and inbreeding coefficient were obtained using the computer program *GENALEX* 6.4 (Peakall and Smouse, 2006). We used the computer program *HP-RARE* (Kalinowski, 2005), to estimate allelic richness corrected for sample size. Population differentiation was described by taking the average pairwise R_{ST} calculated in *ARLEQUIN* 3.5.1.2 (Excoffier and Lischer, 2010) according to the method of Slatkin (1995). We compared expected heterozygosity and population differentiation of northern versus southern populations using a Mann–Whitney *U*-test, implemented in *STATA* 10.1 (Stata Corporation, College Station, TX, USA). Spearman's correlation coefficients were used to evaluate the effects of distance from northeastern range margin and census population size on genetic diversity and differentiation.

The coalescent simulation model *MIGRATE* 3.1.10 with Bayesian inference was used to estimate effective population sizes ($\theta = 4N_e\mu$) and historical migration rates ($M = m/\mu$), where N_e = effective population size, μ = mutation rate, and m = immigration rate (Beerli and Felsenstein, 2001; Beerli, 2006). For more than 95.5% of alleles, allele sizes fell within a distribution that encompassed

each possible repeat length between the two extremes. We therefore used a single-step microsatellite mutation model. Based on low values of M seen in early runs, we reduced the prior of M from 1000 to 100 to speed convergence. All other settings followed *MIGRATE* defaults. Estimates were combined from five replicate runs. Each run began with a random genealogy and visited 500 000 genealogies, recording every 100 steps for a recorded chain length of 5000. Most parameter estimates had effective sample sizes of greater than 1000.

We used *BAYESASS* 1.3 (Wilson and Rannala, 2003) to estimate recent gene flow among populations. Unlike *MIGRATE*, which assumes random mating, *BAYESASS* explicitly incorporates inbreeding into the model. Northern and southern populations were analyzed separately, each for two replicates with different random seeds. *GENALEX* 6.4 was used to conduct Mantel tests for isolation by distance, separately for all populations and for the New England populations only.

RESULTS

Genetic diversity and population structure

Genetic diversity is very low in *I. medeoloides* as revealed by the low polymorphism of microsatellites in the 46 loci initially screened. At the four variable loci analyzed, the mean number of alleles per locus within populations ranged from 1 to 5.75, with genetic diversity especially low in the southern populations (Table 2). When corrected for sampling effort by rarefaction, estimates of allelic richness decreased slightly, but the patterns of variation remained consistent (Table 2; Spearman's $r = 0.98$, $df = 18$). Genotyping confirmed field observations of selfing, with observed heterozygosity ranging from 0.000 to 0.076 and an average inbreeding coefficient of $F = 0.95$. The

Table 2 Demographic population size, genetic diversity, inbreeding coefficient and effective population size of sampled *I. medeoloides* populations, north to south

Population	State	Census size	Sample size	A	A ^s	H _o	H _e	F	θ
1	ME	180	20	3.25	2.43	0.000	0.456	1.000	0.0140
2	NH	107	16	4.25	3.57	0.080	0.728	0.884	0.0807
3	NH	37	13	2.50	2.18	0.000	0.403	1.000	0.0007
4	ME	103	20	4.25	3.36	0.076	0.672	0.888	0.0540
5	ME	70	12	3.25	2.81	0.042	0.533	0.900	0.0007
6	ME	30	19	3.50	2.87	0.026	0.608	0.957	0.0193
7	NH	29	14	2.75	2.11	0.000	0.324	1.000	0.0007
8	NH	55	10	1.00	1.00	0.000	0.000	1.000	0.0007
9	NH	38	10	2.00	1.79	0.025	0.284	0.916	0.0007
10	NH	128	19	2.50	1.73	0.000	0.206	1.000	0.0007
11	NH	43	18	5.75	3.81	0.057	0.685	0.918	0.2273
12	ME	135	17	4.25	3.18	0.029	0.574	0.959	0.0833
13	NH	89	20	3.50	3.04	0.025	0.645	0.966	0.0153
14	NH	16	15	1.25	1.25	0.000	0.124	1.000	0.0007
15	NH	32	10	2.00	1.93	0.000	0.361	1.000	0.0007
16	NH	107	20	1.25	1.24	0.038	0.100	0.624	0.0007
17	VA	11	11	1.25	1.15	0.000	0.041	1.000	0.0007
18	GA	10	10	1.00	1.00	0.000	0.000	1.000	0.0007
19	GA	12	12	1.00	1.00	0.000	0.000	1.000	0.0007
20	GA	13	13	1.25	1.13	0.000	0.036	1.000	0.0007
Mean	North	75.0	15.8	2.95	2.39	0.03	0.42	0.93	0.0313
Mean	South	11.5	11.5	1.12	1.12	0.02	0.34	0.95	0.0007

Abbreviations: A, alleles per locus; A^s, alleles per locus corrected for sample size; F, inbreeding coefficient; GA, Georgia; H_o, observed heterozygosity; H_e, expected heterozygosity; ME, Maine; NH, New Hampshire; VA, Virginia. $\theta = 4N_e\mu$. As the population genetic data were obtained from unusually variable loci, these summary statistics overestimate genetic diversity, and should be used for intra-specific comparisons only.

Table 1 Variable microsatellite loci identified in *I. medeoloides*

Locus	Primer sequence (5'–3')	T _m (°C)	Fragment length (bp)	Number of repeats	Number of alleles
ISME33	TCTCATTGGATCTAGCACAGAG TAGGGTAAGCCTATAACCAAG	60.1 60.1	396–432	14–31	8
ISME35	CCTAGACCTAAAGTTATCAAC ACTAGTGAGAGGAGAGTGGC	55.4 60.5	270–280	44–50	7
ISME36	GGGTTCTATAATAGTATTAGCTTC GTTTGTCGTAATCACAGTATCC	58.3 58.4	218–238	27–37	11
ISME46	GTATATGTAGAGCAACATGCTTC AAGCTTCTCAAGAAATGCCTTC	60.3 58.4	235–249	39–46	8

Repeat motif is uninterrupted CT for all loci. T_m = melting temperature.

high inbreeding coefficient cannot be attributed to undetected null alleles. INEST estimated null-allele frequency to be 0.004–0.005 at each locus and simultaneously estimated $F = 0.96$. Census population size was positively correlated with expected heterozygosity ($r = 0.58$, $n = 20$, $P = 0.007$).

Populations were highly differentiated, with species-wide $R_{ST} = 0.485$ (Table 3, Figure 2). Two of the Georgia populations were separated by only 6 km and were completely diverged, with pairwise $R_{ST} = 1.000$ (Figure 2). Populations in previously glaciated regions had significantly higher expected heterozygosity (Mann–Whitney $U = 197$, $n_1 = 16$, $n_2 = 4$, $P = 0.006$) and lower mean pairwise R_{ST} (Mann–Whitney $U = 136$, $n_1 = 16$, $n_2 = 4$, $P = 0.003$) than the populations in southern, refugial regions (Table 2). For northern populations, expected heterozygosity increased with proximity to the northeastern range limit ($r = -0.57$, $n = 16$, $P = 0.02$; Figure 3). Population differentiation in New England was strongly affected by population size, with smaller populations having higher average pairwise R_{ST} ($r = -0.63$, $n = 16$, $P = 0.008$). A Mantel test found significant isolation-by-distance across the species range ($r = 0.41$, $n = 20$, $P = 0.01$), but not among New England populations ($r = 0.06$, $n = 16$, $P = 0.23$).

Effective population sizes and migration rates

Posterior distributions showed strong convergence on estimates of effective population size for all but one population (Supplementary Figure 1). Demographically small populations had miniscule effective population sizes, with an estimated $\theta = 0.0007$, corresponding to $N_e = 1$ for a mutation rate of $\mu \sim 2 \times 10^{-4}$ (Table 2). For populations with greater than 20 stems, modal values of θ ranged from 0.0007 to

0.23, corresponding to estimated effective population sizes ranging from 1 to 378 individuals. According to the coalescent model implemented in MIGRATE, mutation was much more important than migration in introducing new alleles, with estimated $m/\mu = 0.067$ for all but three of the pairwise migration rates considered. In every case, the estimated immigration rate was fewer than one individual in ten generations. According to output from BAYESASS, the nonimmigration rate in the current generation averaged 95% across all populations, with the upper confidence interval at 99–100% for 17 of the 20 populations (Table 3).

DISCUSSION

Population genetic analysis of microsatellite genotypes in *I. medeoloides* confirmed field observations of predominant self-fertilization, with an average $F = 0.95$ (Table 2). Genetic diversity increased with increasing

Table 3 Genetic differentiation and nonimmigration rates for populations of *I. medeoloides*, north to south

Population	State	Census size	Mean pairwise R_{ST}	Mean nonimmigration rate	Confidence interval
1	ME	180	0.282	0.954	0.886–1.000
2	NH	107	0.268	0.974	0.906–1.000
3	NH	37	0.565	0.829	0.742–0.917
4	ME	103	0.328	0.772	0.692–0.893
5	ME	70	0.562	0.954	0.856–0.998
6	ME	30	0.382	0.979	0.927–0.999
7	NH	29	0.336	0.969	0.902–1.000
8	NH	55	0.549	0.973	0.909–0.999
9	NH	38	0.572	0.968	0.892–0.999
10	NH	128	0.225	0.981	0.934–1.000
11	NH	43	0.271	0.955	0.848–0.999
12	ME	135	0.397	0.976	0.917–0.999
13	NH	89	0.305	0.892	0.799–0.994
14	NH	16	0.613	0.981	0.936–1.000
15	NH	32	0.430	0.969	0.900–0.998
16	NH	107	0.422	0.986	0.944–0.999
17	VA	11	0.858	0.975	0.909–0.999
18	GA	10	0.887	0.975	0.905–0.999
19	GA	12	0.841	0.976	0.922–0.999
20	GA	13	0.824	0.978	0.922–0.999
Mean	North	75.0	0.407	0.945	0.874–0.987
Mean	South	11.5	0.852	0.976	0.914–0.999

Abbreviations: GA, Georgia; ME, Maine; NH, New Hampshire; VA, Virginia. Nonimmigration rates estimated separately for southern and northern populations.

Figure 2 Allele-frequency distributions for a subset of populations at loci ISME35 and ISME36, showing extreme divergence of southern populations. Populations are ordered north to south, and population codes appear on the left.

Figure 3 Expected heterozygosity declines with distance from the northern range limit in New England *I. medeoloides*. Circle size is proportional to census population size.

census size and with proximity to the northeastern range limit. Populations with small census sizes had small effective population sizes, in some cases with an estimated $N_e = 1$, representing the single selfing lineage populating the site. Gene flow between populations is extremely limited in *I. medeoloides*, with an average $R_{ST} = 0.485$. Coalescent analysis inferred that mutation was far more influential than gene flow in introducing new alleles into populations.

Methods of inferring historical gene flow carry assumptions that are violated in actual populations. The coalescent simulations implemented in MIGRATE assume that the populations are at equilibrium, which occurs on the order of $4N_e$ generations. This assumption is suspect for most post-glacial colonists, although it may be reasonably accurate for small populations of *I. medeoloides*. MIGRATE also assumes random mating. A completely self-fertilizing lineage is expected to coalesce in $2N_e$ rather than $4N_e$ generations (Schoen *et al.*, 1996; Nordborg, 2000), causing uncorrected coalescent models to overestimate θ in self-fertilizing populations. Our reported values of θ are therefore at least twice the actual values.

Population differentiation in *I. medeoloides* far surpasses the average $G_{ST} = 0.184$ of orchids estimated from codominant markers as found in a recent review (Forrest *et al.*, 2004). Numerous subsequent studies have also found low genetic differentiation among orchid populations (for example, Chung *et al.*, 2005; Esfeld *et al.*, 2008; Swarts *et al.*, 2009). Self-fertilization may contribute to the high differentiation of *I. medeoloides* populations (Hamrick and Godt, 1996), although *Neotina maculata*, another selfing orchid species, has low population differentiation (Duffy *et al.*, 2009). The tiny and widely separated populations of *I. medeoloides* may also contribute to its strong genetic structure. Small population sizes of orchid species can cause restricted gene flow among populations, despite the great dispersal capacity of the seeds (Tremblay *et al.*, 2005). In the case of *I. medeoloides*, the known 150 populations are distributed over an area of about 75 000 km². In the heart of the species range in northern New England, populations reach an average density of approximately one population per 58 km². Given that a large population occupies an area of about 0.01 km², it is unsurprising that seed dispersal among populations is exceedingly rare.

Populations of *I. medeoloides* in unglaciated regions are especially genetically depauperate and divergent. Allele-frequency distributions for the Georgia populations span the range of allele sizes found

throughout the species (Figure 2), and two populations separated by only 6 km are completely differentiated. It would be unlikely for these allele-frequency distributions to have arisen by independent long-distance dispersal events because the extreme alleles are rarely found in northern populations. It seems more probable that the southernmost populations are remnants of larger populations in the southern Appalachian or nearby refugial regions. Based on lower population sizes and higher-altitude locations of southern populations (600–700 m above sea level compared with 100–200 m above sea level in the north), Holocene warming has presumably degraded habitats in the southern portion of the range. The southernmost populations of *I. medeoloides* therefore conform to expectations for relictual rear edge populations, with their great divergence, apparent antiquity and location in mountainous terrain that permits elevational migration in response to climate change (Hampe and Petit, 2005).

Populations in the glaciated part of the species range stem from relatively recent colonization events, probably within the past 6000 years, when mixed forests became established in northern New England (Jacobson *et al.*, 1986). In contrast to our expectation that stepping-stone dispersal would cause successive bottlenecks, the best predictor of genetic diversity in these populations is proximity to the northern range limit (Figure 3). Simulation models show that the mode of dispersal critically affects the distribution of genetic diversity during species range expansions, such as those following glaciation (Excoffier *et al.*, 2009). By disseminating variants far from their place of origin, long-distance dispersal can maintain diversity during colonization (Bialozyt *et al.*, 2006). Under this scenario, diversity is maintained at the regional rather than population level, with high differentiation among populations. Exponential population growth in newly colonized sites can also preserve genetic diversity near the leading edge, as rare alleles are less likely to be lost to drift (Excoffier *et al.*, 2009). This 'surfing' phenomenon is enhanced by small deme size and low migration between demes (Klopfstein *et al.*, 2006). The lack of isolation-by-distance among New England populations of *I. medeoloides*, together with exceedingly low estimated migration rates, suggest that colonization for this species proceeded by rare long-distance dispersal rather than by a diffusion or stepping-stone process.

Numerous studies have documented declines in genetic diversity toward northern range limits in previously glaciated regions (Hewitt, 2004). In cases where census sizes are smaller at the periphery than in the center of the range, decreased variation at the northern range limit could be attributed either to founder effects or to the expected decline in diversity associated with small census sizes (Leimu *et al.*, 2006; Eckert *et al.*, 2008). Statistical modeling approaches can attempt to disentangle the effects of census size and colonization history (for example, Johansson *et al.*, 2006). Another approach to separate the impacts of colonization from more contemporary forces is to consider species whose census population sizes do not decline toward the northern range limit. For example, census size of the herbaceous perennial *Silene nutans* is uncorrelated with latitude, allowing investigators to conclude that its poleward decline in genetic diversity is due to historical colonization processes rather than more contemporary ones (Van Rossum and Prentice, 2004). Increased genetic diversity toward the northern range limit in previously glaciated regions appears to be an unusual phenomenon. It has been documented in *Alnus rubra*, where coastal refugia lie northward of the present-day distribution (Hamann *et al.*, 1998), and in *Arabidopsis lyrata*, where genetically diverse outcrossing populations lie to the north of selfing, ephemeral populations that establish in transient sand dune and highly human impacted habitats (Mable and Adam, 2007). Enhanced genetic diversity in formerly glaciated regions has

also been found in the North American orchids *Cypripedium parviflorum* (Wallace and Case, 2000) and *Cypripedium reginae* (Kennedy and Walker, 2007). Like *I. medeoloides*, these *Cypripedium* orchids are much more abundant in the northern parts of their ranges, where they are apparently better adapted to current conditions (Wallace and Case, 2000; Kennedy and Walker, 2007). In addition, *C. parviflorum*, like *I. medeoloides*, shows allelic distribution patterns inconsistent with a stepping-stone model of colonization.

Our results are consistent with a scenario in which rare founder events lay the foundation for low within-population variation and high population genetic structure, and subsequent processes magnify these initial tendencies. Gene flow among populations is practically nonexistent, with populations continuing to diverge following initial establishment. Populations consist of a collection of selfing lineages, exacerbating genetic drift (Charlesworth and Pannell, 2001; Siol *et al.*, 2007). As environmental conditions are evidently most favorable near the northern range limit, young populations near the leading edge maintain relatively large census sizes, retaining an increasing number of lineages generated by mutation, and forestalling the inevitable loss of lineages due to genetic drift. Although colonization by long-distance dispersal may have initiated genetic patterning of *I. medeoloides* populations, subsequent demographic processes play the determining role in population genetic diversity and structure.

DATA ARCHIVING

Genotypes have been deposited at Dryad, doi:10.5061/dryad.ds3v4.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank S Cairns of the New Hampshire Heritage Bureau and B Brumback of the New England Wildflower Society for tissue sampling, generous exchange of information and helpful discussions. We thank numerous additional collaborators for tissue sampling, and three anonymous reviewers as well as the associate editor for helpful comments. Financial support was provided by the Maine Department of Conservation, National Science Foundation (DEB-0841482) and the National Institutes of Health (P20 RR-016463) from the INBRE Program of the National Center for Research Resources.

Alsos IG, Eidesen PB, Ehrich D, Skrede I, Westergaard K, Jacobsen GH *et al.* (2007). Frequent long-distance plant colonization in the changing Arctic. *Science* **316**: 1606–1609.

Arditti J, Ghani AKA (2000). Tansley Review 110: numerical and physical properties of orchid seeds and their biological implications. *New Phytol* **145**: 367–421.

Austerlitz F, Jung-Muller B, Godelle B, Gouyon P-H (1997). Evolution of coalescence times, genetic diversity and structure during colonization. *Theor Pop Biol* **51**: 148–164.

Beerli P (2006). Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* **33**: 341–345.

Beerli P, Felsenstein J (2001). Maximum-likelihood estimation of a migration matrix and effective population numbers in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* **98**: 4563–4568.

Bialozyt R, Ziegenhagen B, Petit RJ (2006). Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *J Evol Biol* **19**: 12–20.

Charlesworth D, Pannell JR (2001). Mating systems and population genetic structure in the light of coalescent theory. In: Silvertown J, Antonovics J (eds). *Ecology and Evolution in a Spatial Context*. Blackwell Science: Oxford, pp 73–96.

Chung MY, Nason JD, Chung MG (2005). Spatial genetic structure in populations of the terrestrial orchid *Orchis cyclophila* (Orchidaceae). *Pl Syst Evol* **254**: 209–219.

Chybicki IJ, Burczyk J (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered* **100**: 106–113.

Duffy KJ, Scopece G, Cozzolino S, Fay MF, Smith RJ, Stout JC (2009). Ecology and genetic diversity of the dense-flowered orchid, *Neotinea maculata*, at the centre and edge of its range. *Ann Bot* **104**: 507–516.

Eckert CG, Samis KE, Loughheed C (2008). Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Mol Ecol* **17**: 1170–1188.

Esfeld K, Hensen I, Wesche K, Jakob SS, Tischew S, Blattner FR (2008). Molecular data indicate multiple colonizations of former lignite mining areas in Eastern Germany by *Epipactis palustris* (Orchidaceae). *Biodivers Conserv* **17**: 2441–2453.

Excoffier L, Foll M, Petit RJ (2009). Genetic consequences of range expansions. *Ann Rev Ecol Syst* **40**: 481–501.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res* **10**: 564–567.

Forrest AD, Hollingsworth ML, Hollingsworth PM, Sydes C, Bateman RM (2004). Population genetic structure in European populations of *Spiranthes romanzoffiana* set in the context of other genetic studies on orchids. *Heredity* **92**: 218–227.

Hamann A, El-Kassaby YA, Koshy MP, Namkoong G (1998). Multivariate analysis of allozymic and quantitative trait variation in *Alnus rubra*: geographic patterns and evolutionary implications. *Can J For Res* **28**: 1557–1565.

Hamrick JL, Godt MJW (1996). Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond B Biol Sci* **351**: 1291–1298.

Hampe A, Petit RJ (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecol Lett* **8**: 461–467.

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. *Philos Trans R Soc Lond B Biol Sci* **359**: 183–195.

Ibrahim KM, Nichols RA, Hewitt GM (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* **77**: 282–291.

Jacobson Jr GL, Webb III T, Grimm EC (1986). Changing vegetation patterns of eastern North America during the past 18 000 years: inferences from overlapping distributions of selected pollen types. In: Ruddiman WF, Wright Jr HE (eds) *North America and Adjacent Oceans During the Last Deglaciation*. Geological Society of America: Boulder, Plate 2.

Johansson M, Primmer CR, Merilä J (2006). History vs current demography: explaining the genetic population structure of the common frog (*Rana temporaria*). *Mol Ecol* **15**: 975–983.

Kalinowski ST (2005). HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes* **5**: 187–189.

Kennedy AH, Walker GL (2007). The population genetic structure of the showy lady's slipper orchid (*Cypripedium reginae* Walter) in its glaciated and unglaciated ranges. *Castanea* **72**: 248–261.

Klopfstein S, Currat M, Excoffier L (2006). The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* **23**: 482–490.

Knowles LL, Alvarado-Serrano DF (2010). Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic, and genetic models in montane grasshoppers. *Molecular Ecology* **19**: 3727–3745.

Leimu R, Mutikainen P, Koricheva J, Fischer M (2006). How general are positive relationships between plant population size, fitness, and genetic variation? *J Ecol* **94**: 942–952.

Mable BK, Adam A (2007). Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Mol Ecol* **16**: 3565–3580.

Mehrhoff III LA (1983). Pollination in the genus *Isotria* (Orchidaceae). *Am J Bot* **70**: 1444–1453.

Murren CJ (2003). Spatial and demographic population genetic structure in *Catasetum viridiflavum* across a human-disturbed habitat. *J Evol Biol* **16**: 333–342.

NatureServe (2011). *NatureServe Explorer*. An online encyclopedia of life [web application]. Version 7.1 NatureServe: Arlington, Virginia, Available. <http://www.natureserve.org/explorer> (Accessed: November 11, 2011).

Nordborg M (2000). Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* **154**: 923–929.

Peakall R, Smouse P (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* **6**: 288–295.

Pielou EC (1991). *After the Ice Age: The Return of Life to Glaciated North America*. University of Chicago Press: Chicago.

Schoen DJ, Morgan MT, Bataillon T (1996). How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation. *Phil Trans R Soc Lond B* **351**: 1281–1290.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009). Evolution and ecology of species range limits. *Annu Rev Ecol Syst* **40**: 415–436.

Siol M, Bonin I, Olivier I, Prosperi JM, Ronfort J (2007). Effective population size associated with self-fertilization: lessons from temporal changes in allele frequencies in the selfing annual *Medicago truncatula*. *J Evol Biol* **20**: 2349–2360.

Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.

Swarts ND, Sinclair EA, Krauss SL, Dixon KW (2009). Genetic diversity in fragmented populations of the critically endangered spider orchid *Caledonia huegelii*: implications for conservation. *Conserv Genet* **19**: 1199–1208.

Trapnell DW, Hamrick JL (2004). Partitioning nuclear and chloroplast variation at multiple spatial scales in the neotropical epiphytic orchid, *Laelia rubescens*. *Mol Ecol* **13**: 2655–2666.

Tremblay RL, Ackerman JD, Zimmerman JK, Calvo RN (2005). Variation in sexual reproduction in orchids and its evolutionary consequences: a spasmodic journey to diversification. *Biol J Linn Soc Lond* **84**: 1–54.

USFWS (2007). *Small Whorled Pogonia (Isotria medeoloides) 5-Year Review: Summary and Evaluation*. U.S. Fish and Wildlife Service, New England Office: Concord, New Hampshire.

- Vitt P, Campbell CS (1997). Reproductive biology of *Isotria medeoloides* (Orchidaceae). *Rhodora* **99**: 56–63.
- Van Rossum F, Prentice HC (2004). Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae): population size, geographical position and immigration history. *Biol J Linn Soc* **81**: 357–371.
- Wallace LE, Case MA (2000). Contrasting allozyme diversity between northern and southern populations of *Cypripedium parviflorum* (Orchidaceae): implications for Pleistocene refugia and taxonomic boundaries. *Syst Bot* **25**: 281–296.
- Wilson GA, Rannala B (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**: 1177–1191.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)