MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4721–4734 doi: 10.1111/j.1365-294X.2012.05751.x

Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*

STEPHEN R. KELLER,* KIMBERLY J. GILBERT,†[1] PETER D. FIELDS† and DOUGLAS R. TAYLOR†

*Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Road, Frostburg, MD, 21532, USA, †Department of Biology, University of Virginia, PO Box 400328, Charlottesville, VA, 22904-4328, USA

Abstract

Species invading new ranges are subject to a series of demographic events that can strongly shape genetic diversity. Describing this demographic history is important for understanding where invasive species come from and how they spread, and is critical to testing hypotheses of postinvasion adaptation. Here, we analyse nuclear and chloroplast genetic diversity to study the invasion history of the widespread colonizing weed, *Silene latifolia* (Caryophyllaceae). Bayesian clustering and PCA revealed strong population structure in the native range of Europe, and although genotypes from multiple native sources were present in the introduced range of North America, the spatial distribution of genetic variance was dramatically reorganized. Using approximate Bayesian computation (ABC), we compared support for different invasion scenarios, including the number and size of independent introduction events and the amount of admixture occurring between sources of introduced genotypes. Our results supported independent introductions into eastern and western North America, with the latter forming a bridgehead for a secondary invasion into the Great Lakes region of central North America. Despite small estimated founder population sizes, the duration of the demographic bottleneck after the initial introduction appeared extremely short-lived. This pattern of repeated colonization and rapid expansion has effectively eroded the strong population structure and cytonuclear associations present in Europe, but has retained overall high genetic diversity since invasion. Our results highlight the flexibility of the ABC approach for constructing a narrative of the demographic history of species invasions and provide baseline for future studies of evolutionary changes in introduced *S. latifolia* populations.

Keywords: admixture, approximate Bayesian computation, bottleneck, chloroplast DNA, founder effect, invasion, microsatellite, range expansion, *Silene*

Received 14 February 2012; revision received 12 June 2012; accepted 20 June 2012

Introduction

Biological invasions present a unique opportunity to study evolution over contemporary timescales (Ellstrand & Schierenbeck 2000; Sakai *et al.* 2001; Lee 2002; Keller & Taylor 2008; Prentis *et al.* 2008). Changes in genetic diversity may play a key role in the ability of introduced populations of non-native species to expand their range and become invasive (Lambrinos 2004; Dlugosch & Parker 2008; Keller & Taylor 2008; Wilson *et al.* 2009). Species invading new ranges are subject to a series of demographic events that collectively have great potential to shape genetic diversity. The demography of invasion can result in decreased genetic diversity

Correspondence: Stephen R. Keller, Fax: +1-301-689-7200; E-mail: skeller@umces.edu

[1] Present address: Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, BC, V6T 1Z4, Canada

relative to the source populations of invasion, as in founder effects and genetic bottlenecks arising from the restricted sampling and establishment of few colonists (Dlugosch & Parker 2008). However, invading species may maintain or even increase existing diversity above that found in the source populations via admixture among multiple introductions and recombination between divergent genomes (Kolbe *et al.* 2004; Lavergne & Molofsky 2007). While the genetic impacts of different demographic events are often studied and discussed by invasion biologists as alternative outcomes of a species' introduction history (were invading populations bottlenecked or admixed; introduced in a singular event or over multiple occasions?), they are in fact nonexclusive of each other.

Given the abundant evidence for bottlenecks and multiple introductions among successful invaders (Dlugosch & Parker 2008), we expect that most biological invasions encompass a complex series of demographic outcomes. For example, multiple introductions individually may be bottlenecked to differing degrees, which then experience admixture following secondary contact (e.g. Taylor & Keller 2007; Keller & Taylor 2010).

Our ability to identify introduction sources, detect the occurrence of genetic bottlenecks and/or admixture and estimate the contribution of rapid adaptation and hybridization to the emergence of invasiveness all depend on a more integrated understanding of how genetic diversity has been repeatedly sampled and reorganized during the invasion process (Keller & Taylor 2008; Estoup & Guillemaud 2010). Traditionally, the introduction history of a biological invasion has been inferred from genetic data through a combination of population genetic summary statistics that describe the quantity of diversity (e.g. allelic richness), population subdivision (F_{ST}) or spatial arrangement of genetic variation (AMOVA, genotypic clustering) in native and invasive populations (Genton *et al.* 2005; Schachner *et al.* 2008; Keller *et al.* 2009; Wilson *et al.* 2009; Gaudeul *et al.* 2011). More recently, approximate Bayesian computation (ABC) methods have emerged that allow inference of demographic history from population genetic data. These model-based approaches represent a potentially powerful alternative to traditional population genetic analyses for studying the introduction histories of nonnative species. With ABC, multiple hypotheses representing different demographic scenarios can be proposed a priori, with the genetic data being used to discriminate among the most probable scenarios within a model-based statistical framework (Beaumont *et al.* 2002; Hickerson *et al.* 2006; Knowles 2009; Nielsen & Beaumont 2009). By comparing summary statistics between empirical data and the best-matching simulations from a given scenario of demographic history, ABC can be used to discriminate among different competing models, estimate parameters of biological interest (founding population size, source populations, admixture rate) and incorporate statistical uncertainties such as the sampling completeness of potential native range source populations (Estoup *et al.* 2004; Estoup & Guillemaud 2010; Lombaert *et al.* 2010, 2011).

Here, we apply traditional population genetic analyses and ABC modelling to unravel the invasion history of the European weed *Silene latifolia* (Caryophyllaceae) into North America. Using a combination of multilocus genotypes from 16 nuclear microsatellite loci, along with one chloroplast indel polymorphism diagnostic for two major cytoplasmic lineages present in *S. latifolia*'s native range (Taylor & Keller 2007), we show that the invasion of North America involved a complex history of multiple introductions that admixed diversity from *S. latifolia*'s two native range subpopulations. Admixture in North America has not only reorganized population genetic structure within the nuclear and chloroplast genomes, but also begun to break down the associations between the genomes that characterize the native range. We further show that central North America is most likely a secondary invasion front derived from genotypes introduced to western North America. This complex narrative of the species' demographic history highlights the utility of ABC model-based approaches to discriminate among the different source populations and dispersal corridors during invasive range expansions.

Methods

Population samples and genotyping

Samples of *S. latifolia* were collected from 393 individuals from 40 populations spanning much of the distributional range in both Europe (native) and North America (introduced; Fig. 1a). Genomic DNA was extracted from leaf tissue either dried on silica gel or fresh from plants grown in a glasshouse using collected seeds. Extractions were performed using the methods adapted from Edwards *et al.* (1991) and Slotta *et al.* (2008). Individuals were genotyped at 16 microsatellite markers derived from *S. latifolia* (slat18, slat32, slat33, slat48, slat72, slat85) described by Abdoullaye *et al.* (2010), (SL 8) described by Teixeira & Bernasconi (2007), (SL_eSSR04, SL_eSSR06, SL_eSSR09, SL_eSSR12, SL_eSS16, SL_eSSR20, SL_eSSR29, SL_eSSR30) described by Moccia *et al.* (2009) and one marker derived from *S. vulgaris* (A11) described by Juillet *et al.* (2003).

Microsatellite amplifications were carried out either in single reactions or in multiplexes, depending on the locus, using a QIAGEN multiplex PCR kit following the

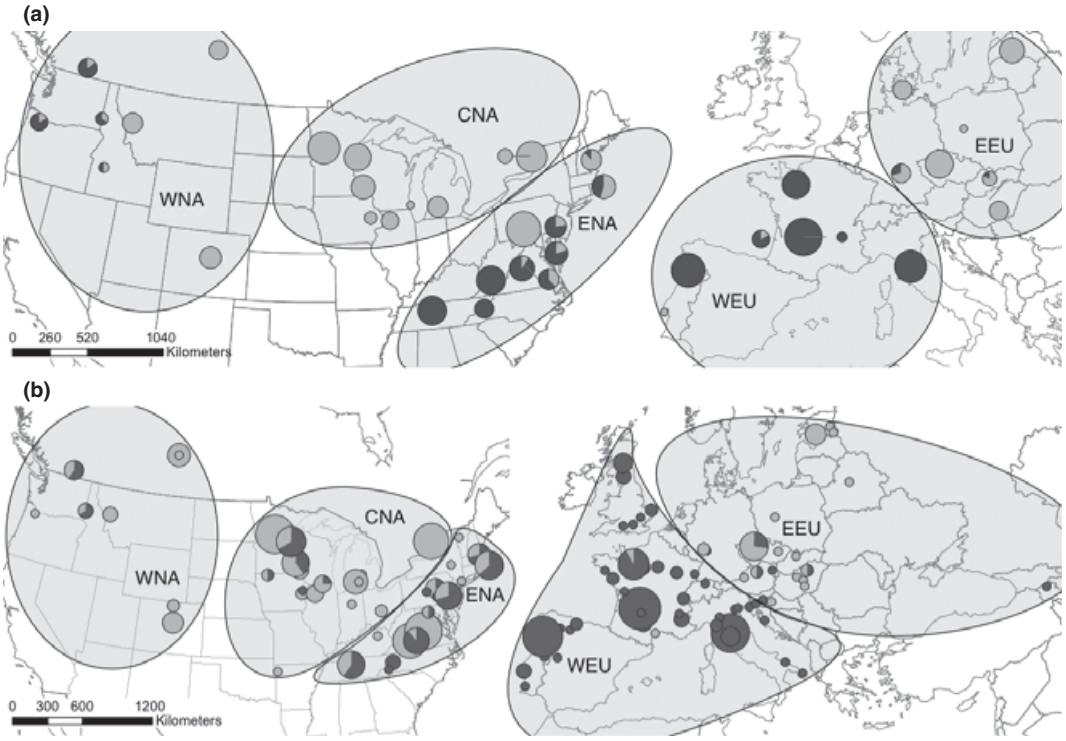

Fig. 1 Populations sampled in Europe and North America for (a) nuclear microsatellite genotyping and (b) chloroplast indel haplotyping. Pie charts show contribution of Eastern (light grey) and Western (dark grey) European ancestry, based on STRUCTURE analysis (nuclear) or indel haplotypes (chloroplast). Sizes are proportional to sample size within each population. Population assignment to regions for approximate Bayesian computation analysis is indicated by shaded areas.

manufacturer's suggested modifications to previously published protocols for each genetic marker. The fluorescently labelled PCR products were combined either in pools of different loci run from single PCRs or simply as the product of an individual multiplex PCR with a loading buffer of HiDi formamide and a size standard of either Genescan 400HD ROX or 500HD LIZ (Applied Biosystems). Samples were separated out on an ABI 3130xl automated sequencer. Genotype scores were determined with GENEMAPPER v3.0 software (Applied Biosystems) using automated scoring and double-checked manually, then binned with the program TANDEM (Matschiner & Salzburger 2009).

To investigate the effects of invasion on the breakdown of population structure and intergenomic associations, we haplotyped all 393 individuals plus an additional 249 individuals from 99 additional populations (Fig. 1b) for a chloroplast indel polymorphism that was developed based on a previous study of cpDNA sequences (Taylor & Keller 2007). Sequence diversity from this marker distinguished two main lineages of *S. latifolia* in the native range—one principally in Western Europe and another in Central and Eastern Europe. One of these diagnostic polymorphisms was a 7-bp insertion/deletion (indel) in the intergenic region between the *trn*S and *trn*G genes. We designed primers that flanked this indel (forward: 5′AAATTAAAATTTCGGGTCTTTC3′; reverse: 5′ATG AAATTCCAATCAGTTGTATAAAG3′) and amplified the region using PCR. Due to an additional neighbouring indel polymorphism within the amplified region, PCR products differed by a final fragment size of only 2 bp (West = 110 bp; East = 112 bp). Therefore, we detected fluorescently labelled fragments using an ABI 3130xl automated sequencer.

Statistical analyses

We examined genetic diversity at the individual, population, region and continent levels by calculating allelic diversity, heterozygosity and *F*-statistics using the software GenoDive version 2.0b20 (Meirmans & Van Tienderen 2004) and hierarchical population structure using the HIERFSTAT package (Goudet 2005) in the statistical software R version 2.13.2 (R Development Core Team 2011). The hierarchical analysis tested for genetic structure among broad-scale geographic regions and among local subpopulations nested within regions.

Five different regions were identified, a priori, based on physiographic provinces and natural breaks in the geographic extent of historical samples. In Europe, we

grouped individuals into Eastern and Western Europe (EEU and WEU, respectively) based on a well-known suture zone running north from the Alps to the North Sea. This phylogeographic boundary corresponds to different routes of postglacial colonization out of isolated Mediterranean refugia following Holocene warming (Hewitt 1996; Petit *et al.* 2003). Additional description of this phylogeographic region as it pertains to *Silene* is provided in the study by Taylor & Keller (2007) and Keller *et al.* (2009).

In North America, we grouped individuals into three geographic regions based on hypothesized invasion foci: an eastern region (ENA) that included the northeastern Atlantic port cities and was bounded to the west by the Mississippi and Ohio rivers and to the north by the St. Lawrence seaway; a central region (CNA) that included the ports and shipping routes around the Great Lakes industrial centres; and a western region (WNA) along the northwest Pacific coast and the intermountain west that included the northwest coastal port cities and the agricultural areas of the prairies. Available evidence from historical collections in University herbaria and in the early botanical literature of North America points to invasion of *S. latifolia* (=*S. alba*, =*Lychnis pratensis*, =*L. verspertina*) having first occurred into ENA, especially around the port cities of Philadelphia, New York, and Boston (Martindale 1876; Antonovics *et al.* 2003; J. Antonovics unpublished data). Later, invasion foci became evident around the inland ports of the Great Lakes, especially along the southern shores of Lake Michigan (Antonovics *et al.* 2003) and along the Saint Lawrence River, which gives shipping access to the Great Lakes (McNeill 1977). This region of relatively high specimen density was separated from the dense collections along coastal ENA by relatively few specimens, thus suggesting that CNA and ENA may represent separate establishment points of the invasion. Lastly, a third geographically disjunct region (WNA) was recognized based on the density of specimens in the Pacific northwest port cities and in the Western United States and Canadian prairies (McNeill 1977), suggestive of a possible separate introduction into the west coast. Collectively, the geographic distribution of historical specimens suggests the possible existence of separate invasion foci as a result of human shipping traffic and commerce, but at face value these scenarios are difficult to discriminate from a scenario of gradual spread punctuated by geographically biased collecting efforts.

We used Bayesian clustering to assign multilocus genotypes into clusters using the program STRUCTURE version 2.2 (Pritchard *et al.* 2000; Falush *et al.* 2003). We implemented the admixture model with allele frequencies correlated and remaining parameters set as default over 10 independent runs for each K (1–10) with 200 000 MCMC iterations and a burn-in period of 50 000 iterations. The number of clusters (K) was chosen based on the method of Evanno *et al.* (2005). Replicate STRUCTURE runs were combined with the software CLUMPP (Jakobsson & Rosenberg 2007) and visualized with *distruct* (Rosenberg 2004). We also explored the clustering of genotypes using a principal components analysis with the R package ade4 version 1.4–16 (Thioulouse & Dray 2007) to support our inferences from the Bayesian clustering.

To examine evidence for admixture and introgression during invasion, we tested for a shift in cytonuclear associations as would occur if intraspecific hybridization resulted in introgression of divergent cytoplasmic haplotypes onto contrasting nuclear backgrounds. We tested for significant shifts in cytonuclear associations using a generalized linear model, with the chloroplast haplotype predicted by continental range and the assignment of individuals to the nuclear STRUCTURE clusters. A significant interaction term indicates a shift in cytonuclear association between native and invaded ranges.

To infer the introduction history of *S. latifolia* into North America from Europe, we performed ABC analyses of demography based on the observed microsatellite data. For the ABC analysis, we pooled samples into 'populations' based on our a priori defined regional geographic groups ($N = 2$ in Europe, abbreviated EU; $N = 3$ in North America, abbreviated NA). To account for the possibility that we failed to sample a region from EU or elsewhere that contributed to the invasion, we also modelled the effects of a sixth unsampled 'ghost' population (see Cornuet *et al.* 2010; Guillemaud *et al.* 2010). These populations were used to evaluate 10 competing scenarios of invasion using the Linux cluster capabilities of DIYABC software version 1.0.4.46 (Cornuet *et al.* 2008; Fig. S1, Supporting information) on the University of Virginia cross campus computing grid. Scenarios varied according to the native range source(s) of introduced genotypes into a particular geographic region, the founding population size, the number of generations that founders remained at small size (i.e. bottleneck duration), the admixture rate if multiple sources were detected, and the final populations size (Table 1). Four of these scenarios incorporated the possibility of the unsampled 'ghost' population (scenarios 7–10; Fig. S1, Supporting information). We chose to test these 10 competing scenarios rather than an exhaustive list of all possible scenarios to focus computational effort on the most likely candidate models, and to eliminate certain sets of models that were unlikely based on our a priori knowledge of the invasion history from historical specimens (e.g. eastern NA

Table 1 Descriptions and prior settings for all parameters used in coding the 10 scenarios for DIYABC. Choices of priors were based on Pleistocene glacial history in Europe, and the approximate period of colonization of North America. Population size parameters are in units of population effective size (N_e), while time parameters (including bottleneck duration) are in units of generations

Parameter	Interpretation	Distribution	Minimum	Maximum	Step
N1, N2, …N6	Population size of EU regions (1–2), NA regions (3–5), and an unsampled 'ghost' region (6)	Uniform	5E3	6E5	1E3
t_a	Time to split of Eastern and Western EU	Uniform	5E3	2E5	5E2
t_b, t_c, t_d	Time to colonization event of NA regions	Uniform	1	3E2	1
N3F, N4F, N5F, N6F	Founding population size of regions	Uniform	1	5E3	10
$r_{3(2,1)}$, $r_{4(2,1)}$, $r_{5(2,1)}$	Admixture rate of EU 1 and 2 in NA regions	Uniform	0	1	1E-2
BDt_b, BDt_c, BDt_d	Bottleneck duration of NA regions	Log-uniform	1	10	1
t_{g2}	Time to split of ghost and eastern EU populations	Uniform	1E3	1E5	1E3
t_{g3}	Time to split of ghost and ancestral populations	Uniform	1E4	5E5	5E3
μ	Mean mutation rate	Uniform	1E-7	1E-4	
μ_i	Individual locus mutation rate*	Gamma	1E-8	1E-3	
μ_{SNI}	Mean single nucleotide insertion/deletion rate	Uniform	1E-10	1E-7	
μ_{SNI}	Individual locus single nucleotide indel rate*	Gamma	1E-11	1E-6	
P	Mean probability that a new mutant allele differs from its ancestor	Uniform	0.1	1.5	
P_i	Individual locus probability that a new mutant allele differs from its ancestor*	Gamma	0.01	2	

*Gamma distribution shape parameter = 2.0

having been invaded first, prior to invasion of central and western NA).

We set uniform priors on most demographic parameters (Table 1) except for bottleneck duration, which was given a log-uniform prior to reflect the expectation that most bottleneck events would be brief. Effective population sizes were given a prior U[5000–600 000] to accommodate a potentially wide range in abundance between native and invasive regions. The divergence between European regions probably occurred as a result of glacial divergence or during postglacial expansion (Taylor & Keller 2007), so we set this prior as U[5000–200 000] which allowed for divergence as early as two glacial cycles ago (the Riss glacial period, *c.* 200 000 ybp) and as recent as during post-Pleistocene expansion. We set the prior on the colonization events in NA as U[1, 300] to allow for the possibility of invasive populations having established prior to the historic record of *Silene* in NA herbarium collections (*c.* early 19th century) and also to allow for the possibility that some regions were colonized only recently. Upper and lower bounds for all priors were chosen and adjusted to encompass the high-probability regions from the posterior distribution based on exploratory modelling results.

Each scenario was simulated based on neutral coalescence for 1×10^6 iterations. A rejection step was then performed to keep only the 0.1% of simulations that most closely matched the observed data based on a series of 10 summary statistics. These included four single-population statistics (mean number of alleles, mean expected heterozygosity, mean allele size variance and mean ratio of number of alleles over the range in alleles sizes), and six pairwise-population statistics (mean individual assignment likelihoods, maximum-likelihood admixture proportion, mean number of alleles across loci, mean expected heterozygosity across loci, mean allele size variance across loci and population pairwise F_{ST}). Summing across all single and pairwise statistics, the total number of summary statistics was 85.

An important step in ABC modelling is model selection, checking and validation. We performed model selection by estimating posterior support for each model using two different methods: a direct estimate, based on the frequency of a given scenario within the 500 data sets closest in summary statistics to the observed, and a logistic regression estimate based on predicting the probability of a scenario from the deviations in the summary statistics among the 1% (50 000) closest simulated and observed data sets (Cornuet *et al.* 2008). To assess confidence in selection of the most probable model, we performed a-posteriori simulations separately for the best and next best models using each model's parameter estimates, and calculated the probability of false-positive and false-negative errors from 500 pseudo-observed data sets. We estimated the false-positive rate of a particular focal scenario to be the proportion of times that model selection resulted in choosing that scenario when in fact the data were simulated under an alternative scenario; similarly, the false-negative rate was estimated from the proportion of times that model selection resulted in choosing the alternative scenario when in fact the data were

simulated under the focal scenario. Lastly, we performed a model checking analysis by comparing the first three axes from a PCA on our observed summary statistics to those obtained from 1000 simulations based on the posterior predictive distribution of the best fitting model (Cornuet *et al.* 2010).

Results

The number of alleles (A) at the 16 microsatellite loci varied from 3 to 34 alleles per locus, with a mean of 14 alleles per locus. Within populations, the mean effective number of alleles (A_E) ranged from 1.8 to 3.7 and was not significantly different between EU and NA (EU: $\bar{A}_E = 3.1$; NA: $\bar{A}_E = 2.9$; t-test: $t = 1.34$, $P > 0.15$). Similarly, within-population expected heterozygosity (H_S) ranged from 0.462 to 0.668 and showed little difference between EU ($\bar{H}_S = 0.568$) and NA ($\bar{H}_S = 0.578$) populations ($t = -0.53$, $P > 0.60$).

While overall changes in population genetic diversity were not apparent since invasion, there was strong evidence that the structuring of diversity at different spatial scales was reorganized. In EU, microsatellite diversity was strongly structured both among regions ($F_{RT} = 0.134$) and among populations within regions ($F_{SR} = 0.147$). NA populations in contrast showed little regional-scale structure ($F_{RT} = 0.025$), although the magnitude of local population structure ($F_{SR} = 0.130$) was similar to EU. Overall structure among all populations across ranges (F_{ST}) was 0.179 for the microsatellites.

A similar breakdown in regional-scale population structure during invasion was observed for the chloroplast locus. Western and Eastern EU were nearly fixed for alternate haplotypes ($F_{RT} = 0.819$) with somewhat less structure at the population level ($F_{SR} = 0.498$). In NA, genetic structure among regions decreased dramatically ($F_{RT} = 0.118$), while among-population divergence was reduced slightly ($F_{SR} = 0.382$), indicating a redistribution of genetic variance during invasion from among regions in EU to within populations in NA. The signal of this change in the scale of population structure is visually evident in the range-wide distributions of the two cpDNA lineages that show a clear segregation of each haplotype in EU with much more mixing within populations and regions in NA (Fig. 1b).

The evolutionary divergence among nuclear genomes, as evidenced by Bayesian clustering, indicated the presence of historical isolation. The Evanno *et al.* (2005) delta-K approach returned high support for two genetic clusters (Fig. S2, Supporting information). Additional partitions beyond $K = 2$ generated noisy clustering results with lower consistency across replicate runs, and a separate PCA analysis of microsatellite genotypes indicated that $K = 2$ captured much of the information available in the data (Fig. 2). Thus, while there is undoubtedly additional genetic substructure present within these *S. latifolia* populations, we proceeded with $K = 2$ as indicative of the major genetic structure consistent with this microsatellite data set.

Among native genotypes, the two genetic clusters were distributed principally between Eastern and Western EU (Figs 1a and 3). Genotypes from both clusters were introduced during invasion, but the relative frequency of the Eastern EU lineage differed significantly between continents (EU: 0.37, NA: 0.63; $G = 27.09$; $P < 0.0001$). Within NA, genotypes with Western EU ancestry occurred principally in southeastern and northwestern NA, while genotypes with Eastern EU ancestry occurred in all three NA regions, but were particularly dominant in frequency in central NA populations. Most genotypes assigned strongly to a single cluster (mean $Q = 0.949$), with the majority of genotypes showing little evidence of introgression between clusters in either North America (mean $Q = 0.938$) or Europe (mean $Q = 0.968$). However, among the 23 genotypes that did possess mixed ancestry ($0.25 \leq Q \leq 0.75$), 19 were from North America (Fig. 3). This indicates that admixture during invasion had increased opportunities for introgression, but that either hybrids are not yet frequent or that repeated backcrossing to a single ancestral cluster has swamped the signal of past introgression.

Cytoplasmic lineages were significantly ($\chi^2 = 262.84$; $P < 0.0001$) associated with different nuclear backgrounds (i.e. STRUCTURE cluster assignment), but the strength of the association had shifted during invasion (Cluster*Continent interaction: $\chi^2 = 12.49$; $P = 0.0004$). Specifically, while there is nearly complete disequilibrium between cpDNA haplotypes and nuclear clusters in *S. latifolia*'s native European range, in North America the Western EU cpDNA lineage has become more frequent on the Eastern EU nuclear background (Fig. 4), probably the result of admixture generating independent assortment of genomes following mating between Eastern and Western EU lineages.

To better understand the introduction history into North America, we used ABC to model 10 competing invasion scenarios, five of which returned non-null support (Fig. S3, Supporting information). Of these, the highest posterior probability was for scenario 3 (direct estimate = 0.4300 [0.0000, 0.8640; 95% credible interval]; logistic regression = 0.6382 [0.5361, 0.7403]), which described a history of multiple introductions with admixture into eastern NA, an independent set of multiple introductions with admixture into western NA and subsequent serial colonization from western to central NA (Fig. S1, Supporting information). The next most probable model was scenario 7, which was similar

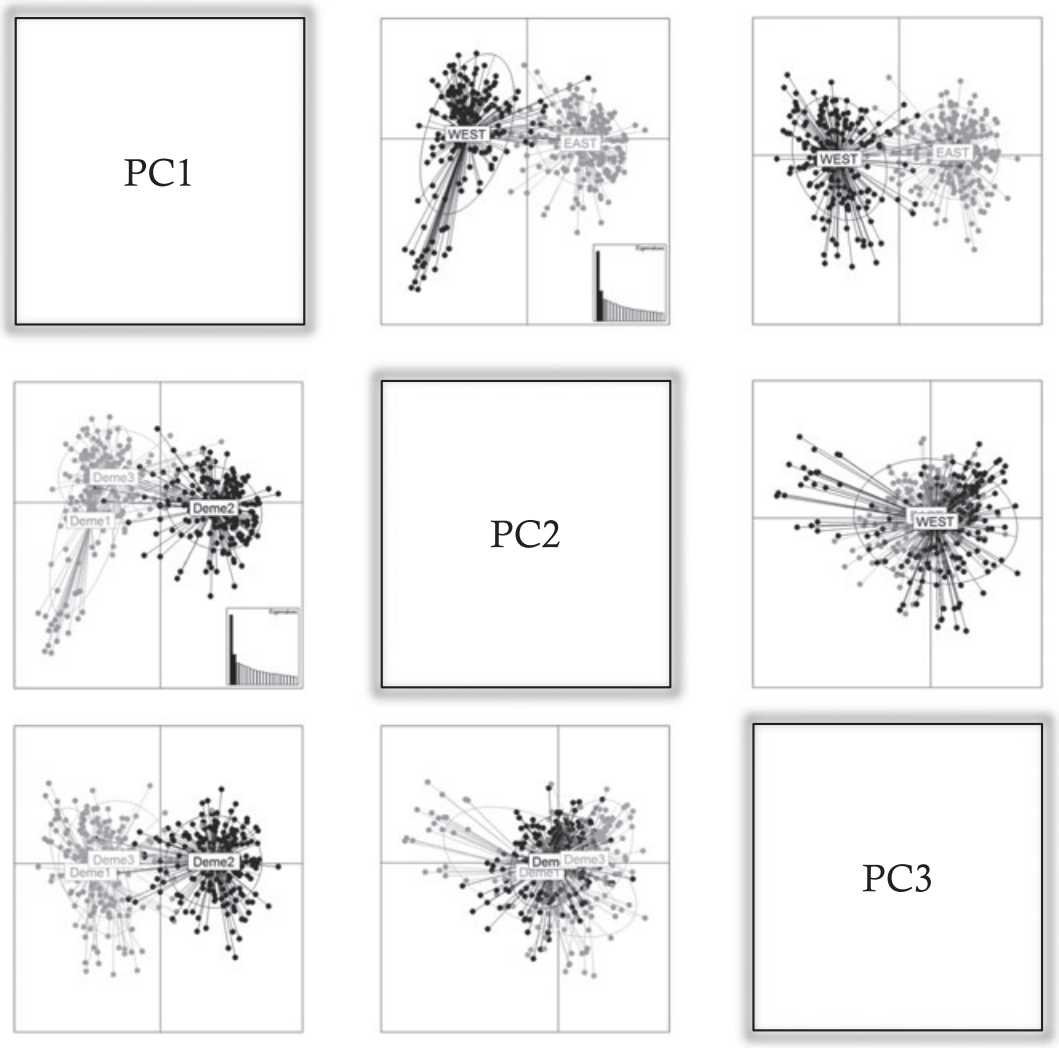

Fig. 2 Principal components analysis showing clustering of genotypes at 18 microsatellite loci. Genotype scores for PCs 1–3 are grouped based on STRUCTURE clustering results for $K = 2$ (above diagonal) and $K = 3$ (below diagonal).

to scenario 3 except that the source for the central NA region was an unsampled 'ghost' population branching off of the Eastern EU lineage. Model support for scenario 7 was slightly lower than for scenario 3 under the direct estimate (0.2240 [0.0000, 0.5894]), but near zero by logistic regression (Fig. S3, Supporting information). A-posteriori simulations to evaluate confidence in scenario choice under each of these two models reaffirmed the selection of scenario 3 over scenario 7 (Table 2). These simulations showed that when the true model was known, we were very unlikely (<13% of simulations) to select scenario 3 when in fact it was false; similarly, we were unlikely (<4% of simulations) to miss selecting scenario 7 if in fact it was true. In other words, based on the size of our sample (numbers of genotypes and loci), model complexity and the contrasting predictions these scenarios make for patterns of genetic variation, there is a high degree of confidence that scenario 3 fits our data better than any other model considered. Model checking analysis showed that the summary statistics from the observed data produced eigenvectors that were within (PC2 and PC3) or at the margins (PC1) of the set of simulated data sets from the posterior predictive distribution based on scenario 3 (Fig. S4, Supporting information).

The introduction history represented by scenario 3 revealed that the *Silene* invasion has been shaped by a combination of demographic bottlenecks and admixture following multiple introductions from Europe (Fig. 5; Table 3). Two of these introductions appear independent of each other: the first into eastern NA that admixed EU lineages at near equal frequencies (0.505) and a second into western NA where admixture was biased towards the Eastern EU lineage (0.865).

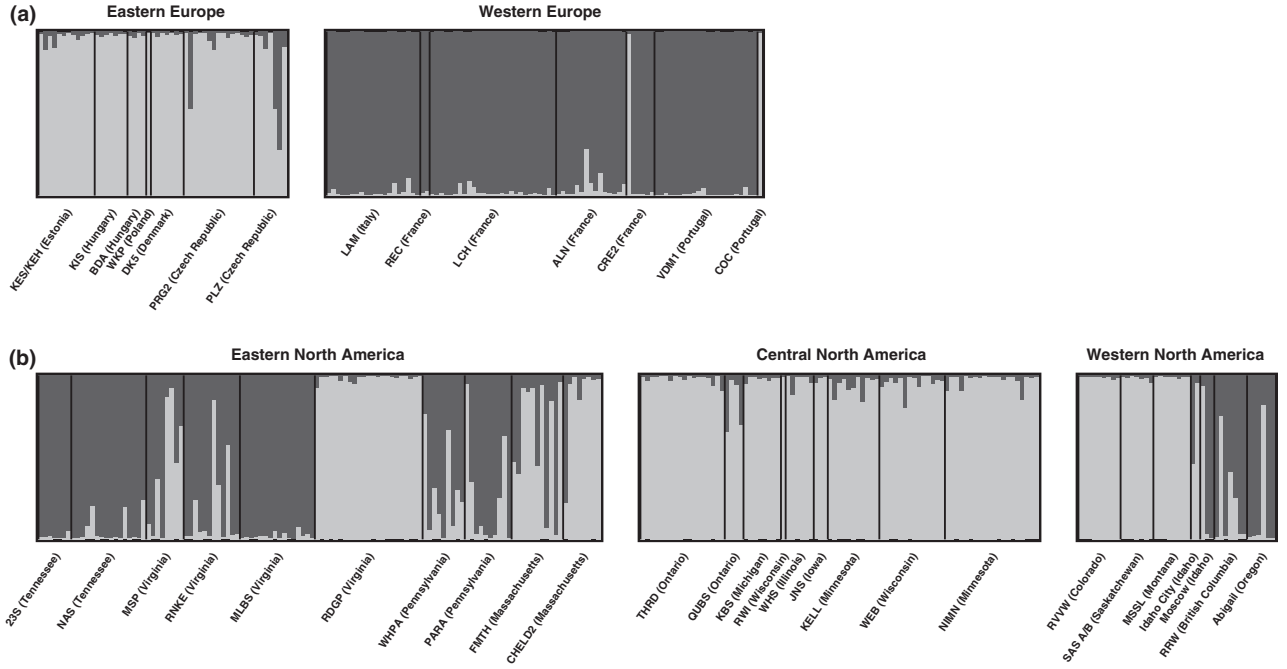

Fig. 3 Bayesian assignments to nuclear genetic clusters with $K = 2$. (a) European populations. (b) North American populations. Population sample locations are listed across the bottom, arranged geographically with light grey representing the Eastern EU cluster and dark grey the Western EU cluster.

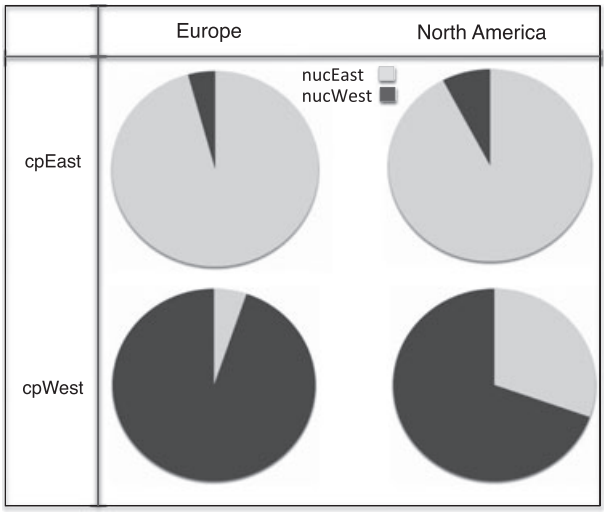

Fig. 4 Change in cytonuclear association of nuclear STRUCTURE clusters (nucEast, nucWest) with chloroplast lineages (cpEast, cpWest) from Europe to North America.

Interestingly, this bias towards Eastern EU became more extreme as founders from western NA formed a secondary invasion front that colonized central NA, an area that is currently represented almost exclusively by Eastern EU nuclear ancestry (Fig. 1b). All three NA regions showed highly reduced founding population sizes during initial establishment, although there was a large amount of uncertainty (Table 3) and in fact our estimates of founding N_e may even be biased upwards (Table S2, Supporting information). The composite estimate of effective population size which takes into account uncertainty due to the mutation rate ($\theta = 4\ N_e\mu$) suggested that the size of the founders relative to their ancestral sources was <2% for each NA region, indicating the possibility of a major demographic bottleneck during colonization. However, the durations of the bottlenecks were brief, and contemporary NA population sizes are 1.9–2.4 times larger on average than populations in Eastern and Western EU (Fig. 5; Table 3). Lastly, while the timing of isolation between Eastern and Western EU lineages that contributed to the invasion (8290 generations ago, 95% credible interval 5040–45 000) suggested divergence during post-Pleistocene range expansion or during the last glacial period while in isolated refugia, contemporary invasion events over the last 300 years were predicted with much less temporal certainty. This is not surprising given how recent these events were relative to the time for new mutations to accumulate.

Discussion

Biological invasions are complex range expansions that sample and re-distribute ancestral genetic diversity into new biogeographic regions. As with other forms of

Table 2 Calculated confidence in top two scenario choices (scenarios 3 and 7) described by both a false-positive rate and a false-negative rate for both the direct estimate and logistic regression. Calculations obtained from 500 simulated pseudo-observed data sets for the top five scenarios considered (scenarios 3, 4, 5, 7 and 10)

Confidence in Scenario Selection				
	Direct estimate		Logistic regression	
Scenario	False negatives	False positives	False negatives	False positives
3	0.252	0.039	0.164	0.129
7	0.010	0.000	0.034	0.019

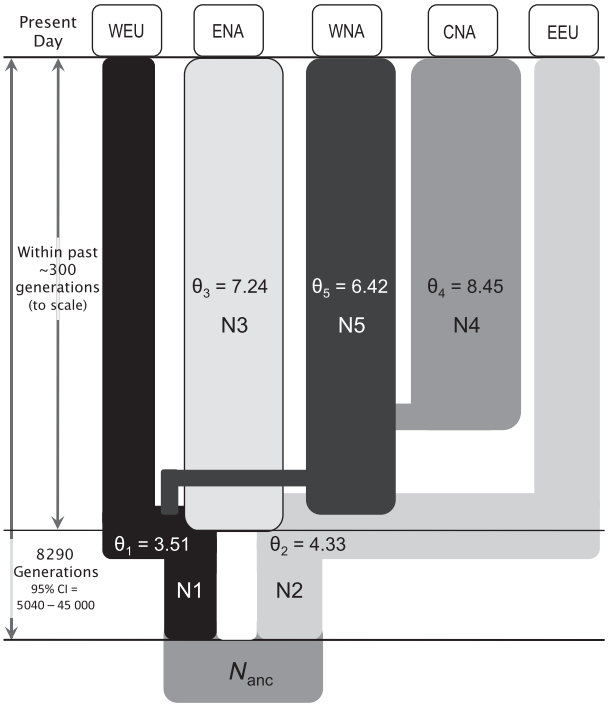

Fig. 5 Inferred scenario of introduction history (scenario 3) into North America. Width of population bars is proportional to effective size (N_e). Note that the timescale on the vertical axis is broken in order to visualize the history from postglacial divergence to contemporary invasion. Within the past 300 generations, the order and timing of population invasions is to scale.

colonization (e.g. postglacial expansion), invasions have tremendous potential to generate founder effects and bottleneck genetic diversity (Petit *et al.* 2005; Dlugosch & Parker 2008; Excoffier *et al.* 2009). However, the genetic consequences of human-facilitated invasions are also unique because the vectors of their introduction often sample propagules across multiple spatially isolated and genetically differentiated native sources, then combine those propagules into the same regions or populations after introduction. This admixture of invading propagules can lead to hybrid vigour (Facon *et al.* 2005; Keller & Taylor 2010; Verhoeven *et al.* 2011), genotypic novelty through recombination (Lavergne & Molofsky 2007; Facon *et al.* 2008), and increased genetic variance (Kolbe *et al.* 2004). Collectively then, there is the prediction that invasion may often be accompanied by evolutionary increases in fitness and adaptive evolution (Sakai *et al.* 2001; Lambrinos 2004; Prentis *et al.* 2008), which may in turn promote secondary invasions elsewhere (Lombaert *et al.* 2010). However, the fitness and adaptation of introduced populations will also be strongly shaped by features of the ancestral sources, such as the degree of pre-adaptation to the introduction site, or constraints of genetic architecture on the response to selection (Keller & Taylor 2008; Colautti *et al.* 2010; Calsbeek *et al.* 2011; Colautti & Barrett 2011). Thus, much of our understanding of the evolutionary trajectories of invasive populations and how they achieve (or are limited by) fitness in their new environment will critically depend on deciphering the sources of introduced propagules and the pathways of their expansion (Keller & Taylor 2008; Estoup & Guillemaud 2010).

Population structure and admixture

The results of our study demonstrate that the invasion of *Silene latifolia* into North America involved a complex sampling and admixture of nuclear and chloroplast diversity from two major ancestral genetic lineages in Western and Eastern EU. The population structure of these lineages was previously observed in phylogeographic analyses of chloroplast sequence (Taylor & Keller 2007) and nuclear AFLPs (Keller *et al.* 2009) and further supported here with nuclear microsatellite genotypes by Bayesian clustering analyses and principal components analysis. At a range-wide scale, the invasion has been biased in favour of the Eastern EU lineage, but Western EU ancestry can still be found scattered throughout the southeastern United States and the Pacific Northwest (Fig. 1). This invasion process has effectively collapsed the hierarchical genetic structure that had built up in the native range since the last postglacial expansion, and has generated a melting pot of genetic diversity within the introduced range.

While multiple introductions have clearly admixed diversity at the range-wide scale in North America, the effects of introgression and recombination between lineages are less obvious. The clearest signal of admixture can be observed in the dissociation of chloroplast and nuclear lineages, with the Western EU cpDNA haplotype introgressing into an Eastern European

Table 3 Posterior median estimates (95% credible interval) of parameter values from the most probable introduction history (scenario 3). Single parameter estimates are in units of effective population size (N_e) or time in generations (t and BD). Admixture coefficients (r) are expressed as proportions. Composite parameter estimates are scaled by the mutation rate (μ). These include the scaled population size ($\theta = N_e\mu$) and scaled coalescent time ($\tau = t\mu$)

Interpretation	Single parameter	Median (95% CI)	Composite parameter	Median (95% CI)
Western EU (WEU)	N1	1.42E+05 (2.69E+04 – 4.22E+05)	θ_1	3.51E+00 (1.56E+00 – 7.56E+00)
Eastern EU (EEU)	N2	1.73E+05 (2.83E+04 – 5.32E+05)	θ_2	4.33E+00 (1.39E+00 – 1.09E+01)
Eastern NA (ENA)	N3	3.03E+05 (3.73E+04 – 5.79E+05)	θ_3	7.24E+00 (8.20E−01 – 2.59E+01)
Central NA (CNA)	N4	3.61E+05 (3.30E+04 – 5.87E+05)	θ_4	8.45E+00 (7.30E−01 – 2.96E+01)
Western NA (WNA)	N5	2.64E+05 (1.92E+04 – 5.79E+05)	θ_5	6.42E+00 (4.8E−01 – 2.82E+01)
ENA founding size	N3F	2.42E+03 (4.02E+02 – 4.83E+03)	θ_{3F}	6.12E−02 (7.8E−03 – 2.19E−01)
CNA founding size	N4F	6.43E+02 (6.10E+01 – 4.17E+03)	θ_{4F}	1.69E−02 (1.30E−03 – 1.36E−01)
WNA founding size	N5F	2.76E+03 (4.84E+02 – 4.87E+03)	θ_{5F}	6.73E−02 (9.50E−03 – 2.48E−01)
WEU/EEU split	t_a	8.29E+03 (5.04E+03 – 4.50E+04)	τ_a	2.45E−01 (1.26E−01 – 8.61E−01)
ENA founding time	t_b	2.58E+02 (1.17E+02 – 2.98E+02)	τ_b	6.33E−03 (2.10E−03 – 1.76E−02)
WNA founding time	t_c	2.47E+02 (1.11E+02 – 2.94E+02)	τ_c	6.33E−03 (1.98E−03 – 1.61E−02)
CNA founding time	t_d	2.08E+02 (4.71E+01 – 2.87E+02)	τ_d	4.92E−03 (1.01E−03 – 1.41E−02)
ENA bottleneck duration	BDt_b	1.25E+00 (5.90E−01 – 8.09E+00)	τ_{BDtb}	3.48E−05 (8.80E−06 – 2.68E−04)
WNA bottleneck duration	BDt_c	1.91E+00 (7.00E−01 – 8.74E+00)	τ_{BDtc}	5.20E−05 (1.19E−05 – 3.41E−04)
CNA bottleneck duration	BDt_d	1.24E+00 (6.05E−01 – 8.19E+00)	τ_{BDtd}	3.46E−05 (9.20E−06 – 2.74E−04)
ENA admixture of W/E EU	$r_{3(2,1)}$	5.05E−01 (4.37E−01 – 5.65E−01)		
WNA admixture of W/E EU	$r_{5(2,1)}$	8.65E−01 (7.84E−01 – 9.23E−01)		

nuclear background. This points to segregation of genetic variance following mating and independent assortment between these historically isolated lineages in a direction that is consistent with assimilation of the Western EU lineage by the Eastern EU lineage as a consequence of the invasion process. Interestingly, the secondary invasion into central NA showed a strong bias towards Eastern EU ancestry in the nuclear but not the chloroplast data. Because seeds (but not pollen) are the propagules that establish new populations, this difference in diversity is not likely attributable to the different dispersal vectors of nuclear (pollen and seed) vs. chloroplast (seed only) genes. Rather, it seems plausible that the Eastern EU nuclear cluster may be spreading preferentially due to a selective advantage, while variation in chloroplast haplotype frequencies reflects a more neutral demography. Alternatively, genetic drift at the wavefront of the expansion (i.e. gene surfing; Excoffier *et al.* 2009) may have biased allele frequencies by chance in favour of the Eastern EU nuclear cluster.

Bayesian analysis of introduction history

Historical records of the invasion of *Silene latifolia* in North America all point to initial colonization in the northeast, but are vague as to the timing and source of invasion into other geographic regions. Because collection efforts in different regions were not well standardized by historical collectors and potentially subject to bias, a reliance solely on herbarium specimens provides little resolution to conclude the order of colonization events of different regions, if these colonization episodes were independent of one another, or what the likely direction of expansion was between regions.

This study highlights some of the advantages and inherent challenges of using ABC to explicitly model the demographic history of introduction events and obtain a more detailed view of the invasion processes and how it shaped genetic diversity. ABC modelling returned strong support for a minimum of four independent introduction events transporting propagules into North America—two into the east coast region (one each from Eastern and Western EU) that admixed the two EU nuclear clusters almost equally, and another two into the west coast that was biased towards the Eastern EU nuclear cluster. These introduction events were permeated by severe but brief demographic bottlenecks of the founding population sizes, averaging <2% of the N_e in each source population, followed by population expansion to nearly double that of the source population N_e.

Our genetic analysis of the demography of the *Silene* invasion confirmed the presence of multiple introductions and provided two important insights of general interest to population geneticists and invasion biologists. First, the twofold increase in N_e between the introduced and native regions provides an estimate, integrated across time and space, of the increase in abundance that accompanies colonization of a new range. Thus, genetic data may help to confirm

(or refute) a biogeographical prediction that invasive species should occur at higher abundance in their introduced vs. native ranges, a prediction that has so far been difficult to address robustly with ecological sampling (Hierro *et al.* 2005). However, the increase in N_e for invasive populations may also reflect a contribution from gene flow among NA regions (see 'Confidence in Scenario Choice and Model Fit' below); thus, the conclusion of increased N_e during invasion must remain tentative at present until a better understanding of the magnitude and effects of gene flow on N_e in NA *Silene* populations can be conducted.

Second, estimates of founding population sizes point to the resilience of introduced populations in the face of colonization bottlenecks. This is particularly interesting in context of the expected lag phase that is predicted to occur between initial establishment and population growth during invasion (Sakai *et al.* 2001). Our analysis indicated the absence of a significant lag phase despite small founder N_e during two separate establishment events. Population genetic theory suggests that when bottlenecks are brief (<10 generations) and populations grow rapidly, the majority of genetic diversity is maintained (Nei *et al.* 1975), which appears to be the case for this *Silene* invasion. However, relatively little is known about the demographic and genetic thresholds that govern the ability of small founder populations to establish and commence rapid growth, especially in species such as *S. latifolia* that are obligate outcrossers and therefore sensitive to demographic and genetic Allee effects (Elam *et al.* 2007; Barrett *et al.* 2008).

The best supported ABC model also suggested that the continental mid-west region surrounding the Great Lakes was not an independent introduction but rather a secondary invasion front derived from introduced populations in western NA. This suggests that the west coast introduction acted as a bridgehead for invasion eastward into the mid-west and the Great Lakes region. This perhaps appears counterintuitive, given the connectedness of the Great Lakes shipping pathway eastward to the Saint Lawrence Seaway and the Atlantic Ocean, but there are many potential vectors for the eastward spread of western agricultural weeds. For example, the Canadian-Pacific railway (constructed in the late 1800s) was a transportation route for grain from the western prairies to the Great Lakes for export through the seaway (National Research Council 2008) and likely served as a route for many weeds thought to have dispersed as contaminants of hay fields and crop seed (McNeill 1977).

Confidence in scenario choice and model fit

Our analysis of false-positive and false-negative rates supported our choice of scenario 3 as the leading model of invasion history and suggested that we had good power to choose scenario 3 when it was true and were unlikely to mistakenly choose scenario 3 when it was false (Table 2). Thus, we have a high degree of confidence that, given the competing models, scenario 3 is the best estimate of *Silene latifolia*'s invasion history. However, while scenario 3 received the greatest posterior support, model checking results based on simulations from the posterior predictive distribution showed that our observed data were at the margins of the posterior distribution based on the first PC of summary statistics (Fig. S4, Supporting information). Thus, our observed data were not perfectly predicted by the estimated parameters of scenario 3, which urges caution against interpreting the demographic parameter estimates too closely.

Interestingly, the observed summary statistics that most consistently deviated from the posterior predictive distribution were population pairwise F_{ST}s (Table S3, Supporting information). This may stem from the absence of a gene flow parameter in our models, as the current implementation of DIYABC does not allow for migration among populations after splitting events. Because our regions are not geographically discreet, gene flow is probably occurring among populations from different regions and may lessen the model fit to the observed data. Failing to account for gene flow between neighbouring regions could affect estimates of N_e, biasing N_e upwards where the relative input from outside alleles to resident polymorphism is high. This effect could be contributing to the higher bias and lower precision observed for N_e in North America regions, where gene flow is probably more prevalent (Table S2, Supporting information).

Implications of introduction history for rapid evolution during invasion

The invasion of *S. latifolia* into North America is one of the best studied examples of an introduced plant evolving life history traits that increase early population growth at the expense of defences against natural enemies (Wolfe 2002; Blair & Wolfe 2004; Wolfe *et al.* 2004, 2007; Keller & Taylor 2008; Bernasconi *et al.* 2009; Keller *et al.* 2009; Verhoeven *et al.* 2011). How does our analysis of the introduction history inform our understanding of these evolutionary changes in traits? First, the contribution of F1 heterosis to increased fitness in *S. latifolia* has been investigated by Wolfe *et al.* (2007) (F1s from interpopulation crosses without regard to region) and Hathaway *et al.* (2009) (F1s from interpopulation crosses between Eastern and Western EU regions). Extrapolating their results with our knowledge of the phylogeographic structure in North America and Europe,

there does not seem to be strong heterosis between the Eastern and Western EU lineages, despite thousands of generations of divergence. Thus, heterosis following admixture of historically isolated lineages does not seem to contribute to invasiveness in *S. latifolia*, unlike its congener *S. vulgaris* which shows similar genetic structure and increased fitness following admixture in North America (Keller & Taylor 2010).

Second, Eastern and Western EU lineages occur in nearly equal frequency in eastern NA, indicating differential pre-adaptation has probably not contributed to each lineage's proliferation within this particular region of introduction. However, the increasing eastern European ancestry as the invasion spread from the west coast introduction into central NA could indicate a role for pre-adaptation to the dry, continental interior climate (or some other set of environmental variables). This hypothesis could be tested experimentally against the null hypothesis of genetic drift by comparing the fitness of genotypes with known EU ancestry when grown in field common gardens (sensu Keller & Taylor 2008, 2010). Thus, knowing the sources and direction of expansion of each invasion front provides critical information for testing competing hypotheses (adaptive vs. nonadaptive) of range expansion and invasion success (Keller & Taylor 2008; Estoup & Guillemaud 2010).

Conclusion

Our population genetic analysis provided several novel insights into the introduction history of *Silene latifolia*, a well-studied invasive plant. The combination of multilocus nuclear genotypes and cytonuclear associations enabled us to detect multiple introductions admixing diversity at both regional and to a lesser extent local population scales. The addition of ABC allowed us to reconstruct a more complex narrative of the demographic history that shaped genetic diversity during the invasion. We showed that (i) a minimum of four introductions from Europe took place that differed in genetic composition, (ii) Central NA populations are the result of a secondary invasion front from western NA and (iii) rapid population growth immediately following the invasion likely alleviated the severity of the demographic bottleneck on genetic diversity. These insights will prove instrumental in designing future experiments aimed at identifying the evolutionary causes of increased fitness acquired during the spread of invasive populations across new environments.

Acknowledgements

We thank Armanda Roco for assistance with genotyping. Matt Fitzpatrick, AE Remy Petit and three anonymous reviewers provided helpful comments that improved the manuscript. We also appreciate the efforts of the many colleagues who have contributed to the *Silene* seed collection over the years. This research was supported by NSF DEB-0349558 and DEB-0919335.

References

Abdoullaye D, Acevedo I, Adebayo AA *et al.* (2010) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 August 2009-30 September 2009. *Molecular Ecology Resources*, **10**, 232–236.

Antonovics J, Hood ME, Thrall PH, Abrams JY, Duthie GM (2003) Herbarium studies of the distribution of anther-smut fungus (*Microbotryum violaceum*) and *Silene* species (Caryophyllaceae) in the eastern United States. *American Journal of Botany*, **90**, 1522–1531.

Barrett SCH, Colautti RI, Eckert CG (2008) Plant reproductive systems and evolution during invasion. *Molecular Ecology*, **17**, 373–383.

Beaumont M, Zhang W, Balding D (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bernasconi GJ, Antonovics J, Biere A *et al.* (2009) Silene as a model system in ecology and evolution. *Heredity*, **103**, 5–14.

Blair AC, Wolfe LM (2004) The evolution of an invasive plant: an experimental study with *Silene latifolia*. *Ecology*, **85**, 3035–3042.

Calsbeek BS, Lavergne S, Patel M, Molofsky J (2011) Comparing the genetic architecture and potential response to selection of invasive and native populations of reed canary grass. *Evolutionary Applications*, **4**, 726–735.

Colautti RI, Barrett SCH (2011) Population divergence along lines of genetic variance and covariance in the invasive plant *Lythrum salicaria* in eastern North America. *Evolution*, **65**, 2514–2529.

Colautti RI, Eckert CG, Barrett SCH (2010) Evolutionary constraints on adaptive evolution during range expansion in an invasive plant. *Proceedings. Biological Sciences/The Royal Society*, **277**, 1799–1806.

Cornuet J-M, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Cornuet J-M, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431–449.

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, **19**, 1349.

Elam DR, Ridley CE, Goodell K, Ellstrand EC (2007) Population size and relatedness affect fitness of a self-incompatible invasive plant. *Proceedings of the National Academy of Sciences USA*, **104**, 549–552.

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences USA*, **97**, 7043–7050.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113–4130.

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad, *Bufo marinus*. *Evolution*, **58**, 2021–2036.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics*, **40**, 481–501.

Facon B, Jarne P, Pointier JP, David P (2005) Hybridization and invasiveness in the freshwater snail *Melanoides tuberculata*: hybrid vigour is more important than increase in genetic variance. *Journal of Evolutionary Biology*, **18**, 524–535.

Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Current Biology*, **18**, 363–367.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Gaudeul MT, Giraud T, Kiss L, Shykoff JA (2011) Nuclear and chloroplast microsatellites show multiple introductions in the worldwide invasion history of common Ragweed, *Ambrosia artemisiifolia*. *PLoS One*, **6**, e17658.

Genton BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*, **14**, 4275–4285.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.

Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.

Hathaway L, Andersson S, Prentic HC (2009) Experimental crosses within European Silene latifolia (Caryophyllaceae): intraspecific differentiation, distance effects, and sex ratio. *Botany*, **87**, 231–240.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hickerson MJ, Stahl EA, Lessios HA (2006) Test for simultaneous divergence using approximate Bayesian computation. *Evolution*, **60**, 2435–2453.

Hierro JL, Maron JM, Callaway RM (2005) A biogeographic approach to plant invasions: the importance of studying exotics in their introduced *and* native range. *Journal of Ecology*, **93**, 5–15.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Juillet N, Freymond H, Degen L, Goudet J (2003) Isolation and characterization of highly polymorphic microsatellite loci in the bladder campion, *Silene vulgaris* (Caryophyllaceae). *Molecular Ecology Resources*, **3**, 358–359.

Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters*, **11**, 852–866.

Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology*, **23**, 1720–1731.

Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR (2009) Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist*, **183**, 678–690.

Knowles LL (2009) Statistical Phylogeography. *Annual Review of Ecology Evolution and Systematic*, **40**, 593–612.

Kolbe JJ, Glor RE, Schettino LRG, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 177–181.

Lambrinos JG (2004) How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecology*, **85**, 2061–2070.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences USA*, **104**, 3883–3888.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology and Evolution*, **17**, 386–391.

Lombaert E, Guillemaud T, Cornuet J-M, Malausa T, Facon B, Estoup A (2010) Bridgehead effect in the worldwide invasion of the biocontrol Harlequin Ladybird. *PLoS One*, **5**, e9743.

Lombaert E, Guillemaud T, Thomas CE *et al.* (2011) Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird *Harmonia axyridis*. *Molecular Ecology*, **20**, 4654–4670.

Martindale IC (1876) The introduction of foreign plants. *Botanical Gazette*, **2**, 55–58.

Matschiner M, Salzburger W (2009) TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics*, **25**, 1982–1983.

McNeill J (1977) The biology of Canadian weeds. 25. Silene alba (Miller) E. H. L. Krause. *Canadian Journal of Plant Science*, **57**, 1103–1114.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Moccia MD, Oger-Desfeux C, Marais GA, Widmer A (2009) A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *BMC Genomics*, **10**, 243.

National Research Council(2008) Great Lakes Shipping, Trade, and Aquatic Invasive Species: Special Report 291. The National Academies Press, Washington, DC.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution*, **29**, 1–10.

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. *Molecular Ecology*, **18**, 1034–1047.

Petit RJ, Aguinagalde I, De Beaulieu JL *et al.* (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, **13**, 288–294.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.
Sakai AK, Allendorf FW, Holt JS *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics*, **32**, 305–332.
Schachner LJ, Mack RN, Novak SJ (2008) *Bromus tectorum* (Poaceae) in midcontinental United States: population genetic analysis of an ongoing invasion. *American Journal of Botany*, **95**, 1584–1595.
Slotta TAB, Brady L, Chao S (2008) High throughput tissue preparation for large-scale genotyping experiments. *Molecular Ecology Resources*, **8**, 83–87.
Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, **61**, 334–345.
Teixeira S, Bernasconi G (2007) High prevalence of multiple paternity within fruits in natural populations of *Silene latifolia*, as revealed by microsatellite DNA analysis. *Molecular Ecology*, **16**, 4370–4379.
Thioulouse J, Dray S (2007) Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. *Journal of Statistical Software*, **22**, 1–14.
Verhoeven KJF, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings. Biological Sciences/The Royal Society*, **278**, 2–8.
Wilson JRU, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution*, **24**, 136–144.
Wolfe LM (2002) Why alien invaders succeed: support for the escape-from-enemy hypothesis. *American Naturalist*, **160**, 705–711.
Wolfe LM, Elzinga JA, Biere A (2004) Increased susceptibility to enemies following introduction in the invasive plant *Silene latifolia*. *Ecology Letters*, **7**, 813–820.
Wolfe LM, Blair AC, Penna BM (2007) Does intraspecific hybridization contribute to the evolution of invasiveness?: an experimental test. *Biological Invasions*, **9**, 515–521.

S.K. is a population geneticist with interests in the feedbacks between genetic diversity and environmental change. K.G. is interested in genetic diversity and population genetic structure across species' ranges. P.F. is interested in the development and application of statistical genetic methods for the inference of ecological and evolutionary processes. D.T. studies molecular population genetics and organelle genome evolution, mostly in plants.

Data accessibility

Microsatellite and chloroplast genotype data, and DIYABC input files: DRYAD entry doi: 10.5061/dryad.9r2h3.

Supporting information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Graphical representation of the 10 competing scenarios used in the ABC analysis.

Fig. S2 Delta-*K* across STRUCTURE runs supporting $K = 2$ (Evanno *et al.* 2005).

Fig. S3 Estimated likelihood of five best competing scenarios from both a direct estimate using the 500 closest datasets and a logistic regression using 1% of closest datasets.

Fig. S4 Model checking results showing the first three axes of a PCA on summary statistics from the 10 simulated demographic scenarios, simulations from the posterior predictive distribution (dark blue), and the observed data (yellow).

Table S1 Population genetic diversity statistics for each geographic region.

Table S2 Bias and precision of parameter estimates from ABC for scenario 3.

Table S3 Deviations of summary statistics for the observed data form the posterior predictive distribution of the most probable invasion history (Scenario 3).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.