

## Recent admixture generates heterozygosity–fitness correlations during the range expansion of an invading species

S. R. KELLER\*, P. D. FIELDS†, A. E. BERARDI† & D. R. TAYLOR†

\*Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD, USA

†Department of Biology, University of Virginia, Charlottesville, VA, USA

### Keywords:

heterosis;  
hybridization;  
introgression;  
invasion;  
*Silene*.

### Abstract

Admixture, the mixing of historically isolated gene pools, can have immediate consequences for the genetic architecture of fitness traits. Admixture may be especially important for newly colonized populations, such as during range expansion and species invasions, by generating heterozygosity that can boost fitness through heterosis. Despite widespread evidence for admixture during species invasions, few studies have examined the demographic history leading to admixture, how admixture affects the heterozygosity and fitness of invasive genotypes, and whether such fitness effects are maintained through time. We address these questions using the invasive plant *Silene vulgaris*, which shows evidence of admixture in both its native Europe and in North America where it has invaded. Using multilocus genotype data in conjunction with approximate Bayesian computation analysis of demographic history, we showed that admixture during the invasion of North America was independent from and much younger than admixture in the native range of Europe. We tested for fitness consequences of admixture in each range and detected a significant positive heterozygosity–fitness correlation (HFC) in North America; in contrast, no HFC was present in Europe. The lack of HFC in Europe may reflect the longer time since admixture in the native range, dissipating associations between heterozygosity at markers and fitness loci. Our results support a key short-term role for admixture during the early stages of invasion by generating HFCs that carry populations past the threat of extinction from inbreeding and demographic stochasticity.

### Introduction

Genetic and evolutionary processes occurring over short time scales are important components of many successful species invasions (Ellstrand & Schierenbeck, 2000; Reznick & Ghalambor, 2001; Lee, 2002; Lambrinos, 2004; Salamin *et al.*, 2010). One of the defining genetic characteristics of contemporary invasions is the occurrence of genetic admixture, typically the result of multiple introductions and introgression among diverse genotypes descended from structured populations in the native range (Kolbe *et al.*, 2004; Dlugosch & Parker, 2008; Hufbauer, 2008; Keller *et al.*, 2012). By mixing

together historically isolated gene pools, admixture during the invasion process can fuel rapid evolutionary changes, including a change in the genetic architecture and variance of traits important to fitness (Calsbeek *et al.*, 2011; Colautti & Barrett, 2011).

An important change in the genetic architecture of fitness traits that is likely to arise from admixture is heterosis, or hybrid vigour, an increase in performance generally attributed to the masking of recessive deleterious alleles in the heterozygous state (Lynch & Walsh, 1998). The positive contribution of heterosis to fitness could be of particular importance to the growth rates of small, recently founded populations. Thus, contemporary human-mediated invasions are a prime example of where we would expect genetic admixture and heterosis to fuel range expansion (Facon *et al.*, 2005, 2011; Excoffier *et al.*, 2009; Keller & Taylor, 2010; Turgeon *et al.*, 2011). However, because the positive fitness

Correspondence: S. R. Keller, Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Rd., Frostburg, MD 21532, USA. Tel.: +1-301-689-7203; fax: +1-301-689-7200; e-mail: skeller@umces.edu

effects of heterosis may be diminished in subsequent generations by segregation and recombination (Lynch, 1991; Lynch & Walsh, 1998), the impact of heterosis is likely to be a short-term catalyst for population growth (e.g. Drake, 2006).

If admixture during invasion boosts fitness by masking deleterious alleles that were fixed or near fixation in the source populations, then this should be evident as a positive heterozygosity–fitness correlation (HFC; David, 1998) within the invaded range. General effect HFCs occur because of genome-wide identity disequilibrium, that is, the correlation in *genotypic* state (heterozygosity and/or homozygosity) among loci, as opposed to the correlation of *allelic* state among loci that causes linkage disequilibrium (Szulkin *et al.*, 2010). Identity disequilibrium can thus result in the heterozygosity of marker loci, even ones that are selectively neutral, being indicative of the overall heterozygosity of the genome and therefore statistically associated with heterozygosity at fitness loci (e.g. QTL) (Hansson & Westberg, 2002; Szulkin *et al.*, 2010). If fitness loci experience either direct dominance or overdominance (i.e. fitness increases when deleterious alleles are sheltered), then identity disequilibrium will produce HFCs observable with marker loci when a structured set of populations becomes admixed, such as during species invasion or range expansion.

In the invasive weed *Silene vulgaris*, signals of admixture are evident in the native European range where introgression between different nuclear genetic lineages is apparent in Central Europe (Keller *et al.*, 2009; Keller & Taylor, 2010). This region is a well-studied suture zone between lineages of many plants and animal species expanding from different Mediterranean glacial refugia (Hewitt, 1996; Petit *et al.*, 2003) and may therefore reflect an admixture event during the northward expansion of *Silene* lineages following Pleistocene glacial retreat, but it is also possible that European admixture has happened more recently as a result of human-mediated dispersal. Admixture is also prevalent between these same lineages in North America, where *S. vulgaris* has been invading since its initial introduction into eastern port cities in the early 1800s (Pursh, 1814; Martindale, 1876). However, it is not known whether admixed North American genotypes are descended directly from admixed European genotypes or whether instead the invasion of North America involved novel bouts of admixture, and whether this contributes to the fitness of invasive genotypes. Keller and Taylor (2010) found an association between fruit production measured under common garden conditions and the degree of genome admixture, measured as heterogeneity in the ancestry coefficients from a Bayesian clustering analysis on AFLP loci. However, because AFLPs are dominant markers, it was not possible to test whether ‘admixture–fitness correlations’ (AFCs) were attributable to an association between fitness and genome-wide heterozygosity *per se*.

Further, the AFC observed was only present among invasive North American genotypes and not within the zone of admixture in Europe. This result is consistent with European admixture having occurred more anciently, allowing time for associations between fitness and markers to dissipate, but AFLPs do not allow for dating the relative ages of different admixture events using population genetic models of demographic history.

Here, we provide the first test of the hypothesis that recent admixture during species invasions increases fitness by generating genome-wide heterozygosity that shelters the genetic load of deleterious recessive alleles. We used the *Silene vulgaris* invasion of North America to generate two predictions regarding the demographic history of admixture events in each range, and the effects of admixture on the genetic architecture of fitness. First, we predicted that admixed genotypes in the invasive range arose recently (post-colonization) and are independent from the admixed genotypes in the native range. We tested this prediction using genotypic data for 15 microsatellite loci and approximate Bayesian computation (ABC) models of population demography to determine the relative timing and independence of admixture events within North America and Europe. Second, we predicted that positive HFCs should exist in the invaded range due to admixed genotypes experiencing identity disequilibrium between typed marker loci and unobserved fitness QTL. We tested this second prediction using common garden measurements of fitness paired with the microsatellite genotypes to test for the presence of HFCs in North America and Europe.

## Materials and methods

### Population samples and genotyping

Samples of *S. vulgaris* were collected from populations spanning much of the distributional range of both Europe (EU; native) and North America (NA; introduced) and were collected as seeds from maternal families of field grown plants or as leaf tissue dried on silica gel (Keller & Taylor, 2010). Genomic DNA was extracted from leaf tissue using Qiagen DNeasy plant mini kits. We attempted to genotype the same individuals used in the AFLP analysis of Keller and Taylor (2010) for 15 codominant microsatellite markers. These included five microsatellite markers derived from *S. vulgaris* (A2, A5, A11, B29 and G3) described by Juillet *et al.* (2003) and ten microsatellite markers derived from *S. latifolia* (SL\_eSSR04, SL\_eSSR06, SL\_eSSR09, SL\_eSSR12, SL\_eSS16, SL\_eSSR20, SL\_eSSR29, SL\_eSSR30) described by Moccia *et al.* (2009). The latter set of markers were checked for cross-species amplification in Moccia *et al.* (2009) and described within their Table 4.

Microsatellite amplifications were carried out either in single or multiplex reactions (depending on the

locus) using the Qiagen Multiplex PCR kit (Qiagen, Valencia, CA, USA). The fluorescently labelled PCR products were combined either in pools of different loci run from singleplex PCRs or 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 (Life Technologies, Grand Island, NY, USA). Samples were then separated out on an ABI 3130xl automated sequencer. Genotype scores were determined with GENEMAPPER v3.0 software (Life Technologies) using automated scoring and manual double checking and then binned using the program TANDEM (Matschiner & Salzburger, 2009).

### Common garden field experiment

The common garden measurements of fitness are the same as those reported in Keller and Taylor (2010). Plants were grown at two common garden sites in North America ('ON' in Ontario, Canada: N 45.8642, W -79.4362; and 'VA' in Virginia, USA: N 37.8577, W -78.8208). The design consisted of four replicate seed offspring from 100 open-pollinated maternal families per range (EU and NA) grown at each site (ON and VA) (total  $N = 1600$ ). Weekly censuses of all plants were conducted for two full growing seasons. During each census, plants were checked for the number of newly produced mature fruits. Fitness was estimated as the sum of fruit production across years per plant, which integrates both individual survivorship and fecundity. Further details concerning the description of the common garden experiment can be found in Keller and Taylor (2010). For the present study, we were able to extract DNA and genotype one randomly chosen individual from 163 of the 200 families planted (74 in EU and 89 in NA) using the 15 microsatellite loci described above.

### Statistical analyses

#### *Genetic structure*

To assess the structure of allele frequency variation in the microsatellite data set, 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 correlated allele frequencies and with all remaining parameters set to the developer's default values. We performed 10 independent runs for each  $K$  (1–10) with 1 000 000 MCMC iterations after a burn-in period of 200 000 iterations. The number of clusters ( $K$ ) was chosen based on the method of Evanno *et al.* (2005) as implemented in the software STRUCTURE HARVESTER (Earl & vonHoldt, 2012). Ancestry coefficients ( $Q$ -values) across runs were combined using CLUMPP (Jakobsson & Rosenberg, 2007) and visualized using *distruct* (Rosenberg, 2004).

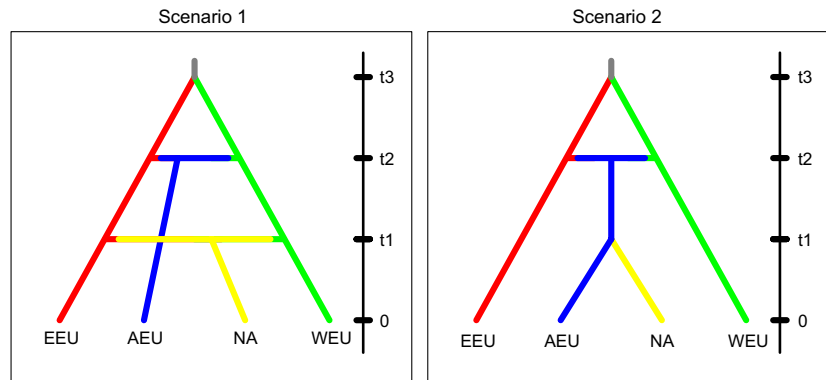
#### *Approximate Bayesian computation modelling of admixture history*

To infer the order and timing of admixture events within the native (EU) and invaded ranges (NA) of *S. vulgaris*, we performed ABC analyses of admixture history based on applying coalescent theory to the microsatellite data. All coalescent simulations were performed using DIYABC version 1.0.4.46 (Cornuet *et al.*, 2008) run on a Linux cluster via the University of Virginia cross-campus computing grid, whereas the ABC calculations were conducted on a Windows 7 machine running the same version.

For the ABC analysis, we organized samples into the following *a priori* defined groups: (1) pure Eastern EU, (2) pure Western EU, (3) admixed EU and (4) NA. Eastern and Western EU groups reflect regional genetic structure in *S. vulgaris* (Keller & Taylor, 2010). The EU admixed group was defined based on the current STRUCTURE analysis of  $Q$ -values, with genotypes having  $0.25 \leq Q \leq 0.75$  assigned to the EU admixed group. All North American genotypes were assigned to the NA group. These groupings were used to evaluate two competing scenarios that gave rise to the admixture observed in both the native and introduced ranges. Modelled scenarios focused on the splitting of *S. vulgaris*' native range into Eastern and Western EU populations, and the subsequent timing and magnitude and origin of the admixed populations in EU and NA (Fig. 1). Scenario 1 described the origin of an admixed European population (AEU) that resulted from introgression between the two ancestral European populations (EEU and WEU). A North American (NA) population then arose as the result of an independent admixture event that occurred between EEU and WEU during the invasion of North America. Scenario 2 is distinguished from Scenario 1 in that the NA population derives directly from the admixed European group (AEU) and thus does not represent an independent admixture event. Most importantly, these two scenarios allow us to determine whether the admixture observed within the European native range of *S. vulgaris* and that of the invaded range have independent demographic histories, and whether the European admixture could be more ancient (Keller & Taylor, 2010). Although it is possible to define many plausible scenarios to describe the history of *S. vulgaris*, these two particular scenarios were formulated in order to specifically address our objective of determining the timing and independence of admixture events in each range.

We set uniform priors on all demographic parameters (Table S1). The upper and lower bounds for priors were adjusted to encompass the high-probability regions from the posterior distribution based on exploratory models and were consistent with our previous research on the invasion history of *S. vulgaris* (Taylor & Keller, 2007; Keller & Taylor, 2010). Effective population sizes were also given a uniform prior [1000–1 000 000] to

**Fig. 1** Two scenarios describing the historical demography of *Silene vulgaris*. EEU = Eastern Europe group; WEU = Western Europe group; AEU = admixed EU, between WEU and EEU; and NA = North American group. Timing of demographic events is indicated along the vertical axis and corresponds to Tables 1 and 2.



accommodate a potentially wide variance in  $N_e$  among groups. The divergence between East and West European groups probably occurred as a result of glacial divergence or during post-glacial expansion (Taylor & Keller, 2007; Keller & Taylor, 2010), so we set this uniform prior as [5000–500 000] which allowed for divergence as early as two glacial cycles ago (Riss glacial period, ca. 200 000 years before present) and as recently as during the Holocene post-glacial expansion. We modelled the temporal origin of the admixed EU group with a uniform prior [1–100 000] and the origin of the admixed NA population with uniform prior [1–500]. This time range allowed for admixture among EU groups to have happened as early as during Pleistocene glacial cycles or as recently as during contemporary time; all *S. vulgaris* in NA result from contemporary (post-Columbian) colonization. Mutation rate parameters were modelled as both mean estimates across loci (uniform prior) and individual locus estimates (gamma prior) and were based on a stepwise mutation model that allowed for deviations from perfect repeats by single nucleotide insertions or deletions (Table S1).

Models for each scenario were simulated based on neutral coalescence for  $1 \times 10^6$  iterations. A rejection step was then performed to keep only the 0.1% of simulations that most closely matched the observed data based on four single-population summary statistics (mean number of alleles, mean expected heterozygosity, mean allele size variance and mean ratio of number of alleles over the range in allele sizes) and six pairwise-population summary 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 54.

We performed ABC 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 generating

summary statistics that most closely matched the observed data, and a logistic regression estimate, based on predicting the probability of a model from the deviations in the summary statistics among the 1% (50 000) closest simulated and observed data sets (Cornuet *et al.*, 2008). In addition to the standard comparisons of summary statistics as part of an ABC analysis, we also included a replicate analysis in which we have included a linear discriminant analysis (LDA) step for summary statistics in order to aid efficient ABC posterior scenario probability estimation (Estoup *et al.*, 2012).

To assess confidence in selection of the most probable model, we performed *a posteriori* simulations using each model's parameter estimates and calculated the probability of false-positive and false-negative errors from 1000 pseudo-observed data sets. We estimated the false-positive rate of a particular scenario to be the proportion of times that model selection resulted in choosing that scenario when in fact the data were simulated under the 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). The LDA step was used as part of our confidence in model choice procedure as well.

#### Multilocus heterozygosity and HFCs

To estimate multilocus heterozygosity for each individual, we calculated three different measures of multilocus heterozygosity using the Rhh package (Alho *et al.*, 2010) in R version 2.14.2 (R Development Core Team, 2011). The three measures consisted of homozygosity by loci (HL; Aparicio *et al.*, 2006), internal relatedness (IR; Amos *et al.*, 2001) and standardized heterozygosity

(SH; Coltman *et al.*, 1999). These measures provide complimentary estimators of identity disequilibrium from a set of marker loci. We found them all to be highly correlated and so we only report results for SH (see Results).

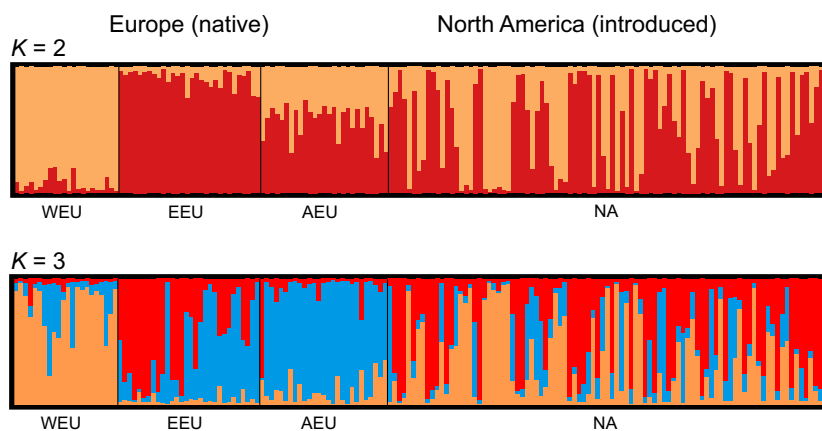
We tested for the presence of HFCs in the global data set using Gaussian linear models fit with the *lm* function in R (R Development Core Team, 2011). Family means of log (fruit production) were predicted by SH, continent of origin and their interaction. Common garden site (ON or VA) was also included as a fixed effect to control for environmental effects on fruit set. To determine whether the presence or magnitude of HFCs varied among different groups, we tested for HFCs using the following partitions of the global data set: (1) All EU individuals; (2) Just admixed EU individuals (based on STRUCTURE ancestry of  $0.25 \leq Q \leq 0.75$ ); (3) Just admixed EU individuals (based on geographical location within a suture zone in Central Europe; Fig. S2); (4) All NA individuals; (5) Just admixed NA individuals (based on STRUCTURE ancestry of  $0.25 \leq Q \leq 0.75$ ); and (6) Just nonadmixed NA individuals (based on STRUCTURE ancestry of  $Q \leq 0.25$  or  $Q \geq 0.75$ ). We included the last group (6) as a sort of negative control, where the lack of genotypic admixture predicts no HFC. To account for multiple testing of the HFC hypothesis, we adjusted our significance threshold to  $\alpha = 0.05/6 = 0.008$ .

## Results

Evidence of evolutionary divergence among nuclear genomes, as estimated by STRUCTURE Bayesian clustering of microsatellite genotypes, described an optimal model of  $K = 3$  clusters based on the  $\Delta K$  method (Fig. S1) (Evanno *et al.*, 2005). These clusters corresponded to a Western EU phylogeographical group (cluster 1) and two additional phylogeographical groups distributed primarily in Central and Eastern Europe and around the Mediterranean Sea (clusters 2

and 3) (Fig. S2). STRUCTURE's estimate of allele frequency divergence was greatest between cluster 1 vs. 2 (0.08) and 3 (0.06), with lesser divergence between clusters 2 and 3 (0.04), indicating that the major axis of divergence lies between clusters 1 vs. 2. A large number of cluster 3 genotypes in EU showed intermediate ancestry coefficients at  $K = 2$ , indicating that this cluster is associated with admixture between WEU and EEU lineages (Fig. 2). In order to simplify ABC demographic modelling of admixture history (see below), we therefore proceeded with analysis using groupings based on the STRUCTURE  $K = 2$  model, delineating Western Europe (WEU, corresponding to cluster 1), Eastern EU (EEU, corresponding to cluster 2) and admixed Europe (AEU, corresponding to admixed ( $0.25 \leq Q \leq 0.75$ ) EU genotypes). Genetic diversity estimated as the number of alleles and expected heterozygosity was generally higher in WEU and lower in EEU, whereas the admixed AEU and NA groups contained nearly the same or higher diversity than WEU (Table S2). All groups showed similar levels of excess homozygosity ( $G_{IS} \sim 0.50$ ), likely indicating population substructure and/or local inbreeding within groups. Divergence among groups ( $F_{ST}$ ) varied from 0.030 to 0.097 and was greatest between WEU and EEU (Table S3).

We modelled the demographic history that gave rise to admixture in the native range (AEU) and invasive range (NA) in order to disentangle the timing and independence of these events. Of the modelled scenarios, Scenario 1 had the highest posterior probability [direct estimate = 0.7060 (0.3067, 1.0000; 95% credible interval); logistic regression suggested zero support for Scenario 2; Fig. S3]. Strong support of Scenario 1 indicates the AEU and NA admixtures were unique, temporally separated events, and thus rejects the hypothesis that NA genotypes are descended from admixture that occurred previously in the native range (Fig. 1). Posterior estimates of the composite timing parameters indicated that European admixture occurred approximately



**Fig. 2** Ancestry assignment from STRUCTURE for models  $K = 2$  and 3. Labels along the horizontal axis indicate phylogeographical groupings used in the approximate Bayesian computation (ABC) modelling of admixture history. These correspond to Western Europe (WEU), Eastern Europe (EEU), admixed Europe (AEU) and North America (NA).

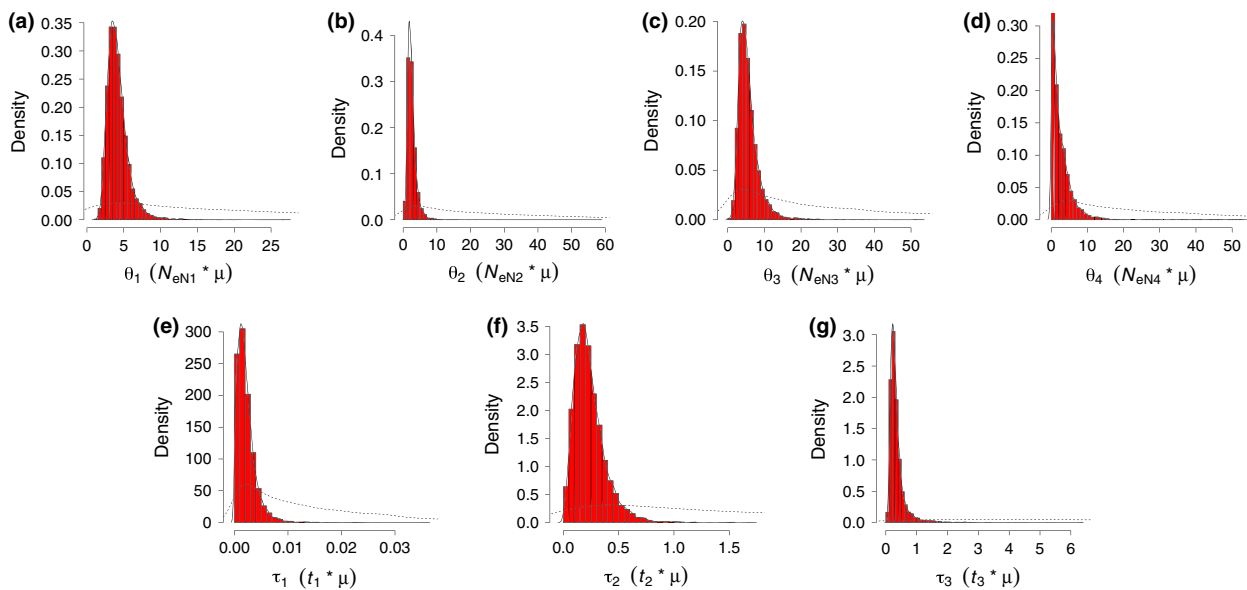
two orders of magnitude earlier (median  $\tau_2 = 2.10 \times 10^{-1}$ ) than admixture during the invasion of North America (median  $\tau_1 = 1.76 \times 10^{-3}$ ), confirming the prediction that native range admixture was more ancient than admixture during invasion (Fig. 3). The median absolute estimate of European admixture timing (ca. 32 000 generations ago; 95% CI 5300–84 900) places this event in the late Pleistocene or early Holocene, shortly after the historical divergence between WEU and EEU, but well before the North American admixture that occurred within the last several hundred generations (ca. 288 generations ago; 95% CI 17.5–491; Table 1).

*A posteriori* simulations affirmed our ability to differentiate Scenarios 1 and 2 based upon the utilized summary statistics (Table 2). Using logistic regression for model selection, both our estimated false-positive and false-negative rates of model selection were <1%, indicating strong confidence in our ability to infer the correct model between these two competing models. Similar results were found based on the direct estimate of model selection (Table 2). Model checking analysis confirmed that ABC estimates of demographic parameters from Scenario 1 were a good fit to the observed microsatellite genotypes, as 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 1 (Fig. S4; Table S4). Additionally, bias and precision assessment

for Scenario 1 parameters indicated most demographic parameters were estimated with good confidence (Table S5).

Multilocus heterozygosity for each genotype, estimated as standardized heterozygosity (SH) by Rhh, ranged from 0 to 2.06 and was not significantly different in mean between EU and NA ranges ( $\overline{SH}_{EU} = 0.97$  and  $\overline{SH}_{NA} = 0.99$ ;  $t = -0.32$ ,  $P = 0.75$ ). Similarly, the variance in SH among genotypes was not significantly different between EU and NA ranges (Levene's test:  $F = 1.49$ ,  $P = 0.07$ ). However, heterozygosity was significantly correlated with the degree of admixture for each genotype ( $r = 0.2297$ , 95% C.I. = 0.0820–0.3675); thus, the SH estimates recovered the predicted increase in genomic heterozygosity expected from introgression between divergent populations.

Linear modelling of the heterozygosity–fitness correlations (HFCs) indicated a significant interaction between SH and continent ( $F = 11.45$ ;  $P < 0.001$ ), and thus, separate models were run by continent of origin to estimate the strength of HFCs in EU and NA. When all EU genotypes were analysed together, there was no correlation between fitness and SH (Table 3). The absence of an HFC in Europe was also true when analysing just EU genotypes that showed admixed *Q*-scores from the STRUCTURE analysis or just those genotypes from the geographical suture zone in Central Europe that had a high frequency of admixture (both  $P > 0.5$ ). In contrast, NA genotypes exhibited a positive and highly significant correlation



**Fig. 3** Bayesian estimates of compound demographic parameters from the best-fit approximate Bayesian computation (ABC) demographic model (Scenario 1). Prior distributions are shown by dotted lines, and posterior distributions by filled bars. Parameters correspond to mutation-scaled estimates of effective population size (a–d), admixture timing for NA (e), AEU (f) and timing of divergence between WEU and EEU (g). Note differences in *x*-axis scale among plots, which are scaled based on the range in the prior distribution for each parameter.

**Table 1** Posterior parameter estimates from the best-fit approximate Bayesian computation (ABC) demographic model (Scenario 1). Time estimates for the original parameters are in units of generations, with the average generation time of *Silene vulgaris* approximately equal to 1.5 (Taylor & Keller, 2007).

Interpretation	Original parameter	Median (95% CI)	Composite parameter	Median (95% CI)
Western EU (WEU)	N1	6.28E+005 (2.21E+005 – 9.55E+005)	$\theta_1$	3.93E+000 (2.20E+000 – 7.81E+000)
Eastern EU (EEU)	N2	3.57E+005 (8.04E+004 – 8.83E+005)	$\theta_2$	2.27E+000 (8.90E–001 – 5.91E+000)
Admixed European (AEU)	N3	8.01E+005 (3.35E+005 – 9.89E+005)	$\theta_3$	5.01E+000 (2.08E+000 – 1.32E+001)
North America (NA)	N4	2.77E+005 (1.84E+004 – 9.48E+005)	$\theta_4$	1.79E+000 (1.10E–001 – 1.05E+001)
WEU/EEU split	t3	4.19E+004 (1.15E+004 – 2.17E+005)	$\tau_3$	2.83E–001 (1.08E–001 – 1.19E+000)
AEU founding time	t2	3.24E+004 (5.30E+003 – 8.49E+004)	$\tau_2$	2.10E–001 (4.40E–002 – 6.00E–001)
NA founding time	t1	2.88E+002 (1.75E+001 – 4.91E+002)	$\tau_1$	1.76E–003 (1.00E–004 – 6.74E–003)
AEU admixture of W/EEU	r2	5.74E–001 (1.38E–001 – 9.05E–001)		
NA admixture of W/EEU	r1	4.93E–001 (3.67E–001 – 6.23E–001)		

**Table 2** Confidence in scenario choice evaluated by computation of false-positive and false-negative rates for both direct estimate and logistic regression methods of model selection.

Scenario	Direct estimate		Logistic regression	
	False negatives	False positives	False negatives	False positives
1	0.036	0	0.004	0.004
2	0	0.036	0.008	0

**Table 3** Heterozygosity–fitness correlations (HFC's) for *Silene vulgaris* genotypes sampled from Europe (native) and North America (introduced). Admixed genotypes based on *Q*-scores were defined as  $0.25 \leq Q \leq 0.75$ , and those based on geography defined as being located within the Central European suture zone between expanding eastern and western refugial populations.

Comparison	<i>N</i>	Estimate (SE)	<i>F</i>	<i>P</i>	<i>R</i> <sup>2</sup>
Europe					
All genotypes	98	–0.46 (0.40)	1.26	0.263	0.005
Admixed genotypes ( <i>Q</i> -scores)	27	0.26 (0.53)	0.01	0.988	0.000
Admixed genotypes (geography)	33	–0.11 (0.49)	0.33	0.565	0.003
North America					
<b>All genotypes</b>	<b>100</b>	<b>1.01 (0.43)</b>	<b>12.88</b>	<b>&lt;0.001</b>	<b>0.043</b>
<b>Admixed genotypes (<i>Q</i>-scores)</b>	<b>25</b>	<b>2.58 (0.79)</b>	<b>17.03</b>	<b>&lt;0.001</b>	<b>0.175</b>
Nonadmixed genotypes ( <i>Q</i> -scores)	66	0.32 (0.51)	2.17	0.143	0.011

See text for further details. Bold values are significant after Bonferroni correction.

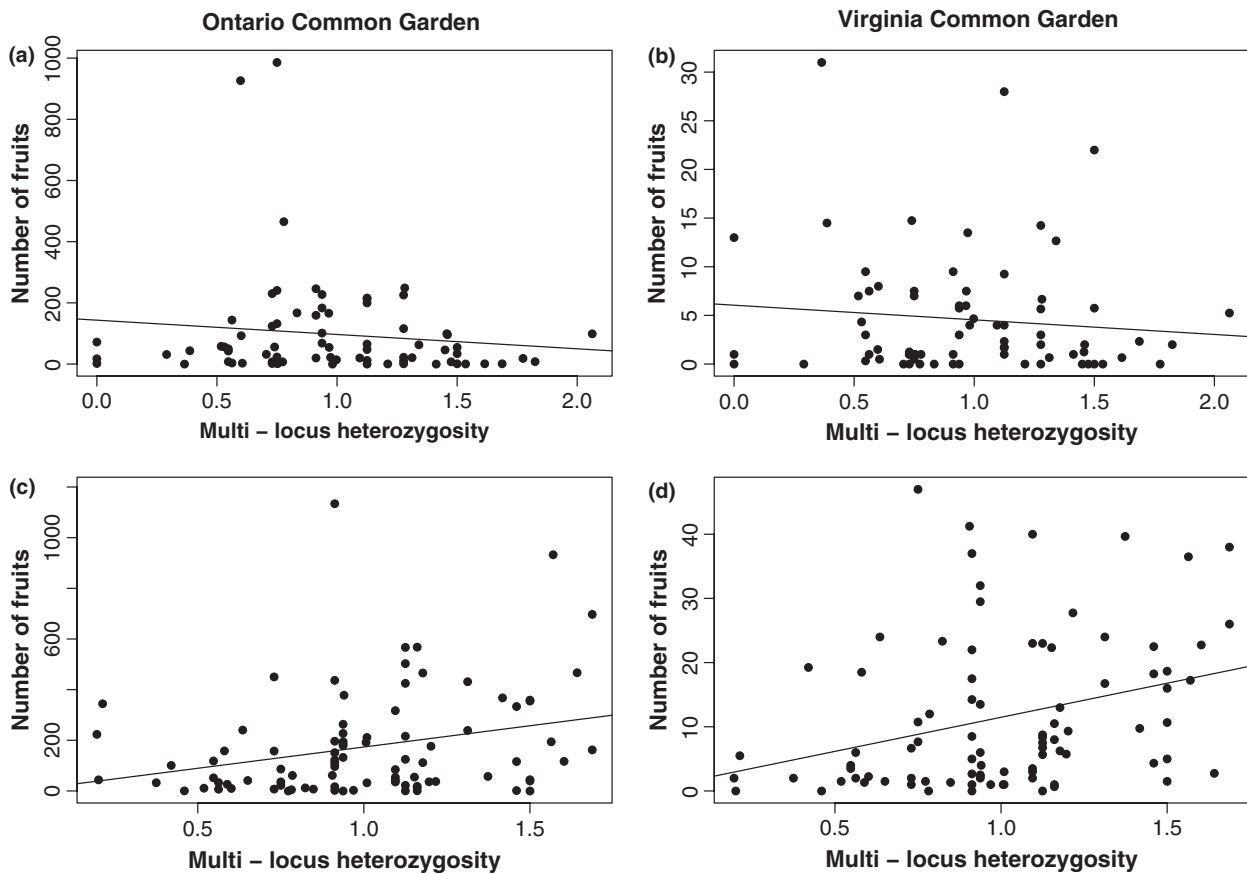
between fitness and SH when analysing all genotypes together ( $P < 0.001$ ; Table 3; Fig. 4). Standardized heterozygosity alone explained ca. 4.3% of the variance in fitness among all North American genotypes. The support for the HFC strengthened when analysing just

NA genotypes that showed admixed *Q*-scores ( $P < 0.001$ ), and increased the variance in fitness explained by SH to 17.5% (Table 3). As predicted, the HFC disappeared when analysing only NA genotypes with nonadmixed *Q*-scores ( $P = 0.143$ ). Importantly, the correlation between SH and fitness was not environmentally sensitive to common garden site (SH\*site interaction,  $P > 0.6$  for both EU and NA all-genotype models), indicating that the signal of positive association between heterozygosity and fitness was statistically repeatable under different growing conditions.

## Discussion

Although genetic admixture is associated with multiple introductions in many species invasions, little is known about how admixture affects the fitness of invasive genotypes. We tested the hypothesis that recent admixture between invasion sources of *Silene vulgaris* introduced to North America was responsible for increases in plant fitness due to a masking of the genetic load. We made two predictions: first, that admixed genotypes in the invasive range arose recently and were independent from the history of admixture in Europe; and second, that fitness should be positively associated with multilocus marker heterozygosity (a significant heterozygosity–fitness correlation, HFC) due to identity disequilibrium between markers and fitness QTL.

Our results support both of these predictions. ABC demographic models of population history unambiguously pointed to admixture in the native European range being much older (late Pleistocene to early Holocene) and independent from the admixture that accompanied the contemporary invasion of North America over the past several 100 years (Scenario 1). Timing of the European admixture between Western and Eastern EU lineages of *S. vulgaris* was coincident with post-glacial expansion and probably reflects historical secondary



**Fig. 4** Heterozygosity–fitness correlations for families deriving from Europe (a and b) and North America (c and d), measured in two common gardens (Ontario and Virginia). Multilocus heterozygosity is based on the standardized heterozygosity (SH) estimate from Rhh. Log (fruit set) is based on family mean estimates within each site. Lines show significant best-fit relationships based on a linear model.

contact between refugial populations that was common in Central Europe following glacial retreat (Petit *et al.*, 2003). This historical admixture did not produce the genotypes that colonized North America, however, which instead arose from separate introductions of WEU and EEU genotypes that became admixed after the invasion.

The magnitude of admixture among genotypes was positively correlated with multilocus heterozygosity, and the latter was positively correlated with fitness (a positive HFC) among NA but not EU genotypes. The effect size of the HFC showed that ca. 4.3% of fitness variance was explained by multilocus heterozygosity when analysing all NA genotypes together and increased to 17.5% when analysing just admixed NA genotypes. These values for HFC are comparatively large relative to a recent meta-analysis of HFCs in other natural populations (ca. <1–3% of fitness variance; Chapman *et al.*, 2009; Szulkin *et al.*, 2010).

What is the likely genetic mechanism underlying the association between heterozygosity and fitness in

*S. vulgaris*? HFC's arise due to identity disequilibrium in genotypic state between markers and fitness loci. As heterozygosity is positively related to both admixture and fitness in *S. vulgaris*, we suggest that during the recent invasion of North America, admixture between WEU and EEU gave rise to genotypes that possessed a decreased frequency of loci harbouring alleles that show identity by descent (IBD). The frequency of IBD at fitness loci will affect the degree to which the genetic load of deleterious mutations is expressed. Thus, by increasing heterozygosity across the genome, admixture during invasion probably reduced the frequency of IBD within 'hybrid' NA genotypes, thereby sheltering the genetic load and increasing fitness (i.e. heterosis; Lynch & Walsh, 1998; Charlesworth & Willis, 2009), as well as generating the identity disequilibrium between fitness and marker loci that lead to the HFC.

Possible alternative explanations to explain the fitness results in NA and EU genotypes are local adaptation and inbreeding depression. The degree of local adaptation (or pre-adaptation) exhibited by



different NA and EU genotypes is likely to affect plant fitness in common gardens and probably contributes to some of the variation in fruit production we observed (see also Keller *et al.*, 2009). However, local adaptation would not be expected to result in a positive heterozygosity–fitness relationship and in some cases may even cause a negative HFC (i.e. selection against locally maladapted hybrids). In addition, the significant HFC we observed in *S. vulgaris* was repeatable across both common gardens, suggesting the HFC was robust to growing environment. Lastly, our previous work showed that fruit production of nonadmixed genotypes was similar between NA and EU, but that admixed NA genotypes had increased fruit production by two-fold over admixed EU genotypes (Fig. 4 in Keller & Taylor, 2010).

Heterozygosity–fitness correlations may also arise due to inbreeding depression when demographic events that increase the frequency of IBD, such as genetic bottlenecks or changes in mating system, lead to decreased fitness in less heterozygous genotypes. Our ABC analysis does suggest a decrease in effective population size in North America (Table 1), which reflects a modest but detectable genetic bottleneck associated with the founding of invasive populations. However, we think that the HFC present among NA genotypes is not due to increased inbreeding for the following reasons. First, if the HFC in North America was caused by increased inbreeding, then we would expect the overall fitness of NA genotypes to be lower than that of EU genotypes, which is not the case (Fig. S5; see also Keller & Taylor, 2010). Additionally, we would expect the HFC to be present among nonadmixed genotypes, as these would have experienced the bottleneck, and thus the effects of increased IBD, which again is not observed (Table 3). Therefore, we conclude that the most likely cause of the HFC in *S. vulgaris* is recent admixture in North America increasing identity disequilibrium between heterozygosity at neutral molecular markers and fitness loci, with the latter benefiting from the increased sheltering of the genetic load of deleterious mutations.

Whereas recent admixture in North America increased fitness among highly heterozygous genotypes, the older admixture in Europe appears to be associated with a loss of high fitness genotypes and a disruption of the heterozygosity–fitness correlation. Given the large number of generations elapsed since the EU admixture, we suspect the loss of fitness in part reflects the loss of heterosis expected with segregation (Lynch, 1991). The lack of an HFC in Europe likely reflects the dissolution of identity disequilibrium between unlinked loci caused by independent assortment and recombination. Given the time since EU admixture, we would expect the loss in identity disequilibrium to cause the heterozygosity at marker loci to become less predictive of heterozygosity

at fitness loci. Thus, both the fitness benefits of heterosis due to admixture and the manifestation of an HFC are likely to be transient and only detectable early in the admixture history. However, with only a single comparison between two admixture events (NA vs. EU), it is impossible to state with certainty that the loss of HFC is directly attributable to time since admixture. A more powerful test of this hypothesis would be to compare the HFC from many populations arrayed along a gradient of time since admixture. Such tests might be possible for invasions where the expansion history is particularly well known.

Our study shows that the demographic history of multiple introductions and admixture associated with many biological invasions can affect heterozygosity across the genome, with corresponding fitness effects on populations observable over ecological timescales – a hypothesis that has so far received limited empirical testing (Keller & Taylor, 2010; Facon *et al.*, 2011; Turgeon *et al.*, 2011; Verhoeven *et al.*, 2011). Comparative tests of HFCs in populations from the native vs. introduced ranges of an invasive species highlight the importance that demographic history plays in generating genetic variation in fitness (Szulkin *et al.*, 2010). What is less clear is whether admixture actually helps drive the range expansion of invasive species forward vs. the alternative scenario that admixture and its fitness effects are simply passive by-products of the invasion process (i.e. secondary contact among different invasion sources may be somewhat inevitable during invasions experiencing high propagule pressure). That is, does the short-term fitness boost from heterosis lead to either 1) greater population densities or 2) an accelerated rate of expansion relative to invading populations not experiencing heterosis?

Drake (2006) argued that heterosis could act as a ‘catapult’ for colonizing populations, contributing to positive population growth rates when populations are small and/or young and most subject to Allee effects or reproductive failure. As populations age, even though the fitness gains from heterosis fade, the increased population growth has lasted long enough to carry populations past the initial threat of extinction from demographic stochasticity. Field surveys that measure the intrinsic rate of increase in invasive populations across a continuum of population age and admixture status would provide an intriguing test of this idea (and the hypothesis of HFC’s fading with time), as would spatially explicit simulation studies that compare the extent and rate of range expansion under contrasting scenarios of admixture and heterosis. It would also be interesting to examine how the prior inbreeding history of populations affects the magnitude of heterosis (and the catapult effect), with one prediction being that the magnitude of heterosis would be greater for more inbred populations, creating a genetic rescue effect of admixture during invasion

(Richards, 2000; Ingvarsson, 2001). This might be particularly relevant when admixture occurs between introduced populations that have undergone severe genetic bottlenecks, as might be common during the initial establishment of invasive species.

Although we have primarily focused on sheltering the genetic load as a positive fitness outcome of admixture, it should be noted that mixing divergent gene pools can bring about a variety of both short- and longer-term evolutionary outcomes (e.g. increased additive genetic variance, transgressive segregation, novel multivariate trait combinations), some of which may have longer-lasting impacts than we observe in EU vs. NA genotypes of *S. vulgaris*. Some genetic consequences of admixture may also be detrimental to invading populations (e.g. outbreeding depression). For example, it is well known that admixture can decrease the fitness of late-generation hybrids (F2 and beyond) if epistatic combinations of positively interacting loci are broken up by recombination and independent assortment (Lynch, 1991; Edmands, 1999). Thus, we might expect that for some introduced species, admixture (or hybridization more generally) could be more permanent, or result in outbreeding depression and reduced population growth rates (Keller & Waller, 2002). Given these different outcomes of admixture on fitness, it would be useful to begin developing predictive metrics of when heterosis or outbreeding depression can be expected for a given invasion (Frankham *et al.*, 2011). Thus, the management of particularly noxious invasive species may stand to directly benefit from genetic studies that connect the effects of admixture with their phenotypic consequences.

## Acknowledgments

This research was funded by NSF DEB # 0919335. The authors have no conflict of interests to declare. We thank the Taylor laboratory group and two anonymous reviewers for insightful comments that greatly improved the manuscript. This is contribution number 4856 from the University of Maryland Center for Environmental Science.

## References

- Alho, J.S., Välimäki, K. & Merilä, J. 2010. R<sub>hh</sub>: an R extension for estimating multilocus heterozygosity and heterozygosity–heterozygosity correlation. *Mol. Ecol. Resour.* **10**: 720–722.
- Amos, W., Wilmer, J.W., Fullard, K., Burg, T.M., Croxall, J.P., Bloch, D. *et al.* 2001. The influence of parental relatedness on reproductive success. *Proc. Biol. Sci.* **268**: 2021–2027.
- Aparicio, J.M., Ortego, J. & Cordero, P.J. 2006. What should we weigh to estimate heterozygosity, alleles or loci? *Mol. Ecol.* **15**: 4659–4665.
- Calsbeek, B., 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. *Evol. Appl.* **4**: 726–735.
- Chapman, J.R., Nakagawa, S., Coltman, D.W., Slaters, J. & Sheldon, B.C. 2009. A quantitative review of heterozygosity–fitness correlations in animal populations. *Mol. Ecol.* **18**: 2746–2765.
- Charlesworth, D. & Willis, J.H. 2009. The genetics of inbreeding depression. *Nat. Rev. Genet.* **10**: 783–796.
- Colautti, R.I. & Barrett, S.C.H. 2011. Population divergence along lines of genetic variance and covariance in the invasive plant *Lythrum salicaria* in eastern North America. *Evolution* **65**: 2514–2529.
- Coltman, D.W., Pilkington, J.G., Smith, J.A. & Pemberton, J.M. 1999. Parasite-mediated selection against inbred Soay Sheep in a free-living, island population. *Evolution* **53**: 1259–1267.
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.-M., Balding, D.J. *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.
- David, P. 1998. Heterozygosity–fitness correlations: new perspectives on old problems. *Heredity* **80**: 531–537.
- Dlugosch, K.M. & Parker, I.M. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol. Ecol.* **17**: 431–449.
- Drake, J.M. 2006. Heterosis, the catapult effect and establishment success of a colonizing bird. *Biol. Lett.* **2**: 304–307.
- Earl, D. & vonHoldt, B. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**: 359–361.
- Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution* **53**: 1757–1768.
- Ellstrand, N.C. & Schierenbeck, K.A. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proc. Natl. Acad. Sci. USA* **97**: 7043–7050.
- Estoup, A., Lombaert, E., Marin, J.-M., Guillemaud, T., Pudlo, P., Robert, C.P. *et al.* 2012. Estimation of demographic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Mol. Ecol. Resour.* **12**: 846–855.
- Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**: 2611–2620.
- Excoffier, L., Foll, M. & Petit, R.J. 2009. Genetic consequences of range expansions. *Annu. Rev. Ecol. Evol. Syst.* **40**: 481–501.
- Facon, B., Jarne, P., Pointier, J.P. & David, P. 2005. Hybridization and invasiveness in the freshwater snail *Melanooides tuberculata*: hybrid vigour is more important than increase in genetic variance. *J. Evol. Biol.* **18**: 524–535.
- Facon, B., Hufbauer, R.A., Tayeh, A., Loiseau, A., Lombaert, E., Vitalis, R. *et al.* 2011. Inbreeding depression is purged in the invasive insect *Harmonia axyridis*. *Curr. Biol.* **21**: 424–427.

- Falush, D., Stephens, M. & Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Frankham, R., Ballou, J.D., Eldridge, M.D.B., Lacy, R.C., Ralls, K., Dudash, M.R. *et al.* 2011. Predicting the probability of outbreeding depression. *Conserv. Biol.* **25**: 465–475.
- Hansson, B. & Westerberg, L. 2002. On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* **11**: 2467–2474.
- Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc.* **58**: 247–276.
- Hufbauer, R.A. 2008. Biological invasions: Paradox lost and paradise gained. *Curr. Biol.* **18**: R246–R247.
- Ingvansson, P. 2001. Restoration of genetic variation lost – The genetic rescue hypothesis. *Trends Ecol. Evol.* **16**: 62–63.
- Jakobsson, M. & Rosenberg, N.A. 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). *Mol. Ecol. Resour.* **3**: 358–359.
- Keller, S.R. & Taylor, D.R. 2010. Genomic admixture increases fitness during a biological invasion. *J. Evol. Biol.* **23**: 1720–1731.
- Keller, L. & Waller, D. 2002. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**: 230–241.
- Keller, S.R., Sowell, D.R., Neiman, M., Wolfe, L.M. & Taylor, D.R. 2009. Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytol.* **183**: 678–690.
- Keller, S.R., Gilbert, K.J., Fields, P.D. & Taylor, D.R. 2012. Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Mol. Ecol.* **21**: 4721–4734.
- Kolbe, J.J., Glor, R.E., Rodriguez-Schettino, L., Chamizo-Lara, A., Larson, A. & Losos, J.B. 2004. Genetic variation increases during biological invasion by a Cuban lizard. *Nature* **431**: 177–181.
- Lambrinos, J.G. 2004. How interactions between ecology and evolution influence contemporary invasion dynamics. *Ecol. Evol.* **85**: 2061–2070.
- Lee, C.E. 2002. Evolutionary genetics of invasive species. *Trends Ecol. Evol.* **17**: 386–391.
- Lynch, M. 1991. The Genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* **45**: 622–629.
- Lynch, M. & Walsh, J.B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, Massachusetts.
- Martindale, I.C. 1876. The introduction of foreign plants. *Bot. Gaz.* **2**: 55–58.
- Matschiner, M. & Salzburger, W. 2009. TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25**: 1982–1983.
- Moccia, M.D., Oger-Desfeux, C., Marais, G.A. & 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.
- Petit, R.J., Aguinalalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R. *et al.* 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* **300**: 1563–1565.
- Pritchard, J., Stephens, M. & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Pursh, F. 1814. *Flora Americae Septentrionalis*. White, London.
- R Development Core Team 2011. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reznick, D.N. & Ghalambor, C.K. 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112**: 183–198.
- Richards, C. 2000. Inbreeding depression and genetic rescue in a plant metapopulation. *Am. Nat.* **155**: 383–394.
- Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**: 137–138.
- Salamin, N., Wuest, R.O., Lavergne, S., Thuiller, W. & Pearn, P.B. 2010. Assessing rapid evolution in a changing environment. *Trends Ecol. Evol.* **25**: 692–698.
- Szulkin, M., Bierne, N. & David, P. 2010. Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64**: 1202–1217.
- Taylor, D.R. & Keller, S.R. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* **61**: 334–345.
- Turgeon, J., Tayeh, A., Facon, B., Lombaert, E., De Clercq, P., Berkvens, N. *et al.* 2011. Experimental evidence for the phenotypic impact of admixture between wild and biocontrol Asian ladybird (*Harmonia axyridis*) involved in the European invasion. *J. Evol. Biol.* **24**: 1044–1052.
- Verhoeven, K.J.F., Macel, M., Wolfe, L.M. & Biere, A. 2011. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc. R. Soc. Lond. B* **278**: 2–8.

## Supporting information

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

**Table S1** Bayesian prior distributions for demographic parameters used in ABC modelling.

**Table S2** Population genetic diversity within STRUCTURE clusters based on  $K = 2$ .

**Table S3** Pair-wise genetic differentiation ( $F_{ST}$ ) among groups.

**Table S4** Deviations of summary statistics for the observed data from the posterior predictive distribution of the most probable invasion history (Scenario 1).

**Table S5** Bias and precision of parameter estimates from ABC for scenario 1.

**Figure S1** Delta-K across 10 replicate STRUCTURE runs per value of K.

**Figure S2** Maps of STRUCTURE clustering at  $K = 2$  and  $K = 3$  in North America and Europe.

**Figure S3** Posterior model probabilities for competing Scenarios 1 and 2 based on a direct estimate and logistic regression.

**Figure S4** Principal component analysis of population genetic summary statistics from ABC demographic modelling.

**Figure S5** Box plots showing mean fruitset differences

between genotypes collected in Europe (EU) and North America (NAM) for two different common garden sites.

Data deposited at Dryad: doi: 10.5061/dryad.vt8t7

*Received 9 August 2013; accepted 4 January 2014*