When plants establish outside their native range, their ability to adapt to the new environment is influenced by both demography and dispersal. However, the relative importance of these two factors is poorly understood. To quantify the influence of demography and dispersal on patterns of genetic diversity underlying adaptation, we used data from a globally distributed demographic research network comprising 35 native and 18 nonnative populations of Plantago lanceolata. Species-specific simulation experiments showed that dispersal would dilute demographic influences on genetic diversity at local scales. Populations in the native European range had stronger spatial genetic structure associated with geographic distance and precipitation seasonality. In contrast, nonnative populations had weaker spatial genetic structure that was not associated with environmental gradients but with higher within-population genetic diversity. Our findings show that dispersal caused by repeated, long-distance, human-mediated introductions has allowed invasive plant populations to overcome environmental constraints on genetic diversity, even without strong demographic changes. The impact of invasive plants may, therefore, increase with repeated introductions, highlighting the need to constrain future introductions of species even if they already exist in an area.
Patterns of genetic diversity across a species’ range arise from a complex interplay between the diversifying effect of demographic variation across landscapes with different selection pressures and the homogenizing effects of dispersal (1–3). On one hand, variability in demographic performance influences genetic diversity through its influence on effective population size (4). Short-lived, highly fecund species generally have higher levels of genetic diversity compared with species that are long lived or have low fecundity (5, 6). On the other hand, dispersal modulates these relationships by facilitating gene flow between populations (7). Gene flow from seed and pollen can increase genetic diversity and reduce genetic differences among populations. While the importance of these forces is widely accepted (8), there is uncertainty about the relative strength of demography and dispersal in shaping genetic structure across global environmental gradients (9, 10).

For invasive species, the situation is even more complex because humans disrupt many of the natural processes that determine genetic diversity (Fig. 1). For example, repeated introductions and long-distance dispersal by humans can release invasive plant species from demographic constraints, such as those imposed by the colonization–competition tradeoff (11). Invasive species might also overcome climatic constraints on phenotypic traits as a result of rapid adaptation to new environments (12) or nonadaptive processes, such as repeated introductions, which can swamp locally adapted phenotypes (13). Thus, emerging evidence suggests that plants in their nonnative range can break ecological “rules” because they are not always constrained by the same biological and climatic forces that operate in their native range.

Some populations of invasive species lose genetic diversity during invasion through founder effects (14), but many have higher genetic diversity outside their native range (15, 16). The mechanisms underlying this phenomenon include admixture (i.e., new genotypes arising from interbreeding among divergent source populations) (17), hybridization (18), rapid mutation (19), and exposure of cryptic genetic variation (20). Such increases in genetic diversity can enhance colonization success (21) and adaptive potential (22) in invasive species. Demographic changes can also improve invasive plant performance (23), which is sometimes associated with release from natural enemies (24).

Unfortunately, demographic and genetic aspects of invasion are...
often analyzed in isolation (25), in part because labor-intensive demographic studies are typically done at one or a few sites, making them severely limited in spatial replication (26). This means that we lack understanding about the relative importance of demographic change and global dispersal on biological invasions (27, 28).

Here, we present a demographically informed analysis of neutral and putatively adaptive genetic diversity in *Plantago lanceolata* L. (Plantaginaceae), a common forb native to Europe and western Asia, which now has a cosmopolitan distribution (Fig. 2). *P. lanceolata* established in its nonnative range through long-distance dispersal by humans (29), repeated introductions (30), and cultivation (31)—all processes that can increase genetic diversity and invasion success (15). The overarching aim of the study was to analyze the influences of local demography and global dispersal patterns on genetic diversity in *P. lanceolata* and determine which of these pathways drives adaptive capacity. This knowledge is necessary to understand how future introduction events will influence the spread of invasive plants. This global analysis of genetic diversity, which integrates field-collected demographic data, was made possible by a spatially distributed demographic research network (PLANTPOPNET).

In addition to demographic data, we sampled DNA from 491 individuals, including outgroups and cultivar lines, and 53 naturally occurring populations across the native European range (*n* = 35) and the nonnative range (*n* = 18) in southern Africa, Australasia, and North America (Fig. 2). To address our main aim, three hypotheses were tested.

1) **Hypothesis 1.** In the absence of dispersal, increases in survival and fecundity will drive increases in genetic diversity. These effects will be diluted by dispersal between populations.

2) **Hypothesis 2.** Patterns of spatial genetic structure among native populations will reflect dispersal limitations across environmental gradients. In the nonnative range, gene flow arising from multiple introductions will disrupt spatial genetic structure observed in the native range.

3) **Hypothesis 3.** Environmental influences on within-population genetic diversity will be explained by demographic variation (density, fecundity, and empirical population growth rate). Repeated introductions into the nonnative range and long-distance dispersal by humans will weaken this relationship (Fig. 1).

A genotypic simulation model, parameterized with empirical demographic data from *P. lanceolata*, was used to test Hypothesis 1. We then coupled field-collected demographic data (density, empirical population growth rate, and fecundity) with single-nucleotide polymorphism (SNP) data (18,166 neutral and 3,024 putatively adaptive SNPs) to test Hypotheses 2 and 3.

![Fig. 2. Global genetic structure in *P. lanceolata*. (A) Colored bars represent the proportion of individual genotypes in each population assigned to one of six genetic clusters identified with fastSTRUCTURE. For clarity, multiple sites were aggregated where overlapping bars had similar assignment probabilities (e.g., southern Ireland, Switzerland). Dark gray points are *P. lanceolata* records from GBIF/BIEN (84, 85). For each nonnative region, the minimum number of propagules (mean ± SE) overall (Propmin) and relative to sample size (Propmin/N), indicates that multiple introductions would be required to produce observed levels of genetic diversity. The number of non-European alleles indicates that more genetic diversity was present in nonnative regions than could be explained by the native sample. (B) Probability of assignment for 491 individuals to six genetic clusters, with individuals grouped by population within region. Three commercial cultivar lines and two outgroups (*P. coronopus* and *P. major*) were included.](https://www.pnas.org/content/doi/10.1073/pnas.2008317117)
The increases in genetic diversity with juvenile survival (Fig. 3) might not confer an adaptive advantage since they reflect genetic diversity arising from neutral demographic processes. The relevance of this result, however, is that there is enough demographic variability in *P. lanceolata* to shape neutral genetic structure, an assumption underlying the hypotheses in the rest of the study. Thus, we can expect juvenile survival to be the dominant demographic parameter underlying differences in *P. lanceolata* genetic diversity when dispersal is limited at local scales. At continental scales, genetic diversity is probably influenced less by juvenile survival when gene flow is high. This might be especially true in the nonnative range where there has been a shorter history of local adaptation (33) and multiple human-mediated introductions (the human activity pathway) (Fig. 1).

**Hypothesis 2: Global Gene Flow from Multiple Introductions Will Disrupt Spatial Genetic Structure.** Admixture analysis of *P. lanceolata* genotypes with fastSTRUCTURE (34) revealed strong genetic structure in the native range and a high degree of admixture in the nonnative range. The number of genetic clusters at Hardy–Weinberg (HW) equilibrium (K) was between 4 and 6 (model complexity maximizing marginal likelihood) and K = 13 (model components used to explain structure in the data), when K = 6, cultivar lines and outgroups (*Plantago coronopus* and *Plantago major*) formed two distinct clusters, and the remaining four clusters were present in the native European range with clear spatial structure (Fig. 2). Greece, Italy, the Islands of the North Atlantic, and Finland made up almost “pure” lines of these four clusters, while other European populations were admixed.

Genotypes of most nonnative populations were admixed, and there was relatively little spatial structure at a global scale (Fig. 2). This was supported by a significantly higher diversity score in the nonnative range (model estimate, SE = 0.34, 0.04) compared with the native range (0.22, 0.03; P = 0.033) (SI Appendix, Fig. S6). Italy and central France were the most similar source material for the dominant genotype in the nonnative populations. Some cultivar stock was identified in the Spanish populations, possibly reflecting the Iberian source of material used to breed cultivars. The cultivars were developed in New Zealand; thus, the presence of cultivar stock in that population might indicate mixing between the naturalized population and pasture plants (Fig. 2). At the upper range of K, further spatial structure was identified in Europe (e.g., at K = 13, Norway was differentiated from Finland), while the nonnative populations still showed admixture of multiple, mostly Mediterranean sources (SI Appendix, Fig. S1). The lack of spatial structure at a global scale was supported by analysis of molecular variance showing that genetic variation between the native and nonnative range was only 2.2%, among individuals within populations was 10.7%, and among populations within ranges was 11.4%. The remaining genetic variation (75.5%) accounted for individual heterozygosity.

The minimum number of colonizing propagules required to produce the observed level of genetic diversity in nonnative regions (*Prop* min) depended on sample size (*r* = 0.99) and ranged from 5.35 in New Zealand to 49.35 in North America (Fig. 2). Multiple introductions were, therefore, required to produce observed levels of genetic diversity in the nonnative ranges. Relative to sample size, *Prop* min ranged from 0.55 to 0.90, indicating that, in each region, more than half the sampled population was required to represent nonnative genetic diversity. *Prop* min was based on the alleles present in the native range, but there were also a number of non-European alleles in each nonnative range (12 to 159) (Fig. 2). Thus, we either failed to sample the full extent of the source population (despite extensive sampling across Europe), or new genotypes were produced after colonization. The latter explanation can arise through transgressive segregation (35) and is one mechanism by which

**Results and Discussion**

**Hypothesis 1: Dispersal between Populations Will Dilute Demographic Effects on Genetic Diversity.** In two simulated populations unconnected by dispersal, with different rates of juvenile survival (σ = 0.1 and 0.2) and female fecundity (seeds per plant, ΦF = 1 to 100), higher juvenile survival led to greater genetic diversity (Fig. 3A). Above the threshold at which populations went extinct (ΦF = 15), genetic diversity increased sharply until ΦF was ~25. Above this point, there was little influence of fecundity on genetic diversity (Fig. 3A). Population size at the end of the simulation was larger with higher juvenile survival (Fig. 3B). Thus, variation in female fecundity seems to have less influence than juvenile survival in determining genetic diversity in *P. lanceolata*. When the two populations were connected by dispersal, differences in heterozygosity persisted until the number of migrants per generation exceeded 50,000 (Fig. 3C and D). This number is realistic in natural populations since reproductive individuals typically produce a minimum 20 to 100 seeds, and migration refers to propagules dispersed before the recruitment process. Male fecundity was kept constant in the model as it is very high in *P. lanceolata* [10,000 to 54,000 pollen grains per anther (32)] and had no influence on genetic diversity.

The simulation result supports our prediction (Hypothesis 1) that demographic processes would influence genetic diversity in *P. lanceolata* when dispersal barriers are present and that dispersal would dilute these effects. The simulation also suggests that juvenile survival is an important parameter controlling heterozygosity. When dispersal barriers are removed, however, gene flow from pollen and seed will swamp local effects of juvenile survival on heterozygosity. We could, therefore, expect demographic effects on genetic diversity to become undetectable at the upper range of pollen and seed movement that occurs in *P. lanceolata*.
invasive species adapt quickly to new environments. However, we also detected private alleles within sites in Europe (SI Appendix, Table S1), and therefore, our sample does not represent the full range of genetic diversity in the species.

Genetic structure measured by $F_{ST}$ (genetic differentiation between all pairs of populations) was stronger among populations in the native range (mean $F_{ST} = 0.16$) than the nonnative range (mean $F_{ST} = 0.09$). To analyze the influence of environmental gradients on $F_{ST}$, we used three separate generalized dissimilarity models, one for each range type: native range, nonnative range, and the global population (native and nonnative combined). The deviance explained by the native model was 74.3% (bootstrap CI = 68.6, 78.3), and two of six variables fitted in the model had a significant influence on $F_{ST}$ (Fig. 4 and SI Appendix, Fig. S2). Genetic distance increased with geographic distance (Fig. 4A), and sites with similar levels of precipitation seasonality were more genetically similar (Fig. 4B) after accounting for other variables in the model (SI Appendix, Fig. S2). No variable significantly affected $F_{ST}$ in the nonnative range (deviance explained = 23.1%, bootstrap CI = 9.4, 34.1) or the global population (deviance explained = 10.9%, bootstrap CI = 7.25, 14.33) (SI Appendix, Fig. S2). Geographic distance was included in each model to account for differences in spatial scale. Thus, if environmental influences on gene flow had persisted in the nonnative range, they should have been detectable. Combined with the admixture analysis, these results support our prediction (Hypothesis 2) that multiple introductions from diverse source populations and long-distance dispersal can weaken environment–genetic structure relationships.

*P. lanceolata* reproduces clonally as well as sexually, and this flexible reproductive mode combined with high admixture in the nonnative range suggests fast expansion after colonization. This might allow the species to overcome ecological constraints without the need for local adaptation (36).

In the native range of *P. lanceolata*, the increase in genetic distance with precipitation seasonality might partially reflect a historic biogeographical pattern (precipitation seasonality was correlated with longitude, $r = 0.47$). Historical processes occurring along both east-west and north–south axes shape contemporary genetic patterns in European plants. For example, glacial refugia in Iberia, Italy, and the Balkans were reflected in highly divergent lines of *Arabidopsis thaliana* south of the alpine barrier (37). In our dataset, the Italian population was genetically distinct, while two eastern sites in Romania were highly differentiated and genetically related to Greece (Fig. 2). François et al. (37) also found evidence for an eastern refuge in *A. thaliana*. Further sampling into the continental Asian range of *P. lanceolata* would help uncover whether the observed patterns arose from movement with agriculture westward across Europe (38, 39) or postglacial colonizers from the Balkans (40).

**Hypothesis 3: Global Gene Flow Will Weaken Demographic Effects on Genetic Diversity within Populations.** We compared a series of linear models, including additive and interactive effects of range (native/nonnative), to address the hypothesis that environmental influences on within-population genetic diversity would differ between the native and nonnative ranges (Dataset S1). Our results offered partial support for Hypothesis 3 because environmental gradients (characterized by mean temperature, temperature seasonality, and mean precipitation) affected population growth rate, fecundity, and neutral and adaptive genetic diversity in native and nonnative ranges of *P. lanceolata* (Fig. 5 and SI Appendix, Fig. S3). Our expectation, however, that genetic responses to the environment could be explained by demographic variation had little support (SI Appendix, Fig. S3). Demographic variables responded to environmental gradients but did not induce a response on genetic diversity when used as predictor variables. Demographic and genetic parameters within populations were best explained by environmental gradients, and in some cases, there were differences in the responses between native and nonnative ranges.

The top-ranked models for population growth rate (Fig. 5A) and fecundity (Fig. 5B) had additive effects of mean temperature, responding similarly in the native and nonnative ranges. Globally, warmer sites tended to have lower population growth rates and higher fecundity. Increases in fecundity can occur to offset lower survival in stressful environments (41), a phenomenon that has been recorded in other studies of *Plantago* (42, 43). There was also an additive effect of temperature seasonality on neutral genetic diversity (Fig. 5C), with highly seasonal sites having greater genetic diversity in the native and nonnative ranges. Mean temperature and temperature seasonality were correlated ($r = −0.36, P = 0.02$) (SI Appendix, Fig. S4). Thus, the observed responses are best thought of as responses to an environmental gradient, with demographic and genetic parameters responding to different aspects of the gradient. High genetic diversity in highly seasonal sites might have been driven by increased fecundity since we found some evidence of a positive relationship between fecundity and genetic diversity (SI Appendix, Fig. S3G and Dataset S1).

Three of the top-ranked models included an interaction between environment and range, showing environmental effects in the native range but not the nonnative range. Both neutral (Fig. 5D) (bootstrap CI = 0.001, 0.010) and adaptive (Fig. 5F) (bootstrap CI = 0.004, 0.021) genetic diversity decreased across a mean precipitation gradient in the native range but not in the nonnative range. Adaptive genetic diversity increased with temperature seasonality but only in the native range (Fig. 5E) (bootstrap CI = −0.021, −0.005). There was also support (a change in Akaike Information Criterion $\Delta AICc < 2$) for nonnative populations having a weaker response to environmental gradients in terms of fecundity (SI Appendix, Fig. S3A and B), population growth rate (SI Appendix, Fig. S3C), and neutral genetic diversity (SI Appendix, Fig. S3D). Taken together, these results suggest that nonnative populations are not constrained by the same environmental forces as their native counterparts.

Population growth rate and neutral and adaptive genetic diversity were all higher in the nonnative range (Fig. 5 and Dataset S1), suggesting that invasive populations have a greater capacity for colonization and adaptation. Higher population growth rates in nonnative populations were probably driven by increases in survival rather than fecundity since fecundity was lower in the nonnative range (Fig. 5B and Dataset S1). Thus, our simulation experiments and our field data indicated stronger effects of survival than of fecundity on genetic diversity and population growth, respectively.

**Fig. 4.** Genetic distance ($F_{ST}$) between pairs of *P. lanceolata* populations in the native European range was explained by two variables: (A) geographic distance and (B) distance in precipitation seasonality (coefficient of variation of annual mean precipitation) between sites. A generalized dissimilarity model indicated that these variables had a significant (adjusted $P < 0.001$) effect on $F_{ST}$ given all other variables in the model (geographic distance, mean temperature, mean precipitation, temperature seasonality, and precipitation seasonality). Deviance explained by the model was 74.3%, and the model splines are shown in SI Appendix, Fig. S2.
Increases in genetic diversity can arise when environmental heterogeneity drives population turnover through increases in sexual reproduction, population growth, and survival (6, 44). In our study, however, population growth was affected by mean temperature, not variability in temperature; cooler sites generally had higher rates of population growth across the first two demographic censuses. This is consistent with previous work showing that high mean temperature was associated with mortality in *P. lanceolata* (42). Thus, we did not find a clear demographic explanation for the effect of temperature seasonality on genetic diversity. Temperature stability might have promoted clonality in *P. lanceolata*, leading to lower genetic diversity (45). However, rates of sexual and clonal reproduction within species are often inversely related (46), and genetic diversity was unaffected by rates of sexual reproduction in our study. The influence of global variation in clonality on genetic diversity needs further investigation, particularly because clonality combined with sexual reproduction can increase invasion success (36).

Our prediction that environmental effects on genetic diversity could be explained by demographic variation had only limited support, even in the native range. Except for a weak increase in neutral genetic diversity with density (SI Appendix, Fig. S3F) and fecundity (SI Appendix, Fig. S3G), there was little direct influence of demographic variables on genetic diversity. There are at least two explanations for this general lack of a demographic relationship. First, genetic structure can arise even under frequent dispersal (44). Thus, although we found strong spatial genetic structure in the native range, it is possible that dispersal was high enough to mask any influence of demography on genetic diversity (the natural dispersal pathway) (Fig. 1). Second, the fine scale of demographic sampling within sites (a few meters²) might not reflect effective population size (47). This fits with our understanding of abiotic filters operating at all scales, while biotic filters, such as inter- and intraspecific interactions affecting demographic performance, generally operate at localized scales (10, 13). *P. lanceolata* is also highly genetically variable within and outside its native range. Thus, the low power within sites might have limited our ability to draw conclusions about demographic influences on genetic diversity. Sampling more individuals per site in the future might reveal stronger effects of fecundity, survival, and population growth on genetic diversity.

In summary, genetic diversity in *P. lanceolata* seems to be shaped predominantly by temperature and precipitation gradients.
related to gene flow and admixture rather than demographic vari-
ation. Our data support the prediction that high dispersal would
dilute demographic effects on genetic diversity (Hypothesis 1). Globally, our analyses suggest that genetic diversity in the nonnative
range is shaped by admixture from multiple source populations and
ongoing introductions, leading to high neutral and adaptive genetic
diversity (Hypothesis 2). Our data suggest that invasive populations
can establish in a broad range of environments without the need
for associated demographic change. Thus, there was little support
for the prediction that demographic variation could explain envi-
ronmental effects on genetic diversity (Hypothesis 3). Our unique
global demographic dataset provides evidence that invasive species
can overcome ecological rules in their nonnative range (11–13).
Reducing long-distance dispersal and further introductions of in-
vasive plants is important, even in areas where they already exist,
as this will limit future increases in genetic diversity and the for-
eration of new genotypes that confer an adaptive advantage in new
environments.

Methods

Study Overview. *P. lanceolata* is a short-lived [mean, max = 2.8, 8 y (48)],
perennial forb native to Europe. It reproduces sexually and vegetatively, with
gynodioecy, self-incompatibility, and protogyny to enhance outcrossing (49).
Flowers are wind-pollinated, and seeds mature in summer. The species occurs
in a wide range of habitats, including seminatural grasslands, roadsides, dis-
turbed sites, abandoned fields, and agricultural land (50). Seeds are dispersed
locally by wind, but seed dispersal distances are estimated to be within cen-
timeters or meters of the mother plant (51). Widespread propagule movement
by humans (29) and repeated introductions as seed contaminants (30)
have led to the global distribution of *P. lanceolata*. It has been present
in Australia since before 1850 (https://www.ala.org.au/), in North America
since before 1832 (30), and for an unknown time in South Africa (52). It is
cultivated as a commercial pasture plant in New Zealand because it grows
well in the mild winter and limits soil nitrification (31). The species is clas-
sified as an invasive taxon outside of its native range (53). We used
field-collected demographic and DNA data from populations of *P.
lanceolata* to analyze spatial variation in demographic rates and genetic
diversity. The demographic data were used to parameterize the simulation part
of the study (Hypothesis 1) and to analyze the demographic influence on
genetic diversity across global environmental gradients (Hypothesis 3). The dataset
comprises to assess call rate (mean = 99%), polymorphic information content
(mean = 22%).

Genotyping. Samples were genotyped at Diversity Arrays Technology P/L
(Canberra, Australia) using double-restriction enzyme complexity reduction and
high-throughput sequencing (DArTseq). Total genomic DNA was
extracted with a NucleoSpin 96 Plant II Core Kit (MACHEREY-NAGEL) and
purified using a Zymo kit (Zymo Research). The enzymes PstI and MseI were
chosen following tests of different enzyme combinations for *P. lanceolata*.
DNA samples were processed in digestion/reaction reactions following Kilian
et al. (57) but substituting the single PstI adaptor for two adaptors corre-
ponding to restriction enzyme-specific overhangs. The PstI adaptor was
modified to include Illumina sequencing primers and variable length
barcodes following Elshire et al. (58). Mixed fragments (PstI–MseI) were
amplified in 30 rounds of PCR using the following reaction conditions:
94 °C for 30 s, 58 °C for 30 s, and 72 °C for 45 s followed by 72 °C for 7 min. After PCR, equimolar amounts of
amplification products from each sample were bulked and applied to c-Bot
(Illumina) bridge PCR followed by single-read sequencing on an Illumina
HiSeq2500 for 77 cycles. Raw sequences were processed using DArTseq
analytical pipelines (DArtTdb) to split samples by barcode and remove
door-quality sequences. Genotypes for codominant SNPs were called de
novo (i.e., without a reference genome) from 69-bp sequences using
DArtTseq proprietary software (DArtTsoft). Replicate samples were pro-
cessed to assess call rate (mean = 79%), reproducibility (mean = 99%),
and polymorphic information content (mean = 22%).

SNP Filtering. Starting with 37,692 SNPs that passed DArTseq quality control, we
filtered the data for minimum minor allele frequency (1%), call rate (50%),
and high-throughput sequencing (DArTseq). Total genomic DNA was
targeted at Diversity Arrays Technology P/L (Canberra, Australia) using
double-restriction enzyme complexity reduction and high-throughput sequencing (DArTseq). Total genomic DNA was
extracted with a NucleoSpin 96 Plant II Core Kit (MACHEREY-NAGEL) and
purified using a Zymo kit (Zymo Research). The enzymes PstI and MseI were
chosen following tests of different enzyme combinations for *P. lanceolata*.
DNA samples were processed in digestion/reaction reactions following Kilian
et al. (57) but substituting the single PstI adaptor for two adaptors corre-
ponding to restriction enzyme-specific overhangs. The PstI adaptor was
modified to include Illumina sequencing primers and variable length
barcodes following Elshire et al. (58). Mixed fragments (PstI–MseI) were
amplified in 30 rounds of PCR using the following reaction conditions:
94 °C for 30 s, 58 °C for 30 s, and 72 °C for 45 s followed by 72 °C for 7 min. After PCR, equimolar amounts of
amplification products from each sample were bulked and applied to c-Bot
(Illumina) bridge PCR followed by single-read sequencing on an Illumina
HiSeq2500 for 77 cycles. Raw sequences were processed using DArTseq
analytical pipelines (DArtTdb) to split samples by barcode and remove
door-quality sequences. Genotypes for codominant SNPs were called de
novo (i.e., without a reference genome) from 69-bp sequences using
DArtTseq proprietary software (DArtTsoft). Replicate samples were pro-
cessed to assess call rate (mean = 79%), reproducibility (mean = 99%),
and polymorphic information content (mean = 22%).

Field Demographic Census and DNA Sampling. PLANTPOPNET is an ongoing
project that began in 2014, and annual censuses of *P. lanceolata*
populations are planned for the long term. Our analysis used data collected
between 2014 and 2017, but not all sites began data collection at the same
time (i.e., year 0 varied among sites) (SI Appendix, Table S1). In most pop-
ulations (61%), year 0 was 2015, and 73% of populations were sampled
twice during this study period (number of annual censuses per population =
1 to 3) (SI Appendix, Table S1). At each census site in year 0, a series of
adjacent 50 × 50-cm quadrats was established along transects until the
quadrats covered 100 individual plants. Researchers established transects
where *P. lanceolata* was present in sufficient numbers for demographic
studies, and therefore, density estimates might reflect upper estimates
across local populations. Quadrats were permanently marked to enable re-
peat censuses from year 1 onward. Each plant was individually tagged, and
all rosettes on each plant were measured according to a standard protocol (56),
which included leaf length, number of flowering stems, inflorescence
length, and stage of seed development.

At each site, fresh leaf tissue from seven to nine individuals was collected
and placed immediately in silica gel (SI Appendix, Table S1). Sampled indi-
viduals were close to (5 to 20 m) but outside of census plots and were sepa-
rated from each other by 5 to 10 m. Thus, we avoided damage to
permanently marked individuals in the census population, ensured that sam-
ples were closely related to the census population, and minimized the chance
of sampling clones. We included two samples each from one population of
*P. coronopus* (Spain) and four populations of *P. major* (Australia × 2, Ireland ×
1, Romania × 1) as outgroups. To investigate if naturally occurring populations
were influenced by genetic stock from commercial pasture lines, we included
nine individuals from each of three cultivar lines derived from *P. lanceolata*:
Agritonic, Ceres Tonic, and Tonic Plantain. The whole dataset thus included
491 individuals. The data are publicly available (https://zenodo.org/record/ 3626288).

SNP Filtering. Starting with 37,692 SNPs that passed DArTseq quality control, we
filtered the data for minimum minor allele frequency (1%), call rate (50%),
and high-throughput sequencing (DArTseq). Total genomic DNA was
targeted at Diversity Arrays Technology P/L (Canberra, Australia) using
double-restriction enzyme complexity reduction and high-throughput sequencing (DArTseq). Total genomic DNA was
extracted with a NucleoSpin 96 Plant II Core Kit (MACHEREY-NAGEL) and
purified using a Zymo kit (Zymo Research). The enzymes PstI and MseI were
chosen following tests of different enzyme combinations for *P. lanceolata*.
DNA samples were processed in digestion/reaction reactions following Kilian
et al. (57) but substituting the single PstI adaptor for two adaptors corre-
ponding to restriction enzyme-specific overhangs. The PstI adaptor was
modified to include Illumina sequencing primers and variable length
barcodes following Elshire et al. (58). Mixed fragments (PstI–MseI) were
amplified in 30 rounds of PCR using the following reaction conditions:
94 °C for 30 s, 58 °C for 30 s, and 72 °C for 45 s followed by 72 °C for 7 min. After PCR, equimolar amounts of
amplification products from each sample were bulked and applied to c-Bot
(Illumina) bridge PCR followed by single-read sequencing on an Illumina
HiSeq2500 for 77 cycles. Raw sequences were processed using DArTseq
analytical pipelines (DArtTdb) to split samples by barcode and remove
door-quality sequences. Genotypes for codominant SNPs were called de
novo (i.e., without a reference genome) from 69-bp sequences using
DArtTseq proprietary software (DArtTsoft). Replicate samples were pro-
cessed to assess call rate (mean = 79%), reproducibility (mean = 99%),
and polymorphic information content (mean = 22%).

Detecting Loci under Putative Selection. Neutrality was an assumption un-
derlying the population structure models that we used; thus, we investigated
if SNPs were putatively under selection using one population-level method

Smith et al.
PNAS Latest Articles | 7 of 10
BayesScan and two individual-level methods (PCAdapt and LFMM). BayesScan uses a Markov chain Monte Carlo algorithm to examine outlier loci against background values of population differentiation ($F_{ST}$) among predefined populations (65). PCAdapt and LFMM both define background population structure as $K$ principal components derived from individual genotypes (66, 67). In PCAdapt, each SNP is regressed against each principal component. LFMM uses the principal components as latent factors in a Gaussian mixed model, where the genotype matrix is modeled as a function of an environmental matrix (67). While BayesScan is suitable for our population-level sampling design, PCAdapt and LFMM are more reliable for species with complex, hierarchical population structure (e.g., multiple divergence events) and are less sensitive to admixed individuals and outliers in the data (68, 69). Thus, we considered outliers identified in any of the three methods to be putatively under selection.

For BayesScan, we set the prior odds at 200 (appropriate for the number of markers in our data (70)), ran the model using default parameters (100,000 iterations with a thinning interval of 10, a burn in of 50,000 and 20 pilot runs of 5,000 iterations), and checked the distribution of the log likelihood across iterations to ensure model convergence (SI Appendix, Fig. 55). For both individual-level methods, we examined scree plots to determine $K$ and used the first 10 components that captured the majority of population structure in the data (SI Appendix, Fig. 55). We defined the LFMM environmental matrix using the four 30-s BioClim variables described above and three additional variables: elevation (meters above sea level measured at the site) and two variables extracted from Climond (71) at 5-min resolution (annual mean moisture index and seasonality in moisture [coefficient of variation of annual mean moisture]). To control for false discovery rate, we calculated $q$ values from $P$ values and classed SNPs as outliers where $q < 0.05$ for BayesScan and PCAdapt and $q < 0.1$ for LFMM (to account for the small number of SNPs in the analysis). With this method, 678 outlier loci were identified across all four regions. We used linear mixed model to evaluate whether there was a difference in $DS$ between the native and nonnative range, with site fitted as a random effect.

To determine whether multiple introductions of $P$. lanceolata had occurred in nonnative regions (Australia, Japan, New Zealand, North America, and South Africa), we estimated the minimum number of propagules required to produce the observed level of genetic diversity in nonnative regions ($\text{Prop}_{\text{min}}$) (77). We defined the source population as all of Europe because nonnative individuals were usually composed of admixed genotypes from multiple European populations. For each nonnative region, we calculated the number of alleles not present in Europe and removed these from the reference panel of nonnative alleles. Individuals from the native range were then randomly cumulatively sampled without replacement. $\text{Prop}_{\text{min}}$ was the number of individuals sampled at the point when all alleles in the nonnative panel were represented (https://zenodo.org/record/3626288). We repeated the process 1,000 times to obtain a mean and SE. We also calculated the number of unique alleles in each of the 53 sites as a measure of uniqueness.

To assess the influence of environmental gradients on spatial genetic structure, we used generalized dissimilarity models (78, 79). We fitted one model for the native range, a second for the nonnative range, and a third for the global dataset (native and nonnative). We calculated genetic dissimilarity as $F_{ST}$ between all pairs of populations in GENEPop 4.6.8 (80). Environmental distances between all pairs of populations $i$ and $j$ were calculated from the four BioClim variables ($x_i − x_j$) (79). For each of the three datasets, we fitted geographic distance and all environmental distances as $F_{ST}$ predictors in a single global model (Table 1). The importance of each environmental variable was assessed by comparing the fitted model with 500 models with a permuted environmental matrix (79). Thus, the effect of each environmental variable can be interpreted independently, and differences in spatial scale are accounted for by the geographic distance variable. $P$ values were Bonferroni adjusted across all terms within each model. We used deviance explained to assess goodness of fit of the three models. Given sample size differences between the three datasets, we used a bootstrap estimate from 10,000 replicates of the deviance explained to assess the accuracy of the model fit. We assumed the deviance explained to be accurate if bootstrap 95% CI did not include zero.

Demographic and Dispersal Effects on Genetic Diversity (Hypothesis 3). We used linear regression to determine whether environmental influences on within-population genetic diversity could be explained by demographic variation and whether this effect would be weakened by mass dispersal into the nonnative range (Hypothesis 3). The observation level for all analyses was the population, and the number of observations was 44 (i.e., all populations with genetic and demographic data) (SI Appendix, Table S1).

Genetic diversity was calculated as allelic richness in hierfstat (81) separately for the neutral (18,166 SNPs) and adaptative (5,024 SNPs) datasets. Allelic richness was highly correlated with expected heterozygosity ($H_e$; $r = 0.98$), and because it was standardized for sample size, it eliminated a weak correlation that we observed between $H_e$ and sample size. We characterized the environment using the four BioClim variables. For demography, we used
three variables that can influence genetic diversity (Table 1): population density (rosettes per meter$^2$), fecundity, and empirical population growth rate. For fecundity, we used reproductive effort estimated as the rosette-level inflorescence length $\times$ number of flowering stems per meter$^2$. Empirical population growth rate was calculated as $r = \log(N_t/N_0)$, indicating the strength and direction of change in rosettes per meter$^2$ in the first 2 y of the study (for 38 of the 44 populations with 2 y of data) (SI Appendix, Table S1). Thus, $r$ reflects the combined influence of fecundity and survival (the variables explored in simulation Experiment 1). We used rosette-level data for all metrics to reduce potential observer bias in assessing clonality, but plant- and rosette-level metrics were highly correlated ($r = 0.94$). Fecundity was log transformed to address a strongly skewed distribution, and all predictors were standardized prior to analysis ($x = \text{mean}(x) \pm \text{SD}(x)$).

We tested environmental and demographic effects separately to determine which variables best described variation in genetic diversity. The analysis comprised two stages. First, we analyzed the effect of each environmental variable on genetic diversity. Here, we also modeled the environmental effect on demography (i.e., the three demographic variables as response terms) to establish a baseline for environmental influences on demographic rates. Second, we examined whether each demographic variable influenced genetic diversity. In both stages, we analyzed environmental and demographic interactions with range (native/nonnative). Because of data limitations ($n = 44$), it was not possible to fit complex models with multiple interaction terms, and therefore, we modeled each predictor separately.

To determine the importance of each environmental or demographic predictor, we used AICc to compare model fit across five alternative model forms: a null model (no predictor variation), a predictor-only model, a range-only model, an additive model (predictor + range), and an interactive (predictor $\times$ range). We considered a model to have support from the data if it improved the fit over the null model by $\Delta$AICc $> 2$ (82). Among models that outfitted the null, those within $\Delta$AICc $\leq 2$ of each other were considered to have equal support from the data. In these cases, we presented the top-ranked model in the main document and supported models in SI Appendix. To interpret interaction models in light of sample size differences between the native (30) and nonnative (14) ranges (e.g., a strong response in the native range and no response in the nonnative range), we obtained a bootstrap 95% CI from 10,000 bootstrap replicates of the interaction coefficient using the adjusted bootstrap percentile method.

ACKNOWLEDGMENTS. Jan van Groenendael helped design the PLANTPOPNET network. Leander Anderdeg, Lauchlan Fraser, Jennifer Gremer, Emily Griffoul, Adrian Oprea, Richard Shefferson, and Danielle Sherman provided data. Mavee Harrison assisted with field work. Valuable discussions with Alan Stewart and Andrej Kilar improved our knowledge of Plantago cultivation and SNP data generation, respectively. This research was supported by a Science Foundation Ireland grant to Y.M.B. (European Research Council Development Programme 15/ERC12803). A.L.S. was supported by a Marie Sklodowska-Curie Individual Fellowship (746191) under the European Union Horizon 2020 Programme for Research and Innovation. Additional support came from: Catalan Institution for Research and Advanced Studies (CREA) (Academia Award to S.M.-B.), Spanish Government (Ministerio de Economía y Competitividad BFU2015-64001-P/MINECO/FEDER to S.M.-B.), Estonian Ministry of Education and Research (Institutional Research Funding IUT20–29 to M.P.), European Regional Development Fund (Centre of Excellence EChange to M.P.), New Zealand Ministry for Business Innovation (Employment’s Strategic Science Investment Fund to R.G.) and Academy of Finland (285746 to S.R.).

Table 1. Demographic variables used to analyze population processes that are important to genetic diversity

<table>
<thead>
<tr>
<th>Demographic variable measured</th>
<th>Used as a proxy for</th>
<th>Relevance to genetic diversity</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic variable measured</td>
<td>Used as a proxy for</td>
<td>Relevance to genetic diversity</td>
<td>Formula</td>
</tr>
<tr>
<td>Density</td>
<td>Population size</td>
<td>Effective population size</td>
<td>$N_{i+1}/N_i$</td>
</tr>
<tr>
<td>Reproductive effort per unit area</td>
<td>Fecundity</td>
<td>Fitness</td>
<td>$N_{i+1}/N_i$</td>
</tr>
<tr>
<td>Empirical population growth rate</td>
<td>Combined effects of survival and fecundity</td>
<td>Fitness</td>
<td>$N_{i+1}/N_i$</td>
</tr>
</tbody>
</table>

The relevance of demographic variables to genetic diversity is outlined in Fig. 1 and described in detail by Ellegren and Galtier (4).