MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2455–2468 doi: 10.1111/j.1365-294X.2012.05486.x

Rapid evolution of an adaptive cyanogenesis cline in introduced North American white clover (*Trifolium repens* L.)

NICHOLAS J. KOOYERS and KENNETH M. OLSEN

Department of Biology, Washington University in St Louis, One Brookings Drive, Campus Box 1137, St. Louis, MO 63130, USA

Abstract

White clover is polymorphic for cyanogenesis (HCN production after tissue damage), and this herbivore defence polymorphism has served as a classic model for studying adaptive variation. The cyanogenic phenotype requires two interacting biochemical components; the presence/absence of each component is controlled by a simple Mendelian gene (*Ac/ac* and *Li/li*). Climate-associated cyanogenesis clines occur in both native (Eurasian) and introduced populations worldwide, with cyanogenic plants predominating in warmer locations. Moreover, previous studies have suggested that epistatic selection may act within populations to maintain cyanogenic (AcLi) plants and acyanogenic plants that lack both components (acli plants) at the expense of plants possessing a single component (Acli and acLi plants). Here, we examine the roles of selection, gene flow and demography in the evolution of a latitudinal cyanogenesis cline in introduced North American populations. Using 1145 plants sampled across a 1650 km transect, we determine the distribution of cyanogenesis variation across the central United States and investigate whether clinal variation is adaptive or an artefact of population introduction history. We also test for the evidence of epistatic selection. We detect a clear latitudinal cline, with cyanogenesis frequencies increasing from 11% to 86% across the transect. Population structure analysis using nine microsatellite loci indicates that the cline is adaptive and not a by-product of demographic history. However, we find no evidence for epistatic selection within populations. Our results provide strong evidence for rapid adaptive evolution in these introduced populations, and they further suggest that the mechanisms maintaining adaptive variation may vary among populations of a species.

Keywords: adaptive cline, balanced polymorphism, cyanogenesis, epistatic selection, parallel evolution, *Trifolium repens*

Received 20 October 2011; revision received 25 November 2011; accepted 6 December 2011

Introduction

Clinal variation, a shift in phenotypes or allele frequencies over geographic space, has frequently been used as a model to investigate how species respond to heterogeneous selection pressures in the presence of gene flow (e.g. Hiesey *et al.* 1942; Antonovics & Bradshaw 1970; Lenormand *et al.* 1999; Saccheri *et al.* 2008). Clines are useful model systems because the patchy complexity of landscapes is reduced into a simpler gradient where it is easier to associate adaptive traits and their underlying genetic variation with environmental variation. Recent studies have examined the genetic basis of adaptive clines in diverse traits, including pigmentation (Hoekstra *et al.* 2004; Mullen & Hoekstra 2008; Gross *et al.* 2009; Steiner *et al.* 2009; Rosenblum *et al.* 2010), morphology (Huey *et al.* 2000; McKechnie *et al.* 2010), life history (Caicedo *et al.* 2004) and physiology (Storz *et al.* 2007; Zhen & Ungerer 2008). These studies have

Correspondence: Kenneth M. Olsen, Fax: 314 935 4432;
E-mail: kolsen@wustl.edu

provided new insights into the spatial and temporal scales over which local adaptation can take place. Specifically, this research indicates that local adaptation can occur on short time scales (e.g. Umina *et al.* 2005; Saccheri *et al.* 2008), that adaptive clines can evolve in parallel in different portions of a species range (e.g. Oakeshott *et al.* 1982; Huey *et al.* 2000; Phifer-Rixey *et al.* 2008) and that this process of parallel cline evolution can occur through multiple, independent molecular mechanisms (e.g. Steiner *et al.* 2009; Rosenblum *et al.* 2010).

For populations occurring along an adaptive cline, genotype frequencies within a given population are potentially determined by two types of processes: those reflecting coarse-grained (regional-scale) factors and those reflecting fine-grained (within-population) dynamics. Coarse-grained factors operate in all adaptive clines and are the primary focus of most studies of clinal variation; these include interpopulation gene flow and selection associated with regional-scale environmental heterogeneity. Fine-grained factors that can affect adaptive polymorphism levels within populations of a cline include disruptive (diversifying) selection and directional selection corresponding to fine-scale spatial or temporal environmental heterogeneity (e.g. Weinig *et al.* 2003; Caicedo *et al.* 2004; Kroymann & Mitchell-Olds 2005; Todesco *et al.* 2010; reviewed by Charlesworth 2006; Hedrick 2006; Hudson & Kane 2009). One particular genetic mechanism by which polymorphism can be maintained within populations is epistatic selection, whereby selection favours specific allele combinations at unlinked loci. For example, selection for alternative flowering time phenotypes in Arabidopsis has been shown to arise through epistatic selection for specific allele combinations at two unlinked genes (Caicedo *et al.* 2004).

In addition to coarse- and fine-grained selective processes, clines can also potentially reflect nonadaptive processes, such as independent population introductions and/or range expansion (Clegg *et al.* 2002; Vasemagi 2006; Keller & Taylor 2008). Comparison of putatively selected genes with unlinked neutral markers can be used to assess the degree to which observed clinal variation is actually adaptive *vs.* a product of demographic history and resulting population structure.

Despite a well-developed theoretical literature on cline evolution (e.g. Fisher 1937; Haldane 1948; Endler 1977), very few studies have empirically examined the relative roles of coarse-grained and fine-grained selective factors in shaping allele or trait frequencies along adaptive clines. To do so requires a well-characterized adaptive polymorphism, preferably with a simple genetic basis, which also shows clinal variation. This study takes advantage of a well-studied polymorphism for cyanogenesis (HCN release in response to tissue damage) in white clover (*Trifolium repens* L.) to examine the roles of coarse-grained and fine-grained selective factors, as well as the potential role of demographic history, in shaping clinal variation.

Cyanogenesis in white clover

The white clover cyanogenesis polymorphism is considered a textbook example of selectively maintained adaptive variation (e.g. Dirzo & Sarukhan 1984; Silvertown & Charlesworth 2001). Cyanogenic and acyanogenic plants occur together in many natural populations, and the relative proportions of the two phenotypes correlate closely with environmental variation. Specifically, there is a higher frequency of cyanogenic plants in warmer climates, creating a pattern of climate-associated clines in both native (Eurasian) and introduced populations around the world. Previous studies have documented latitudinal cyanogenesis clines across Europe (Daday 1954a) and North America (Daday 1958), and altitudinal clines in the Swiss Alps (Daday 1954b), Wales (de Araujo 1976), the French Alps (Till-Bottraud *et al.* 1988), the Himalayas (Majumdar *et al.* 2004) and the Pacific Cascades (Ganders 1990). Although white clover occurs in temperate regions worldwide today, it has been introduced into most of its current range within the last 500 years. Given that this is a potentially long-lived perennial species, the recurrent and rapid emergence of cyanogenesis clines suggests intense climate-related selection associated with the cyanogenesis polymorphism. However, the population structure of white clover in these regions is unknown, and a formal comparison to neutral genetic variation is necessary before concluding that cyanogenesis clinal variation is adaptive.

It is generally hypothesized that white clover cyanogenesis clines are products of an ecological trade-off between selection for resistance to herbivores and selection against the costs of cyanogenesis in colder climates (reviewed by Hughes 1991). Ecological studies conducted over the last several decades have provided substantial evidence that cyanogenic plants are differentially protected from small generalist herbivores, including slugs, snails, insects and voles (e.g. Ennos 1981; Dirzo & Harper 1982a,b; Kakes 1989; Pederson & Brink 1998; Saucy *et al.* 1999; Viette *et al.* 2000). The predominance of acyanogenic plants in colder climates has been proposed to reflect the decreased benefits of energetic investment in chemical defence in areas of low herbivore pressure (Kakes 1989). Consistent with this hypothesis, acyanogenic plants in greenhouse and field experiments have been reported to flower earlier

and more prolifically than cyanogenic plants in the absence of herbivores (Daday 1965; Kakes 1989), suggesting a competitive advantage under these conditions. An alternative 'autotoxicity' hypothesis proposes that cyanogenic plants may be directly selected against in colder climates, because frost-induced tissue damage could lead to cyanide poisoning within the plant (Daday 1965; Brighton & Horne 1977; but see Olsen & Ungerer 2008).

The cyanogenic phenotype in clover requires two components that are separated in intact tissue and brought together with tissue damage that causes cell rupture: cyanogenic glucosides (stored in the vacuoles of photosynthetic tissue) and their hydrolysing enzyme linamarase (stored in the cell wall) (reviewed by Hughes 1991). Acyanogenic phenotypes can arise through the absence of either or both of the two required components. Classical genetics studies (e.g. Corkill 1942; Atwood & Sullivan 1943) revealed that two independently segregating Mendelian genes can account for these biochemical polymorphisms. The gene *Ac* describes the presence/absence of cyanogenic glucosides, and *Li* describes the presence/absence of linamarase. For both genes, a homozygous recessive genotype corresponds to the absence of the cyanogenic component. Thus, plants with at least one dominant (functional) allele at both *Ac* and *Li* are cyanogenic (the AcLi cyanogenesis phenotype, or 'cyanotype'), while the other cyanotypes (acLi, Acli and acli) lack one or both components and correspond to acyanogenic phenotypes. The molecular bases of both biochemical polymorphisms have been recently described: *Ac/ac* corresponds to a gene presence/absence polymorphism at the locus *CYP79D15,* which encodes the P450 protein catalysing the first step in the cyanogenic glucoside biosynthetic pathway (Olsen *et al.* 2008); *Li/li* corresponds to an unlinked gene presence/absence polymorphism at *Li*, which encodes the linamarase glycoprotein precursor (Olsen *et al.* 2007). These loci are known to occur in different linkage groups, and genetic mapping is in progress (K. Olsen, unpublished observations).

The fact that the clover cyanogenesis polymorphism arises through two independently segregating simple Mendelian genes provides a unique opportunity to study the role of epistatic selection in the evolution of this adaptive polymorphism. Plants that produce one cyanogenic component without the other may be maladaptive because they bear some of the energetic costs of cyanogenesis but without the herbivore defence benefits (Kakes 1989). Thus, in polymorphic populations, there may be epistatic selection favouring plants that are either cyanogenic (AcLi cyanotype) or lacking in both cyanogenic components (acli cyanotype) at the expense of acLi and Acli cyanotypes. Consistent with this hypothesis, some polymorphic mid-cline populations in Europe have been reported to show strong linkage disequilibrium (LD) between the two unlinked cyanogenesis genes, with a statistical excess of AcLi and acli cyanotypes (i.e. positive LD) among adult plants (Ennos 1982; Kakes 1987); this pattern potentially suggests pervasive epistatic selection against Acli and acLi plants. If epistatic selection were a general phenomenon affecting cyanotype frequencies in white clover populations, this fine-grained selective force would be predicted to generate positive LD among adult plants in most polymorphic populations.

In this study, we examine a continent-wide latitudinal transect of introduced U.S. white clover populations to investigate the roles of coarse-grained environmental heterogeneity, fine-grained epistatic selection and demographic history in shaping cyanogenesis cline evolution. We address the following questions: (i) Is the U.S. latitudinal cyanogenesis cline adaptive, or is it an artefact of white clover's introduction history into North America? and (ii) If the cline is adaptive, can cyanogenesis frequencies in U.S. populations be attributed solely to coarse-grained dynamics (interpopulation gene flow and regional-scale environmental gradients), or does epistatic selection within populations also play a role? We find evidence for striking clinal variation in these introduced populations that is not a by-product of demographic history. However, unlike populations in the native European range, we find no evidence for epistatic selection or other fine-scale selective processes within these North American populations. These findings suggest that coarse-grained factors alone are sufficient to maintain the *Ac/ac* and *Li/li* adaptive cyanogenesis polymorphisms in North America.

Materials and methods

Sampling

Trifolium repens has been widely introduced from its native Eurasian range into mesic temperate regions worldwide (Kjærgaard 2003). It is cultivated as a forage crop and also occurs commonly in lawns, roadsides and other mowed or grazed areas. The species is a self-incompatible perennial, which spreads vegetatively through stolons. For this study, introduced white clover populations were sampled along a latitudinal transect spanning the central United States (Fig. 1). Seven regions were sampled: New Orleans, LA; Jackson, MS; Grenada, MS; Memphis, TN; Cape Girardeau, MO; St. Louis, MO; and Wausau, WI. Collections in each region were made within a geographical area of approximately 20×20 km (except in the New Orleans region where sampled sites broadly surround Lake

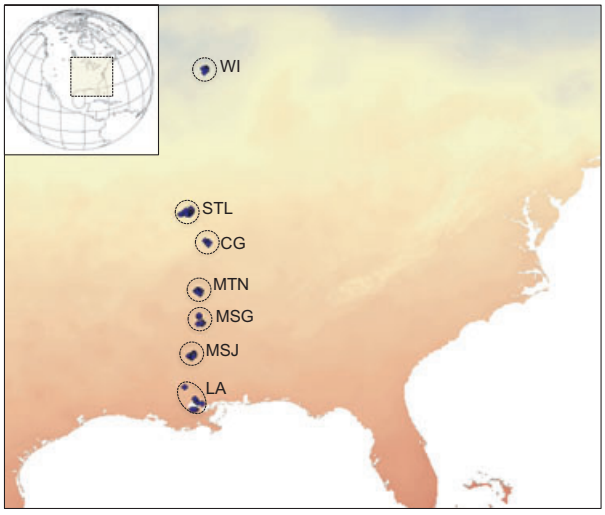

Fig. 1 Locations of sampled clover populations. Each dot corresponds to a sampled subpopulation within one of seven populations (south to north): New Orleans, LA (LA); Jackson, MS (MSJ); Grenada, MS (MSG); Memphis, TN (MTN); Cape Girardeau, MO (CG); St. Louis, MO (STL); Wausau, WI (WI). The blue to red gradient indicates minimum winter temperature as determined from the BIOCLIM data set (Hijmans *et al.* 2005).

Pontchartrain). In each region, plants were initially collected in 15 separate sampling sites ('subpopulations'), each site encompassing an area of approximately 50 × 50 m; ten stolon cuttings were collected per subpopulation, yielding 150 individuals per region and 1050 individuals (Table S1, Supporting information). An additional 15 subpopulations were later sampled in the St. Louis region, yielding a total of 120 subpopulations (1200 individuals in total, 1145 of which survived transplantation and were used in analyses). Most sampling sites were in public parks, schools and athletic fields; within each sampling site, sampled individuals were separated by at least 3 m to prevent the collection of multiple ramets from a single genet (Gliddon & Saleem 1985). Plants from New Orleans, St. Louis and Wausau were collected in early summer 2008, and remaining regions were collected in early summer 2009. Collected stolon cuttings were planted in MetroMix360 soil medium in 4'' pots and maintained on a mist bench in the greenhouse of Washington University for 3–5 days, after which plants were grown under standard greenhouse conditions. Because we observed no genetic differentiation among subpopulations within regions (see Results), we refer to groups of subpopulations within regions as 'populations' hereafter.

Geographical distances among subpopulations were calculated from GPS coordinates for each sampling site using a modified Haversine Formula as implemented in GenAlEx 6 (Peakall & Smouse 2006). Distances among populations were calculated by the same method but using the geographical mean centre of subpopulations for each population, as determined with the MeanCenter function in ArcMap 9.3.1 (ESRI, Redlands, CA). Because cyanogenesis allele frequencies may be more closely correlated with local climate than latitude (Daday 1965; Hughes 1991), we also estimated minimum winter temperature for each subpopulation using GIS data (extracted from bio6 in the BIOCLIM dataset; Hijmans *et al.* 2005). Minimum winter temperature for each population was calculated by averaging values across subpopulations.

Cyanogenesis gene analyses

Cyanotypes were determined for collected plants using Feigl-Anger HCN detection assays (Feigl & Anger 1966), following the protocol described by Olsen *et al.* (2007, 2008). Briefly, fresh leaves were collected from each plant to provide six approximately identical leaf tissue samples (typically three leaves used per sample). Pairs of duplicate samples were then tested for HCN production under one of three treatments: leaf tissue tested alone (to identify AcLi cyanotypes); leaf tissue plus exogenous cyanogenic glucoside solution (to identify acLi cyanotypes); or leaf tissue plus exogenous linamarase solution (to identify Acli cyanotypes). Leaves were frozen, thawed and mashed with a pipet tip to ensure cell rupture. Samples were incubated with test paper for four hours at 37 °C, as longer incubations can potentially generate false positives for acLi cyanotypes (resulting from the gradual hydrolysis of exogenously added cyanogenic glucosides; data not shown). We tested for experimental error by rephenotyping 40 plants and confirmed complete reproducibility. Any plants showing discrepancies in results between the two duplicate samples for each test treatment were retested until we were confident of the cyanotype. We also tested whether Feigl-Anger results were appropriate proxies for the presence/absence of the *CYP79D15* and *Li* loci using PCR assays for the presence/absence of each gene (see Olsen *et al.* 2007, 2008 for PCR protocols); a sample of 54 plants representing all cyanotypes was tested. We found a false-negative error rate (i.e. failure to detect a cyanogenic component despite the presence of the corresponding gene) of <5% (data not shown). Thus, for the rest of this study, we use the classical gene terminology (*Ac/ac* and *Li/li*) with the assumption that the dominant and recessive alleles correspond at the molecular level to the presence and absence of each cyanogenesis gene.

Cyanotype frequencies (AcLi, Acli, acLi and acli) and inferred *Ac* and *Li* genotype and allele frequencies were calculated for individual subpopulations and for popu-

lations. Frequencies of homozygous recessive genotypes (*acac* or *lili*) can be calculated directly from cyanotype data, as they correspond to the absence of individual cyanogenic components. Allele and genotype frequencies for *Ac/ac* and *Li/li* were inferred using homozygous recessive genotype frequencies under the assumption of Hardy–Weinberg equilibrium ($p^2 + 2pq + q^2 = 1$, where q^2 is the frequency of homozygous recessive genotypes). Calculations were performed both by hand and using GenAlEx 6.

Neutral marker analyses

To assess population structure across the sampled transect, we genotyped a subset of individuals from each population using primer pairs for nine microsatellite loci that have been found to be polymorphic in white clover (C. Brummer, pers. comm.; Noble Foundation, Ardmore, OK) (Table S2, Supporting information). We selected 4–7 subpopulations for each population at random and genotyped all sampled individuals that survived transplantation within them, yielding approximately 50 individuals genotyped per population (349 individuals in total, representing 38 subpopulations; Table S3, Supporting information). DNA was extracted with DNeasy Plant Mini kits (Qiagen, Valencia, CA) following the manufacturer's protocols using 50 mg young leaf tissue from each plant. PCR conditions followed the protocol developed by Brummer and colleagues (Noble Foundation), including use of a PIGtail (Brownstein *et al.* 1996) and an M13 fluorescently labelled tag (Schuelke 2000). Amplified products were multiplexed and diluted to 1:300 before being genotyped on an ABI 3100 sequencer at Washington University in St. Louis. All microsatellite alleles were scored using a ROX 400HD ladder, and the resulting fragment lengths were analysed using GeneMapper version 3.7 (Applied Biosystems, Foster City, CA). We hand-checked every allele call and coded any sample with more than two strong peaks as missing data. Microsatellite reactions that did not amplify were repeated up to three times before being coded as missing data. All loci had <5% missing data.

Genotypes were exported from GeneMapper and converted to GenePop format using GMCONVERT 0.32 (Faircloth 2006). Population tags were added to the GenePop file, and CONVERT 1.31 (Glaubitz 2004) was used to produce Arlequin and revised GenePop files. CONVERT 1.31 was also used to tabulate the number of private alleles in each subpopulation. We used MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.* 2004) to check for the evidence of null alleles within each population via 1000 Monte Carlo simulations of the alleles in each population. Further, we checked for deviations from Hardy–Weinberg equilibrium using exact tests implemented in Arlequin 3.5.1.2 (Excoffier *et al.* 2005). We used 95% confidence intervals followed by Bonferroni corrections as criteria for significant evidence of null alleles.

To assess the distribution of neutral variation among subpopulations and populations, we used both Fst-like and Rst-like AMOVA settings in Arlequin 3.5.1.2 (Excoffier *et al.* 2005) and tested for population differentiation based on a frequency distribution created from 1000 permutations of the data. For all subpopulations and populations, we also calculated pairwise Fst (Reynolds *et al.* 1983) and Rst (Slatkin 1995) in Arlequin 3.5.1.2 and pairwise φpt using GenAlEx 6. These measures all provided quantitatively similar estimates of population differentiation (see Results). Unlike Fst and Rst, φpt can be calculated for both codominant markers (such as microsatellites) and dominant markers (such as the cyanogenesis genes), and so we focused on this measure of differentiation in the comparisons described elsewhere.

Comparison of population differentiation at neutral markers and cyanogenesis genes

To test for signatures of isolation by distance in microsatellite data and the cyanogenesis genes, we performed Mantel tests and partial Mantel tests using the program zt (Bonnet & Van de Peer 2002). Statistical significance of correlations between matrices was determined through Fisher's exact tests following 100 000 permutations of the matrices (see Bonnet & Van de Peer 2002 for exact methods and formulas). We tested for isolation by distance in neutral markers by comparing the geographical distance matrix between subpopulations with a linearized pairwise φpt matrix computed from microsatellite genotype frequencies. We investigated clinal variation in the cyanogenesis genes by comparing the geographical distance matrix with linearized pairwise φpt matrices constructed from cyanotype frequencies and from the individual *Ac/ac* and *Li/li* genes. Because we did not have microsatellite genotypes for every individual for which we had cyanotype data, we used only those populations where we had both microsatellite genotypes and cyanotypes (38 subpopulations in total). To explicitly test for greater isolation by distance at the cyanogenesis genes compared to neutral markers, we used partial Mantel tests to test for correlations between geographical distance and differentiation at cyanogenesis loci while controlling for neutral marker genetic distances. We also used a matrix of pairwise differences in minimum winter temperature in place of the geographical distance matrix for comparison to the microsatellite and cyanogenesis gene matrices. Further, to test for any evidence of adaptive divergence at *Ac/ac* and *Li/li* with-

out regard to geography, we compared pairwise φpt values between the cyanogenesis genes and the microsatellite loci. Under neutral expectations, microsatellite pairwise φpt values should be equal to pairwise φpt values at each cyanogenesis gene; however, if the two cyanogenesis genes are under selection for adaptive divergence, we expect greater values for the cyanogenesis genes.

Testing for epistatic selection

To investigate the potential role of epistatic selection in shaping cyanotype frequencies, we tested for linkage disequilibrium (LD) between the *Ac/ac* and the *Li/li* genes. A statistically significant excess of AcLi and acli cyanotypes and a corresponding deficit of Acli and acLi cyanotypes (i.e. positive LD) would be consistent with the patterns of epistatic selection previously reported for some white clover populations (Ennos 1982; Kakes 1987). Because LD between unlinked genes can also arise as an artefact of pooling subpopulations that are genetically differentiated, we tested for LD both at the population level (pooling across subpopulations) and within individual subpopulations. LD between *Ac* and *Li* was calculated from cyanotype frequencies using Hill's (1974) maximum likelihood estimator ($\hat{D}$), following the method employed previously in white clover (Ennos 1982; Kakes 1987). This method for calculating LD does not depend on our inferred allele or genotype frequencies, only on the empirically observed cyanotype frequencies.

Results

Cyanogenesis genes show latitudinal clinal variation across the United States

Both the *Ac/ac* and *Li/li* genes were polymorphic in most subpopulations sampled in the seven surveyed populations (Table 1; see also Table S1, Supporting information). Across the sampled transect, a strong latitudinal cline is apparent in the frequency of cyanogenic plants; AcLi cyanotype frequencies were 85.8% in the New Orleans, LA population and steadily declined to 11.0% in the Wausau, WI population (Table 1; Fig. 2). Among the three acyanogenic cyanotypes, weak clinal variation is apparent for the Acli and acli cyanotypes, with both showing an increase in frequency towards higher latitudes. In contrast, the frequency of acLi cyanotypes is highest near the centre of the transect and declines to the north and south (Fig. 2).

We used linear regressions to investigate spatial distributions of the *Ac/ac* and *Li/li* allele frequencies and the cyanogenic phenotype (AcLi cyanotype) as a function of sampling site across the transect. Correlations are highly statistically significant ($P < 0.001$ in all cases), both for regressions against geographical distance along the transect and against minimum winter temperature (Figs 3 and S1, Supporting information). As a further assessment of clinal variation at the cyanogenesis genes, we performed Mantel tests to compare matrices of pairwise differences between sampling sites (using either geographical location or minimum winter temperature) with matrices of pairwise differentiation at the cyanogenesis genes (using φpt values for *Ac/ac*, *Li/li*, or cyanotype). All comparisons yielded statistically significant associations ($P < 0.05$ in all tests; Table 2).

Microsatellite variation and population structure

All microsatellite loci were polymorphic in every subpopulation sampled (Table S3, Supporting information). The number of alleles per locus ranged from 3 to 16, with 90 alleles detected in total. Additional summary statistics are presented in Table S3 (Supporting information). Not all loci were in Hardy–Weinberg equilibrium within every population. After implementing Bonferroni corrections for multiple tests, four of nine

Table 1 Summary of cyanogenesis polymorphism data by population

Region	*N*	Latitude (°N)	Average minimum winter temp. (°C)	AcLi cyanotype frequency	*Ac* allele frequency	*Li* allele frequency	$\hat{D}$
New Orleans, LA (LA)	141	30.22	4.59	0.86	0.72	0.73	−0.01 (ns)
Jackson, MS (MSJ)	146	32.36	1.22	0.47	0.46	0.41	−0.02 (ns)
Grenada, MS (MSG)	142	33.83	−0.8	0.42	0.33	0.44	−0.05 (ns)
Memphis, TN (MTN)	147	35.02	−1.26	0.33	0.29	0.43	0.00 (ns)
Cape Girardeau, MO (CG)	135	37.31	−5.36	0.24	0.34	0.27	0.03 (ns)
St. Louis, MO (STL)	298	38.65	−6.76	0.15	0.24	0.16	−0.01 (ns)
Wausau, WI (WI)	136	44.93	−6.77	0.11	0.30	0.13	0.01 (ns)

Allele frequencies for *Ac* and *Li* were calculated with the assumption of Hardy–Weinberg equilibrium within populations. Linkage disequilibrium values were calculated directly from observed cyanotype frequencies. No linkage disequilibrium values are statistically significantly different from zero.

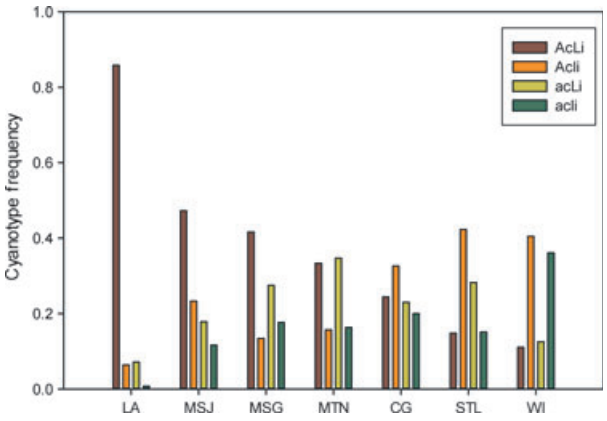

Fig. 2 Cyanotype frequencies in each population along the sampled U.S. latitudinal transect. Population abbreviations correspond to Fig. 1.

loci displayed an excess of homozygotes in at least 15% of the subpopulations surveyed. Analysis with MICRO-CHECKER (Van Oosterhout *et al*. 2004) indicated that this excess was consistent with the presence of null alleles at these loci (Tables S3 and S4, Supporting information). We therefore ran all subsequent analysis with two data sets, the first including all microsatellites and the second excluding the four loci with greatest evidence of null alleles (TRSSRB03C06, TRSSRA05E05, TRSSRA04G02, TRSSRA02B11; see Table S2, Supporting information). There were no qualitative differences in any analysis, and the quantitative differences did not affect any of the conclusions of this study. Further, when we used the null allele correction method of Chapuis & Estoup (2007), we observed no differences in downstream analysis. Results reported below are for all nine microsatellite loci.

AMOVAs using both Rst-based genetic distances (which incorporate information on repeat number differences among alleles) and Fst-based measures (based on allele identity only) reveal that there is little overall geographical structure in neutral marker variation across the sampled North American transect. Almost all of the variation is distributed within individual subpopula-

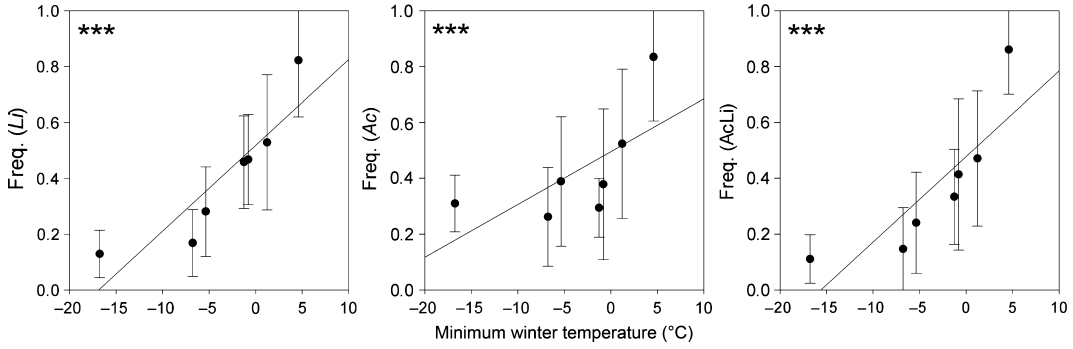

Fig. 3 Regression of population frequencies of the *Li* allele (A), *Ac* allele (B) or the AcLi cyanotype (C) against average minimum winter temperature for the region where samples were collected. Statistical significance of linear regressions is indicated as follows: *$0.01 < P < 0.05$, **$0.01 < P < 0.001$ and ***$P < 0.001$.

Table 2 Results of Mantel and partial Mantel tests comparing differentiation at the cyanogenesis genes or microsatellite loci with geographical and climatic data for sampled populations

Matrix 1	Matrix 2	Control Matrix	*r*	*P*
Geography	*Ac/ac*	—	0.239	0.0017
Geography	*Li/li*	—	0.345	0.0003
Geography	AcLi cyanotype	—	0.303	0.0020
Geography	Microsatellites	—	0.483	0.0001
Minimum Winter Temperature	*Ac/ac*	—	0.203	0.0051
Minimum Winter Temperature	*Li/li*	—	0.335	0.0004
Minimum Winter Temperature	AcLi cyanotype	—	0.280	0.0008
Minimum Winter Temperature	Microsatellites	—	0.476	0.0001
Geography	*Ac/ac*	Microsatellites	0.153	0.0152
Geography	*Li/li*	Microsatellites	0.185	0.0047
Geography	AcLi cyanotype	Microsatellites	0.164	0.0151
Minimum Winter Temperature	*Ac/ac*	Microsatellites	0.112	0.0570
Minimum Winter Temperature	*Li/li*	Microsatellites	0.176	0.0083
Minimum Winter Temperature	AcLi cyanotype	Microsatellites	0.139	0.0363

Table 3 Hierarchical analysis of molecular variance (AMOVA) based on nine microsatellite markers

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	6	932.05	1.23	5.19***
Among subpopulations within populations	31	1023.80	0.6	2.54
Within subpopulations	660	14478.88	21.93	92.27***
Total	697	16434.72	23.78	

Variances were calculated using the 'Rst-like' AMOVA option in Arlequin 3.5.1.2 (Excoffier *et al*. 2005). Asterisks denote statistical significance of variation components as follows: *$0.01 < P < 0.05$, **$0.01 < P < 0.001$ and ***$P < 0.001$.

tions rather than among subpopulations within populations or among populations (92.27% for Rst-based AMOVA; Table 3). Within populations, there is no more variation partitioned among subpopulations than would be expected by chance, indicating an absence of any detectable population structure at this geographical scale (Table 3). Lack of population structure within populations is also evident from very low pairwise genetic distances between subpopulations within populations (mean φpt = 0.030). Pairwise genetic distances among all subpopulations across the entire transect also reveal very low differentiation, even between sites located more than 1600 km apart (mean Rst = 0.066, mean φpt = 0.050; see Table S4, Supporting information). Nonetheless, the AMOVAs reveal significantly more variation distributed among populations than would be expected by chance (5.19%, $P < 0.0001$ for Rst-based AMOVA; Table 3), a pattern consistent with some degree of population differentiation on a continent-wide scale. In addition, we find a statistically significant correlation between pairwise genetic distances (φpt) among subpopulations and geographical distances (Mantel's $r = 0.473$, $P = 0.01$), indicating some isolation by distance across the transect.

To test whether the isolation by distance detected at the microsatellite markers is sufficient to account for the cyanogenesis gene clinal variation, we used partial Mantel tests, where a pairwise geographical distance matrix (or a pairwise minimum winter temperature difference matrix) was compared with pairwise linear φpt matrices from *Ac/ac*, *Li/li* and AcLi cyanotype while controlling for a pairwise linear φpt matrix from microsatellite data. We find that after controlling for population structure, there is still statistically significant clinal variation in *Ac* (Mantel's $r = 0.15$, $P = 0.015$), *Li* (Mantel's $r = 0.19$, $P = 0.005$) and the AcLi cyanotype (Mantel's $r = 0.16$, $P = 0.015$) (Table 2, Fig. 4). Similarly, we find that pairwise φpt values among subpopulations are statistically significantly greater for the cyanogenesis genes than for microsatellite loci (*Ac*: $t = 16.0$, d.f. = 702, $P < 0.0001$; *Li*: $t = 17.6$, d.f. = 702, $P < 0.0001$; Fig. S2, Supporting information), consistent with selection favouring adaptive divergence at the cyanogenesis genes. These findings indicate that the strong clinal variation observed at the cyanogenesis genes is not an artefact of demographic history.

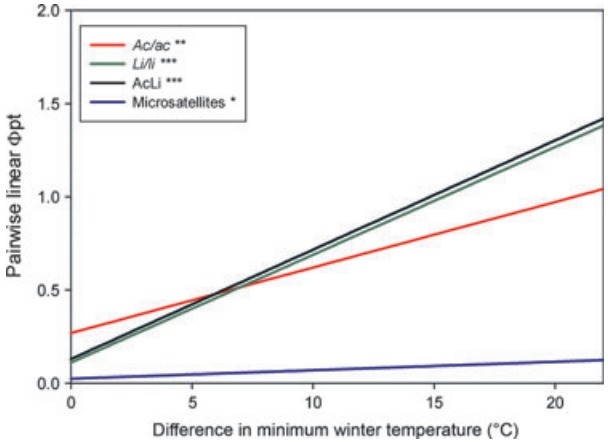

Fig. 4 Regressions of linear pairwise φpt values between sampling sites for microsatellite data, *Ac/ac*, *Li/li* and AcLi cyanotype against pairwise minimum winter temperature data. Statistical significance of regressions is indicated as follows: *$0.01 < P < 0.05$, **$0.01 < P < 0.001$ and ***$P < 0.001$.

No evidence for epistatic selection within populations

To assess the potential role of epistatic selection in maintaining the cyanogenesis polymorphism, we measured the degree of linkage disequilibrium (LD) between the *Ac/ac* and *Li/li* genes (calculated as $\hat{D}$; Hill 1974) for each sampled population and for individual subpopulations within populations. In contrast to native European populations, where a positive association between the dominant *Ac* and *Li* alleles has been reported for some populations (Ennos 1982; Kakes 1987), none of the North American populations have $\hat{D}$ values that are statistically significant from zero (Table 1). At the level of individual subpopulations, 11 of 120 sampled sites show statistically significant values of $\hat{D}$; however, all of these have an excess of Acli and acLi cyanotypes (i.e. negative linkage disequilibrium

(LD) between *Ac* and *Li*; Table S1, Supporting information), which is the opposite of the pattern expected under the epistatic selection hypothesis proposed for white clover.

Discussion

The white clover cyanogenesis polymorphism has long served as a model system for understanding how adaptive variation is maintained within and between populations. Here, we have investigated the evolutionary processes underlying the establishment and maintenance of cyanogenesis clinal variation in introduced North American populations. In the native Eurasian species range, both adaptation in response to coarse-grained spatial environmental heterogeneity and fine-grained epistatic selection have been proposed to maintain the *Ac/ac* and *Li/li* cyanogenesis polymorphisms (Daday 1954a, 1958, 1965; Ennos 1982; Kakes 1987). We find that while the North American latitudinal cline is consistent with adaptation and is not an artefact of introduction history (Figs 2 and 4; Table 3), there is no evidence for within-population epistatic selection as a factor in the maintenance of cyanogenesis variation (Tables 1 and S3, Supporting information). Thus, our findings suggest that coarse-grained spatial environmental heterogeneity alone is sufficient to explain the origin and persistence of the cyanogenesis polymorphism in the United States. Below, we discuss the implications of our findings for studies of adaptive cline evolution and specifically for the clover cyanogenesis polymorphism.

Rapid adaptive cline evolution

White clover was introduced into the United States with European colonization within the last 500 years. Thus, the continent-wide cyanogenesis cline documented in this study has evolved in this perennial species in a relatively few generations. *Trifolium repens* is an obligately outcrossing, self-incompatible species, and, as a common human commensal, it has an exceedingly large effective population size in North America (Kjærgaard 2003). These characteristics would all be expected to facilitate the rapid evolution of adaptive clinal variation in this species. A number of other studies have also documented the establishment of adaptive clines in relatively short periods of time (Antonovics & Bradshaw 1970; Huey *et al.* 2000; Umina *et al.* 2005; Dlugosch & Parker 2008; Montague *et al.* 2008; Saccheri *et al.* 2008). Like white clover, these species generally have large effective population sizes, little population structure and outcrossing mating systems (e.g. *Drosophila melanogaster*). All of these characteristics are expected to lead to efficient responses to selection within populations (Charlesworth 2009), as long as interpopulation gene flow is not so great as to overwhelm local adaptation (Endler 1977).

Our study is not the first to document the North American white clover cyanogenesis cline. Daday (1958) surveyed 3203 plants from 38 populations between southern Louisiana and Alaska, 36 of which were polymorphic at both *Ac* and *Li*. That study revealed similar clinal variation to the patterns we observe, although Daday observed universally higher *Ac* and *Li* allele frequencies at a given minimum winter temperature, as well as greater variation in cyanogenesis frequencies among the northern populations that he sampled.

It is important to note that observations of clinal variation in putatively adaptive traits do not by themselves demonstrate adaptive cline evolution. For recently established populations in particular, the history of population introduction can easily create clines that are reflections of demographic history rather than local adaptation (Cavalli-Sforza *et al.* 1993; Clegg *et al.* 2002; Vasemagi 2006; Keller & Taylor 2008; Keller *et al.* 2009). In the case of white clover, the close correlation between climate and cyanogenesis in the native Eurasian range could have led to the differential introduction of acyanogenic, cold-adapted clover from northern Europe into the northern United States and warm-adapted, cyanogenic clover from the Mediterranean into the southern United States. If this introduction pattern did in fact occur, a near-panmictic level of subsequent gene flow has all but eliminated any initial population structure (see Tables 3 and S4, Supporting information). Thus, the cyanogenesis cline observed across our sampled latitudinal transect (Fig. 2) can be most easily attributed to local adaptation along the latitudinal gradient.

Historically, the presence of similar phenotypic clines, either in different geographical portions of a species range or in several different species over the same range, has been taken as evidence of adaptation (e.g. Endler 1986; Huey *et al.* 2000). While several other climate-associated cyanogenesis clines have been described in white clover populations around the world (Daday 1954a,b, 1958; de Araujo 1976; Till-Bottraud *et al.* 1988; Ganders 1990; Majumdar *et al.* 2004), most of this documentation occurred before the advent of molecular techniques that could be used for assessing population structure with neutral markers. We suggest that understanding population structure is one of the necessary prerequisites for drawing any definitive conclusions about adaptive divergence in these populations.

No evidence for epistatic selection or other fine-grained processes within North American populations

If epistatic selection were acting to maintain cyanogenesis variation within North American white clover

populations, as has previously been proposed for other populations (Ennos 1982; Kakes 1987), one would expect a statistically significant excess of AcLi and acli cyanotypes relative to Hardy–Weinberg expectations (i.e. positive LD between *Ac* and *Li*). We reject epistatic selection as a mechanism for maintaining polymorphism within our sampled populations: we find no evidence for significant LD in any population (Table 1). Similarly, LD is not significantly different from zero in most subpopulations sampled, and those showing significant $\hat{D}$ values indicate negative rather than positive LD (Table S1, Supporting information); this pattern is inconsistent with the epistatic selection hypothesis proposed for white clover.

We also find no evidence of a role for other fine-scale factors in the maintenance of the cyanogenesis polymorphism. For example, if fine-scale spatial environmental heterogeneity were favouring adaptive divergence among subpopulations within populations, subpopulations within a given population would be expected to show some degree of adaptive divergence at the cyanogenesis genes. However, no differentiation is detected on this spatial scale (Table S1, Supporting information). Moreover, when subpopulations within populations are pooled, we find no evidence of LD between the cyanogenesis genes, which would be predicted with the pooling of subpopulations that are adaptively diverged (Nordborg & Tavare 2002). Thus, if there are any fine-scale spatial differences in selection among the clover subpopulations, the selection is either too weak or the gene flow too great to allow for local adaptation.

In order to compare our LD observations to previous studies, we compiled data from the literature documenting cyanogenesis polymorphisms in white clover populations from around the world, and we calculated $\hat{D}$ for these populations. Across nine studies, which collectively examined cyanogenesis variation in 14 world regions, 12 of the surveyed regions contain one or more populations showing statistically significant LD (Table 4). Among these 12 regions, 11 show a bias towards positive LD, consistent with epistatic selection for AcLi and acli cyanotypes. However, fewer than one-third of polymorphic populations show statistically significant positive LD (29.4% of populations on average across the 11 regions), and only two regions show a majority of populations with positive LD (Table 4). To test whether our LD estimates for these other studies might be biased downwards by the inclusion of populations with low polymorphism levels, we also calculated $\hat{D}$ using only those populations where all *Ac/ac* and *Li/li* alleles were all present at a frequency of at least 0.15. This recalculation does not result in a statistically significant increase in the percentage of populations showing positive LD (*t*-test; $P > 0.05$). Comparisons of native vs. introduced populations also do not reveal any obvious patterns with respect to LD; proportions of

Table 4 Summary of linkage disequilibrium detected between *Ac/ac* and *Li/li* within polymorphic populations in 10 studies. Positive LD corresponds to an excess of AcLi and acli cyanotypes over Hardy–Weinberg expectations

Study	Area*	No. polymorphic populations†	Average N per polymorphic population	Populations with significant nonzero LD *N*, (%)	Populations with significant positive LD *N*, (%)
Daday (1954a)	Eurasia	33	133.5	7 (21.2)	7 (21.2)
Daday (1954b)	Swiss Alps	5	115.6	0 (0)	0 (0)
de Araujo (1976)	Wales	17	40.8	1 (5.9)	1 (5.9)
Daday (1958)	Asia	8	143.5	2 (25.0)	2 (25.0)
Daday (1958)	Africa*	7	84.9	0 (0)	0 (0)
Daday (1958)	S. America*	5	119.8	1 (20.0)	1 (20.0)
Daday (1958)	N. America*	36	85.5	4 (11.1)	3 (8.3)
Daday (1958)	Australia*	27	103.7	8 (29.6)	8 (29.6)
Daday (1958)	New Zealand*	3	72.3	2 (66.6)	2 (66.6)
Ennos (1982)	England	10	94.2	2 (20.0)	2 (20.0)
Frasier (1986)	Canada*	7	58.7	1 (14.3)	0 (0)
Kakes (1987)	Netherlands	16	49.8	7 (43.8)	7 (43.8)
Till-Bottraud *et al.* (1988)	French Alps	13	104.8	3 (23.1)	3 (23.1)
Majumdar *et al.* (2004)	Himalayas	5	1258.2	5 (100)	3 (60.0)
Present study	N. America*	7	163.6	0 (0)	0 (0)

*Indicates non-native (introduced) populations.
†Sampled populations that were polymorphic at both cyanogenesis genes.

populations showing positive LD range from 0% to ≥60% for both categories of populations. Taken together, these findings suggest that while epistatic selection may be a factor in the evolution of the cyanogenesis polymorphism in some populations, it does not appear to be a universal phenomenon in white clover.

For the specific case of North American populations, the absence of any detectable LD in the present study differs somewhat from the observations of Daday (1958), who surveyed 36 polymorphic North American populations, three of which (8.3%) show evidence of significantly positive LD (Table 4). Whereas Daday's North American sampling focused largely on populations occurring west of the Rocky Mountains, the present study sampled populations exclusively within the Mississippi basin. This difference in sampling could potentially account for some of the differences in LD observed between the two studies. Another possibility is that the earlier North American study may have unintentionally pooled some populations that were adaptively differentiated at the cyanogenesis genes, thereby creating positive LD as an artefact of admixture (Nordborg & Tavare 2002). We judge this latter explanation as unlikely because the population sampling scheme employed by Daday in his classic studies is very similar to our own (e.g. Daday 1954a, 1958).

More generally, the apparent presence of positive LD in some, but not all, polymorphic white clover populations (Table 4) suggests that the importance of epistatic selection in the cyanogenesis polymorphism may vary from one population to another. We propose two nonmutually exclusive explanations that could account for this pattern. First, differences in herbivore communities and abundances may affect which components of the cyanogenic response are under selection in a given location. For example, Ennos (1982) notes that some mollusks possess β-glucosidase enzymes in their guts that could function in the same way as linamarase in hydrolysing cyanogenic glucosides and inducing a deterrent effect. Thus, clover populations in a community rich in mollusks may be equally well protected whether they are Acli or AcLi cyanotypes. Under such conditions, no epistatic selection would be expected.

A second possibility is that the cyanogenic components may have functions other than in herbivore deterrence, and that there may be variation among populations in selective pressure on these functions. Both linamarase and cyanogenic glucosides have been proposed to serve as nitrogen storage and transport compounds (Gleadow & Woodrow 2002; Piotrowski 2008; Møller 2010), functions whose importance could vary tremendously depending on factors such as the length of the growing season or other climatic variables. Regional differences in selective pressure for these other functions would thus be expected to affect how the cyanogenesis polymorphisms are maintained. Consistent with this explanation, our review of published white clover cyanogenesis data indicates that very few populations are entirely fixed for the acyanogenic phenotype, even in extremely cold locations such as Northern Canada and Alaska (e.g. Fraser 1986). This observation suggests that there may be undiscovered functions of the cyanogenesis components that could be affecting cyanotype frequencies.

Conclusions

Understanding how polymorphism is maintained in heterogeneous environments has been widely studied both theoretically and empirically (Levene 1953; Hedrick 2006). A major conclusion of these studies is that generalizations are rare because the complexity of natural systems, including animal–plant interactions, will often produce multidimensional selection pressures that vary across geographic mosaics (Linhart 1991; Thompson 2005). This study identifies rapid evolution of an adaptive cyanogenesis cline across the central United States and finds that the mechanism of maintaining polymorphism may differ from that in the native region. Specifically, epistatic selection, previously proposed as a force that maintains this chemical defence polymorphism, does not appear to be operating in these North American white clover populations. Our study raises interesting new questions about the maintenance of adaptive polymorphism in this classic ecological genetic system. Ultimately, we suggest that alternative mechanisms of maintaining the *Ac/ac* and *Li/li* polymorphisms may be at play in different populations within this species, and that understanding the ecological differences between these populations (including both biotic and abiotic factors) would provide much-needed insight into this system.

Acknowledgements

We appreciate the assistance of AgResearch New Zealand and specifically Brent Barrett and Andrew Griffith for providing information regarding molecular work. We thank Yangling Wei and other members of the Brummer Laboratory (Noble Foundation) for providing data on microsatellite markers and for technical advice in genotyping. We thank Mike Dyer and other staff of the Washington University greenhouse for assistance in growing clover samples. This study was greatly improved by Briana Gross, Jared Strasburg, Allison Miller and the Olsen laboratory group, who took time to review an earlier version of this manuscript. Funding was provided by an NSF CAREER award to KMO (DEB-0845497), and by an NSF Doctoral Dissertation Improvement Grant to NJK (DEB-1110588).

References

Antonovics J, Bradshaw AD (1970) Evolution in closely adjacent plant populations. 8. Clinal patterns at a mine boundary. *Heredity*, **25**, 349–362.

de Araujo AM (1976) The relationship between altitude and cyanogenesis in white clover (*Trifolium repens*, L.). *Heredity*, **37**, 291–293.

Atwood SS, Sullivan JT (1943) Inheritance of a cyanogenic glucoside and its hydrolyzing enzyme in *Trifolium repens*. *Journal of Heredity*, **34**, 311–320.

Bonnet P, Van de Peer Y (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **7**, 1–12.

Brighton F, Horne MT (1977) Influence of temperature on cyanogenic polymorphisms. *Nature*, **265**(5593), 437–438.

Brookfield J (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non-templated nucleotide addition by tag DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques*, **20**, 1004–1006.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between Arabidopsis *FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences, USA*, **101**, 15670–15675.

Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human-evolution. *Science*, **259**, 639–646.

Chapuis M, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution*, **24**, 621–631.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, **2**, 379–384.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, **10**, 195–205.

Clegg SM, Degnan SM, Kikkawa J, Moritz C, Estoup A, Owens IPF (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences, USA*, **99**, 8127–8132.

Corkill L (1942) Cyanogenesis in white clover (*Trifolium repens* L.). V. The inheritance of cyanogenesis. *New Zealand Journal of Science and Technology*, **23**, 178B–193B.

Daday H (1954a) Gene frequencies in wild populations of *Trifolium repens* L. I. Distribution by latitude. *Heredity*, **8**, 61–78.

Daday H (1954b) Gene frequencies in wild populations of *Trifolium repens* II. Distribution by altitude. *Heredity*, **8**, 377–384.

Daday H (1958) Gene frequencies in wild populations of *Trifolium repens* L. III. World distribution. *Heredity*, **12**, 169–184.

Daday H (1965) Gene frequencies in wild population of *Trifolium repens* IV. Mechanism of natural selection. *Heredity*, **20**, 355–365.

Dirzo R, Harper JL (1982a) Experimental studies on slug-plant interactions III. Differences in the acceptability of individual plants of Trifolium repens to slugs and snails. *Journal of Ecology*, **70**, 101–117.

Dirzo R, Harper JL (1982b) Experimental studies on slug-plant interactions IV. The performance of cyanogenic and acyanogenic morphs of Trifolium repens in the field. *Journal of Ecology*, **70**, 119–138.

Dirzo R, Sarukhan J (1984) *Perspectives on Plant Population Ecology*. Sinauer & Associates, Sunderland, Massachusetts.

Dlugosch KM, Parker IM (2008) Invading populations of an ornamental shrub show rapid life history evolution despite genetic bottlenecks. *Ecology Letters*, **11**, 701–709.

Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey.

Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, New Jersey.

Ennos RA (1981) Detection of selection in populations of white clover (*Trifolium repens* L). *Biological Journal of the Linnean Society*, **15**, 75–82.

Ennos RA (1982) Association of the cyanogenic loci in white clover. *Genetical Research*, **40**, 65–72.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Faircloth BC (2006) GMCONVERT: file conversion for GENEMAPPER output files. *Molecular Ecology Notes*, **6**, 968–970.

Feigl F, Anger V (1966) Replacement of benzidine by copper ethylacetoacetate and tetra base as a spot-test reagent for hydrogen cyanide and cyanogen. *Analyst*, **91**, 282–284.

Fisher RA (1937) The wave of advance of advantageous genes. *Annals of Human Genetics*, **7**, 355–369.

Fraser J (1986) Cyanogenesis in naturalised white clover (*Trifolium repens* L.) populations in Atlantic Canada. *Progress Report, Clovers and Special Purpose Legumes Research*, **19**, 12–14.

Ganders FR (1990) Altitudinal clines for cyanogenesis in introduced populations of white clover near Vancouver, Canada. *Heredity*, **64**, 387–390.

Glaubitz JC (2004) CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, **4**, 309–310.

Gleadow RM, Woodrow IE (2002) Constraints on effectiveness of cyanogenic glycosides in herbivore defense. *Journal of Chemical Ecology*, **28**, 1301–1313.

Gliddon C, Saleem M (1985) Gene-flow in *Trifolium repens*—and expanding genetic neighbourhood. In: *Genetic Differentiation and Dispersal in Plants* (ed P. Jacquard, G. Heim and J. Antonovics), pp. 293–309. Springer-Verlag, Berlin, Germany.

Gross JB, Borowsky R, Tabin CJ (2009) A novel role for *Mc1r* in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics*, **5**, e1000326.

Haldane JBS (1948) The theory of a cline. *Journal of Genetics*, **48**, 277–284.

Hedrick PW (2006) Genetic polymorphism in heterogeneous environments: the age of genomics. *Annual Review of Ecology Evolution and Systematics*, **37**, 67–93.

Hiesey WH, Clausen J, Keck DD (1942) Ecological aspects of evolution: relations between climate and intraspecific variation in plants. *The American Naturalist*, **76**, 5–22.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for

global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hill WG (1974) Estimation of linkage disequilibrium in randomly mating populations. *Heredity*, **33**, 229–239.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.

Hudson ME, Kane NC (2009) Plant genomes do a balancing act. *Molecular Ecology*, **18**, 2743–2745.

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science*, **287**, 308–309.

Hughes MA (1991) The cyanogenic polymorphism in *Trifolium repens* L (white clover). *Heredity*, **66**, 105–115.

Kakes P (1987) On the polymorphism for cyanogenesis in natural populations of *Trifolium repens* L. in the Netherlands I. Distribution of the genes Ac and Li. *Acta Botanica Neerlandica*, **36**, 59–69.

Kakes P (1989) An analysis of the costs and benefits of the cyanogenic system in *Trifolium repens* L. *Theoretical and Applied Genetics*, **77**, 111–118.

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

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

Kjærgaard T (2003) A plant that changed the world: the rise and fall of clover 1000-2000. *Landscape Research*, **28**, 41–49.

Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature*, **435**, 95–98.

Lenormand T, Bourguet D, Guillemaud T, Raymond M (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature*, **400**, 861–864.

Levene H (1953) Genetic equilibrium when more than one ecological niche is available. *The American Naturalist*, **87**, 331–333.

Linhart YB (1991) Disease, parasitism and herbivory: multidimensional challenges in plant evolution. *Trends in Ecology & Evolution*, **6**, 392–396.

Majumdar S, De KK, Banerjee S (2004) Influence of two selective factors on cyanogenesis polymorphism of *Trifolium repens* L. in Darjeeling Himalaya. *Journal of Plant Biology*, **47**, 124–128.

McKechnie SW, Blacket MJ, Song SV *et al*. (2010) A clinally varying promoter polymorphism associated with adaptive variation in wing size in *Drosophila*. *Molecular Ecology*, **19**, 775–784.

Møller BL (2010) Functional diversifications of cyanogenic glucosides. *Current Opinion in Plant Biology*, **13**, 338–347.

Montague JL, Barrett SCH, Eckert CG (2008) Re-establishment of clinal variation in flowering time among introduced populations of purple loosestrife (*Lythrum salicaria*, Lythraceae). *Journal of Evolutionary Biology*, **21**, 234–245.

Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–1569.

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, **18**, 83–90.

Oakeshott J, Gibson J, Anderson P, Knibb W, Anderson D, Chambers G (1982) Alcohol-dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution*, **36**, 86–96.

Olsen KM, Ungerer MC (2008) Freezing tolerance and cyanogenesis in white clover (*Trifolium repens* L. Fabaceae). *International Journal of Plant Sciences*, **169**, 1141–1147.

Olsen KM, Sutherland BL, Small LL (2007) Molecular evolution of the *Li/li* chemical defence polymorphism in white clover (*Trifolium repens* L.). *Molecular Ecology*, **16**, 4180–4193.

Olsen KM, Hsu S-C, Small LL (2008) Evidence on the molecular basis of the *Ac/ac* adaptive cyanogenesis polymorphism in white clover (*Trifolium repens* L.). *Genetics*, **179**, 517–526.

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

Pederson GA, Brink GE (1998) Cyanogenesis effect on insect damage to seedling white clover in a bermudagrass sod. *Agronomy Journal*, **90**, 208–210.

Phifer-Rixey M, Heckman M, Trussell GC, Schmidt PS (2008) Maintenance of clinal variation for shell colour phenotype in the flat periwinkle *Littorina obtusata*. *Journal of Evolutionary Biology*, **21**, 966–978.

Piotrowski M (2008) Primary or secondary? Versatile nitrilases in plant metabolism. *Phytochemistry*, **69**, 2655–2667.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation of the co-ancestry coefficient: basis for a short-term genetic-distance. *Genetics*, **105**, 767–779.

Rosenblum EB, Roempler H, Schoeneberg T, Hoekstra HE (2010) Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences, USA*, **107**, 2113–2117.

Saccheri IJ, Rousset F, Watts PC, Brakefield PM, Cook LM (2008) Selection and gene flow on a diminishing cline of melanic peppered moths. *Proceedings of the National Academy of Sciences, USA*, **105**, 16212–16217.

Saucy F, Studer J, Aerni V, Schneiter B (1999) Preference for acyanogenic white clover (*Trifolium repens*) in the vole *Arvicola terrestris*: I. Experiments with two varieties. *Journal of Chemical Ecology*, **25**, 1441–1454.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Silvertown J, Charlesworth D (2001) *Introduction to Plant Population Biology*, 4th edn. Blackwell Science, Oxford, UK.

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

Steiner CC, Roempler H, Boettger LM, Schoeneberg T, Hoekstra HE (2009) The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution*, **26**, 35–45.

Storz JF, Sabatino SJ, Hoffmann FG *et al*. (2007) The molecular basis of high-altitude adaptation in deer mice. *PLoS Genetics*, **3**, 448–459.

Thompson JN (2005) *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago, Illinois.

Till-Bottraud I, Kakes P, Dommee B (1988) Variable phenotypes and stable distribution of the cyanotypes of

Trifolium repens L. in Southern France. *Acta Oecologica*, **9**, 393–404.

Todesco M, Balasubramanian S, Hu TT *et al*. (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature*, **465**, 632–636.

Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science*, **308**, 691–693.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Vasemagi A (2006) The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. *Genetics*, **173**, 2411–2414.

Viette M, Tettamanti C, Saucy F (2000) Preference for acyanogenic white clover (*Trifolium repens*) in the vole *Arvicola terrestris*: II. Generalization and further investigations. *Journal of Chemical Ecology*, **26**, 101–122.

Weinig C, Dorn LA, Kane NC *et al*. (2003) Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. *Genetics*, **165**, 321–329.

Zhen Y, Ungerer MC (2008) Clinal variation in freezing tolerance among natural accessions of *Arabidopsis thaliana*. *New Phytologist*, **177**, 419–427.

The authors' research interests center on plant evolutionary biology. N.K. is a PhD student in the EEPB graduate program at Washington University in St. Louis. His research focuses on understanding the genetic basis of adaptation and ecological trade-offs that maintain polymorphism across space and time. He is particularly interested in adaptations involving plant secondary compounds and plant-animal interactions. K.O. is an associate professor in the Biology department of Washington University. His laboratory's research includes studies on genetic mechanisms of adaptation and the evolution of crops and their relatives.

Data accessibility

Microsatellite Data—DRYAD doi:10.5061/dryad.7q3m05fm.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Subpopulation information and cyanogenesis polymorphism data.

Table S2 Details of microsatellite markers and observed polymorphism data.

Table S3 Summary of microsatellite polymorphism data per subpopulation.

Table S4 Pairwise population measures of genetic differentiation among subpopulations using all microsatellite data.

Fig. S1 Regression of population frequencies of the *Li* allele (A), *Ac* allele (B) or the AcLi cyanotype (C) against geographical distances among populations, using the southernmost population as the reference point (0 km). Statistical significance of linear regressions is indicated as follows: $*0.01 < P < 0.05$, $**0.01 < P < 0.001$ and $***P < 0.001$.

Fig. S2 Regressions of pairwise measures of population differentiation for (A) *Ac/ac* and (B) *Li/li* against population differentiation for nine microsatellite loci. Points above the diagonal lines indicate greater differentiation between sampling sites for the cyanogenesis genes than for neutral markers. Statistical significance of one to one plots is indicated as follows: $*0.01 < P < 0.05$, $**0.01 < P < 0.001$ and $***P < 0.001$.

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