MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 6011–6028 doi: 10.1111/mec.12993

Ecotypes of an ecologically dominant prairie grass (*Andropogon gerardii*) exhibit genetic divergence across the U.S. Midwest grasslands' environmental gradient

MIRANDA M. GRAY,* PAUL ST. AMAND,† NORA M. BELLO,‡ MATTHEW B. GALLIART,§ MARY KNAPP,¶ KAREN A. GARRETT,** THEODORE J. MORGAN,§ SARA G. BAER,†† BRIAN R. MARICLE,‡‡ EDUARD D. AKHUNOV** and LORETTA C. JOHNSON§

*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA, †USDA-ARS, Wheat Genetics Research Unit, Manhattan, KS 66506, USA, ‡Department of Statistics, Kansas State University (KSU), Manhattan, KS 66506, USA, §Division of Biology, KSU, Manhattan, KS 66506, USA, ¶Department of Agronomy, KSU, Manhattan, KS 66506, USA, **Department of Plant Pathology, KSU, Manhattan, KS 66506, USA, ††Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901, USA, ‡‡Department of Biological Sciences, Fort Hays State University, Hays, KS 67601, USA

Abstract

Big bluestem (*Andropogon gerardii*) is an ecologically dominant grass with wide distribution across the environmental gradient of U.S. Midwest grasslands. This system offers an ideal natural laboratory to study population divergence and adaptation in spatially varying climates. Objectives were to: (i) characterize neutral genetic diversity and structure within and among three regional ecotypes derived from 11 prairies across the U.S. Midwest environmental gradient, (ii) distinguish between the relative roles of isolation by distance (IBD) vs. isolation by environment (IBE) on ecotype divergence, (iii) identify outlier loci under selection and (iv) assess the association between outlier loci and climate. Using two primer sets, we genotyped 378 plants at 384 polymorphic AFLP loci across regional ecotypes from central and eastern Kansas and Illinois. Neighbour-joining tree and PCoA revealed strong genetic differentiation between Kansas and Illinois ecotypes, which was better explained by IBE than IBD. We found high genetic variability within prairies (80%) and even fragmented Illinois prairies, surprisingly, contained high within-prairie genetic diversity (92%). Using BAYENV2, 14 top-ranked outlier loci among ecotypes were associated with temperature and precipitation variables. Six of seven BAYESCAN F_{ST} outliers were in common with BAYENV2 outliers. High genetic diversity may enable big bluestem populations to better withstand changing climates; however, population divergence supports the use of local ecotypes in grassland restoration. Knowledge of genetic variation in this ecological dominant and other grassland species will be critical to understanding grassland response and restoration challenges in the face of a changing climate.

Keywords: climate change, genome scan, isolation by environment, outlier analyses, restoration, tallgrass prairie

Received 9 January 2013; revision received 29 October 2014; accepted 31 October 2014

Introduction

One of the main goals of evolutionary biology is to understand factors that contribute to population genetic divergence (Mayr 1963), ultimately leading to formation of new species (Coyne & Orr 2004; Nosil 2012). Habitats

Correspondence: Loretta C. Johnson, Fax: 785 532 6653; E-mail: johnson@ksu.edu

are often both temporally and spatially variable, and this can result in divergent selection across environments, and may lead to adaptive genetic divergence (Dobzhansky 1937; Nosil & Crespi 2004). Indeed, populations in heterogeneous environments have the potential to undergo local adaptation to a specific environment if the frequency of locally beneficial alleles increases within the population (Conner & Hartl 2004). Such adaptive loci would be expected to show excess differentiation (i.e. 'outliers') among populations compared to the rest of the genome that is evolving neutrally (Wright 1949; Beaumont & Nichols 1996; Excoffier *et al.* 1999). Moreover, the study of genetic diversity across environments allows for insight into the role of environmental drivers in adaptive differentiation. Frequency of outlier loci can be related to environmental gradients (Freedman *et al.* 2010; Manel *et al.* 2012) to study the nature of local adaptation (Leimu & Fischer 2008; Savolainen *et al.* 2013).

Within the last decade, the study of genetic variation across heterogeneous environments has advanced considerably due to the use of landscape genetic approaches (Holdenregger *et al.* 2010; Manel *et al.* 2010; Sork *et al.* 2010; Lee & Mitchell-Olds 2011; Joost *et al.* 2013). Concurrently, an increase in population genomic data combined with interest in identifying adaptive loci has spurred the development of analytical tools (as reviewed in De Mita *et al.* 2013; Jones *et al.* 2013 and Lotterhos & Whitlock 2014). Recent studies have utilized outlier loci approaches in nonmodel, ecologically relevant species to assess the relationship between outlier loci and environment and to identify candidate loci driving adaptation (Hancock *et al.* 2011; Lee & Mitchell-Olds 2011; reviewed in Tonsor 2012). Work on adaptive loci has shed new light on the role of regional climate (Hancock *et al.* 2011; Chen *et al.* 2012; De La Torre *et al.* 2014; Yoder *et al.* 2014; Zhou *et al.* 2014) and altitudinal differences (Gonzalo-Turpin & Hazard 2009; Poncet *et al.* 2010; Manel *et al.* 2012; Anderson *et al.* 2013) on adaptive divergence of plant species. These studies take on greater importance in the face of a rapidly changing climate (Jump & Penuelas 2005; Reusch & Wood 2007; Temunovic *et al.* 2013), as plants must either adapt genetically on contemporary timescales (Hoffman & Sigro 2011), adjust phenotypically (Franks *et al.* 2014), migrate, or suffer extinction (Shaw & Etterson 2012).

Population divergence may occur due to factors other than selection, such as a reduction in gene flow across a landscape (Wright 1943). Traditionally, landscape genomics has focused on isolation by distance (IBD) as a main driver of divergence (Jenkins *et al.* 2010). More recently, however, problems with IBD (Miermans 2012) and disentangling the roles of distance and demographic history from ecology ('isolation by environment' or IBE) have come to the forefront (Gaggiotti *et al.* 2009; Bradburd *et al.* 2013; Wang *et al.* 2013; Sexton *et al.* 2014). To this end, newer and more powerful methodologies that take into account evolutionary non-independence between populations are increasingly being utilized (Carl & Kuhn 2007; Bradburd *et al.* 2013; Frichot *et al.* 2013; Wang 2013; Wang *et al.* 2013). A comprehensive meta-analysis by Shafer & Wolf (2013) comparing the relative strengths of IBD vs. IBE in ecological speciation found that ecologically induced divergent selection is widespread in nature, across timescales and taxa. Furthermore, Lee & Mitchell-Olds (2011) observed *Boechera stricta* intraspecific genetic differentiation was more attributed to environmentally based selection (specifically, a water availability gradient) than to IBD. The interplay of IBD and IBE in species' genetic divergence thus appears to be complex and system dependent.

Big bluestem (*Andropogon gerardii*) is one of the most ecologically dominant C_4 grasses (Epstein *et al.* 1998) of the U.S. Midwest grassland. The species occurs in every state east of the Rocky Mountains and eastern Canada but attains biomass dominance (80% cover, Risser *et al.* 1981) in the tallgrass prairie. In spite of its importance, studies of big bluestem intraspecific variation are few and have focused on local (Avolio *et al.* 2011) and regional geographical scales (Illinois or Arkansas: Gustafson *et al.* 1999; Ohio: Selbo & Snow 2005; Carolinas: Tompkins *et al.* 2011) or on cultivars (Gustafson *et al.* 1999). However, most genetic differentiation studies (except Rouse *et al.* 2011) have focused on regions outside the current centre of dominance (Tompkins *et al.* 2011; Price *et al.* 2012). This is despite the fact that the dominant distribution of big bluestem spans one of the sharpest environmental gradients of the U.S. This gradient is characterized by strong historical precipitation variation ranging from 58 to 116 cm mean annual rainfall/year from central Kansas to southern Illinois over a span of 1150 km. As the existing tallgrass prairie formed >10 000 years ago, since the last glaciation (Axelrod 1985), there has likely been adequate time for climatic and ecological selection pressures to be exerted on populations. Such unique circumstances provide an ideal natural laboratory to study population divergence and adaptation.

Studies of climate-linked genetic variation in foundation species are timely (Sork *et al.* 2010). Specifically, for the U.S. Midwest region, climate predictions include increased frequency of drought (IPCC 2013). Recently, this region experienced the worst drought in >50 years (NOAA 2012). Thus, it is imperative to characterize genetic variation across current climate gradients to better predict how this species may respond to future climates–either through adaptive evolution, range

expansion (Shaw & Etterson 2012) or with human-assisted migration in restoration plantings (Jump & Penuelas 2005). Furthermore, spatial genetic approaches are instrumental to the discovery of genetic differentiation that may help inform restoration of grasslands in the United States and beyond in the face of climate change, with only 4% of historical prairie remaining (Samson & Knopf 1994). The largest continuous expanse of prairie occurs in Kansas (Samson & Knopf 1994) while the eastern extent of this ecosystem in Illinois consists of small patches of virgin prairie due to row crop agriculture and fragmentation (Robertson 1996; Corbett 2004). Big bluestem is one of the main species used in U.S. grassland restorations, including 3.6 million ha in a five-state area of the Midwest (Conservation Reserve Programme, http://www.nrcs.usda.gov/programs/crp). Thus, genetic studies are critical to inform land managers on genetic suitability of plant populations used for restoration (Gustafson *et al.* 2001, 2002, 2004a,b; Jones 2003; Rice & Emery 2003) and for possible mitigation against climate change (Harris *et al.* 2006; Nicotra *et al.* 2010). Our study addresses levels of genetic diversity across the dominant range of big bluestem and the suitability of natural populations for restoration.

Here, we use a landscape genomics approach (Manel *et al.* 2010; Joost *et al.* 2013; Sork *et al.* 2013) to focus on divergent selection of a widely distributed prairie grass across the spatially variable environmental gradient of the U.S. Midwest grasslands. Objectives were to: (i) characterize neutral genetic diversity and structure within and among three regional ecotypes derived from 11 prairies across the U.S. Midwest environmental gradient, (ii) distinguish between the relative roles of IBD vs. IBE on ecotype divergence, (iii) identify outlier loci under selection and (iv) assess the association between outlier loci and climate. We hypothesized big bluestem populations genetically diverged across the U.S. Midwest environmental gradient, due to a combination of regional climate, geographical distance and prairie fragmentation. Given the strong precipitation gradient and its importance in regulating growth and performance of grasses (Sala *et al.* 1988; Knapp *et al.* 2001), we expected aspects of precipitation to be most associated with outlier loci differentiating ecotypes.

Methods

Seed collection

Seeds were collected in autumn 2008 from 11 prairie populations across the U.S. Midwest. These were the same prairies from which seeds were collected for ecotype reciprocal garden experiments reported elsewhere (L. C. Johnson, S. G. Baer & B. R. Maricle, unpublished data). The 11 source prairies of varying sizes were partitioned across three main ecotype regions: central Kansas (CKS), eastern Kansas (EKS) and southern Illinois (SIL) (Table 1, map overlap in Fig. 3A). All sampled prairies are protected parks and/or research areas (with the exception of two private properties) and received no prior ploughing or restoration with cultivars. Prairies were occasionally burned and historically grazed (L. C. Johnson, personal communication with land managers). Across the tallgrass prairie landscape, several soil characteristics vary locally, although the dominant textures include silt loam and silty clay loam. Specific soil types for the source prairies as determined by http://www.websoilsurvey.sc.egov.usda.gov are included in Table 1. Large volumes (hundreds of grams) of seed were collected from multiple locations and time points within each prairie. Seeds collected within a prairie were mixed and subsampled to attain an unbiased representation of the natural variation within each prairie.

Sample preparation and DNA isolation

Approximately 3.5 g of seeds per prairie population were rubbed to remove chaff and sown in flats. Seedlings were well-watered and grown in a greenhouse at 25 °C with a 12-h photoperiod. After 2 months, seedlings were transplanted into 10 × 10-cm pots with Metro-Mix 510 potting soil until 75–100 mg of young leaves per plant could be collected for DNA isolation. Leaf tissue was lyophilized in a freeze-drier (ModulyoD-115; Thermo Savant, Holbrook, NY, USA) and ground to a fine powder with 4.0-mm stainless steel beads (Abbott Ball Company Inc., Hartford, CT, USA) using a Mixer Mill 400 (Retsch Inc., Newton, PA, USA) at 25–30 cycles/s for 15 min. DNA was then isolated using the CTAB protocol (Doyle & Doyle 1987) and resuspended in 50–100 μL Tris-HCl (10 mM) + Triton X-100 (0.003125%) buffer (pH 8.0). Quality and quantity of DNA was verified using a spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) with OD requirements of 260/280 ~2.0 and 260/230 ≥ 1.80 for genotyping. Samples were checked for lack of degradation on 0.8% agarose gels.

AFLP genotyping

Our AFLP protocol followed aspects of Rouse *et al.* (2011) specific to big bluestem. DNA restriction digestion and adapter ligation steps were combined and comprised of: ~300 ng genomic DNA (~25 ng/μL), 5 units of *Eco*RI HF (0.25 μL; New England Biolabs) and 5 units of *Mse*I (0.5 μL; New England Biolabs), 100 units of T4 DNA ligase (0.25 μL; New England Biol-

Table 1 Sampled prairies and related environmental information. Table includes number of plants genotyped per prairie, county, prairie size in hectares, soil classification, GPS coordinates and 10 site environmental descriptors: prairie elevation (m), number of precipitation events >1.25 cm/year, total precipitation in the driest year (cm), annual mean precipitation (cm), seasonal mean precipitation (cm), annual diurnal temperature variation (°C), seasonal diurnal temperature variation (°C), annual mean temperature (°C), seasonal mean temperature (°C) and a temperature severity index expressed as the fraction of days with maximum temperatures >35°C divided by total number of days. Climatic data are average data collected between 1961 and 2011. All weather stations were within 20 km of each prairie

Region	Prairie site* (no. plants sampled)	County (prairie size in hectares)	Weather station	Soil classification	Long (W)	Lat (N)	Elevation (m)	Precip events > 1.25 (cm)	Precip driest year (cm)	Annual mean precip (cm)	Seasonal mean precip (cm)	Annual diurnal (°C)	Seasonal diurnal (°C)	Annual mean temp (°C)	Seasonal mean temp (°C)	Temp severity index
Central Kansas	WEB (40)	Rooks (356)	Webster Dam	Wakeen-Harney silty loam	99.32	39.24	606	17	25.96	58.70	39.35	15	15.4	18.8	19.8	0.054
	SAL (30)	Ellis (880)	Hays 1S	Bogue-Armo clay loam	99.14	39.02	641	16	36.32	58.04	37.70	14.6	14.6	12.0	20.7	0.040
	CDB (33)	Trego (850)	Cedar Bluff Dam	Armo clay loam	99.46	38.45	688	16	32.18	53.31	35.97	14.4	14.5	18.7	19.5	0.055
Eastern Kansas	KON (22)	Riley/Geary (1557)	Manhattan 6SW	Benfield-Florence complex silty clay loam	96.36	39.05	366	22	68.89	88.47	56.54	12.7	12.4	12.8	21.0	0.034
	TAL (23)	Chase (4409)	Tallgrass Park	Cline Sogne silty clay loam	96.33	38.25	392	21	59.77	82.82	49.19	12.8	12.2	19.0	20.8	0.037
	CAR (47)	Pottawatomie (99)	Wamego	Benfield-Florence complex silty clay loam	96.38	39.2	389	23	52.35	87.20	53.34	16.6	13.1	13.0	21.4	0.032
	TOW (28)	Riley (61)	Tuttle Creek	Irwin silty clay loam	96.37	39.13	379	21	45.11	81.12	50.70	13.1	13.2	11.7	20.3	0.039
Illinois	DES (55)	Jackson (~0.4)	Carbondale	Orthents silt loam	89.14	37.51	119	33	67.41	115.92	53.53	12.3	12.6	13.2	27.4	0.014
	TM (43)	Effingham, Monroe, Fayette (28)	Salem	Cisne silty loam	88.5	38.46	160	25	70.01	107.57	51.49	11.7	12.4	12.8	24.5	0.014
	WAL (42)	Jasper (5)	Charleston	Atlas silty clay loam	88.09	38.59	150	27	69.18	104.04	50.80	10.8	11.7	13.4	26.1	0.010
	FUL (15)	Monroe (214)	Sparta	Menfro silty clay loam	89.48	37.58	215	31	69.38	111.27	55.14	11.9	12.6	13.3	27.3	0.018

*WEB = Webster Prairie; SAL = Saline Prairie; CDB = Cedar Bluff; KON = Konza Prairie; TAL = Tallgrass Prairie National Preserve; CAR = Carnahan; TOW = Top of the World; DES = DeSoto Prairie; TM = Twelve-Mile Prairie; WAL = Walters' Prairie; FUL = Fults Hill Prairie.

abs), 2 μL of 10× ligase buffer (New England Biolabs), 1.0 μL of each adaptor pair (5 pm/μL of *Eco*RI adaptors; 50 pm/μL of MseI adaptors; Integrated DNA Technologies) and 13 μL ddH_2O for a total reaction volume of 30 μL. The restriction–ligation mixture was incubated at room temperature overnight to ensure the complete digestion–ligation. Restricted-ligated DNAs were diluted 10×.

Pre-amplification reactions used primers complementary to the DNA restriction site and adapter pair with an additional one base pair overhang (*Eco*RI = 5′-AGA CTGCGTACCAATTC-A-3′ and *Mse*I = 5′-GATGAGTCC TGAGTAA-C-5′). Individual pre-amplification PCRs consisted of a final volume of 40 μL and included: 10 μL diluted restricted-ligated DNA template, 1.2 μL of each primer (10 μM), 6 μL 5× PCR buffer (Promega), 3 μL $MgCl_2$ (25 mM; Promega), 0.64 μL dNTPs (5 mM each), 0.75 units of Go Taq Flexi DNA polymerase (0.15 μL; Promega) and 17.75 μL ddH_2O. PCR steps were as follows: 20 °C, 5 s; ramp from 20 to 70 °C (0.2 °C/s); 70 °C, 2 min; 94 °C, 1 min; then 30 cycles of 94 °C, 30 s; 56 °C, 1 min; 72 °C, 1 min; followed by 72 °C, 10 min; 15 °C, 5 min. Pre-amplified template was diluted 20×.

A selective PCR was performed using two primer sets with three additional bases (primer set 1: 5′GAT GAGTCCTGAGTAA-CTG-3′ + 5′HEX-AGACTGCGTAC CAATTC-ACC-3′; primer set 2: 5′GATGAGTCCTGAG TAACGC-3′ + 5′6FAM-AGACTGCGTACCAATTC-AA A-3′). We chose these two selective primer pairs after examining the quality of genotype profiles resulting from eight primer combinations (data not shown). Each selective PCR had a 20.5 μL final volume and consisted of: 1.5 μL diluted pre-amplified template, 1.62 μL M-side primer (10 μM, M-CTG or M-CGC), 1.62 μL fluorescently labelled E-side primer (10 μM, 5′-6HEX or 5′-6FAM), 4 μL 5× PCR buffer (Promega), 2 μL $MgCl_2$ (25 mM; Promega), 0.8 μL dNTPs (5 mM each), 1 unit Go Taq Flexi DNA polymerase (0.2 μL; Promega) and 8.76 μL ddH_2O. The touchdown PCR profile was as follows: 95 °C for 2 min; 13 cycles of 65 °C for 30 s (−0.7 °C/cycle), 72 °C for 90 s and 94 °C for 30 s; 23 cycles of 56 °C for 30 s, 72 °C for 90 s and 94 °C for 30 s; 72 °C for 10 min and 15 °C for 5 min. To optimize the efficiency (overall band intensity) of primer set 2 (M-CGC + 6-FAM), the touchdown PCR was modified to 60 °C rather than 65 °C. The selective PCR was diluted 10×. A solution of 9.5 μL formamide + 0.5 μL GeneScan-500 LIZ internal size standard (Applied Biosystems, Foster City, CA, USA) was added to 1.5 μL diluted selective template. Samples were loaded onto an ABI Prism 3730 DNA Analyzer (Applied Biosystems) with a 50-cm capillary and electrokinetic injection voltage of 1 kV applied for 10 s. Lower injection voltage and shorter injection time improved the resolution of AFLP bands of similar molecular weights. This method also improved the repeatability of longer fragments observed in genotype profiles as well as prevented oversaturation of peak intensities.

Marker scoring and error rate estimation

Non-normalized profiles were scored using GeneMarker software version 1.97 (SoftGenetics LLC, State College, PA, USA). AFLP panels were autocreated with a 1.0 base pair total width; afterwards, bins were manually checked and adjusted to retain only smoothly shaped peaks. Irreproducible peaks or irregularly shaped peaks were discarded. We scored only peaks above 100 relative fluorescent units, as this was reliably above the noise of negative controls included in the study (recommended by Bonin *et al.* 2004). Band sizes between 80 and 500 base pairs were scored.

We took necessary precaution to ensure AFLP reproducibility (Crawford *et al.* 2012). To verify the consistency of the AFLP technique, a set of four reference DNAs were included in each successive AFLP reaction to ensure between-run reproducibility. In addition, 2–3 independent restriction–ligations were performed on one DNA sample per prairie and genotyped (11 total replicates) to calculate an overall error rate. The replicate samples comprised 4% of total genotyped samples. Replication at the restriction–ligation stage was implemented as it is the most critical step of the AFLP reaction and can result in band presence/absence artefacts (Mueller & Wolfenbarger 1999). The AFLP technical error rate estimation was calculated by dividing total number of mismatched bands by the total number of AFLP bands produced overall in the AFLP fingerprint (Bonin *et al.* 2004).

Prairie genetic diversity and structure

In the final genotyping data set, 15–55 plants per prairie were included (Table 1). Marker statistics, diversity and diversity analyses were calculated in GenAlEx version 6.56 (Peakall & Smouse 2006). Relatedness among all individuals was depicted using an unrooted neighbour-joining tree where pairwise genetic distance among individuals was calculated using the Dice coefficient of dissimilarity (Dice 1945). We also performed an analysis of molecular variance (AMOVA), pooling the data in two ways: (i) by prairie, with the starting null hypothesis that the eleven prairies could be considered together as one large, randomly mating population and (ii) at a larger scale depicting the three ecotypes (CKS, EKS and SIL), adjusting the null hypothesis such that each of the regions were considered as separate, panmictic

populations. The latter was performed based on the neighbour-joining tree suggesting regional genetic differentiation. The AMOVA consisted of 999 random permutations to test these two hypotheses. We also performed a principal coordinate analysis (PCoA), sorting data by prairie and by regional ecotype. The full AFLP marker data set as well as outlier loci were analysed using STRUCTURE version 2.3.3 with 20 000 burn-in and 500 000 Markov chain Monte Carlo (MCMC) steps (Falush *et al.* 2007). Admixture was included in the model and uncorrelated allele frequencies assumed. STRUCTURE HARVESTER (Earl & vonHoldt 2012) was used for the calculation of delta *K* (Evanno *et al.* 2005). Clusters were permutated using CLUMPP (Jakobsson & Rosenberg 2007) and bar plots visualized in DISTRUCT (Rosenberg 2004).

Disentangling the relative contribution of geographical and environmental distance to differentiation

We sought to measure isolation by geographical and environmental distances (IBD vs. IBE) and assess their relative effects on genetic differentiation between populations. To this end, we implemented a modified version of the hierarchical Bayesian model proposed by Bradburd *et al.* (2013), namely Bayesian Estimation of Differentiation in Alleles by Spatial Structure and Local Ecology, as implemented in the R package BEDASSLE. We used the complete data set of 378 plants across 11 prairie populations genotyped at 387 loci. In keeping with Bradburd *et al.* 2013's approach, the binomial distribution on the response variable was defined in terms of frequency of the presence of AFLP marker alleles. To accommodate data overdispersion, we used the beta-binomial modelling approach.

To characterize and reduce the dimensionality of environmental variables and define ecological distance across populations (as required by BEDASSLE), we conducted a principal component analysis on the 10 environmental variables (Table 1) in their original scales. The loadings (i.e. correlations) of the eigenvectors with the environmental variables were inspected to weight the contribution of each environmental variable to each principal component, in particular the first one, which accounted for 99.8% of the variability in the environmental variables across populations. Next, the scores of the first principal component corresponding to each population were computed as surrogates for the environmental variables. Pairwise ecological distances between populations were computed as the difference in scores of the first principal component for the corresponding populations. Pairwise geographical distances (in kilometers) input into BEDASSLE were calculated for all pairs of the 11 populations. Both pairwise distance variables were normalized (i.e. divided) by their standard deviations before model inclusion. After acceptance rates for all parameters fell within the range of 20–70%, as recommended by Bradburd *et al.* (2013), the MCMC was run for 5×10^6 iterations, and the chain thinned every 50 iterations. Trace plots were checked for convergence.

Detection of outlier loci

To ensure robustness in the detection of outlier loci, we used the method proposed by Günther & Coop (2013) as implemented in BAYENV2, which relaxes the assumption of genetic independence among populations. This method corrects for demographic processes that may have led to population divergence while controlling for false positives (Günther & Coop 2013; Lotterhos & Whitlock 2014). In BAYENV2, the 384 polymorphic AFLP loci served as 'control loci' to estimate covariance matrices across four independent runs of 10^6 iteration each (Blair *et al.* 2014). To ensure MCMC convergence, visual inspection of the four covariance matrices was performed. Correlation matrices were generated using the *cov2cor* function in R (R Development Core Team 2011) and compared with pairwise population matrices to confirm high F_{ST} values corresponded with low correlations among populations. All AFLP loci were then tested to identify loci that deviate from the null model of population structure by estimating the test statistic X^TX. Empirical ranks of the X^TX statistic for each marker and the top 3% differentiated outliers were identified across four independent runs of the covariance matrix. We then repeated the covariance matrix estimation, this time removing top-ranked outlier loci from the 'control loci' set to confirm identification of the same top-ranked outliers.

In parallel to BAYENV2 outlier analyses, we also used BAYESCAN 2.1 to detect F_{ST} outlier loci (Foll & Gaggiotti 2008). We acknowledge that this approach has been recently demonstrated to suffer from inflation of false positives, especially under scenarios of IBD or demographic histories such as population range expansion (Lotterhos & Whitlock 2014). Thus, we intended results from BAYESCAN to serve as a cross-check for consistency with those of BAYENV2, while also providing a benchmark for comparison and interpretation with current literature. The BAYESCAN data set was reduced to 325 marker loci after discarding alleles at <2% frequency as recommended by Foll & Gaggiotti (2008). Data were entered by regional ecotype and run parameters included 20 pilot runs of length 5 and 50 K data burn-in, a thinning interval of 10 and a sample size of 5 K. The prior odds for the neutral model was set to 10, but the inbreeding coefficient (F_{IS} prior) allowed to vary between 0.0 and 1.0 (1.0 representing complete inbreeding). Although big bluestem is a self-incompatible

species, a floating F_{IS} prior value was used to avoid introducing biases into F_{ST} estimation (O. Gaggiotti, personal correspondence). The two models that are compared in BayeScan are a neutral model and a model with selection. The BayeScan algorithm was independently repeated three times, and outlier loci selected according to their repeatability across runs and q-values ≥ 0.5 for substantial evidence of selection. The q-value is the FDR analogue of the P-value. A threshold of 5% was chosen (meaning those outliers having a q-value less than 5% are expected to be false positives).

Statistical modelling of the association between outlier loci presence and environment

To identify associations between outlier loci and the environmental gradient, we conducted multivariate logistic regression analyses on each selected outlier locus. All multivariate logistic regression models were fitted using the LOGISTIC procedure of SAS (Version 9.3; SAS Institute, Cary, NC, USA). Prior to data analyses, preliminary screening of environmental variables was implemented to (i) prevent multicollinearity among explanatory variables and to (ii) identify and exclude any explanatory variables for which a quasi-complete separation of data points (i.e. extreme category problem or perfect discrimination; Agresti 2002) was detected. Environmental variables (Table 1) were entered into a stepwise model selection process. For all climatic data, we referred to the National Oceanic and Atmospheric Administration (NOAA) database daily weather records from 1960 to 2011 and extracted pertinent variables to plant growth.

For each outlier locus, we implemented a stepwise selection approach with significance levels for entry and exclusion of 0.05 to identify the most relevant subset of environmental variables (Collett 2003). Outcomes from logistic regression modelling are typically presented in terms of estimated odds ratios (*OR*) and corresponding 95% confidence intervals (CI) per unit increase in the associated predictor variables (Agresti 2002). The *OR* describes the magnitude of the association between a given predictor (i.e. environmental variable) and the odds of a binary response (i.e. presence of outlier marker in a plant genotype), assuming all other selected explanatory variables are held constant. Usually, *OR*s > 1 suggest a positive association between the predictor and the odds of the response, whereas the opposite is true when $OR < 1$. A $(1-\alpha)$% CI on the *OR* that does not include the null value of $OR = 1$ indicates evidence for an α-significant association between the predictor and the odds of the response. To facilitate the interpretation of *OR*s, we also calculated the expected per cent increase (or decrease) in the odds of the presence of that particular outlier locus per unit increase of the environmental predictor, assuming all other selected explanatory variables were held constant.

Results

AFLP genotyping results

A total of 387 AFLP loci and 384 polymorphic bands were identified (mean = 194 bands per primer, st.dev = 47). Most markers were present at ≥25% frequency, with 8% of the data set represented by low frequency alleles (<2% frequency). The overall error rate was 9.2% and thus within the error range typically reported for AFLP studies of 2–10% (Avolio *et al.* 2011; Rouse *et al.* 2011; Price *et al.* 2012).

Ecotype genetic differentiation and structure

The unrooted neighbour-joining tree demonstrated genotypic differentiation among regional ecotypes, with greatest similarity observed between CKS and EKS (Fig. 1). The SIL ecotype was split into several unique branches, largely separated from Kansas prairies. A number of tree branches also included individuals from several prairie sites, indicating among site genetic similarities. Nei's pairwise genetic distance ranged from 0.01 to 0.08 between prairies, indicating mild genetic differentiation across prairies; however, the highest genetic distances were between prairies from different regions (Table 2). A similar trend was observed in the PCoA of the genetic relationships between individuals, with two main genetic clusters formed by SIL and Kansas (CKS and EKS) regional ecotypes (Fig. 2A,B). Kansas and SIL ecotypes were mostly discriminated along the first PCoA axis (38%), with first and second axes representing 61% of the total variation.

Population structure across the U.S. Midwest grasslands landscape was detected in agreement with the PCA (Fig. 3A); most notably, distinct genetic structure was observed between the Kansas (CKS and EKS) and SIL ecotypes (Fig. 3B), with support for $K = 6$ clusters (Evanno *et al.* 2005; Figs S1 and S2, Supporting information). The model converged to this result during both short- and long-chain lengths (MCMC = 10 K and MCMC = 500 K steps, with a burn-in of 10 and 20 K, respectively). Most prairie sites were predominated by a single genetic cluster, with some highly admixed individuals within each prairie (Fig. 3B). Kansas (EKS and CKS) and SIL genetic groups mostly are not overlapping (with the exception of Fult's Hill prairie in SIL that better aligns with prairies from Kansas), supporting genetic differentiation and structuring between regional ecotypes.

Furthermore, several regional or ecotype-specific AFLP markers were identified (four in SIL and six private to EKS), all of which were found segregating at an overall frequency >2% (Table S1, Supporting information). No private markers were found segregating in CKS.

Genetic diversity

When considering all 11 prairie populations, the AMOVA significantly partitioned the most variation within prairies (80%) than across prairies (12%) (Table S2, Supporting information, $P < 0.001$). The remaining total variation (8%) was partitioned between ecotypes from CKS, EKS and SIL regions. When pooling genotype data by regional ecotype, within-ecotype variation was significant, ranging from 84% to 92% ($P < 0.001$). Despite small size and fragmentation of Illinois prairies, these prairies still retained high genetic variation (92% of total variation).

IBE vs. IBD

Isolation by environment was assessed based on environmental predictors, which were subjected to PCA for dimensional reduction. The first principal component on the environmental variables described 99.8% of the variability and accurately separated the prairies into three groups corresponding to regional ecotypes of CKS, EKS and SIL (Fig. S3, Supporting information). The posterior median of the effect size ratio of environmental distances (expressed as first PC scores) to the effect size of geographical distances was 51.2, and the 95% highest posterior density interval was (11.1, 176.3). Departure of posterior effect size from its null value (=1) indicates that genetic differentiation among prairie populations was more heavily influenced by environmental variables than by geographical distance. It is noted that the PC score used to summarize environmental variables is, by definition, a dimensional and thus lacks a meaningful scale; however, the relative contribution of each environmental variable to the first PC score can be considered. The first PC score defining ecological distance was most heavily influenced by elevation (0.99 score units/m) and secondarily, by annual mean precipitation (−0.107 score units/mm). This means that a one-unit difference in environmental distance between two populations expressed in terms of the first PC score (and corresponding to ~1 m in elevation or 9 mm in annual precipitation) had a similar impact on genetic differentiation as ~51 km of geographical distance.

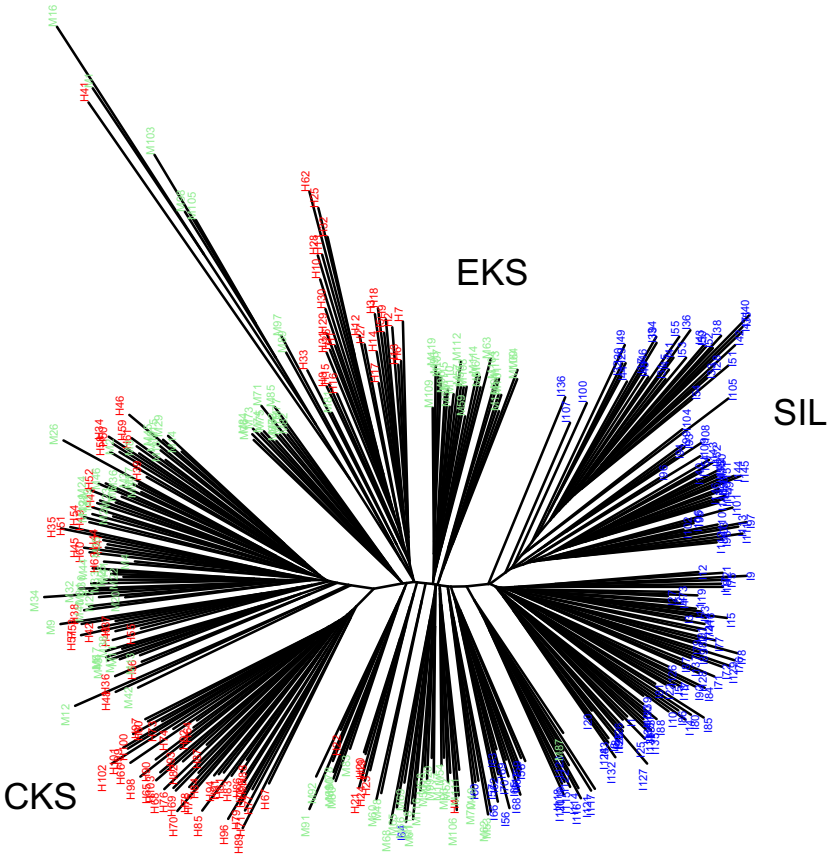

Fig. 1 Unrooted neighbour-joining tree of genetic dissimilarity across individuals. The neighbour-joining tree was built using the Dice coefficient of dissimilarity. Branch tips are colour-coded according to the regional ecotype (Red = Central Kansas; Green = Eastern Kansas; Blue = Southern Illinois.

Outlier loci suggest diversifying selection among ecotypes

Using BAYENV2 and accounting for demographic processes and nonindependence among populations, we identified 14 top-ranked outlier markers (based on X^TX). Importantly, four independent runs of 1 million iterations gave nearly the same ranking and X^TX result for each independent run suggesting convergence. The

Table 2 Pairwise Nei's unbiased genetic distances between 11 sampled prairies. Distances are calculated as: −1*Ln (Nei's Identity) (Nei 1978)

	DES*	FUL*	TM*	WAL*	CAR†	KON†	TAL†	TOW†	CDB‡	SAL‡	WEB‡
DES*	0.000										
FUL*	0.054	0.000									
TM*	0.012	0.048	0.000								
WAL*	0.019	0.044	0.008	0.000							
CAR†	0.047	0.071	0.058	0.060	0.000						
KON†	0.053	0.023	0.048	0.043	0.056	0.000					
TAL†	0.042	0.034	0.038	0.033	0.035	0.022	0.000				
TOW†	0.051	0.030	0.050	0.047	0.043	0.012	0.020	0.000			
CDB‡	0.048	0.038	0.058	0.057	0.029	0.028	0.025	0.020	0.000		
SAL‡	0.053	0.071	0.066	0.066	0.004	0.054	0.037	0.043	0.026	0.000	
WEB‡	0.068	0.076	0.066	0.064	0.032	0.056	0.034	0.048	0.049	0.035	0.000

*Southern Illinois.
†Eastern Kansas.
‡Central Kansas ecotype regions.

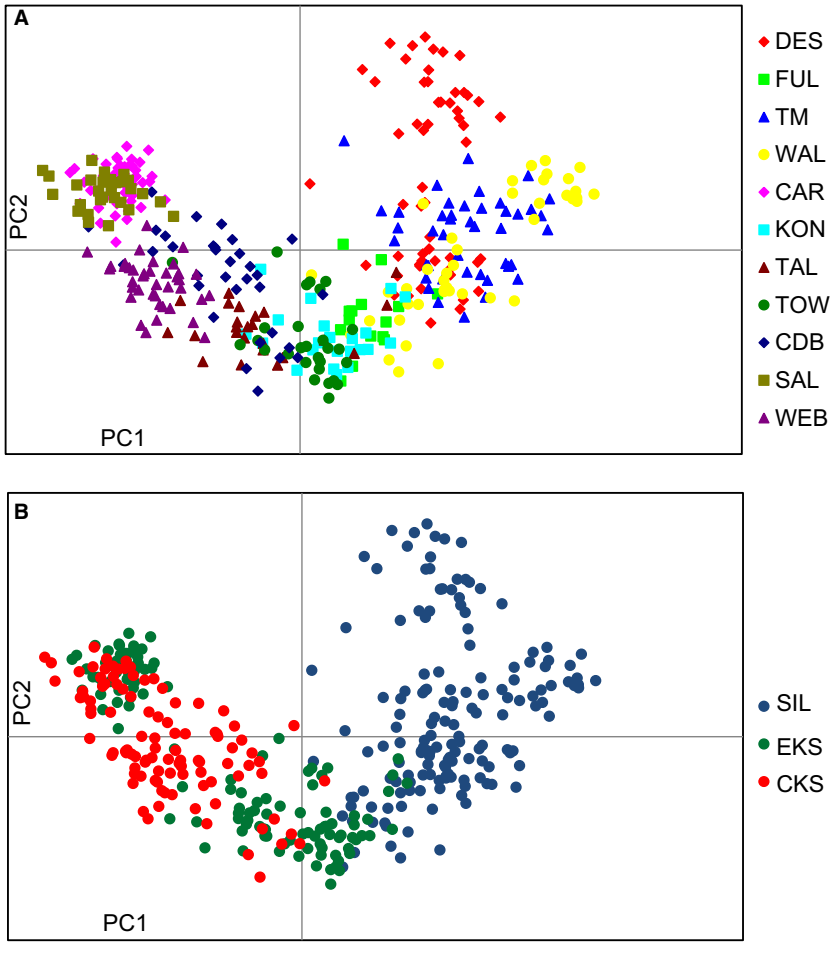

Fig. 2 Genetic principal coordinate analysis of individuals within (A) prairies and (B) regional ecotypes based on the presence/absence of 387 AFLP loci across 378 big bluestem individuals. Abbreviations and symbols correspond to (A) individual prairies listed in Table 1 and (B) regional ecotypes (Red = Central Kansas; Green = Eastern Kansas; Blue = Southern Illinois). Kansas prairies are differentiated from Illinois prairies in the first two axes (axis 1 = 38% and axis 2 = 23% of the variation explained; total variation explained = 61%).

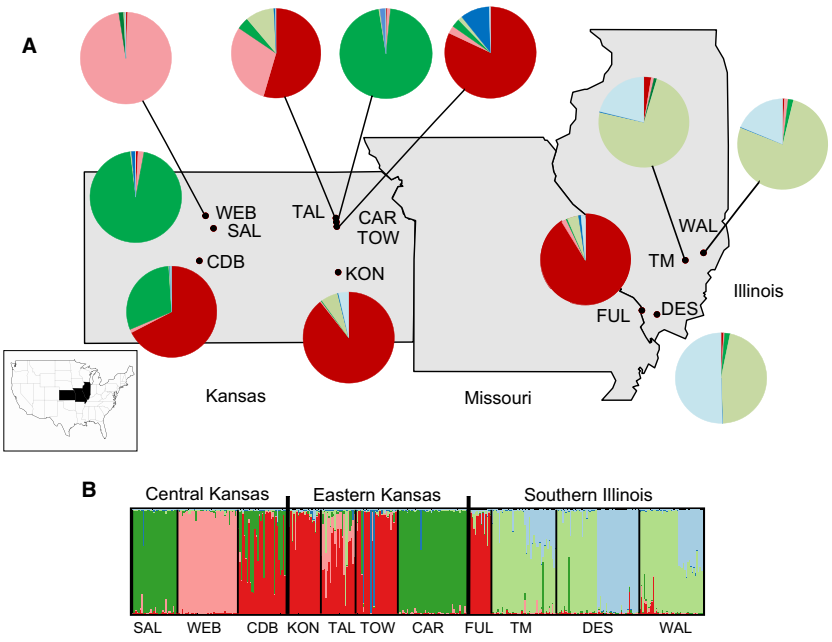

Fig. 3 STRUCTURE (A) individual membership pie charts overlaid across the U.S. Midwest environmental gradient and (B) bar plot labelled by regional ecotype and by prairie. The most likely genetic grouping solution, $K = 6$, is shown. Each colour indicates one genetic group, and each bar represents percentage membership to genetic group(s). Mixed membership indicates admixture.

identity and rankings of the markers with highest X^TX are provided in Table 3.

In turn, the BAYESCAN 2.1 analysis yielded seven F_{ST} outliers (2% of the total number of AFLP marker loci) across independent runs of the algorithm, six of which overlapped with the outliers identified using BAYENV2 (described above). An average overall species F_{ST} of 0.1 was determined. All loci were deemed 'high outliers' under diversifying selection and were highly differentiated among ecotypes with locus-specific F_{ST} = 0.3–0.5 (Fig. 4). Additionally, in pairwise comparisons of locus-specific F_{ST} values between regional ecotypes in BAYESCAN (data not shown), the EKS vs. SIL comparison yielded five highly differentiated markers, four of which were also outlier loci differentiating the three regional ecotypes. The EKS vs. CKS and CKS vs. SIL outlier analyses identified one outlier in each case. BAYESCAN outliers and their commonality with BAYENV2 outliers are provided in Table 3.

Association between AFLP locus presence and environmental predictors

All outlier loci were associated with two or more environmental variables (Table 3). More specifically, top-ranked outliers were significantly related to temperature severity (14 of 14) and annual mean temperature (11 of 14). Importantly, seasonal mean precipitation and seasonal mean temperature had large effect sizes for six outlier loci each, suggesting importance of seasonal factors on ecotype differentiation. Table 3 shows associations between outlier loci and environmental variables using *ORs* and corresponding 95% confidence intervals. Take for example, the significant association of outlier M228 and prairie elevation, with an estimated *OR* of 1.037 which can be interpreted as an expected 3.7% increase in the odds of the presence of this outlier for every 1 m increase in elevation, provided the remaining model variables are held constant.

In general, explanatory variables related to some aspect of precipitation were estimated to have 'large effects' (i.e. associated with large increases or decreases) on the odds of observing an outlier such as seasonal mean precipitation (M250), precipitation amount in the driest year (M232) and number of heavy precipitation events (i.e. >1.25 cm per event) per year (M371). In these cases, every unit increase in the corresponding precipitation-related predictors was expected to more than triple, or even quadruple, the odds of observing these outlier loci. In summary, outlier loci were linked to multiple aspects of both temperature and precipitation across the environmental gradient of the Midwest grasslands.

Discussion

Habitats are often both temporally and spatially variable and ultimately may lead to species' genetic differentiation. We highlight here population divergence and ecotypic variation of a foundation prairie grass across the environmental gradient of U.S. Midwest grasslands. Despite large geographical distances between regional populations and fragmentation of the prairie ecosystem

Table 3 Estimated odds ratios *(ORs)* describing extent of the environmental association of AFLP outlier markers identified in BAYENV2$_{(BE)}$, and both BAYENV2 and BAYESCAN 2.1$_{(BE,BS)}$. The best fitting environmental predictors were selected using a stepwise variable selection approach. Percentage increase (or decrease) in the odds of the outlier being present in the plant genotype per unit increase of the environmental predictor, given all other variables are held constant, is provided in ( ). Confidence intervals of estimated *ORs* are provided in [ ]. For large *OR* estimates, estimated fold-change is presented in place of percentage increase or decrease in odds (i.e. >3× corresponds to the odds of marker presence more than tripling per unit change in the predictor variable). BAYENV2 outliers are listed according to ranked highest–lowest differentiation (X^TX)

Outlier	Elevation (m)	Precip events >1.25 (cm)	Precip in driest year (cm)	Ann mean temp (°C)	Ann diurnal (°C)	Ann mean precip (cm)	Seasonal mean temp (°C)	Seasonal diurnal (°C)	Seasonal mean precip (°C)	Temp sev index
M228$_{(BE,BS)}$	1.037 (+3.7%) [1.024, 1.052]	—	—	—	0.668 (−33.2%) [0.505, 0.817]	—	1.086 (+8.6%) [1.037, 1.146]	—	1.429 (+42.9%) [1.241, 1.677]	1.036 (+3.6%) [1.018, 1.059]
M232$_{(BE,\ BS)}$	0.761 (−23.9%) [0.691, 0.821]	68.217 (>60×) [20.944, 279.496]	3.080 (>3×) [2.233, 4.613]	—	0.198 (−80.2%) [0.099, 0.323]	0.013 (−98.7%) [0.003, 0.042]	—	1.163 (+16.3%) [1.113, 1.240]	—	0.988 (−1.2%) [0.977, 1.000]
M252$_{(BE,BS)}$	—	—	—	0.933 (−6.7%) [0.911, 0.954]	—	—	1.588 (+58.8%) [1.428, 1.794]	0.951 (−4.9%) [0.940, 0.961]	—	1.109 (+10.9%) [1.084, 1.139]
M292$_{(BE,BS)}$	0.952 (−4.8%) [0.929, 0.967]	—	1.389 (+38.9%) [1.222, 1.697]	0.928 (−7.2%) [0.889, 0.957]	0.764 (−23.6%) [0.566, 0.955]	—	—	—	0.300 (−70.0%) [0.180, 0.413]	1.059 (+5.9%) [1.041, 1.085]
M17$_{(BE)}$	0.966 (+3.4%) [0.955, 0.976]	—	—	1.054 (−5.4%) [1.038, 1.071]	—	0.841 (+15.9) [0.782, 0.901]	—	—	—	1.014 (−1.4%) [1.006, 1.021]
M248$_{(BE)}$	0.979 (+2.1%) [0.963, 0.993]	1.891 (−89.1%) [1.557, 2.335]	—	0.914 (+8.6%) [0.886, 0.939]	—	0.621 (+37.9%) [0.512, 0.731]	1.320 (−32.0%) [1.233, 1.436]	0.939 (+6.1%) [0.924, 0.951]	—	1.089 (−8.9%) [1.069, 1.114]
M242$_{(BE)}$	0.962 (+3.8%) [0.947, 0.974]	2.781 (−178.1%) [1.546, 6.760]	1.650 (−65.0%) [1.483, 1.870]	0.918 (+8.2%) [0.889, 0.945]	—	—	—	—	0.340 (+66.0%) [0.247, 0.442]	1.089 (−8.9%) [1.067, 1.119]
M241$_{(BE)}$	—	1.939 (−93.9%) [1.522, 2.653]	1.599 (−59.9%) [1.375, 2.036]	0.903 (+9.7%) [0.871, 0.929]	0.687 (+31.3%) [0.580, 0.784]	—	—	—	0.502 (+49.8%) [0.363, 0.633]	1.095 (−9.5%) [1.067, 1.139]
M256$_{(BE,BS)}$	1.029 (+2.9%) [1.019, 1.042]	1.871 (+87.1%) [1.483, 2.531]	—	0.977 (−2.3%) [0.960, 0.993]	0.723 (−27.7%) [0.643, 0.798]	—	—	—	—	1.026 (+2.6%) [1.016, 1.037]
M238$_{(BE)}$	—	—	0.883 (+11.7%) [0.776, 0.994]	0.912 (+8.8%) [0.885, 0.935]	1.188 (−18.8%) [1.043, 1.359]	—	1.255 (−25.5%) [1.183, 1.346]	0.936 (+6.4%) [0.912, 0.956]	—	1.078 (−7.8%) [1.060, 1.098]
M371$_{(BE,BS)}$	—	4.159 (>4×) [2.951, 6.526]	1.536 (+53.5%) [1.350, 1.816]	0.823 (−17.7%) [0.769, 0.865]	0.477 (−52.3%) [0.351, 0.603]	—	—	—	0.288 (−71.2%) [0.190, 0.391]	1.102 (+10.2%) [1.076, 1.136]
M237$_{(BE)}$	0.983 (+2.7%) [0.973, 0.992]	0.590 (+54.0%) [0.460, 0.725]	—	1.023 (−2.3%) [1.007, 1.041]	1.284 (−28.4%) [1.177, 1.414]	—	—	—	—	0.969 (−95.9%) [0.960, 0.977]
M1$_{(BE)}$	0.891 (10.9%) [0.862, 0.915]	2.748 (−174.8%) [1.775, 4.675]	1.599 (−59.9%) [1.373, 1.936]	—	—	0.130 (−30%) [0.059, 0.242]	1.540 (−54%) [1.340, 1.837]	—	3.548 (−254.8%) [2.080, 6.953]	1.075 (−7.5%) [1.058, 1.100]
M268$_{(BE)}$	0.959 (+4.1%) [0.945, 0.970]	—	—	1.073 (−7.3%) [1.049, 1.101]	—	0.741 (25.9%) [0.650, 0.833]	0.942 (+5.8%) [0.894, 0.989]	1.034 (−3.4%) [1.022, 1.051]	—	0.966 (+3.4%) [0.952, 0.979]

divergence across neutral and non-neutral outlier loci is more strongly related to factors of regional climate (IBE) than geographical isolation. We show genomewide markers under divergent selection among ecotypes are associated with several temperature and precipitation-related environmental predictors, especially seasonal rainfall and especially precipitation that has 'large effect' on outlier presence. The high genetic diversity within and among populations may enable this foundation grass to withstand environmental change and should guide restoration efforts.

High genetic diversity maintained despite population structure across U.S. Midwest grasslands

Partitioning genetic variation found within prairie vs. across prairies is informative to population processes, spatial genetic differentiation and restoration genetics (Jones 2003). When genetic variation of big bluestem was partitioned within prairies, across prairies and across regions, the highest genetic variation (80%, $P < 0.001$) was observed within prairies (Table S2, Supporting information) across the expanse of the Midwest. We detected high levels of diversity even in the small, fragmented Illinois prairies when we had originally expected reduced genetic diversity in these remnant prairies as they are possibly more prone to genetic drift (Wright 1938). It is unclear whether the observed high diversity in Illinois remnant prairies is a legacy of the once expansive eastern tall grass prairie prior to conversion to agriculture and landscape fragmentation. Nevertheless, results of high within-prairie diversity in Illinois agree with other big bluestem studies 86% in Wisconsin and Northeast U.S. prairies (Price *et al.* 2012) and 89% within-prairie diversity in Illinois and Arkansas prairies (Gustafson *et al.* 1999). Similar patterns of high within-population genetic diversity have been observed in other outcrossing prairie grasses, namely switchgrass (Morris *et al.* 2013; Mutegi *et al.* 2014). In summary, high genetic diversity observed within prairies may provide sufficient genetic material on which selection can act and may play a role in partially buffering these populations in a changing environment (Shaw & Etterson 2012).

High within-prairie genetic diversity can be expected for several reasons. Big bluestem is highly self-incompatible, with low to inviable seed production following selfing (Norrmann *et al.* 1997). This is consistent with previous studies showing increased genetic variation as a result of obligate outcrossing (Gustafson *et al.* 1999; Bomblies *et al.* 2010; Price *et al.* 2012; Mutegi *et al.* 2014). Furthermore, the complex polyploid genome of big bluestem (Norrmann *et al.* 1997; Keeler 2004) may have consequences for allelic variation and genetic diversity. Rouse *et al.* (2011) refute this hypothesis as AFLP profile dissimilarity is not related to ploidy when genotyping plants of different ploidy levels. Thus, it would seem that high genetic diversity may be attributed primarily to the outcrossing nature of big bluestem rather than ploidy variation.

Interestingly, population genetic structure exists across big bluestem regional ecotypes (Figs 2 and 3). Illinois populations remain mostly distinct from Kansas populations based on STRUCTURE analysis (with the exception of Fult's Prairie). Furthermore, results agreed with genetic distance-based methods such as neighbour-joining and PCoA, which showed major clusters representing Illinois and Kansas ecotypes. While ploidy differences, as observed in big bluestem (Keeler 2004), can complicate analysis of population structure, this is now accommodated in recent software (Falush *et al.* 2007), making it unlikely that ploidy differences across plants used in this study are solely driving genetic

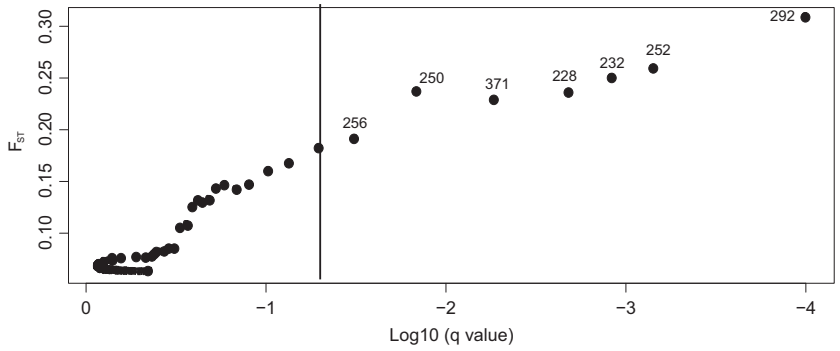

Fig. 4 F_{ST} outlier analysis of 325 polymorphic AFLP markers in BAYESCAN 2.1. Plot shows F_{ST} vs. $\log_{10}$-transformed q-values (the minimum FDR at which a locus become significant). Data were organized according to the regional ecotype (Central Kansas, Eastern Kansas and Southern Illinois). The observed global $F_{ST} = 0.1$. Seven marker loci are 'high outliers' with greater genetic differentiation than expected under neutrality (FDR = 0.05, vertical line shows significance cut-off). Loci are under positive or diversifying selection ($\alpha > 0$).

structure among ecotypes. Moreover, ecotypes grown from seed for this study are mostly 6× based on flow cytometry (L. C. Johnson & J. Gaffney unpublished data). Therefore, we do not expect ploidy bias towards observed population structure across the environmental gradient. Genetic structure is also unlikely to be an artefact of the seed collection method as we collected seed within prairie at multiple locations and times, making it unlikely we sampled genetic clones.

IBE prominent over IBD

The prominence of IBE suggests factors related to the environment play a greater role in divergence of bluestem populations than geographical isolation. Possible mechanisms responsible for IBE are selection pressures from historical climate that have been in place for *c*. 10 000 years (Axelrod 1985), namely the more than twofold difference in precipitation from central KS to Illinois, as well as the corresponding range of temperatures, both average and extreme. However, environmental factors can shape gene flow (i.e. environment affecting phenological differences among populations) and ultimately, constrain gene flow. Thus, IBE may not be due solely to selection but could be confounded or even explained by the impact of environmental factors on gene flow. While less strong than IBE, IBD could be explained by geographical distance, increasing fragmentation from other land uses (such as agriculture, forest and residential, GLCCD 1998) and small prairie size (Table 1) moving eastward. All of these could effectively disrupt gene flow in the Midwest.

Outlier loci linked to climate variables

In identifying outlier loci, we sought to determine how selection may play a role in shaping genetic ecotypic differentiation along sharp environmental clines. All seven loci identified in BayeScan as undergoing putative diversifying selection (Fig. 4) were associated with environmental predictors across the U.S. Midwest environmental gradient (Table 3), suggesting these regions of the genome seem to be diverging and that climate may play a role.

Most outliers (13 of 14 BAYENV2) were associated with precipitation related predictors, probably due to the steep gradient in precipitation along our sampled region. In addition, all outliers were associated with temperature-related environmental predictors, suggesting that temperature may also be exerting spatially divergent pressure on ecotypes. Although seasonal mean precipitation is associated with few outliers (6 of 14), it has a 'large effect' on whether each of these is observed. This result suggests that perhaps ecotypes may be more challenged by seasonal rainfall amounts or drought events during the growing season than by annual precipitation, which includes periods of plant dormancy. Looking ahead, this presents a problematic scenario given that climate change predictions (IPCC 2013) for the U.S. Midwest forecast extreme events of drought during the summer growing season of C4 grasses. In summary, results reveal a more integrated and complex relationship of outlier marker presence with multiple environmental predictors, rather than the presence of a single, major driving factor as was originally hypothesized of mean annual precipitation.

Genetic divergence and ecotype local adaptation

Genetic divergence studies have also been related to parallel phenotypic divergence among ecotypes. Grass ecotypic differentiation was first reported in the seminal studies of McMillan (1956) and more recently, in switchgrass (Aspinwall *et al.* 2013; Lowry *et al.* 2014). In our case, we identified and associated diversifying selection that may be informative to the phenomenon of local adaptation observed in big bluestem ecotypes in an on-going reciprocal garden study (L. C. Johnson, S. G. Baer & B. R. Maricle, unpublished data). In this complementary study, gardens were seeded at four sites (including central Kansas, eastern Kansas and Illinois) that span 1150 km of the U.S. Midwest prairie. We seeded the same ecotypes studied here (CKS, EKS and SIL) and identified that climatic differences across this environmental gradient appear to have exerted strong selection, resulting in phenotypically based local adaptation to 'home' environments. Specifically, we found local adaptation of the CKS and SIL ecotypes to their home environments, including differences in reproductive timing. While our interpretation is limited due to the fact that these are not the same exact seed genotyped here, the phenomenon of local adaptation in the these ecotypes suggest that in spite of gene flow, large population sizes and an outbreeding mating system, climatic selection pressures are potentially strong enough to result in local adaptation. Additionally, local adaptation, in spite of gene flow, has indeed been observed in other systems (Sambatti & Rice 2006; Gonzalo-Turpin & Hazard 2009). Finally, on the basis of these strong phenotypic (M. B. Galliart, J. T. Olsen, H. M. Tetreault, S. Sabates, J. Bryant, A. De La Cruz, L. Wilson, D. Gibson, N. M. Bello, T. J. Morgan, S. G. Baer, B. R. Maricle & L. C. Johnson, unpublished data), ecological (L. C. Johnson, S. G. Baer & B. R. Maricle, unpublished data) and genetic differences (this study) among bluestem ecotypes, we recognize each of the ecotypes as being distinct from one another.

Comparison of AFLP genome scan with next-generation sequencing methods

Technical capabilities to acquire more comprehensive sequencing data have dramatically increased in recent years, particularly with the advent of next-generation massive parallel sequencing technologies. Here, we employed an AFLP genome scan; however, we acknowledge that questions of both adaptative and neutral adaptive divergence can be probed more comprehensively using DNA sequencing methods such as genotyping by sequencing (GBS, Elshire *et al.* 2011) and double-digestion RAD-seq (Peterson *et al.* 2012), in which thousands of informative loci are generated, rather than hundreds. In our AFLP genome scan, we identified 3.6% of total polymorphic AFLP loci to be outliers. The percentage of outliers detected in our study is in line with current next-generation sequencing models. For instance, Larson *et al.* (2014) in a GBS study in Chinook salmon identified 6.7% of total 10 K SNPs as outliers while Hess *et al.* (2012) identified 3.6% of loci as outliers in Pacific lamprey using restriction site-associated DNA sequencing. Reassuringly, in a preliminary GBS study in big bluestem generating 4 K SNPs (M. B. Galliart & L. C. Johnson, unpublished data), we found similar frequency of outliers, genetic structure and differentiation as in this study using only 384 AFLP loci. The percentage of outlier loci detected in our study was also in agreement with those uncovered in recent AFLP genome scans in alpine plants (9%), bitter vine (2.9%), periwinkles (5%) and mussel species (2%) (Poncet *et al.* 2010; Tice & Carlon 2011; Wang *et al.* 2012; Gosset & Bierne 2013, respectively).

Implications for restoring threatened tallgrass prairie in changing climates

Tallgrass prairie restoration efforts will benefit from understanding how much underlying genetic diversity exists in ecotypes of this foundation grass species. Widely used to improve environmental quality or re-create historical plant assemblages, this study demonstrates that big bluestem populations possess high genetic diversity within regions and within populations, including small, isolated populations (e.g. in Illinois). These results support recommendations to use local ecotypes in restoration (Gustafson *et al.* 2001; McKay *et al.* 2005), particularly if retaining historical genetic structure is the goal of restoration. Introducing genetic mixtures and correspondingly high genetic diversity has been proposed as a restoration strategy to mitigate the effects of climate change (Jump & Penuelas 2005; Broadhurst *et al.* 2008; Nicotra *et al.* 2010). High within-prairie genetic variation and local selection (Avolio *et al.* 2011) may enable the persistence of big bluestem populations under predicted greater climatic variability (IPCC 2013). Mixing populations would increase genetic variation of propagules, potentially buffering the effect of climate change in mesic regions if dry-adapted ecotypes are included. However, the relative success of different populations in these mixtures is unknown, and ecological context is an important consideration, as nonlocal seed in restoration can pose genetic risks to extant populations (Hufford & Mazer 2003; McKay *et al.* 2005; Cremieux *et al.* 2009; Schiffers *et al.* 2013).

Our study has relevance for other grasslands worldwide, as these regions are among the most threatened of biomes in need of protection and restoration (Hoekstra *et al.* 2005). Investigations of genetic variation in ecologically dominant foundation species within the current and changing climate of the U.S. Midwest may help make meaningful predictions regarding grassland response and restoration in the face of climate change.

Acknowledgements

We extend gratitude to park and landowners for permission to collect seed used in this publication. We thank Dr. Bai and the Kansas State University DNA Sequencing and Genotyping Facility for technical support, Dr. Shichen Wang for computational assistance, Richard Wynia (USDA Plant Materials Centre, Manhattan, KS) for field support, Lauren Wheeler for aid in leaf collection, Dr. Adam Sparks for assistance with R scripts, Gideon Bradburd for assistance with BEDASSLE and Dr. Torsten Günther for detailed information and code for the implementation of Bayenv2. Dr. Christopher Toomajiian, Dr. Oscar Gaggiotti, Dr. Victoria Sork, Renato C. Nali and three anonymous reviewers provided valuable suggestions on earlier versions of this manuscript. Funding was provided from the USDA (Grant no. 2008-35100-04545), KSU GK-12 Programme Graduate Student Fellowship (NSF DGE-0841414) to M.M.G, as well as from the Kansas Native Plant Society/Grassland Heritage Foundation scholarship and Sigma Xi Grants-in-Aid of Research to M.M.G. This is Kansas Agricultural Experiment Station number 13-099-J.

References

Agresti A (2002) *Categorical Data Analysis*, 2nd edn. Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, New Jersey.

Anderson JT, Lee C, Rushworth CA, Colautti RI, Mitchell-Olds T (2013) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology*, **22**, 699–708.

Aspinwall MJ, Lowry DB, Taylor SH *et al.* (2013) Genotypic variation in traits linked to climate and aboveground productivity in a widespread C4 grass: evidence for a functional trait syndrome. *New Phytologist*, **199**, 966–980.

Avolio ML, Chang CC, Smith MD (2011) Assessing fine-scale genotypic structure of a dominant species in native grasslands. *American Midland Naturalist*, **165**, 211–224.

Axelrod DI (1985) Rise of the grassland biome, central North America. *Botanical Review*, **51**, 163–201.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 1619–1626.

Blair LM, Granka JM, Feldman MW (2014) On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, **8**, 1–13.

Bomblies K, Yant L, Laitinen RA *et al.* (2010) Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000890.

Bonin A, Bellemain E, Eidesen PB, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**, 3261–3273.

Bradburd GS, Ralph PL, Coop GM (2013) Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, **67**, 3258–3273.

Broadhurst LM, Lowe A, Coates DJ *et al.* (2008) Seed supply for broadscale restoration: maximizing evolutionary potential. *Evolutionary Applications*, **1**, 587–597.

Carl G, Kuhn I (2007) Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling*, **207**, 159–170.

Chen J, Kallman T, Ma X *et al.* (2012) Disentangling the roles of history and local selection in shaping clinal variation of allele frequencies and gene expression in Norway spruce (*Picea abies*). *Genetics*, **191**, 865–881.

Collett D (2003) *Modeling Binary Data*, 2nd edn. Chapman & Hall/CRC Press, London.

Conner JK, Hartl DL (2004) *A Primer of Ecological Genetics*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Corbett EA (2004) *A comparison of Illinois remnant prairies, 1976–1988*. University of Nebraska-Lincoln, North American Prairie Conference.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Crawford LA, Koscinski D, Keyghobadi N (2012) A call for more transparent reporting of error rates: the quality of AFLP data in ecological and evolutionary research. *Molecular Ecology*, **21**, 5911–5917.

Cremieux L, Bischoff A, Muller-Scharer H, Steinger T (2009) Gene flow from foreign provenances into local plant populations: fitness consequences and implications for biodiversity restoration. *American Journal of Botany*, **97**, 94–100.

De La Torre AR, Roberts DR, Aitken SN (2014) Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular Ecology*, **23**, 2046–2059.

De Mita S, Thuillet A, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383–1399.

Dice LR (1945) Measures of the amount of ecological association between species. *Ecology*, **26**, 297–302.

Dobzhansky TG (1937) *Genetics and the Origin of Species*. Columbia University Press, New York, New York.

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Epstein HE, Lauenroth WK, Burke IC, Coffin DP (1998) Regional productivities of plant species in the Great Plains of the United States. *Plant Ecology*, **134**, 173–195.

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

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, **7**, 574–578.

Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–993.

Franks SJ, Weber JJ, Aitken SN (2014) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications*, **7**, 123–139.

Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.

Frichot E, Schoville S, Bouchard G, Francois O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699.

Gaggiotti OE, Bekkevold D, Jorgensen HBH *et al.* (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. *Evolution*, **63**, 2939–2951.

GLCCD (1998) *Global land cover characteristics database, Version 1.2*. Available from http://edc2.usgs.gov/glcc/globdoc1_2.php.

Gonzalo-Turpin H, Hazard L (2009) Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*. *Molecular Ecology*, **97**, 742–751.

Gosset CC, Bierne N (2013) Differential introgression from a sister species explains high F_{ST} outlier loci within a mussel species. *Journal of Evolutionary Biology*, **26**, 14–26.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Gustafson DJ, Gibson DJ, Nickrent DL (1999) Random amplified polymorphic DNA variation among remnant big bluestem (*Andropogon gerardii* Vitman) populations from Arkansas' Grant Prairie. *Molecular Ecology*, **8**, 1693–1701.

Gustafson DJ, Gibson DJ, Nickrent DL (2001) Characterizing three restored *Andropogon gerardii* Vitman (big bluestem) populations established with Illinois and non-Illinois seed: established plants and their offspring. *Proc. 17th N.A. Prairie Conference*, 118–124.

Gustafson DJ, Gibson DJ, Nickrent DL (2002) *Genetic Diversity and Competitive Abilities of Dalea purpurea (Fabaceae) from Remnant and Restored Grasslands*. Publications Paper 4 Southern Illinois University, Carbondale, Illinois.

Gustafson DJ, Gibson DJ, Nickrent DL (2004a) Competitive relationships of *Andropogon gerardii* (Big bluestem) from remnant and restored native populations and select cultivated varieties. *Functional Ecology*, **18**, 451–457.

Gustafson DJ, Gibson DJ, Nickrent DL (2004b) Conservation genetics of two co-dominant grass species in an endangered grassland ecosystem. *Journal of Applied Ecology*, **41**, 389–397.

Hancock AM, Brachi B, Faure N *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, **334**, 83–86.

Harris JA, Hobbs RJ, Higgs E, Aronson J (2006) Ecological restoration and global climate change. *Restoration Ecology*, **14**, 170–176.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2012) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**, 2898–2916.

Hoekstra JM, Boucher TM, Ricketts TH, Roberts C (2005) Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, **8**, 23–29.

Hoffman AA, Sigro M (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.

Holdenregger R, Buehler D, Gugerli F, Manel S (2010) Landscape genetics of plants. *Trends in Plant Science*, **15**, 675–683.

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends in Ecology and Evolution*, **18**, 147–155.

IPCC (2013) *Climate change 2013: physical science basis*. Contribution of Working Group I to the Fifth Assessment Report of the IPCC.

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

Jenkins DG, Carey M, Czerniewska J *et al.* (2010) A meta-analysis of isolation by distance: relic or reference standard for landscape genetics? *Ecography*, **33**, 315–320.

Jones TA (2003) The restoration gene pool concept: beyond the native versus non-native debate. *Restoration Ecology*, **11**, 281–290.

Jones MR, Forester BR, Teufel AI *et al.* (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–3468.

Joost S, Vuilleumier S, Jensen JD *et al.* (2013) Uncovering the genetic basis of adaptive change: on the intersection of landscape genomics and theoretical population genetics. *Molecular Ecology*, **22**, 3659–3665.

Jump AS, Penuelas J (2005) Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters*, **8**, 1010–1020.

Keeler KH (2004) Impact of intraspecific polyploidy in *Andropogon gerardii* (Poaceae) populations. *The American Midland Naturalist*, **152**, 63–74.

Knapp AK, Briggs JM, Koelliker JK (2001) Frequency and extent of water limitation to primary production in a mesic temperate grassland. *Ecosystems*, **4**, 19–28.

Larson WA, Seeb LW, Everett MV, Waples RK, Templin WD, Seeb JE (2014) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, **7**, 355–369.

Lee CR, Mitchell-Olds T (2011) Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. *Molecular Ecology*, **20**, 4631–4642.

Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. *PLoS ONE*, **3**, 1–8.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization of the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Lowry DB, Behrman KD, Grabowski P, Morris GP, Kiniry JR, Juenger TE (2014) Adaptation between ecotypes and along environmental gradients in *Panicum virgatum*. *The American Naturalist*, **183**, 682–692.

Manel S, Joost S, Epperson BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.

Manel S, Gugerli F, Thuiller W *et al.* , IntraBioDiv Consortium (2012) Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, **21**, 3729–3738.

Mayr E (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge, Massachusetts.

McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?"—a review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, **13**, 432–440.

McMillan C (1956) Ecotypic differentiation within four North American prairie grasses. II. Behavioral variation within transplanted community fractions. Ecotypic differentiation within four North American prairie grasses. II. Behavioral variation within transplanted community fractions. *American Journal of Botany*, **52**, 55–65.

Miermans PG (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Morris GP, Grabowski PP, Borevitz JO (2013) Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology*, **20**, 4938–4952.

Mueller UG, Wolfenbarger LL (1999) AFLP genotyping and fingerprinting. *Tree*, **14**, 389–394.

Mutegi E, Stottlemyer AL, Snow AA, Sweeney PM (2014) Genetic structure of remnant populations and cultivars of switchgrass (*Panicum virgatum*) in the context of prairie conservation and restoration. *Restoration Ecology*, **22**, 223–231.

National Oceanic & Atmospheric Administration (2012) *U.S. Department of Commerce U.S. Drought Monitor*.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.

Nicotra AB, Atkin OK, Bonser SP *et al.* (2010) Plant phenotypic plasticity in a changing climate. *Trends in Plant Science*, **15**, 684–692.

Norrmann GA, Quarin CL, Keeler KH (1997) Evolutionary implications of meiotic chromosome behavior, reproductive biology, and hybridization in 6X and 9X cytotypes of *Andropogon gerardii* (Poaceae). *American Journal of Botany*, **84**, 201–207.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, New York, New York.

Nosil P, Crespi BJ (2004) Does gene flow constrain adaptive divergence or vice versa? A test using ecomorphology and sexual isolation in *Timema cristinae* walking-sticks. *Evolution*, **58**, 102–112.

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

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for De Novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Poncet BN, Herrmann D, Gugerli F *et al.* (2010) Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology*, **19**, 2896–2907.

Price D, Salon P, Casler MD (2012) Big bluestem gene pools in the Central and Northeastern United States. *Crop Science*, **52**, 189–200.

R Development Core Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org/.

Reusch TBH, Wood TE (2007) Molecular ecology of global change. *Molecular Ecology*, **16**, 3973–3992.

Rice KJ, Emery NC (2003) Managing microevolution: restoration in the face of global change. *Frontiers in Ecology and the Environment*, **1**, 469–478.

Risser PG, Birney EC, Blocker HD, May SW, Parton WJ, Wiens JA (1981) *The True Prairie Ecosystem*. Hutchinson and Ross, Stroudsburg, Pennsylvania.

Robertson KR (1996) *50 Years of change in Illinois hill prairies*. Available from http://www.inhs.uiuc.edu/~kenr/hillprairie.html.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rouse MN, Saleh AA, Seck A *et al.* (2011) Genomic resistance gene homolog diversity of the dominant tallgrass prairie species across the U.S. Great Plains precipitation gradient. *PLoS ONE*, **6**, e17641.

Sala OE, Parton WJ, Joyce LA, Lauenroth WK (1988) Primary production of the central grassland region of the United States: spatial pattern and major controls. *Ecology*, **69**, 40–45.

Sambatti J, Rice KJ (2006) Local adaptation, patterns of selection, and gene flow in the Californian serpentine sunflower (*Helianthus exilis*). *Evolution*, **60**, 696–710.

Samson F, Knopf F (1994) Prairie conservation in North-America. *BioScience*, **44**, 418–421.

Savolainen O, Lascoux M, Merila J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–820.

Schiffers K, Bourne EC, Lavergne S, Thuiller W, Travis JMJ (2013) Limited evolutionary rescue of locally adapted populations facing climate change. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **368**, 1–10.

Selbo SM, Snow AA (2005) Flowering phenology and genetic similarity among local and recently introduced populations of *Andropogon gerardii* in Ohio. *Restoration Ecology*, **13**, 441–447.

Sexton JP, Hangartner SB, Hoffman AA (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, **68**, 1–15.

Shafer A, Wolf J (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Shaw RG, Etterson JR (2012) Rapid climate change and the rate of adaptive insight from experimental quantitative genetics. *New Phytologist*, **195**, 752–765.

Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Temunovic M, Frascaria-lacoste N, Franjic J, Satovic Z, Fernandez-Manjarres JF (2013) Identifying refugia from climate change using coupled ecological and genetic data in a transitional Mediterranean-temperate tree species. *Molecular Ecology*, **22**, 2128–2142.

Tice KA, Carlon DB (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology*, **24**, 1814–1825.

Tompkins RD, Stringer WC, Bridges WC Jr (2011) An outcrossing reciprocity study between remnant big bluestem (*Andropogon gerardii*) populations in the Carolinas. *Ecological Restoration*, **29**, 339–345.

Tonsor SJ (2012) Population genomics and the causes of local differentiation. *Molecular Ecology*, **21**, 5393–5395.

Wang IJ (2013) Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, **67**, 3403–3411.

Wang T, Chen G, Zan Q, Wang C, Y-j Su (2012) AFLP genome scan to detect genetic structure and candidate loci under selection for local adaptation of the invasive weed *Mikania micrantha*. *PLoS ONE*, **7**, e41310.

Wang IJ, Glor RE, Losos JB (2013) Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecology Letters*, **16**, 175–182.

Wright S (1938) Size of population and breeding structure in relation to evolution. *Science*, **87**, 430–431.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wright S (1949) The genetic structure of populations. *Annals of Human Genetics*, **15**, 323–354.

Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P (2014) Genomic signature of adaptation to climate in *Medicago truncatula*. *Genetics*, **196**, 1263–1275.

Zhou Y, Zhang L, Liu J, Wu G, Salvolainen O (2014) Clinal adaptation and ecological divergence between two closely related pine species in Southeast China. *Molecular Ecology*, **23**, 3502–3522.

L.C.J., P.S., E.D.A., T.J.M., K.A.G., S.J.B. and B.R.M. conceived of the experiment. M.M.G. performed experiment and data analyses. P.S. provided assistance with AFLP development and analysis. N.B. implemented marker-environment statistical models. M.K. created the climatic

database. M.B.G. generated covariance and X^TX matrices. M.M.G., N.B. and L.C.J. wrote the manuscript. All co-authors approved the final manuscript.

Data accessibility

Scripts and data package uploaded to Dryad: doi: 10.5061/dryad.tp96r.

Supporting information

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

Table S1 Private bands occurring across regions.

Table S2 Analysis of molecular variance statistical summary.

Fig. S1 Plot of mean likelihood estimated probability for K = 2-9.

Fig. S2 Delta *K* calculation shown for *K* = 2–9.

Fig. S3 Principal component analysis of environmental variables from source prairie populations.