The LINNEAN SOCIETY of London BOTANICAL Journal of the Linnean Society

Botanical Journal of the Linnean Society, 2016, **181**, 99–114. With 2 figures

Hybridization and divergence in multi-species oak (*Quercus*) communities

ALEXIS R. SULLIVAN[1,2], SANDRA A. OWUSU[2], JAIME A. WEBER[3], ANDREW L. HIPP[3,4] and OLIVER GAILING[2]*

[1]*Department of Ecology and Environmental Sciences, Umeå University, Umeå 901 87, Sweden*
[2]*School of Forest Resources and Environmental Science, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA*
[3]*The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532, USA*
[4]*Department of Botany, The Field Museum, 1400 S Lake Shore Drive, Chicago, IL 60605, USA*

Received 22 October 2015; revised 10 January 2016; accepted for publication 26 January 2016

Oaks (*Quercus*: Fagaceae) commonly interbreed yet retain their morphological, genetic and ecological distinctiveness. Post-zygotic isolation mechanisms, such as ecologically dependent selection on adaptive loci, may therefore limit introgression. To test this hypothesis, we quantified hybridization and genetic divergence across the contact zone of four red oaks (*Quercus* section *Lobatae*) in the Great Lakes region of North America using a suite of 259 amplified fragment length polymorphisms and 27 genic and genomic microsatellite markers. First, we identified hybrids using genetic structure analysis and confirmed the reliability of our assignments via simulations. Then, we identified candidate loci for species maintenance with three complementary tests for selection and obtained partial gene sequences linked to an outlier locus and three other loci. We detected evidence of recent hybridization among all species and considerable gene flow between *Q. ellipsoidalis* and *Q. velutina*. Overall, *c.* 20% of *Q. velutina* had recent ancestry from *Q. ellipsoidalis*, whereas nearly 30% of *Q. ellipsoidalis* had a *Q. velutina* ancestor. Most loci were negligibly to weakly differentiated among species, but two gene-linked microsatellites deviated significantly from neutral expectations in multiple, complementary outlier tests. Both outlier loci were located in the same 15-cM bin on an existing *Q. robur* linkage map, a region under divergent selection in other oak species. Adaptive loci in this highly differentiated genomic region may contribute to ecological divergence among species and limit introgression. © 2016 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2016, **181**, 99–114

ADDITIONAL KEYWORDS: genetic divergence – introgression – outlier loci – *Quercus* section *Lobatae* – red oaks – selective sweep.

INTRODUCTION

Hybridization is thought to be common among oaks (*Quercus*: Fagaceae) and species boundaries at morphological and genetic markers are notoriously ambiguous (Burger, 1975; Van Valen, 1976; Donoghue, 1985). This interspecific gene flow may increase intraspecific adaptive genetic variation by transferring traits and alleles among species (Arnold, 2004; Arnold & Martin, 2010). Conversely, hybridization can erode species coherence and lead to the homogenization of previously distinct gene pools (Seehausen, 2006; Vonlanthen *et al.*, 2012). However, recurrent hybridization among oaks has apparently not led to a loss of genetic or adaptive distinctiveness (Whittemore & Schaal, 1991; Dodd & Afzal-Rafii, 2004; Petit *et al.*, 2004; Curtu, Gailing & Finkeldey, 2007; Peñaloza-Ramírez *et al.*, 2010). Oaks thus present an opportunity to study how species coherence can be maintained with ongoing gene flow (Arnold & Bennett, 1993; Dodd & Afzal-Rafii, 2004), despite the fact that hybridization has been shown to counteract species divergence in other taxa (Seehausen, 2006; Vonlanthen *et al.*, 2012).

Similarly to interfertile European white oaks (*Quercus* section *Quercus*, e.g. Curtu *et al.*, 2007), genetic differentiation within and among species is weak in the red oaks (*Quercus* section *Lobatae*), a

*Corresponding author. E-mail: ogailing@mtu.edu

North American clade comprising *c*. 195 species (Aldrich *et al.*, 2003; Dodd & Afzal-Rafii, 2004; Peñaloza-Ramírez *et al.*, 2010; Moran, Willis & Clark, 2012). Many red oaks may have a relatively recent origin (Aldrich & Cavender-Bares, 2011), which can lead to overlapping allele frequency distributions among species (e.g. Sullivan *et al.*, 2013) and sharing of plastid haplotypes (Zhang, Hipp & Gailing, 2015) through some combination of incomplete lineage sorting and interspecific gene flow (Muir & Schlötterer, 2005; Lexer, Kremer & Petit, 2006). Among the red oaks of eastern North America, hybridization and introgression are reported with varying frequencies among different age classes. Parentage analysis of seedlings indicated hybridization rates > 20% (Moran *et al.*, 2012), whereas adult hybrids are apparently much more rare, on the order of 1–5% (Jensen *et al.*, 1993; Aldrich *et al.*, 2003; Hipp & Weber, 2008; Moran *et al.*, 2012; Lind & Gailing, 2013). Similarly, gene flow is apparently more restricted between some species than others: hybridization was limited between all age classes of neighbouring stands of northern pin oak (*Q. ellipsoidalis* E.J. Hill) and northern red oak (*Q. rubra* L.; Lind & Gailing, 2013; Lind-Riehl, Sullivan & Gailing, 2014; Collins, Sullivan & Gailing, 2015) but appears more common between *Q. ellipsoidalis* and black oak (*Q. velutina* Lam.; Hipp & Weber, 2008). Extensive gene flow may occur among *Q. rubra*, *Q. velutina* and scarlet oak (*Q. coccinea* Münch.; Jensen, 1977; Moran *et al.*, 2012), although other reports indicate more restricted gene flow between *Q. rubra* and other species (Hipp & Weber, 2008; Owusu *et al.*, 2015).

Barriers to gene flow in hybridizing species may be maintained through differential environmental pressures (Dodd & Afzal-Rafii, 2004; Hamilton & Aitken, 2013). Consistent with this hypothesis, red oak assemblages are frequently stratified along water and nutrient availability clines and differ in their adaptations to drought, light availability and temperature extremes as reflected in traits such as leaf conductance, root depth, xylem anatomy and leaf shape (Abrams, 1990, 1996). Although species occurrences overlap, especially near range extents (A. L. Hipp, pers. observ.), *Q. ellipsoidalis* generally occupies the most xeric sites, followed by *Q. coccinea*, *Q. velutina* with *Q. rubra*, the latter being mesophytic (Seidel, 1972; Day & Monk, 1974; Hinckley *et al.*, 1978; Bahari, Pallardy & Parker, 1985; Abrams, 1990). Given the divergence of these four oak species in drought-tolerance traits (reviewed by Abrams, 1990), intermediate phenotypes could be subjected to ecologically dependent selection (e.g. Rundle, 2002), which could limit effective gene flow. Pre-zygotic mechanisms such as differences in flowering time and pollen incompatibilities contribute to reproductive isolation among interfertile white oaks (*Quercus* section *Quercus*; Abadie *et al.*, 2012; Lepais *et al.*, 2013). Evidence is currently lacking for the red oaks, but vegetative bud burst (a trait which is correlated with flowering time; Chesnoiu *et al.*, 2009) differed significantly between *Q. rubra* and *Q. ellipsoidalis* seedlings in a common garden experiment (Gailing, 2013), although this pattern was not observed among adult trees in field conditions (Collins *et al.*, 2015). Both the strength and the direction of pre-/post-zygotic barriers to gene flow are probably environmentally modulated. Given recent and ongoing changes in global temperature and precipitation patterns, flowering times may become more synchronous and introgressed alleles from drought-tolerant to more mesic species may confer a selective advantage (e.g. Whitney, Randell & Rieseberg, 2010).

We examined interspecific gene flow and divergence in five population pairs of *Q. ellipsoidalis* and *Q. velutina* as well as in *Q. rubra* and *Q. coccinea* populations using 27 gene-linked [expressed sequence tag– simple sequence repeat (EST-SSR)] and genomic (gSSR) microsatellite markers and 259 amplified fragment length polymorphism (AFLP) markers. We focused on *Q. velutina* and *Q. ellipsoidalis* because they tend to have different microhabitat distributions but are suspected to hybridize widely (Hipp & Weber, 2008). We tested: (1) the power and accuracy of the AFLP and microsatellite panel to distinguish purebred *Q. ellipsoidalis* and *Q. velutina* from hybrids; (2) the frequency and distribution of first and later generation hybrids between *Q. velutina* and *Q. ellipsoidalis*; and (3) the evidence of divergent selection among all four species using three complementary outlier loci tests.

MATERIAL AND METHODS

Plant material, identification and DNA isolation

The contemporary distributions of *Q. ellipsoidalis* and *Q. coccinea* are largely parapatric with only a narrow contact zone in northern Indiana and southern Michigan (Hipp & Weber, 2008). In contrast, the ranges of *Q. ellipsoidalis* and *Q. coccinea* are both entirely circumscribed by *Q. rubra* and overlap broadly with *Q. velutina* (Fig. 1). *Quercus velutina* and *Q. ellipsoidalis* co-occur across a broad region in the northern and southern extents of their ranges, respectively, and in isolated populations scattered throughout the Lower Peninsula of Michigan (Fig. 1).

We utilized published and new plant material (Supporting Information, Table S1) to develop a

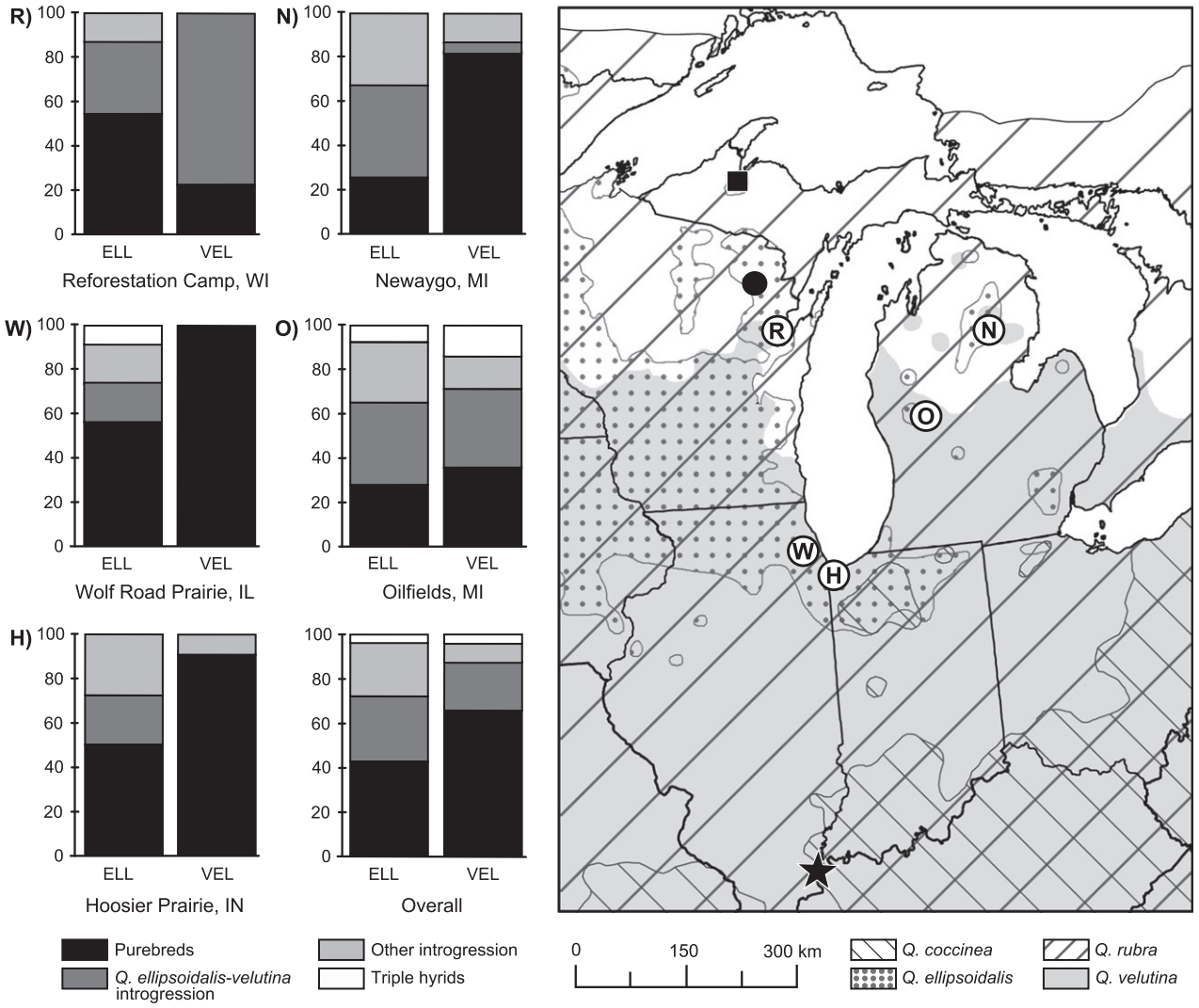

Figure 1. Number of purebreds ($q \geq 0.90$), introgressive forms ($0.90 < q > 0.10$) and triple hybrids ($q < 0.80$ and $q \geq 0.10$ in two other species clusters) in neighbouring *Quercus ellipsoidalis* and *Q. velutina* populations. Population pairs are denoted by black letters in white circles, which correspond to the bar plots showing species assignments and to the population abbreviations in Figure 2. *Quercus rubra* populations are represented by black circles, the *Q. coccinea* population by a star and the reference *Q. ellipsoidalis* population by a square. Approximate species ranges are from Little (1971).

dataset comprising five neighbouring population pairs of *Q. ellipsoidalis* and *Q. velutina* broadly spanning their contact zone. To allow testing for possible introgression from *Q. rubra* and *Q. coccinea* into *Q. velutina* and *Q. ellipsoidalis*, which could be a confounding variable if present but undetected, we also included a previously studied *Q. coccinea* population (Hipp & Weber, 2008) and two populations of *Q. rubra* (Lind & Gailing, 2013; Sullivan *et al.*, 2013; Lind-Riehl *et al.*, 2014) in the analysis. We also included a *Q. ellipsoidalis* population that has been previously shown to mostly comprise purebreds in all age classes (Lind & Gailing, 2013; Lind-Riehl *et al.*, 2014; Collins *et al.*, 2015) to improve admixture estimates between *Q. velutina* and *Q. ellipsoidalis.* All populations in sympatry with *Q. velutina* were used also in our search for genes potentially involved in species maintenance. In total, our analysis included 202 adult [diameter at breast height (DBH) > 13 cm] individuals from 14 populations, comprising 106 *Q. ellipsoidalis* from six populations, 70 *Q. velutina* from five populations, 66 *Q. rubra* from two populations and 20 *Q. coccinea* individuals from a single population in southern Illinois (Supporting Informa-

tion, Table S1, Fig. 1). Vouchers were deposited at the herbarium of The Morton Arboretum and accession numbers are given in Supporting Information (Table S1; accessible online, http://quercus.mortonarb.org/).

For both new and published material, initial species assignments were based on leaf and acorn traits outlined in Jensen (1997), tree habit and site characteristics (e.g. mesic vs. xeric; geographical location). Based on the results of earlier assignment analyses (Hipp & Weber, 2008; Hipp, 2010; Hipp, Weber & Srivastava, 2010), we gave particular focus to the morphological differences in the buds and acorns of *Q. velutina*, *Q. ellipsoidalis* and *Q. coccinea*. Although species morphological traits occur along a continuum and individual trees may show a mixture of characteristics of two or more species, the end buds tend to be densely canescent in *Q. velutina* but silky pubescent in *Q. coccinea* and *Q. ellipsoidalis* and the inner surface of the cupule is generally densely pubescent in *Q. velutina* but sparsely pubescent to glabrous in *Q. ellipsoidalis* and *Q. coccinea*. Morphological species assignments were made independently by two authors (A.L.H. and J.A.W.) and putative hybrids were included in the genetic assignment analysis.

Marker analysis

We selected eight gSSRs and 19 gene-linked EST-SSRs on the basis of linkage group position, putative function and annealing temperature (Steinkellner *et al.*, 1997; Aldrich *et al.*, 2002; Sullivan *et al.* 2013) PCRs, capillary electrophoresis and fragment scoring were done as previously described (Sullivan *et al.*, 2013).

Quercus rubra populations and the *Q. ellipsoidalis* Ford Forestry Center population (Supporting Information, Table S1) were only analysed with the panel of 27 microsatellite markers. The remaining samples were additionally characterized by 259 AFLP markers. Total genomic DNA was digested by *Eco*RI/*Mse*I and selectively amplified using fluorescently labelled *Eco*RI-ACG and *Mse*I-CTC primers following the procedure described by Hipp & Weber (2008). PCR products were purified using the CleanSeq dye-terminator removal system (Agencourt Bioscience) and visualized on an ABI3730 genetic analyser. AFLP chromatographs were analysed in GeneMapper v. 3.7 using the automatic binning system and manually edited to correct for inconsistencies in fragment-calling. Each sample was selectively amplified and electrophoretically resolved twice, allowing us to test for locus drop-out and to distinguish between background noise and true polymorphisms. Samples were scored blindly. Markers ambiguous in any individual and markers with < 100% reproducibility were excluded from the entire dataset.

Microsatellite markers are prone to null alleles and scoring errors. Therefore, we used the program MicroChecker v. 2.2.3 (van Oosterhout *et al.*, 2004) to assess the likelihood that a given marker in a given population was misscored due to the presence of stutter bands, allelic dropout or null alleles. Analyses were performed using both the adjusted genotypes supplied by MicroChecker and the original, unadjusted dataset. Neither hybridization estimates nor outlier loci results differed between the two. As null allele correction itself may introduce error and bias (Dąbrowski *et al.*, 2014, 2015), we present results based on the unadjusted dataset. Weir and Cockerham's estimator of Wright's fixation index (F_{ST}; Weir and Cockerham, 1984) and its analogue R_{ST} for microsatellites (Slatkin, 1995), the inbreeding coefficient (F_{IS}), expected heterozygosity (H_e), observed heterozygosity (H_o) and allele frequencies for microsatellites were estimated using Genepop v. 4.2 (Raymond & Rousset, 1995). Allelic richness was computed in Hp-Rare v. 1.0 and rarefied to the smallest sample size to allow for comparisons among unequal sample sizes (Kalinowski, 2004, 2005). For AFLPs, we calculated the percentage of polymorphic loci and estimated H_e using the Bayesian method with non-uniform priors on allele frequencies (Zhivotovsky, 1999) in AFLPSurv 1.0 (Vekemans *et al.*, 2002).

Identification of hybrids and introgressive forms

Structure v. 2.3.4 was used to infer population structure and assign individuals to populations (Pritchard, Stephens & Donnelly, 2000). The number of populations K was estimated by running ten replicate simulations for each potential value of K = 1–6 under the admixture model without any prior morphological species information. Parameters were estimated from 500 000 Markov chain Monte Carlo (MCMC) generations following a 500 000-generation burn-in. We selected the optimal K by comparing log-likelihood probabilities and by calculating the change in log-likelihoods over different values of K (ΔK, Evanno, Regnaut & Goudet, 2005). Both metrics were calculated in Structure Harvester (Earl, 2012). After determining the optimal K value, we calculated final assignments from 10^6 MCMC generations following 10^6 generations of burn-in. Barplot diagrams were prepared using distruct v. 1.1 (Rosenberg, 2004).

Structure was first run on the AFLP and microsatellite datasets separately. Genotypic cluster assignments were congruent between datasets (r^2 = 0.93, 0.88 and 0.85 for *Q. coccinea*, *Q. ellipsoidalis* and *Q. velutina*, respectively; no AFLPs were ampli-

fied in *Q. rubra*), so we combined the AFLPs and microsatellites in a third run. The power of STRUCTURE to identify hybrids is expected to be low when differentiation is weak (Vähä & Primmer, 2006) and our focal species pair, *Q. ellipsoidalis* and *Q. velutina*, are the most weakly differentiated of the four (F_{ST} = 0.033; see Results below). To assess our capacity to distinguish hybrids from purebred *Q. ellipsoidalis* and *Q. velutina*, we generated the expected multilocus genotypes of 20 hybrid individuals in ten separate $K = 4$ runs from the allele frequencies of individuals that (1) were morphologically typical and (2) had q values ≥ 0.90 in the combined AFLP and microsatellite STRUCTURE run. Hybrid genotypes at the 27 SSR markers were generated with HYBRIDLAB v. 1.0 (Nielsen, Bach & Kotlicki, 2006). AFLP genotypes were simulated as one binomial trial per genotype, where the probability for the presence of each band was set equal to the average band frequency between *Q. ellipsoidalis* and *Q. velutina*. Similarly, we generated both *Q. ellipsoidalis*- and *Q. velutina*-like F_2 AFLP and microsatellite genotypes by crossing simulated F_1 hybrids with the observed pure individuals. Finally, we simulated ten sets of 20 new purebred individuals using the observed microsatellite and AFLP genotype frequencies to test STRUCTURE's ability to correctly identify pure species in this dataset. Based on these results, we considered individuals with a q coefficient (i.e. the proportion of ancestry from each cluster) of $q \geq 0.90$ as purebreds, $0.90 > q > 0.10$ as introgressive forms (e.g. Vähä & Primmer, 2006) and to be triple hybrids (i.e. involving three species) if q values were < 0.80 with $q \geq 0.10$ in two other species clusters (e.g. Dodd & Afzal-Rafii, 2004; Lepais *et al.*, 2009).

Outlier detection

We employed the F_{ST}-based program LOSITAN (Antao *et al.*, 2008) to detect SSR markers with differentiation higher than expected under neutrality (Beaumont & Nichols, 1996). LOSITAN simulates a neutral confidence envelope describing the neutral expectation of Wright's F_{ST} vs. H_e under an island model of migration (Beaumont & Nichols, 1996; Antao *et al.*, 2008). Outlier loci falling outside the upper bounds of the confidence envelope exhibit patterns of genetic diversity consistent with divergent selection (Beaumont & Nichols, 1996; Beaumont, 2005). Runs were conducted using a 0.95 confidence envelope, a 0.10 false discovery rate, the stepwise mutation model and 10^6 iterations following initial estimation of neutral F_{ST}. Due to the stochastic nature of the simulated neutral envelope, we ran three replicate simulations for each pairwise comparison and report only loci identified as outliers in all three replicates.

Model-based approaches to identify selection are powerful (Narum & Hess, 2011) and widely employed but may yield false positives because of population structure (Lotterhos & Whitlock, 2014) and/or heterogeneity in locus mutation rates (Beaumont & Nichols, 1996; Beaumont, 2005). Therefore, we combined the model-based F_{ST}-outlier method with two statistical tests developed for use with microsatellite markers to reduce the type I error rate (e.g., Vasemägi, Nilsson & Primmer, 2005). In contrast to F_{ST}-based methods, the LnRH and LnRV statistics do not invoke a particular demographic model but instead identify microsatellite loci with significantly lower heterozygosity or variance in repeat size, respectively, in one species relative to another (Schlötterer, 2002; Schlötterer & Dieringer, 2005). Such differential reduction in genetic diversity may be indicative of a selective sweep (Schlötterer, 2002; Schlötterer & Dieringer, 2005). We consider markers identified as outliers by at least two of the three tests to be most promising loci for species maintenance but note this method will fail to detect an alternatively fixed locus as heterozygosity and size variance will be low in both species (e.g. Lind-Riehl *et al.*, 2014).

Outlier locus sequencing

We designed primer pairs for portions of the genes linked to the EST-SSRs identified as under directional selection by multiple outlier-detection tests (GOT040 and POR016; see Results below). To support the assumption that linkage disequilibrium in *Quercus* decays slowly enough for EST-SSRs to reflect coding sequence variation, we also designed primers for coding regions linked to: (1) the most strongly differentiated EST-SSR (FIR039); (2) a marker with high interspecific F_{ST}/R_{ST}, which was also found to be an outlier in a Mediterranean oak complex (VIT107, Goicoechea *et al.*, 2015); and (3) a putatively neutral EST-SSR with low interspecific F_{ST}/R_{ST} (FIR031; Table 1). For these five loci, we reassembled the corresponding EST sequences (Durand *et al.*, 2010) using CAP3 (Huang & Madan, 1999) and designed primers in PRIMER3 (Rozen & Skaletsky, 1999; Supporting Information, Table S2). Locus names were assigned on the basis of putative homology (Supporting Information, Table S2).

Amplicons were generated for two to seven purebreds ($q \geq 0.90$) per species (see Table 3) in a 12.5-μL PCR mix comprising 3.5 mM $MgCl_2$, 200 μM dNTPs, 0.25 U Invitrogen Platinum *Taq* polymerase, 1× Invitrogen PCR buffer (Life Technologies), 6.25 pmol primer and *c.* 3.6 ng DNA. Cycling

Table 1. Observed heterozygosity (H_o), gene diversity (H_e), rarefied number of alleles (N_a, $N = 18$ individuals), fixation index (F_{ST}) and inbreeding coefficient (F_{IS}) for 27 markers characterized in purebred ($q \geq 0.90$) *Quercus rubra*, *Q. ellipsoidalis*, *Q. coccinea* and *Q. velutina*

	Q. rubra					*Q. ellipsoidalis*				*Q. coccinea*				*Q. velutina*			
Locus	F_{ST}/R_{ST}	N_a	H_o	H_e	F_{IS}	N_a	H_o	H_e	F_{IS}	N_a	H_o	H_e	F_{IS}	N_a	H_o	H_e	F_{IS}
Parameter estimates by locus and species																	
FIR024	**0.15/0.07**	5	0.50	0.53	0.08	5	0.65	0.62	−0.04	5	0.44	0.45	0.05	4	0.40	0.52	0.24
FIR030	**0.09/0.04**	7	0.45	0.43	−0.03	4	0.51	0.54	0.07	3	0.39	0.64	0.42	4	0.48	0.55	0.13
FIR031	**0.06/0.01**	16	0.90	0.89	0.01	10	0.77	0.85	0.11	10	0.67	0.82	0.21	14	0.80	0.89	0.11
FIR039	**0.22/0.53**	5	0.55	0.72	0.26	4	0.43	0.42	−0.03	3	0.61	0.53	−0.13	3	0.24	0.22	−0.08
FIR053	**0.14/0.26**	5	0.65	0.70	0.10	3	0.25	0.31	0.19	3	0.50	0.57	0.15	6	0.73	0.75	0.02
FIR104	**0.14/0.15**	3	0.35	0.34	−0.02	4	0.67	0.59	−0.13	4	0.56	0.64	0.16	5	0.74	0.71	−0.03
GOT004	**0.10/0.08**	6	0.60	0.80	0.28	10	0.80	0.82	0.05	7	0.61	0.69	0.15	9	0.56	0.78	**0.30**
GOT009	0.03/0.00	10	0.75	0.85	0.14	7	0.80	0.71	−0.11	7	0.83	0.69	−0.19	8	0.82	0.79	−0.03
GOT021	0.04/0.04	2	0.10	0.10	−0.03	2	0.14	0.13	−0.04	2	0.06	0.05	−0.03	3	0.34	0.31	−0.08
GOT037	**0.04/0.29**	13	0.63	0.90	**0.32**	9	0.75	0.85	0.13	12	0.72	0.88	0.20	8	0.66	0.81	0.20
GOT040	**0.19/0.32**	8	0.85	0.75	−0.10	1	0.02	0.02	−0.01	2	0.06	0.05	−0.03	4	0.22	0.25	0.14
PIE039	**0.07/0.26**	8	0.75	0.71	−0.04	4	0.57	0.66	0.15	7	0.56	0.56	0.04	5	0.54	0.59	0.10
PIE040	**0.09/0.14**	4	0.50	0.53	0.08	4	0.64	0.62	−0.02	3	0.47	0.48	0.04	4	0.58	0.62	0.07
PIE099	**0.03/0.26**	10	0.60	0.87	0.33	11	0.86	0.87	0.02	11	0.83	0.83	0.03	10	0.84	0.84	0.01
PIE125	**0.13/0.10**	5	0.65	0.67	0.06	4	0.44	0.59	0.26	5	0.56	0.51	−0.06	5	0.54	0.58	0.08
POR003	**0.04/0.08**	9	0.85	0.83	0.00	11	0.86	0.87	0.02	12	1.00	0.88	−0.11	10	0.90	0.80	−0.11
POR016	**0.08/0.06**	1	0.00	0.00	0.00	2	0.05	0.04	−0.01	3	0.39	0.32	−0.18	3	0.28	0.27	−0.04
QpZAG15	**0.05/0.07**	13	0.90	0.86	−0.02	12	0.89	0.85	−0.03	11	0.78	0.85	0.11	15	0.96	0.92	−0.03
VIT057	0.03/0.07	3	0.25	0.30	0.19	3	0.35	0.42	0.19	3	0.44	0.44	0.01	3	0.52	0.53	0.03
VIT107	**0.13/0.78**	12	0.80	0.84	0.07	5	0.73	0.71	−0.02	4	0.44	0.54	0.21	7	0.46	0.78	**0.42**
1P10	0.04/**0.13**	10	0.90	0.81	−0.08	8	0.82	0.81	0.00	9	0.67	0.74	0.13	8	0.72	0.78	0.09
2P24	**0.13/0.22**	11	0.95	0.87	−0.07	8	0.73	0.71	−0.01	5	0.56	0.45	−0.22	8	0.74	0.77	0.05
3A05	**0.27/0.44**	5	0.45	0.62	0.29	5	0.45	0.64	0.30	4	0.33	0.33	0.03	6	0.78	0.72	−0.07
3D15	**0.08/0.16**	8	0.85	0.74	−0.13	7	0.66	0.75	0.13	9	0.94	0.83	−0.12	10	0.84	0.85	0.02
quru-GA-0C11	**0.02/0.18**	10	0.80	0.88	0.12	11	0.89	0.89	0.01	13	1.00	0.90	−0.09	13	0.90	0.92	0.03
quru-GA-0E09	**0.02/0.17**	16	0.55	0.92	**0.43**	15	0.89	0.91	0.04	12	0.72	0.89	0.22	17	0.86	0.94	0.09
quru-GA-1F07	**0.03/0.26**	19	0.95	0.93	0.00	12	0.93	0.89	−0.03	16	0.89	0.91	0.05	16	0.68	0.93	**0.28**
Parameter estimates by species																	
	0.09/0.18	8	0.63	0.68	**0.06**	7	0.61	0.63	0.03	7	0.59	0.61	0.01	8	0.64	0.68	**0.06**

Parameter estimates were bulked by species. Significant values are denoted in bold type (sequential Bonferroni corrected $P < 0.05$).

conditions consisted of 94 °C for 1 min, followed by 30 cycles of 30 s at 94 °C, 30 s at 59 °C, 30 s at 68 °C and a final extension at 68 °C for 10 min. PCR products were purified using the QIAQuick purification kit (Qiagen) and directly sequenced at the Nevada Genomics Center using the Big Dye Terminator cycle sequencing kit v. 3.1 (Applied Biosystems). Forward and reverse PCR primers were used for sequencing.

Sequences were analysed for the presence of single nucleotide polymorphisms (SNPs) and insertion-deletion (INDEL) polymorphisms in Codon Code Aligner v. 4.0 (Codon Code). All electropherograms were visually inspected to ensure accurate identification of heterozygous SNPs. Sequences were aligned using MUSCLE (Edgar, 2004) as implemented in MEGA6 (Tamura *et al.*, 2013). Exons were identified using the GENESEQER@PLANTGDB server, which analyses spliced EST and cDNA alignments to identify putative coding regions of a genomic query (Schlueter, Dong & Brendel, 2003). We used the conceptual translations to check for non-synonymous, fixed sites between species, but our small sample size did not justify other tests of selection. We calculated genetic distance (N_{ST}) and identified variable sites in DNASP v. 5 (Librado & Rozas, 2009). As we included a single population of *Q. coccinea* but multiple populations for the other species, we do not report sequence (π) or haplotype (h) diversity. Rather, our goal was to search for non-synonymous differences between species and provide support for the assumption that patterns of diversity at EST-SSRs reflect diversity at linked genes.

RESULTS

Genetic diversity and marker characterization

Genetic variation in the four species was high and F_{ST} among them relatively low (Table 1). Differentiation as measured by R_{ST} indicated congruent but overall higher levels of differentiation, particularly at loci with high F_{ST} and at the eight genomic microsatellites (Table 1). Higher R_{ST} than F_{ST} estimates across most microsatellite loci suggest a low incidence of homoplasy and a generally stepwise mode of microsatellite evolution (Hardy *et al.*, 2003). Gene diversity (H_e) and mean number of alleles per locus (N_a) at the common set of 27 SSR markers was slightly higher in *Q. rubra* and *Q. velutina* (both species $N_a = 8$, $H_e = 0.68$) than in *Q. ellipsoidalis* ($N_a = 7$, $H_e = 0.63$) and *Q. coccinea* ($N_a = 7$, $H_e = 0.61$). Although we sampled *Q. ellipsoidalis* and *Q. velutina* populations across the Great Lakes region, we only included two populations of *Q. rubra* and one of *Q. coccinea* and may have underestimated their diversity as a result.

Genotypic frequencies significantly departed from Hardy–Weinberg equilibrium after Bonferroni corrections at two markers in *Q. rubra* and three in *Q. velutina* (Table 1). These markers potentially have null alleles according to MICROCHECKER as indicated by the excess of homozygotes in all size classes. Overall, *Q. velutina* and *Q. rubra* populations showed a small but significant excess of homozygotes (both $F_{IS} = 0.06$), apparently due to null alleles at a few loci. No significant departure from Hardy–Weinberg equilibrium was observed in *Q. ellipsoidalis* or *Q. coccinea*, although MICROCHECKER found evidence for null alleles at two markers in each of these species.

STRUCTURE power and error tests using simulated AFLP and SSR genotypes

Using a $q \geq 0.90$ threshold for pure species, 99.0% of the newly simulated *Q. velutina* genotypes were correctly identified. Our type I error rate for *Q. ellipsoidalis* was slightly higher: 96.0% of simulated pure genotypes were identified as pure while 4.0% were incorrectly identified as introgressive forms. No simulated pure genotypes were classified as F_1 hybrids or triple hybrids or determined to be misclassified ($q \geq 0.90$ in a genetic cluster different from the morphological species cluster). STRUCTURE detected some degree of interspecific gene flow ($0.90 > q > 0.10$ in two clusters) in 96.5% of the simulated F_1 hybrid genotypes. Over half (55.5%) of the simulated F_1 hybrids were accurately identified whereas 41.0% were misclassified as introgressive forms (Supporting Information, Table S3). Among simulated F_2 genotypes, STRUCTURE could more effectively distinguish *Q. ellipsoidalis*-like introgressive forms from pure *Q. ellipsoidalis* than *Q. velutina*-like forms from pure *Q. velutina*. Thirty-one per cent of the simulated *Q. ellipsoidalis* backcrosses were correctly identified, whereas only 12% of the *Q. velutina*-like backcrosses were correctly identified. While this suite of AFLP and microsatellite markers can accurately identify purebred *Q. ellipsoidalis* and *Q. velutina* and detect interspecific gene flow, our power to distinguish F_1 hybrids from later introgressive forms is limited. In light of these results, we group F_1 hybrids and backcrosses together in a single category ($0.90 > q > 0.10$).

Differentiation and admixture among the four species

STRUCTURE analysis of 27 microsatellites and 259 AFLPs confirmed the presence of four distinct genetic clusters (Supporting Information, Fig. S1), which correspond to *Q. coccinea*, *Q. ellipsoidalis*,

Q. rubra and *Q. velutina* (Fig. 2). About 9% of the total genetic variation was distributed among species ($P < 0.001$) after excluding probable hybrids and introgressive forms. Microsatellite F_{ST} values ranged from 0.02 to 0.27 and followed an L-shaped curve with most markers negligibly to weakly differentiated (Table 1). After excluding two outlier microsatellites (see 'outlier analysis' below), pairwise differentiation ranged from 3% between *Q. velutina* and *Q. ellipsoidalis* to 8% between *Q. rubra* and *Q. coccinea* (Table 2).

Despite their close phylogenetic relationship (A. L. Hipp unpubl. data; Owusu *et al.*, 2015) but consistent with their parapatric ranges, we did not find evidence of extensive hybridization between *Q. coccinea* and *Q. ellipsoidalis*. Ten of the 106 *Q. ellipsoidalis* individuals, six of which are from their narrow zone of sympatry, showed evidence of recent introgression from *Q. coccinea* (Fig. 2, populations H3 and W7; Supporting Information, Table S4). Three *Q. velutina* individuals showed evidence of a recent *Q. coccinea* ancestor, but we note that two (both located in Michigan) had q values close to the purebred threshold ($q = 0.896$ and 0.869).

Although the range of *Q. coccinea* is entirely circumscribed by *Q. rubra*, we found no evidence of recent introgression in our *Q. coccinea* population (Supporting Information, Table S4, Fig. 2). However, a single *Q. rubra* individual in the Wisconsin population, located far outside the zone of sympatry (Fig. 1), apparently had a recent *Q. coccinea* ancestor (Fig. 2, population LW13). Gene flow from *Q. rubra* into *Q. ellipsoidalis* and *Q. velutina* was also more limited: ten of the 106 *Q. ellipsoidalis* and four of the 70 *Q. velutina* individuals had a recent *Q. rubra* ancestor (Supporting Information, Table S4, Fig. 2). Unexpectedly, we found evidence for considerable gene flow from *Q. velutina* into the *Q. rubra* population from Lakewood, Wisconsin (Fig. 2, population LW13). Six of the 26 individuals, or 23% of the population, had q coefficients consistent with hybridization with *Q. velutina*.

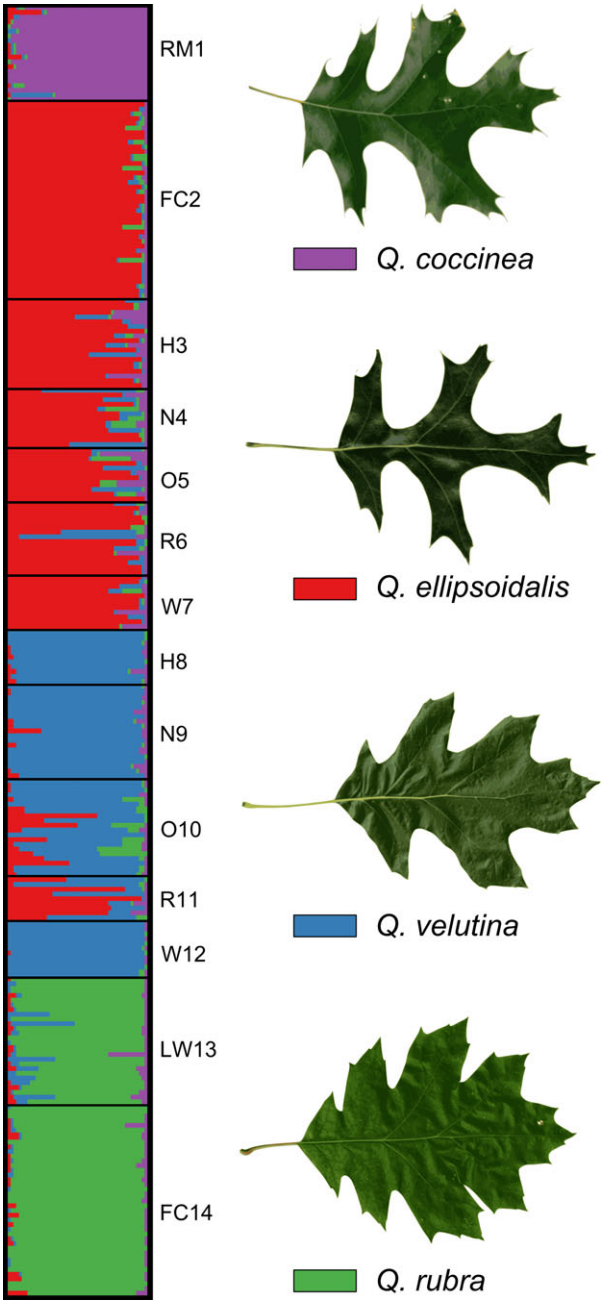

Figure 2. Genetic assignment results for five population pairs of *Quercus ellipsoidalis* and *Q. velutina* and populations of *Q. rubra*, *Q. coccinea* and *Q. ellipsoidalis*. Numbers correspond to the populations in Supporting Information (Table S1).

Patterns of hybridization in *Q. ellipsoidalis* and *Q. velutina* population pairs

Gene flow from *Q. velutina* into *Q. ellipsoidalis* was extensive and asymmetric at all five localities across their zone of sympatry (Fig. 1, Supporting Information, Table S5), with 19–42% of *Q. ellipsoidalis* having a recent *Q. velutina* ancestor. Gene flow from *Q. velutina* was highest in the two populations in the Lower Peninsula of Michigan (Newaygo 42% and Oilfields 36%), located outside the core distribution of either species. More purebred *Q. ellipsoidalis* individuals were detected on the western side of Lake Michigan, and introgression from *Q. velutina* and also from *Q. rubra* and *Q. coccinea*, declined from south to north. Overall, 30% of *Q. ellipsoidalis* individuals showed evidence of introgression from *Q. velutina* in their zone of sympatry and another 22% had a recent *Q. rubra* or *Q. coccinea* ancestor

Table 2. Multilocus pairwise differentiation (F_{ST}) between four hybridizing species of *Quercus* section *Lobatae*

Species	*Q. coccinea*	*Q. ellipsoidalis*	*Q. rubra*	*Q. velutina*
Q. coccinea				
Q. ellipsoidalis	0.057			
Q. rubra	0.088	0.042		
Q. velutina	0.067	0.033	0.051	

All pairs were significantly differentiated ($P < 0.001$).

(Supporting Information, Tables S1, S2). The *Q. ellipsoidalis* population located outside the zone of sympatry with *Q. velutina* and *Q. coccinea*, but circumscribed by *Q. rubra* stands, comprised 90% purebreds and all four putative introgressive forms had a *Q. rubra* ancestor (Supporting Information, Table S4, Fig. 2, population FC2).

Introgression from *Q. ellipsoidalis* into the gene pool of *Q. velutina* was overall less extensive but ranged from 0 to 78% among the five populations. Overall, 65.7% ($N = 46$) of the 70 *Q. velutina* individuals were purebreds and 15 individuals (21.4%) had a recent *Q. ellipsoidalis* ancestor. A smaller proportion of individuals also had ancestry from *Q. rubra* ($N = 3$; 4.3%) and *Q. coccinea* ($N = 3$; 4.3%) and three were triple hybrids (4.3%). Hybridization with *Q. ellipsoidalis* was largely confined to two sites near the northern range limit of *Q. velutina*. Together, the Reforestation Camp site in Wisconsin and the Oilfields population in Michigan harboured 93% of the *Q. velutina*-like introgressive forms (Fig. 1, Supporting Information, Table S5). At the Oilfields site, three additional individuals appeared to be triple hybrids from a cross between an F_1 *Q. velutina* × *Q. ellipsoidalis* hybrid and a purebred *Q. rubra*. The two *Q. velutina* populations located well within their core range showed no evidence of gene flow from *Q. ellipsoidalis* (Hoosier Prairie and Wolf Road Prairie, Fig. 1, Supporting Information, Table S5).

Microsatellite outlier analysis

Nine different microsatellite loci demonstrated patterns of differentiation concordant with divergent selection across the four pairwise species comparisons according to at least one selection-detection method (Supporting Information, Table S6). Although the model-based F_{ST}-outlier test is underpinned by more biological assumptions (Lotterhos & Whitlock, 2014), this test and the two model-free statistics, LnRH and LnRV, identified a similar number of outlier loci (Supporting Information, Table S6). Given the potentially high type I error of outlier tests, we consider markers identified by at least two methods in a given pairwise comparison as the most likely to be located in or near a region involved in adaptive genetic divergence. Two candidate EST-SSR markers were identified using this criterion, GOT040 and POR016, which are both located on the same 15-cM linkage group bin in *Q. robur* (Durand *et al.*, 2010). Both GOT040 and POR016 were outlier loci between *Q. ellipsoidalis* and *Q. velutina* and the outlier status of GOT040 was supported by all three tests (Supporting Information, Table S6).

No diagnostic alleles were found at GOT040 or POR016: all four species shared the most common allele at both markers (Supporting Information, Fig. S2). Instead, these two markers exhibited patterns of diversity consistent with a selective sweep. Expected heterozygosity and the number of alleles at GOT040 were markedly reduced in *Q. coccinea* and *Q. ellipsoidalis* relative to *Q. rubra* and *Q. velutina* (Table 1, Supplementary Information, Fig. S2). Variation at POR016 was restricted in *Q. rubra* and *Q. ellipsoidalis*, although gene diversity and the number of alleles at this marker were also relatively low in *Q. coccinea* and *Q. velutina* (Supporting Information, Fig. S2, Table 1). Notably, all five *Q. ellipsoidalis* populations across the Great Lakes were nearly monomorphic at both markers, whereas the neighbouring *Q. velutina* populations had comparably higher variation (Supporting Information, Fig. S2).

Sequence analysis

We successfully sequenced gene fragments linked to four of the five target microsatellites: the outlier locus GOT040, two markers with high interspecific differentiation that were not consistently identified as outliers (VIT107 and FIR039) and the weakly differentiated EST-SSR FIR031. Attempts to sequence the gene linked to POR016, the second outlier locus, failed. Although a single band was produced on a 2.5% agarose gel, direct sequencing revealed that multiple loci of the same size were amplified even after PCR and primer optimization. Sequences were uploaded to GenBank as follows (see below for locus names): Qu*APR* (KT899980–KT899990); Qu*RPS16* (KT907445–KT907456); Qu*HD* (KT946628–KT946643); and Qu*LHCA2* (KT907457–KT907472).

We sequenced 80% of the coding sequence of the gene associated with the outlier microsatellite GOT040, which encodes a 40S ribosomal 16S-like protein (Qu*RSP16*). No non-synonymous polymorphisms were observed, but 20% of the observed variation was distributed among species ($N_{ST} = 0.20$).

Pairwise haplotype differentiation was minimal between *Q. coccinea* and *Q. ellipsoidalis*, which were also negligibly differentiated at the GOT040 microsatellite and reached up to 45.7% differentiation between *Q. rubra* and *Q. coccinea* (Table 3), for which GOT040 was identified as outlier.

The marker with highest F_{ST} among the four oak species (FIR039) was located in the 5′ untranslated region (UTR) of a gene encoding a histone deacetylase-like protein (Qu*HD*). Sequence differentiation (N_{ST} = 0.50) at this locus was higher than at Qu*RSP16*, although these values are not directly comparable as the majority of the amplified Qu*HD* fragment was located in an intron (Table 3). However, 90% of the amplified sequence linked to VIT107 was located in the coding sequence of a light-harvesting compound-like (Qu*LHCA2*) protein and this locus had the second highest N_{ST} values (Table 3). Despite this high differentiation, we also did not observe any non-synonymous mutations in Qu*HD* or Qu*LHCA2*. No differentiation was observed between *Q. velutina*, *Q. ellipsoidalis* and *Q. coccinea* in the sequenced portion of the adenylylsulphate reductase-like gene (Qu*APR*; 70% coding) linked to FIR031, a weakly differentiated microsatellite (Table 1), although *Q. rubra* was moderately differentiated from the other three species.

Consistent with the out-crossing breeding system of *Quercus*, the relationship between pairwise microsatellite F_{ST} and N_{ST} across all four loci was moderate (r = 0.41, Spearman's rho, $P < 0.05$). For example, although *Q. velutina* and *Q. coccinea* were weakly differentiated at the GOT040 microsatellite, estimated N_{ST} was the highest for this pairwise comparison (F_{ST} = 0.03, N_{ST} = 0.47). Conversely, although F_{ST} was high (F_{ST} = 0.45) between *Q. velutina* and *Q. rubra* at FIR039, the corresponding sequence was only moderately differentiated (N_{ST} = 0.15).

DISCUSSION

Our genetic structure analysis indicated extensive, asymmetric gene flow into *Q. ellipsoidalis*, a Great Lakes endemic, from the widely distributed *Q. velutina* throughout their zone of sympatry. Considered together, 25% of the 136 *Q. velutina* and *Q. ellipsoidalis* individuals were hybrids or later introgressive forms, a rate higher than commonly found in white oaks (*Quercus* section *Quercus*; e.g. Whittemore & Schaal, 1991; Curtu *et al.*, 2007; Lagache *et al.*, 2013; but see Lepais *et al.*, 2009) but similar to values in other red oak complexes (Dodd & Afzal-Rafii, 2004; Peñaloza-Ramírez *et al.*, 2010; Moran *et al.*, 2012; Owusu *et al.*, 2015). Despite the relatively high number of markers employed here, we were unable to distinguish F_1 hybrids from later introgressive forms. Nevertheless, we also detected hybrid forms involving more than one parental species, which necessarily implies a lasting contribution

Table 3. Variation in the coding sequence of a gene linked to a microsatellite with an allele frequency spectrum consistent with a selective sweep (GOT040) and at sequences linked to two other highly differentiated microsatellites (FIR039 and VIT107) in four *Quercus* spp. The sequence linked to FIR031 was not identified as under divergent selection and was weakly differentiated in all pairwise comparisons.

Locus	Homology	L.G.	bp	N_{ST}	% Coding	Species	*N*	*s*
GOT040	40S ribosomal protein s16 (Qr*RPS16*-like)	6	362	0.20	100	*Q. coccinea*	3	0
						Q. ellipsoidalis	2	2
						Q. rubra	4	3
						Q. velutina	3	2
FIR039	Histone deacetylase-like (Qr*HD*-like)	2	796	0.50	33	*Q. coccinea*	3	0
						Q. ellipsoidalis	5	5
						Q. rubra	5	7
						Q. velutina	7	5
VIT107	Light-harvesting compound 2-like (Qr*LCHA2*-like)	3	384	0.32	90	*Q. coccinea*	5	1
						Q. ellipsoidalis	3	4
						Q. rubra	5	2
						Q. velutina	3	1
FIR031	Adenosine 5′-phosphosulfate reductase (Qr*APR*-like)	11	300	0.06	70	*Q. coccinea*	2	2
						Q. ellipsoidalis	2	2
						Q. rubra	4	3
						Q. velutina	3	2

L.G., linkage group (Durand *et al.*, 2010); bp, number of base pairs sequenced; N_{ST}, differentiation among species haplotypes; *N*, number genotypes per species; *s*, number of segregating sites.

to the gene pool (Dodd & Afzal-Rafii, 2004). In addition, our STRUCTURE power tests suggest that earlier studies in *Quercus* or other species in which interspecific differentiation is characteristically low may have underestimated the prevalence of recent hybridization (e.g. Lind & Gailing, 2013).

Barriers to hybridization in *Q. ellipsoidalis* were particularly weak, with fewer than half the analysed individuals being purebreds. Although *Q. velutina* made the single largest contribution to the *Q. ellipsoidalis* gene pool, recent ancestry from either *Q. rubra* or *Q. coccinea*, the two other red oak species in the region, was detected in almost one-quarter of the individuals. In the narrow band of Indiana and Illinois where all four species occur, five *Q. ellipsoidalis* individuals, or 28% of the population, showed evidence of a recent *Q. coccinea* ancestor. However, we also detected *Q. coccinea* ancestry far outside this zone of sympatry: for example, at the Newaygo site located *c.* 200 km north-east of the documented range of *Q. coccinea*, three individuals showed evidence of a recent *Q. coccinea* ancestor. Similarly, two individuals in the *Q. coccinea* population in southern Illinois were putative hybrids with *Q. ellipsoidalis*, which is restricted to the extreme northern portion of the state. These results may be false positives or reflect shared ancestral variation, but the species identity of the red oaks in this region has been of considerable taxonomic debate (Hipp & Weber, 2008). Species boundaries delimited on the bases of morphology may not be an accurate reflection of genetic affinities (e.g. Kremer *et al.*, 2002; Gailing, Lind & Lilleskov, 2012), especially near range extents, and *Q. coccinea* may occur further north than presently recognized.

The proportion of hybrids varied across the five *Q. ellipsoidalis* – *Q. velutina* population pairs. Spatial heterogeneity in hybridization rates could arise through pre-zygotic mechanisms such as differences in species densities (Lepais *et al.*, 2009; Lagache *et al.*, 2013) and variability in pollen production and flowering time (Abadie *et al.*, 2012). Consistent with frequency-dependent hybridization dynamics (Lepais *et al.*, 2009), introgressive *Q. ellipsoidalis*-like individuals outnumbered purebreds by nearly three-fold at the two sites in the Lower Peninsula of Michigan, where *Q. ellipsoidalis* occupies a patchy distribution and may be rare in the landscape. Similarly, introgression into *Q. velutina* was largely confined to two sites located near the northern extent of the species range. However, extensive introgression from *Q. velutina* into *Q. ellipsoidalis* was also found at a site in western Wisconsin where *Q. ellipsoidalis* is locally abundant (Owusu *et al.*, 2015). Similarly, introgression of *Q. velutina* alleles into the Wisconsin *Q. rubra* population is unlikely to be due to a higher availability of heterospecific pollen because *Q. velutina* is only a rare occurrence in this region. A more detailed understanding of local densities, pollen dispersal and intrinsic barriers to gene flow is needed to understand the importance of pre-zygotic factors in shaping red oak hybrid zones (e.g. Lepais *et al.*, 2009, 2013; Abadie *et al.*, 2012; Lagache *et al.*, 2013), especially between *Q. velutina* and *Q. ellipsoidalis*.

Post-zygotic selection has probably also contributed to the spatial pattern of hybridization reported here through differences in habitat availability (Moore, 1977; Arnold & Bennett, 1993) and environmentally dependent selection on introgressed alleles (Arnold, 2004; Dodd & Afzal-Rafii, 2004). Ecological niche differentiation coupled with decreasing abundance of hybrids from acorns to adults (Moran *et al.*, 2012) also support selection as a probable factor modulating introgression in oaks. *Quercus ellipsoidalis* is highly drought tolerant (Abrams, 1988; 1990) and genetic surveys of oaks near Lake Superior have found that *Q. ellipsoidalis* is largely restricted to extremely dry, sandy pine barrens (Lind & Gailing, 2013; Lind-Riehl *et al.*, 2014). In the western Great Lakes, *Q. velutina* is more common on savannas and is less water efficient than *Q. ellipsoidalis* (Abrams, 1990, 1996). *Quercus rubra* is the least drought tolerant of the four species (Seidel, 1972; Hinckley *et al.*, 1978; Bahari *et al.*, 1985) but also the fastest growing (Ashton & Larson 1996; Gailing, 2013). Experimental comparison of growth, water use efficiency and stress responses of *Q. velutina*, *Q. ellipsoidalis*, *Q. rubra* and their hybrids could quantify relative fitness and identify possible transfer of adaptations (e.g. Arnold, 2004). As many other species depend on oaks for food and shelter, changes in species/hybrid composition could affect ecosystem function and stability (e.g. Whitham *et al.*, 1999).

Hybridization is expected to result in largely homogeneous genomes, except for small, highly differentiated regions harbouring ecologically significant variation (e.g. Scotti-Saintagne *et al.*, 2004a; Nadeau *et al.*, 2012; Goicoechea *et al.*, 2015). Using multiple selection-detection methods, we identified two microsatellite markers located within 15 cM of each other on the *Q. robur* linkage map (Durant *et al.*, 2010) as candidates for adaptive divergence and species maintenance: POR016, a trinucleotide repeat in the 5′ UTR of a putative serine/threonine-protein phosphatase and GOT040, located in the 3′ UTR of a gene encoding a 40S ribosomal protein s16. Neither of these genes have been well characterized, but both microsatellites were unusually differentiated between *Q. velutina* and *Q. ellipsoidalis* and at least one of the two was an outlier locus in all other pairwise species comparisons. In addition, GOT040

and POR016 were both outlier loci in another *Q. ellipsoidalis* and *Q. rubra* population pair located in Michigan (Lind-Riehl *et al.*, 2014).

We did not find a clear fixation of alternative alleles in either the microsatellites or the corresponding nucleotide sequences, which would more clearly implicate divergent selection (e.g. Lind-Riehl *et al.*, 2014). Instead, these loci appear to be involved in a selective sweep, where diversity at GOT040 was strongly reduced in *Q. coccinea* and *Q. ellipsoidalis* and POR016 in *Q. ellipsoidalis* and *Q. rubra*. Microsatellites can be involved in gene regulation and may be the direct target of strong selection (Li *et al.*, 2004), but it is perhaps more likely that Qu*RPS16*, GOT040 and POR016 are in linkage disequilibrium with another, ecologically important locus. The chromosomal region harbouring GOT040 and POR016 has been found to contain outlier loci among hybridizing white oaks species (Scotti-Saintagne *et al.*, 2004a; A. L. Curtu, pers. comm.), a gene induced by osmotic stress in *Q. suber* L. (Porth *et al.*, 2005) and quantitative trait loci for several leaf morphological characteristics (Saintagne *et al.* 2004; Gailing *et al.*, 2013) and height growth in *Q. robur* and *Q. petraea* (Scotti-Saintagne *et al.*, 2004b). The recent availability of the *Quercus* genome assembly (Plomion *et al.*, 2016; https://w3.pierroton.inra.fr/QuercusPortal/index.php) will aid characterization of this and other divergent chromosomal regions and enable detailed analyses of the causes and consequences of introgression.

CONCLUSIONS

Hybridization has the capacity to generate new species (Rieseberg *et al.*, 2003) but may erode the boundaries of others (Vonlanthen *et al.*, 2012). Despite high levels of hybridization and introgression, many oaks still comprise distinct, stable species. Consequently, oaks provide an emerging model system for studying the consequences of hybridization and the mechanisms maintaining species continuity in the face of gene flow. We report some of the most extensive hybridization discovered so far in *Quercus* between the Great Lakes endemic *Q. ellipsoidalis* and the widely distributed *Q. velutina* throughout their zone of sympatry. In addition, we identify a chromosomal region marked by high differentiation at two microsatellites and a linked nucleotide sequence, which may underlie ecologically significant traits involved in species maintenance. Although divergent selection on adaptive loci may help to prevent the complete homogenization of red oak gene pools, the high level of effective gene flow detected here suggests that introgression may also allow these ecologically divergent species to transfer adaptations.

ACKNOWLEDGEMENTS

We are grateful to L. Parker and D. L. Hinson from the US Forest Service for their help in locating the *Q. rubra* stand in Wisconsin. K. Feldheim and staff of The Pritzker Laboratory assisted in generating the AFLP data. Helpful comments from Erika Hersch-Green and anonymous reviewers improved the clarity and accuracy of the manuscript. This research was funded by grants from the USDA McIntire Stennis Fund (1309051) and the NSF Plant Genome Research Program (NSF 1025974) to O.G., from the Hanes Trust to A.L.H. and O.G., a Franklin Research Grant from the American Philosophical Society to A.L.H., and a Michigan Technological University Ecosystem Science Center Graduate Student Research Grant to A.R.S. A.R.S. was supported by the US Department of Education and the European Commission's Directorate General for Education and Culture States' ATLANTIS Program, the Northern Institute of Applied Climate Science (USDA Forest Service) and the USDA McIntire Stennis Fund. Additional support was provided by Michigan Tech's Biotech Research Center.

REFERENCES

Abadie P, Roussel G, Dencausse B, Bonnet C, Bertocchi E, Louvet KM, Kremer A, Garnier-Géré B. 2012. Strength, diversity and plasticity of postmating reproductive barriers between two hybridizing oak species (*Quercus robur* L. and *Quercus petraea* (Matt) Liebl.). *Journal of Evolutionary Biology* **25:** 157–173.

Abrams MD. 1988. Comparative water relationships of three successional hardwood species in central Wisconsin. *Tree Physiology* **4:** 263–273.

Abrams MD. 1990. Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiology* **7:** 227–238.

Abrams MD. 1996. Distribution, historical development and ecophysiological attributes of oak species in the eastern United States. *Annals of Forest Science* **53:** 487–512.

Aldrich PR, Cavender-Bares J. 2011. *Quercus*. In: Kole C, ed. *Wild crop relatives: genomic and breeding resources*. Heidelberg: Springer, 89–129.

Aldrich PR, Michler CH, Sun W, Romero-Severson J. 2002. Microsatellite markers for northern red oak (Fagaceae: *Quercus rubra*). *Molecular Ecology Notes* **2:** 472–474.

Aldrich PR, Parker GR, Michler CH, Romero-Severson J. 2003. Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex

in an Indiana old-growth forest. *Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiére* **33:** 2228–2237.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench to detect molecular adaptation based on an F_{ST}-outlier method. *BMC Bioinformatics* **9:** 323.

Arnold ML. 2004. Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right? *The Plant Cell* **16:** 562–570.

Arnold ML, Bennett BD. 1993. Natural hybridization in Louisiana irises: genetic variation and ecological determinants. In: Harrison RG, ed. *Hybrid zones and the evolutionary process*. New York: Oxford University Press, 115–139.

Arnold ML, Martin NH. 2010. Hybrid fitness across time and habitats. *Trends in Ecology & Evolution* **25:** 530–536.

Ashton MS, Larson BC. 1996. Germination and seedling growth of *Quercus* (section *Erythrobalanus*) across openings in a mixed-deciduous forest of southern New England, USA. *Forest Ecology and Management* **80:** 81–94.

Bahari ZA, Pallardy SG, Parker WC. 1985. Photosynthesis, water relations, and drought adaptation in six woody species of oak-hickory forests in central Missouri. *Forest Science* **31:** 557–569.

Beaumont MA. 2005. Adaptation and speciation: what can F_{ST} tell us? *Trends in Ecology & Evolution* **20:** 435–440.

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.

Burger WC. 1975. The species concept in *Quercus*. *Taxon* **24:** 45–50.

Chesnoiu EN, Sofletea N, Curtu AL, Toader A, Radu R, Enescu M. 2009. Bud burst and flowering phenology in a mixed oak forest from Eastern Romania. *Annals of Forest Research* **52:** 199–206.

Collins E, Sullivan AR, Gailing O. 2015. Limited effective gene flow between two interfertile red oak species. *TREES – Structure and Function* **29:** 1135–1148.

Curtu A, Gailing O, Finkeldey R. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology* **7:** 218.

Dąbrowski MJ, Bornelöv S, Kruczyk M, Baltzer N, Komorowski J. 2015. 'True' null allele detection in microsatellite loci: a comparison of methods, assessment of difficulties and survey of possible improvements. *Molecular Ecology Resources* **15:** 477–488.

Dąbrowski MJ, Pilot M, Kruczyk M, Żmihorski M, Umer HM, Gliwicz J. 2014. Reliability assessment of null allele detection: inconsistencies between and within different methods. *Molecular Ecology Resources* **14:** 361–373.

Day FP, Monk CD. 1974. Vegetation patterns on a southern Appalachian watershed. *Ecology* **55:** 1064–1074.

Dodd RS, Afzal-Rafii Z. 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* **58:** 261–269.

Donoghue MJ. 1985. A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist* **88:** 172–181.

Durand J, Bodénès C, Chancerel E, Frigerio J-M, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn H-P, Villani F, Mattioni C, Cherubini M, Goicoechea PG, Herrán A, Ikaran Z, Cabané C, Ueno S, Alberto F, Dumoulin P-Y, Guichoux E, de Daruvar A, Kremer A, Plomion C. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* **11:** 570.

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

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

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.

Gailing O. 2013. Differences in growth, survival and phenology in *Quercus rubra* and *Q. ellipsoidalis* seedlings. *Dendrobiology* **70:** 73–81.

Gailing O, Bodénès C, Finkeldey R, Kremer A, Plomion C. 2013. Genetic mapping of EST-derived simple sequence repeats (EST-SSRs) to identify QTL for leaf morphological characters in a *Quercus robur* full-sib family. *Tree Genetics & Genomes* **9:** 1361–1367.

Gailing O, Lind J, Lilleskov EA. 2012. Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. elliposidalis* E. J. Hill populations in contrasting environments. *Plant Systematics and Evolution* **298:** 1533–1545.

Goicoechea PG, Herrán A, Durand J, Bodénès C, Plomion C, Kremer A. 2015. A linkage disequilibrium perspective on the genetic mosaic of speciation in two hybridizing Mediterranean white oaks. *Heredity* **114:** 373–386.

Hamilton JA, Aitken SN. 2013. Genetic and morphological structure of a spruce hybrid (*Picea sitchensis*× *P. glauca*) zone along a climatic gradient. *American Journal of Botany* **100:** 1651–1662.

Hardy OJ, Charbonnel N, Fréville H, Heuertz M. 2003. Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics* **163:** 1467–1482.

Hinckley T, Aslin R, Aubuchon R, Metcalf C, Roberts J. 1978. Leaf conductance and photosynthesis in four species of the oak-hickory forest type. *Forest Science* **24:** 73–84.

Hipp AL. 2010. Hill's oak: the taxonomy and dynamics of a Western Great Lakes endemic. *Arnoldia* **67:** 2–14.

Hipp AL, Weber JA. 2008. Taxonomy of Hill's oak (*Quercus ellipsoidalis*: Fagaceae): evidence from AFLP data. *Systematic Botany* **33:** 148–158.

Hipp AL, Weber JA, Srivastava A. 2010. Who am I this time? The affinities and misbehaviors of Hill's oak (*Quercus ellipsoidalis*). *International Oak Journal* **21:** 27–36.

Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Research* **9:** 868–877.

Jensen RJ. 1977. A preliminary numerical analysis of the red oak complex in Michigan and Wisconsin. *Taxon* **26:** 399–407.

Jensen RJ. 1997. *Quercus* Linnaeus sect. *Lobatae* Loudon, the red or black oaks. In: Flora of North America Editorial Committee, ed. *Flora of North America north of Mexico, Vol. 3*. New York and Oxford: Oxford University Press, 447–468.

Jensen RJ, Hokanson SC, Isebrands JG, Hancock JF. 1993. Morphometric variation in oaks of the apostle islands in Wisconsin – evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). *American Journal of Botany* **80:** 1358–1366.

Kalinowski ST. 2004. Counting alleles with rarefaction: private alleles and hierarchical sampling designs. *Conservation Genetics* **5:** 539–543.

Kalinowski ST. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* **5:** 187–189.

Kremer A, Dupouey JL, Deans JD, Cottrell J, Csaikl U, Finkeldey R, Espinel S, Jensen J, Kleinschmit J, Van Dam B, Ducousso A, Forrest I, Lopez de Heredia U, Lowe AJ, Tutkova M, Munro RC, Steinhoff S, Badeau V. 2002. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands. *Annals of Forest Science* **59:** 777–787.

Lagache L, Klein EK, Guichoux E, Petit RJ. 2013. Fine-scale environmental control of hybridization in oaks. *Molecular Ecology* **22:** 423–436.

Lepais O, Petit RJ, Guichoux E, Lavabre JE, Alberto F, Kremer A, Gerber S. 2009. Species relative abundance and direction of introgression in oaks. *Molecular Ecology* **18:** 2228–2242.

Lepais O, Roussel G, Hubert F, Kremer A, Gerber S. 2013. Strength and variability of postmating reproductive isolating barriers between four European white oak species. *Tree Genetics & Genomes* **9:** 841–853.

Lexer C, Kremer A, Petit RJ. 2006. Comment: shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology* **15:** 2007–2012.

Li Y-C, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution* **21:** 991–1007.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451–1452.

Lind J, Gailing O. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* **9:** 707–722.

Lind-Riehl JF, Sullivan AR, Gailing O. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. *Annals of Botany* **113:** 967–975.

Little EL Jr. 1971. *Atlas of United States trees, Volume 1: conifers and important hardwoods*. US Department of Agriculture Miscellaneous Publication 1146. Available at: http://esp.cr.usgs.gov/data/little/

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

Moore WS. 1977. An evaluation of narrow hybrid zones in vertebrates. *Quarterly Review of Biology* **52:** 263–277.

Moran EV, Willis J, Clark JS. 2012. Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae). *American Journal of Botany* **99:** 92–100.

Muir G, Schlötterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* **14:** 549–561.

Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW, Jiggins CD. 2012. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B* **367:** 343–353.

Narum SR, Hess JE. 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources* **11:** 184–194.

Nielsen EE, Bach LA, Kotlicki P. 2006. HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes* **6:** 971–973.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. Microchecker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4:** 535–538.

Owusu SA, Sullivan AR, Weber JA, Hipp AL, Gailing O. 2015. Taxonomic relationships and gene flow in four North American *Quercus* species (*Quercus* section *Lobatae*). *Systematic Botany* **40:** 510–521.

Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K. 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of Botany* **105:** 389–399.

Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161:** 151–164.

Plomion C, Aury J-M, Amselem J, Alaeitabar T, Barbe V, Belser C, Bergès H, Bodénès C, Boudet N, Boury C, Canaguier A, Couloux A, Da Silva C, Duplessis S, Ehrenmann F, Estrada-Mairey B, Fouteau S, Francillonne N, Gaspin C, Guichard C, Klopp C, Labadie K, Lalanne C, Le Clainche I, Leplé J-C, Le

Provost G, Leroy T, Lesur I, Martin F, Mercier J, Michotey C, Murat F, Salin F, Steinbach D, Faivre-Rampant P, Wincker P, Salse J, Quesneville H, Kremer A. 2016. Decoding the oak genome: public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources* **16:** 254–265.

Porth I, Scotti-Saintagne C, Barreneche T, Kremer A, Burg K. 2005. Linkage mapping of osmotic stress induced genes of oak. *Tree Genetics & Genomes* **1:** 31–40.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155:** 945–959.

Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86:** 248–249.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301:** 1211–1216.

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

Rozen S, Skaletsky H. 1999. Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz SA, eds. *Bioinformatics methods and protocols*. Totowa, NJ: Humana Press, 365–386.

Rundle HD. 2002. A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution* **56:** 322–329.

Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A. 2004. Distribution of genomic regions differentiating oak species assessed by QTL detection. *Heredity* **92:** 20–30.

Schlötterer C. 2002. A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160:** 753–763.

Schlötterer C, Dieringer D. 2005. A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In: Nurminsky D, ed. *Selective sweep*. New York: Kluwer Academic, 55–64.

Schlueter SD, Dong Q, Brendel V. 2003. GeneSeqer@ PlantGDB: gene structure prediction in plant genomes. *Nucleic Acids Research* **31:** 3597–3600.

Scotti-Saintagne C, Bodénès C, Barreneche T, Bertocchi E, Plomion C, Kremer A. 2004b. Detection of quantitative trait loci controlling bud burst and height growth in *Quercus robur* L. *Theoretical and Applied Genetics* **109:** 1648–1659.

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénès C, Kornel B, Kremer A. 2004a. Genome scanning for interspecific differentiation between two closely related oak species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168:** 1615–1626.

Seehausen O. 2006. Conservation: losing biodiversity by reverse speciation. *Current Biology* **16:** R334–R337.

Seidel KW. 1972. Drought resistance and internal water balance of oak seedlings. *Forest Science* **18:** 34–40.

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

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glössl J. 1997. Identification and characterization of (GA/CT) n-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology* **33:** 1093–1096.

Sullivan AR, Lind JF, McCleary T, Romero-Severson J, Gailing O. 2013. Development and characterization of genomic and gene-based microsatellite markers in North American red oak species. *Plant Molecular Biology Reporter* **31:** 231–239.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30:** 2725–2729.

Vähä JP, Primmer CR. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology* **15:** 63–72.

Van Valen L. 1976. Ecological species, multispecies, and oaks. *Taxon* **25:** 233–239.

Vasemägi A, Nilsson J, Primmer CR. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Molecular Biology and Evolution* **22:** 1067–1076.

Vekemans X, Beauwens T, Lemaire M, Roldán-Ruiz I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11:** 139–151.

Vonlanthen P, Bittner D, Hudson A, Young K, Müller R, Lundsgaard-Hansen B, Roy D, Di Piazza S, Largiader CR, Seehausen O. 2012. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482:** 357–362.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

Whitham TG, Martinsen GD, Floate KD, Dungey HS, Potts BM, Keim P. 1999. Plant hybrid zones affect biodiversity: tools for a genetic-based understanding of community structure. *Ecology* **80:** 416–428.

Whitney KD, Randell RA, Rieseberg LH. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist* **187:** 230–239.

Whittemore AT, Schaal BA. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the United States of America* **88:** 2540–2544.

Zhang R, Hipp A, Gailing O. 2015. Sharing of chloroplast haplotypes among red oak species suggests interspecific gene flow between neighboring populations. *Botany-Botanique* **93:** 691–700.

Zhivotovsky LA. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology* **8:** 907–913.

SUPPORTING INFORMATION

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

Figure S1. Optimal number of STRUCTURE clusters (*K*) according to log-likelihood probabilities (A) and Δ*K* (B, Evanno *et al.*, 2005).
Figure S2. Allele frequency distribution at the outlier microsatellite loci GOT040 (A) and POR016 (B) in *Quercus velutina*, *Q. rubra*, *Q. coccinea* and *Q. ellipsoidalis*.
Table S1. Populations, sample size (*N*), geographical coordinates and material/genotype sources used in this study.
Table S2. Sequencing primers, expected product size, annealing temperature and putative homology of conceptually translated gene products.
Table S3. STRUCTURE power and error test results for detection of *Q. ellipsoidalis*, *Q. velutina* and their hybrids and backcrosses.
Table S4. Average genetic admixture coefficients (*q*) for each population at $K = 4$ as identified by STRUCTURE based on 27 microsatellite and 259 AFLP markers in four *Quercus* spp.
Table S5. Number of purebred ($q \geq 0.90$), introgressive forms ($0.90 > q > 0.10$) and triple hybrids ($0.80 > q$ with ≥ 0.10 in more than other two species clusters) in neighbouring *Q. ellipsoidalis* and *Q. velutina* populations.
Table S6. Results of pairwise LnRH, LnRV and F_{ST}-based outlier tests in four *Quercus* spp.