NRC Research Press

ARTICLE

A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny[1]

Kasey K. Pham, Andrew L. Hipp, Paul S. Manos, and Richard C. Cronn

Abstract: Owing to high rates of introgressive hybridization, the plastid genome is poorly suited to fine-scale DNA barcoding and phylogenetic studies of the oak genus (*Quercus*, Fagaceae). At the tips of the oak plastome phylogeny, recent gene migration and reticulation generally cause topology to reflect geographic structure, while deeper branches reflect lineage divergence. In this study, we quantify the simple and partial effects of geographic proximity and nucleome-inferred phylogenetic history on oak plastome phylogeny at different evolutionary scales. Our study compares pairwise phylogenetic distances based on complete plastome sequences, pairwise phylogenetic distances from nuclear restriction site-associated DNA sequences (RADseq), and pairwise geographic distances for 34 individuals of the white oak clade representing 24 North American and Eurasian species. Within the North American white oak clade alone, phylogenetic history has essentially no effect on plastome variation, while geography explains 11%–21% of plastome phylogenetic variance. However, across multiple continents and clades, phylogeny predicts 30%–41% of plastome variation, geography 3%–41%. Tipwise attenuation of phylogenetic informativeness in the plastome means that in practical terms, plastome data has little use in solving phylogenetic questions, but can still be a useful barcoding or phylogenetic marker for resolving questions among major clades.

Key words: gene flow, hybridization, partial Mantel test, plastome sequencing, restriction-site associated DNA (RADseq).

Résumé : En raison de l'abondance d'hybridation introgressive, le génome plastidique convient peu à des études fines de codage à barres de l'ADN et aux études phylogénétiques chez les chênes (genre *Quercus*, Fagacées). Aux extrémités de la phylogénie du plastome du chêne, des événements récents de migration génique et de réticulation font généralement en sorte que la topologie reflète la structure géographique, tandis que les branches ancrées plus profondément reflètent davantage la divergence des lignées. Dans ce travail, les auteurs ont quantifié les effets simples et partiels de la proximité géographique et de l'historique phylogénétique, déduit en étudiant le génome nucléaire, sur la phylogénie du plastome du chêne à différentes échelles évolutives. Ce travail compare les distances phylogénétiques basées sur les séquences complètes des plastomes, les distances calculées sur la base de séquences RADseq nucléaires et les distances géographiques chez 34 individus du clade du chêne blanc représentant 24 espèces nord-américaines et eurasiennes. Au sein du seul clade du chêne blanc nord-américain, l'historique phylogénétique n'avait essentiellement aucun effet sur la variation observée au sein du plastome, tandis que la géographie expliquait entre 11 et 21 % de la variance génétique du plastome. Cependant, à l'échelle de plusieurs continents et clades, la phylogénie permettait d'expliquer entre 30 et 41 % de la variation plastomique et la géographie entre 3 et 41 %. L'atténuation de la valeur informative de la phylogénie aux extrémités de l'arbre signifie qu'en termes pratiques les données du plastome sont peu utiles pour répondre à des questions phylogénétiques, mais qu'elles demeurent utiles comme marqueurs, en tant que codes à barres ou en phylogénie, pour résoudre des questions au sein de principaux clades. [Traduit par la Rédaction]

Mots-clés : flux génique, hybridation, test de Mantel partiel, séquençage du plastome, séquençage de l'ADN associé aux sites de restriction (RADseq).

Introduction

Oaks have long been notorious among biologists for both high intraspecific morphological variation and low reproductive barriers (e.g., Wiegand 1935; Muller 1952; Burger 1975; Hardin 1975). Van Valen (1976) famously wrote, "It may well be that *Quercus macrocarpa* in Quebec exchanges many more genes with local *Q. bicolor* than it does with *Q. macrocarpa* in Texas" (Van Valen 1976, pg. 235). Other studies have since confirmed that while oak species with overlapping ranges exchange alleles through introgressive hybridization (Whittemore and Schaal 1991; Cavender-Bares and Pahlich 2009; Peñaloza-Ramírez et al. 2010), these events are rare relative to the rate of intraspecific gene flow (Muir et al. 2000; Hipp and Weber 2008; Cavender-Bares and Pahlich 2009; Gerber et al. 2014). This asymmetry of rates helps to maintain species boundaries even in sympatry and the genetic coherence of oak species over large geographic ranges. It appears that high rates of

Received 24 October 2016. Accepted 21 March 2017.

Corresponding Editor: Juan P. Jaramillo-Correa.

K.K. Pham. The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532-1293, USA; Department of Plant Biology, Michigan State University, East Lansing, MI 48824-1312, USA.
A.L. Hipp.* The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532-1293, USA; The Field Museum, 1400 S Lake Shore Drive, Chicago, IL 60605, USA.
P.S. Manos. Department of Biology, Duke University, Durham, NC 27708-0338, USA.
R.C. Cronn. Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331-4401, USA.

Corresponding authors: K.K. Pham (email: kase.khanh.pham@gmail.com); A.L. Hipp (email: ahipp@mortonarb.org).

*Andrew L. Hipp currently serves as a Guest Editor; peer review and editorial decisions regarding this manuscript were handled by Juan P. Jaramillo-Correa.

[1]This paper is part of a Special Issue entitled The Evolution of Tree Diversity.

intraspecific gene exchange ensure that introgressant alleles remain at low frequencies (Petit and Excoffier 2009), increasing their chance of loss through drift. Thus gene flow among conspecific populations counterbalances the homogenizing effect of hybridization in the oak nuclear genome, making nuclear DNA a reliable source of phylogenetic data (Eaton et al. 2015; Hipp 2015).

The plastid genome, long the workhorse of angiosperm phylogenetics (Chase et al. 1993; Ruhfel et al. 2014; Shaw et al. 2014 and references therein), is haploid and non-recombinant, and it has a smaller effective population size than nuclear alleles. These factors increase the rate of fixation of haplotypes within populations and species via drift. Moreover, maternally inherited plastids typically have lower rates of among-population migration, reducing their effectiveness in maintaining species cohesion (Petit and Excoffier 2009). For these reasons, the plastome often tracks patterns of geographically structured interspecific gene flow that dominates in lineages at recent phylogenetic scales (e.g., Rieseberg and Soltis 1991; Rieseberg et al. 1991; Whittemore and Schaal 1991; Sang et al. 1997; Petit and Excoffier 2009) and can reveal evidence of hybridization history at even deep phylogenetic scales (Folk et al. 2016). These and related studies demonstrate that while phylogeny frequently tracks geography at broad scales, localized interspecific gene flow is often detectable at finer scales through conflict between plastomic and genomic phylogenies.

Forest trees in particular frequently demonstrate plastome divergence patterns that conflict with known lineage divergence history (phylogeny), reflecting either the sorting of plastome lineages in large populations across large geographic regions or contemporary hybridization among interfertile species (Hamzeh and Dayanandan 2004; Shaw et al. 2012). Numerous studies of oaks have demonstrated that plastome haplotypes are shared freely among species of the same section, and that these haplotypes tend to cluster by geography instead of population divergence history or phylogenetic relationships (Whittemore and Schaal 1991; Petit et al. 1993; Dumolin-Lapegue et al. 1997; Simeone et al. 2016). At deeper phylogenetic scales, however, the plastome appears to track the monophyly of species within the major oak clades, whereas nuclear data provide evidence for higher order relationships among those clades (Manos et al. 1999; Oh and Manos 2008). The plastome may be thought of as complementary to the nuclear genome in oaks (Manos et al. 1999), providing resolution for ambiguous nuclear relationships near the base of the clade while reconstructing patterns of contemporary interspecific gene flow and introgression. The latter may be more weakly captured in the nuclear genome, where recombining loci differentially track segregation of ancestral alleles, population divergence, and ongoing gene flow (Via and West 2008).

Phylogeny and gene flow are typically considered separate targets of inquiry: studies investigating phylogeny generally treat gene flow as a nuisance factor, while studies of gene flow frequently ignore the contribution of phylogenetic history to variation observed in their molecular genetic datasets. Our study presents a novel approach to estimating the relative contributions of both phylogenetic signal and geographic proximity (a covariate of local gene flow) to patterns of oak plastome variation at different phylogenetic levels. We use whole plastome assemblies, nuclear restriction site-associated DNA sequencing (RADseq) data, and geographic source information to examine the joint and partial effects of population divergence history and local gene flow on the plastome phylogeny in white oaks (*Quercus* L. section *Quercus*) and intermediate oaks (*Quercus* section *Protobalanus* (Trelease) A. Camus).

Materials and methods

Sampling

Forty-three RADseq and 45 plastome samples were drawn from the distribution of white and intermediate oaks (sections *Quercus* and *Protobalanus*, respectively) across North America to detect phylogeographic structure at a continental scale; 33 of these samples overlapped between datasets. Three Eurasian white oak species (*Quercus mongolica* Fisch. ex. Ledeb.; *Q. petraea* (Mattuschka) Libel.; *Q. robur* L.) were included for both datasets to assess geographic and phylogenetic structure between North America and Eurasia based on RADseq and plastome data. An additional 51 white oak samples focused predominantly on four Eurasian white oak species (the three aforementioned, plus *Q. dentata* Thunb.) were also surveyed for plastome variation to more exhaustively assess plastome phylogenetic and geographically structured genetic variation among North American, European, and Asian white oaks. The total number of plastome sequences reported here is 91. A previously published red oak plastid genome (*Q. rubra* L.; NCBI NC_020152) was included as an outgroup, and one RADseq sample from a different individual was added as the RADseq outgroup. Thus the total number of RADseq samples presented in this paper is 47. Geographic coordinates (latitude and longitude) were collected for source trees using GPS.

Restriction-site associated DNA sequencing (RADseq)

RADseq DNA extraction, library preparation, and sequencing were conducted as presented previously by A. Hipp and colleagues (Hipp et al. 2014; Eaton et al. 2015; Cavender-Bares et al. 2015). Briefly, DNA for all RADseq samples was extracted from fresh or frozen material using the DNeasy plant extraction protocol (DNeasy, Qiagen, Valencia, Calif.). DNA extractions were gel quantified in agarose by visual comparison with the New England BioLabs 100 bp DNA Ladder (NEB, Ipswich, Mass.). Extraction concentrations ranged from 5 to 10 ng DNA/μL extraction. RAD sequencing library preparation was conducted at Floragenex, Inc. (Portland, Oreg., USA) following the methods of Baird et al. (2008) with *Pst*I. RAD libraries were barcoded by individual and multiplexed on an Illumina Genome Analyzer IIx. Sequencing reads were 100 bp in length; after removal of the barcode and recognition sequence, analyzed sequences were 85 bp long. Quality, read lengths, and base composition of FASTQ data were assessed in R v. 3.3.1 ("Bug in your hair"; R Core Team 2016) using the ShortRead package (Morgan et al. 2009).

RADseq clustering

Data were analyzed using the PyRAD pipeline (Eaton 2014; www.dereneaton.com/software). In this pipeline, sequences were clustered first by individual, and highly similar sequences are clustered into "stacks." In pyRAD, these stacks were generated using VSEARCH (Rognes et al. 2016), which allows sequences within clusters to vary in indels, nucleotide polymorphisms, and sequencing strand (direction). After clustering, heterozygosity and sequencing error were jointly estimated from the base counts observed across all sequences, sites, and clusters using the likelihood equation of Lynch (2008). Heterozygotes were inferred by a binomial probability based on these parameters. Bases that could not be assigned with ≥95% probability were treated as unknown (N). Any locus possessing more than two haplotypes within individuals after correcting for sequencing errors was discarded, under the assumption that it included one or more paralogous sequences. For each individual, each locus was summarized as a consensus sequence, and consensus sequences were clustered among individuals to generate a data matrix for each locus. Owing to variation in sequencing coverage and mutation at the restriction site defining RAD loci, the resulting data matrix was not complete for all loci in all individuals. The following parameters were specified for pyRAD: Minimum depth of reads per within-sample cluster: 6; maximum number of sites in a read that can have a quality score of less than 20: 4; clustering threshold (percent similarity): 0.85; minimum number of samples in each across-sample cluster: 4; maximum number of individuals with a shared heterozygous site in an across-sample cluster: 3. All other settings used default values.

RADseq data matrix

After clustering, consensus sequences of resulting RADseq loci were produced in R (R Core Team 2016) using the RADami package (Hipp 2014) and mapped to a complete *Q. rubra* oak plastome using Bowtie2 (Langmead and Salzberg 2012) with default settings. Loci with a positive match to the plastome were removed using RADami. Samples of the western North American *Q. gambelii* Nutt. were initially included, but preliminary analyses of the dataset demonstrated that *Q. gambelii* samples in combination with *Q. macrocarpa* Michx. destabilize the white oak RADseq phylogeny, suggesting a complex genomic composition that could be the product of hybridization and that is beyond the scope of the current paper. This issue is addressed in a separate paper in this volume (McVay et al. 2017*a*), but in brief, inclusion of *Q. gambelii* has the effect of pulling *Q. lobata* Née toward *Q. macrocarpa* and relatives, at odds with all nuclear data. Samples of *Q. gambelii* were therefore removed from the RADseq dataset for final analyses. RADseq loci were subsetted for alternative analyses presented below and concatenated for analysis in RAxML (Stamatakis 2014) using RADami and exported as a FASTA file. Concatenation order was the numerical sort order of the locus names, which are assigned arbitrarily during the clustering pipeline implemented in pyRAD.

Plastome sequencing and plastome assembly

Individual DNA aliquots (0.5–1 g) were sheared to a median length of ~300 bp and converted into individually indexed sequencing libraries using Illumina TruSeq v.2 kits at the USDA Forest Service (Corvallis, Oreg., USA), as described in Schroeder et al. (2016). Briefly, sequencing was performed using the Illumina MiSeq with 2 × 150 bp paired-end reads to produce individual de novo genome reference assemblies for representative North American, European, and Asian species (*Q. alba*, *Q. petraea*, and *Q. mongolica*, respectively); and the Illumina HiSeq with 100 bp single-end reads to sequence individual samples for reference-guided read mapping and genome assembly. All reactions used version 3 sequencing chemistry. Information on raw clusters, sequence yield, and approximate target sequence (plastid genome) coverage depth is provided in the supplementary data (Supplement S2[2]).

Plastome data matrix

Raw read quality filtering, de novo assembly, and reference-guided assembly for these samples are described in Schroeder et al. (2016). Briefly, raw read quality filtering was accomplished using Trimmomatic v0.30 (Bolger et al. 2014) to remove reads with a mean Phred score less than 33. Reference-guided read mapping and plastome assembly were performed using CLC Genomics Workbench v. 7.5.1 (CLC-Bio; Aarhus, Denmark) and the 135 258 bp *Q. mongolica* draft plastid genome reference (accession MOR360/QUMO5_CH_1; Schroeder et al. 2016). This chloroplast genome reference contains only one of the two inverted repeat regions that are present in red oak chloroplast genomes (Alexander and Woeste 2014), so our reference size (135.2 kb) is smaller than a complete oak chloroplast genome (~161 kb). Reads were mapped to the reference using a length fraction of 0.9, a similarity fraction of 0.94, a mismatch cost of 2, and insertion/deletion costs of 6. A minimum depth of 5 or greater was applied to call reference and non-reference bases; sites with a depth less than 5 were coded as *N*. Ingroup plastome assemblies were aligned in MAFFT v7.149b (Katoh and Standley 2013), using gap opening and extension penalties of 2.0 and 0.1, respectively. The ingroup matrix was aligned to the *Q. rubra* outgroup using MAFFT profile-profile alignment. The resulting alignment was edited manually in Geneious version R8 (http://www.geneious.com, Kearse et al. 2012), primarily to remove large insertions present in the ingroup plastome alignment. The resulting alignment was processed in gblocks (Castresana 2000), using default parameters for DNA to remove positions where differences in the ingroup and outgroup sequences resulted in erroneous or ambiguous alignments. Ambiguous positions (Ns) were retained in alignments and analysis except where gblocks removed them. Unlike in the RADseq matrix, *Q. gambelii* was retained in the plastome matrix, as intragenomic recombination or genealogical discordance is not an issue in the plastome.

Phylogenetic analyses

Phylogenetic analyses were conducted under maximum likelihood using RAxML v. 8.2.4 on the RADseq and plastome alignments separately. Given the size and complexity of the genomic data, we utilized the relatively parameter-rich GTR + gamma model, under the GTRCAT approximation. For each dataset, bootstrap proportions were generated from 100 non-parametric bootstrap replicates and mapped to the internal branches of the ML tree. For computational reasons, we utilized fast bootstrapping (-f a option in RaxML) for all analyses presented in this paper. A comparison of fast bootstrapping with thorough bootstrapping (-b in RAxML) on the full plastome tree (Fig. 1) shows no systematic difference between bootstrapping methods (Supplement S1[2]). Trees were rooted using *Q. rubra* for both datasets.

Modeling the effect of geography and population divergence history on plastome phylogeny

The relative predictive power of phylogeny and geography on plastome similarity was estimated in R using the packages ape (Paradis et al. 2004) and phytools (Revell 2012) and the morton R library (https://github.com/andrew-hipp/morton). Cophenetic pairwise distances (branch length distances, estimated in RAxML according to the GTR+gamma nucleotide substitution model as described above) between all tips on a single tree were calculated for both the RADseq and plastome trees as our estimates of plastome pairwise dissimilarity (Δ_{plast}) and nuclear DNA pairwise dissimilarity (Δ_{nuc}). Latitude and longitude data for the samples were used to calculate the geographic distance between samples (Δ_{geog}) using the Haversine formula as implemented in the morton library. Multiple and partial correlation coefficients were estimated, with Mantel permutations to assess significance (Mantel 1967; Smouse et al. 1986; Diniz-Filho et al. 2013), on the following regression models: plastome variation predicted by both gene flow and nuclear genetic variation ($\Delta_{plast} \sim \Delta_{geog} + \Delta_{nuc}$); plastome variation predicted by gene flow alone ($\Delta_{plast} \sim \Delta_{geog}$); plastome variation predicted by nuclear genetic variation alone ($\Delta_{plast} \sim \Delta_{nuc}$); and nuclear genetic variation predicted by gene flow alone ($\Delta_{nuc} \sim \Delta_{geog}$). Only the position of samples in the response matrix was permuted, under the assumption that the prediction matrices were known without error (Smouse et al. 1986).

To assess how a dataset's phylogenetic depth affects the partial effects of lineage divergence history and gene flow on plastome variation, the dataset was pruned to several different phylogenetic levels and the Mantel tests described above were performed on each. The pruned datasets are as follows. (1) All oaks sampled (n = 33 samples), in which trees are pruned to only tips shared between the RADseq and plastome trees. This dataset contains both Eurasian and North American oaks in sections *Quercus* (the white oaks) and *Protobalanus* (the intermediate oaks). It spans the widest geographic range and largest phylogenetic scale. (2) The North American oaks (n = 30 samples): excludes the Eurasian white oaks from all oaks sampled. This exclusion set allows us to test whether geographic signal in the plastome is amplified by studying populations separated by an ocean. (3) The North American white oaks (n = 25 samples): excludes the intermediate oaks (section *Protobalanus*) from the North American oak dataset. This exclusion set allows us to examine the effects of removing a large

[2]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/gen-2016-0191.

Fig. 1. Plastome phylogeny, all samples. Maximum likelihood phylogeny with non-parametric bootstrap values indicated by branch thickness (thickest branches: bootstrap >90%; medium thickness: bootstrap ≤90% – >50%; thinnest branches: bootstrap ≤50%). Colored blocks indicate the four main plastome lineages. Sample names follow Table 1.

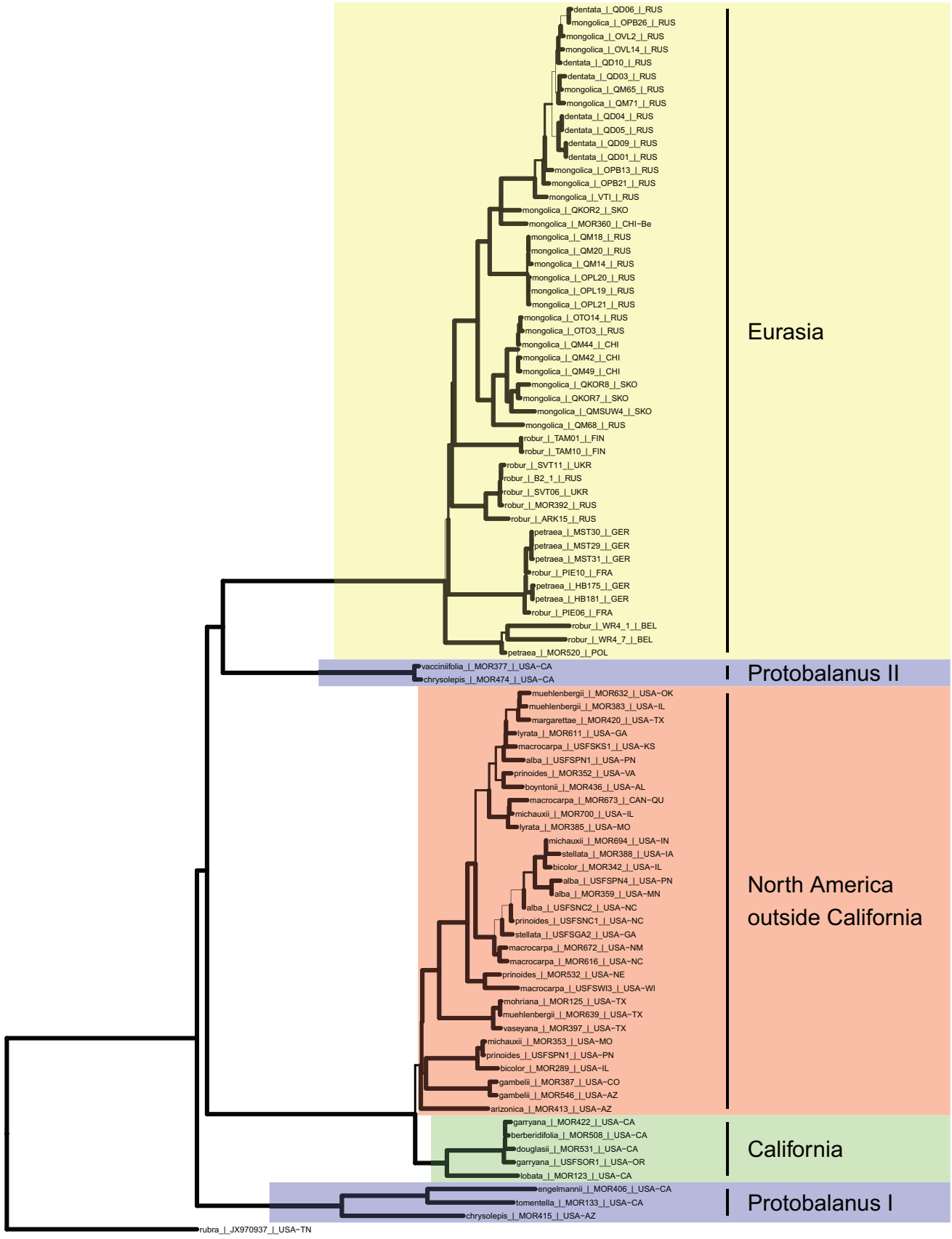


Table 1. Samples metadata.

Specimen code	Species	Herbarium accession	Latitude	Longitude	Section	Continent	Country	State/Province	RADseq SRA BioSample ID	cpDNA SRA BioSample ID
OAK-MOR-359	*Quercus alba* L.	174462	45.367983	−93.219304	*Quercus*	North America	USA	Minnesota	SAMN06446142	SAMN03264816
PM-19	*Quercus alba* L.	177669	36.021818	−79.016068	*Quercus*	North America	USA	North Carolina	SAMN06446143	—
OAK-MOR-204	*Quercus alba* L.	175336	39.9347718	−89.8016103	*Quercus*	North America	USA	Illinois	SAMN06446144	—
OAK-MOR-76	*Quercus alba* L.	174513	41.866545	−86.349921	*Quercus*	North America	USA	Michigan	SAMN06446145	—
QUAL_NC_2	*Quercus alba* L.	—	35.25357	−82.19709	*Quercus*	North America	USA	North Caolina	—	SAMN03264817
QUAL_PN_1	*Quercus alba* L.	—	41.38389	−79.054	*Quercus*	North America	USA	Pennsylvania	—	SAMN03264818
QUAL_PN_4	*Quercus alba* L.	—	41.50643	−79.24511	*Quercus*	North America	USA	Pennsylvania	—	SAMN03264819
OAK-MOR-413	*Quercus arizonica* Sarg.	174229	31.44959506	−109.2431153	*Quercus*	North America	USA	Arizona	SAMN06446146	SAMN03264796
OAK-MOR-508	*Quercus berberidifolia* Liebm.	174692	38.41167	−122.04917	*Quercus*	North America	USA	California	SAMN06446147	SAMN03264799
OAK-MOR-342	*Quercus bicolor* Willd.	174534	41.740864	−87.860335	*Quercus*	North America	USA	Illinois	SAMN06446148	SAMN03264820
OAK-MOR-289	*Quercus bicolor* Willd.	175338	38.9048	−94.8926	*Quercus*	North America	USA	Illinois	SAMN06446149	SAMN03264821
OAK-MOR-436	*Quercus boyntonii* Beadle	175278	34.014264	−86.00664	*Quercus*	North America	USA	Alabama	SAMN06446150	SAMN03264814
OAK-MOR-415	*Quercus chrysolepis* Liebm.	174230	31.43347498	−109.2431153	*Protobalanus*	North America	USA	Arizona	SAMN06446151	SAMN03264797
OAK-MOR-474	*Quercus chrysolepis* Liebm.	174682	38.757996	−120.548892	*Protobalanus*	North America	USA	California	SAMN06446152	SAMN03264807
Qd_01	*Quercus dentata* Thunb.	—	42.801233	131.245433	*Quercus*	Europe	RUS	—	—	SAMN03653768
Qd_03	*Quercus dentata* Thunb.	—	42.801267	131.245467	*Quercus*	Europe	RUS	—	—	SAMN03653764
Qd_04	*Quercus dentata* Thunb.	—	42.801283	131.245483	*Quercus*	Europe	RUS	—	—	SAMN03653765
Qd_05	*Quercus dentata* Thunb.	—	42.8013	131.2455	*Quercus*	Europe	RUS	—	—	SAMN03653766
Qd_06	*Quercus dentata* Thunb.	—	42.801317	131.245517	*Quercus*	Europe	RUS	—	—	SAMN03653769
Qd_09	*Quercus dentata* Thunb.	—	42.801367	131.245567	*Quercus*	Europe	RUS	—	—	SAMN03653767
Qd_10	*Quercus dentata* Thunb.	—	42.801383	131.245583	*Quercus*	Europe	RUS	—	—	SAMN03653762
OAK-MOR-531	*Quercus douglasii* Hook. & Arn.	174681	38.696153	−120.887322	*Quercus*	North America	USA	California	SAMN06446153	SAMN03264808
OAK-MOR-406	*Quercus engelmannii* Greene	174583	32.594883	−116.843812	*Quercus*	North America	USA	California	SAMN06446154	SAMN03264801
OAK-MOR-387	*Quercus gambelii* Nutt.	174463	38.861158	−105.172682	*Quercus*	North America	USA	Colorado	SAMN06446155	SAMN03264809
OAK-MOR-546	*Quercus gambelii* Nutt.	174218	31.75663448	−109.2431153	*Quercus*	North America	USA	Arizona	SAMN06446156	SAMN03264795
QUGA4_OR_1	*Quercus garryana* Douglas ex. Hook.	—	43.3050	−123.2412	*Quercus*	North America	USA	Oregon	—	SAMN03264824
OAK-MOR-422	*Quercus garryana* Douglas ex. Hook.	174699	38.678479	−120.812513	*Quercus*	North America	USA	California	SAMN06446157	SAMN03264823
OAK-MOR-123	*Quercus lobata* Née	174703	36.61667	−120.85	*Quercus*	North America	USA	California	SAMN06446158	SAMN03264805
OAK-MOR-385	*Quercus lyrata* Walter	174540	36.720972	−90.112917	*Quercus*	North America	USA	Missouri	SAMN06446159	SAMN03264826
OAK-MOR-611	*Quercus lyrata* Walter	174529	30.7084	−84.8564	*Quercus*	North America	USA	Georgia	SAMN06446160	SAMN03264825
OAK-MOR-356	*Quercus macrocarpa* Michx.	174544	37.149225	−94.443003	*Quercus*	North America	USA	Missouri	SAMN06446161	—
OAK-MOR-357	*Quercus macrocarpa* Michx.	174520/17	41.486798	−87.799832	*Quercus*	North America	USA	Illinois	SAMN06446162	—
OAK-MOR-673	*Quercus macrocarpa* Michx.	175524	45.5129	−73.5506	*Quercus*	North America	CAN	Quebec	SAMN06446163	SAMN03264829
OAK-MOR-672	*Quercus macrocarpa* Michx.	175314	33.6	−105.4	*Quercus*	North America	USA	New Mexico	SAMN06446164	SAMN03264812
QUMA2_KS_1	*Quercus macrocarpa* Michx.	—	39.0286	−94.9404	*Quercus*	North America	USA	Kansas	—	SAMN03264827
QUMA2_WI_3	*Quercus macrocarpa* Michx.	—	42.8751	−88.3276	*Quercus*	North America	USA	Wisconsin	—	SAMN03264830
OAK-MOR-420	*Quercus margarettae* (Ashe) Small	174705	29.616474	−97.572742	*Quercus*	North America	USA	Texas	SAMN06446165	SAMN03264804
OAK-MOR-353	*Quercus michauxii* Nutt.	174516	36.644722	−89.285	*Quercus*	North America	USA	Missouri	SAMN06446166	SAMN03264833
OAK-MOR-694	*Quercus michauxii* Nutt.	174537	38.905693	−86.035881	*Quercus*	North America	USA	Indiana	SAMN06446167	SAMN03264832
OAK-MOR-700	*Quercus michauxii* Nutt.	175231	37.15382	−89.34699	*Quercus*	North America	USA	Illinois	SAMN06446168	SAMN03264831
PM143	*Quercus michauxii* Nutt.	—	35.99555	−79.05416	*Quercus*	North America	USA	North Carolina	SAMN06446169	—
PM155	*Quercus michauxii* Nutt.	—	36.015223	−78.923317	*Quercus*	North America	USA	North Carolina	SAMN06446170	—
OAK-MOR-125	*Quercus mohriana* Buckley ex. Rydb.	174704	31.979167	−104.754722	*Quercus*	North America	USA	Texas	SAMN06446171	SAMN03264800
OAK-MOR-360	*Quercus mongolica* Fisch. ex. Ledeb.	174518	40.58	116.7375	*Quercus*	Europe	CHI	Beijing Shi	SAMN06446172	SAMN03264834
Opb_13	*Quercus mongolica* Fisch. ex. Ledeb.	—	50.187167	138.591	*Quercus*	Europe	RUS	—	—	SAMN03653770
Opb_21	*Quercus mongolica* Fisch. ex. Ledeb.	—	50.187167	138.591	*Quercus*	Europe	RUS	—	—	SAMN03653771
Opb_26	*Quercus mongolica* Fisch. ex. Ledeb.	—	50.187167	138.591	*Quercus*	Europe	RUS	—	—	SAMN03653772
Opl_19	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.374833	133.892	*Quercus*	Europe	RUS	—	—	SAMN03653773

Table 1 (*continued*).

Specimen code	Species	Herbarium accession	Latitude	Longitude	Section	Continent	Country	State/Province	RADseq SRA BioSample ID	cpDNA SRA BioSample ID
Opl_20	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.374833	133.892	*Quercus*	Europe	RUS	—	—	SAMN03653774
Opl_21	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.374833	133.892	*Quercus*	Europe	RUS	—	—	SAMN03653775
Oto_03	*Quercus mongolica* Fisch. ex. Ledeb.	—	49.0497	131.860333	*Quercus*	Europe	RUS	—	—	SAMN03653777
Oto_14	*Quercus mongolica* Fisch. ex. Ledeb.	—	49.0497	131.860333	*Quercus*	Europe	RUS	—	—	SAMN03653776
Ovl_02	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.073167	131.909833	*Quercus*	Europe	RUS	—	—	SAMN03653779
Ovl_04	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.073167	131.909833	*Quercus*	Europe	RUS	—	—	SAMN03653780
Ovl_14	*Quercus mongolica* Fisch. ex. Ledeb.	—	43.073167	131.909833	*Quercus*	Europe	RUS	—	—	SAMN03653778
Qkor_02	*Quercus mongolica* Fisch. ex. Ledeb.	—	37.5667	126.9781	*Quercus*	Europe	SKO	—	—	SAMN03653781
Qkor_07	*Quercus mongolica* Fisch. ex. Ledeb.	—	37.5667	126.9781	*Quercus*	Europe	SKO	—	—	SAMN03653782
Qkor_08	*Quercus mongolica* Fisch. ex. Ledeb.	—	—	—	*Quercus*	Europe	SKO	—	—	SAMN03653803
Qm_14	*Quercus mongolica* Fisch. ex. Ledeb.	—	45.083333	133.947194	*Quercus*	Europe	RUS	—	—	SAMN03653784
Qm_18	*Quercus mongolica* Fisch. ex. Ledeb.	—	46.083028	133.945889	*Quercus*	Europe	RUS	—	—	SAMN03653785
Qm_20	*Quercus mongolica* Fisch. ex. Ledeb.	—	46.082194	133.945556	*Quercus*	Europe	RUS	—	—	SAMN03653786
Qm_42	*Quercus mongolica* Fisch. ex. Ledeb.	—	45.429692	127.092031	*Quercus*	Europe	CHI	—	—	SAMN03653800
Qm_44	*Quercus mongolica* Fisch. ex. Ledeb.	—	45.429531	127.091156	*Quercus*	Europe	CHI	—	—	SAMN03653801
Qm_49	*Quercus mongolica* Fisch. ex. Ledeb.	—	45.428283	127.090569	*Quercus*	Europe	CHI	—	—	SAMN03653802
Qm_65	*Quercus mongolica* Fisch. ex. Ledeb.	—	48.619717	135.90783	*Quercus*	Europe	RUS	—	—	SAMN03653787
Qm_68	*Quercus mongolica* Fisch. ex. Ledeb.	—	48.619717	135.90783	*Quercus*	Europe	RUS	—	—	SAMN03653788
Qm_71	*Quercus mongolica* Fisch. ex. Ledeb.	—	48.619717	135.90783	*Quercus*	Europe	RUS	—	—	SAMN03653789
Qm_SUW_4	*Quercus mongolica* Fisch. ex. Ledeb.	—	37.5667	126.9781	*Quercus*	Europe	SKO	—	—	SAMN03653783
OAK-MOR-352	*Quercus montana* Willd.	174846	37.141444	–79.995722	*Quercus*	North America	USA	Virginia	SAMN06446173	SAMN03264838
OAK-MOR-693	*Quercus montana* Willd.	174514	37.52575	–80.249722	*Quercus*	North America	USA	Virginia	SAMN06446174	—
OAK-MOR-575	*Quercus montana* Willd.	175353	35.429102	–82.251822	*Quercus*	North America	USA	North Carolina	SAMN06446175	—
OAK-MOR-383	*Quercus muehlenbergii* Engelm.	174533	41.210569	–88.017575	*Quercus*	North America	USA	Illinois	SAMN06446176	SAMN03264810
OAK-MOR-632	*Quercus muehlenbergii* Engelm.	175345	35.623899	–99.008733	*Quercus*	North America	USA	Oklahoma	SAMN06446177	SAMN03264815
OAK-MOR-639	*Quercus muehlenbergii* Engelm.	174690	31.97917	–104.75417	*Quercus*	North America	USA	Texas	SAMN06446178	SAMN03264803
OAK-MOR-520	*Quercus petraea* (Matt.) Liebl.	174539	51.816667	19.883333	*Quercus*	Europe	POL	—	SAMN06446179	SAMN03264835
HB_175	*Quercus petraea* (Matt.) Liebl.	—	53.700539	10.719256	*Quercus*	Europe	GER	—	—	SAMN03653790
HB_181	*Quercus petraea* (Matt.) Liebl.	—	53.700539	10.719256	*Quercus*	Europe	GER	—	—	SAMN03653791
MST_29	*Quercus petraea* (Matt.) Liebl.	—	51.909356	7.740925	*Quercus*	Europe	GER	—	—	SAMN03653792
MST_30	*Quercus petraea* (Matt.) Liebl.	—	51.909356	7.740925	*Quercus*	Europe	GER	—	—	SAMN03653793
MST_31	*Quercus petraea* (Matt.) Liebl.	—	51.909356	7.740925	*Quercus*	Europe	GER	—	—	SAMN03653794
OAK-MOR-532	*Quercus prinoides* Willd.	175238	40.076297	–95.720956	*Quercus*	North America	USA	Nebraska	SAMN06446180	SAMN03264811
QUPR2_NC_1	*Quercus prinoides* Willd.	—	35.2535	–82.197	*Quercus*	North America	USA	North Carolina	—	SAMN03264836
QUPR2_PN_1	*Quercus prinoides* Willd.	—	41.5073	–79.2442	*Quercus*	North America	USA	Pennsylvania	—	SAMN03264837
OAK-MOR-392	*Quercus robur* L.	174567	55.838647	37.600483	*Quercus*	Europe	RUS	—	SAMN06446181	SAMN03264839
ARK_15	*Quercus robur* L.	—	54.416667	56.666667	*Quercus*	Europe	RUS	—	—	SAMN03653809
B2_01	*Quercus robur* L.	—	52.674197	58.742517	*Quercus*	Europe	RUS	—	—	SAMN03653795
PIE_06	*Quercus robur* L.	—	44.737028	–0.776389	*Quercus*	Europe	FRA	—	—	SAMN03653797
PIE_10	*Quercus robur* L.	—	44.737028	–0.776389	*Quercus*	Europe	FRA	—	—	SAMN03653796
SVT_06	*Quercus robur* L.	—	49.052833	37.9295	*Quercus*	Europe	UKR	—	—	SAMN03653807
SVT_11	*Quercus robur* L.	—	49.052833	37.9295	*Quercus*	Europe	UKR	—	—	SAMN03653808
TAM_01	*Quercus robur* L.	—	59.974439	23.435525	*Quercus*	Europe	FIN	—	—	SAMN03653798
TAM_10	*Quercus robur* L.	—	59.974439	23.435525	*Quercus*	Europe	FIN	—	—	SAMN03653799
WR4_01	*Quercus robur* L.	—	53.685833	23.792833	*Quercus*	Europe	BEL	—	—	SAMN03653805
WR4_07	*Quercus robur* L.	—	53.685833	23.792833	*Quercus*	Europe	BEL	—	—	SAMN03653806
OAK-MOR-498	*Quercus rubra* L.	174465	—	—	*Lobatae*	North America	USA	Illinois	SAMN06446182	—
OAK-MOR-388	*Quercus stellata* Wangenh.	174535	40.5849	–91.2756	*Quercus*	North America	USA	Iowa	SAMN06446183	SAMN03264841

Table 1 (*concluded*).

Specimen code	Species	Herbarium accession	Latitude	Longitude	Section	Continent	Country	State/Province	RADseq SRA BioSample ID	cpDNA SRA BioSample ID
OAK-MOR-340	*Quercus stellata* Wangenh.	174543	38.034944	−91.520306	*Quercus*	North America	USA	Missouri	SAMN06446184	—
OAK-MOR-616	*Quercus stellata* Wangenh.	175244	35.4	−82.4	*Quercus*	North America	USA	North Carolina	SAMN06446185	SAMN03264813
QUST_GA_2	*Quercus stellata* Wangenh.	—	33.9025	−83.3824	*Quercus*	North America	USA	Georgia	—	SAMN03264840
OAK-MOR-133	*Quercus tomentella* Engelm.	174688	34.023206	−119.765766	*Protobalanus*	North America	USA	California	SAMN06446186	SAMN03264802
OAK-MOR-377	*Quercus vaccinifolia* Kellogg	174684	38.811462	−120.123212	*Protobalanus*	North America	USA	California	SAMN06446187	SAMN03264806
OAK-MOR-397	*Quercus vaseyana* Buckley	174580	29.970985	−101.109021	*Quercus*	North America	USA	Texas	SAMN06446188	SAMN03264798

Note: SRA BioSample ID references NCBI sample ID for each sequence.

source of phylogenetic signal in the plastome. While there is one North American white oak (*Q. engelmannii* Greene) that has been shown to harbor a *Protobalanus* plastome (Manos et al. 1999), there is no evidence of ongoing gene flow between these two, and we know of no other cases of interbreeding between white oaks and intermediate oaks. (4) The North American oaks without California white oaks (n = 21 samples): excludes California oaks from the North American oak dataset. The California white oaks excluding *Q. engelmannii* are a well-defined clade (Sork et al. 2016) unlikely to exchange alleles with eastern North American oaks, making the California oaks another potential source of phylogenetic signal in the plastome. This exclusion set is analogous to the North American white oak dataset. The inclusion of the intermediate oaks but exclusion of California oaks tests whether gene flow and phylogenetic signal is comparable with either of the two western North American clades present in our dataset. (5) The eastern North American white oaks (n = 19 samples): excludes the Californian oaks and a clade of derived Mexican and southwestern North American oaks (represented in our study by *Q. arizonica* Sarg., *Q. engelmannii*, *Q. vaseyana* Buckley, and *Q. mohriana* Buckley ex. Rydb.) from the North American white oak dataset. This exclusion set is paraphyletic but concentrates the partial Mantel tests on just the potentially interbreeding eastern North American oaks, a group that has been particularly challenging to oak systematists (e.g., Hardin 1975). All exclusion sets are summarized in Supplement S3[2].

Results

RADseq clustering and data matrix

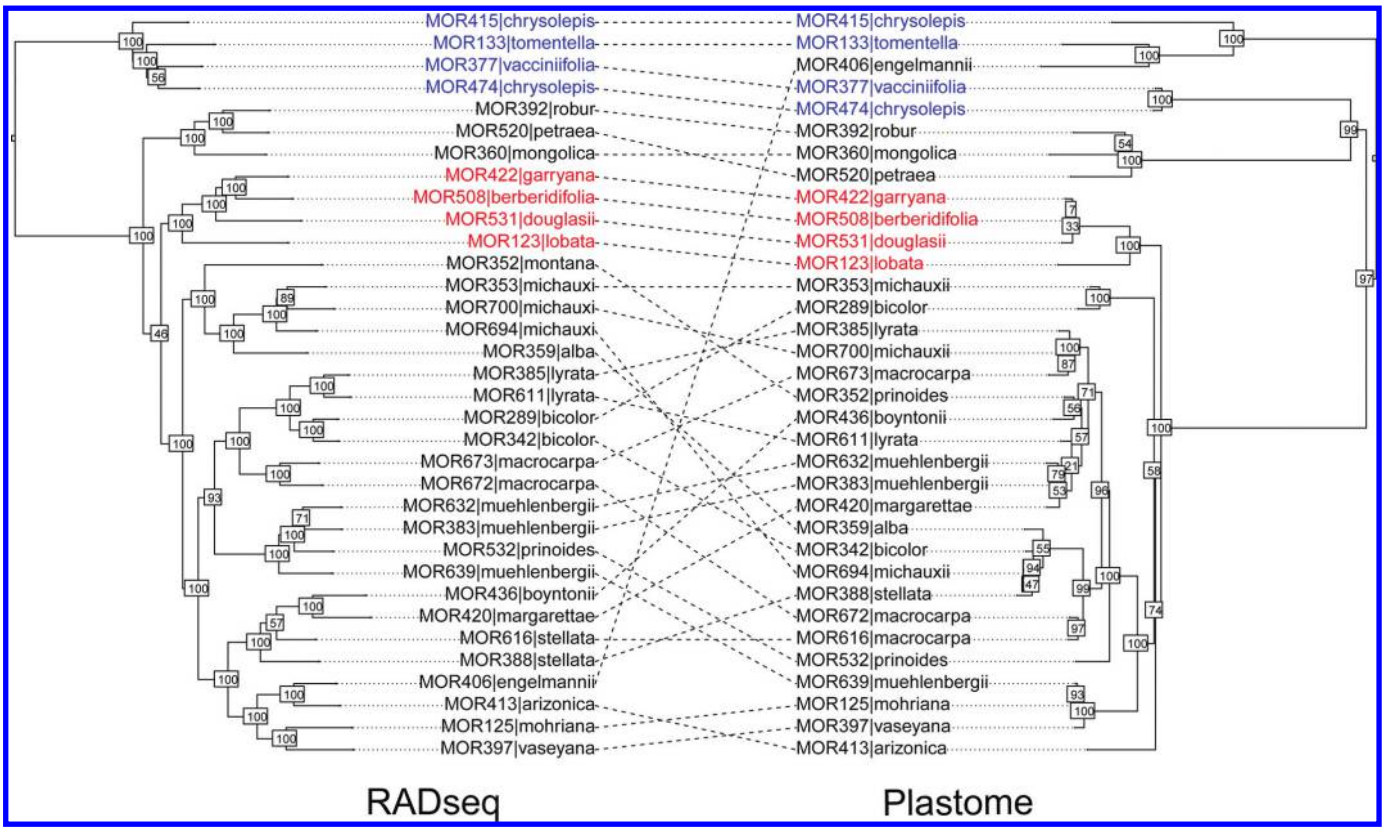
RADseq sequencing runs produced 8.43×10^5 to 3.49×10^6 sequences per individual, with a mean of $2.16 \times 10^6 \pm 1.87 \times 10^6$. Sequencing quality averages nearly 40 (out of 40) across bases 1–60, with high quality averaged across all runs (mean quality score = 36.7 ± 0.536). Including the outgroup *Q. rubra*, the RADseq library comprises 48 individuals and 6.92×10^4 loci, spanning 80–118 bp, with an average length of 87.6 ± 2.90 bp. The number of loci returned for single individuals ranges from 1.81×10^4 to 3.58×10^4, with a mean of $2.66 \times 10^4 \pm 4.07 \times 10^3$. The number of variable sites per locus averages 7.58 ± 5.62.

Seven loci (loci 508, 5924, 11871, 34638, 23273, 51846, 69261) mapped back to the *Q. rubra* plastid reference genome and were omitted from locus concatenation. In the *Q. rubra* plastome utilized in this study, there are 13 *Pst*I cut sites, suggesting that closer to 26 RADseq loci ought to have mapped back to the genome. However, several of these are in the inverted repeat region and may as a consequence have been collapsed into loci. Alternatively, our mapping may have underestimated the number of loci mapping back to the plastome. At any rate, even without removal of these seven loci, the plastome is a negligible contributor to the RADseq dataset, which we safely treat as a nuclear dataset in this study. Two *Q. gambelii* individuals (accession numbers MOR-387 and MOR-546) were also omitted from concatenation (see discussion above). Not including the outgroup, the final data matrix contains 45 individuals belonging to 25 species, and 6.06×10^6 nucleotide positions.

Plastome data matrix

Individual shotgun sequencing used for plastome assembly produced between 2.46×10^6 and 18.33×10^6 reads per individual, yielding 109×10^6 to 1765×10^6 bases (mean values: 8.17×10^6 reads and 763.5×10^6 bp per individual). Sequencing quality averaged 34.9 ± 0.1.73 for 101 bp reads. Reference-guided mapping to the *Q. mongolica* draft plastid genome reference (MOR360) showed that an average of 3.47% of sequence reads could be mapped to the reference (range: 0.36%–9.33%); this produced assemblies that averaged 204× depth and were nearly complete, with all sequences showing fewer than 350 Ns (mean = 84). Assembly information

Fig. 2. Nuclear RADseq phylogeny (left panel) and plastome phylogeny (right panel), containing all samples listed in Table 1. Corresponding tips are connected between trees. Bootstrap values are shown in boxes at the nodes of the clades to which they correspond. Tips are color-coded by clade; intermediate oaks are blue, california white oaks are red, and all other white oaks are black.



can be found in the supplementary data (Supplement S2[2]). Alignment to the *Q. rubra* reference plastid genome yielded a data matrix with 92 individuals, 26 species, and an overall length of 136 790 bp. Further refinement with gblocks removed 12 854 bp, resulting in a final matrix 132 936 bp in length. Alignments between white oaks (*Quercus* section *Quercus*) revealed 1852 ingroup variable sites across all genomes (including ambiguities), while adding members of *Quercus* section *Protobalanus* yielded an additional 306 ingroup variable sites (2158 total including ambiguities). The profile-to-profile plastome alignment is 136 790 bp long once aligned with the *Q. rubra* outgroup sequence, and contains 91 individuals comprising 26 species not including the outgroup. gblocks removes 3854 bp from the alignment, resulting in a final matrix that is 132 936 bp long.

Phylogenetic analyses

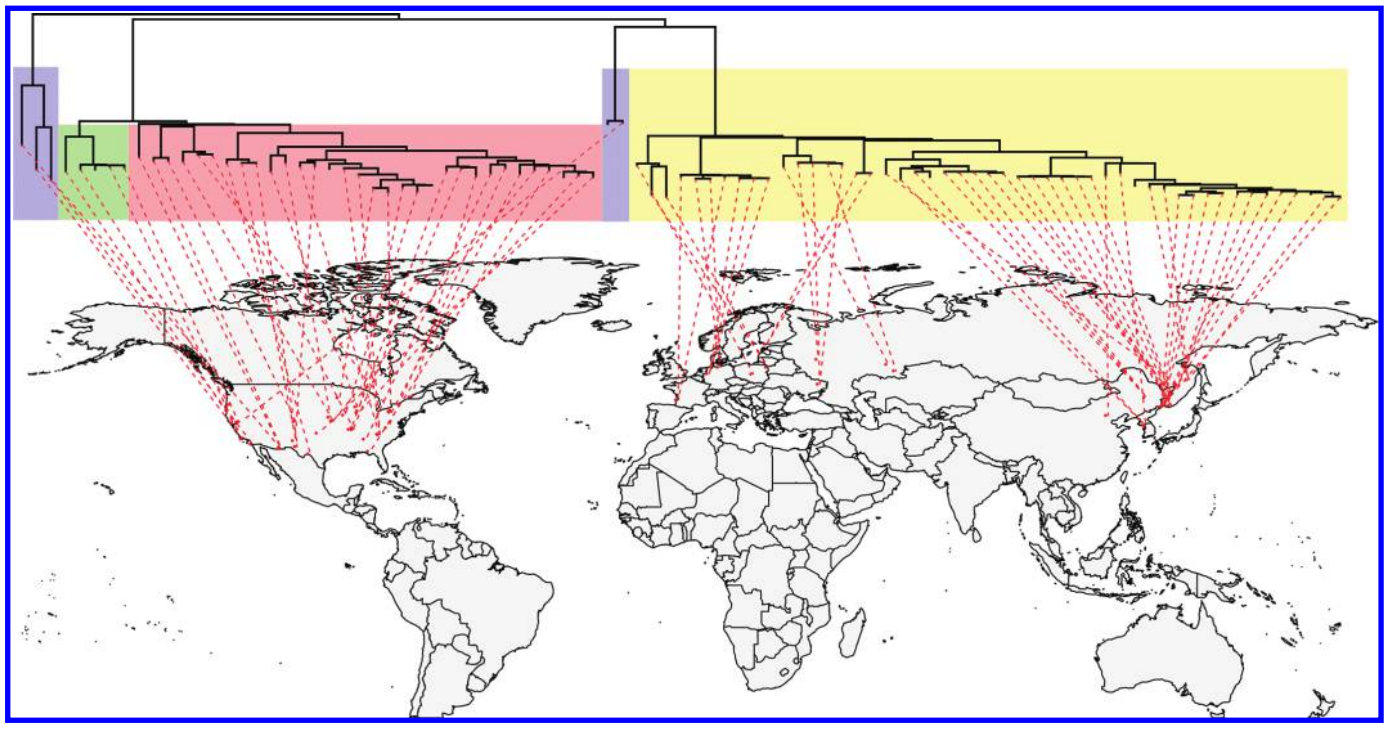
The plastome ML tree (Fig. 1) exhibits four deep plastome lineages, each supported with 100% bootstrap support, and each confined to either North America (three lineages) or Eurasia (one lineage). Sections *Quercus* and *Protobalanus* are both non-monophyletic in this tree, and one member of section *Quercus* (*Q. engelmannii* Greene) falls within an otherwise *Protobalanus*-dominated plastome lineage, as reported previously (Manos et al. 1999). Section *Protobalanus* exhibits a deep polyphyly that was also reported in Manos et al. (1999), with two lineages that show no evidence of recent plastome sharing (Fig. 1): species in this section segregate for comparatively ancient plastomes that conflict with sectional circumscription. Numerous species within the white oaks are non-monophyletic in the plastome tree, with accessions clustering coarsely by geography. For example, *Q. muehlenbergii* Engelm. is represented by three individuals in the plastome analysis (Fig. 1), and two of these samples (MOR-632 and MOR-383) do not resolve as closely related to the third sample (MOR-639). While this species together with *Q. prinoides* Willd. form a clade in the RADseq tree (Fig. 2), samples MOR-632 and MOR-383 were collected from the eastern United States (Illinois, Oklahoma) and cluster with other samples from the same geographic region (e.g., *Q. lyrata* from Georgia, *Q. macrocarpa* from Kansas, *Q. alba* from Pennsylvania). In contrast, the outlier sample (MOR-639) was collected from the Guadalupe Mountains of Texas, and it resolves with other samples (*Q. mohriana* Buckley ex. Rydb. and *Q. vaseyana* Buckley) that were collected from the same geographic region in southwest Texas.

The RADseq ML tree (Fig. 3, left) produced by the RADseq matrix shows a clear split between sections *Quercus* and *Protobalanus* (100% bootstrap support for both clades). Within the white oaks (section *Quercus*), the California oaks are sister to the rest of the clade (bootstrap = 100%). The Eurasian oaks form a clade sister to all white oaks except for the California clade (bootstrap = 100%). All taxa but *Q. macrocarpa* Michx., *Q. stellata* Wangenh., *Q. muehlenbergii*, and *Q. chrysolepis* Liebm. are monophyletic, with 100% bootstrap support for the bipartition grouping accessions representing each species. With the exception of the placement of the Eurasian white oaks, which fall sister to a small clade of eastern North American white oaks in Hipp et al. (2014), the topology of the tree concurs with this previous skeletal phylogeny of the oaks based on RADseq data, and all clades are strongly supported.

Effect of geography and population divergence history on plastome phylogeny

At the broadest phylogenetic and geographic scale (all oaks), geography and phylogeny are both strong, significant partial predictors of plastome variation (geography partial correlation: $r^2 = 0.413$, $p < 0.002$; phylogeny partial correlation: $r^2 = 0.382$, $p < 0.002$; phylogeny simple correlation: $r^2 = 0.311$, $p < 0.002$; Table 2). In the North American oaks, geography decreases in

Fig. 3. Plastome phylogeny, all samples except outgroup, mapped to sample locations. Maximum likelihood plastome phylogeny (see Fig. 1), excluding the *Quercus rubra* outgroup. Colored blocks correspond to lineages in Fig. 1.



predictive power (r^2 = 0.028, p = 0.062; Table 2), but phylogeny remains a strongly significant predictor of plastome variation (r^2 = 0.407, p < 0.002; Table 2). The North American white oak dataset and the North American oak dataset without California each excludes one western North American clade (section *Protobalanus* and the California white oaks, respectively) that exhibits little hybridization with non-California taxa. The two exclusion sets, however, exhibit different patterns of plastome prediction. Exclusion of section *Protobalanus* alone results in a dramatic decrease in the predictive power of phylogeny (r^2 = 0.001, p = 0.621; Table 2). However, exclusion of the California white oaks results in an increase in predictive power of phylogeny from the North American oaks (r^2 = 0.301), though both phylogenetic and geographic signal are significant (p < 0.002 for both). This result appears, however, to be due solely to the presence of the Mexican clade species *Q. engelmannii*, for its exclusion reduces partial correlations of phylogeny and geography to non-significance (Supplement S4[2]). At the finest phylogenetic scale (the eastern North American white oaks), geographic signal and phylogeny are both non-significant predictors of plastome variation (Table 2).

Discussion

Phylogenetic incongruence between nuclear and plastid genome genealogies have been documented for diverse taxa over decades, using a variety of gene marker systems (Maddison 1997; Wendel and Doyle 1998). One common outcome of these studies is that organelle genome variation often tracks geographic genetic structure, while the nuclear genome appears to track evolutionary relationships as predicted by morphology or ecology (Whittemore and Schaal 1991). Our study is unique in its attempt to partition the effects of phylogeny and geography on the plastome nucleotide variation at varying phylogenetic depths in a lineage that is known to experience contemporary hybridization. We show that increasing the depth of the phylogeny considered increases the partial effect of phylogeny on the plastome topology, and that at the continental scale in North America, geography outweighs phylogeny as a predictor of plastome topology. We find, moreover, that the strong phylogeographic structure found in previous studies of European oaks (e.g., Dumolin-Lapegue et al. 1997) is not demonstrated in the eastern North American oaks, the classic oak syngameon (Hardin 1975; Burger 1975; Van Valen 1976). The deepest unambiguous split in the nuclear phylogeny, the split between the white oaks (section *Quercus*) and the intermediate oaks (section *Protobalanus*), is the single most important factor explaining concordance between the plastome phylogeny and population divergence history as encoded in the RADseq data. Inclusion of that split with the California species, however, reduces the fit between the RADseq and the plastome data, presumably because historic hybridization between section *Protobalanus* and California white oaks (as evidenced by incongruence in the placement of *Q. engelmannii*) erodes phylogenetic signal. Thus, joint analysis of nuclear and plastome data captures the interaction among biogeography, ancestral lineage divergences, and contemporary gene flow in the oaks, but within major clades of the genus, combined analysis is a poor method of inferring phylogenetic history.

Phylogenetic analyses

The recovery of phylogenetic structure among major clades and geographic signal at fine scales in the plastome tree is consistent with past studies, which have also found non-monophyly of species but geographic structuring of populations at fine scales based on plastome sequences (Dumolin-Lapegue et al. 1997; Petit et al. 1993; Manos et al. 1999). The RADseq tree also returned expected relationships at fine interspecific levels, confirming relationships and monophyly of most species inspected (Pearse and Hipp 2009; Hipp et al. 2014). The paraphyly of *Q. stellata* in the RADseq tree derives, we suspect, from a relatively high rate of hybridization in this species (cf. Nixon and Muller 1997) that we do not investigate further in the current paper.

Effect of geography and population divergence history on plastome phylogeny

Hybridization in oaks is primarily limited to crosses within sections, with the exception of the single inferred case of historical

Table 2. Partial regressions of plastome variation on nuclear phylogeny and geographic distance, using Mantel permutations to assess significance.

Dataset	plast~nuc+geo	plast~nuc.geo	plast~geo.nuc	plast~geo	plast~nuc	nuc~geo
(1) All oaks sampled	$r^2 = 0.596$	$r^2 = 0.382$, $p < 0.002$	$r^2 = 0.413$, $p < 0.002$	$r^2 = 0.346$, $p < 0.002$	$r^2 = 0.311$, $p < 0.002$	$r^2 = 0.011$, $p = 0.460$
(2) North American oaks	$r^2 = 0.464$	$r^2 = 0.407$, $p < 0.002$	$r^2 = 0.028$, $p = 0.062$	$r^2 = 0.097$, $p < 0.002$	$r^2 = 0.449$, $p < 0.002$	$r^2 = 0.084$, $p < 0.002$
(3) North American white oaks	$r^2 = 0.116$	$r^2 = 0.001$, $p = 0.621$	$r^2 = 0.112$, $p < 0.002$	$r^2 = 0.115$, $p < 0.002$	$r^2 = 0.004$, $p = 0.517$	$r^2 = 0.077$, $p < 0.002$
(4) North American oaks without California white oaks	$r^2 = 0.526$	$r^2 = 0.301$, $p < 0.002$	$r^2 = 0.162$, $p < 0.002$	$r^2 = 0.322$, $p < 0.002$	$r^2 = 0.435$, $p < 0.002$	$r^2 = 0.204$, $p < 0.002$
(5) Eastern North American white oaks	$r^2 = 0.013$	$r^2 = 0.011$, $p = 0.188$	$r^2 = 0.003$, $p = 0.779$	$r^2 = 0.002$, $p = 0.817$	$r^2 = 0.011$, $p = 0.156$	$r^2 = 0.002$, $p = 0.588$

Note: Models tested are represented using "~" to indicate "is predicted by" and "." to indicate "partial regression", where the predictor before the "." is the partial coefficient being estimated, and the predictor after the "." is the covariate. Abbreviations: plast, the plastome phylogenetic pairwise distances; nuc, nuclear (RADseq) phylogeny pairwise distances; geo, geographic (great circle) pairwise distances. Shaded cells are significant at $p < 0.002$.

hybridization between western North American *Protobalanus* and *Quercus* first reported in Manos et al. (1999) involving the white oak *Q. engelmannii* and corroborated in our study. We therefore expect the plastome to track deeper (more ancient) phylogenetic relationships, as seen in our results by the predictive power of the RADseq tree on the plastome topology in the all oaks sampled dataset. At the scale of all oaks sampled, the plastome topology is significantly predicted by geography in both simple and partial regressions, with the partial correlation coefficient (r^2 = 0.407) slightly larger than the simple correlation coefficient (r^2 = 0.338). This increase of 7% in variance explained from the simple to partial regression may reflect the fact that at the broadest phylogenetic scale in this study, there is substantial geographic structure embedded within the phylogenetic structure, and that a portion of this variation is only captured when phylogeny is controlled for using partial regression.

Removing Eurasian white oaks from this analysis has minimal effect on the proportion of plastome variation explained by the RADseq phylogeny (from r^2 = 0.382 to 0.407). The phylogenetic placement of Eurasian white oaks has varied among published studies, with the group resolving as monophyletic sister to a subset of eastern North American white oaks (RADseq nuclear markers; Hipp et al. 2014; McVay et al. 2017*b*), weakly monophyletic within the white oaks (six nuclear genes; Hubert et al. 2014), or strongly monophyletic and sister to the remainder of the white oaks (AFLP markers; Pearse and Hipp 2009; also plastid genomes in this study). The plastid genome and the nuclear genome consistently recover the Eurasian white oaks as a strongly supported clade; it is their placement within *Quercus* section *Quercus* that is uncertain in our study, and the deep divergence between the Eurasian white oaks and their sister taxa in the plastid genome tree (e.g., Fig. 1) contrast the more recent divergences indicated by the nuclear data (Fig. 2, left panel). As a consequence of these long branches, removing the Eurasian white oaks has the effect of focusing comparisons on the deepest divergences in both trees, such as the divergence of *Protobalanus* and the white oaks, and between Californian and Eastern North American oaks.

The deepest plastome divergences distinguish the two *Protobalanus* lineages, the Eurasian white oak lineage, and the North American white oak lineage. Thus, excluding both *Protobalanus* and the Eurasian white oaks from analysis removes the effect of nuclear phylogeny on the plastome topology (partial effect of RADseq variance on plastome variance decreases from r^2 = 0.382 to 0.001). Major clades are recoverable in the plastome tree because there are stronger reproductive barriers among sections than within sections and among geographic regions. Removing both the strongest geographic outliers (viz., the Eurasian white oaks) and the members of the section *Protobalanus* restricts comparisons to a single major clade on a single continent. The plastome divergence between the California and eastern North American white oaks, which presumably have a very limited history of gene flow, is not nearly as deep as the break with either *Protobalanus* or the Eurasian white oaks, perhaps reflecting a history of gene flow among clades that postdates initial divergence. Thus the North American oaks, even with California included, shows much weaker correlation between the plastome topology and the RADseq phylogeny than the datasets that include the deepest plastome divergences.

An interesting finding from our genome-scale survey is evidence for two comparatively ancient plastome lineages in section *Protobalanus*. This recapitulates a finding of Manos et al. (1999) with far greater genome-scale resolution, albeit with shallower taxon sampling. In the prior study, members of *Protobalanus* formed two monophyletic lineages that were incompletely resolved in a basal polytomy. Sampling of *Protobalanus* in Manos et al. (1999) was sufficient to demonstrate that two divergent plastome types were segregating in at least two of the sampled species (*Q. chrysolepis*, *Q. palmeri* Engelm.). This unusual level of polymorphism may reflect a very ancient gene exchange event between

Protobalanus and a white oak lineage, or lineage sorting of ancient haplotypes: in our study, the divergence between *Protobalanus* plastomes corresponds to the crown node of the entire ingroup, and in Manos et al. (1999, pg. 337), the two plastome clades differed by 13 restriction sites. The origin and taxonomic distribution of these two ancient plastome clades bears further study. Our plastome tree is also congruent with Manos et al. (1999) in placing *Q. engelmannii* Greene, a California white oak, in the same plastome group as *Q. tomentella* Engelm. Given the crown depth of each *Protobalanus* plastome clade (much younger than the divergence time between the *Protobalanus* plastomes), this placement almost certainly reflects more recent gene flow between white oaks and intermediate oaks. Stronger within-species sampling in a geographic context within the intermediate and California white oaks will be needed to evaluate how prevalent gene flow is between these sections.

Our analysis of the eastern North American white oaks contrasts sharply with previous studies on the European oaks (e.g., Dumolin-Lapegue et al. 1997), which have found strong phylogeographic signal. In the eastern North American oaks, plastome variation correlates poorly with both geographic and phylogeny (Table 2), unlike at the largest phylogenetic scales, where both geographic and phylogenetic signal are strong predictors because of the geographic structure embedded in phylogenetic signal (congruence of the two signals). In this classic eastern North American oak syngameon, it is remarkable that we do not find consistent geographic clustering across deep splits in the plastome tree within North America (Fig. 3). While our sampling should be sufficient to detect, for example, the Mississippi River discontinuity exhibited by numerous animals and a few plant species (Soltis et al. 2006), no node on the plastome tree supports a clean east–west phylogeographic split. The capacity for long-distance gene flow in oaks and their large population sizes may blur any signal of such phylogeographic breaks in eastern North America, where Pleistocene refugia were likely less discrete than in Europe.

In this study, we quantified the proportion of evolutionary history explained by plastome variation and contemporary geographic distributions. It is important to note, however, that for oaks and other temperate zone trees, current geographic distributions can be accurately modeled as a response to climatic variables (Rehfeldt et al. 2006; Roberts and Hamann 2012). This suggests that the correlations we observe between phylogeny and geography most likely have their origins in both adaptation to climate and the biogeography of diversification. The distributions of oaks and other temperate zone tree species have been in a continuous state of flux during Pleistocene-era climate change (Hewitt 1996; Roberts and Hamann 2015), with some species moving continental-scale distances of latitude and longitude (e.g., *Q. alba*, Gugger et al. 2013; *Q. petraea* and *Q. robur*, Taberlet et al. 1998), and others migrating shorter geographic distances across elevational gradients (e.g., *Q. garryana*, Crookston 2016; *Q. lobata*, Gugger et al. 2013). This climate-driven migratory history is likely to be a major source of the "geographic signal" and "phylogenetic noise" we observe in eastern North American white oaks. Geographic signals may be enhanced via changes in population sizes (e.g., contraction during glacial maxima; expansion during interglacial episodes) or the exchange of haplotypes via rare hybridization events between species as they migrate into and out of common refugia (e.g., Taberlet et al. 1998). Climate envelope modelling and paleoclimate reconstructions are now being integrated to understand the role of past climates on contemporary species diversity (Roberts and Hamann 2015), and such analyses may help to dissect the individual contributions of phylogeny, geography, and climate to contemporary genetic variation in such complex groups as the white oaks.

Finally, our findings have relevance to ongoing efforts to identify DNA barcodes (Hollingsworth et al. 2016), informative genetic markers that are diagnostic for oak species and related hybrids. Owing to the ecological and economic importance of *Quercus* globally, there have been a few attempts to develop simple diagnostic markers for oak species identification for use in ecological and landscape inventories (Simeone et al. 2012), conservation management, and screening oak wood products suspected as being derived from exploited sources (Schroeder et al. 2016). For example, *Q. mongolica* is listed as a CITES Appendix III species owing to pressure from illegal logging in the far-east of the Russian Federation (U.S. Department of Justice 2015); this status restricts trade of *Q. mongolica*, and brings all commercial species of white oak under greater scrutiny for labeling accuracy. Owing to their high similarity, *Q. mongolica* cannot be discriminated from other white oak species by wood anatomy, leaving DNA-based detection as one of the few available tools for identifying the taxonomic source of wood and enforcing illegal logging laws (Dormontt et al. 2015).

Our survey of 91 complete plastid genomes (Fig. 1) shows that organelle genomes are frequently shared across diverse evolutionary lineages and species. Indeed, of the 15 species represented by multiple sequences in our study, plastome monophyly was only observed in one case (*Q. gambelii*), and this seems likely to be a consequence of limited sampling. This result, combined with other barcoding efforts in *Quercus* (Piredda et al. 2010; Simeone et al. 2012), suggests that plastid genome sequences—either as subsets of barcode genes, or as the complete genomic "ultrabarcode" (Coissac et al. 2016)—are highly unlikely to offer the precision required for confident identification of source species, unless they are accompanied by additional information from the nuclear genome. In this regard, oak joins a growing list of woody species that cannot be "barcoded" with a single genome (Arca et al. 2012; Clement and Donoghue 2012; Parks et al. 2012; Percy et al. 2014) and that will require additional information from multiple genomes and multiple loci for accurate species identification. Our study does, however, confirm that a substantial proportion of variation in the plastome can be explained by geography (e.g., Table 2; Figs. 1, 2). This should make it possible to develop markers and databases that can address questions of geographic origin, as has recently been accomplished for the Asian white oaks *Q. mongolica* and *Q. dentata* (Schroeder et al. 2016).

Acknowledgements

This study was funded by National Science Foundation awards to A.L.H. (Awards 1146488, 1255901) and P.S.M. (Award 1146102); USDA Forest Service International Programs Office (R.C.C.); the German Federal Ministry of Food and Agriculture and the Deutsche Bundesstiftung Umwelt; and the US Agency for International Development and US Department of State (R.C.C.). We especially acknowledge Marlene Hahn (The Morton Arboretum, Lisle, IL, USA) for sample preparation and curation of metadata and specimens for the RADseq dataset; Tara Jennings (Pacific Northwest Research Station, Corvallis, OR, USA) for sample preparation and sequence assembly of the plastid genome dataset; and Mark Dasenko, Matthew Peterson, and Chris Sullivan (Oregon State University Center for Genome Research and Biocomputing, Corvallis, OR, USA) for assistance with sequencing. We thank Brad Kinder, Alex Moad, and Shelley Gardner (US Forest Service, Washington D.C.) for project coordination to obtain Asian oak samples; Bernd Degen (Johann Heinrich von Thünen Institut, Grosshansdorf, Germany) and Yulai Yanbaev (Bashkir State University, Ufa, Russia) for coordinating Russian oak collections; and a long list of collaborators who provided specimens for this study (identified in supplement). Finally, we thank Aaron Liston (Oregon State University) and others for their input on this project and the resulting manuscript, and the editors and reviewers of *Genome* for their insightful comments on the first version of this paper.

References

Alexander, L.W., and Woeste, K.E. 2014. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. Tree Genet. Genomes, **10**(4): 803–812. doi:10.1007/s11295-013-0681-1.

Arca, M., Hinsinger, D.D., Cruaud, C., Tillier, A., Bousquet, J., and Frascaria-Lacoste, N. 2012. Deciduous trees and the application of universal DNA barcodes: A case study on the circumpolar *Fraxinus*. PLoS ONE, **7**:e34089. doi:10.1371/journal.pone.0034089. PMID:22479532.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., et al. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE, **3**(10): e3376. doi:10.1371/journal.pone.0003376. PMID:18852878.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics, **30**: 2114–2120. doi:10.1093/bioinformatics/btu170. PMID:24695404.

Burger, W.C. 1975. The species concept in *Quercus*. Taxon, **24**: 45–50. doi:10.2307/1218998.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17**: 540–552. doi:10.1093/oxfordjournals.molbev.a026334. PMID:10742046.

Cavender-Bares, J., and Pahlich, A. 2009. Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). Am. J. Bot. **96**: 1690–1702. doi:10.3732/ajb.0800315. PMID:21622355.

Cavender-Bares, J., Gonzalez-Rodriguez, A., Eaton, D.A.R., Hipp, A., Buelke, A., and Manos, P. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): A genomic and population genetic approach. Mol. Ecol. **24**: 3668–3687. doi:10.1111/mec.13269. PMID:26095958.

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., et al. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. **80**(3): 528–580. doi:10.2307/2399846.

Clement, W.L., and Donoghue, M.J. 2012. Barcoding success as a function of phylogenetic relatedness in *Viburnum*, a clade of woody angiosperms. BMC Evol. Biol. **12**: 73. doi:10.1186/1471-2148-12-73. PMID:22646220.

Coissac, E., Hollingsworth, P.M., Lavergne, S., and Taberlet, P. 2016. From barcodes to genomes: extending the concept of DNA barcoding. Mol. Ecol. **25**: 1423–1428. doi:10.1111/mec.13549. PMID:26821259.

Crookston, N. 2016. Research on forest climate change: Potential effects of global warming on forests and plant climate relationships in Western North America and Mexico. [Online.] Available from http://charcoal.cnre.vt.edu/climate/ [accessed: 10/17/2016].

Diniz-Filho, J.A.F., Soares, T.N., Lima, J.S., Dobrovolski, R., Landeiro, V.L., de Campos Telles, M.P., et al. 2013. Mantel test in population genetics. Genet. Mol. Biol. **36**(4): 475–485. doi:10.1590/S1415-47572013000400002. PMID:24385847.

Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., et al. 2015. Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biol. Conserv. **191**: 790–798. doi:10.1016/j.biocon.2015.06.038.

Dumolin-Lapegue, S., Demesure, B., Fineschi, S., Le Corre, V., and Petit, R.J. 1997. Phylogeographic structure of white oaks throughout the European continent. Genetics, **146**: 1475–1487. PMID:9258689.

Eaton, D.A.R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics, **30**(13): 1844–1849. doi:10.1093/bioinformatics/btu121. PMID:24603985.

Eaton, D.A.R., Hipp, A.L., González-Rodríguez, A., and Cavender-Bares, J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution, **69**: 2587–2601. doi:10.1111/evo.12758. PMID:26299374.

Folk, R.A., Mandel, J.R., and Freudenstein, J.V. 2016. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. Syst. Biol. **66**(3): 320–337. doi:10.1093/sysbio/syw083. PMID:27637567.

Gerber, S., Chadœuf, J., Gugerli, F., Lascoux, M., Buiteveld, J., Cottrell, J., et al. 2014. High rates of gene flow by pollen and seed in oak populations across Europe. PLoS ONE, **9**: e85130. doi:10.1371/journal.pone.0085130. PMID:24454802.

Gugger, P.F., Ikegami, M., and Sork, V.L. 2013. Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Nee. Mol. Ecol. **22**: 3598–3612. doi:10.1111/mec.12317. PMID:23802553.

Hamzeh, M., and Dayanandan, S. 2004. Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast *trn*T-*trn*F region and nuclear rDNA. Am. J. Bot. **91**(9): 1398–1408. doi:10.3732/ajb.91.9.1398. PMID:21652373.

Hardin, J.W. 1975. Hybridization and introgression in *Quercus alba*. J. Arnold Arboretum, **56**: 336–363.

Hewitt, G.M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. **58**: 247–276. doi:10.1006/bijl.1996.0035.

Hipp, A.L. 2014. RADami: R Package for Phylogenetic Analysis of RADseq Data. R package version 1.0-3. Available from https://cran.r-project.org/web/packages/RADami/index.html [accessed 5 July 2016].

Hipp, A.L. 2015. Should hybridization make us skeptical of the oak phylogeny? International Oaks, **26**: 9–18.

Hipp, A.L., and Weber, J.A. 2008. Taxonomy of Hill's oak (*Quercus ellipsoidalis* E.J. Hill): evidence from AFLP data. Syst. Bot. **33**: 148–158. doi:10.1600/036364408783887320.

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., and Manos, P.S. 2014. A framework phylogeny of the American oak clade based on sequenced RAD Data. PLoS ONE, **9**(4): e93975. doi:10.1371/journal.pone.0093975. PMID:24705617.

Hollingsworth, P.M., Li, D.-Z., van der Bank, M., and Twyford, A.D. 2016. Telling plant species apart with DNA: from barcodes to genomes. Philos. Trans. R. Soc. B Biol. Sci. **371**: 20150338. doi:10.1098/rstb.2015.0338.

Hubert, F., Grimm, G.W., Jousselin, E., Berry, V., Franc, A., and Kremer, A. 2014. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. Syst. Biodivers. **12**(4): 405–423. doi:10.1080/14772000.2014.941037.

Katoh, K., and Standley, D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. **30**(4): 772–780. doi:10.1093/molbev/mst010. PMID:23329690.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, **28**(12): 1647–1649. doi:10.1093/bioinformatics/bts199. PMID:22543367.

Langmead, B., and Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods, **9**: 357–359. doi:10.1038/nmeth.1923. PMID:22388286.

Lynch, M. 2008. Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol. Biol. Evol. **25**(11): 2409–2419. doi:10.1093/molbev/msn185. PMID:18725384.

Maddison, W.P. 1997. Gene trees in species trees. Syst. Biol. **46**: 23–536. doi:10.1093/sysbio/46.3.523.

Manos, P.S., Doyle, J.J., and Nixon, K.C. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Mol. Phylogenet. Evol. **12**(3): 333–349. doi:10.1006/mpev.1999.0614. PMID:10413627.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27**(2): 209–220. PMID:6018555.

McVay, J.D., Hauser, D., Hipp, A.L., and Manos, P.S. 2017a. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in Western North America. Genome, **60**(This issue): doi:10.1139/gen-2016-0206.

McVay, J.D., Hipp, A.L., and Manos, P.S. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proc. R. Soc. B, **284**(1854): 20170300. doi:10.1098/rspb.2017.0300. PMID:28515204.

Morgan, M., Anders, S., Lawrence, M., Aboyoun, P., Pagès, H., and Gentleman, R. 2009. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics, **25**: 2607–2608. doi:10.1093/bioinformatics/btp450. PMID:19654119.

Muir, G., Fleming, C.C., and Schlötterer, C. 2000. Species status of hybridizing oaks. Nature (Lond.), **405**: 1016. doi:10.1038/35016640.

Muller, C.H. 1952. Ecological control of hybridization in *Quercus*: a factor in the mechanism of evolution. Evolution, **6**(2): 147–161. doi:10.2307/2405620.

Nixon, K.C., and Muller, C.H. 1997. *Quercus* Linnaeus sect. *Quercus*. *In* Flora of North America, Vol. 3. *Edited by* Flora of North America Editorial Committee. Oxford University Press, New York. pp. 471–506.

Oh, S.-H., and Manos, P.S. 2008. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. Taxon, **57**: 434–451.

Paradis, E., Claude, J., and Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics, **20**: 289–290. doi:10.1093/bioinformatics/btg412. PMID:14734327.

Parks, M., Cronn, R., and Liston, A. 2012. Separating the wheat from the chaff: mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (Pinaceae). BMC Evol. Biol. **12**: 100. doi:10.1186/1471-2148-12-100. PMID:22731878.

Pearse, I.S., and Hipp, A.L. 2009. Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. Proc. Natl. Acad. Sci. U.S.A. **106**(43): 18097–18102. doi:10.1073/pnas.0904867106. PMID:19841257.

Peñaloza-Ramírez, J.M., González-Rodríguez, A., Mendoza-Cuenca, L., Caron, H., Kremer, A., and Oyama, K. 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. Ann. Bot. **105**: 389–399. doi:10.1093/aob/mcp301. PMID:20056653.

Percy, D., Argus, G.W., Cronk, Q.C., Fazekas, A.J., Kesanakurti, P.R., Burgess, K.S., et al. 2014. Understanding the spectacular failure of DNA barcoding in willows (*Salix*): Does this result from a trans-specific selective sweep? Mol. Ecol. **23**: 4737–4756. doi:10.1111/mec.12837. PMID:24944007.

Petit, R.J., and Excoffier, L. 2009. Gene flow and species delimitation. Trends Ecol. Evol. **24**(7): 386–393. doi:10.1016/j.tree.2009.02.011. PMID:19409650.

Petit, R.J., Wagner, D.B., and Kremer, A. 1993. Ribosomal DNA and chloroplast DNA polymorphisms in a mixed stand of *Quercus robur* and *Q. petraea*. Ann. For. Sci. **50**: 41s–47s. doi:10.1051/forest:19930703.

Piredda, R., Simeone, M.C., Attimonelli, M., Bellarosa, R., and Schirone, B. 2010. Prospects of barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species identity. Mol. Ecol. Res. **11**: 72–83. doi:10.1111/j.1755-0998.2010.02900.x.

R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/ [accessed 19 October 2016].

Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., and Evans, J.S. 2006. Empirical analyses of plant-climate relationships for the western United States. Int. J. Plant Sci. **167**: 1123–1150. doi:10.1086/507711.

Revell, L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. **3**: 217–223. doi:10.1111/j.2041-210X.2011.00169.x.

Rieseberg, L.H., and Soltis, D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. Evol. Trends Plants, **5**(1): 65–84.

Rieseberg, L.H., Beckstrom-Sternberg, S.M., Liston, A., and Arias, D.M. 1991. Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* (Asteraceae). Syst. Bot. **16**: 50–76. doi:10.2307/2418973.

Roberts, D.R., and Hamann, A. 2012. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. Global Ecol. Biogeogr. **21**: 121–133. doi:10.1111/j.1466-8238.2011.00657.x.

Roberts, D.R., and Hamann, A. 2015. Glacial refugia and modern genetic diversity of 22 western North American tree species. Proc. R. Soc. B Biol. Sci. **282**: 2903. doi:10.1098/rspb.2014.2903.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ Preprints, **4**: e2409v1. Available from https://doi.org/10.7287/peerj.preprints.2409v1 [accessed 19 October 2016].

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. 2014. From algae to angiosperms—inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. **14**: 23. doi:10.1186/1471-2148-14-23. PMID:24533922.

Sang, T., Crawford, D.J., and Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogography of *Paeonia* (Paeoniaceae). Am. J. Bot. **84**: 1120–1136. doi:10.2307/2446155. PMID:21708667.

Schroeder, H., Cronn, R., Yanbaev, Y., Jennings, T., Mader, M., Degen, B., and Kersten, B. 2016. Development of molecular markers for determining continental origin of wood from white oaks (*Quercus* L. sect. *Quercus*). PLoS ONE, **11**(6): e0158221. doi:10.1371/journal.pone.0158221. PMID:27352242.

Shaw, J., Craddock, J.H., and Binkley, M.A. 2012. Phylogeny and phylogeography of North American *Castanea* Mill. (Fagaceae) using cpDNA suggests gene sharing in the southern Appalachians (*Castanea* Mill., Fagaceae). Castanea, **77**(2): 186–211. doi:10.2179/11-033.

Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., and Morris, A.B. 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am. J. Bot. **101**(11): 1987–2004. doi:10.3732/ajb.1400398. PMID:25366863.

Simeone, M.C., Piredda, R., Papini, A., Vessella, F., and Schirone, B. 2012. Application of plastid and nuclear markers to DNA barcoding of Euro-Mediterranean oaks (*Quercus*, Fagaceae): Problems, prospects and phylogenetic implications. Bot. J. Linn. Soc. **172**: 478–499. doi:10.1111/boj.12059.

Simeone, M.C., Grimm, G.W., Papini, A., Vessella, F., Cardoni, S., Tordoni, E., et al. 2016. Plastome data reveal multiple geographic origins of *Quercus* Group *Ilex*. PeerJ, **4**: e1897. doi:10.7717/peerj.1897. PMID:27123376.

Smouse, P.E., Long, J.C., and Sokal, R.R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35**(4): 627–632. doi:10.2307/2413122.

Soltis, D.E., Morris, A.B., McLachlan, J.S., Manos, P.S., and Soltis, P.S. 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. **15**: 4261–4293. doi:10.1111/j.1365-294X.2006.03061.x. PMID:17107465.

Sork, V.L., Riordin, E., Gugger, P.F., Fitz-Gibbon, S., Wei, X., and Ortego, J. 2016. Phylogeny and introgression of California scrub white oaks (*Quercus* section *Quercus*). International Oaks, **27**: 61–74.

Stamatakis, A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, **30**(9): 1312–1313. doi:10.1093/bioinformatics/btu033. PMID:24451623.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., and Cosson, J.-F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. **7**: 453–464. doi:10.1046/j.1365-294x.1998.00289.x. PMID:9628000.

U.S. Department of Justice. 2015. Statement of facts, United States of American v. Lumber Liquidators. Case document 2:15-cr-00126-RAJ-LRL (2015). Available from http://www.justice.gov/opa/file/787141/download.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. Taxon, **25**: 233–239. doi:10.2307/1219444.

Via, S., and West, J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. **17**(19): 4334–4345. doi:10.1111/j.1365-294X.2008.03921.x. PMID:18986504.

Wendel, J.F., and Doyle, J. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. *In* Molecular systematics of plants II: DNA sequencing. *Edited by* D. Soltis, P. Soltis, and J. Doyle. Kluwer, Boston, Mass. pp. 265–296.

Whittemore, A.T., and Schaal, B.A. 1991. Interspecific gene flow in sympatric oaks. Proc. Natl. Acad. Sci. U.S.A. **88**: 2540–2544. doi:10.1073/pnas.88.6.2540. PMID:11607170.

Wiegand, K.M. 1935. A taxonomist's experience with hybrids in the wild. Science, **81**: 161–166. doi:10.1126/science.81.2094.161. PMID:17817007.

This article has been cited by:

1. Min Deng, Xiao-Long Jiang, Andrew L. Hipp, Paul S. Manos, Marlene Hahn. 2018. Phylogeny and biogeography of East Asian evergreen oaks (Quercus section Cyclobalanopsis ; Fagaceae): Insights into the Cenozoic history of evergreen broad-leaved forests in subtropical Asia. *Molecular Phylogenetics and Evolution* **119**, 170-181. [Crossref]
2. Andrew L. Hipp, Paul S. Manos, Antonio González-Rodríguez, Marlene Hahn, Matthew Kaproth, John D. McVay, Susana Valencia Avalos, Jeannine Cavender-Bares. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist* **217**:1, 439-452. [Crossref]
3. Andrew L. Hipp, Santiago C. Gonzalez-Martinez, Juan P. Jaramillo-Correa. 2017. The Evolution of Tree Diversity: Proceedings of the 2016 IUFRO Genomics and Forest Tree Genetics Conference, Phylogenetics and Genomic Evolution Session, Arcachon, France. *Genome* **60**:9, v-vi. [Citation] [Full Text] [PDF] [PDF Plus]
4. Thomas Denk, Guido W. Grimm, Paul S. Manos, Min Deng, Andrew L. Hipp. An Updated Infrageneric Classification of the Oaks: Review of Previous Taxonomic Schemes and Synthesis of Evolutionary Patterns 13-38. [Crossref]