GENOMIC IDENTITY OF WHITE OAK SPECIES IN AN EASTERN NORTH AMERICAN SYNGAMEON[1]

Andrew L. Hipp,[2,3] Alan T. Whittemore,[4] Mira Garner,[2] Marlene Hahn,[2] Elisabeth Fitzek,[5] Erwan Guichoux,[6] Jeannine Cavender-Bares,[7] Paul F. Gugger,[8] Paul S. Manos,[9] Ian S. Pearse,[10] and Charles H. Cannon[2]*

Abstract

The eastern North American white oaks, a complex of approximately 16 potentially interbreeding species, have become a classic model for studying the genetic nature of species in a syngameon. Genetic work over the past two decades has demonstrated the reality of oak species, but gene flow between sympatric oaks raises the question of whether there are conserved regions of the genome that define oak species. Does gene flow homogenize the entire genome? Do the regions of the genome that distinguish a species in one part of its range differ from the regions that distinguish it in other parts of its range, where it grows in sympatry with different species? Or are there regions of the genome that are relatively conserved across species ranges? In this study, we revisit seven species of the eastern North American white oak syngameon using a set of 80 single-nucleotide polymorphisms (SNPs) selected in a previous study because they show differences among, and consistency within, the species. We test the hypothesis that there exist segments of the genome that do not become homogenized by repeated introgression, but retain distinct alleles characteristic of each species. We undertake a range-wide sampling to investigate whether SNPs that appeared to be fixed based on a relatively small sample in our previous work are fixed or nearly fixed across the range of the species. Each of the seven species remains genetically distinct across its range, given our diagnostic set of markers, with relatively few individuals exhibiting admixture of multiple species. SNPs map back to all 12 *Quercus* linkage groups (chromosomes) and are separated from each other by an average of 7.47 million bp (± 8.74 million bp, SD), but are significantly clustered relative to a random null distribution, suggesting that our SNP toolkit reflects genome-wide patterns of divergence while potentially being concentrated in regions of the genome that reflect a higher-than-average history of among-species divergence. This application of a DNA toolkit designed for the simple problem of identifying species in the field has two important implications. First, the eastern North American white oak syngameon is composed of entities that most taxonomists would consider "good species." Second, and more fundamentally, species in the syngameon are genetically coherent because characteristic portions of the genome remain divergent despite a history of introgression. Understanding the conditions under which some loci diverge while others introgress is key to understanding the origins and maintenance of global tree diversity.

Key words: Cohesion species, DNA genotyping toolkit, hybridization, introgression, *Quercus*, single-nucleotide polymorphism (SNP), syngameon.

[1] This paper is based on a talk presented by the first author at the Missouri Botanical Garden's 65th Annual Fall Symposium. Portions of the introduction and conclusion are adapted from an essay published by the first author in the *Journal of International Oaks* (Hipp, 2015). The manuscript benefited from comments by Thibault Leroy. Data and analyses presented here have not been published elsewhere. The majority of collections and all of M. G.'s time were funded by USDA Agreement Number 58-8020-5-005, project number 8020-21000-070-03S, to A. L. H. and A. T. W. Leaf silhouettes were digitized in part by M. Kaproth. SNP data were generated with financial support of The Morton Arboretum's Center for Tree Science. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

[2] The Morton Arboretum, Herbarium, 4100 Illinois Route 53, Lisle, Illinois 60532, U.S.A.

[3] Integrative Research Center, The Field Museum, 1400 S Lake Shore Drive, Chicago, Illinois 60605, U.S.A.

[4] U.S. National Arboretum, 3501 New York Avenue NE, Washington, D.C. 20002, U.S.A.

[5] Universität Bielefeld, Department of Computational Biology, Center for Biotechnology - CeBiTec, Universitaetsstrasse 27, 33615 Bielefeld, Germany.

[6] BIOGECO, INRA, Université Bordeaux, 33610 Cestas, France.

[7] Department of Ecology, Evolution and Behavior, University of Minnesota, 1479 Gortner Avenue, Saint Paul, Minnesota 55108, U.S.A.

[8] University of Maryland Center for Environmental Science, Appalachian Laboratory, 301 Braddock Road, Frostburg, Maryland 21532, U.S.A.

[9] Department of Biology, Duke University, 130 Science Drive, Durham, North Carolina 27708, U.S.A.

[10] U.S. Geological Survey, Fort Collins Science Center, 2150 Centre Avenue, Fort Collins, Colorado 80526, U.S.A.

* Author for correspondence: ahipp@mortonarb.org

Version of record first published online on 5 September 2019 ahead of Fall 2019 issue.
doi: 10.3417/2019434

Hybridization and introgressive gene flow in oaks have long suggested the question of what constitutes an oak species. The 1867 edition of Gray's *Manual of the Botany of the Northern United States* (Gray, 1867), for example, reports five hybrids in oaks,[1] and Wiegand (1935) notes that in this edition, "we find hybrids scarcely mentioned except in one genus, *Quercus*." In the early 20th century, studies of character segregation in first- and second-generation oak hybrids suggested that adaptive gene flow might contribute to range extensions in the southern live oak *Q. virginiana* Mill. (Ness, 1918; Allard, 1932; Yarnell & Palmer, 1933). The roughly 100 years following Gray's 1867 edition saw a number of seminal papers, mostly dealing with the effects of interspecific hybridization on oak species origins, coherence, and evolutionary trajectories (e.g., Engelmann, 1876; Palmer, 1948; Muller, 1952).

In the mid-1970s, a trio of now-classic papers that focused on the eastern North American white oak syngameon set the stage for contemporary studies of oak species coherence. In 1975, James Hardin published an article in the *Journal of the Arnold Arboretum* reporting evidence of widespread gene flow among 16 white oaks of eastern North America (Hardin, 1975). At about the same time, a pair of articles in *Taxon* argued that gene flow in oaks is dominated by localized gene flow among individuals that are closely enough related to exchange genes, irrespective of species, rather than among populations within species (Burger, 1975; Van Valen, 1976). Because of ongoing gene flow and introgression, Burger and Van Valen argued, oak species cannot be defined by reproductive isolation. Rather, oak species represent ecologically discrete lineages with distinct evolutionary trajectories. "Species," Van Valen wrote, "are maintained for the most part ecologically, not reproductively." He and Burger both argued that local gene flow among sympatric populations of different species may exceed gene flow between geographically distant populations of single species, and that the capacity for interbreeding cannot therefore be the criterion by which we recognize oak species. Burger went so far as to suggest erecting subgenera or sections that are equivalent to reproductive species, but allowing our named species in oaks to represent ecologically and morphologically defined evolutionary lineages. The idea that gene flow is often insufficient to cause species to cohere across their range had been discussed previously (Ehrlich & Raven, 1969), but Burger and Van Valen seem to be making a stronger claim: oak species are delimited not reproductively, but ecologically. A measured skepticism about oak species is not uncommon among botanists even today, unsurprising in the face of ample evidence of introgression and gene flow (e.g., Whittemore & Schaal, 1991; Dumolin-Lapegue et al., 1997; Dumolin-Lapegue & Petit, 1999; Petit et al., 2003; Dodd & Afzal-Rafii, 2004; Tovar-Sánchez & Oyama, 2004; Craft & Ashley, 2006; Lexer et al., 2006; Curtu et al., 2007; Hipp & Weber, 2008; Chybicki & Burczyk, 2010; Moran et al., 2012).

In the past two decades, the increased availability of single-locus DNA markers has stimulated investigation into the processes that maintain distinct species in the presence of interspecific hybridization (Kremer & Hipp, in press). It is notable that different studies using single-locus DNA markers have shown strikingly different patterns. Studies utilizing chloroplast DNA markers have generally yielded clear evidence of introgressive exchange of markers, with little if any clustering of individuals by species (Whittemore & Schaal, 1991; Dumolin-Lapegue et al., 1997, 1999; Petit et al., 1997, 2003; Manos et al., 1999; Belahbib et al., 2001; Pham et al., 2017). Studies utilizing nuclear markers, on the other hand, have typically demonstrated that gene flow among species (Dodd & Afzal-Rafii, 2004; Gömöry & Schmidtová, 2007; de Casas et al., 2007; Eaton et al., 2015) is balanced by gene flow within species, promoting species cohesion (Whittemore & Schaal, 1991; Muir et al., 2000; Muir & Schlötterer, 2005; Lexer et al., 2006; Leroy et al., 2017, 2018).

Next-generation DNA sequencing (NGS) has made it practical to test more rigorous models of introgression history in oaks using much larger numbers of loci (e.g., Eaton et al., 2015; Leroy et al., 2017). Additionally, NGS has enabled economical development of genotyping toolkits for smaller applications. In a recent paper, we utilized a large RAD-seq dataset for white oaks (McVay et al., 2017b; Hipp et al., 2018) to develop a low-cost single-nucleotide polymorphism (SNP) genotyping kit for eastern North American white oaks (Fitzek et al., 2018). We demonstrated our 80-marker SNP kit to be effective at identifying 15 species and F_1 hybrids and validated it in a garden setting, where we found hybridization between non-native species in the collection and the native white oaks of the surrounding woodlands.

[1] The history of Gray's reports of hybrids is instructive. The first edition (Gray & Sullivant, 1848) included two hybrids in the genus *Quercus*, each reported to be "founded on" a single tree or individual. In the 1857 through 1862 editions (Gray, 1857, 1859, 1862), this number increased to three, which Gray described as "the following remarkable forms, by some regarded as species." Gray's language changes between 1848 and 1862—years flanking the publication of *Origin of Species*—from suggesting that these hybrids are mere sports to suggesting that they might be species of hybrid origin. Gray was a great supporter of Darwin and had an avid correspondence with him even before publication of *Origin* (Browne, 2010), and Gray's change in language undoubtedly reflects a change in his view of the evolutionary implications of hybridization.

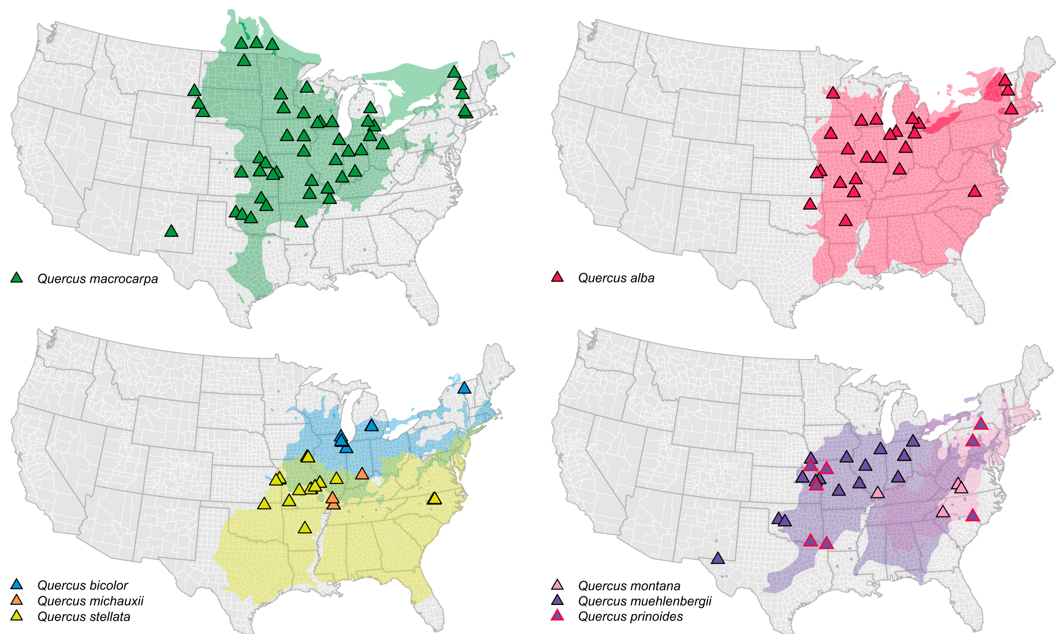

Figure 1. Sampling map. Sites were sampled to roughly cover the range of the taxa as known; on each panel, collections are overlaid on the range maps for each species following Little (1971, 1977, 1979), except for *Quercus prinoides* Willd., for which a base map was not available.

In the current study, we test this marker set in natural populations across a range-wide sample of seven eastern North American white oaks. These species are components of a classic syngameon, where there is good documentation of interspecific hybridization in many combinations (Hardin, 1975) and introgressive exchange of chloroplast haplotypes (Whittemore & Schaal, 1991; Pham et al., 2017). We investigate whether the species are genetically cohesive at these 80 loci or a subset thereof, representing areas of the genome that have presumably been shielded from introgression across the range of the species. We also map these markers back to a chromosome-level assembly of the *Quercus robur* L. genome (Plomion et al., 2018) to investigate whether they are distributed across the genome or, conversely, whether genetic cohesion of the eastern North American white oaks is concentrated in a few genomic islands of differentiation. Our study provides a first framework investigation of the eastern North American white oak syngameon using a genome-wide sample of molecular markers, laying the groundwork for future studies of introgression and species cohesion in the group.

Materials and Methods

Sampling and Genotyping

Data were initially collected from 184 individuals of seven eastern North American white oak species, encompassing a wide geographic range for each species. In this study, *Quercus muehlenbergii* Engelm. and *Q. prinoides* Willd. are separated in name only, as our RAD-seq data failed to distinguish the species (McVay et al., 2017b; Hipp et al., 2018), and consequently no SNPs were identified to distinguish them from each other (Fitzek et al., 2018). The species status of these two bears investigation with broader sampling. Throughout the remainder of this paper, we will refer to these two together as *Q. muehlenbergii/prinoides*, not because we are making a claim that they are not distinct taxonomically, but to reinforce that they are grouped for analysis. Samples represent unique adults with seven exceptions, for which a second extraction of each individual was genotyped as a technical replicate. Individuals were selected to be typical of the species morphologically, not to be a random sample of all potential pure and introgressed individuals. Twenty-one individuals for which fewer than 90% of loci amplified successfully were removed from analysis and are not discussed further in this paper, leaving a final set of 163 individuals analyzed (Fig. 1, Table 1).

To reduce the opportunity for hybridization with taxa from outside the natural range of each species, samples were preferentially selected from wild populations or from trees grown in gardens from seeds of known wild provenance (as discussed in Fitzek et al., 2018; Hipp et al., 2018); five individuals were analyzed from cultivated material (Table 1). Sample size per species ranges from seven to nine in *Quercus montana* Willd.

Table 1. Samples included in the study. Locality and coordinate data indicate source populations for both wild and cultivated material, where material is of a cultivated source and no state or county data are provided. Replicates indicate technical replicates extracted from the same individual: individuals with the same replicate code are identical.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE000321		unknown	unknown	*Quercus alba*	Marlene Hahn	CA-DAV-MH48	cultivated	cultivated	Royal Botanical Gardens, Hamilton, Ontario, Wentworth Co., W Flamboro Lap., *R. E. Halward s.n.*, 24 Sep. 1964	cultivated
QUE000128.a	** A **	41.8649	−86.3508	*Q. alba*	Marlene Hahn	IL-MOR-MH086	MI	Berrien	along St. Joseph River along River Trail on Fernwood Botanical Garden NW of Niles	wild
QUE000128.b	** A **	41.8649	−86.3508	*Q. alba*	Marlene Hahn	IL-MOR-MH086	MI	Berrien	along St. Joseph River along River Trail on Fernwood Botanical Garden NW of Niles	wild
QUE000596		39.9348	−89.8016	*Q. alba*	Marlene Hahn	IL-SH-162	IL	Menard	StarHill Forest spontaneous, Petersburg	wild
QUE000151		45.3680	−93.2193	*Q. alba*	Carol DeVries	IL-MOR-MH109	MN	Anoka	on a farm	wild
QUE000700		36.0218	−79.0161	*Q. alba*	Paul Manos	PM-19	NC	Orange	near crossroad Cornwallis Rd. & Murphy School Rd.	wild
QUE001805		42.0421	−93.6057	*Q. alba*	Mira Garner	IA-MG-262	IA	Story	Ames, Veenker Memorial Golf Course (ISU)	wild
QUE001815		40.7063	−91.7939	*Q. alba*	Mira Garner	IA-MG-270	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001841		37.9732	−92.7623	*Q. alba*	Mira Garner	MO-MG327	MO	Camden	Ha Ha Tonka State Park	wild
QUE001918		34.8030	−92.3260	*Q. alba*	Mira Garner	AR-MG387	AR	Pulaski	Little Rock, Burns Park	wild
QUE001932		37.1573	−91.3650	*Q. alba*	Mira Garner	MO-MG401	MO	Shannon	Eminence, Buttin Rock Access, near trailer park	wild
QUE001884		36.2181	−95.9000	*Q. alba*	Mira Garner	OK-MG353	OK	Tulsa	Tulsa, Mohawk Park	wild
QUE002075		39.8393	−88.3677	*Q. alba*	Ian Pearse	Chickenbristle 4	IL	Douglas	property of Bob Pearse	wild
QUE002091		38.7363	−86.4143	*Q. alba*	Mira Garner	IN-MG612	IN	Lawrence	Spring Mill State Park, Trail 5	wild
QUE002108		40.4591	−85.5092	*Q. alba*	Mira Garner	IN-MG629	IN	Grant	Taylor Wilderness, Taylor University	wild
QUE002121		41.6574	−87.0605	*Q. alba*	Mira Garner	IN-MG642	IN	Porter	Indiana Dunes State Park	wild
QUE002130		42.0161	−73.3353	*Q. alba*	Paul Gugger	QUAL-1029	CT	Litchfield	1-19 Sand Rd, North Canaan, CT	wild
QUE002138		43.6047	−73.1804	*Q. alba*	Paul Gugger	QUAL-1037	VT	Rutland	D&H Trail, Castleton, VT	wild
QUE002155		44.4464	−73.2202	*Q. alba*	Paul Gugger	QUAL-1054	VT	Chittenden	1-225 Industrial Pkwy., Burlington, VT	wild
QUE002210		38.2172	−91.0864	*Q. alba*	Mira Garner	MO-MG586	MO	Crawford	Meramec State Park, Sullivan, MO	wild
QUE002253		38.9803	−94.8053	*Q. alba*	Mira Garner	KS-MG433	KS	Johnson	Shawnee Mission Park	wild
QUE002282		38.8096	−95.1927	*Q. alba*	Mira Garner	KS-MG462	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE002337		42.3047	−83.7508	*Q. alba*	Mira Garner	MI-MG674	MI	Washtenaw	Barton Nature Area, Ann Arbor, off trail	wild
QUE002355		42.7666	−84.3911	*Q. alba*	Mira Garner	MI-MG692	MI	Ingham	Lake Lansing Park, East Lansing, picnic area	wild
QUE002366		43.0168	−90.1142	*Q. alba*	Mira Garner	WI-MG703	WI	Iowa	Governor Dodge State Park, Dodgeville, Cox Hollow Campground	wild
QUE002399		43.0171	−88.4353	*Q. alba*	Mira Garner	WI-MG736	WI	Waukesha	University of Wisconsin, Waukesha Field Station, Ottawa	wild
QUE002493		41.5522	−84.3590	*Q. alba*	Mira Garner	OH-MG830	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE000643		41.1922	−87.4463	*Q. bicolor*	Marlene Hahn	IL-SH-030	IN	Lake	Mohawk Club, Schneider	wild
QUE000618		38.8931	−94.8322	*Q. bicolor*	Marlene Hahn	IL-SH-184	KS	Johnson		wild
QUE000136		41.7409	−87.8603	*Q. bicolor*	Marlene Hahn	IL-MOR-MH094	IL	Cook	along the Des Plaines River near Willow Springs	wild
QUE001813		40.7048	−91.7963	*Q. bicolor*	Mira Garner	IA-MG-268	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE002153		44.4001	−73.2375	*Q. bicolor*	Paul Gugger	QUBI-1052	VT	Chittenden	1136 Bay Rd., Shelburne, VT	wild
QUE002196		38.2267	−91.0830	*Q. bicolor*	Mira Garner	MO-MG572	MO	Crawford	Meramec State Park, Sullivan, MO, campground	wild
QUE002360		42.7951	−84.3927	*Q. bicolor*	Mira Garner	MI-MG697	MI	Ingham	Lake Lansing Park, East Lansing, edge of marsh	wild
QUE002361		42.7653	−84.3825	*Q. bicolor*	Mira Garner	MI-MG698	MI	Ingham	Lake Lansing Park, East Lansing, edge of marsh	wild
QUE002528		42.1843	−87.9163	*Q. bicolor*	Mira Garner	IL-MG865	IL	Lake	Ryerson Woods Conservation Area, trail behind cabins	wild
QUE002539		41.8245	−87.9333	*Q. bicolor*	Mira Garner	IL-MG876	IL	DuPage	Fullersburg Woods Nature Preserve	wild
QUE000671		40.4554	−86.9165	*Q. macrocarpa*	Bethany Hayward Brown	IL-SH-58	IN	Wabash	West Lafayette	wild
QUE000623		40.1923	−96.6650	*Q. macrocarpa*	Marlene Hahn	IL-SH-189	NE	Gage	Blue River, via NSA 2000	wild
QUE000619		36.6467	−89.3021	*Q. macrocarpa*	Marlene Hahn	IL-SH-185	MO	Mississippi	Big Oak Tree State Park	wild
QUE000640		38.3507	−87.8226	*Q. macrocarpa*	Marlene Hahn	IL-SH-027	IL	Wabash	Beall Woods State Park	wild
QUE000107		41.4868	−87.7998	*Q. macrocarpa*	Marlene Hahn	IL-MOR-MH003 (A/B)	IL	Cook	near Sauk Lake in Sauk Trail Forest Preserve	wild
QUE000617		48.3076	−98.7287	*Q. macrocarpa*	Marlene Hahn	IL-SH-183	ND	Ramsey		wild
QUE000673		39.7799	−96.0153	*Q. macrocarpa*	Bethany Hayward Brown	IL-SH-060	KS	Nemaha		wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE000622		35.6239	−99.0087	*Q. macrocarpa*	Marlene Hahn	IL-SH-188	OK	Custer		wild
QUE000672		45.5029	−104.4767	*Q. macrocarpa*	Bethany Hayward Brown	IL-SH-59	MT	Carter		wild
QUE000620		33.6067	−105.3631	*Q. macrocarpa*	Marlene Hahn	IL-SH-186	NM	Lincoln	Capitan Mtns.	wild
QUE000624		45.5039	−73.5545	*Q. macrocarpa*	Marlene Hahn	IL-SH-190	Quebec		Montreal	wild
QUE001759		43.1070	−89.8083	*Q. macrocarpa*	Mira Garner	WI-MG230	WI	Dane	Pleasant Valley Conservancy	wild
QUE001804		42.0421	−93.6062	*Q. macrocarpa*	Mira Garner	IA-MG-261	IA	Story	Ames, Veenker Memorial Golf Course (ISU)	wild
QUE001814		40.7056	−91.7942	*Q. macrocarpa*	Mira Garner	IA-MG-269	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001863		36.8450	−96.4253	*Q. macrocarpa*	Mira Garner	OK-MG282	OK	Osage	Pawhuska, Tallgrass Prairie Preserve	wild
QUE001894		35.4438	−98.3545	*Q. macrocarpa*	Mira Garner	OK-MG363	OK	Caddo	Hinton, Red Rock Canyon State Park	wild
QUE001916		34.8038	−92.3263	*Q. macrocarpa*	Mira Garner	AR-MG385	AR	Pulaski	Little Rock, Burns Park	wild
QUE001933		37.1569	−91.3647	*Q. macrocarpa*	Mira Garner	MO-MG402	MO	Shannon	Eminence, Buttin Rock Access, near river	wild
QUE001783		41.9736	−91.7239	*Q. macrocarpa*	Mira Garner	IA-MG-239	IA	Linn	Cedar Rapids, Cherokee Park	wild
QUE001880		36.2204	−95.8985	*Q. macrocarpa*	Mira Garner	OK-MG349	OK	Tulsa	Tulsa, Mohawk Park	wild
QUE001907		35.1769	−97.4497	*Q. macrocarpa*	Mira Garner	OK-MG376	OK	Cleveland	Norman, Oliver's Woods, University of Oklahoma campus	wild
QUE001937		43.9029	−91.6400	*Q. macrocarpa*	Mira Garner	MN-MG493	MN	Winona	Winona, Prairie Moon Nursery, Wiscoy Co-op, in woods near side of path	wild
QUE001951		49.7138	−95.2439	*Q. macrocarpa*	Mira Garner	MB-MG507	Manitoba		Whiteshell Provincial Park	wild
QUE001963		49.7614	−99.1604	*Q. macrocarpa*	Mira Garner	MB-MG519	Manitoba		Spruce Woods Provincial Park	wild
QUE001971		49.8578	−97.2491	*Q. macrocarpa*	Mira Garner	MB-MG527	Manitoba		Assiniboine Forest, near trail	wild
QUE001982		46.0259	−91.1429	*Q. macrocarpa*	Mira Garner	WI-MG538	WI	Sawyer	Round Lake, Chequamegon-Nicolet National Forest	wild
QUE002057		38.9636	−98.5891	*Q. macrocarpa*	Ian Pearse	Minooka 2	KS	Russell	Minooka Park Recreation Area	wild
QUE002074		39.8393	−88.3677	*Q. macrocarpa*	Ian Pearse	Chickenbristle 3	IL	Douglas	property of Bob Pearse	wild
QUE002081		37.5149	−89.4445	*Q. macrocarpa*	Mira Garner	IL-MG602	IL	Jackson	Oakwood Bottoms, Shawnee National Forest, along State Rd. 3	wild
QUE002082		37.5164	−89.4454	*Q. macrocarpa*	Mira Garner	IL-MG603	IL	Jackson	Oakwood Bottoms, Shawnee National Forest, along State Rd. 3	wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE002085		38.7362	−86.4126	*Q. macrocarpa*	Mira Garner	IN-MG606	IN	Lawrence	Spring Mill State Park, near trail/lake	wild
QUE002102		40.4591	−85.5041	*Q. macrocarpa*	Mira Garner	IN-MG623	IN	Grant	Taylor Wilderness, Taylor University	wild
QUE002129		41.9623	−73.3130	*Q. macrocarpa*	Paul Gugger	QUMAC-1028	CT	Litchfield	Litchfield Co., US-CT	wild
QUE002133		42.1666	−73.4121	*Q. macrocarpa*	Paul Gugger	QUMAC-1032	MA	Berkshire	16-18 Creamery Rd., Egremont, MA	wild
QUE002137		43.6034	−73.1811	*Q. macrocarpa*	Paul Gugger	QUMAC-1036	VT	Rutland	D&H Trail, Castleton, VT	wild
QUE002154		44.4007	−73.2376	*Q. macrocarpa*	Paul Gugger	QUMAC-1053	VT	Chittenden	Shelburne Bay, Shelburne, VT	wild
QUE002208		38.2283	−91.0824	*Q. macrocarpa*	Mira Garner	MO-MG584	MO	Crawford	Meramec State Park, Sullivan, MO, campground	wild
QUE002251		38.9801	−94.8052	*Q. macrocarpa*	Mira Garner	KS-MG431	KS	Johnson	Shawnee Mission Park	wild
QUE002284		38.8087	−95.1939	*Q. macrocarpa*	Mira Garner	KS-MG464	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild
QUE002295		39.1071	−96.6077	*Q. macrocarpa*	Mira Garner	KS-MG475	KS	Riley	Konza Prairie, Manhattan, KS, near nature trail along King's Creek	wild
QUE002302		39.1034	−96.5962	*Q. macrocarpa*	Mira Garner	KS-MG482	KS	Riley	Konza Prairie, Manhattan, KS, along King's Creek	wild
QUE002326		43.8532	−83.9230	*Q. macrocarpa*	Mira Garner	MI-MG663	MI	Bay	Pinconning Park, campground	wild
QUE002336		42.3084	−83.7567	*Q. macrocarpa*	Mira Garner	MI-MG673	MI	Washtenaw	Barton Nature Area, Ann Arbor, picnic area	wild
QUE002356		42.7651	−84.3890	*Q. macrocarpa*	Mira Garner	MI-MG693	MI	Ingham	Lake Lansing Park, East Lansing, edge of marsh	wild
QUE002367		43.0169	−90.1148	*Q. macrocarpa*	Mira Garner	WI-MG704	WI	Iowa	Governor Dodge State Park, Dodgeville, Cox Hollow Campground	wild
QUE002400		43.0161	−88.4351	*Q. macrocarpa*	Mira Garner	WI-MG737	WI	Waukesha	University of Wisconsin, Waukesha Field Station, Ottawa	wild
QUE002424		44.3639	−93.9354	*Q. macrocarpa*	Mira Garner	MN-MG761	MN	Le Sueur	Ottawa Bluffs	wild
QUE002436		45.5300	−94.2364	*Q. macrocarpa*	Mira Garner	MN-MG773	MN	Stearns	Quarry Park State Natural Area, Waite Park	wild
QUE002471		40.7312	−83.0933	*Q. macrocarpa*	Mira Garner	OH-MG808	OH	Crawford	Daughmer Prairie Savannah State Nature Preserve	wild
QUE002480		41.5507	−84.3597	*Q. macrocarpa*	Mira Garner	OH-MG817	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE002579		43.8168	−103.2501	*Q. macrocarpa*	Jeannine Cavender-Bares	JCB-SD-US16-1010	SD	Custer	E of Custer State Park driving eastward toward Hermosa on US16 Alt-W	wild
QUE002585		44.4673	−103.8498	*Q. macrocarpa*	Jeannine Cavender-Bares	JCB-SD-SPF-1016	SD	Lawrence	heading toward Spearfish	wild
QUE002717	** B **	38.9057	−86.0359	*Q. michauxii*	Elisabeth Fitzek	645-48*2	IN	Jackson		wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE000121	** B **	38.9057	−86.0359	*Q. michauxii*	Laurie Glaysher	IL-MOR-MH079	IN	Jackson		wild
QUE002718		unknown	unknown	*Q. michauxii*	Elisabeth Fitzek	476-42*1	cultivated	cultivated	graft from University of Washington Botanic Gardens	cultivated
QUE002719		36.6447	−89.2850	*Q. michauxii*	Elisabeth Fitzek	539-96*3	MO	Mississippi	in picnic ground at Big Oak Tree State Park	wild
QUE002679		36.6447	−89.2850	*Q. michauxii*	Carol DeVries	IL-MOR-MH250	MO	Mississippi	in picnic ground at Big Oak Tree State Park	wild
QUE002680		36.6447	−89.2850	*Q. michauxii*	Marilyn Carle	IL-MOR-MH251	MO	Mississippi	in picnic ground at Big Oak Tree State Park	wild
QUE002720	** C **	36.6447	−89.2850	*Q. michauxii*	Elisabeth Fitzek	539-96*5	MO	Mississippi	in picnic ground at Big Oak Tree State Park	wild
QUE000105	** C **	36.6447	−89.2850	*Q. michauxii*	Ken Potenberg	IL-MOR-MH001	MO	Mississippi	in picnic ground at Big Oak Tree State Park	wild
QUE000588		37.1538	−89.3470	*Q. michauxii*	Bethany Hayward Brown	IL-SH-154	IL		Horseshoe Lake, Olive Branch	wild
QUE001116		35.9956	−79.0542	*Q. michauxii*	Paul Manos	PM143	NC	Orange	Johnston Mill Preserve	wild
QUE001128		36.0152	−78.9233	*Q. michauxii*	Paul Manos	PM155	NC	Durham	Edith Street, Durham, NC	wild
QUE002722		37.1414	−79.9957	*Q. montana*	Elisabeth Fitzek	606-2000*3	VA	Franklin	Cahas Mtn.	wild
QUE002723	** D **	37.5258	−80.2497	*Q. montana*	Elisabeth Fitzek	602-2000*2	VA	Craig	at picnic area across from entrance to county rd. 1772 leading to Potts Mtn., Jefferson National Forest, Craig, Virginia; lat. 37.525750N, long. 80.249722W, elev. 926 m	wild
QUE000122	** D **	37.5258	−80.2497	*Q. montana*	Evelyn Means	IL-MOR-MH080	VA	Craig	at picnic area across from entrance to county rd. 1772 leading to Potts Mtn., Jefferson National Forest, Craig, Virginia; lat. 37.525750N, long. 80.249722W, elev. 926 m	wild
QUE002724		37.5258	−80.2497	*Q. montana*	Elisabeth Fitzek	602-2000*1	VA	Craig	at picnic area across from entrance to county rd. 1772 leading to Potts Mtn., Jefferson National Forest, Craig, Virginia; lat. 37.525750N, long. 80.249722W, elev. 926 m	wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE002725		37.5258	−80.2497	*Q. montana*	Elisabeth Fitzek	602-2000*3	VA	Craig	at picnic area across from entrance to county rd. 1772 leading to Potts Mtn., Jefferson National Forest, Craig, Virginia; lat. 37.525750N, long. 80.249722W, elev. 926 m	wild
QUE000639		37.6179	−88.7048	*Q. montana*	Marlene Hahn	IL-SH-26	IL	Saline	near Stonefort	wild
QUE000111		37.1414	−79.9957	*Q. montana*	Marlene Hahn	IL-MOR-MH007	VA	Franklin	Cahas Mtn.	wild
QUE000576		35.4291	−82.2518	*Q. montana*	Bethany Hayward Brown	IL-SH-116	NC		Chimney Rock Park	wild
QUE002726		unknown	unknown	*Q. muehlenbergii*	Elisabeth Fitzek	704-46*2	IN			wild
QUE002727		unknown	unknown	*Q. muehlenbergii*	Elisabeth Fitzek	704-63*3	IN			wild
QUE000152		40.6715	−95.7047	*Q. muehlenbergii*	Chris Courtney	IL-MOR-MH110	IA	Fremont	8 mi. E of Nebraska City, S of IA 2, on top of E bluff overlooking Missouri River in Waubonsie State Park	wild
QUE000587		39.9352	−89.8023	*Q. muehlenbergii*	Bethany Hayward Brown	IL-SH-153	IL	Menard	Petersburg, Starhill Forest Arboretum, spontaneous plant (presumably local population)	wild
QUE000145		41.2106	−88.0176	*Q. muehlenbergii*	Marlene Hahn	IL-MOR-MH103	IL	Will	in campground N of the Kankakee River under spontaneous plants 1/2 mi. W of the Kankakee Co. line	wild
QUE000670		35.6239	−99.0087	*Q. muehlenbergii*	Bethany Hayward Brown	IL-SH-57	OK	Custer		wild
QUE000322		31.9792	−104.7542	*Q. muehlenbergii*	Marlene Hahn	CA-DAV-MH49	TX	Culberson	Guadalupe Mtns., McKittrick Canyon, 0.6 mi. downstream from Rock House	wild
QUE001819		40.7048	−91.7959	*Q. muehlenbergii*	Mira Garner	IA-MG-274	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001840		37.9730	−92.7622	*Q. muehlenbergii*	Mira Garner	MO-MG326	MO	Camden	Ha Ha Tonka State Park	wild
QUE001893		35.4481	−98.3535	*Q. muehlenbergii*	Mira Garner	OK-MG362	OK	Caddo	Hinton, Red Rock Canyon State Park	wild
QUE002086		38.7363	−86.4125	*Q. muehlenbergii*	Mira Garner	IN-MG607	IN	Lawrence	Spring Mill State Park, Trail 5	wild
QUE002098		38.7374	−86.4126	*Q. muehlenbergii*	Mira Garner	IN-MG619	IN	Lawrence	Spring Mill State Park, roadside near Nature Center	wild
QUE002101		40.4589	−85.5038	*Q. muehlenbergii*	Mira Garner	IN-MG622	IN	Grant	Taylor Wilderness, Taylor University, off 8th St.	wild
QUE002105		40.4592	−85.5081	*Q. muehlenbergii*	Mira Garner	IN-MG626	IN	Grant	Taylor Wilderness, Taylor University	wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE002189		38.5114	−90.5592	*Q. muehlenbergii*	Mira Garner	MO-MG565	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002247		38.9789	−94.8050	*Q. muehlenbergii*	Mira Garner	KS-MG427	KS	Johnson	Shawnee Mission Park	wild
QUE002250		38.9739	−94.8051	*Q. muehlenbergii*	Mira Garner	KS-MG430	KS	Johnson	Shawnee Mission Park	wild
QUE002285		38.8087	−95.1939	*Q. muehlenbergii*	Mira Garner	KS-MG465	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild
QUE002303		39.1016	−96.5998	*Q. muehlenbergii*	Mira Garner	KS-MG483	KS	Riley	Konza Prairie, Manhattan, KS	wild
QUE002304		39.1079	−96.6047	*Q. muehlenbergii*	Mira Garner	KS-MG484	KS	Riley	Konza Prairie, Manhattan, KS, along nature trail	wild
QUE002481		41.5510	−84.3585	*Q. muehlenbergii*	Mira Garner	OH-MG818	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE002699		40.0499	−95.7298	*Q. prinoides*	Chris Courtney	IL-MOR-MH270	NE	Richardson	Rock Creek bluffs, 3 mi. S of Salem	wild
QUE002695		39.8189	−94.0103	*Q. prinoides*	Satish Sachdev	IL-MOR-MH266	MO		NW section of Missouri	wild
QUE002728		40.0383	−95.7565	*Q. prinoides*	Elisabeth Fitzek	120-2001*2	NE	Richardson	plant grown from wild seed SW of Salem	wild
QUE002729		40.0383	−95.7565	*Q. prinoides*	Elisabeth Fitzek	120-2001*3	NE	Richardson	plant grown from wild seed SW of Salem	wild
QUE002730		unknown	unknown	*Q. prinoides*	Elisabeth Fitzek	218-77*2	cultivated	cultivated	seed from MOR accession 742-51	cultivated
QUE002731		unknown	unknown	*Q. prinoides*	Elisabeth Fitzek	218-77*3	cultivated	cultivated	seed from MOR accession 742-51	cultivated
QUE002689		unknown	unknown	*Q. prinoides*	Sarah Packard	IL-MOR-MH260	cultivated	cultivated	seed from MOR accession 742-51	cultivated
QUE002701		40.7925	−77.8621	*Q. prinoides*		IL-MOR-MH272	PA	Centre	state college grounds	wild
QUE002694	** E **	40.7925	−77.8621	*Q. prinoides*	Marilyn Carle	IL-MOR-MH265	PA	Centre	state college grounds	wild
QUE000133	** E **	40.7925	−77.8621	*Q. prinoides*	Ken Potenberg	IL-MOR-MH091	PA	Centre	state college grounds	wild
QUE002693		39.8189	−94.0103	*Q. prinoides*	Edie Moran	IL-MOR-MH264	MO	Daviess	NW section of Missouri	wild
QUE002696		39.8189	−94.0103	*Q. prinoides*	Chris Courtney	IL-MOR-MH267	MO	Daviess	NW section of Missouri	wild
QUE002697		40.7925	−77.8621	*Q. prinoides*	Charlene Kubic	IL-MOR-MH268	PA	Centre	state college grounds	wild
QUE002698		40.7925	−77.8621	*Q. prinoides*	Chris Courtney	IL-MOR-MH269	PA	Centre	state college grounds	wild
QUE002700		40.0499	−95.7298	*Q. prinoides*	Charlene Kubic	IL-MOR-MH271	NE	Richardson	Rock Creek bluffs, 3 mi. S of Salem	wild
QUE000565		38.4676	−95.1365	*Q. prinoides*	Andrew Hipp	IL-SH-105	KS	Franklin	3.5 mi. NW of Lane	wild
QUE000678		40.0763	−95.7210	*Q. prinoides*	Marlene Hahn	IL-SH-96	NE	Richardson	SW of Salem	wild
QUE000753		42.0015	−76.5991	*Q. prinoides*	E. A. Cope	PM93	NY	Chemung		wild
QUE002683		40.6487	−91.6733	*Q. stellata*	Satish Sachdev	IL-MOR-MH254	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002732				*Q. stellata*	Elisabeth Fitzek	11-86*2	MO	Lawrence		wild
QUE002733		37.0922	−93.8381	*Q. stellata*	Elisabeth Fitzek	11-86*3	MO	Lawrence		wild
QUE002734	** F **	38.0349	−91.5203	*Q. stellata*	Elisabeth Fitzek	1137-2004*2	MO	Phelps	along I-44 & RR at Rosati	wild
QUE000143	** F **	38.0349	−91.5203	*Q. stellata*	Marlene Hahn	IL-MOR-MH101	MO	Phelps	along I-44 & RR at Rosati	wild

Table 1. Continued.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	Collector number	State of origin	County of origin	Locality of origin	Source
QUE002735		38.0349	−91.5203	*Q. stellata*	Elisabeth Fitzek	1137-2004*3	MO	Phelps	along I-44 & RR at Rosati	wild
QUE002704		40.6487	−91.6733	*Q. stellata*	Ken Potenberg	IL-MOR-MH275	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002703		40.6487	−91.6733	*Q. stellata*	Ken Potenberg	IL-MOR-MH274	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002706	** G **	40.6487	−91.6733	*Q. stellata*	Bethany Hayward Brown	IL-MOR-MH277	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE000137	** G **	40.6487	−91.6733	*Q. stellata*	Marlene Hahn	IL-MOR-MH095	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE000608		38.7272	−88.7795	*Q. stellata*	Bethany Hayward Brown	IL-SH-174	IL	Marion	Forbes State Recreation Area	wild
QUE000638		37.9995	−91.6092	*Q. stellata*	Marlene Hahn	IL-SH-25	MO	Phelps	Hillview Haven, St. James	wild
QUE000692		35.9767	−78.9866	*Q. stellata*	Paul Manos	PM11	NC	Durham	Durham Co., 3658 Pineview Circle	wild
QUE001118		36.0187	−78.9253	*Q. stellata*	Paul Manos	*P. S. Manos 1907*	NC	Durham	Watts Hillandale tree intersection of Carolina Ave. & Woodrow	wild
QUE001839		37.9729	−92.7622	*Q. stellata*	Mira Garner	MO-MG325	MO	Camden	Ha Ha Tonka State Park	wild
QUE001862		36.8491	−96.4152	*Q. stellata*	Mira Garner	OK-MG281	OK	Osage	Pawhuska, Tallgrass Prairie Preserve	wild
QUE001915		34.8041	−92.3267	*Q. stellata*	Mira Garner	AR-MG384	AR	Pulaski	Little Rock, Burns Park	wild
QUE002187		38.5109	−90.5599	*Q. stellata*	Mira Garner	MO-MG563	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002188		38.5115	−90.5592	*Q. stellata*	Mira Garner	MO-MG564	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002209		38.2181	−91.0836	*Q. stellata*	Mira Garner	MO-MG585	MO	Crawford	Meramec State Park, Sullivan, MO	wild
QUE002219		38.2179	−91.0921	*Q. stellata*	Mira Garner	MO-MG595	MO	Crawford	Meramec State Park, Sullivan, MO, Deer Hollow Trail head	wild
QUE002252		38.9801	−94.8050	*Q. stellata*	Mira Garner	KS-MG432	KS	Johnson	Shawnee Mission Park	wild
QUE002283		38.8096	−95.1927	*Q. stellata*	Mira Garner	KS-MG463	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild

Table 2. Sample sizes, sample distances and ranges, and overall species ranges. Sample distance (D) maximum and median were calculated from Table 1 using the Haversine formula. Species ranges were inferred from range maps of Little (1971, 1977, 1979) for all species except *Quercus prinoides* Willd., which was estimated by visual inspection of maps published in *Flora of North America* (Nixon, 1997).

	N	Sample D max (km)	Sample D median (km)	Sample latitude	Species latitude	Sample longitude	Species longitude
Q. macrocarpa	52	3005.3	888.8	33.6, 49.9	28, 52.7	−105.4, −73.2	−104.4, −66.1
Q. alba	26	2120.1	695	34.8, 45.4	29.6, 46.5	−95.9, −73.2	−96.3, −69.1
Q. muehlenbergii	21	2098.3	543	32, 41.6	24.8, 44.7	−104.8, −84.4	−105.2, −72.2
Q. stellata	21	1565.6	325.6	34.8, 40.6	27.6, 41.8	−96.4, −78.9	−101.4, −70
Q. prinoides	17	1618.9	185.9	38.5, 42	34.1, 42.9	−95.8, −76.6	−99.8, −70
Q. bicolor	10	1889.8	453.4	38.2, 44.4	35.2, 46.4	−94.8, −73.2	−96.4, −70
Q. michauxii	9	939	380.5	36, 38.9	28.8, 41	−89.3, −78.9	−95.5, −74.3
Q. montana	7	771.3	277.8	35.4, 37.6	32, 44.6	−88.7, −80	−90, −70.5

and *Q. michauxii* Nutt. to 38 to 52 in *Q. muehlenbergii/prinoides* and *Q. macrocarpa* Michx., respectively (Table 1). The distance between the most widely separated populations sampled within each species ranges from 771 km in *Q. montana* to 3005 km in *Q. macrocarpa* (Table 2). Moreover, aside from samples of *Q. macrocarpa* at the westernmost and northernmost edges of its range (Fig. 1), almost all samples in our study were collected from within the range of at least one other species. Consequently, while our study does not encompass the entire range of each species, the samples cover a wide geographic range within each species, with the opportunity for crossing among congeners. Locations for source populations of all samples for which source information was available were plotted over published range maps. Range maps were plotted from shapefiles (Prasad & Iverson, 2003) generated from previously published range maps of North American trees (Little, 1971, 1977, 1979) over the "county" and "state" base maps provided in maps v. 3.3.0 (Becker et al., 2018) for R v. 3.4.2, "Short Summer" (R-Development-Core-Team, 2004). All plotting was done in R using the ggplot2 (Wickham, 2009) and ggmap (Kahle & Wickham, 2013) packages, using proj4 (Urbanek, 2012) for map projections.

Samples were genotyped using an 80-SNP DNA toolkit developed to distinguish 15 eastern North American white oaks (as described in Fitzek et al., 2018). Briefly, an extensive RAD-seq dataset comprising multiple exemplars of all 15 species (McVay et al., 2017b) was surveyed for SNP variation, using pairwise F_{ST} to identify SNPs that were (1) fixed or nearly fixed between species and (2) flanked by at least 20 bp of conserved sequence, which could be used for primer design. Multiplexes of up to 40 primers for potential SNPs were designed using the Assay Design 4.0 Suite (Agena Biosciences, San Diego, California, U.S.A.), which is optimized for MassARRAY analysis (Bradić et al., 2012). Samples were genotyped using the iPLEX Gold chemistry following Gabriel et al. (2009) on a MassARRAY system (Agena Biosciences) at the Genomic Platform of Bordeaux with the help of Adline Delcamp. Data analysis was completed using MassARRAY Typer Analyzer 4.0.26.75 (Agena Biosciences). We manually checked each marker clustering to detect potential ambiguous genotype assignation or unusable SNP. The results were exported as a genotype table for downstream analyses. After genotyping, five SNPs were removed from analysis because they failed to amplify in more than 30% of individuals.

The oak genome was not yet available when this DNA toolkit was published, but since then a chromosome-level genome has been published for *Quercus robur* (Plomion et al., 2018), a white oak closely related to the species for which this toolkit was developed. To evaluate the genomic independence of the loci we used in this study, all RAD-seq loci used to develop the 80 SNPs were mapped to the oak genome using BLASTN (Altschul et al., 1990; Camacho et al., 2009) with a threshold EValue of 0.0001. Each RAD-seq locus was identified as mapping to a single position on a chromosome, multiple positions, or not mapping. All SNPs were designed from distinct RAD-seq loci save two (CL_2457_66 and CL_2457_32), which both come from a single RAD-seq locus that maps to position 36,055,433 on *Q. robur* chromosome 12. The two SNPs identified in this locus were designed to distinguish *Q. stellata* from the remaining taxa and should not be considered independent of one another. They are not strongly decisive and do not figure prominently in downstream analyses in this study or in Fitzek et al. (2018).

Genomic clustering of loci was evaluated by calculating intervals between loci on each chromosome and comparing these to a simulated null distribution. The null distribution was simulated based on 10,000 replicate datasets of 59 loci drawn at random from the 41,898 uniquely mapped *Pst*I RAD-seq loci from the larger study from which our SNPs were developed (Hipp

et al., 2019). Three test statistics were evaluated: mean interval length between all loci on all chromosomes; number of intervals < 1E04 bp; and number of intervals < 1E06 bp. Code for performing this test is archived in <https://github.com/andrew-hipp/white-oak-syngameon>.

DATA ANALYSIS: EVALUATING SPECIES COHESION

We define species cohesion operationally in this study using two criteria: (1) clustering of all plants sampled from each species in genetic space, exclusive of other species, and irrespective of geography; and (2) minimal evidence of genetic admixture between species at some conserved region of the genome (in this case, based on preselected markers). By this definition, clustering of individuals by geography instead of by species would be evidence against species cohesion, as would sharing of alleles between individuals of different species, when those alleles are nearly fixed between the species. This operational definition corresponds with practices widely used by plant systematists to define "good species" (Rieseberg et al., 2006) as well as statistical methods traditionally used to infer patterns and degree of interspecific introgression (Anderson, 1949). It puts off for the time being possible empirical and philosophical issues with cohesion species as a concept (Barker, 2007; Barker & Wilson, 2010) as well as questions about the mechanisms by which species cohere (Morjan & Rieseberg, 2004).

We assess criterion 1, clustering in genetic space, using the unweighted pair group method with arithmetic mean (UPGMA) (Sokal & Michener, 1958), a clustering method that aggregates individuals based on a pairwise distance matrix, in this case a Euclidean distance matrix based on allele counts within individuals, where each allele is present as 0, 1, or 2 copies per individual. UPGMA is well suited to within-species comparisons of genetic data or other comparisons of data that are truly ultrametric, where it performs reasonably well as an estimator of genetic relatedness (Felsenstein, 2004). In our study, UPGMA has the desirable property of apportioning genetic variance to branches, so that we can assess whether the variance in our data is better assigned to among-species or within-species differences. Because our markers are designed with extreme bias toward among-species differences, we do not attempt to quantify variance components using analysis of molecular variance (AMOVA) (Excoffier et al., 1992) and urge that the clustering results not be interpreted as estimating these variance components. We compare UPGMA results with non-metric multidimensional scaling (NMDS) ordination on the same data matrix. We present results from the three-dimensional ordination because it suffices to discriminate the species in our study.

We assess criterion 2 using the Bayesian population genetic clustering algorithm implemented in STRUCTURE v 2.3.4 (Pritchard et al., 2000). We utilized the admixture model with correlated allele frequencies and λ fixed at 1.0, allowing *K* (the number of populations) to range from 1 to 12. For each value of *K*, we ran 10 replicate Markov chain Monte Carlo (MCMC) runs of 1E06 generations following a 1E05 generation burn-in. We followed the method of Evanno et al. (2005) to identify the most probable value of *K* based on the maximum value of Δ*K*, but given the problematic nature of identifying *K* with hierarchical data, we report the configurations recovered under multiple values of *K*. We utilized STRUCTURE HARVESTER (<http://taylor0.biology.ucla.edu/structureHarvester/>) (Earl & vonHoldt, 2012) to calculate the Evanno statistics and CLUMPP v 1.1.2 for 64-bit Linux (Jakobsson & Rosenberg, 2007) to average STRUCTURE run replicates for each value of *K*. We visualized results using DISTRUCT v. 1.1 (Rosenberg, 2004).

To evaluate whether the entire SNP toolkit is necessary to discriminate among the species we are studying and to identify SNPs that might be fixed within species, we calculated the absolute number and proportion of individuals within each species possessing each polymorphism observed. With the caveat that sampling is uneven across species (ranging from N = 7 in *Quercus montana* to N = 52 in *Q. macrocarpa*), the resulting heatmap (Fig. 2, Supplementary Figure S1) and the table underlying it (Supplementary Table S1) estimate the decisiveness of each SNP relative to species identification in this species group: the summed proportion of individuals by species that have a given SNP estimates that SNP's decisiveness, where a sum of 1.0 or 2.0 (for *Q. muehlenbergii/prinoides*) indicates a locus that is alone decisive for a taxon for the samples we have genotyped. The reduced set may have practical benefit for both cost and because the combinability of primer pairs plays a crucial role in multiplexing (Fitzek et al., 2018). Decisiveness was overlaid on the mapped SNPs to identify whether loci that are fixed or nearly fixed within species are genomically clustered (Table 3).

All data and code required to reproduce analyses presented here are archived in <https://github.com/andrew-hipp/white-oak-syngameon>.

RESULTS

In the full dataset of 184 individuals for 80 loci, missing data per individual averaged 2.56 ± 4.10% (SD) loci, and missing data per locus averaged 14.6 ± 26.8% (SD) individuals. In the dataset cleaned to 163 individuals for 75 loci, excluding individuals with > 10% missing loci and loci with > 30% missing individuals, missing data dropped to 1.19 ± 1.13%

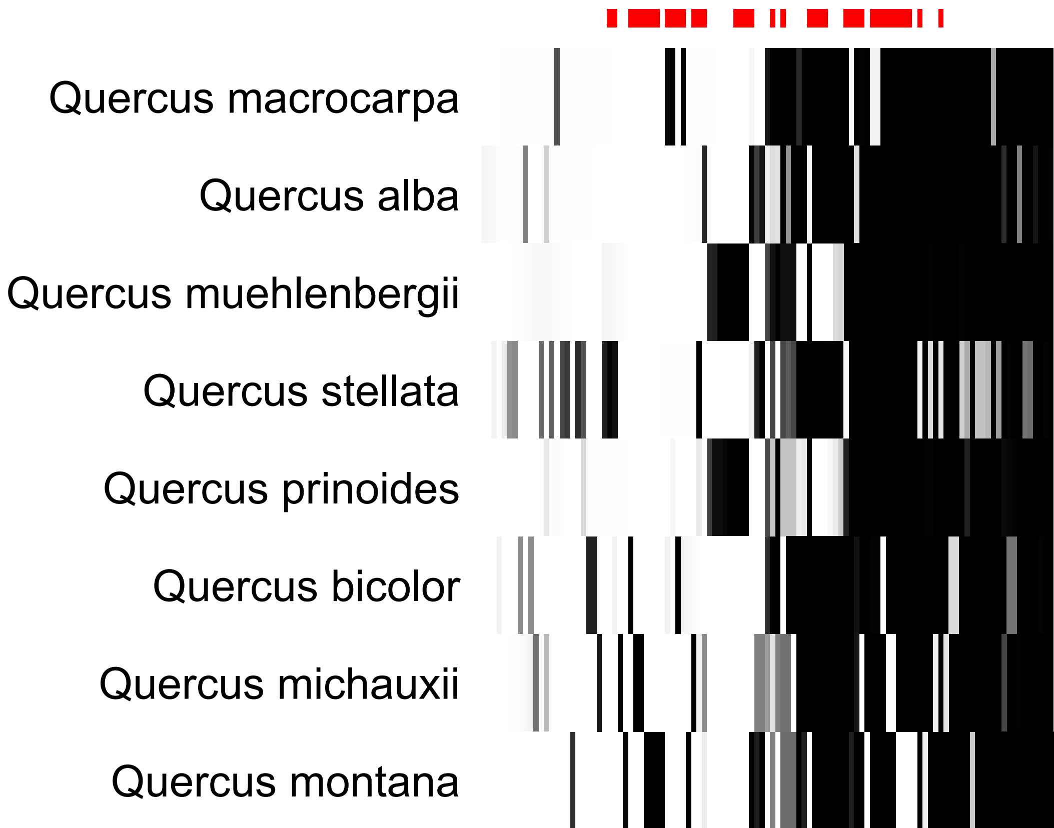

Figure 2. Single-nucleotide polymorphism (SNP) heatmap by species. Darkness of cells indicates the percent of individuals of a given named species possessing the indicated nucleotide. Red bars along the top of the figure indicate SNPs in 20 loci that we hand-selected because they were highly decisive for the species represented in the present study. Names of all loci are included in Supplementary Figure S1.

missing loci per individual and 5.60 ± 13.9% missing individuals per locus. Of the 75 cleaned loci, 20 were monomorphic and 55 had two or more polymorphisms. Among seven pairs of technical replicates, a total of 38 differences were found. Of these, 37 were differences in whether a locus amplified or not; only one difference in allele call was found (for locus CL_55087_OAK-MOR340_32, G/T in *Quercus stellata* QUE002706 vs. G/G in specimen QUE000137). Thus, among 7 × 75 = 525 replicated sites, only one genotyping error (0.17%) and 37 loci that failed to amplify in one of the two replicates (6.43%) were detected.

Seven loci exhibit only a single SNP for exactly one species in our dataset—one in *Quercus alba* L., two in *Q. michauxii*, four in *Q. montana*—and three exhibit a single SNP in *Q. muehlenbergii/prinoides*. An additional 10 SNPs exhibit a summed proportion between 0.95 and 1.05, suggesting relatively high decisiveness for *Q. stellata* (two SNPs) and *Q. bicolor* (three SNPs). Based on these, we handpicked 20 SNPs that suffice to diagnose the species in our study (Fig. 2, Supplementary Figure S1, red bars).

Using all loci, the UPGMA (Fig. 3, Supplementary Figure S2) and NMDS ordination (Fig. 4) both clearly separate individuals by species, except for *Quercus prinoides* and *Q. muehlenbergii*, which our SNP genotyping primers were not designed to distinguish from one another. Thus there are seven distinct clusters recognized in this study. Individuals of these clusters separate with no overlap in three-dimensional genetic ordination space (Fig. 4; note that while some species overlap in one or two dimensions, none overlap in all three) and UPGMA stem lengths that equal or exceed the species crown depth for four of the clusters (*Q. macrocarpa*, *Q. bicolor*, *Q. muehlenbergii/prinoides*, and *Q. montana*) and, for the other three, stem lengths that are approximately equal to (*Q. stellata*, *Q. michauxii*) or substantially less than (*Q. alba*) the crown height. Using the 20 handpicked loci, our SNP genotyping toolkit successfully distinguishes species from one another using UPGMA (Fig. 3, Supplementary Figure S2).

Bayesian admixture analysis in STRUCTURE favors a $K = 4$ solution using the ΔK statistic of Evanno et al. (2005). Given the susceptibility of STRUCTURE and

Table 3. Map positions and decisiveness of single-nucleotide polymorphisms (SNPs) that map to a unique position on one of the *Quercus robur* L. chromosomes. The 60 SNPs that map back to one of the 12 *Q. robur* chromosomes inferred in Plomion et al. (2018) are shown here with their start position on the *Q. robur* chromosome to which they map and their decisiveness. Shading changes from top to bottom of the table indicate different oak chromosomes. All loci mapped with identity > 95%, locus length > 70 bp, and E-value < 1.0 × 10^{-30}. The table demonstrates that the most decisive loci in our toolkit are distributed across all chromosomes except 5, 7, and 11, and separated by an average of 7.47 million bp ± 8.74 million bp (SD). Five pairs of loci are < 10,000 bp from one another—loci at positions LG2, 2.72E07; LG2, 5.65E07; LG3, 2.99E07; and LG12, 2.11E07, two pairs of loci—and bear further investigation as potentially falling within genomic islands of differentiation. Mapping details from BLASTN and mapping information from non-uniquely mapping loci and loci that did map are in Supplemental Table S3.

Query (locus)	*LG*	Start (bp)	Dist. (bp)	Decisiveness
locus_11631_48	01	1.32E+07		
locus_17927_52	01	1.92E+07	6.04E+06	*
newl_21880_27	01	2.02E+07	9.71E+05	***
locus_821_26	01	4.61E+07	2.59E+07	
newl_17339_35	01	4.85E+07	2.40E+06	**
CL_42027_	02	9.97E+06		**
CL_6426_61	02	1.46E+07	4.62E+06	
locus_4492_52	02	2.38E+07	9.21E+06	
locus_20180_49	02	2.72E+07	3.37E+06	
locus_3962_56	02	2.72E+07	1.90E+03	
CL_49075_43	02	4.71E+07	1.99E+07	
CL_35240	02	5.65E+07	9.40E+06	**
locus_12538_49	02	5.65E+07	9.16E+03	***
locus_23517_52	02	6.66E+07	1.01E+07	
newl_23554_	02	7.19E+07	5.26E+06	
locus_3169_44	02	7.55E+07	3.68E+06	
locus_9121_49	02	9.24E+07	1.68E+07	
CL_55087__32	02	9.38E+07	1.43E+06	*
locus_8059_35	03	2.99E+07		
CL_11069_58	03	2.99E+07	2.65E+03	*
locus_8717_53	03	3.81E+07	8.25E+06	
locus_7123_50	03	4.03E+07	2.18E+06	**
locus_5882_32	03	5.25E+07	1.22E+07	
locus_8617_30	04	3.15E+07		**
locus_5229_56	05	5.16E+07		
locus_29214_32	06	1.28E+07		***
locus_10977_45	06	2.04E+07	7.58E+06	
CL_54979_	06	3.53E+07	1.49E+07	
CL_12923_	06	4.47E+07	9.42E+06	**
locus_7834_43	06	4.58E+07	1.02E+06	**
newl_27648_32	07	1.25E+07		
locus_5482_34	07	2.57E+07	1.32E+07	
locus_27412_25	07	3.67E+07	1.10E+07	
locus_30948_43	08	9.66E+05		***
locus_5422_58	08	4.43E+07	4.34E+07	
locus_26761_43	08	5.11E+07	6.78E+06	***
locus_24383_42	08	6.07E+07	9.62E+06	
locus_10104_41	08	6.92E+07	8.51E+06	**
locus_28457_43	09	1.90E+07		**
newl_16979_31	09	2.81E+07	9.05E+06	
locus_1378_30	09	3.61E+07	8.08E+06	***
locus_30512_25	10	2.75E+06		**
locus_2085_53	10	3.57E+07	3.29E+07	
locus_20667_37	11	2.92E+07		
locus_14289_31	11	3.82E+07	9.02E+06	
CL_48165	12	1.46E+07		***
locus_11302_50	12	1.61E+07	1.55E+06	**
locus_31722_39	12	1.75E+07	1.34E+06	***
locus_9837_55	12	2.05E+07	3.04E+06	
newl_25158_45	12	2.11E+07	5.67E+05	
locus_26885_29	12	2.11E+07	3.37E+03	***
locus_792_52	12	2.11E+07	5.29E+04	
locus_8226_51	12	2.11E+07	7.99E+03	***
locus_25236_45	12	2.12E+07	9.02E+04	**
newl_15918_ PM11_41	12	2.41E+07	2.83E+06	
locus_10802_36	12	2.99E+07	5.81E+06	
locus_17368_30	12	3.54E+07	5.49E+06	
CL_2457_32	12	3.61E+07	6.97E+05	
CL_2457_OAK-MOR-340_66	12	3.61E+07		*same locus*
locus_4850_29	12	3.92E+07	3.12E+06	

Abbreviations: Decisiveness, decisiveness of the SNP for identifying one species or a pair of species (cf. Fig. 2); Dist. (bp), distance in base pairs from the start of the locus to the end of the locus adjacent to it on the same chromosome; LG, linkage group (chromosome number), following Plomion et al. (2018); Query (locus), the RAD-seq locus SNP abbreviation from Fitzek et al. (2018); Start (bp), start position of the RAD-seq locus on the *Q. robur* chromosome.

*Within 0.200 of 1.000 or 2.000.

**Within 0.100 of 1.000 or 2.000.

***An SNP whose decisiveness is exactly 1.000 or 2.000 for the sample studied here (i.e., diagnostic for one or two species).

particularly the ΔK statistic to the highest hierarchical level of genetic structure in a dataset, we find the $K = 4$ solution not a useful description of genetic structure in our phylogenetically structured dataset. To the contrary, the $K = 4$ clustering does the best job at separating species by clade, following well-supported phylogenetic relationships (Hipp et al., 2018), viz. four clusters comprising *Quercus macrocarpa* and *Q. bicolor*; *Q. alba*, *Q. michauxii*, and *Q. montana*; and *Q. stellata* and *Q. muehlenbergii/prinoides* each on their own (Fig. 5). Given our phylogenetically structured sample, it is not surprising that ΔK favors a configuration that splits individuals among clades above the species level. STRUCTURE continues to distinguish species up until $K = 8$, with seven species pairs yielding individuals admixed 10% or more based on our markers (Figs. 5, 6). Notably, it is not until $K = 8$ that the seven species are distinguished from each other, perhaps due to high genetic variation within species that is not adequately

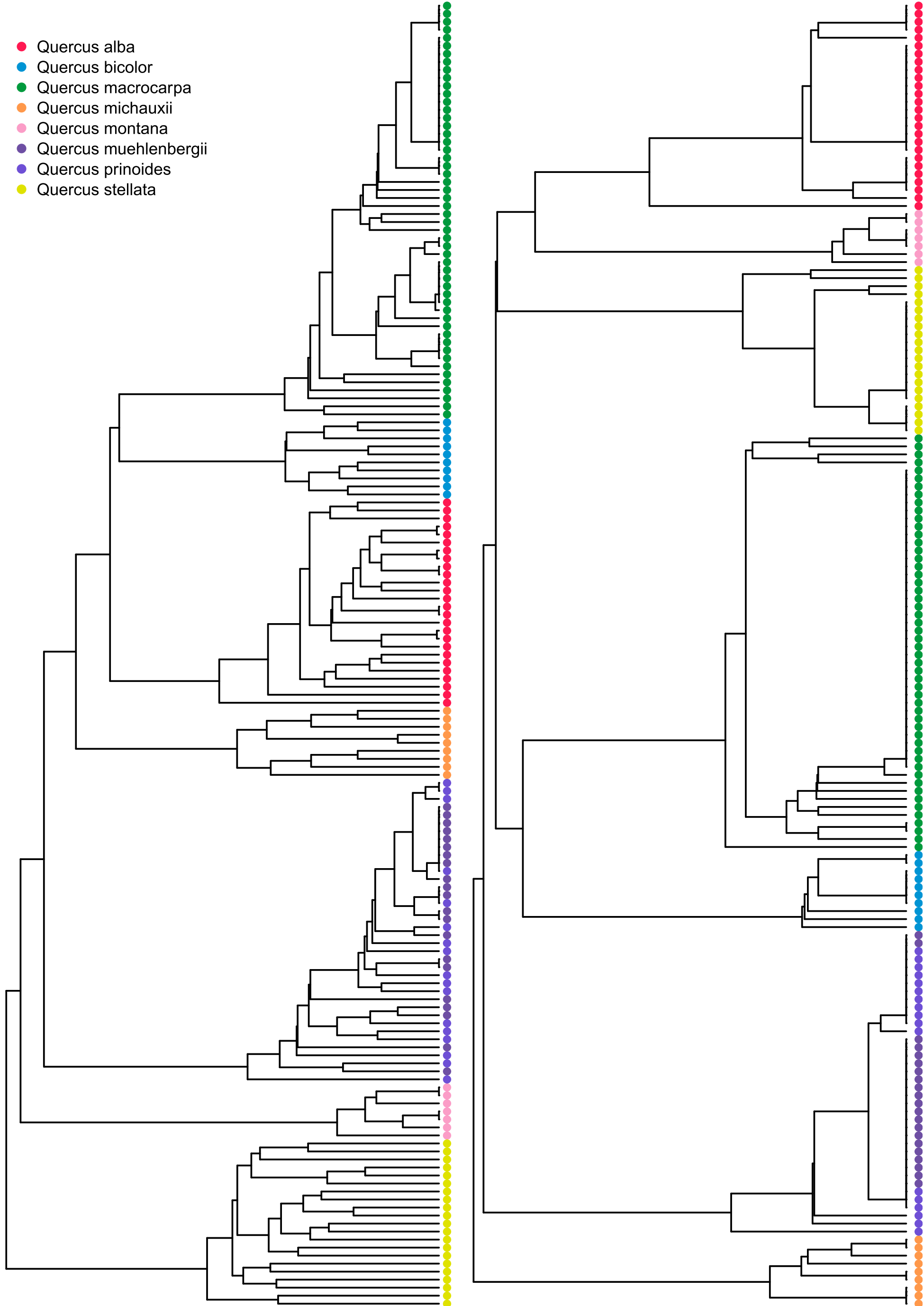

Figure 3. Unweighted pair group method with arithmetic mean (UPGMA), all loci (A) and 20 loci (B). UPGMA was conducted on a Euclidean distance matrix calculated from a three-state nucleotide matrix, where each nucleotide present for each SNP is coded as 0 = absent, 1 = 1 copy (i.e., individual is heterozygous for that single-nucleotide polymorphism [SNP]), 2 = 2 copies (i.e., individual is homozygous for that SNP). (A) UPGMA clustering based on all 75 loci. (B) UPGMA clustering based on 20 loci hand-selected for their decisiveness in the species sample represented here (cf. Fig. 2, red bars). Full sample names are included in Supplemental Figure S2.

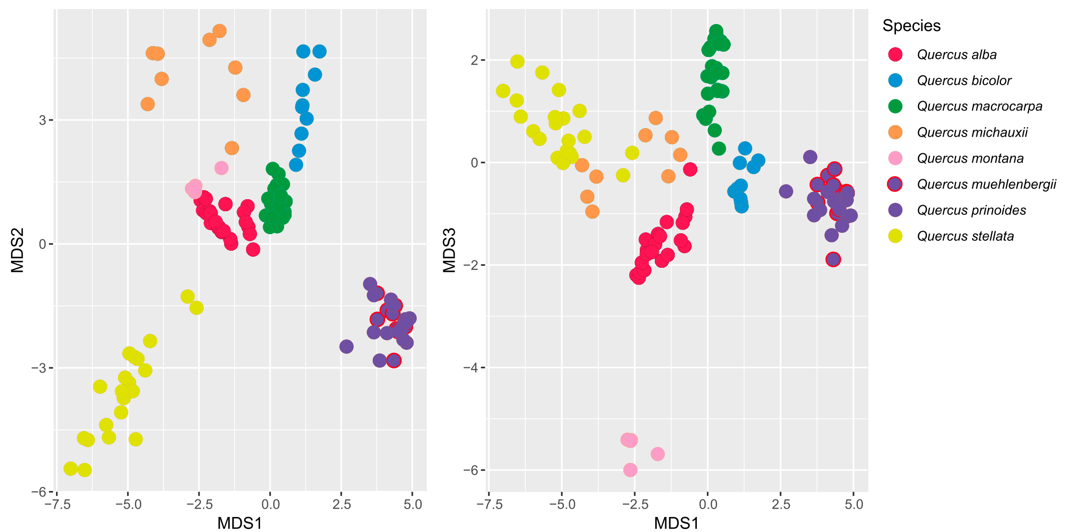

Figure 4. Non-metric multidimensional scaling (NMDS) ordination, 75 loci. NMDS was conducted in three dimensions for the same Euclidean distance matrix used in Figure 3. NMDS ordination final stress was 0.08607 and failed to reach convergent solutions in 20 iterations, but all replicate ordination attempts distinguished all pairs of species in at least one dimension, as seen in this figure.

resolved with these markers. One individual identified as *Q. alba* in the field shows evidence of introgression from both *Q. macrocarpa* and *Q. bicolor*. In the $K = 8$ configuration, *Q. bicolor* gives the appearance of being uniformly admixed with *Q. montana* at a relatively low level (9/10 individuals < 10% admixed). However, this appears to be artifactual, as the phenomenon is absent in the $K = 6$, 7, and 9 configurations, all of which show genetic separation between *Q. bicolor* and *Q. montana*. In the $K = 8$ configuration, *Q. alba* resolves as a mix of two genotypes, which we combine in estimating the number of individuals admixed at 5%, 10%, 15%, or 20% (Supplementary Table S2, Fig. 6).

Of the 79 RAD-seq loci used to design our SNP toolkit—79 rather than 80 because two of our SNPs derive from a single locus—59 map to a unique position on one chromosome (hereafter referred to as "uniquely mapped loci"), nine map to multiple locations in the genome, and 11 do not map to any location in the genome (Table 3, Supplementary Table S3). The uniquely mapped loci demonstrate that decisiveness is spread across the genome: 25 loci diagnostic for one or two species are found on nine out of the 12 *Quercus* chromosomes (Table 3). Moreover, distances between loci within chromosomes are mostly > 1 million bp (37/47 interlocus distances), and only 11% (5/47 interlocus distances) are < 10,000 bp. Distances between uniquely mapped loci average 7.47 million bp (± 8.74 million bp, SD). These are all significantly clustered relative to a random draw of SNPs, under which only 0.909 interlocus distances < 10,000 bp are expected ($P < 0.0001$), 4.70 interlocus distances < 1,000,000 bp ($P = 0.0123$), and mean interlocus distance is expected to be 9.440×10^6 ($P < 0.0001$). Only two of the 11 RAD-seq loci that did not map to the genome exhibit moderate decisiveness (0.81–0.869, where 1.0 or 2.0 indicates loci that are perfectly decisive for one or two species, respectively). Three of the nine loci that map to multiple locations are highly decisive (1.000–1.021).

Discussion

Our study demonstrates that with a relatively small amount of curated data—just 20 SNPs chosen to maximize genetic distinctiveness—we are able to distinguish seven genetically cohesive taxa. The fact that we are able to identify fixed or nearly fixed SNPs across wide geographic ranges in several species suggests that introgression is distributed heterogeneously along the genome, with some areas of the genome strongly protected against introgression on a species-pair by species-pair basis. Given that several of the apparently fixed SNPs are limited to our species with smallest sample size—one in *Quercus bicolor* (N = 10), two in *Q. michauxii* (N = 9), four in *Q. montana* (N = 7)—the question of whether they are truly fixed bears further investigation. However, *Q. muehlenbergii/prinoides* is represented by 38 individuals in our dataset and three fixed SNPs, suggesting that the high-frequency proportional representation of SNPs in some species may not be an artifact of low sample size. We interpret this finding as evidence that these seven species are genetically cohesive across their ranges at least at a small number of regions of the genome, even in the face of introgression.

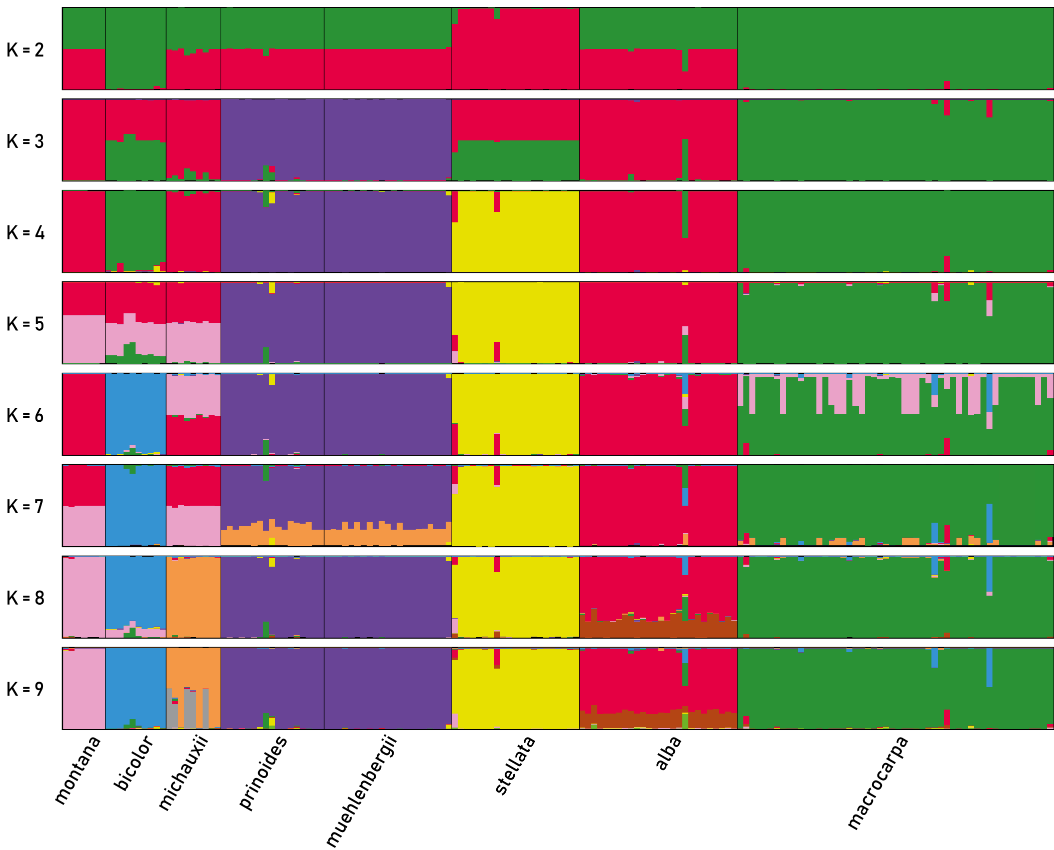

Figure 5. Bayesian admixture analysis conducted in STRUCTURE, assuming $K = 2$ to $K = 9$ populations. STRUCTURE analyses were conducted under the admixture model with correlated allele frequencies, from $K = 1$ to $K = 12$. Values of K above 9 provide no additional information on population structure and are consequently not shown here. All figures represent averages over 10 independent runs of 1E06 generations each following 1E05 burn-in generations; runs were aggregated for display using the "greedy" algorithm in CLUMPP.

It is somewhat remarkable that we are able to distinguish seven interbreeding oak species with just 20 handpicked markers. By comparison, the now-classic study demonstrating genetic distinctiveness of *Quercus petraea* (Matt.) Liebl. and *Q. robur* L. utilized 20 microsatellites for just those two species (Muir et al., 2000). Other studies using five (Craft & Ashley, 2006), six (Moran et al., 2012), or even 15 variable microsatellites (Aldrich et al., 2003) have by contrast failed to find consistent genetic differentiation between two to three co-occurring white or red oaks (for a counter example of relatively clean differentiation based on only 11 microsatellites, see Cavender-Bares & Pahlich, 2009). All used markers selected for variability rather than for segregation by species. Larger numbers of loci (as low as 27 to 28 in, e.g., Owusu et al., 2015; Sullivan et al., 2016) tend to pick up divergent neutral markers or markers under divergent selection (Lind-Riehl et al., 2014b; Sullivan et al., 2016). This suggests that a moderate-sized but random sample of loci will often reflect regions of the genome that either are not yet differentiated between species (Muir & Schlötterer, 2005, 2006) or are subject to ancient or contemporary gene flow (Lexer et al., 2006). Because the loci that bear the stamp of population divergence history for one species pair may record introgression history for other species pairs (Crowl et al., 2019; Hipp et al., 2019), we would not expect any particular small set of loci to adequately describe species description across the oak phylogeny. In the current study, however, we have demonstrated that a small number *can* suffice to distinguish numerous species in a multispecies syngameon.

The SNPs we have utilized may be linked to loci under strong selection. They may as a consequence not be representative of the genome as a whole. As discussed in the paper in which these SNPs were published (Fitzek et al., 2018), we selected SNPs by querying a RAD-seq dataset for loci that had pairwise $F_{ST} > 0.95$. Such outlier loci can tell much more refined stories about population divergence than loci that are not under

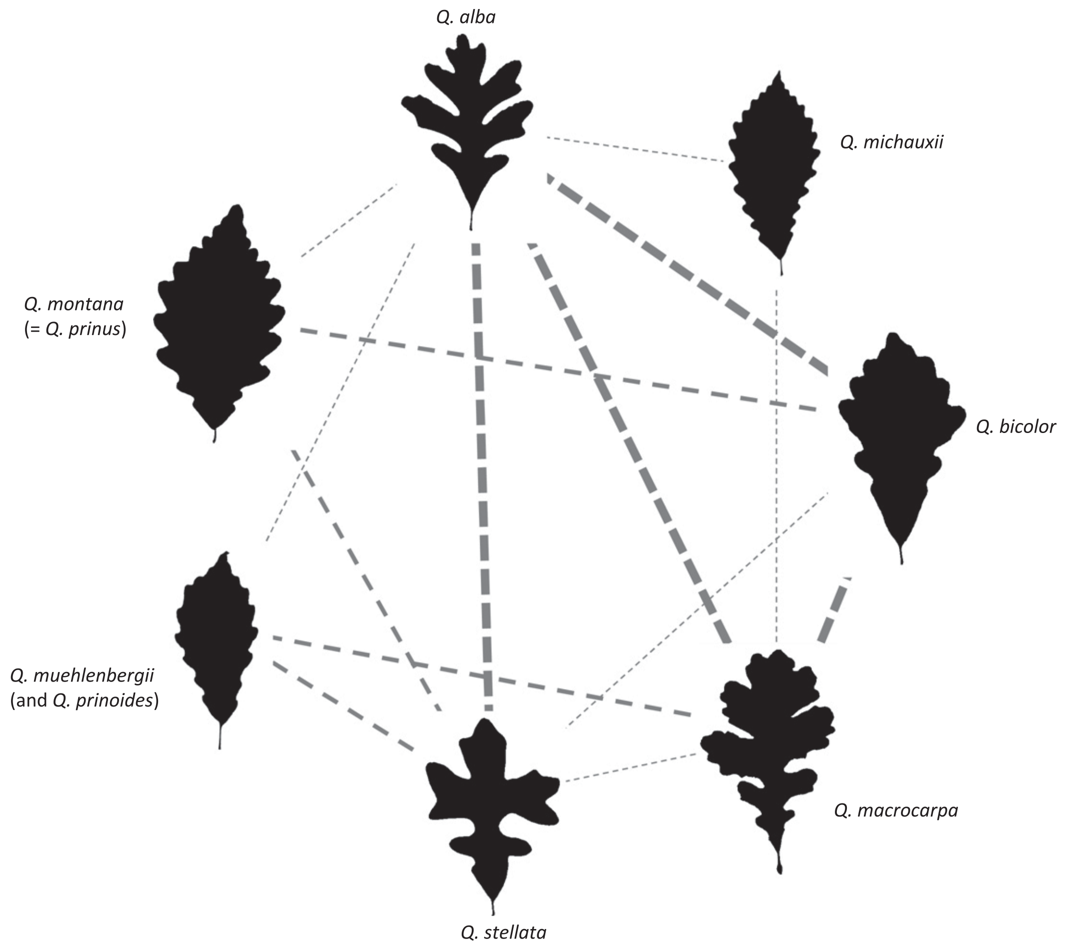

Figure 6. The white oak syngameon of eastern North America sensu Hardin (1975), including only the species investigated in the current study. The figure replicates the 16-species figure of Hardin (1975: fig. 1), including only the subset of seven species we investigated in the current study (treating *Quercus muehlenbergii* Engelm. and *Q. prinoides* Willd. as one), with lines indicating hybridizations that Hardin inferred from morphological study. Thin-dashed lines indicate hybridizations identified by Hardin but not by us; medium-dashed lines were identified by both Hardin and us, at an admixture level of 0.10 to 0.19 for at least one specimen; and thick-dashed lines indicate admixture levels of 0.20 or higher for at least one specimen. Vouchers for leaf silhouettes are *Q. alba* L.: *P. S. Manos 1838* (MOR 177669); *Q. michauxii* Nutt.: *P. S. Manos 1843* (MOR 177659); *Q. bicolor* Willd.: *P. S. Manos 1847* (MOR 177662); *Q. macrocarpa*: IL-MOR-MH108 (MOR 174544); *Q. stellata* Wagenh.: *P. S. Manos 1835* (MOR 177663); *Q. muehlenbergii*: PM-98; *Q. montana* Willd.: *P. S. Manos 1860* (MOR 177731).

such strong selection (Scotti-Saintagne et al., 2004; Guichoux et al., 2013; Lind-Riehl et al., 2014b) and may thus pick up on divergence histories that are not clear from a broader sample of loci. These selected genes may occur in islands of differentiation distributed across the genome (Scotti-Saintagne et al., 2004; Goicoechea et al., 2015) and have the potential to explain genetic cohesion across species ranges even when populations diverge at neutral loci (Morjan & Rieseberg, 2004) or to distinguish species that are exchanging genes more frequently across the remainder of the genome (Gailing & Curtu, 2014; Lind-Riehl et al., 2014; Oney-Birol et al., 2018; Hipp, 2018). This gives them practical utility as a species identification toolkit. A genome-scale investigation, as has been conducted in the European white oaks (Leroy et al., 2017, 2018), would be required to characterize the genomic architecture of differentiation among these species and address the question of whether species differences are concentrated in divergent loci under strong selection. For the time being, our study suggests that a relatively small number of selected genes may suffice to *diagnose*—not define—species, even in the face of ongoing introgression.

Despite the low sampling of loci in our study, we do find significant clustering on the genome of loci within 1

Mbp of each other ($P = 0.008$) or within 10 Kbp of each other ($P < 0.001$). This supports earlier studies that have found significant clustering of high-F_{ST} loci (Scotti-Saintagne et al., 2004) as well as linkage disequilibrium (LD) among loci separated by as much as 20 centimorgans (cM) (Goicoechea et al., 2015). While the *Pst*I RAD-seq loci used to design these SNPs are widespread on the genome, they are not randomly distributed, but rather are situated at higher-than-expected frequency within coding genes (Hipp et al., 2019). However, our simulated distribution accounts for this, as it is drawn from the larger RAD-seq dataset from which our SNPs were developed. Thus the clustering of our SNPs appears to reflect genomic clustering of outlier loci that distinguish species of the eastern North American white oak syngameon. The causes, consequences, and scale of these genomic islands of differentiation among eastern North American white oaks bear investigation using higher sampling of individuals and loci.

We expect our power to detect complex patterns of introgression in a multispecies hybrid zone to be compromised by the low locus sampling of this SNP toolkit. Nonetheless, our study demonstrates that even without attempting to find hybrids, potentially biasing ourselves against detecting introgression, and even without using the large numbers of loci generally favored for hybridization studies, we are able to identify introgressants involving several pairs of species from a sampling of natural populations (Figs. 5, 6). The fact that we have selected loci to be fixed or nearly fixed within species may aid in detecting first-generation hybrids. At the same time, by selecting genes with high pairwise F_{ST}, we effectively designed our SNPs within outlier loci (by definition, loci with higher-than-expected F_{ST}), which may overestimate divergence between species and underestimate the proportion of the genome that is subject to introgression. The pairs that we found to be admixed at the 10% level for at least one individual were also found by Hardin to hybridize (Fig. 6; cf. fig. 1 in Hardin, 1975). It remains to be seen using genomic markers that are not subject to the ascertainment bias in our study what the actual frequency and average percent of admixture are for these species.

Conclusions

Oaks have been a bugbear of evolutionary biology since Darwin's time, raising significant questions about what species are and how we can make sense of speciation in the face of ongoing gene flow (Arnold, 2016). Our work builds on studies that, in aggregate, suggest that oak species are genetically coherent across their ranges (Muir et al., 2000; Hipp & Weber, 2008; Cavender-Bares & Pahlich, 2009; Hauser et al., 2017) despite a history of introgression (Eaton et al., 2015; McVay et al., 2017a; Kim et al., 2018). We concur with Hardin (1975: 360), who wrote, "Neither Baranski (1975) nor I agree with Minckler (1965), who thinks that hybridization may mask evidence of races within white oak."

Our study does not, however, speak to the *frequency* of hybridization, because our markers are selected for fixation or near-fixation within species. This bias may afford the markers increased utility to identify early-generation hybrids but make them poor estimators of genome-wide rates of genetic exchange. It is important to note, in fact, that we could have told the story of introgression with a different handpicked set of 20 or 80 SNPs: different regions of the genome—i.e., regions that reflect either gene flow or population divergence history more strongly—will suggest different histories of speciation and introgression. Both histories are embedded in the genome, and both are equally real. We cannot consequently assess Muller's (1952: 148) claim that "the bulk of claims of hybridity [in *Quercus*] are based upon trivial variations of the sort one may encounter in a relatively pure population of a single species." What we can say is that the eastern North American white oak syngameon is composed of entities that most taxonomists would consider "good species."

It is equally important to note that while our study demonstrates that there exist loci that distinguish species in the white oak syngameon across their ranges, it leaves open the question of *which* regions of the genome are responsible for species cohesion in oaks. As increasing evidence suggests that forest tree syngameons may be common, especially in the tropics (Cannon & Lerdau, 2015; Caron et al., 2019; Kenzo et al., 2019), the forces shaping how and the degree to which different regions of the genome capture different aspects of population divergence and gene flow history will be a central question—perhaps *the* central question—of tree biodiversity for the coming decade.

Literature Cited

Aldrich, P. R., G. R. Parker, C. H. Michler & J. Romero-Severson. 2003. Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest. Canad. J. Forest Res. 33: 2228–2237.

Allard, H. A. 1932. A progeny study of the so-called oak species *Quercus saulii*, with notes on other probable hybrids found in or near the District of Columbia. Bull. Torrey Bot. Club 59: 267–277.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers & D. J. Lipman. 1990. Basic local alignment search tool. J. Molec. Biol. 215: 403–410.

Anderson, E. 1949. Introgressive Hybridization. John Wiley & Sons, Inc., New York.

Arnold, M. L. 2016. Divergence with Genetic Exchange. Oxford University Press, Oxford.

Baranski, M. J. 1975. An Analysis of Variation within White Oak (*Quercus alba* L.). North Carolina Agricultural Experiment Station, Raleigh.

Barker, M. J. 2007. The empirical inadequacy of species cohesion by gene flow. Philos. Sci. 74: 654–665.

Barker, M. J. & R. A. Wilson. 2010. Cohesion, gene flow, and the nature of species. J. Philos. 107: 61–79.

Becker, R. A., A. R. Wilks, R. Brownrigg, T. P. Minka & A. Deckmyn. 2018. maps: Draw Geographical Maps, version 3.3.0. <https://CRAN.R-project.org/package=maps>, accessed 13 August 2019.

Belahbib, N., M. H. Pemonge, A. Ouassou, H. Sbay, A. Kremer & R. J. Petit. 2001. Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. Molec. Ecol. 10: 2003–2012.

Bradić, M., J. Costa & I. M. Chelo. 2012. Genotyping with Sequenom. Pp. 193–210 *in* V. Orgogozo & M. V. Rockman (editors), Molecular Methods for Evolutionary Genetics. Humana Press, Totowa, New Jersey.

Browne, J. 2010. Asa Gray and Charles Darwin: Corresponding naturalists. Harvard Pap. Bot. 15: 209–220.

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

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer & T. L. Madden. 2009. BLAST+: Architecture and applications. BMC Bioinformatics 10: 421.

Cannon, C. H. & M. Lerdau. 2015. Variable mating behaviors and the maintenance of tropical biodiversity. Frontiers Genet. 6: 183.

Caron, H., J. F. Molino, D. Sabatier, P. Léger, P. Chaumeil, C. Scotti-Saintagne, J. M. Frigério, et al. 2019. Chloroplast DNA variation in a hyperdiverse tropical tree community. Ecology and Evolution. <https://doi.org/10.1002/ece3.5096>, accessed 30 July 2019.

de Casas, R. R., E. Cano, L. Balaguer, E. Pérez-Corona, E. Manrique, C. García-Verdugo & P. Vargas. 2007. Taxonomic identity of *Quercus coccifera* L. in the Iberian Peninsula is maintained in spite of widespread hybridisation, as revealed by morphological, ISSR and ITS sequence data. Flora, Morphol. Distrib. Funct. Ecol. Pl. 202: 488–499.

Cavender-Bares, J. & A. Pahlich. 2009. Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). Amer. J. Bot. 96: 1690–1702.

Chybicki, I. J. & J. Burczyk. 2010. Realized gene flow within mixed stands of *Quercus robur* L. and *Q. petraea* (Matt.) L. revealed at the stage of naturally established seedling. Molec. Ecol. 19: 2137–2151.

Craft, K. J. & M. V. Ashley. 2006. Population differentiation among three species of white oak in northeastern Illinois. Canad. J. Forest Res. 26: 206–215.

Crowl, A. A., P. S. Manos, J. D. McVay, A. R. Lemmon, E. M. Lemmon & A. L. Hipp. 2019. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). New Phytol. doi:10.1111/nph.15842.

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

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

Dumolin-Lapegue, S. A. K. & R. J. Petit. 1999. Are chloroplast and mitochondrial DNA variation species independent in oaks? Evolution 53: 1406–1413.

Dumolin-Lapegue, S., B. Demesure, S. Fineschi, V. L. Come & R. J. Petit. 1997. Phylogeographic structure of white oaks throughout the European continent. Genetics 146: 1475–1487.

Earl, D. A. & B. M. vonHoldt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet. Resources 4: 359–361.

Eaton, D. A. R., A. L. Hipp, A. González-Rodríguez & J. Cavender-Bares. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution 69: 2587–2601.

Ehrlich, P. R. & P. H. Raven. 1969. Differentiation of populations: Gene flow seems to be less important in speciation than the neo-Darwinians thought. Science 165: 1228–1232.

Engelmann, G. 1876. The oaks of the United States. Trans. Acad. Sci. St. Louis 3: 539–543.

Evanno, G., S. Regnaut & J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: A simulation study. Molec. Ecol. 14: 2611–2620.

Excoffier, L., P. E. Smouse & J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131: 479–491.

Felsenstein, J. 2004. Inferring Phylogenies. Sinauer Associates, Inc., Sunderland, Massachusetts.

Fitzek, E., A. Delcamp, E. Guichoux, M. Hahn, M. Lobdell & A. L. Hipp. 2018. A nuclear DNA barcode for eastern North American oaks and application to a study of hybridization in an arboretum setting. Ecol. Evol. 8: 5837–5851.

Gabriel, S., L. Ziaugra & D. Tabbaa. 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. Curr. Protoc. Human Genet. 60: 2.12.1–2.12.18.

Gailing, O. & A. L. Curtu. 2014. Interspecific gene flow and maintenance of species integrity in oaks. Ann. Forest Res. 57: 5–18.

Goicoechea, P. G., A. Herran, J. Durand, C. Bodenes, C. Plomion & A. Kremer. 2015. A linkage disequilibrium perspective on the genetic mosaic of speciation in two hybridizing Mediterranean white oaks. Heredity 114: 373–386.

Gömöry, D. & J. Schmidtová. 2007. Extent of nuclear genome sharing among white oak species (*Quercus* L. subgen. *Lepidobalanus* (Endl.) Oerst.) in Slovakia estimated by allozymes. Pl. Syst. Evol. 266: 253–264.

Gray, A. 1857. Manual of the Botany of the Northern United States. Including Virginia, Kentucky, and all East of the Mississippi: Arranged According to the Natural System. G. P. Putnam & Co., New York.

Gray, A. 1859. Manual of the Botany of the Northern United States: Including Virginia, Kentucky, and all East of the Mississippi Arranged According to the Natural System. Ivison & Phinney, New York.

Gray, A. 1862. Manual of the Botany of the Northern United States, 3rd revised edition. Ivison, Phinney & Co., Chicago.

Gray, A. 1867. Manual of the Botany of the Northern United States, Including the District East of the Mississippi and North of North Carolina and Tennessee, Arranged According to the Natural System. Ivison, Blakeman, Taylor & Co., New York.

Gray, A. & W. S. Sullivant. 1848. A Manual of the Botany of the Northern United States, from New England to Wisconsin and South to Ohio and Pennsylvania Inclusive (the Mosses and Liverworts by Wm. S. Sullivant) Arranged According to the Natural System. J. Munroe, J. Chapman, Boston and London.

Guichoux, E., P. Garnier-Géré, L. Lagache, T. Lang, C. Boury & R. J. Petit. 2013. Outlier loci highlight the direction of introgression in oaks. Molec. Ecol. 22: 450–462.

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

Hauser, D. A., A. Keuter, J. D. McVay, A. L. Hipp & P. S. Manos. 2017. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). Amer. J. Bot. 104: 1581–1595.

Hipp, A. L. 2015. Should hybridization make us skeptical of the oak phylogeny? Int. Oak J. 26: 9–18.

Hipp, A. L. 2018. Pharaoh's dance: The oak genomic mosaic. PeerJ Preprints 6: E27405v1.

Hipp, A. L. & J. A. Weber. 2008. Taxonomy of Hill's oak (*Quercus ellipsoidalis*: Fagaceae): Evidence from AFLP data. Syst. Bot. 33: 148–158.

Hipp, A. L., P. S. Manos, A. González-Rodríguez, M. Hahn, M. Kaproth, J. D. McVay, S. V. Avalos, et al. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. New Phytol. 217: 439–452.

Hipp, A. L., P. S. Manos, M. Hahn, M. Avishai, C. Bodénès, J. Cavender-Bares, A. Crowl, et al. 2019. Genomic landscape of the global oak phylogeny. bioRxiv <https://doi.org/10.1101/587253>, accessed 30 July 2019.

Jakobsson, M. & N. A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806.

Kahle, D. & H. Wickham. 2013. ggmap: Spatial visualization with ggplot2. R J. 5: 144–161.

Kenzo, T., K. Kamiya, K. M. Ngo, N. Faizu, S. K. Y. Lum, S. Igarashi, Y. Norichika, et al. 2019. Overlapping flowering periods among *Shorea* species and high growth performance of hybrid seedlings promote hybridization and introgression in a tropical rainforest of Singapore. Forest Ecol. Managem. 435: 38–44.

Kim, B. Y., X. Wei, S. Fitz-Gibbon, K. E. Lohmueller, J. Ortego, P. F. Gugger & V. L. Sork. 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). Molec. Ecol. 27: 4556–4571.

Kremer, A. & A. L. Hipp. Oaks: An evolutionary success story. New Phytol. (in press).

Leroy, T., C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. P. Paiva, C. Dossat, et al. 2017. Extensive recent secondary contacts between four European white oak species. New Phytol. 214: 865–878.

Leroy, T., Q. Rougemont, J. L. Dupouey, C. Bodenes, C. Lalanne, C. Belser, K. Labadie, et al. 2018. Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. New Phytol. <https://doi.org/10.1111/nph.16039>, accessed 30 July 2019.

Lexer, C., A. Kremer & R. J. Petit. 2006. Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. Molec. Ecol. 15: 2007–2012.

Lind-Riehl, J. F., A. R. Sullivan & O. Gailing. 2014. Evidence for selection on a CONSTANS-like gene between two red oak species. Ann. Bot. 113: 967–975.

Little, E. L. 1971. Conifers and Important Hardwoods, Vol. 1: Atlas of United States Trees. U.S. Department of Agriculture, Forest Service, Washington, D.C.

Little, E. L. 1977. Minor Eastern Hardwoods, Vol. 4: Atlas of United States Trees. U.S. Department of Agriculture, Forest Service, Washington, D.C.

Little, E. L. 1979. Checklist of United States Trees (Native and Naturalized). U.S. Department of Agriculture, Washington, D.C.

Manos, P. S., J. J. Doyle & K. C. Nixon. 1999. Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). Molec. Phylogen. Evol. 12: 333–349.

McVay, J. D., D. Hauser, A. L. Hipp & P. S. Manos. 2017a. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. Genome 60: 733–742.

McVay, J. D., A. L. Hipp & P. S. Manos. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. Proc. Biol. Sci. 284: 20170300.

Minckler, L. S. 1965. White oak (*Quercus alba* L.). Pp. 631–637 *in* Silvics of Forest Trees of the United States. Agriculture Handbook 271. U.S. Department of Agriculture, Washington, D.C.

Moran, E. V., J. Willis & J. S. Clark. 2012. Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae). Amer. J. Bot. 99: 92–100.

Morjan, C. L. & L. H. Rieseberg. 2004. How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. Molec. Ecol. 13: 1341–1356.

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

Muir, G. & C. Schlötterer. 2006. Moving beyond single-locus studies to characterize hybridization between oaks (*Quercus* spp.). Molec. Ecol. 15: 2301–2304.

Muir, G., C. C. Fleming & C. Schlötterer. 2000. Species status of hybridizing oaks. Nature (London) 405: 1016.

Muller, C. H. 1952. Ecological control of hybridization in *Quercus*: A factor in the mechanism of evolution. Evolution 6: 147–161.

Ness, H. 1918. Hybrids of the live oak and overcup oak. J. Heredity 9: 263–268.

Nixon, K. C. 1997. *Quercus*. *In* Flora of North America Editorial Committee (editors), Flora of North America North of Mexico, Vol. 3. Oxford University Press, New York.

Oney-Birol, S., S. Fitz-Gibbon, J. M. Chen, P. F. Gugger & V. L. Sork. 2018. Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (*Quercus* sect. *Quercus*). BMC Genet. 19: 88.

Owusu, S. A., A. R. Sullivan, J. A. Weber, A. L. Hipp & O. Gailing. 2015. Taxonomic relationships and gene flow in four North American *Quercus* species (*Quercus* section *Lobatae*). Syst. Bot. 40: 510–521.

Palmer, E. J. 1948. Hybrid oaks of North America. J. Arnold Arbor. 29: 1–48.

Petit, R., E. Pineau, B. Demesure, R. Bacilieri, A. Ducousso & A. Kremer. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. Proc. Natl. Acad. Sci. U.S.A. 94: 9996–10001.

Petit, R., C. Bodenes, A. Ducousso, G. Roussel & A. Kremer. 2003. Hybridization as a mechanism of invasion in oaks. New Phytol. 161: 151–164.

Pham, K. K., A. L. Hipp, P. S. Manos & R. C. Cronn. 2017. A time and a place for everything: Phylogenetic history and geography as joint predictors of oak plastome phylogeny. Genome 60: 720–732.

Plomion, C., J. M. Aury, J. Amselem, T. Leroy, F. Murat, S. Duplessis, S. Faye, et al. 2018. Oak genome reveals facets of long lifespan. Nat. Pl. 4: 440–452.

Prasad, A. M. & L. R. Iverson. 2003. Little's Range and FIA Importance Value Database for 135 Eastern US Tree

Species. Northeastern Research Station, USDA Forest Service, Delaware, Ohio.

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

R-Development-Core-Team. 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rieseberg, L. H., T. E. Wood & E. J. Baack. 2006. The nature of plant species. Nature (London) 440: 524–527.

Rosenberg, N. A. 2004. DISTRUCT: A program for the graphical display of population structure. Molec. Ecol. Notes 4: 137–138.

Scotti-Saintagne, C., S. Mariette, I. Porth, P. G. Goicoechea, T. Barreneche, C. Bodenes, K. Burg, et al. 2004. Genome scanning for interspecific differentiation between two closely related oak species. [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.] Genetics 168: 1615–1626.

Sokal, R. R. & C. D. Michener. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. 38: 1409–1438.

Sullivan, A. R., S. A. Owusu, J. A. Weber, A. L. Hipp & O. Gailing. 2016. Hybridization and divergence in multispecies oak (*Quercus*) communities. Bot. J. Linn. Soc. 181: 99–114.

Tovar-Sánchez, E. & K. Oyama. 2004. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: Morphological and molecular evidence. Amer. J. Bot. 91: 1352–1363.

Urbanek, S. 2012. proj4: A simple interface to the PROJ.4 cartographic projections library.

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

Whittemore, A. T. & B. A. Schaal. 1991. Interspecific gene flow in sympatric oaks. Proc. Natl. Acad. Sci. U.S.A. 88: 2540–2544.

Wickham, H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Wiegand, K. M. 1935. A taxonomist's experience with hybrids in the wild. Science 81: 161–166.

Yarnell, S. H. & E. J. Palmer. 1933. Inheritance in an oak species hybrid. J. Arnold Arbor. 14: 68–75.

ISSN 0026-6493 (PRINT); ISSN 2162-4372 (ONLINE)