bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Conserved DNA polymorphisms distinguish species in the eastern North American white oak syngameon: Insights from an 80-SNP oak DNA genotyping toolkit [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], Charles H. Cannon[2]

[1] This paper is based on a talk presented by the first author at the Missouri Botanical Garden 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 ALH and ATW. 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, IL 60532, USA

[3] The Field Museum, Department of Botany, Chicago, IL 60605, USA

[4] U.S. National Arboretum, Washington, DC 20002, USA

[5] Universität Bielefeld, Department of Computational Biology, 33615 Bielefeld, GERMANY

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

[7] University of Minnesota, Minneapolis MN 55455, USA

[8] Maryland Center for Environmental Science, Appalachian Laboratory, Frostburg, MD 21532, USA

[9] Duke University, Durham NC 27708, USA

[10] U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO 80526, USA

[*] Author for correspondence: ahipp@mortonarb.org

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 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 rangewide 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 base pairs (± 8.74 million bp, s.d.), 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 an important implication: the eastern North American white oak syngameon is composed of entities that most taxonomists would consider "good species," and species in the syngameon retain their genetic cohesion because characteristic portions of the genome do not become homogenized despite a history of introgression.

Keywords. Cohesion species; DNA genotyping toolkit; hybridization; introgression; *Quercus alba*, *Quercus bicolor*; *Quercus macrocarpa*; *Quercus stellata*; single nucleotide polymorphism (SNP); syngameon

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 *Quercus virginiana* (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 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, A., & Petit, 1999; Petit *et al.*, 2003; Dodd & Afzal-Rafii, 2004; Tovar-Sánchez & Oyama, 2004; Craft & Ashley, 2006; Lexer, Kremer, & Petit,

[1] The history of Gray's reports of hybrids is instructive. The first edition (Gray & Sullivant, 1848) included two hybrids in the genus *Quercus*, both 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.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2006; Curtu, Gailing, & Finkeldey, 2007; Hipp & Weber, 2008; Chybicki & Burczyk, 2010; Moran, Willis, & Clark, 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, Accepted pending revision). 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, Doyle, & Nixon, 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, Fleming, & Schlötterer, 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, Hipp, & Manos, 2017b; Hipp *et al.*, 2018) to develop a low-cost SNP genotyping kit for eastern North American white oaks (Fitzek *et al.*, 2018). We demonstrated our 80-marker SNP kit to be effective for 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. In the current study, we test this marker set in natural populations across a rangewide 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 genomewide sample of molecular markers, laying the groundwork for future studies of introgression and species cohesion in the group.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

MATERIALS AND METHODS

Sampling and genotyping

Data were initially collected from 184 individuals of seven eastern North American white oak species, collected from 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 SNPs were consequently not designed to separate these two (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 reflect the fact 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 7–9 in *Quercus montana* Willd. and *Q. michauxii* Nutt. to 38–52 in *Q. muehlenbergii / prinoides* and *Q. macrocarpa* 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 *Quercus 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 range maps for *Q. macrocarpa*, the most wide-ranging species in our study; *Q. bicolor*, the most widespread northern species; and *Q. stellata*, the most widespread southern species. 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.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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), which is optimized for MassARRAY analysis (Bradić, Costa, & Chelo, 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, 5 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 chromsome, 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 *Quercus robur* chromosome 12. The two SNPs identified in this RAD-seq were designed to distinguish *Quercus 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 chromosomes 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

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 any proportion of the genome of individuals of a putative species that is shared with individuals of other putative species. This operational definition corresponds with practices widely used by plant systematists to define "good species" (Rieseberg, Wood, & Baack, 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 AMOVA (Excoffier, Smouse, & Quattro, 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.

Criterion 2 we assess using the Bayesian population genetic clustering algorithm implemented in STRUCTURE v 2.3.4 (Pritchard, Stephens, & Donnelly, 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 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 ΔK, but given the problematic nature of identifying K with hierarchical data, we report the structures 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).

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 *Q. montana* to N = 52 in *Q. macrocarpa*), the resulting heatmap (Fig. 2) and the table underlying it (Supplemental 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 (s.d.) loci, and missing data per locus averaged 14.6% ± 26.8 (s.d.) 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% 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 7 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_OAKMOR340_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 *Q. alba*, two in *Q. michauxii*, four in *Q. montana*—and three exhibit a single SNP in *Q. muehlenbergii / prinoides.* An additional ten SNPs exhibit a summed proportion between 0.95 and 1.05, suggesting relatively high decisiveness for *Q. stellata* (2 SNPs) and *Q. bicolor* (3 SNPs). Based on these, we hand-picked 20 SNPs that suffice to diagnose the species in our study (Fig. 2, red bars along left edge).

Using all loci, the UPGMA (Fig. 3a) 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

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 hand-picked loci, our SNP genotyping toolkit successfully distinguishes species from one another using UPGMA (Fig. 3b).

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 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 7 species pairs yielding individuals admixed 10% or more based on our markers (Figs. 5, 6). Notably, it is not until $K = 8$ that the 7 species are distinguished from each other, perhaps due to high genetic variation within species that is not adequately 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 artefactual, 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% (Supplemental 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 eleven do not map to any location in the genome (Table 3; Supplemental 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 twelve *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 averages 7.47 million bp (± 8.74 million bp, s.d.). These are all significantly clustered relative to a random draw of SNPs, under which only 0.909 interlocus

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 eleven 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 these apparently-fixed SNPs are limited to our species with smallest sample size—one in *Q. alba* (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.

It is somewhat remarkable that we are able to distinguish seven interbreeding oak species with just 20 hand-picked markers. By comparison, the now-classic study demonstrating genetic distinctiveness of *Q. 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 fifteen variable microsatellites (Aldrich *et al.*, 2003) have by contrast failed to find consistent genetic differentiation between two to three co-occuring 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 27–28 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, Sullivan, & Gailing, 2014b; Sullivan *et al.*, 2016). This suggests that a moderate-sized but random sample of loci will often reflect regions of the genome that are either not yet differentiated between species (Muir & Schlötterer, 2005, 2006) or 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.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 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 differentiate species that are exchanging genes more frequently across the remainder of the genome (Lind-Riehl, Sullivan, & Gailing, 2014a; Gailing & Curtu, 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 disequilbrium (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, sited 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 (only 20 selected SNPs). Nonetheless, our study demonstrates that even without attempting to find hybrids, potentially biasing ourselves against detecting introgression, and even without employing 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-exected F_{ST}), which

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 is for these species.

CONCLUSIONS

Oaks have been a bugbear of systematics 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), 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 hand-picked set of 20 or 80 SNPs: the "right" regions of the genome—by which we mean those regions that favor one particular gene-flow / genetic coherence process over another—will tell one story or the other. Both stories are embedded in the genome, and both are equally real. We cannot consequently assess Muller's (1952) 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 (Caron *et al.*; Cannon & Lerdau, 2015; Kenzo *et al.*, 2019), the forces shapping 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.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

LITERATURE CITED

Aldrich PR, Parker GR, Michler CH & Romero-Severson J. **2003**. Whole-tree silvic identifications and the microsatellite genetic structure of a red oak species complex in an Indiana old-growth forest. *Canadian Journal of Forest Research* **33**: 2228–2237.

Allard HA. **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. *Bulletin of the Torrey Botanical Club* **59**: 267–277.

Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ. **1990**. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403–410.

Anderson E. **1949**. *Introgressive hybridization*. New York: John Wiley & Sons, Inc.

Arnold ML. **2016**. *Divergence With Genetic Exchange*. Oxford: Oxford University Press.

Baranski MJ. **1975**. *An analysis of variation within white oak (*Quercus alba *L.)*. Raleigh: North Carolina Agricultural Experiment Station.

Barker MJ. **2007**. The Empirical Inadequacy of Species Cohesion by Gene Flow. *Philosophy of Science* **74**: 654–665.

Barker MJ & Wilson RA. **2010**. Cohesion, Gene Flow, and the Nature of Species. *The Journal of Philosophy* **107**: 61–79.

Becker RA, Wilks AR, Brownrigg R, Minka TP & Deckmyn A. **2018**. *maps: Draw Geographical Maps*.

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

Bradić M, Costa J & Chelo IM. **2012**. Genotyping with Sequenom. In: *Molecular Methods for Evolutionary Genetics*. Humana Press, 193–210.

Browne J. **2010**. Asa Gray and Charles Darwin: Corresponding Naturalists. *Harvard Papers in Botany* **15**: 209–220.

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

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

Cannon CH & Lerdau M. **2015**. Variable mating behaviors and the maintenance of tropical biodiversity. *Frontiers in Genetics* **6**.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Caron H, Molino JF, Sabatier D, Léger P, Chaumeil P, Scotti Saintagne C, Frigério JM, Scotti I, Franc A & Petit RJ. Chloroplast DNA variation in a hyperdiverse tropical tree community. *Ecology and Evolution* **0**.

de Casas RR, Cano E, Balaguer L, Pérez-Corona E, Manrique E, García-Verdugo C & Vargas P. **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 - Morphology, Distribution, Functional Ecology of Plants* **202**: 488–499.

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

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

Craft KJ & Ashley MV. **2006**. Population differentiation among three species of white oak in northeastern Illinois. *Canadian Journal of Forest Research* **26**: 206–215.

Crowl AA, Manos PS, McVay JD, Lemmon AR, Lemmon EM & Hipp AL. **2019**. Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*). *New Phytologist* **doi:10.1111/nph.15842**.

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

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

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

Dumolin-Lapegue S, A. K & Petit RJ. **1999**. Are Chloroplast and Mitochondrial DNA Variation Species Independent in Oaks? *Evolution* **53**: 1406–1413.

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Engelmann G. **1876**. The oaks of the United States. *Transactions of the Academy of Sciences St. Louis* **3**: 539–543.

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

Excoffier L, Smouse PE & Quattro JM. **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*. Sunderland, Maryland: Sinauer Associates, Inc.

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

Gabriel S, Ziaugra L & Tabbaa D. **2009**. SNP Genotyping Using the Sequenom MassARRAY iPLEX Platform. *Current Protocols in Human Genetics* **60**: 2.12.1-2.12.18.

Gailing O & Curtu AL. **2014**. Interspecific gene flow and maintenance of species integrity in oaks. *Annals of Forest Research* **57**: 5-18–18.

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

Gömöry D & Schmidtová J. **2007**. Extent of nuclear genome sharing among white oak species (Quercus L. subgen. Lepidobalanus (Endl.) Oerst.) in Slovakia estimated by allozymes. *Plant Systematics and Evolution* **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*. New York: G. P. Putnam & Co.

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*. New York: Ivison & Phinney.

Gray A. **1862**. *Manual of the botany of the northern United States, third revised edition*. Chicago: Ivison, Phinney, & Co.

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*. New York: Ivison, Blakeman, Taylor & Co.

Gray A & Sullivant WS. **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*

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by Wm. S. Sullivant) arranged according to the natural system. Boston and London: J. Munroe, J.Chapman.

Guichoux E, Garnier-Géré P, Lagache L, Lang T, Boury C & Petit RJ. **2013**. Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology* **22**: 450–462.

Hardin JW. **1975**. Hybridization and introgression in *Quercus alba*. *Journal of the Arnold Arboretum* **56**: 336–363.

Hauser DA, Keuter A, McVay JD, HIpp AL & Manos PS. **2017**. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). *American Journal of Botany* **104**: 1581–1595.

Hipp AL. **2015**. Should hybridization make us skeptical of the oak phylogeny? *International Oak Journal* **26**: 9–18.

Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Avalos SV & Cavender-Bares J. **2018**. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist* **217**: 439–452.

Hipp AL. **2018**. Pharaoh's Dance: the oak genomic mosaic. *PeerJ Preprints* **6**: e27405v1.

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl A, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Sork VL & Valencia-Avalos S. **2019**. Genomic landscape of the global oak phylogeny. *bioRxiv*: 587253.

Hipp AL & Weber JA. **2008**. Taxonomy of Hill's Oak (*Quercus ellipsoidalis*: Fagaceae): Evidence from AFLP Data. *Systematic Botany* **33**: 148–158.

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

Kahle D & Wickham H. **2013**. ggmap: Spatial Visualization with ggplot2. *The R Journal* **5**: 144–161.

Kenzo T, Kamiya K, Ngo KM, Faizu N, Lum SKY, Igarashi S, Norichika Y & Ichie T. **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 Ecology and Management* **435**: 38–44.

Kim BY, Wei X, Fitz□Gibbon S, Lohmueller KE, Ortego J, Gugger PF & Sork VL. **2018**. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Molecular Ecology* **27**: 4556–4571.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kremer A & Hipp AL. **Accepted pending revision**. Oaks: an evolutionary success story. *New Phytologist*.

Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, Paiva JAP, Dossat C, Aury JM, Plomion C & Kremer A. **2017**. Extensive recent secondary contacts between four European white oak species. *New Phytologist* **214**: 865–878.

Leroy T, Rougemont Q, Dupouey JL, Bodenes C, Lalanne C, Belser C, Labadie K, Provost GL, Aury JM, Kremer A & Plomion C. **2018**. Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *bioRxiv*: 246637.

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

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

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

Little EL. **1971**. *Atlas of United States trees, volume 1, Conifers and important hardwoods*. Washington, D.C.□: U.S. Dept. of Agriculture, Forest Service.

Little EL. **1977**. *Atlas of United States trees, volume 4, Minor eastern hardwoods*. Washington, D.C.□: U.S. Dept. of Agriculture, Forest Service.

Little EL. **1979**. *Checklist of United States trees (native and naturalized).* Washington, D.C.: U.S. Department of Agriculture.

Manos PS, Doyle JJ & Nixon KC. **1999**. Phylogeny, Biogeography, and Processes of Molecular Differentiation in Quercus Subgenus Quercus (Fagaceae). *Molecular Phylogenetics and Evolution* **12**: 333–349.

McVay JD, Hauser D, Hipp AL & Manos PS. **2017a**. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* **60**: 733–742.

McVay JD, Hipp AL & Manos PS. **2017b**. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc. R. Soc. B* **284**: 20170300.

Minckler LS. **1965**. White oak (*Quercus alba* L.). *Silvics of forest trees of the United States* **271**: 631–637.

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

Morjan CL & Rieseberg LH. **2004**. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**: 1341–1356.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

Muller CH. **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. *Journal of Heredity* **9**: 263–268.

Oney-Birol S, Fitz-Gibbon S, Chen JM, Gugger PF & Sork VL. **2018**. Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (*Quercus* sect. *Quercus*). *BMC Genetics* **19**: 88.

Owusu SA, Sullivan AR, Weber JA, Hipp AL & Gailing O. **2015**. Taxonomic Relationships and Gene Flow in Four North American Quercus Species (Quercus section Lobatae). *Systematic Botany* **40**: 510–521.

Palmer EJ. **1948**. Hybrid oaks of North America. *Journal of the Arnold Arboretum* **29**: 1–48.

Petit R, Pineau E, Demesure B, Bacilieri R, Ducousso A & Kremer A. **1997**. Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA* **94**: 9996–10001.

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

Pham KK, Hipp AL, Manos PS & Cronn RC. **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, Aury JM, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N, Labadie K, Provost GL, Lesur I, Bartholomé J, Faivre-Rampant P, Kohler A, Leplé JC, Chantret N, Chen J, Diévart A, Alaeitabar T, Barbe V, Belser C, Bergès H, Bodénès C, Bogeat-Triboulot MB, Bouffaud ML, Brachi B, Chancerel E, Cohen D, Couloux A, Silva CD, Dossat C, Ehrenmann F, Gaspin C, Grima-Pettenati J, Guichoux E, Hecker A, Herrmann S, Hugueney P, Hummel I, Klopp C, Lalanne C, Lascoux M, Lasserre E, Lemainque A, Desprez-Loustau ML, Luyten I, Madoui MA, Mangenot S, Marchal C, Maumus F, Mercier J, Michotey C, Panaud O, Picault N, Rouhier N, Rué O, Rustenholz C, Salin F, Soler M, Tarkka M, Velt A, Zanne AE, Martin F, Wincker P, Quesneville H, Kremer A & Salse J. **2018**. Oak genome reveals facets of long lifespan. *Nature Plants* **4**: 440–452.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prasad AM & Iverson LR. **2003**. *Little's range and FIA importance value database for 135 eastern US tree species*. Delaware, Ohio: Northeastern Research Station, USDA Forest Service.

Pritchard JK, Stephens M & Donnelly P. **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*. Vienna.

Rieseberg LH, Wood TE & Baack EJ. **2006**. The nature of plant species. *Nature (London)* **440**: 524–527.

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

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes C, Burg K & Kremer A. **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 RR & Michener CD. **1958**. A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin* **38**: 1409–1438.

Sullivan AR, Owusu SA, Weber JA, Hipp AL & Gailing O. **2016**. Hybridization and divergence in multi-species oak (*Quercus*) communities. *Botanical Journal of the Linnean Society* **181**: 99–114.

Tovar-Sánchez E & Oyama K. **2004**. Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: morphological and molecular evidence. *American Journal of Botany* **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 AT & Schaal BA. **1991**. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences USA* **88**: 2540–2544.

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

Wiegand KM. **1935**. A taxonomist's experience with hybrids in the wild. *Science* **81**: 161–166.

Yarnell SH & Palmer EJ. **1933**. Inheritance in an oak species hybrid. *Journal of the Arnold Arboretum* **14**: 68–75.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE CAPTIONS

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*, for which a base map was not available.

Figure 2. 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 side of the figure indicate SNPs in 20 loci we hand-selected because they were highly decisive for the species represented in the present study.

Figure 3. 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 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).

Figure 4. NMDS ordination, 75 loci. Non-metric multidimensional scaling was conducted in three dimensions for the same Euclidean distance matrix utilized in the UPGMA figure reported above. 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.

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.

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 (his Fig. 1), including only the subset of seven species we investigated in the current study (treating *Q. muehlenbergii* and *Q. prinoides* 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 lines indicate admixture levels of 0.20 or higher for at least one specimen. Vouchers for leaf silhouettes are *Q. alba*: PS Manos 1838 [MOR 177669]; *Q. michauxii*: PS Manos 1843 [MOR 177659]; *Q. bicolor*: PS Manos 1847 [MOR 177662]; *Q. macrocarpa*: IL-MOR-MH108 [MOR 174544]; *Q. stellata*: PS Manos 1835 [MOR 177663]; *Q. muehlenbergii*: PM-98; *Q. montana*: PS Manos 1860 [MOR 177731].

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

[note to editors: Table 1 was provided as PDF and XLSX; please format for inclusion in text]

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 *Q. prinoides*, which was estimated by visual inspection of maps published in Flora of North America (Nixon, 1977).

[note to editors: Table 2 was provided as PDF and XLSX; please format for inclusion in text]

Table 3. Map positions and decisiveness of SNPs that map to a unique position on one of the *Quercus robur* chromosomes. The 60 SNPs that map back to one of the 12 *Quercus 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, abbreviated as follows: '***' indicates a SNP whose decisiveness is exactly 1.000 or 2.000 for the sample studied here (i.e., diagnostic for one or two species); '**' if it is within 0.100 of 1.000 or 2.000; or '*' if it is within 0.200 of 1.000 or 2.000. All loci mapped with identity > 95%, locus length > 70 bp, and E-value $< 1.0 \times 10^{-30}$. 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 (s.d.). Four pairs of loci are < 10,000 bp from one another (indicated by bold italics in the table under "Dist. (bp)" and may bear further investigation as possible 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. Abbreviations: Query (locus) = the RAD-seq locus SNP abbreviation from Fitzek et al. 2018; LG = linkage group (chromosome number), following Plomion et al. 2018; Start (bp) = start position of the RAD-seq locus on the *Q. robur* chromosome; 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; Decisiveness = decisiveness of the SNP for identifying one species or a pair of species (cf. Fig. 2).

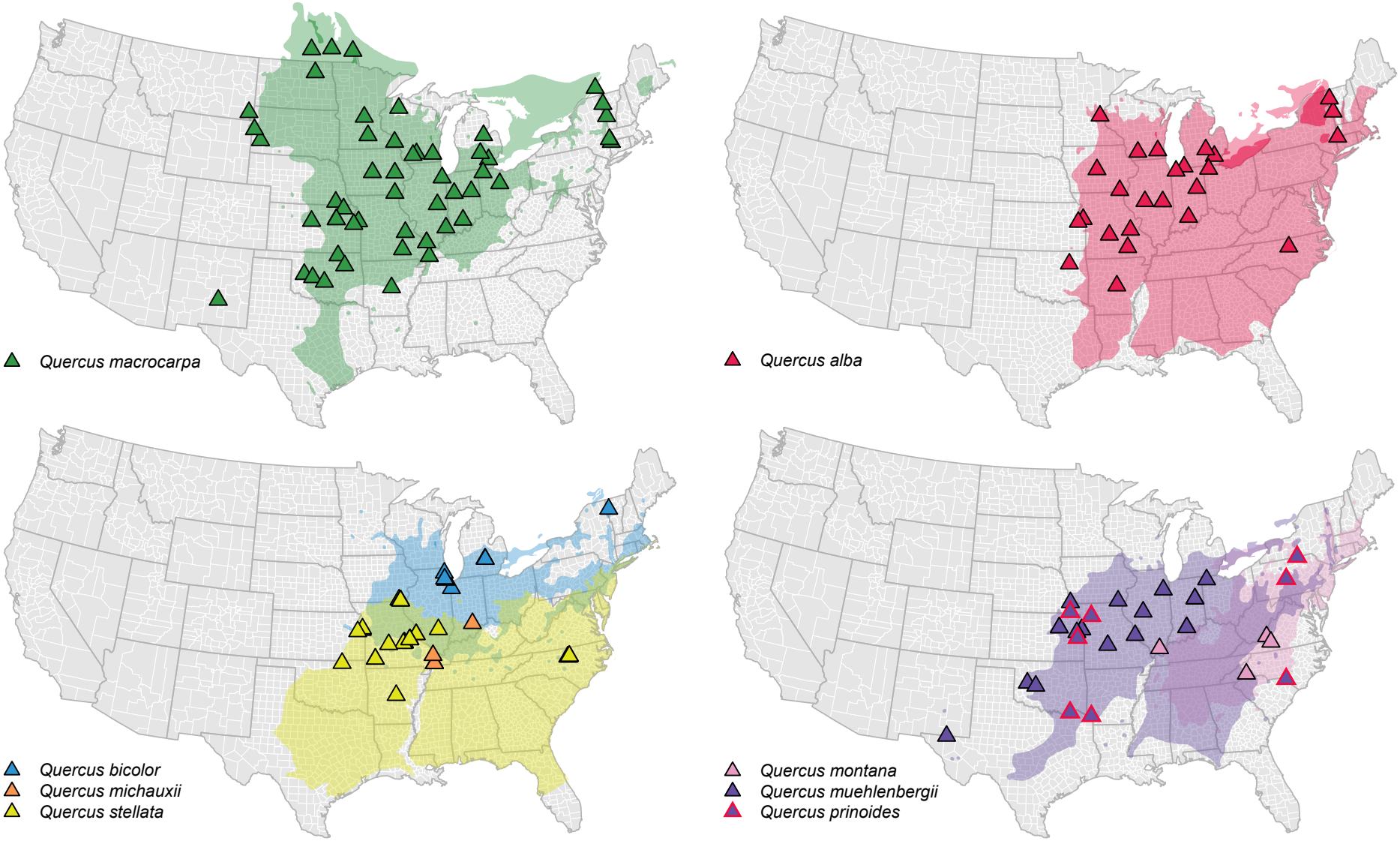
Query (locus)	*LG*	Start (bp)	Dist. (bp)	Decisiveness
locus_11631_48	01	1.32E+07		
locus_17927_52	01	1.92E+07	6.04E+06	*
newl_21880_27	01	2.02E+07	9.71E+05	***

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

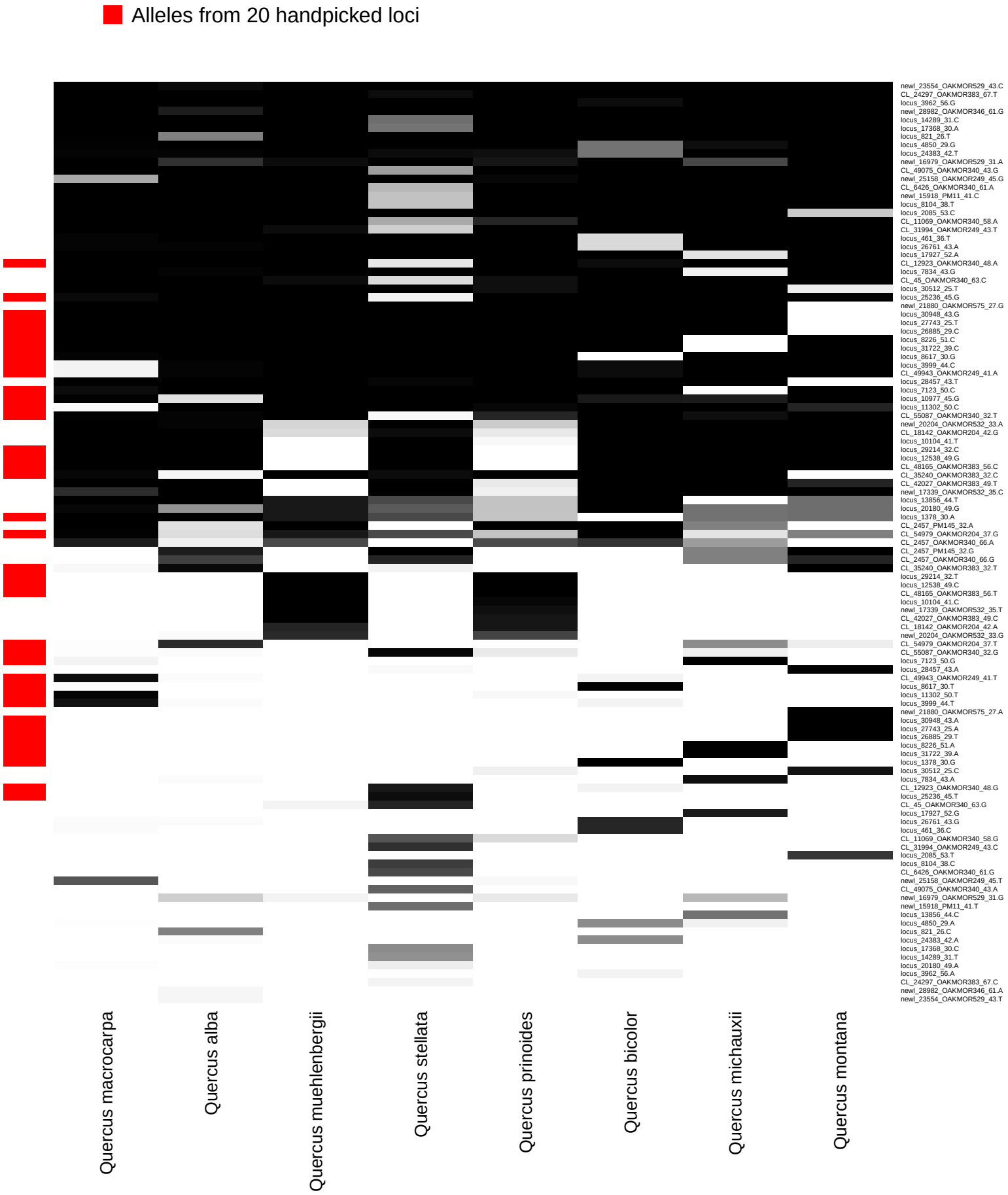
locus_821_26	01	4.61E+07	2.59E+07	
newl_17339_35	01	4.85E+07	2.40E+06	**
CL_42027_	02	9.97E+06		**
CL_6426_61	02	1.46E+07	4.62E+06	
locus_4492_52	02	2.38E+07	9.21E+06	
locus_20180_49	02	2.72E+07	3.37E+06	
locus_3962_56	02	2.72E+07	1.90E+03	
CL_49075_43	02	4.71E+07	1.99E+07	
CL_35240	02	5.65E+07	9.40E+06	**
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	***

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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	

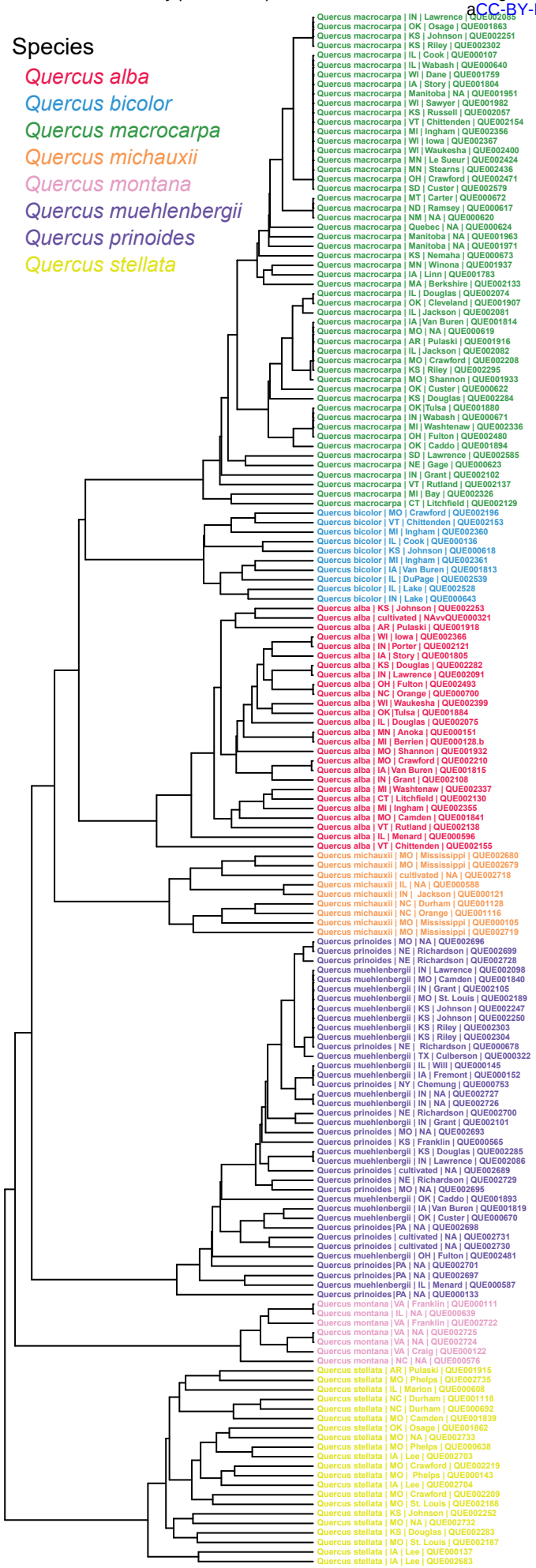

Quercus macrocarpa
Quercus alba
Quercus bicolor
Quercus michauxii
Quercus stellata
Quercus montana
Quercus muehlenbergii
Quercus prinoides

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

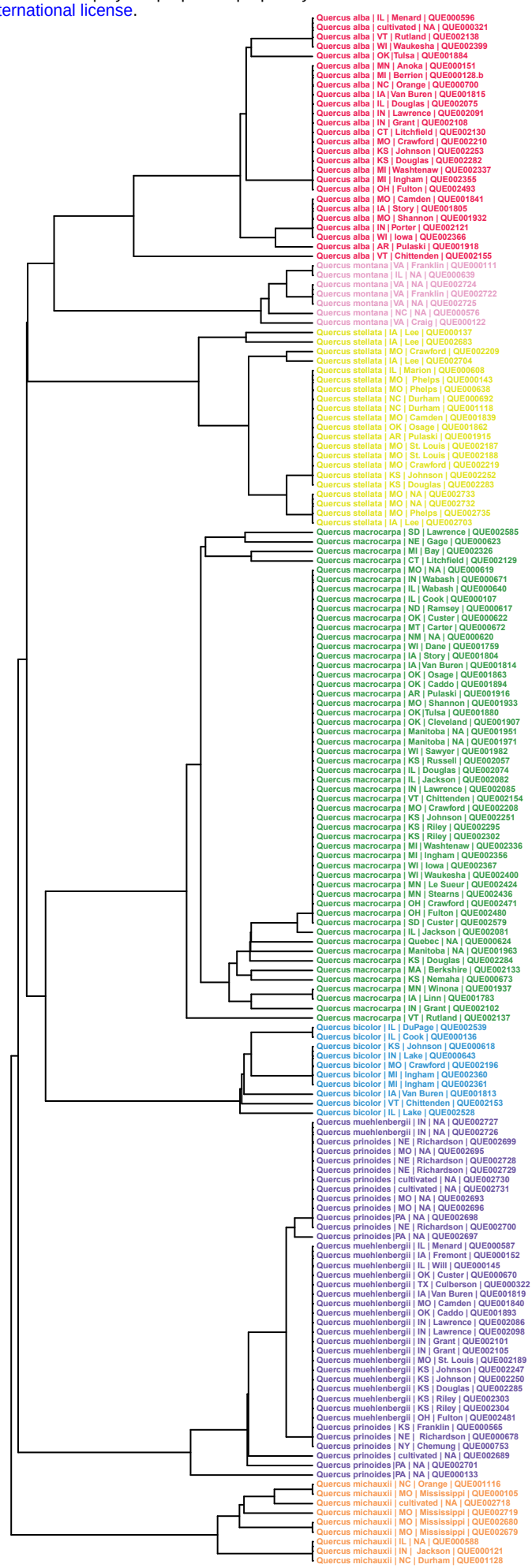


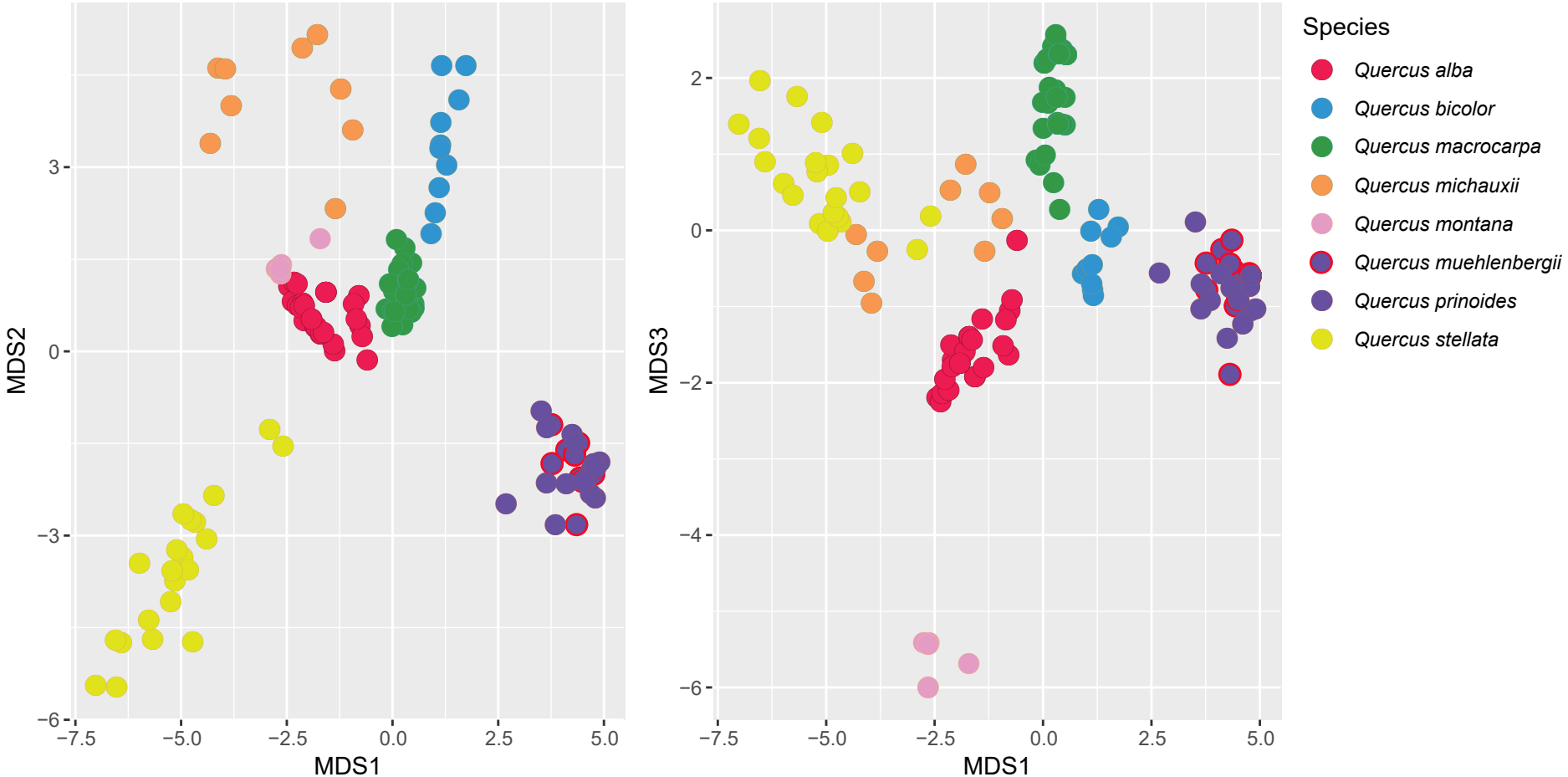
bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A



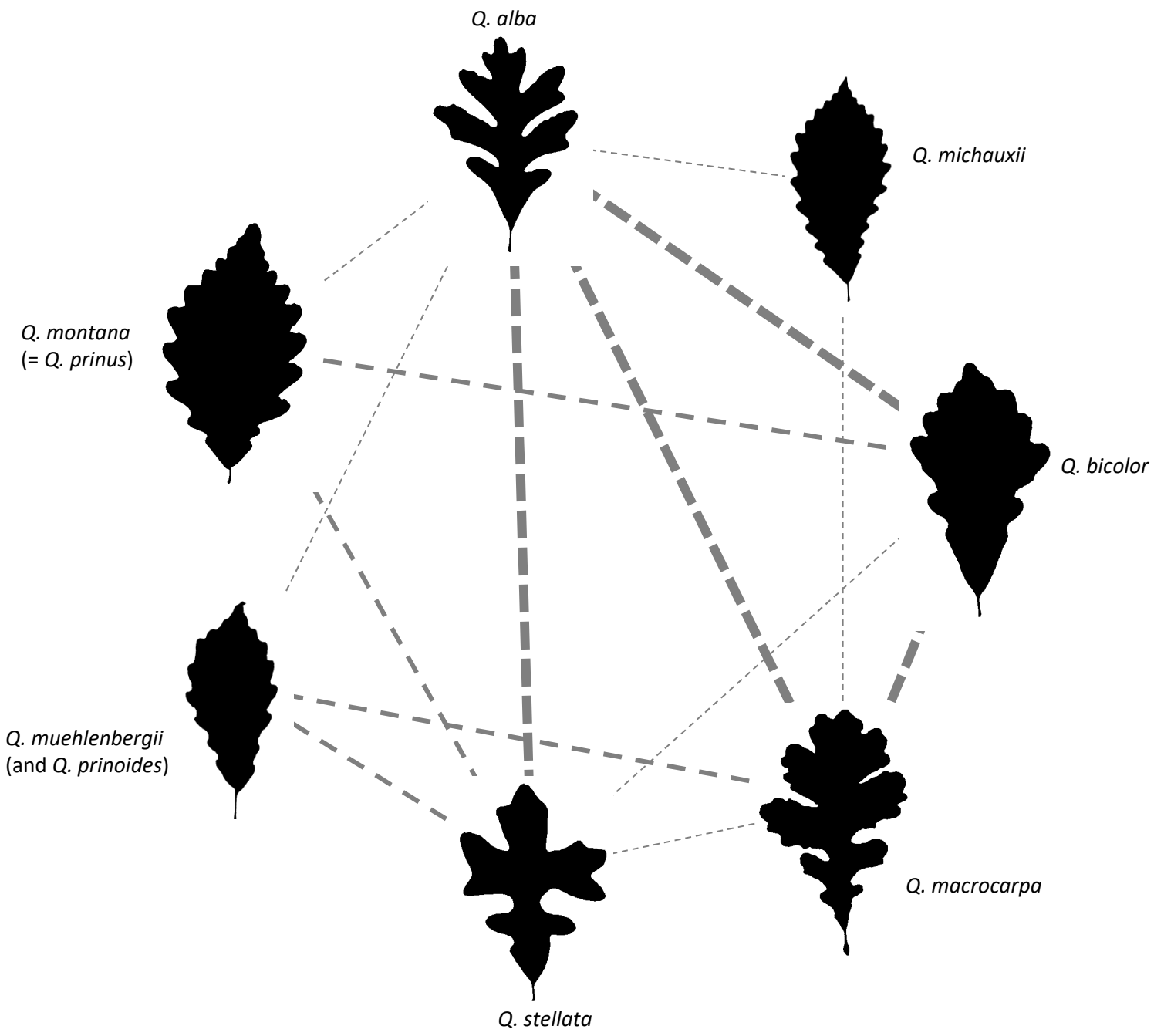
B




MDS2
MDS1
MDS3
MDS1
3
0
−3
−6
2
0
−2
−4
−6
−7.5
−5.0
−2.5
0.0
2.5
5.0
Species
Quercus alba
Quercus bicolor
Quercus macrocarpa
Quercus michauxii
Quercus montana
Quercus muehlenbergii
Quercus prinoides
Quercus stellata

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Specimen	Replicates	Latitude	Longitude	Species	Primary collector	collectorNumber	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: Went	cultivated
QUE000128.a	** A **	41.8649	-86.3508	*Quercus alba*	Marlene Hahn	IL-MOR-MH086	MI	Berrien	Along St. Joseph River along River Trail on Fernw	wild
QUE000128.b	** A **	41.8649	-86.3508	*Quercus alba*	Marlene Hahn	IL-MOR-MH086	MI	Berrien	Along St. Joseph River along River Trail on Fernw	wild
QUE000596		39.9348	-89.8016	*Quercus alba*	Marlene Hahn	IL-SH-162	IL	Menard	StarHill Forest spont. Petersburg.	wild
QUE000151		45.3680	-93.2193	*Quercus alba*	Carol DeVries	IL-MOR-MH109	MN	Anoka	on a farm	wild
QUE000700		36.0218	-79.0161	*Quercus alba*	Paul Manos	PM-19	NC	Orange	near crossroad Cornwallis Rd and Murphy School	wild
QUE001805		42.0421	-93.6057	*Quercus alba*	Mira Garner	IA-MG-262	IA	Story	Ames, Veenker Memorial Golf Course(ISU)	wild
QUE001815		40.7063	-91.7939	*Quercus alba*	Mira Garner	IA-MG-270	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001841		37.9732	-92.7623	*Quercus alba*	Mira Garner	MO-MG327	MO	Camden	Ha Ha Tonka State Park	wild
QUE001918		34.8030	-92.3260	*Quercus alba*	Mira Garner	AR-MG387	AR	Pulaski	Little Rock, Burns Park	wild
QUE001932		37.1573	-91.3650	*Quercus alba*	Mira Garner	MO-MG401	MO	Shannon	Eminence, Buttin Rock Access, near trailer park	wild
QUE001884		36.2181	-95.9000	*Quercus alba*	Mira Garner	OK-MG353	OK	Tulsa	Tulsa, Mohawk Park	wild
QUE002075		39.8393	-88.3677	*Quercus alba*	Ian Pearse	Chickenbristle 4	IL	Douglas	Property of Bob Pearse	wild
QUE002091		38.7363	-86.4143	*Quercus alba*	Mira Garner	IN-MG612	IN	Lawrence	Spring Mill State Park, Trail 5	wild
QUE002108		40.4591	-85.5092	*Quercus alba*	Mira Garner	IN-MG629	IN	Grant	Taylor Wilderness, Taylor University	wild
QUE002121		41.6574	-87.0605	*Quercus alba*	Mira Garner	IN-MG642	IN	Porter	Indiana Dunes State Park	wild
QUE002130		42.0161	-73.3353	*Quercus alba*	Paul Gugger	QUAL-1029	CT	Litchfield	1-19 Sand Rd, North Canaan, CT, US	wild
QUE002138		43.6047	-73.1804	*Quercus alba*	Paul Gugger	QUAL-1037	VT	Rutland	D&H Trl, Castleton, VT, US	wild
QUE002155		44.4464	-73.2202	*Quercus alba*	Paul Gugger	QUAL-1054	VT	Chittenden	1-225 Industrial Pkwy, Burlington, VT, US	wild
QUE002210		38.2172	-91.0864	*Quercus alba*	Mira Garner	MO-MG586	MO	Crawford	Meramec State Park, Sullivan, MO	wild
QUE002253		38.9803	-94.8053	*Quercus alba*	Mira Garner	KS-MG433	KS	Johnson	Shawnee Mission Park	wild
QUE002282		38.8096	-95.1927	*Quercus alba*	Mira Garner	KS-MG462	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild
QUE002337		42.3047	-83.7508	*Quercus alba*	Mira Garner	MI-MG674	MI	Washtenaw	Barton Nature Area, Ann Arbor, Off Trail	wild
QUE002355		42.7666	-84.3911	*Quercus alba*	Mira Garner	MI-MG692	MI	Ingham	Lake Lansing Park, East Lansing, Picnic Area	wild
QUE002366		43.0168	-90.1142	*Quercus alba*	Mira Garner	WI-MG703	WI	Iowa	Governor Dodge State Park, Dodgeville, Cox Hollo	wild
QUE002399		43.0171	-88.4353	*Quercus alba*	Mira Garner	WI-MG736	WI	Waukesha	University of Wisconsin - Waukesha Field Station,	wild
QUE002493		41.5522	-84.3590	*Quercus alba*	Mira Garner	OH-MG830	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE000643		41.1922	-87.4463	*Quercus bicolor*	Marlene Hahn	IL-SH-030	IN	Lake	Mohawk Club; Schneider.	wild
QUE000618		38.8931	-94.8322	*Quercus bicolor*	Marlene Hahn	IL-SH-184	KS	Johnson		wild
QUE000136		41.7409	-87.8603	*Quercus bicolor*	Marlene Hahn	IL-MOR-MH094	IL	Cook	Along the Des Plaines River near Willow Springs	wild
QUE001813		40.7048	-91.7963	*Quercus bicolor*	Mira Garner	IA-MG-268	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE002153		44.4001	-73.2375	*Quercus bicolor*	Paul Gugger	QUBI-1052	VT	Chittenden	1136 Bay Rd, Shelburne, VT, US	wild
QUE002196		38.2267	-91.0830	*Quercus bicolor*	Mira Garner	MO-MG572	MO	Crawford	Meramec State Park, Sullivan, MO, Campground	wild
QUE002360		42.7951	-84.3927	*Quercus bicolor*	Mira Garner	MI-MG697	MI	Ingham	Lake Lansing Park, East Lansing, Edge of marsh	wild
QUE002361		42.7653	-84.3825	*Quercus bicolor*	Mira Garner	MI-MG698	MI	Ingham	Lake Lansing Park, East Lansing, Edge of marsh	wild
QUE002528		42.1843	-87.9163	*Quercus bicolor*	Mira Garner	IL-MG865	IL	Lake	Ryerson Woods Conservation Area, Trail behind c	wild
QUE002539		41.8245	-87.9333	*Quercus bicolor*	Mira Garner	IL-MG876	IL	DuPage	Fullersburg Woods Nature Preserve	wild
QUE000671		40.4554	-86.9165	*Quercus macrocarpa*	Bethany Hayward Brown	IL-SH-58	IN	Wabash	West Lafayette	wild
QUE000623		40.1923	-96.6650	*Quercus macrocarpa*	Marlene Hahn	IL-SH-189	NE	Gage	Blue River,via NSA 2000	wild
QUE000619		36.6467	-89.3021	*Quercus macrocarpa*	Marlene Hahn	IL-SH-185	MO		Big Oak Tree State Park	wild
QUE000640		38.3507	-87.8226	*Quercus macrocarpa*	Marlene Hahn	IL-SH-027	IL	Wabash	Beall Woods State Park	wild
QUE000107		41.4868	-87.7998	*Quercus macrocarpa*	Marlene Hahn	IL-MOR-MH003 (A/B)	IL	Cook	Near Sauk Lake in Sauk Trail Forest Preserve	wild
QUE000617		48.3076	-98.7287	*Quercus macrocarpa*	Marlene Hahn	IL-SH-183	ND	Ramsey		wild
QUE000673		39.7799	-96.0153	*Quercus macrocarpa*	Bethany Hayward Brown	IL-SH-060	KS	Nemaha		wild
QUE000622		35.6239	-99.0087	*Quercus macrocarpa*	Marlene Hahn	IL-SH-188	OK	Custer		wild
QUE000672		45.5029	-104.4767	*Quercus macrocarpa*	Bethany Hayward Brown	IL-SH-59	MT	Carter		wild
QUE000620		33.6067	-105.3631	*Quercus macrocarpa*	Marlene Hahn	IL-SH-186	NM		Capitan Mountains	wild
QUE000624		45.5039	-73.5545	*Quercus macrocarpa*	Marlene Hahn	IL-SH-190	Quebec		Montreal	wild
QUE001759		43.1070	-89.8083	*Quercus macrocarpa*	Mira Garner	WI-MG230	WI	Dane	Pleasant Valley Conservancy	wild
QUE001804		42.0421	-93.6062	*Quercus macrocarpa*	Mira Garner	IA-MG-261	IA	Story	Ames, Veenker Memorial Golf Course(ISU)	wild
QUE001814		40.7056	-91.7942	*Quercus macrocarpa*	Mira Garner	IA-MG-269	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001863		36.8450	-96.4253	*Quercus macrocarpa*	Mira Garner	OK-MG282	OK	Osage	Pawhuska, Tallgrass Prairie Preserve	wild
QUE001894		35.4438	-98.3545	*Quercus macrocarpa*	Mira Garner	OK-MG363	OK	Caddo	Hinton, Red Rock Canyon State Park	wild
QUE001916		34.8038	-92.3263	*Quercus macrocarpa*	Mira Garner	AR-MG385	AR	Pulaski	Little Rock, Burns Park	wild
QUE001933		37.1569	-91.3647	*Quercus macrocarpa*	Mira Garner	MO-MG402	MO	Shannon	Eminence, Buttin Rock Access, near river	wild
QUE001783		41.9736	-91.7239	*Quercus macrocarpa*	Mira Garner	IA-MG-239	IA	Linn	Cedar Rapids, Cherokee Park	wild
QUE001880		36.2204	-95.8985	*Quercus macrocarpa*	Mira Garner	OK-MG349	OK	Tulsa	Tulsa, Mohawk Park	wild
QUE001907		35.1769	-97.4497	*Quercus macrocarpa*	Mira Garner	OK-MG376	OK	Cleveland	Norman, Oliver's Woods, University of Oklahoma c	wild
QUE001937		43.9029	-91.6400	*Quercus macrocarpa*	Mira Garner	MN-MG493	MN	Winona	Winona, Prairie Moon Nursery, Wiscoy Co-op, in w	wild
QUE001951		49.7138	-95.2439	*Quercus macrocarpa*	Mira Garner	MB-MG507	Manitoba		Whiteshell Provincial Park	wild
QUE001963		49.7614	-99.1604	*Quercus macrocarpa*	Mira Garner	MB-MG519	Manitoba		Spruce Woods Provincial Park	wild
QUE001971		49.8578	-97.2491	*Quercus macrocarpa*	Mira Garner	MB-MG527	Manitoba		Assiniboine Forest, near trail	wild
QUE001982		46.0259	-91.1429	*Quercus macrocarpa*	Mira Garner	WI-MG538	WI	Sawyer	Round Lake, Chequamegon-Nicolet National Fores	wild
QUE002057		38.9636	-98.5891	*Quercus macrocarpa*	Ian Pearse	Minooka 2	KS	Russell	Minooka Park Recreation Area	wild
QUE002074		39.8393	-88.3677	*Quercus macrocarpa*	Ian Pearse	Chickenbristle 3	IL	Douglas	Property of Bob Pearse	wild
QUE002081		37.5149	-89.4445	*Quercus macrocarpa*	Mira Garner	IL-MG602	IL	Jackson	Oakwood Bottoms, Shawnee National Forest, Alon	wild
QUE002082		37.5164	-89.4454	*Quercus macrocarpa*	Mira Garner	IL-MG603	IL	Jackson	Oakwood Bottoms, Shawnee National Forest, Alon	wild
QUE002085		38.7362	-86.4126	*Quercus macrocarpa*	Mira Garner	IN-MG606	IN	Lawrence	Spring Mill State Park, Near trail/lake	wild
QUE002102		40.4591	-85.5041	*Quercus macrocarpa*	Mira Garner	IN-MG623	IN	Grant	Taylor Wilderness, Taylor University	wild
QUE002129		41.9623	-73.3130	*Quercus macrocarpa*	Paul Gugger	QUMAC-1028	CT	Litchfield	Litchfield County, US-CT, US	wild
QUE002133		42.1666	-73.4121	*Quercus macrocarpa*	Paul Gugger	QUMAC-1032	MA	Berkshire	16-18 Creamery Rd, Egremont, MA, US	wild
QUE002137		43.6034	-73.1811	*Quercus macrocarpa*	Paul Gugger	QUMAC-1036	VT	Rutland	D&H Trl, Castleton, VT, US	wild
QUE002154		44.4007	-73.2376	*Quercus macrocarpa*	Paul Gugger	QUMAC-1053	VT	Chittenden	Shelburne Bay, Shelburne, VT, US	wild
QUE002208		38.2283	-91.0824	*Quercus macrocarpa*	Mira Garner	MO-MG584	MO	Crawford	Meramec State Park, Sullivan, MO, Campground	wild
QUE002251		38.9801	-94.8052	*Quercus macrocarpa*	Mira Garner	KS-MG431	KS	Johnson	Shawnee Mission Park	wild
QUE002284		38.8087	-95.1939	*Quercus macrocarpa*	Mira Garner	KS-MG464	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild
QUE002295		39.1071	-96.6077	*Quercus macrocarpa*	Mira Garner	KS-MG475	KS	Riley	Konza Prairie, Manhattan, KS, Near nature trail-alc	wild
QUE002302		39.1034	-96.5962	*Quercus macrocarpa*	Mira Garner	KS-MG482	KS	Riley	Konza Prairie, Manhattan, KS, Along King's Creek	wild
QUE002326		43.8532	-83.9230	*Quercus macrocarpa*	Mira Garner	MI-MG663	MI	Bay	Pinconning Park, Campground	wild
QUE002336		42.3084	-83.7567	*Quercus macrocarpa*	Mira Garner	MI-MG673	MI	Washtenaw	Barton Nature Area, Ann Arbor, Picnic Area	wild
QUE002356		42.7651	-84.3890	*Quercus macrocarpa*	Mira Garner	MI-MG693	MI	Ingham	Lake Lansing Park, East Lansing, Edge of marsh	wild
QUE002367		43.0169	-90.1148	*Quercus macrocarpa*	Mira Garner	WI-MG704	WI	Iowa	Governor Dodge State Park, Dodgeville, Cox Hollo	wild
QUE002400		43.0161	-88.4351	*Quercus macrocarpa*	Mira Garner	WI-MG737	WI	Waukesha	University of Wisconsin - Waukesha Field Station,	wild

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

QUE002424		44.3639	-93.9354	*Quercus macrocarpa*	Mira Garner	MN-MG761	MN	Le Sueur	Ottawa Bluffs	wild
QUE002436		45.5300	-94.2364	*Quercus macrocarpa*	Mira Garner	MN-MG773	MN	Stearns	Quarry Park State Natural Area, Waite Park	wild
QUE002471		40.7312	-83.0933	*Quercus macrocarpa*	Mira Garner	OH-MG808	OH	Crawford	Daughmer Prairie Savannah State Nature Preserv	wild
QUE002480		41.5507	-84.3597	*Quercus macrocarpa*	Mira Garner	OH-MG817	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE002579		43.8168	-103.2501	*Quercus macrocarpa*	Jeannine Cavender-Bares	JCB-SD-US16-1010	SD	Custer	East of Custer State Park driving eastward towards	wild
QUE002585		44.4673	-103.8498	*Quercus macrocarpa*	Jeannine Cavender-Bares	JCB-SD-SPF-1016	SD	Lawrence	heading towards Spearfish	wild
QUE002717	** B **	38.9057	-86.0359	*Quercus michauxii*	Elisabeth Fitzek	645-48*2	IN	Jackson		wild
QUE000121	** B **	38.9057	-86.0359	*Quercus michauxii*	Laurie Glaysher	IL-MOR-MH079	IN	Jackson		wild
QUE002718		Unknown	Unknown	*Quercus michauxii*	Elisabeth Fitzek	476-42*1	Cultivated	Cultivated	Graft from University of Washington Botanic Garde	cultivated
QUE002719		36.6447	-89.2850	*Quercus michauxii*	Elisabeth Fitzek	539-96*3	MO	Mississippi	In picnic ground at Big Oak Tree State Park.	wild
QUE002679		36.6447	-89.2850	*Quercus michauxii*	Carol DeVries	IL-MOR-MH250	MO	Mississippi	In picnic ground at Big Oak Tree State Park.	wild
QUE002680		36.6447	-89.2850	*Quercus michauxii*	Marilyn Carle	IL-MOR-MH251	MO	Mississippi	In picnic ground at Big Oak Tree State Park.	wild
QUE002720	** C **	36.6447	-89.2850	*Quercus michauxii*	Elisabeth Fitzek	539-96*5	MO	Mississippi	In picnic ground at Big Oak Tree State Park.	wild
QUE000105	** C **	36.6447	-89.2850	*Quercus michauxii*	Ken Potenberg	IL-MOR-MH001	MO	Mississippi	In picnic ground at Big Oak Tree State Park.	wild
QUE000588		37.1538	-89.3470	*Quercus michauxii*	Bethany Hayward Brown	IL-SH-154	IL		Horeshoe Lake; Olive Branch	wild
QUE001116		35.9956	-79.0542	*Quercus michauxii*	Paul Manos	PM143	NC	Orange	Jonston Mill Preserve	wild
QUE001128		36.0152	-78.9233	*Quercus michauxii*	Paul Manos	PM155	NC	Durham	Edith Street Durham, NC	wild
QUE002722		37.1414	-79.9957	*Quercus montana*	Elisabeth Fitzek	606-2000*3	VA	Franklin	Cahas Mt.	wild
QUE002723	** D **	37.5258	-80.2497	*Quercus montana*	Elisabeth Fitzek	602-2000*2	VA	Craig	At picnic area across from entrance to county road	wild
QUE000122	** D **	37.5258	-80.2497	*Quercus montana*	Evelyn Means	IL-MOR-MH080	VA	Craig	At picnic area across from entrance to county road	wild
QUE002724		37.5258	-80.2497	*Quercus montana*	Elisabeth Fitzek	602-2000*1	VA		At picnic area across from entrance to county road	wild
QUE002725		37.5258	-80.2497	*Quercus montana*	Elisabeth Fitzek	602-2000*3	VA		At picnic area across from entrance to county road	wild
QUE000639		37.6179	-88.7048	*Quercus montana*	Marlene Hahn	IL-SH-26	IL		near Stonefort	wild
QUE000111		37.1414	-79.9957	*Quercus montana*	Marlene Hahn	IL-MOR-MH007	VA	Franklin	Cahas Mt	wild
QUE000576		35.4291	-82.2518	*Quercus montana*	Bethany Hayward Brown	IL-SH-116	NC		Chimney Rock Park	wild
QUE002726		Unknown	Unknown	*Quercus muehlenbergii*	Elisabeth Fitzek	704-46*2	IN			wild
QUE002727		Unknown	Unknown	*Quercus muehlenbergii*	Elisabeth Fitzek	704-63*3	IN			wild
QUE000152		40.6715	-95.7047	*Quercus muehlenbergii*	Chris Courtney	IL-MOR-MH110	IA	Fremont	8 mi. east of Nebraska City, south of IA 2, on top o	wild
QUE000587		39.9352	-89.8023	*Quercus muehlenbergii*	Bethany Hayward Brown	IL-SH-153	IL	Menard	Petersburg. Starhill Forest Spont.	wild
QUE000145		41.2106	-88.0176	*Quercus muehlenbergii*	Marlene Hahn	IL-MOR-MH103	IL	Will	In campground north of the Kankakee River under	wild
QUE000670		35.6239	-99.0087	*Quercus muehlenbergii*	Bethany Hayward Brown	IL-SH-57	OK	Custer		wild
QUE000322		31.9792	-104.7542	*Quercus muehlenbergii*	Marlene Hahn	CA-DAV-MH49	TX	Culberson	Guadalupe Mountains:McKittrick Canyon: 0.6 Mile	wild
QUE001819		40.7048	-91.7959	*Quercus muehlenbergii*	Mira Garner	IA-MG-274	IA	Van Buren	Bonaparte, Lindsay Wilderness	wild
QUE001840		37.9730	-92.7622	*Quercus muehlenbergii*	Mira Garner	MO-MG326	MO	Camden	Ha Ha Tonka State Park	wild
QUE001893		35.4481	-98.3535	*Quercus muehlenbergii*	Mira Garner	OK-MG362	OK	Caddo	Hinton, Red Rock Canyon State Park	wild
QUE002086		38.7363	-86.4125	*Quercus muehlenbergii*	Mira Garner	IN-MG607	IN	Lawrence	Spring Mill State Park, Trail 5	wild
QUE002098		38.7374	-86.4126	*Quercus muehlenbergii*	Mira Garner	IN-MG619	IN	Lawrence	Spring Mill State Park, Roadside near Nature Cent	wild
QUE002101		40.4589	-85.5038	*Quercus muehlenbergii*	Mira Garner	IN-MG622	IN	Grant	Taylor Wilderness, Taylor University, Off Eighth Str	wild
QUE002105		40.4592	-85.5081	*Quercus muehlenbergii*	Mira Garner	IN-MG626	IN	Grant	Taylor Wilderness, Taylor University	wild
QUE002189		38.5114	-90.5592	*Quercus muehlenbergii*	Mira Garner	MO-MG565	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002247		38.9789	-94.8050	*Quercus muehlenbergii*	Mira Garner	KS-MG427	KS	Johnson	Shawnee Mission Park	wild
QUE002250		38.9739	-94.8051	*Quercus muehlenbergii*	Mira Garner	KS-MG430	KS	Johnson	Shawnee Mission Park	wild
QUE002285		38.8087	-95.1939	*Quercus muehlenbergii*	Mira Garner	KS-MG465	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild
QUE002303		39.1016	-96.5998	*Quercus muehlenbergii*	Mira Garner	KS-MG483	KS	Riley	Konza Prairie, Manhattan, KS	wild
QUE002304		39.1079	-96.6047	*Quercus muehlenbergii*	Mira Garner	KS-MG484	KS	Riley	Konza Prairie, Manhattan, KS, Along nature trail	wild
QUE002481		41.5510	-84.3585	*Quercus muehlenbergii*	Mira Garner	OH-MG818	OH	Fulton	Goll Woods State Nature Preserve	wild
QUE002699		40.0499	-95.7298	*Quercus prinoides*	Chris Courtney	IL-MOR-MH270	NE	Richardson	Rock Creek bluffs, 3 miles south of Salem	wild
QUE002695		39.8189	-94.0103	*Quercus prinoides*	Satish Sachdev	IL-MOR-MH266	MO		Northwestern section of Missouri.	wild
QUE002728		40.0383	-95.7565	*Quercus prinoides*	Elisabeth Fitzek	120-2001*2	NE	Richardson	plant grown from wild seed southwest of Salem.	wild
QUE002729		40.0383	-95.7565	*Quercus prinoides*	Elisabeth Fitzek	120-2001*3	NE	Richardson	plant grown from wild seed southwest of Salem.	wild
QUE002730		Unknown	Unknown	*Quercus prinoides*	Elisabeth Fitzek	218-77*2	Cultivated	Cultivated	Seed from MOR accession 742-51	cultivated
QUE002731		Unknown	Unknown	*Quercus prinoides*	Elisabeth Fitzek	218-77*3	Cultivated	Cultivated	Seed from MOR accession 742-51	cultivated
QUE002689		Unknown	Unknown	*Quercus prinoides*	Sarah Packard	IL-MOR-MH260	Cultivated	Cultivated	Seed from MOR accession 742-51	cultivated
QUE002701		40.7925	-77.8621	*Quercus prinoides*	NA	IL-MOR-MH272	PA		State College grounds	wild
QUE002694	** E **	40.7925	-77.8621	*Quercus prinoides*	Marilyn Carle	IL-MOR-MH265	PA		State College grounds	wild
QUE000133	** E **	40.7925	-77.8621	*Quercus prinoides*	Ken Potenberg	IL-MOR-MH091	PA		State College grounds	wild
QUE002693		39.8189	-94.0103	*Quercus prinoides*	Edie Moran	IL-MOR-MH264	MO		Northwestern section of Missouri.	wild
QUE002696		39.8189	-94.0103	*Quercus prinoides*	Chris Courtney	IL-MOR-MH267	MO		Northwestern section of Missouri.	wild
QUE002697		40.7925	-77.8621	*Quercus prinoides*	Charlene Kubic	IL-MOR-MH268	PA		State College grounds	wild
QUE002698		40.7925	-77.8621	*Quercus prinoides*	Chris Courtney	IL-MOR-MH269	PA		State College grounds	wild
QUE002700		40.0499	-95.7298	*Quercus prinoides*	Charlene Kubic	IL-MOR-MH271	NE	Richardson	Rock Creek bluffs, 3 miles south of Salem	wild
QUE000565		38.4676	-95.1365	*Quercus prinoides*	Andrew Hipp	IL-SH-105	KS	Franklin	3.5 mi. NW of Lane	wild
QUE000678		40.0763	-95.7210	*Quercus prinoides*	Marlene Hahn	IL-SH-96	NE	Richardson	SW of Salem	wild
QUE000753		42.0015	-76.5991	*Quercus prinoides*	EA Cope	PM93	NY	Chemung		wild
QUE002683		40.6487	-91.6733	*Quercus stellata*	Satish Sachdev	IL-MOR-MH254	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002732		37.0922	-93.8381	*Quercus stellata*	Elisabeth Fitzek	11-86*2	MO		Lawrence	wild
QUE002733		37.0922	-93.8381	*Quercus stellata*	Elisabeth Fitzek	11-86*3	MO		Lawrence	wild
QUE002734	** F **	38.0349	-91.5203	*Quercus stellata*	Elisabeth Fitzek	1137-2004*2	MO	Phelps	Along I-44 and RR at Rosati	wild
QUE000143	** F **	38.0349	-91.5203	*Quercus stellata*	Marlene Hahn	IL-MOR-MH101	MO	Phelps	Along I-44 and RR at Rosati	wild
QUE002735		38.0349	-91.5203	*Quercus stellata*	Elisabeth Fitzek	1137-2004*3	MO	Phelps	Along I-44 and RR at Rosati	wild
QUE002704		40.6487	-91.6733	*Quercus stellata*	Ken Potenberg	IL-MOR-MH275	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002703		40.6487	-91.6733	*Quercus stellata*	Ken Potenberg	IL-MOR-MH274	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE002706	** G **	40.6487	-91.6733	*Quercus stellata*	Bethany Hayward Brown	IL-MOR-MH277	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE000137	** G **	40.6487	-91.6733	*Quercus stellata*	Marlene Hahn	IL-MOR-MH095	IA	Lee	in the Donnellson Unit of Shimek State Forest	wild
QUE000608		38.7272	-88.7795	*Quercus stellata*	Bethany Hayward Brown	IL-SH-174	IL	Marion	Forbes State Recreation Area	wild
QUE000638		37.9995	-91.6092	*Quercus stellata*	Marlene Hahn	IL-SH-25	MO	Phelps	Hillview Haven; St. James	wild
QUE000692		35.9767	-78.9866	*Quercus stellata*	Paul Manos	PM11	NC	Durham	Durham county; 3658 Pineview Circle	wild
QUE001118		36.0187	-78.9253	*Quercus stellata*	Paul Manos	PS Manos 1907	NC	Durham	Watts Hillandale tree Intersection of Carolina Aven	wild
QUE001839		37.9729	-92.7622	*Quercus stellata*	Mira Garner	MO-MG325	MO	Camden	Ha Ha Tonka State Park	wild
QUE001862		36.8491	-96.4152	*Quercus stellata*	Mira Garner	OK-MG281	OK	Osage	Pawhuska, Tallgrass Prairie Preserve	wild
QUE001915		34.8041	-92.3267	*Quercus stellata*	Mira Garner	AR-MG384	AR	Pulaski	Little Rock, Burns Park	wild
QUE002187		38.5109	-90.5599	*Quercus stellata*	Mira Garner	MO-MG563	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002188		38.5115	-90.5592	*Quercus stellata*	Mira Garner	MO-MG564	MO	St. Louis	Tyson Research Center, Eureka, MO	wild
QUE002209		38.2181	-91.0836	*Quercus stellata*	Mira Garner	MO-MG585	MO	Crawford	Meramec State Park, Sullivan, MO	wild

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

QUE002219	38.2179	-91.0921	*Quercus stellata*	Mira Garner	MO-MG595	MO	Crawford	Meramec State Park, Sullivan, MO, Deer Hollow Tr	wild
QUE002252	38.9801	-94.8050	*Quercus stellata*	Mira Garner	KS-MG432	KS	Johnson	Shawnee Mission Park	wild
QUE002283	38.8096	-95.1927	*Quercus stellata*	Mira Garner	KS-MG463	KS	Douglas	Breidenthal Woods/Baldwin Woods	wild

bioRxiv preprint doi: https://doi.org/10.1101/602573; this version posted July 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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