

1 **Full title:** Genomic landscape of the global oak phylogeny
2 **Short title:** Genomic landscape of the global oak phylogeny

3 Andrew L. Hipp^{1,2}, Paul S. Manos³, Marlene Hahn¹, Michael Avishai⁴, Cathérine Bodénès⁵, Jeannine
4 Cavender-Bares⁶, Andrew Crowl³, Min Deng⁷, Thomas Denk⁸, Sorel Fitz-Gibbon⁹, Oliver Gailing¹⁰,
5 M. Socorro González-Elizondo¹¹, Antonio González-Rodríguez¹², Guido W. Grimm¹³, Xiao-Long
6 Jiang⁷, Antoine Kremer⁵, Isabelle Lesur⁵, John D. McVay^{3,14}, Christophe Plomion⁵, Hernando
7 Rodríguez-Correa¹², Ernst-Detlef Schulze¹⁵, Marco C. Simeone¹⁶, Victoria L. Sork⁹, Susana Valencia-
8 Avalos¹⁷

9 ¹ The Morton Arboretum, Lisle IL 60532-1293, USA

10 ² The Field Museum, Chicago IL 60605, USA

11 ³ Duke University, Durham NC 27708, USA

12 ⁴ Previously of The Hebrew University of Jerusalem, Botanical Garden, Jerusalem, Israel

13 ⁵ INRA, UMR1202 BIOGECO, Cestas F-33610, France

14 ⁶ University of Minnesota, Minneapolis MN 55455, USA

15 ⁷ Shanghai Chenshan Plant Science Research Center, Chinese Academy of Sciences, Shanghai 201602,
16 China

17 ⁸ Swedish Museum of Natural History, 10405 Stockholm, Sweden

18 ⁹ University of California, Los Angeles, Los Angeles CA 90095, USA

19 ¹⁰ Büsgen-Institute, Georg-August-University Göttingen, D-37077 Göttingen, Germany

20 ¹¹ CIIDIR Unidad Durango, Instituto Politécnico Nacional, Durango, México

21 ¹² Escuela Nacional de Estudios Superiores Unidad Morelia, Universidad Nacional Autónoma de
22 México

23 ¹³ Unaffiliated, 45100 Orléans, France; <http://orcid.org/0000-0003-0674-3553>

24 ¹⁴ Current affiliation : Division of Plant Industry, Florida Department of Agriculture and Consumer
25 Services, Gainesville FL 32608, USA

26 ¹⁵ Max Planck Institute for Biogeochemistry, Hans-Knoell-Str. 10, 07745 Jena, Germany

27 ¹⁶ Università degli Studi della Tuscia, Viterbo, Italy

28 ¹⁷ Herbario de la Facultad de Ciencias, Departamento de Biología Comparada, Universidad Nacional
29 Autónoma de México, Circuito Exterior, s.n., Ciudad Universitaria, Coyoacán, CP 04510, México City,
30 México

31 **Author for correspondence:** Andrew L. Hipp / Tel: +1 630 725 2094 / Email: ahipp@mortonarb.org

32 **Total word count** 6,474

33 • Introduction: 559

34 • Materials and Methods: 1,291

35 • Results: 1,770

36 • Discussion: 2,770

37 • Acknowledgements: 84

38

39 Tables: 1

40 Figures: 6 [all color]

41 **Summary**

- 42 • The tree of life is highly reticulate, with the history of population divergence buried amongst
43 phylogenies deriving from introgression and lineage sorting. In this study, we test the
44 hypothesis that there are regions of the oak (*Quercus*, Fagaceae) genome that are broadly
45 informative about phylogeny and investigate global patterns of oak diversity.
- 46 • We utilize fossil data and restriction-site associated DNA sequencing (RAD-seq) for 632
47 individuals representing ca. 250 oak species to infer a time-calibrated phylogeny of the world's
48 oaks. We use reversible-jump MCMC to reconstruct shifts in lineage diversification rates,
49 accounting for among-clade sampling biases. We then map the > 20,000 RAD-seq loci back to
50 a recently published oak genome and investigate genomic distribution of introgression and
51 phylogenetic support across the phylogeny.
- 52 • Oak lineages have diversified among geographic regions, followed by ecological divergence
53 within regions, in the Americas and Eurasia. Roughly 60% of oak diversity traces back to four
54 clades that experienced increases in net diversification due to climatic transitions or ecological
55 opportunity.
- 56 • The support we find for the phylogeny contrasts with high genomic heterogeneity in
57 phylogenetic signal and introgression. Oaks are phylogenomic mosaics, and their diversity may
58 in fact depend on the gene flow that shapes the oak genome.

59
60 **Keywords:** Diversification rates, Genomic mosaicism, Oaks, Introgression, Phylogenomics, *Quercus*,
61 Restriction-site associated DNA sequencing (RAD-seq), Tree diversity

62
63

64 **Introduction**

65 The tree of life exhibits reticulation from its base to its tips (Folk *et al.*, 2018; Quammen, 2018). Oaks
66 (*Quercus* L., Fagaceae) are no exception (Hipp, 2018), and in fact the genus is rife with case-studies in
67 localized gene flow (e.g. Hardin, 1975; Whittemore & Schaal, 1991; McVay *et al.*, 2017a; Kim *et al.*,
68 2018), and ancient introgression (Crowl *et al.*, In review; McVay *et al.*, 2017b; Kim *et al.*, 2018). Oaks
69 have in fact been held up as a paradigmatic syngameon (Hardin, 1975; Van Valen, 1976; Dodd &
70 Afzal-Rafii, 2004; Cannon & Scher, 2017; Boecklen, 2017), a system of interbreeding species in which
71 incomplete reproductive isolation may facilitate adaptive gene flow and species migration (Petit *et al.*,
72 2003; Dodd & Afzal-Rafii, 2004). The oak genome (Plomion *et al.*, 2018) consequently tracks
73 numerous unique species-level phylogenetic histories that result from lineage sorting and differential
74 rates of introgression (Anderson, 1953; Eaton *et al.*, 2015; McVay *et al.*, 2017b; Edelman *et al.*, 2018).
75 Oak genomes are mosaics of disparate phylogenetic histories (cf. Pääbo, 2003). Given the prevalence
76 of hybridization in trees globally (Petit & Hampe, 2006; Cannon & Lerda, 2015), understanding how
77 these stories line up with one another, and whether there are regions of the genome that track a
78 common story, is essential to understanding the prevalence of adaptive gene flow and the phylogenetic
79 history of forest trees.

80 Restriction-site associated DNA sequencing (RAD-seq; Miller *et al.*, 2007a,b; Lewis *et al.*,
81 2007; Baird *et al.*, 2008; Ree & Hipp, 2015) has revolutionized our understanding of oak phylogeny in
82 the past five years (Jiang *et al.*, In review; Hipp *et al.*, 2014, 2018; Cavender-Bares *et al.*, 2015; Eaton
83 *et al.*, 2015; Hipp, 2017; Fitz-Gibbon *et al.*, 2017; Pham *et al.*, 2017; Ortego *et al.*, 2018; Deng *et al.*,
84 2018; Kim *et al.*, 2018). Its ties to the genome, however, have not been fully exploited because of the
85 lack of an assembled genome. While earlier studies have explored the effects of gene identity on
86 phylogenetic informativeness (Hipp *et al.*, 2014) and genomic heterogeneity in phylogenetic vs.
87 introgressive signals (McVay *et al.*, 2017b,a), they have not had access to the oak genome sequence.
88 As a consequence, we do not understand the distribution of genomic breakpoints between introgressive
89 and divergent histories. Moreover, no studies to date have brought together a comprehensive sampling
90 of taxa to investigate the history of diversification across the genus.

91 In this paper, we integrate data from the recently published *Quercus robur* genome (Plomion *et*
92 *al.*, 2016, 2018) with previously published RAD-seq data for 427 sequenced oak individuals across the
93 tree of life and new RAD-seq data for an additional 205 individuals to investigate the global oak
94 phylogenomic mosaic for approximately 60% of the world's *Quercus* species. We test the hypothesis
95 that there are regions of the genome that are uniformly informative about *Quercus* phylogeny, regions
96 that make oak lineages what they are. Furthermore, using a time-calibrated one-tip-per species tree

97 novel to this study for ca. 60% of known species, we test the hypothesis that the high diversity of oaks
98 in Mexico and eastern China is a consequence of high diversification rates. Finally, we show that the
99 consensus of the evolutionary histories of more than 20,000 RAD-seq loci matches our understanding
100 of oak evolution based on morphological information from extant and fossil species in spite of broadly
101 conflicting individual locus genealogies.

102

103 **Materials and Methods**

104 *Previously published RAD-seq and new RAD-seq: sequencing and clustering*

105 Data from previously published RAD-seq phylogenies were analyzed alongside new RAD-seq data for
106 a total of 632 individuals (Table S1). RAD-seq data were generated as described in the previous
107 studies. New data were from library preparations conducted at Floragenex, Inc. (Portland, OR, USA)
108 following the methods of Baird *et al.* (2008) with *Pst*I, barcoded by individual, and sequenced on an
109 Illumina Genome Analyzer Iix at Floragenex, or an Illumina HiSeq 2500 or HiSeq 4000 at the
110 University of Oregon Genomic Facility.

111 FASTQ files were demultiplexed and filtered to remove sequences with more than 5 bases of
112 quality score < 20 and assembled into loci for phylogenetic analysis using ipyrad 0.7.23 (Eaton, 2014)
113 at 85% sequence similarity. Consensus sequences for each individual for each locus were then
114 clustered across individuals, retaining loci present in at least 4 individuals and possessing a maximum
115 of 20 SNPs and 8 indels across individuals. The dataset was filtered to loci with a minimum of 15
116 individuals each, for a total of 58,985 loci. Data were imported into R using the RADami package
117 (Hipp *et al.*, 2014) for downstream analysis.

118 RAD-seq loci were mapped back to the latest version of the *Quercus robur* haploid genome
119 (haplome 2.3; <https://urgi.versailles.inra.fr/Data/Genome/Genome-data-access>) (Plomion *et al.*, 2018).
120 The oak genome is made of 12 pseudomolecules (*i.e.* chromosomes) and a set of 538 unassigned
121 scaffolds. Mapping was performed using Blast+ 2.8.1 (Camacho *et al.*, 2009). We filtered alignments
122 based on expect (E) values ($E\text{-value} \leq 10^{-5}$), alignment length ($\geq 80\%$ of the length of the loci) and
123 percent identity ($\geq 80\%$). For each locus, the best alignment was kept. All sequence data analyzed in
124 this paper are available as FASTQ files from NCBI's Short Read Archive (Table S1), and aligned loci
125 and additional data and scripts for all analysis are available from [https://github.com/andrew-](https://github.com/andrew-hipp/global-oaks-2019)
126 [hipp/global-oaks-2019](https://github.com/andrew-hipp/global-oaks-2019). Analysis details are in the Supplement (Methods S1).

127

128 *Phylogenetic analysis*

129 Maximum likelihood phylogenetic analyses were conducted in RAxML v8.2.4 (Stamatakis, 2014)
130 using the GTRCAT implementation of the general time reversible model of nucleotide evolution
131 (Stamatakis, 2006), with branch support assessed using RELL bootstrapping (Minh *et al.*, 2013). For
132 the phylogeny including all tips (Fig. S1), analysis was unconstrained, and we used the taxonomic
133 disparity index (TDI) of Pham *et al.* (2016) to identify the extent of non-monophyly by species.
134 Topology within the white oaks of sections *Ponticae*, *Virentes*, and *Quercus* (hereafter in the paper
135 “white oaks *s.l.*,” contrasted with “white oaks *s.str.*” for just section *Quercus*) was observed to be at
136 odds with previous close studies (Crowl *et al.*, In review; McVay *et al.*, 2017b,a; Hipp *et al.*, 2018) that
137 have shown the topology of the white oaks *s.l.* to be sensitive to taxon and locus sampling. For dating,
138 samples were pruned to one sample per named species, favoring samples with the most loci, except for
139 species in which variable position of samples from different populations was deemed to represent
140 cryptic diversity, in which case more than one exemplar was retained. The singletons tree was
141 estimated in RAxML using a phylogenetic constraint (Manos, 2016; McVay *et al.*, 2017b; Hipp *et al.*,
142 2018) available in the supplemental methods and supplemental data. The remainder of the tree was
143 unconstrained and conforms closely to previous topologies.

144 We utilized neighbor-net (Bryant & Moulton, 2004) to visualize overall patterns of molecular
145 genetic diversity. Likelihood-based methods (e.g., Solís-Lemus & Ané, 2016; Solís-Lemus *et al.*, 2017;
146 Wen *et al.*, 2018; Zhang *et al.*, 2018) that we have utilized on smaller oak datasets (Crowl *et al.*, In
147 review; Eaton *et al.*, 2015; Hauser *et al.*, 2017; McVay *et al.*, 2017b,a) proved computationally
148 intractable for the current dataset. Consequently, we utilized a splits network inferred with SPLITSTREE
149 v. 14.3 (Huson & Bryant, 2006) based on the maximum-likelihood (GTR+gamma) pairwise distance
150 matrix estimated in RAxML and the same datasets utilized for the singletons tree. Full phylogenetic
151 analysis details are in the Supplement (Methods S1).

152

153 *Calibration of singletons tree*

154 Branch lengths on the tree were inferred using penalized likelihood under both a relaxed model, where
155 rates are uncorrelated among branches (Paradis, 2013), and a correlated rates model (which
156 corresponds to the penalized likelihood approach of Sanderson, 2002), as implemented in the chronos
157 function of ape v 5.1 (Paradis *et al.*, 2004) of R v 3.4.4 (“Someone to Lean On”) (R-Development-
158 Core-Team, 2004). Nodes were calibrated in two different ways, either using eight fossil calibrations,
159 corresponding to the crown of the genus and seven key clades (Fig. S2a; Table 1), or more
160 conservatively as stem ages, using a subset of five fossils (Fig. S2b; Table 1). The two calibrations
161 (referred to as the ‘crown calibration’ and ‘stem calibration’ respectively) bracket what we consider to

162 be plausible age ranges for the tree. A separate estimate of the best fit λ for the correlated clock model
163 was made using cross-validation as implemented in the `chronopl` function of `ape`, and that value of λ
164 was used for both the relaxed and correlated clocks. Comparison of Δ IC was used to identify the best
165 fit model for each value of λ . Analysis details are in the Supplement (Methods S1)

166 Transitions in lineage diversification rates were estimated using the speciation-extinction model
167 implemented in Bayesian Analysis of Macroevolutionary Mixtures (BAMM) (Rabosky, 2014); the
168 BAMMtools R package was used for configuration and analysis of MCMC. Priors were set using the
169 `setBAMMpriors` function. Analyses were run for 4E06 generations, saving every 2000 generations,
170 with four chains per MCMC analysis. To visualize changes in standing diversity over time for the
171 different sections, we plotted lineage through time (LTT) plots by section against $\delta^{18}\text{O}$ levels reported
172 in Zachos *et al.* (2001) as a temperature proxy. Analysis details are in the Supplement (Methods S1).

173

174 *Investigating the genomic landscape of oak evolutionary history*

175 Introgressive status of loci for two known introgression events involving the Eurasian white oaks
176 (McVay *et al.*, 2017b) and the western North American lobed-leaf white oaks (McVay *et al.*, 2017a)
177 was assessed by calculating the likelihood of phylogenies inferred for each locus under the constraint
178 of the inferred divergence history (species tree) and the gene flow history at odds with that divergence
179 history, as inferred in the studies cited above. These two cases are of particular interest because they
180 are well studied, and lineage sorting has been ruled out in the above studies as an explanation of
181 incongruence between the alternative topologies we test. Position of loci with a relative support of at
182 least 2 log-likelihood points for one history relative to the other were mapped back to the *Quercus*
183 *robur* genome (Plomion *et al.*, 2018). Analysis details are in the Supplement (Methods S1).

184 To identify relative phylogenetic informativeness of loci, two tests were conducted based on the
185 singletons tree. First, the ML topology was estimated in RAxML for each of 2,762 mapped, rootable
186 loci of at least 10 individuals that resolved at least one bipartition. Overall, locus trees resolved an
187 average of 4.48 (+/- 1.83 s.d.) nodes, with a maximum of 15 and a median of 4. These were compared
188 with the total-evidence tree using quartet similarities using the `tqDist` algorithm (Sand *et al.*, 2014) in
189 the `Quartet` package (Smith, 2019). We used as our similarity metric the number of quartets resolved
190 the same way for both the locus tree and the whole singletons tree divided by the sum of quartets
191 resolved the same or differently. Then, these same locus trees were mapped back to the singletons tree
192 using `phyparts` (Smith *et al.*, 2015), which identifies for all branches on a single tree how many
193 individual locus trees support or reject that branch. We tested for genomic autocorrelation in

194 phylogenetic signal using spline correlograms (Bjørnstad & Falck, 2001; Bjørnstad, 2008), with each
195 chromosome tested independently. Analysis details are in the Supplement (Methods S1).

196

197 **Results**

198 *RAD-seq data matrix*

199 RAD-seq library preps and sequencing yielded a mean of $1.685E06 \pm 1.104E06$ (s.d.) raw reads per
200 individual; of these, > 99.8% ($1.683E06 \pm 1.104E06$) passed quality filters. The total number of
201 clusters per individual prior to clustering across individuals was $101,895 \pm 58,810$, with a mean depth
202 of 17.2 ± 11.2 sequences per individual and cluster. Clusters with more than 10,000 sequences per
203 individual were discarded. Mean estimated heterozygosity by individual was 0.0135 ± 0.0027 , and
204 sequencing error rate was 0.0020 ± 0.0004 . After clustering, a total of 49,991 loci were present in at
205 least 15 individuals each. Each individual in the final dataset possesses $6.48\% \pm 2.48\%$ of all clustered
206 loci. The total data matrix is 4.352×10^6 aligned nucleotides in width. The singletons dataset is
207 composed of 22,432 loci present in at least 15 individuals, making up a dataset of 1.970×10^6 aligned
208 nucleotides.

209

210 *All-tips tree*

211 The all-tips tree (Fig. S1) comprises 246 named *Quercus* species, of which 99 have a single sample.
212 The remaining 147 species have an average of 3.54 ± 2.72 (s.d.) samples each. 97 of the 147 species
213 with more than one sample cohere for all samples, and only 13 have a taxonomic disparity index (TDI,
214 Pham *et al.*, 2016) of 10 or more (Table S3), suggesting taxonomic problems beyond difficulties
215 distinguishing very close relatives. All but four are Mexican species or species split between the
216 southwestern U.S. and Mexico (see Discussion). Of the others, the largest TDI values are for *Q. stellata*
217 and *Q. parvula* of North America, *Q. hartwissiana* and *Q. petraea* of western Eurasia, all with a
218 complicated taxonomic history.

219 The topology of the all-tips tree closely matches previous analyses based on fewer taxa (McVay
220 *et al.*, 2017b; Hipp *et al.*, 2018; Deng *et al.*, 2018) for all sections except sections *Quercus* and
221 *Virentes*. Unlike previous analyses, the all-tips topology embeds the long-branched section *Virentes*
222 within section *Quercus*, sister to a clade comprising the SW US and Mexican clade and the *Stellatae*
223 clade. This appears to be an artefact of clustering, as prior analyses of the same taxa do not reveal this
224 topology, and unconstrained analysis of these taxa also recovers this aberrant topology. As a
225 consequence, we consider the large-scale topology of the white oaks *s.l.* not to be reliable in the all-tips

226 tree, and as this topology is well resolved in prior works (McVay *et al.*, 2017b,a), we constrain the
227 singletons topology as described in the methods section.

228

229 *Topology and timing of the oak phylogeny*

230 Between the correlated and relaxed models of molecular rate heterogeneity, the correlated rates model
231 (i.e., the penalized likelihood approach of Sanderson 2002) is consistently favored using Δ IC except at
232 λ of 0, when the models are identical (Table S4). Though dating estimates differ little from $\lambda = 0$ to $\lambda =$
233 10 (not shown, but reproducible using code archived for this paper), cross-validation shows lowest
234 sensitivity of taxon-removal on dating estimates at $\lambda = 1$.

235 Analyses with the crown-age calibrations (Fig. 1, S3a) suggest an older origin of most sections than
236 proposed in prior studies (e.g., Cavender-Bares *et al.*, 2015; Hipp *et al.*, 2018; Deng *et al.*, 2018), in
237 part because in the current study we had access to a more comprehensive picture of the fossil record in
238 oaks, including fossils used as age priors that predate those used in earlier studies. Section *Virentes* in
239 our analysis has a crown age of ca. 30 Ma, whereas Cavender-Bares *et al.* (2015) estimated the crown
240 age at 11 Ma. Even under the stem-age calibrations (Fig. 1; Fig. S3b, c), we estimate the crown age of
241 *Virentes* at close to the Oligocene-Miocene boundary (ca. 23 Ma), nearly twice as old as prior
242 estimates. Sections *Quercus* and *Lobatae* had an Oligocene crown constraint (31 Ma) in our previous
243 work (Hipp *et al.*, 2018); in the current study, they were constrained to a mid-Eocene origin (45–48
244 Ma) for the crown calibration, while the stem calibration recovers a late-Eocene origin for the red oaks
245 (39 Ma) while the white oaks float down to a mid-Oligocene crown age (28 Ma). In the previous study
246 of section *Cyclobalanopsis*, a minimum age of 33 Ma was set as a constraint at the root of subgenus
247 *Cerris*, leading to a late Oligocene crown age for section *Cyclobalanopsis* (Deng *et al.*, 2018); by
248 contrast we recover an early Eocene crown age (38 Ma) for the group under the crown calibration, late
249 Eocene (36 Ma) under the stem calibration. Given the high fossil density in *Quercus* (Table 1 and
250 references therein; also reviewed in part in Denk & Grimm, 2009; Grímsson *et al.*, 2015; Denk *et al.*,
251 2017), the potential for alternative interpretations of their placement, and disparity among alternative
252 methods for modeling (Paradis, 2013; Donoghue Philip C. J. & Yang Ziheng, 2016), we leave an
253 investigation of a broader range of dating scenarios to later studies.

254 White oaks *s.str.* are estimated in the crown-calibration analysis to have arrived in Eurasia some
255 point in the Oligocene, close to the split between the section *Ponticae* sisters, which despite their
256 morphological similarity appear to have diverged from one another nearly twice as long ago as the
257 crown age of the Mexican white oaks; under the stem-calibration, the Eurasian white oaks are
258 approximately half the crown age of the *Ponticae*. By contrast with the two species of sect. *Ponticae*,

259 the Mexican white oak ancestor gave rise to an estimated 80 species in approximately half the time.
260 The Roburoids had divided into a European and an East Asian clade by the early Miocene under the
261 crown calibration, the late Miocene under the stem calibration.

262 Under the diversification scenarios implied by both the crown and the stem calibrations (Fig. 1, 2),
263 there are four relatively recent and nearly simultaneous upticks in diversification: white oaks of Mexico
264 and Central America; the red oaks of Mexico and Central America; the Eurasian (Roburoid) white
265 oaks; and the *Glauca*, *Semiserrata*, and *Acuta* clades of section *Cyclobalanopsis*. In addition, the
266 Eurasian white oaks and the southeastern U.S. white oaks (the *Stellatae* clade) and red oaks (the
267 *Laurofoliae* clade) show a lesser increase in diversification rate in both analyses, and the clade of
268 section *Ilex* that includes the Himalayan and Mediterranean species shows an uptick in diversification
269 rate in the stem calibration. This result is robust to missing taxa, as we find essentially the same clades
270 increasing in rate even assuming the 40% of missing taxa in our study were missing at random from the
271 tree (Fig. S3a-c), with the addition of a portion of section *Ilex* and some of the eastern North American
272 taxa as high-rate clades under the global sampling proportions model.

273

274 *Genomic arrangement of RAD-seq loci*

275 A total of 39,860 loci aligned to at least one position on the oak genome. The 12
276 “pseudochromosomes” (inferred linkage groups, corresponding to the 12 *Quercus* chromosomes) as
277 well as 360 scaffolds that did not map to the linkage groups were targeted by these loci. A total of
278 19,468 loci mapped to a unique position on a scaffold placed to one of the 12 oak genome
279 pseudochromosomes, an average of $1,622.3 \pm 575.4$ (s.d.) per chromosomes. Of these, $31.7\% \pm 8.1\%$
280 overlapped with the boundaries of a gene model (Fig. 3), despite the fact that only 10.1% of the 716
281 Mb of the *Quercus robur* genome that fall within the 12 pseudochromosomes fall within the endpoints
282 of a gene model.

283 For the tests of introgression, 2,422 loci had taxon sampling appropriate to testing for
284 introgression involving *Q. macrocarpa* and *Q. lobata* (the *Dumosae* alternative topologies); 2,228 were
285 suitable to testing for introgression involving the Roburoid white oaks and *Q. pontica* (the Roburoid
286 alternative topologies); and 728 were suitable to testing both. Because we were interested in
287 investigating genomic overlap in support for different areas of the species tree, we limited ourselves to
288 the 728 loci that were potentially informative about both situations. Of these, 418 mapped to one
289 position on one of the *Quercus robur* pseudochromosomes; and of these, 297 exhibited a log-likelihood
290 difference of at least 2.0 between the better and more poorly supported topology for the *Dumosae*
291 hypothesis or the Roburoid hypothesis, or both (Fig. 4). There was no correlation between the Roburoid

292 and Dumosae hypotheses ($r = -0.0286$, $p = 0.4878$), meaning that loci that support or reject either of the
293 Roburoid hypotheses do not correlate with a particular Dumosae hypothesis. Moreover, whether or not
294 a locus is located within one of the *Q. robur* gene models has no effect on whether it recovers the
295 introgression or the divergence history for the Roburoid oaks ($F_{1,366} = 0.6494$, $p = 0.4209$) or the
296 Dumosae ($F_{1,415} = 0.0377$, $p = 0.8461$).

297 Quartet similarity—the number of taxon quartets with a topology shared between trees over the
298 total number of quartets that both trees are informative about—between the RAD-seq individual-locus
299 trees and the singletons tree (Fig. S4) is similarly uninfluenced by presence in one of the gene models
300 presented in the *Quercus robur* genome (Plomion *et al.*, 2018) ($F_{1,2542} = 0.0495$, $p = 0.8239$) and shows
301 no evidence of genomic auto-correlation (Fig. S5). Rather, loci that support the tree are distributed
302 across the genome. The same is true using locus trees to investigate the support for selected nodes of
303 the phylogeny, all strongly supported (bootstrap support > 95% for all nodes tested; Fig. S1) (Fig. 5).
304 The 2762 RAD-seq locus trees made 4,745 branch-level support claims and 27,283 conflict claims on
305 the singletons tree, of which 6,409 total claims pertain to the nodes investigated, ranging from 107 to
306 1,055 per node (427.3 ± 273.7 ; Fig. 5). The locus-by-locus incongruence is high at this level: the
307 proportion of loci concordant with each node averages 0.2395 ± 0.2523 , but the range is high, from
308 0.6879 for the genus as a whole to as low as 0.0075 for the Mexican red oaks and 0.0088 for the
309 Mexican white oaks (Table S5). There is no genomic autocorrelation in support vs. rejection of nodes
310 in the singletons tree by individual locus trees (as inferred using phyparts; Smith *et al.*, 2015) (Fig. S6),
311 but the correlation between the crown age of clades investigated and the proportion of loci concordant
312 with the crown age is positive and moderately significant ($r = 0.4996$, $p = 0.0579$; Fig. S7). Three
313 clades stand out as outliers for high proportion of loci supporting divergence (outside the 95%
314 regression CI): the genus as a whole, and sections *Cerris* and *Ilex*. This widespread genomic
315 incongruence is reflected in broad network-like reticulation in the neighbor-net tree at the base of most
316 clades (Fig. 6).

317

318 **Discussion**

319 Our analyses demonstrate that the diversity of oaks we observe today reflects deep geographic
320 separation of major clades within the first 15 million years after the origin of the genus, and that
321 standing species diversity arose mostly within the last 10 million years, predominantly in four rapidly
322 diversifying clades that together account for ca. 60% of the diversity of the genus. Previous work has
323 demonstrated American oak diversity was shaped in large part by ecological opportunity, first by the
324 space left by tropical forests as they receded from North America, then by migration into the mountains

325 of Mexico (Hipp *et al.*, 2018; Cavender-Bares *et al.*, 2018). The current study deepens this
326 understanding by demonstrating two increases in diversification rates in Eurasia: one in the Eurasian
327 white oaks, which arrived from eastern North America 7.5 to 18 Ma to low continental oak diversity,
328 and no closely related oaks; and one in the southeast Asian section *Cyclobalanopsis*, driven by
329 changing climates and the Himalayan orogeny (Deng *et al.*, 2018). At the same time, our work
330 demonstrates widespread genomic incongruence in phylogenetic history, with alternative phylogenetic
331 histories interleaved across all linkage groups. Contrary to our hypothesis at the outset of this study,
332 there appear to be no regions of the genome that on their own define the entire oak phylogeny. Instead,
333 the primary divergence history of oaks (Crowl *et al.*, In review; McVay *et al.*, 2017b) knits together
334 and emerges from a patchwork of histories that comprise the oak genome.

335

336 *Topology and timing of the global oak phylogeny*

337 Our work indicates that by the mid-Eocene (45 Ma), all *Quercus* sections (*vide* Denk *et al.*, 2017),
338 representing eight major clades of the genus, had originated with the possible exception of section
339 *Quercus*, which under the stem calibrations scenario arose at the Eocene-Oligocene boundary (33 Ma).
340 Following this compressed interval of crown radiation, diversification rates spiked in the late Miocene
341 to Pliocene, ca. 10 Ma (Fig. 2), primarily in southeast Asia, Mexico, and the white oaks of Eurasia. The
342 eight fossil calibrations that we utilize here, and the two alternative methods of calibrating the tree (Fig.
343 S3a-c), bracket what we consider to be a wide range of the plausible diversification times for the genus;
344 so that while additional calibrations and a wider range of rate models bear investigation, we consider
345 this overall finding for the shape and timing of oak diversification to be reasonable.

346 While *Quercus* arose at around the early Eocene climatic optimum (the earliest known *Quercus*
347 fossil is pollen from Sankt Pankratz, Austria, 47°45' N latitude, ca. 56 Ma; Hofmann *et al.*, 2011), early
348 fossils range as far north as Axel Heiberg Island in far northern Canada, which at 79° (both modern and
349 paleolatitude in early Eocene; Scotese, 2014) is 20° further north than the northernmost oak populations
350 today. As it followed the cooling climate southward, the genus remained largely a lineage of the
351 northern temperate zone with some species of sections *Virentes*, *Lobatae*, and *Quercus* inhabiting
352 tropical climates; but even these possess physiological adaptations that reflect their temperate ancestry
353 (Cavender-Bares, 2019). In Eurasia, section *Cyclobalanopsis* dominates in subtropical evergreen
354 broadleaf forests (Deng *et al.*, 2018), but the sister sections *Cerris* and *Ilex* are temperate to
355 Mediterranean. This climatic conservatism structures the geographic distribution of oak clades at
356 several levels. Geographic patterns among and within major clades in the American oaks (subg.
357 *Quercus*) have already been studied in detail, with geographic differentiation among the western U.S.,

358 the eastern U.S., and the southwestern U.S. and Mexico / Central America in each of two sections
359 approximately simultaneously (Hipp *et al.*, 2018). The current phylogeny makes clear that in the
360 Eurasian white oaks of sect. *Quercus*, the Roburoid clade, the morphologically distinctive
361 Mediterranean, dry-adapted species often treated as subsection *Galliferae* (e.g., Tschan & Denk, 2012)
362 are distributed among all four subclades, suggesting that adaptations to the Mediterranean climate are
363 convergent within the Roburoid clade; as discussed below under *Rapid diversification of the Eurasian*
364 *white oaks*, it is geography rather than ecology or morphology that defines clades: species within clades
365 are mostly separated by ecology, not geography. Likewise, the western Eurasian members of section
366 *Ilex* form an inclusive subtree, in which the two widespread Mediterranean species *Q. coccifera* and *Q.*
367 *ilex* are clearly separated and placed sister to the montane Asian clade. The geographically most distant
368 species of the section are also genetically most distinct (Fig. 6). Even within clades, geographic
369 structuring is evident. In section *Cerris*, for example, the east and west Eurasian species group in sister
370 clades; within these latter, the western Mediterranean *Q. crenata* and *Q. suber* ‘corkish oaks’, the Near
371 East ‘Aegilops’ oaks (*Q. brantii*, *Q. ithaburensis*, *Q. macrolepis*), and the remaining central-eastern
372 Mediterranean members of the section are clearly separated. Within sect. *Quercus*, the North American
373 Prinoids and Albae form a grade, reflecting diversification in North America predating dispersal of the
374 Roburoid ancestor back to Eurasia. Once established in Eurasia, this lineage then diverged into East
375 Asian and western Eurasian sister clades, ca. 10 My after isolation from its North American ancestors.
376 Geography is imprinted in the oak phylogeny across clades, time periods, and continents.

377 Despite the older crown-age inferences in the current study in comparison to the RAD-seq
378 studies of 2015–2018, relative dates in the present study confirm earlier results that the American oaks
379 increased in diversification rate as they entered Mexico (in both red oaks and white oaks). It broadens
380 this perspective with a global sample, providing evidence that the relative diversification rate of the
381 *Glauca*, *Acuta*, and *Semiserrata* clades of the semitropical southeast Asian section *Cyclobalanopsis* is
382 comparable to if not higher than the Mexican diversification. To a lesser extent, the Eurasian white
383 oaks (the Roburoid clade) also show an increased rate of diversification. It is worth noting that the
384 crown age of the Roburoid clade as a whole may be younger than our inferences, as fossil data raise
385 some questions as to whether the Old World Roburoids were already isolated by the early Oligocene.
386 Eocene sect. *Quercus* from Axel Heiberg Island (Canada), for example, appears to be closely allied
387 with East Asian white oaks, and *Quercus furuhjelmi* from the Paleogene of Alaska and central Asia
388 might belong to any of the modern New World or Old World white oak lineages, as might the early
389 Oligocene *Quercus kodairae* and *Q. kobatakei* from Japan (Camus, 1936, 1938; Tanai & Uemura,
390 1994; Menitsky, 2005; Denk & Grimm, 2010; Tschan & Denk, 2012). Whereas previous analysis of

391 *Fagus* (Fagaceae) found an unambiguous deep split between North American and Eurasian beech
392 species that was also backed by fossils (Renner *et al.*, 2016), the fossil data we have to date do not
393 conclusively pin down the divergence between the North American and Eurasian white oaks. By
394 contrast, the inferred early Miocene split between western Eurasian and East Asian white oaks is
395 compatible with fossil evidence (Denk & Grimm, 2010), lending support to the observed increase in
396 diversification rates observed in this study.

397

398 *Taxonomy of the Mexican and Central American oaks*

399 The general high species-coherence we observe in the all-tips tree provides strong evidence that oak
400 species, in general, are genetically coherent biological entities. The fact that 97 of the 147 species with
401 more than one sample cohere for all samples provides the broadest test to date of species coherence in
402 oaks. Among the species that do not exhibit coherence, the majority are from Mexico. Two sets of
403 examples suggest that the Mexican oaks, while having been the focus of extensive taxonomic study
404 (e.g., Trelease, 1924; Spellenberg & Bacon, 1996; Spellenberg *et al.*, 1998; González-Villarreal, 2003;
405 Valencia-A., 2004; de Beaulieu & Lamant, 2010), may harbor even higher species diversity than
406 current estimates. The examples of *Quercus laeta* (González-Elizondo *et al.*, In prep.) and *Q. konzattii*
407 (McCauley *et al.*, In revision; McCauley & Oyama, In prep.) exemplify a problem likely to be common
408 in Mexican oaks. Both species have samples from northern and central to southern Mexico.

409 Researchers working with them have noticed that northern and southern populations differ and may
410 constitute separate species as our molecular data suggest. These samples are from two centers of
411 Mexican oak diversity (Torres-Miranda *et al.*, 2011, 2013; Rodríguez-Correa *et al.*, 2015) and may
412 reflect even higher species diversity in areas already known for high diversity. Interestingly, the
413 observed divergence between northern Mexico and the Jalisco and Oaxaca samples in these examples
414 appear to correlate with the formation of the Tepic-Zacoalco rift 5.5 Ma in the Jalisco block (Ferrari &
415 Rosas-Elguera, 2000) and not with climatic transitions during the Pleistocene, which has been argued
416 to be more a period of population movement than of speciation in the neotropics (Bennett *et al.*, 2012).
417 Notably, one of the youngest groups in the white oaks is located in the Sierra Madre Occidental, which
418 harbors great habitat diversity in relatively small areas (Torres-Morales *et al.*, 2010). The rugged and
419 relatively young topography, a product of magmatism and subduction processes that lasted up through
420 12 Ma (Ferrari *et al.*, 2018), and the convergence of temperate and tropical climates shaped the high
421 diversification rates.

422 Several other cases of confusing taxonomy involving Mexican and Central American species
423 are less clear. For example, the sect. *Lobatae* complex involving *Q. eugeniifolia*, *Q. benthamii*, *Q.*

424 *cortesii* and *Q. lowilliamsii*, has a history of extensive taxonomic complication (Quezada Aguilar *et al.*,
425 2016). The current work provides evidence that the species constitute a complex meriting more
426 attention and draws attention to the possibility that Central American oak diversity and the role of
427 Central American geology in Neotropical oak diversification has been underestimated (Cárdenes-Sandí
428 *et al.*, 2019), overshadowed as they have been by interest in the Mexican oak diversification (Quezada
429 Aguilar *et al.*, 2017). In the white oaks *s.str.* (sect. *Quercus*), cases such as *Q. insignis* and *Q.*
430 *corrugata* seem even more obscure. Field observations (HG-C) suggest subtle differences between *Q.*
431 *insignis*, a species of conservation concern from Jalisco, Oaxaca, Chiapas and Veracruz (Jerome,
432 2018), and *Q. corrugata* (from Chiapas and Oaxaca), but our molecular data are inconclusive. In
433 general, taxonomy of the recently diverged or still-diverging Mexican species is particularly
434 complicated because of extensive hybridization and introgression, even among relatively distantly
435 related species (Spellenberg, 1995; Bacon & Spellenberg, 1996; González-Rodríguez *et al.*, 2004;
436 Bacon *et al.*, 2011) and the dynamics of recent or ongoing speciation.

437

438 *Rapid diversification of Eurasian white oaks*

439 Among the long-studied oaks of Eurasia (e.g., Camus, 1936, 1938, 1952; Schwarz, 1993; Menitsky,
440 2005), the data presented here point to the important role of ecological and morphological convergence
441 among unrelated oaks. Phylogeny of the Eurasian white oaks (the Roburoid clade of section *Quercus*)
442 has not previously been addressed in detail, despite their importance to our understanding of oak
443 biodiversity and biology (cf. Kremer *et al.*, 1991; Dumolin-Lapegue *et al.*, 1997; Petit *et al.*, 1997;
444 Leroy *et al.*, 2017 and references therein). Previous work has sampled a maximum of 14 Roburoid
445 species (Hubert *et al.*, 2014), but not recovered the monophyly of the clade, much less relationships
446 among species. Our study includes 23 of the estimated 25 Roburoid white oak species, the strongest
447 sampling to date. The late Miocene increase in diversification rate inferred in our study at the base of
448 the western Eurasian white oaks clade is a particularly exciting finding, as it is one of only four major
449 upticks in diversification inferred in our study. Our sampling of northern temperate white and red oaks
450 is almost complete, and we have accounted for sampling bias in our diversification analyses, making it
451 unlikely that the increase in diversification rate detected here is artefactual. The fact that the Roburoids
452 are a northern temperate clade makes their radiation notable.

453 The unexpected increase in diversification rate in the Roburoids parallels the sympatric
454 diversification of red and white oaks in North America, with divergence within clades and geographic
455 regions accompanying convergence between clades (Cavender-Bares *et al.*, 2018). As in the Mexican
456 oak diversification (Torres-Miranda *et al.*, 2011; Rodríguez-Correa *et al.*, 2015), the western Eurasian

457 white oaks are ecologically diverse, ranging from lowland swamp to Mediterranean scrub, steppe and
458 from mesic lowland forests to subalpine timberline (de Beaulieu & Lamant, 2010). The European
459 Roburoid clades are not readily diagnosable morphologically, and the morphological and ecological
460 convergence among clades has led to taxonomic confusion. The morphologically distinctive
461 Mediterranean, dry-adapted species (subsection *Galliferae*; cf. Tschan & Denk, 2012), for example, are
462 distributed among all four subclades. Conversely, Roburoid clades 1 to 4 show geographic sorting
463 whereas differentiation within clades commonly reflects ecological and climatic niche evolution along
464 with morphological adaptations (e.g. from deciduous large lobed leaves to small, brevideciduous,
465 unlobed leaves). Our study thus demonstrates that across the genus, ecological diversification within
466 clades has shaped diversification.

467

468 *Genomic landscape of the global oak phylogeny*

469 The current study uses mapped phylogenomic markers to demonstrate that the oak tree of life is etched
470 broadly across the genome. Previous work demonstrated that approximately 19% of RAD-seq loci were
471 associated with ESTs (Hipp *et al.*, 2014), but that the EST-associated RAD-seq loci analyzed alone did
472 not yield a topology that was different or differently supported than the RAD-seq loci not associated
473 with EST markers, and that they were not differently apportioned to the base or the tips of the
474 phylogeny (which might have suggested that RAD-seq loci associated with coding regions were more
475 or less conservative or more or less homoplasious than the remainder). In the current study, 6,099
476 (31.3%) of RAD-seq loci in our dataset that map uniquely to one position in the genome do so in or
477 overlapping with a predicted gene in the *Quercus robur* genome (as expected from a methylation-
478 sensitive restriction enzyme; Rabinowicz *et al.*, 2005; Pegadaraju *et al.*, 2013). Our work demonstrates
479 that gene-based RAD-seq loci do not differ from non-gene-based RAD-seq loci in similarity to the
480 consensus tree or on introgression rates in the Roburoids and the Dumosae. Gene identity tells us little
481 or nothing about how reliably a region of the genome records phylogenetic history.

482 At the same time, non-significant correlation between loci that strongly differentiate alternative
483 topologies in the Dumosae and Roburoids suggests that these stories segregate nearly independently on
484 the genome. There is also no evidence of genomic autocorrelation of phylogenetic informativeness in
485 our study, despite the fact that our study has more mapped markers that significantly differentiate
486 topologies in at least one of these parts of the tree than a previous study investigating genomic
487 architecture of differentiation at the species level ($N = 158$ mapped markers with known G_{ST} ; Scotti-
488 Saintagne *et al.*, 2004). Our hypothesis that there are particular genes or regions of the genome that
489 define the oak phylogeny globally appears to be false: rather, the phylogenetic history of oaks is

490 defined by different genes in different lineages, making evolutionary history of oaks a phylogenetic and
491 genomic mosaic. The effort to find a single best suite of genes for phylogenetic or population genetic
492 inference across the oak genus is thus unlikely to be successful, though markers can clearly be designed
493 for individual clades (Guichoux *et al.*, 2011; Fitzek *et al.*, 2018). What is perhaps most remarkable is
494 that this heterogeneity of histories covarying independently along the oak genome yields, in aggregate,
495 an evolutionary history of the complex genus that mirrors the morphological and ecological diversity of
496 living and fossil oak species.

497

498 **Conclusion**

499 Questions about the genomic architecture of population differentiation and speciation are generally
500 asked at fine scales (Leroy *et al.*, 2017, 2018), at the point at which population level processes directly
501 shape genomic differentiation. But microevolution—comprising processes at the population level—
502 leaves an imprint in the phylogeny; when such impressions persist, they can often be detected using
503 topological methods that may be sensitive even to introgression along internal phylogenetic branches
504 (Eaton *et al.*, 2015; Solís-Lemus & Ané, 2016; McVay *et al.*, 2017b). With multiple Fagaceae genomes
505 now becoming available (Staton *et al.*, 2015; Plomion *et al.*, 2016, 2018; Sork *et al.*, 2016; Ramos *et*
506 *al.*, 2018), we may soon be able to detangle the mosaic history of oaks and understand what story each
507 gene tells. The current study makes clear that the phylogeny we unravel will neither be unitary nor told
508 by a small subset of the genome, as the regions of the genome capturing the divergence history for one
509 clade are not the regions capturing the divergence history of another. Understanding phylogenetic
510 history in the face of this variation is only one problem. It will be followed by a greater one: how do we
511 interpret the history of oak diversification in space and time if it is really a collection of diverse
512 histories from different regions of the genome, all reflecting different evolutionary pathways, all
513 equally real?

514

515 **Acknowledgements**

516 Funding for this project was provided by U.S. National Science Foundation Awards 1146488 to ALH,
517 1146102 to PSM and 1146380 to JCB; Swedish Research Grants 2015-03986 to TD and 2009-00000 to
518 GWG; support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig
519 funded by the German Research Foundation (FZT 118) to E-DS and GWG; and The Morton
520 Arboretum Center for Tree Science. This paper is dedicated to the memory of Michael Avishai (1935-
521 2018), founder of the Jerusalem Botanical Gardens and cherished colleague.

522

523

524 **Author Contributions**

525 ALH, PSM, JCB, MD, AK, CP, and AG-R conceived and designed the study. ALH, PSM, MH, MA,
526 JCB, MD, TD, OG, MSG-E, AG-R, GWG, X-LJ, JDM, HR-C, MCS, VLS, and SV-A collected,
527 identified, and curated samples. ALH, PSM, MH, JCB, AC, MD, TD, AG-R, GWG, X-LJ, JDM, VLS
528 generated and analyzed phylogenetic data. CB, AK, IL, CP generated and analyzed genomic data.
529 ALH, PSM, TD and GWG drafted the manuscript. All authors wrote and edited the manuscript.

530

531

532 **References**

533 **Anderson E. 1953.** Introgressive Hybridization. *Biological Reviews* **28**: 280–307.

534 **Bacon JR, Dávila-Aranda PD, Spellenberg R, González-Elizondo MS. 2011.** The taxonomic status
535 of the Mexican oak *Quercus undata* (Fagaceae, *Quercus*, section *Quercus*). *Revista Mexicana de*
536 *Biodiversidad* **82**: 1123–1131.

537 **Bacon JR, Spellenberg R. 1996.** Hybridization in two distantly related Mexican black oaks *Quercus*
538 *conzattii* and *Quercus eduardii* (Fagaceae: *Quercus*: section *Lobatae*). *SIDA, Contributions to Botany*
539 **17**: 17–41.

540 **Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA,**
541 **Johnson EA. 2008.** Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers.
542 *PLOS ONE* **3**: e3376.

543 **de Beaulieu A le H, Lamant T. 2010.** *Guide Illustré des Chênes*. Geer, Belgium: Edilens.

544 **Bennett K, Bhagwat S, Willis K. 2012.** Neotropical refugia. *The Holocene* **22**: 1207–1214.

545 **Bjørnstad ON. 2008.** *ncf: spatial nonparametric covariance functions*.

546 **Bjørnstad ON, Falck W. 2001.** Nonparametric spatial covariance functions: Estimation and testing.
547 *Environmental and Ecological Statistics* **8**: 53–70.

548 **Boecklen WJ. 2017.** Topology of syngameons. *Ecology and Evolution* **7**: 10486–10491.

549 **Botsyun S, Sepulchre P, Donnadieu Y, Risi C, Licht A, Rugenstein JKC. 2019.** Revised
550 paleoaltimetry data show low Tibetan Plateau elevation during the Eocene. *Science* **363**: eaaq1436.

551 **Bryant D, Moulton V. 2004.** Neighbor-Net: An Agglomerative Method for the Construction of
552 Phylogenetic Networks. *Molecular Biology and Evolution* **21**: 255–265.

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

- 555 **Camus AA. 1936.** *Monographie du genre Quercus. Tome I. Genre Quercus. Sous-genre*
556 *Cyclobalanopsis et sous-genre Euquercus (Section Cerris et Mesobalanu).* Paris: Editions Paul
557 Lechevalier.
- 558 **Camus AA. 1938.** *Monographie du genre Quercus. Tome II. Genre Quercus. Sous-genre Euquercus*
559 *(sections Lepidobalanus et Macrobalanus).* Paris: Editions Paul Lechevalier.
- 560 **Camus AA. 1952.** *Monographie du genre Quercus. Tome III. Genre Quercus. Sous-genre Euquercus*
561 *(sections Protobalanus et Erythrobalanus).* Paris: Editions Paul Lechevalier.
- 562 **Cannon CH, Lerdau M. 2015.** Variable mating behaviors and the maintenance of tropical
563 biodiversity. *Frontiers in Genetics* **6**.
- 564 **Cannon CH, Scher CL. 2017.** Exploring the potential of gametic reconstruction of parental genotypes
565 by F1 hybrids as a bridge for rapid introgression. *Genome* **60**: 713–719.
- 566 **Cárdenes-Sandí GM, Shadik CR, Correa-Metrio A, Gosling WD, Cheddadi R, Bush MB. 2019.**
567 Central American climate and microrefugia: A view from the last interglacial. *Quaternary Science*
568 *Reviews* **205**: 224–233.
- 569 **Cavender-Bares J. 2019.** Diversification, adaptation, and community assembly of the American oaks
570 (*Quercus*), a model clade for integrating ecology and evolution. *New Phytologist* **221**: 669–692.
- 571 **Cavender-Bares J, Gonzalez-Rodriguez A, Eaton DAR, Hipp AL, Beulke A, Manos PS. 2015.**
572 Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): A genomic and
573 population genetics approach. *Molecular Ecology* **24**: 3668–3687.
- 574 **Cavender-Bares J, Kothari S, Meireles JE, Kaproth MA, Manos PS, Hipp AL. 2018.** The role of
575 diversification in community assembly of the oaks (*Quercus* L.) across the continental U.S. *American*
576 *Journal of Botany* **105**: 565–586.
- 577 **Crowl AA, McVay JD, Manos PS, Hipp AL, Lemmon A, Lemmon E. In review.** Uncovering the
578 genomic signature of ancient introgression between white oak lineages (*Quercus*) using anchored
579 enrichment. *New Phytologist*.
- 580 **Deng M, Jiang X-L, Hipp AL, Manos PS, Hahn M. 2018.** Phylogeny and biogeography of East
581 Asian evergreen oaks (*Quercus* section *Cyclobalanopsis*; Fagaceae): Insights into the Cenozoic history
582 of evergreen broad-leaved forests in subtropical Asia. *Molecular Phylogenetics and Evolution* **119**:
583 170–181.
- 584 **Denk T, Grimm GW. 2009.** Significance of Pollen Characteristics for Infrageneric Classification and
585 Phylogeny in *Quercus* (Fagaceae). *International Journal of Plant Sciences* **170**: 926–940.
- 586 **Denk T, Grimm GW. 2010.** The oaks of western Eurasia: Traditional classifications and evidence
587 from two nuclear markers. *Taxon* **59**: 351–366.
- 588 **Denk T, Grimm GW, Grímsson F, Zetter R. 2013.** Evidence from ‘Köppen signatures’ of fossil
589 plant assemblages for effective heat transport of Gulf Stream to subarctic North Atlantic during
590 Miocene cooling. *Biogeosciences* **10**: 7927–7942.

- 591 **Denk T, Grimm GW, Manos PS, Deng M, Hipp AL. 2017.** An Updated Infrageneric Classification
592 of the Oaks: Review of Previous Taxonomic Schemes and Synthesis of Evolutionary Patterns. In: Tree
593 Physiology. Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L.
594 Springer, Cham, 13–38.
- 595 **Denk T, Grimsson F, Zetter R. 2010.** Episodic migration of oaks to Iceland: Evidence for a North
596 Atlantic ‘land bridge’ in the latest Miocene. *American Journal of Botany* **97**: 276–287.
- 597 **Dodd RS, Afzal-Rafii Z. 2004.** Selection and dispersal in a multispecies oak hybrid zone. *Evolution*
598 **58**: 261–269.
- 599 **Donoghue Philip C. J., Yang Ziheng. 2016.** The evolution of methods for establishing evolutionary
600 timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**: 20160020.
- 601 **Dumolin-Lapegue S, Demesure B., Fineschi S, Come V. L, Petit RJ. 1997.** Phylogeographic
602 structure of white oaks throughout the European continent. *Genetics* **146**: 1475–1487.
- 603 **Eaton DAR. 2014.** PyRAD: assembly of de novo RADseq loci for phylogenetic analyses.
604 *Bioinformatics (Oxford, England)* **30**: 1844–1849.
- 605 **Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J. 2015.** Historical introgression
606 among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**:
607 2587–2601.
- 608 **Edelman NB, Frandsen P, Miyagi M, Clavijo BJ, Davey J, Dikow R, Accinelli GG, Belleghem SV,
609 Patterson NJ, Neafsey DE, et al. 2018.** Genomic architecture and introgression shape a butterfly
610 radiation. *bioRxiv*: 466292.
- 611 **Ferrari L, Orozco-Esquivel T, Bryan SE, López-Martínez M, Silva-Fragoso A. 2018.** Cenozoic
612 magmatism and extension in western Mexico: Linking the Sierra Madre Occidental silicic large
613 igneous province and the Comondú Group with the Gulf of California rift. *Earth-Science Reviews* **183**:
614 115–152.
- 615 **Ferrari L, Rosas-Elguera J. 2000.** Late Miocene to Quaternary extension at the northern boundary of
616 the Jalisco Block, western Mexico: The Tepic-Zacoalco Rift revised. In: Special Paper 334: Cenozoic
617 tectonics and volcanism of Mexico. Geological Society of America, 41–63.
- 618 **Fitzek E, Delcamp A, Guichoux E, Hahn M, Lobdell M, Hipp AL. 2018.** A nuclear DNA barcode
619 for eastern North American oaks and application to a study of hybridization in an Arboretum setting.
620 *Ecology and Evolution* **8**: 5837–5851.
- 621 **Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork V. 2017.** Phylogenomic inferences from
622 reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of
623 California white oaks (*Quercus* subgenus *Quercus*). *Genome* **60**: 743–755.
- 624 **Folk RA, Soltis PS, Soltis DE, Guralnick R. 2018.** New prospects in the detection and comparative
625 analysis of hybridization in the tree of life. *American Journal of Botany* **105**: 364–375.

- 626 **González-Rodríguez A, Arias DM, Valencia-A. S, Oyama K. 2004.** Morphological and RAPD
627 analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks.
628 *American Journal of Botany* **91**: 401–409.
- 629 **González-Villarreal LM. 2003.** Two new species of oak (Fagaceae, *Quercus* sect. *Lobatae*) from the
630 Sierra Madre del Sur, Mexico. *Brittonia* **55**: 49–60.
- 631 **Grímsson F, Zetter R, Grimm GW, Pedersen GK, Pedersen AK, Denk T. 2015.** Fagaceae pollen
632 from the early Cenozoic of West Greenland: revisiting Engler’s and Chaney’s Arcto-Tertiary
633 hypotheses. *Plant Systematics and Evolution* **301**: 809–832.
- 634 **Guichoux E, Lagache L, Wagner S, Léger P, Petit RJ. 2011.** Two highly validated multiplexes (12-
635 plex and 8-plex) for species delimitation and parentage analysis in oaks (*Quercus* spp.). *Molecular*
636 *Ecology Resources* **11**: 578–585.
- 637 **Hardin JW. 1975.** Hybridization and introgression in *Quercus alba*. *Journal of the Arnold Arboretum*
638 **56**: 336–363.
- 639 **Hauser DA, Keuter A, McVay JD, Hipp AL, Manos PS. 2017.** The evolution and diversification of
640 the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). *American*
641 *Journal of Botany* **104**: 1581–1595.
- 642 **Hipp AL. 2017.** American oaks in an evolutionary context. In: Jerome D, Beckman E, Kenny L,
643 Wenzell K, Kua C-S, Westwood M, eds. *The Red List of US Oaks*. Lisle: USDA Forest Service and
644 The Morton Arboretum, 10–11.
- 645 **Hipp AL. 2018.** Pharaoh’s Dance: the oak genomic mosaic. *PeerJ Preprints* **6**: e27405v1.
- 646 **Hipp AL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, Manos PS. 2014.** A framework
647 phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE* **9**: e93975.
- 648 **Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Avalos SV,**
649 **Cavender-Bares J. 2018.** Sympatric parallel diversification of major oak clades in the Americas and
650 the origins of Mexican species diversity. *New Phytologist* **217**: 439–452.
- 651 **Hofmann C-C, Mohamed O, Egger H. 2011.** A new terrestrial palynoflora from the
652 Palaeocene/Eocene boundary in the northwestern Tethyan realm (St. Pankraz, Austria). *Review of*
653 *Palaeobotany and Palynology* **166**: 295–310.
- 654 **Hubert F, Grimm GW, Jouselin E, Berry V, Franc A, Kremer A. 2014.** Multiple nuclear genes
655 stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and Biodiversity* **12**: 405–423.
- 656 **Huson DH, Bryant D. 2006.** Application of Phylogenetic Networks in Evolutionary Studies.
657 *Molecular Biology and Evolution* **23**: 254–267.
- 658 **Jerome D. 2018.** *Quercus insignis*. *IUCN Red List of Threatened Species* **2018**: e.T194177A2302931.
- 659 **Jiang X-L, Deng M, Hipp AL, Su T, Yan M-X, Zho Z-K. In review.** East Asian origins of European
660 holly oaks via the Tibet-Himalayas. *Journal of Biogeography*.

- 661 **Kim BY, Wei X, Fitz-Gibbon S, Lohmueller KE, Ortego J, Gugger PF, Sork VL. 2018.** RADseq
662 data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species
663 (*Quercus* sect. *Quercus*: Fagaceae). *Molecular Ecology* **27**: 4556–4571.
- 664 **Kremer A, Petit R, Zanetto A, Fougère V, Ducouso A, Wagner D, Chauvin C. 1991.** Nuclear and
665 organelle gene diversity in *Quercus robur* and *Q. petraea*. In: Müller-Starck G, Ziehe M, eds. Genetic
666 Variation in European Forest Trees. Frankfurt-am-Main: Sauerländer's Verlag, 141–166.
- 667 **Leroy T, Rougemont Q, Dupouey J-L, Bodenes C, Lalanne C, Belser C, Labadie K, Provost GL,
668 Aury J-M, Kremer A, et al. 2018.** Massive postglacial gene flow between European white oaks
669 uncovered genes underlying species barriers. *bioRxiv*: 246637.
- 670 **Leroy T, Roux C, Villate L, Bodénès C, Romiguier J, Paiva JAP, Dossat C, Aury J-M, Plomion
671 C, Kremer A. 2017.** Extensive recent secondary contacts between four European white oak species.
672 *New Phytologist* **214**: 865–878.
- 673 **Lewis ZA, Shiver AL, Stiffler N, Miller MR, Johnson EA, Selker EU. 2007.** High-Density
674 Detection of Restriction-Site-Associated DNA Markers for Rapid Mapping of Mutated Loci in
675 *Neurospora*. *Genetics* **177**: 1163–1171.
- 676 **Manos PS. 2016.** Systematics and biogeography of the American oaks. *International Oaks* **27**: 23–36.
- 677 **McCauley RA, Cortés-Palomec AC, Oyama K. In revision.** Species diversification in a lineage of
678 Mexican red oak (*Quercus* section *Lobatae* subsection *Racemiflorae*) – the interplay between distance,
679 habitat, and hybridization. *Tree Genetics & Genomes*.
- 680 **McCauley RA, Oyama K. In prep.** A re-evaluation of taxonomy in *Quercus* section *Lobatae*
681 subsection *Racemiflorae*, resurrection of the name *Q. pennivenia* and description of a new taxon *Q.*
682 *huicholensis*.
- 683 **McVay JD, Hauser D, Hipp AL, Manos PS. 2017a.** Phylogenomics reveals a complex evolutionary
684 history of lobed-leaf white oaks in western North America. *Genome* **60**: 733–742.
- 685 **McVay JD, Hipp AL, Manos PS. 2017b.** A genetic legacy of introgression confounds phylogeny and
686 biogeography in oaks. *Proc. R. Soc. B* **284**: 20170300.
- 687 **Menitsky YL. 2005.** *Oaks of Asia*. □: Science Publishers. Boca Raton: CRC Press.
- 688 **Miller M, Atwood T, Eames BF, Eberhart J, Yan Y-L, Postlethwait J, Johnson E. 2007a.** RAD
689 marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biology* **8**: R105.
- 690 **Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007b.** Rapid and cost-effective
691 polymorphism identification and genotyping using restriction site associated DNA (RAD) markers.
692 *Genome Research* **17**: 240–248.
- 693 **Minh BQ, Nguyen MAT, von Haeseler A. 2013.** Ultrafast Approximation for Phylogenetic Bootstrap.
694 *Molecular Biology and Evolution* **30**: 1188–1195.
- 695 **Ortego J, Gugger PF, Sork VL. 2018.** Genomic data reveal cryptic lineage diversification and
696 introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist* **218**: 804–818.

- 697 **Pääbo S. 2003.** The mosaic that is our genome. *Nature* **421**: 409–412.
- 698 **Paradis E. 2013.** Molecular dating of phylogenies by likelihood methods: A comparison of models and
699 a new information criterion. *Molecular Phylogenetics and Evolution* **67**: 436–444.
- 700 **Paradis E, Claude J, Strimmer K. 2004.** APE: Analyses of Phylogenetics and Evolution in R
701 language. *Bioinformatics* **20**: 289–290.
- 702 **Pegadaraju V, Nipper R, Hulke B, Qi L, Schultz Q. 2013.** De novo sequencing of sunflower genome
703 for SNP discovery using RAD (Restriction site Associated DNA) approach. *BMC Genomics* **14**: 556.
- 704 **Petit R, Bodenes C, Ducouso A, Roussel G, Kremer A. 2003.** Hybridization as a mechanism of
705 invasion in oaks. *New Phytologist* **161**: 151–164.
- 706 **Petit RJ, Hampe A. 2006.** Some Evolutionary Consequences of Being a Tree. *Annual Review of*
707 *Ecology, Evolution, and Systematics* **37**: 187–214.
- 708 **Petit R, Pineau E, Demesure B, Bacilieri R, Ducouso A, Kremer A. 1997.** Chloroplast DNA
709 footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*
710 **94**: 9996–10001.
- 711 **Pham KK, Hahn M, Lueders K, Brown BH, Bruederle LP, Bruhl JJ, Chung K-S, Derieg NJ,**
712 **Escudero M, Ford BA, et al. 2016.** Specimens at the Center: An Informatics Workflow and Toolkit
713 for Specimen-Level Analysis of Public DNA Database Data. *Systematic Botany* **41**: 529–539.
- 714 **Pham KK, Hipp AL, Manos PS, Cronn RC. 2017.** A Time and a Place for Everything: Phylogenetic
715 history and geography as joint predictors of oak plastome phylogeny. *Genome* **60**: 720–732.
- 716 **Plomion C, Aury J-M, Amselem J, Alaeitabar T, Barbe V, Belser C, Bergès H, Bodénès C,**
717 **Boudet N, Boury C, et al. 2016.** Decoding the oak genome: public release of sequence data, assembly,
718 annotation and publication strategies. *Molecular Ecology Resources* **16**: 254–265.
- 719 **Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillonne N,**
720 **Labadie K, Provost GL, et al. 2018.** Oak genome reveals facets of long lifespan. *Nature Plants* **4**:
721 440–452.
- 722 **Quammen D. 2018.** *The Tangled Tree: A Radical New History of Life*. New York: Simon and
723 Schuster.
- 724 **Quezada Aguilar ML, Rodas-Duarte L, Marroquín-Tintí AA. 2016.** *Diversidad de encinos en*
725 *Guatemala; una alternativa para bosques enérgicos, seguridad alimentaria y mitigación al cambio*
726 *climático. Fase II. Jutiapa, Jalapa y Santa Rosa*. Instituto de Investigaciones Químicas y Biológicas
727 (IIQB). Centro de Estudios Conservacionistas (CECON), Facultad de Ciencias Químicas y Farmacia.
728 Universidad de San Carlos de Guatemala.
- 729 **Quezada Aguilar ML, Rodas-Duarte R, Marroquín-n-Tintí- AA. 2017.** Contribución al
730 conocimiento de los encinos (*Quercus*: Fagaceae) en los departamentos de Alta Verapaz, Baja Verapaz
731 y Petén, Guatemala. *Ciencia, Tecnología y Salud* **3**: 115–126.

- 732 **Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O’Shaughnessy AL,**
733 **Nascimento LU, McCombie WR, Martienssen RA. 2005.** Differential methylation of genes and
734 repeats in land plants. *Genome Research* **15**: 1431–1440.
- 735 **Rabosky DL. 2014.** Automatic Detection of Key Innovations, Rate Shifts, and Diversity-Dependence
736 on Phylogenetic Trees. *PLoS ONE* **9**: e89543.
- 737 **Ramos AM, Usié A, Barbosa P, Barros PM, Capote T, Chaves I, Simões F, Abreu I,**
738 **Carrasquinho I, Faro C, et al. 2018.** The draft genome sequence of cork oak. *Scientific Data* **5**:
739 180069.
- 740 **R-Development-Core-Team. 2004.** *R: A language and environment for statistical computing*. Vienna.
- 741 **Ree RH, Hipp AL. 2015.** Inferring phylogenetic history from restriction site associated DNA
742 (RADseq). In: Hoerandl E, Appelhaus M, eds. *Next Generation Sequencing in Plant Systematics*.
743 Koenigstein: Koeltz Scientific Books, 181–204.
- 744 **Renner SS, Grimm GW, Kapli P, Denk T. 2016.** Species relationships and divergence times in
745 beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model.
746 *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**: 20150135.
- 747 **Rodríguez-Correa H, Oyama K, MacGregor-Fors I, González-Rodríguez A. 2015.** How Are Oaks
748 Distributed in the Neotropics? A Perspective from Species Turnover, Areas of Endemism, and Climatic
749 Niches. *International Journal of Plant Sciences* **176**: 222–231.
- 750 **Sand A, Holt MK, Johansen J, Brodal GS, Mailund T, Pedersen CNS. 2014.** tqDist: a library for
751 computing the quartet and triplet distances between binary or general trees. *Bioinformatics* **30**: 2079–
752 2080.
- 753 **Sanderson MJ. 2002.** Estimating absolute rates of molecular evolution and divergence times: a
754 penalized likelihood approach. *Molecular Biology and Evolution* **19**: 101–109.
- 755 **Schwarz O. 1993.** *Quercus L. Flora Europaea* **I**: 72–76.
- 756 **Scotese CR. 2014.** *Atlas of Paleogene Paleogeographic Maps (Mollweide Projection), Maps 8-15,*
757 *Volume 1, The Cenozoic, PALEOMAP Atlas for ArcGIS*. Evanston, IL: PALEOMAP Project.
- 758 **Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodenes C, Burg K,**
759 **Kremer A. 2004.** Genome Scanning for Interspecific Differentiation Between Two Closely Related
760 Oak Species [*Quercus robur* L. and *Q. petraea* (Matt.) Liebl.]. *Genetics* **168**: 1615–1626.
- 761 **Smith MR. 2019.** Quartet: comparison of phylogenetic trees using quartet and bipartition measures.
762 doi: 10.5281/zenodo.2536318.
- 763 **Smith SA, Moore MJ, Brown JW, Yang Y. 2015.** Analysis of phylogenomic datasets reveals
764 conflict, concordance, and gene duplications with examples from animals and plants. *BMC*
765 *Evolutionary Biology* **15**: 150.
- 766 **Solís-Lemus C, Ané C. 2016.** Inferring Phylogenetic Networks with Maximum Pseudolikelihood
767 under Incomplete Lineage Sorting. *PLOS Genetics* **12**: e1005896.

- 768 **Solís-Lemus C, Bastide P, Ané C. 2017.** PhyloNetworks: A Package for Phylogenetic Networks.
769 *Molecular Biology and Evolution* **34**: 3292–3298.
- 770 **Sork VL, Fitz-Gibbon ST, Puiu D, Crepeau M, Gugger PF, Sherman R, Stevens K, Langley CH,**
771 **Pellegrini M, Salzberg SL. 2016.** First Draft Assembly and Annotation of the Genome of a California
772 Endemic Oak *Quercus lobata* Née (Fagaceae). *G3: Genes, Genomes, Genetics* **6**: 3485–3495.
- 773 **Spellenberg R. 1995.** On the hybrid nature of *Quercus basaseachicensis* (Fagaceae, sect. *Quercus*).
774 *Sida, Contributions To Botany* **16**: 427–437.
- 775 **Spellenberg R, Bacon JR. 1996.** Taxonomy and Distribution of a Natural Group of Black Oaks of
776 Mexico (*Quercus*, Section *Lobatae*, Subsection *Racemiflorae*). *Systematic Botany* **21**: 85–99.
- 777 **Spellenberg R, Bacon JR, González Elizondo MS. 1998.** Los encinos (*Quercus*, Fagaceae) en un
778 transecto sobre la Sierra Madre Occidental. *Boletín del Instituto de Botánica de la Universidad de*
779 *Guadalajara* **5**: 357–387.
- 780 **Stamatakis A. 2006.** Phylogenetic models of rate heterogeneity: a high performance computing
781 perspective. In: Proceedings 20th IEEE International Parallel Distributed Processing Symposium. 8 pp.
- 782 **Stamatakis A. 2014.** RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large
783 Phylogenies. *Bioinformatics* **30**: 1312–1313.
- 784 **Staton M, Zhebentyayeva T, Olukolu B, Fang GC, Nelson D, Carlson JE, Abbott AG. 2015.**
785 Substantial genome synteny preservation among woody angiosperm species: comparative genomics of
786 Chinese chestnut (*Castanea mollissima*) and plant reference genomes. *BMC Genomics* **16**: 744.
- 787 **Tanai T, Uemura K. 1994.** Lobed oak leaves from the Tertiary of East Asia with reference to the oak
788 phytogeography of the northern hemisphere. *Transactions and proceedings of the Paleontological*
789 *Society of Japan. New series* **1994**: 343–365.
- 790 **Torres-Miranda A, Luna-Vega I, Oyama K. 2011.** Conservation biogeography of red oaks (*Quercus*,
791 section *Lobatae*) in Mexico and Central America. *American Journal of Botany* **98**: 290–305.
- 792 **Torres-Miranda A, Luna-Vega I, Oyama K. 2013.** New Approaches to the Biogeography and Areas
793 of Endemism of Red Oaks (*Quercus* L., Section *Lobatae*). *Systematic Biology* **62**: 555–573.
- 794 **Torres-Morales L, García-Mendoza DF, López-González C, Muñoz-Martínez R. 2010.** Bats of
795 Northwestern Durango, Mexico: Species Richness at the Interface of Two Biogeographic Regions. *The*
796 *Southwestern Naturalist* **55**: 347–362.
- 797 **Trelease W. 1924.** The American Oaks. *Memoirs of the National Academy of Sciences* **20**: 1–255.
- 798 **Tschan GF, Denk T. 2012.** Trichome types, foliar indumentum and epicuticular wax in the
799 Mediterranean gall oaks, *Quercus* subsection *Galliferae* (Fagaceae): implications for taxonomy,
800 ecology and evolution. *Botanical Journal of the Linnean Society* **169**: 611–644.
- 801 **Valencia-A. S. 2004.** Diversidad del género *Quercus* (Fagaceae) en México. *Boletín de la Sociedad*
802 *Botánica de México* **75**: 33–53.
- 803 **Van Valen L. 1976.** Ecological species, multispecies, and oaks. *Taxon* **25**: 233–239.

804 **Walker JD, Geissman JW, Bowring SA, Babcock LE. 2018.** *Geologic Time Scale v. 5.0: Geological*
805 *Society of America,*
806 https://www.geosociety.org/GSA/Education_Careers/Geologic_Time_Scale/GSA/timescale/home.aspx.

807 **Wen D, Yu Y, Zhu J, Nakhleh L. 2018.** Inferring Phylogenetic Networks Using PhyloNet. *Systematic*
808 *Biology* **67**: 735–740.

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

811 **Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001.** Trends, Rhythms, and Aberrations in
812 Global Climate 65 Ma to Present. *Science* **292**: 686–693.

813 **Zhang C, Ogilvie HA, Drummond AJ, Stadler T. 2018.** Bayesian Inference of Species Networks
814 from Multilocus Sequence Data. *Molecular Biology and Evolution* **35**: 504–517.

815

816

817 **Figure Captions**

818 **Fig. 1.** Singletons tree, calibrated using eight crown calibration fossils (solid lines) or 5 stem-
819 calibration fossils (dotted lines). Single exemplars per species were analyzed using maximum
820 likelihood; multiple samples are included for some species to represent cryptic or undescribed diversity
821 (e.g., in *Quercus arizonica*, *Q. laeta*, *Q. conzattii*) or named infraspecies (e.g., varieties of *Quercus*
822 *agrifolia* and *Q. parvula*). Labels to the right of the tree indicate subgenera (black) and sections
823 (medium gray) following the latest taxonomy for the genus (Denk *et al.*, 2017). Branch colors represent
824 net diversification rates estimated using reversible-jump MCMC in BAMM (Rabosky, 2014),
825 integrating over uncertainty in the timing and location of shifts in lineage diversification rates.
826 rjMCMC was conducted with explicit lineage-specific sampling proportions specified, and thus
827 accounts for the relatively low species sampling in the Mexican / Central American oaks and the
828 southeast Asian section *Cyclobalanopsis*. All bootstrap values > 100 except for nodes marked with *,
829 which are all 80-99 except for two: the common ancestor of *Q. costaricensis* and *Q. humboldtii* and the
830 MRCA of *Q. myrsinifolia* and *Q. salicina* both have bootstrap values < 5.

831
832 **Fig. 2.** Lineages-through-time (LTT) plot showing the diversification of five species-rich lineages
833 (sections *Cerris*, *Cyclobalanopsis*, *Ilex*, *Lobatae*, *Quercus*) within genus *Quercus* for the preferred
834 (early fossils treated as crown group representatives) and conservative dating (dimmed lines, fossils
835 treated as stem group taxa). Major tectonic events on the Northern Hemisphere (formation of the
836 Qinghai-Tibetan Plateau, QTP (Scotese, 2014; Botsyun *et al.*, 2019); closure of the Turgai Sea) and
837 global climate context (based on marine stable isotope data; Zachos *et al.*, 2001) shown for
838 comparison. Timing of onset of Arctic glaciation and viability of the North Atlantic Land Bridge for
839 oak migration are reviewed in Denk *et al.* (2010, 2013) and literature therein. Background colors
840 indicate Cenozoic epochs/periods (following Walker *et al.* (2018); from left to right: Paleocene,
841 Eocene, Oligocene, Miocene, Quaternary).

842
843 **Fig. 3.** RAD-seq loci by chromosome. RAD-seq loci mapping to a unique position on one of the twelve
844 *Quercus robur* genome pseudochromosomes are included in this figure and analyses reported in the
845 paper. Chromosome length is based on total sequence length of scaffolds assigned to the *Q. robur*
846 pseudochromosomes. Genomic position of loci overlapping vs not overlapping a gene was determined
847 by detecting overlap of the RAD-seq locus start and end points with start and end points of the 25,808
848 gene models reported for the *Q. robur* genome (Plomion *et al.*, 2018).

849

850 **Fig. 4.** Genomic distribution of loci favoring alternative placements of the Roburoid white oaks and of
851 *Quercus lobata* / *Quercus macrocarpa*. The 19,468 RAD-seq loci that map to a single position on one
852 of the *Quercus robur* pseudochromosomes are represented by gray bands; chromosomal areas of darker
853 gray have a denser mapping of RAD-seq loci. Mapped beside the chromosomes are the positions of
854 325 RAD-seq loci with a log-likelihood difference of at least 2 between trees constrained to be
855 monophyletic for the Roburoids vs those placing the Roburoids with *Q. pontica* (194 loci); those
856 differing by at least 2 between trees constrained to be monophyletic for both the Dumosae and the
857 Prinoids vs those placing *Q. lobata* or *Q. macrocarpa* in the opposing clade (290 loci); or both (159
858 loci). These two hypotheses were selected because the topological differences have been demonstrated
859 in prior studies (Crowl *et al.*, In review; McVay *et al.*, 2017b,a) to be a consequence of introgression,
860 not lineage sorting alone. The relative mapping of these loci thus provides a study in the distribution of
861 loci that are informative about population divergence history vs. ancient introgression in two closely
862 related clades. The mismatch between loci ($r = -0.0286$, $p = 0.4878$) suggests that introgression is not
863 genomically conserved.

864

865 **Fig. 5.** Loci congruent vs. discordant with key nodes of phylogeny. An average of 123.9 (± 178.9)
866 RAD-seq locus trees are informative about each of the 15 named clades represented in this figure. Dark
867 bands indicate RAD-seq loci that support a node; light bands indicate loci that conflict with it.

868

869 **Fig. 6.** Neighbor-net, planar (meta-)phylogenetic network based on pairwise ML distances. Members of
870 the major clades with unambiguous (tree) support (cf. Fig. 1) are clustered. All currently accepted
871 sections are color-coded; edge bundles defining neighborhoods corresponding to sections and infra-
872 sectional clades are colored accordingly. Main biogeographic splits within each section are indicated by
873 dotted gray lines. The graph depicts the variance in inter- and intra-sectional genetic diversity patterns.
874 The most genetically unique clades within each subgenus (sect. *Lobatae* for subgenus *Quercus*; sect.
875 *Cerris* for subgenus *Cerris*) are placed on the right side of the graph; the distance to the spider-web-like
876 center of the graph, which in this case may represent the point-of-origin (being also the mid-point
877 between all tips and the connection of both subgenera) reflects the corresponding phylogenetic root-tip
878 distances observed in the ML tree. Tree-like portions may be indicative of bottleneck situations in the
879 formation of a clade; fan-like portions reflect potential genetic gradients developed during unhindered
880 radiation (geographic expansion; note e.g. the position of Texan white and red oaks; strict West-East
881 ordering within sect. *Ilex*), i.e. absence of major evolutionary bottlenecks.

882

883 **Table 1.** Fossil calibrations used in this study, with nodes indicated as most recent common ancestor of
884 selected taxa. Max and min indicate maximum and minimum ages for calibrations in Ma. Crown
885 calibration node and stem calibration node indicate the taxa whose MRCA are the calibration points for
886 the crown and stem calibration analyses respectively. References cited in Table S2.

Node	Max	Min	Crown calibration node	Stem calibration node
Quercus – genus	56	56	Quercus	Quercus Notholithocarpus
section Lobatae	47.87	47.87	Quercus_agrifolia Quercus_emoryi	Quercus_agrifolia Quercus_arizonica
section Cyclobalanopsis	48.32	48.32	Quercus_gilva Quercus_acuta	Quercus_gilva Quercus_rehderiana
section Quercus	45	45	Quercus_lobata Quercus_arizonica	Quercus_pontica Quercus_arizonica
section Ilex	47.8	37.8	Quercus_franchetii Quercus_rehderiana	
section Ilex – in part	35.5	33.4	Quercus_rehderiana Quercus_semecarpifolia	
section Cerris – in part	34	30	Quercus_chenii Quercus_acutissima	Quercus_franchetii Quercus_cerris
section Cerris – European clade	23	20.5	Quercus_crenata Quercus_cerris	

887

888 Supplement

889 **Fig. S1.** All-tips tree split by page (separate PDF)

890 **Fig. S2a.** Fossil calibration points: crown calibrations

891 **Fig. S2b.** Fossil calibration points: stem calibrations

892 **Fig. S3a.** Crown calibrations, global sampling estimate (60%)

893 **Fig. S3b.** Stem calibrations with rates, assuming clade-specific sampling proportions.

894 **Fig. S3c.** Stem calibrations, global sampling estimate (60%)

895 **Fig. S4.** Quartet similarity between individual loci and the full, all-tips tree, mapped to chromosomes

896 **Fig. S5.** Splines by chromosomes - quartets

897 **Fig. S6.** Splines by chromosomes - phyparts

898 **Fig. S7.** Phypart components

899

900 **Table S1.** Sampling table (separate XLSX)

901 **Table S2.** Citations for fossil calibrations.

902 **Table S3.** Taxonomic disparity index (TDI) for all unique species

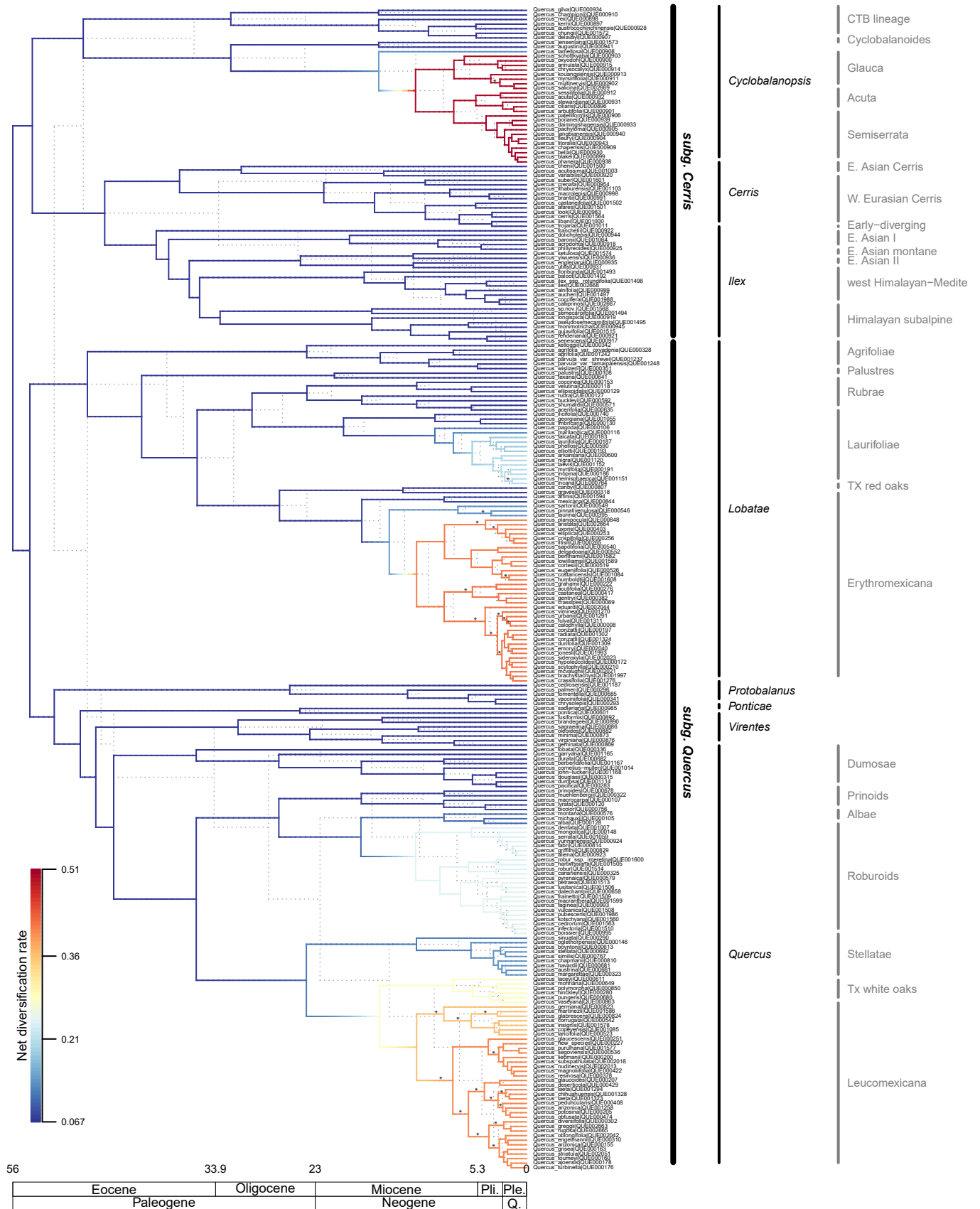
903 **Table S4.** \square IC values for alternative calibrations

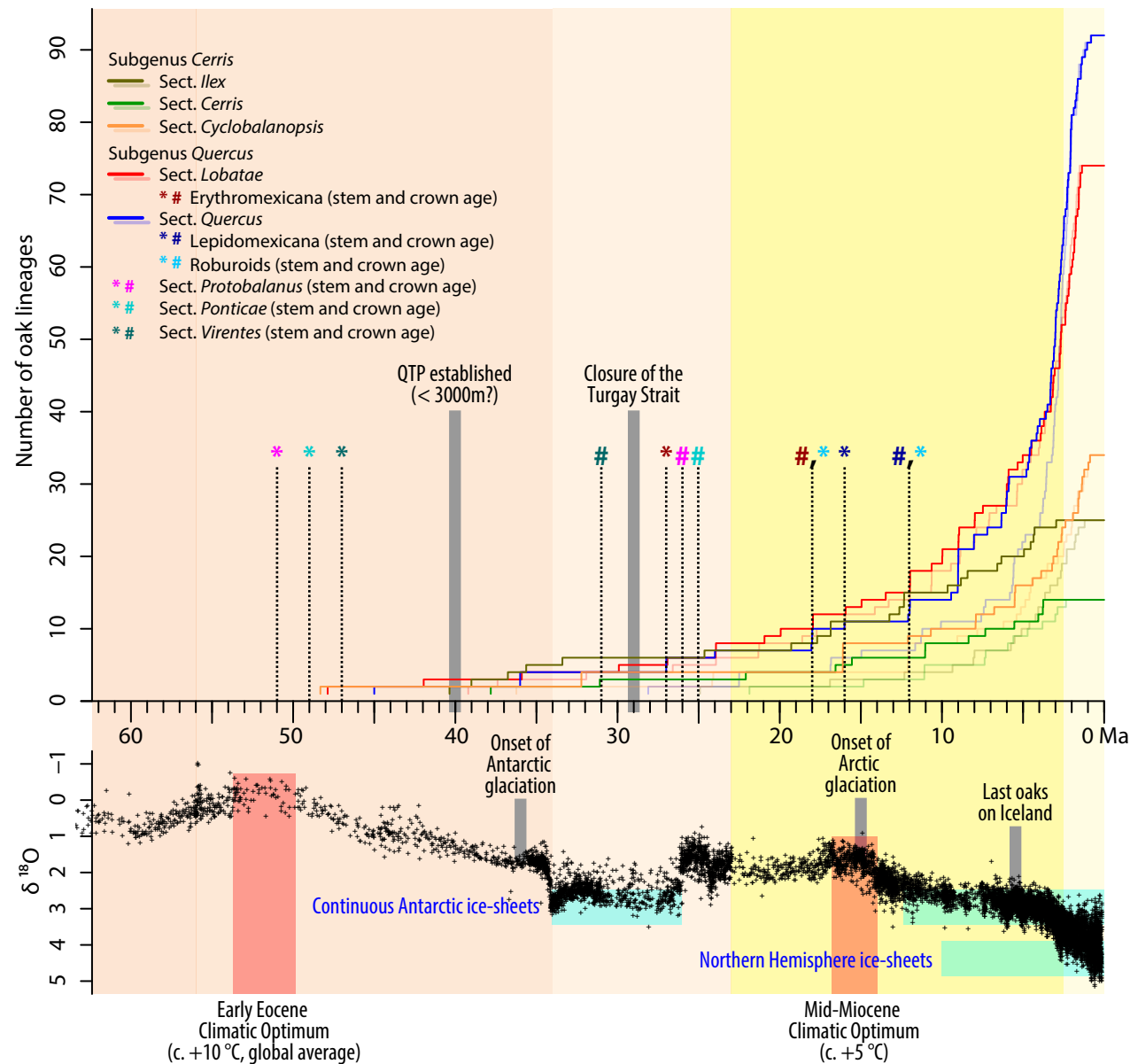
904 **Table S5.** Phypart components and clade ages

905

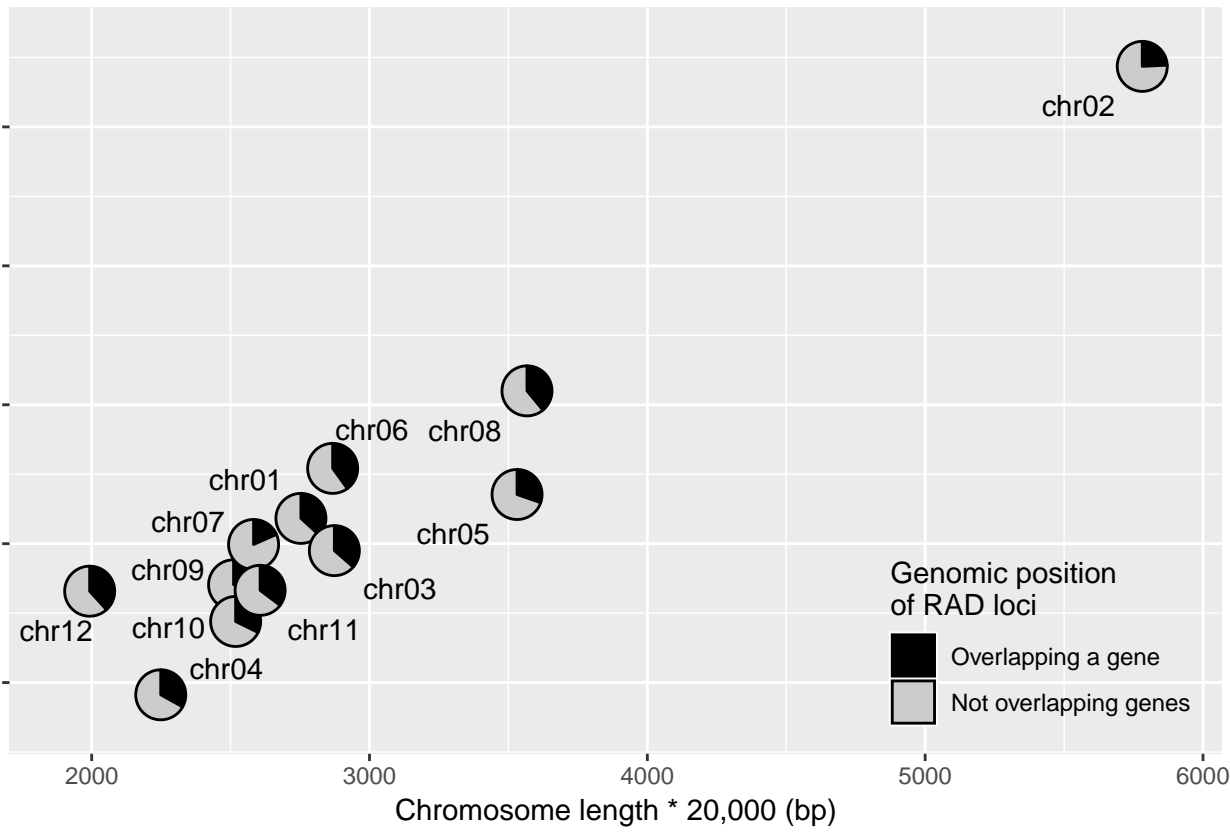
906 **Methods S1.** Analysis details.

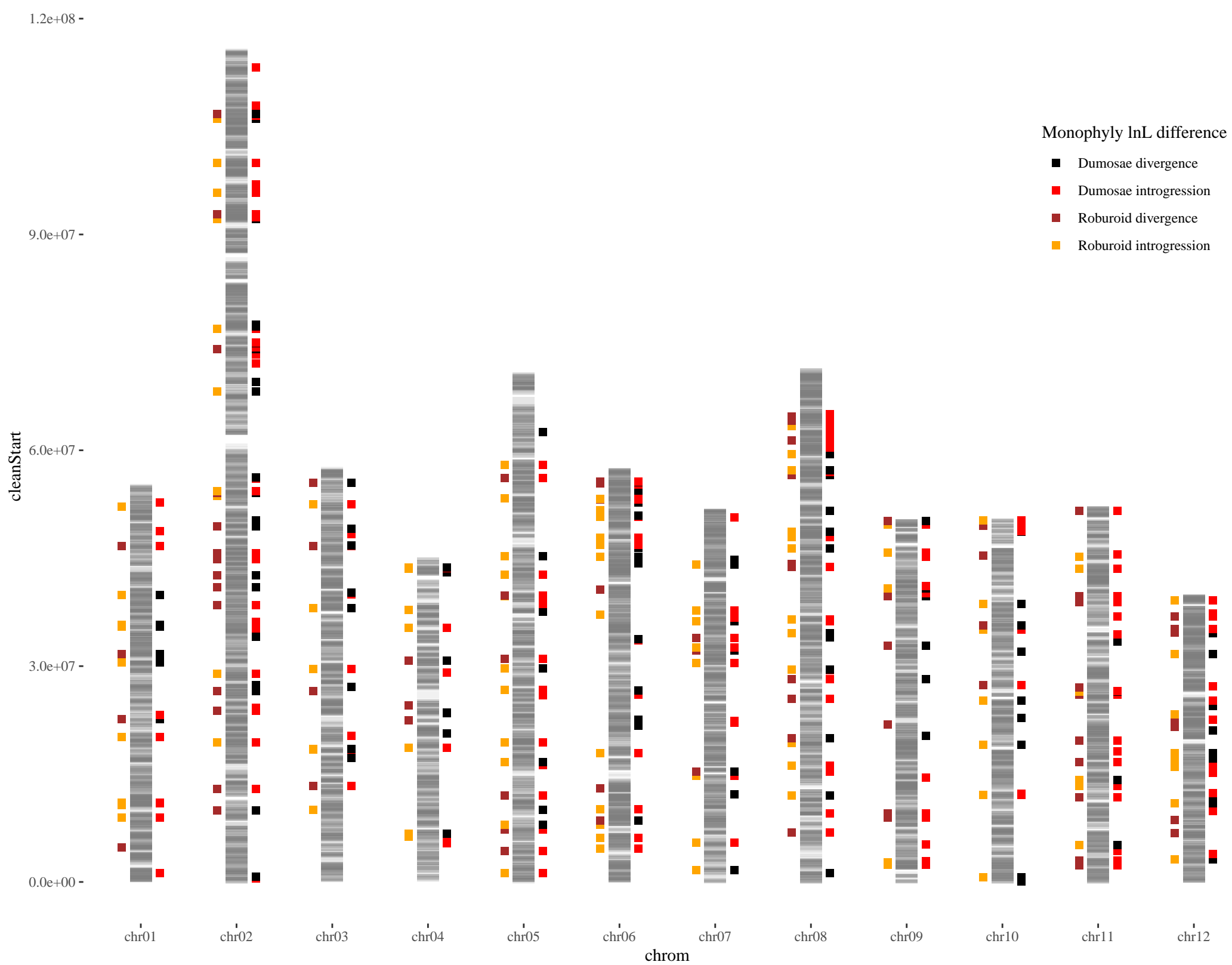
907



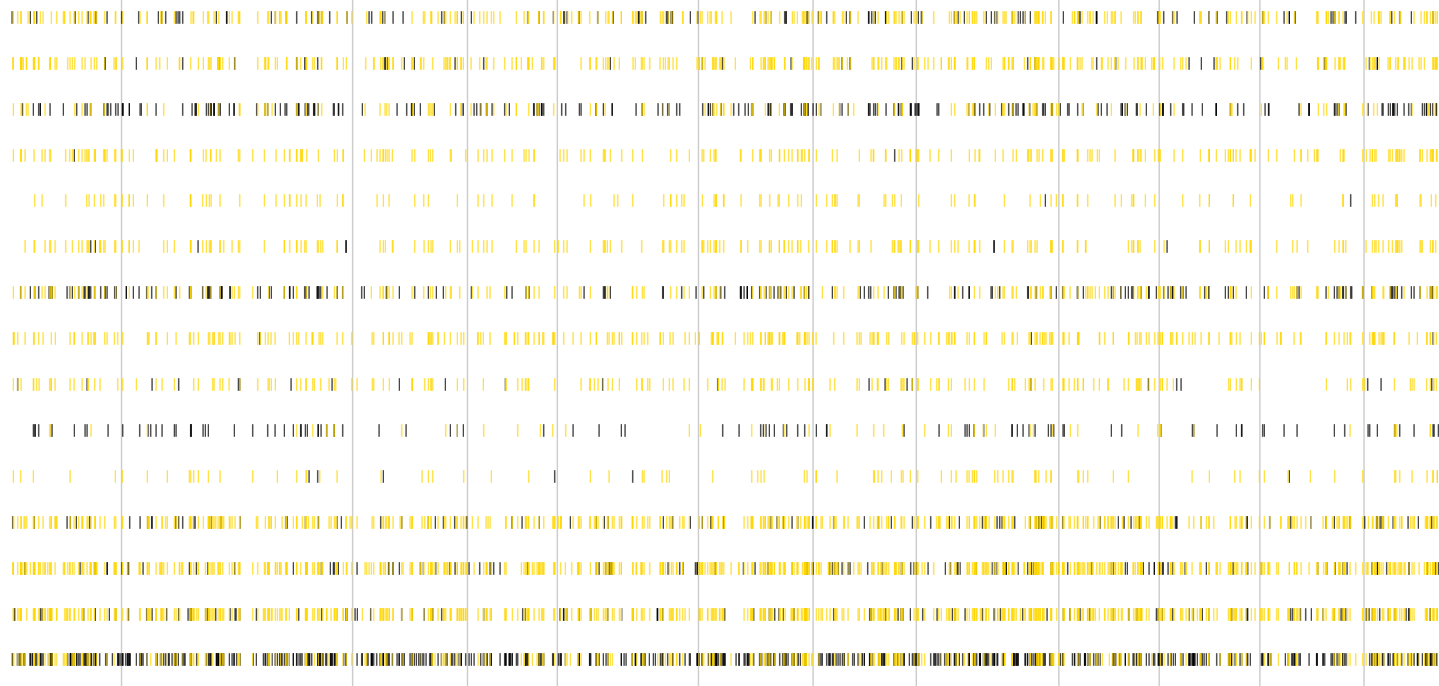


Number of RAD loci mapped to chromosome





15. sect. Cyclobalanopsis -
 14. sect. Ilex -
 13. sect. Cerris -
 12. Mexican red oaks -
 11. Rubrae -
 10. Laurifoliae -
 09. sect. Lobatae -
 08. Mexican white oaks -
 07. Roburoid white oaks -
 06. sect. Virentes -
 05. sect. Protobalanus -
 04. sect. Quercus -
 03. subg. Cerris -
 02. subg. Quercus -
 01. Quercus -



Node support

┆ concord

┆ conflict

