MOLECULAR ECOLOGY

Supplemental Information for:

Incorporating interspecific interactions into phylogeographic models: A case study with Californian oaks

Joaquín Ortego and L. Lacey Knowles

Contents:

SUPPLEMENTAL METHODS

METHODS S1 Genomic library preparation

We used a mixer mill to grind ~50 mg of frozen leaf tissue in tubes with a tungsten bead and performed DNA extraction and purification with NucleoSpin Plant II kits (Macherey-Nagel, Düren, Germany). We processed genomic DNA into four genomic libraries (72 individuals/library) using the double-digestion restriction-site associated DNA sequencing procedure (ddRADseq) described in Peterson et al., (2012). In brief, we digested DNA with the restriction enzymes *Mse*I and *Eco*RI (New England Biolabs, Ipswich, MA, USA) and ligated Illumina adaptors including unique 7-bp barcodes to the digested fragments of each individual. We pooled ligation products and size-selected them between 350-450 bp with a Pippin Prep machine (Sage Science, Beverly, MA, USA). We amplified the fragments by PCR with 12 cycles using the iProof™ High-Fidelity DNA Polymerase (BIO-RAD, Hercules, CA, USA) and sequenced the libraries in single-read 150-bp lanes on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (SickKids, Toronto, ON, Canada).

METHODS S2 Processing of genomic data

We used the different programs distributed as part of the STACKS v. 1.35 pipeline (*process_radtags*, *ustacks*, *cstacks*, *sstacks*, and *populations*) to assemble our sequences into *de novo* loci and call genotypes (Hohenlohe et al., 2010; Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). We demultiplexed and filtered reads for overall quality using the program *process_radtags*, retaining reads with a Phred score > 10 (using a sliding window of 15%), no adaptor contamination, and that had an unambiguous barcode and restriction cut site. We screened raw reads for quality with FASTQC v. 0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed all sequences to 129-bp using SEQTK (Heng Li, https://github.com/lh3/seqtk) in order to remove low-quality reads near the 3´ ends. We assembled filtered reads *de novo* into putative loci with the program *ustacks*. We set the minimum stack depth (*m*) to three and allowed a maximum distance of two nucleotide mismatches (*M*) to group reads into a "stack". We used the "removal" (*r*) and "deleveraging" (*d*) algorithms to eliminate highly repetitive stacks and resolve over-merged loci, respectively. We identified single nucleotide polymorphisms (SNPs) at each locus and called genotypes using a multinomial-based likelihood model that accounts for sequencing errors, with the upper bound of the error rate (ε) set to 0.2 (Hohenlohe et al., 2010; Catchen et al., 2011; Catchen et al., 2013). Then, we built a catalogue of loci using the *cstacks* program, with loci recognized as homologous across individuals if the number of nucleotide mismatches between consensus sequences (*n*) was ≤2. Finally, we matched each individual data against this catalogue using the program *sstacks* and exported output files in different formats for subsequent analyses using the program *populations*. For all downstream analyses, we exported only the first SNP per RAD locus and retained loci with a minimum stack depth ≥ 5 (*m* = 5), a minimum minor allele frequency (MAF) ≥ 0.01 (*min_maf* = 0.01) and that were represented in all populations (*p* = 8 for *Q. berberidifolia* and *p* = 10 for *Q. chrysolepis*) and the 80% of the individuals within each population (*r* = 0.8).

TABLE S1 Geographical location and genetic statistics (P, H_O, H_E and π) for the studied populations of California scrub oak (*Quercus berberidifolia*) and canyon live oak (*Q. chrysolepis*).

						All positions				Variant positions			
Code	Species	Locality	*n*	Latitude	Longitude	*P*	H_O	H_E	π	*P*	H_O	H_E	π
CLE	*Q. berberidifolia*	Clearlake	8	39.024420	-122.466510	0.9995	0.0007	0.0007	0.0007	0.9408	0.0901	0.0874	0.0933
LIC	*Q. berberidifolia*	Lick Observatory	8	37.329710	-121.490370	0.9995	0.0007	0.0007	0.0008	0.9393	0.0868	0.0920	0.0984
FIG	*Q. berberidifolia*	Figueroa Mountain	8	34.719980	-119.957033	0.9995	0.0007	0.0008	0.0008	0.9357	0.0926	0.0972	0.1038
THR	*Q. berberidifolia*	Three Points	8	34.752000	-118.715050	0.9995	0.0007	0.0007	0.0008	0.9368	0.0917	0.0960	0.1026
GAB	*Q. berberidifolia*	San Gabriel Mountains	8	34.228150	-117.670360	0.9995	0.0007	0.0008	0.0008	0.9326	0.0965	0.1010	0.1079
ELS	*Q. berberidifolia*	Lake Elsinore	8	33.648550	-117.410180	0.9995	0.0008	0.0008	0.0009	0.9296	0.1049	0.1064	0.1136
HEM	*Q. berberidifolia*	Hemet	7	33.705890	-116.754840	0.9994	0.0009	0.0008	0.0009	0.9275	0.1109	0.1073	0.1158
GUA	*Q. berberidifolia*	Guatay Mountain	8	32.854420	-116.576190	0.9995	0.0007	0.0008	0.0009	0.9304	0.0957	0.1049	0.1121
SHA	*Q. chrysolepis*	Shasta	8	40.604770	-122.502300	0.9992	0.0011	0.0012	0.0013	0.8965	0.1477	0.1513	0.1616
TAH	*Q. chrysolepis*	Tahoe	8	39.281970	-120.988660	0.9992	0.0011	0.0011	0.0012	0.9019	0.1451	0.1428	0.1525
SON	*Q. chrysolepis*	Sonoma	8	38.678220	-123.136930	0.9993	0.0011	0.0010	0.0011	0.9062	0.1374	0.1325	0.1415
YOS	*Q. chrysolepis*	Yosemite	8	37.715669	-119.677205	0.9993	0.0010	0.0011	0.0011	0.9069	0.1322	0.1373	0.1466
KIN	*Q. chrysolepis*	Kings Canyon	8	36.741340	-119.031300	0.9992	0.0011	0.0011	0.0012	0.9021	0.1367	0.1428	0.1526
HAS	*Q. chrysolepis*	Hastings	8	36.358960	-121.551000	0.9992	0.0011	0.0011	0.0012	0.9011	0.1398	0.1446	0.1544
FIG	*Q. chrysolepis*	Figueroa Mountain	8	34.724470	-119.950080	0.9992	0.0011	0.0011	0.0012	0.9009	0.1404	0.1432	0.1530
GAB	*Q. chrysolepis*	San Gabriel Mountains	8	34.356780	-117.743150	0.9992	0.0011	0.0011	0.0012	0.9021	0.1380	0.1445	0.1543
BER	*Q. chrysolepis*	San Bernardino Mountains	8	34.130280	-116.982500	0.9993	0.0010	0.0011	0.0012	0.9047	0.1346	0.1403	0.1499
LAG	*Q. chrysolepis*	Laguna Mountain	8	32.849540	-116.485350	0.9992	0.0011	0.0011	0.0012	0.8998	0.1392	0.1453	0.1553

n, number of analyzed individuals; Average values across loci are presented for major allele frequency (*P*), nucleotide diversity (π), and observed (H_O) and expected (H_E) heterozygosity. Genetic statistics were calculated in STACKS for all positions (polymorphic and non-polymorphic) and only variant (polymorphic) positions considering loci that were represented in all populations (*p* = 8 for *Q. berberidifolia* and *p* = 10 for *Q. chrysolepis*) and the 80% of individuals within populations (*r* = 0.8).

TABLE S2 Environmental niche modeling (ENM) of Californian oaks (*Quercus* sp.) using species-specific model parameter tuning. Table shows the parameters of the best model selected using the Akaike's information criterion corrected for small sample sizes (AICc) and the variables retained sorted from higher to lower values of permutation importance. Variables in bold are those that cumulatively contributed > 50% to the model based on the permutation importance statistic.

Species	*n*	*FC*	*RM*	MTSS	Environmental variables
Q. berberidifolia	1,173	HPLTQ	3.0	0.061	**BIO19**,**BIO14**,BIO4,BIO11,BIO18,BIO15,BIO8,slope,BIO10,BIO3,BIO2,BIO9
Q. durata	660	T	4.0	0.059	**BIO15**,**BIO19**,BIO5,slope,BIO4,BIO18,BIO3,BIO9,BIO14,BIO1,BIO6
Q. dumosa	160	HPLTQ	7.5	0.026	**BIO4**,BIO17,BIO9,BIO3,BIO11,BIO19,BIO5,BIO15,BIO8,slope,BIO2,BIO18
Q. pacifica	158	T	15.0	0.043	**BIO2**,BIO5,BIO18,BIO1,BIO4,BIO15,BIO6,slope,BIO19
Q. cornelius-mulleri	224	T	5.0	0.071	**BIO14**,**BIO4**,BIO15,BIO18,BIO8,BIO12,BIO3,BIO11,BIO9,BIO2,BIO10,slope
Q. john-tuckeri	252	HPLTQ	2.5	0.119	**BIO18**,**BIO14**,**BIO6**,BIO4,BIO19,BIO15,BIO5,slope,BIO8,BIO3,BIO9,BIO2,BIO1
Q. douglasii	1,567	HPLTQ	5.0	0.097	**BIO18**,**BIO19**,BIO14,BIO9,BIO2,BIO4,BIO3,slope,BIO11,BIO8,BIO15,BIO5
Q. lobata	1,226	HPLTQ	2.0	0.095	**BIO18**,BIO19,BIO6,BIO8,BIO5,BIO3,BIO4,BIO1,BIO15,slope,BIO9,BIO2
Q. garryana	1,830	T	2.0	0.132	**BIO19**,**BIO17**,**BIO6**,**BIO18**,BIO1,BIO8,BIO9,BIO15,BIO4,BIO5,BIO3,BIO2,slope
Q. engelmannii	449	H	3.5	0.031	**BIO10**,**BIO19**,BIO4,BIO18,BIO14,BIO3,BIO11,BIO15,BIO2,slope
Q. turbinella	343	T	1.5	0.239	**BIO19**,**BIO18**,**BIO11**,BIO8,BIO15,slope,BIO7,BIO17,BIO2,BIO9,BIO10,BIO3
Q. sadleriana	246	T	5.5	0.026	**BIO3**,**BIO8**,BIO14,BIO4,BIO9,BIO19,BIO15,BIO18
Q. chrysolepis	2,541	T	4.0	0.177	**BIO14**,**BIO19**,BIO15,slope,BIO8,BIO2,BIO11,BIO3,BIO5,BIO18,BIO7,BIO9
Q. tomentella	123	T	8.5	0.024	**BIO5**,BIO18,BIO3,slope,BIO2,BIO15,BIO7,BIO16,slope
Q. vaccinifolia	261	HPLTQ	3.0	0.042	**BIO18**,**BIO14**,BIO19,BIO10,BIO15,BIO3,BIO8,BIO6,BIO2,BIO9
Q. palmeri	132	T	2.0	0.135	**BIO19**,BIO14,BIO3,BIO6,BIO9,BIO8,BIO15,BIO18,slope,BIO2
Q. agrifolia	1,967	T	5.0	0.056	**BIO19**,**BIO4**,BIO6,BIO14,BIO18,BIO8,BIO15,BIO1,BIO5,slope,BIO3,BIO2
Q. wislizeni	244	T	2.0	0.185	**BIO19**,BIO17,BIO15,BIO8,slope,BIO18,BIO11,BIO3,BIO4,BIO9,BIO2,BIO5
Q. kelloggii	2,301	T	5.0	0.203	**BIO19**,BIO14,BIO8,BIO10,BIO3,BIO15,BIO18,slope,BIO2,BIO7

n, number of occurrence records used for ENM after filtering duplicates; *FC*, feature class for variable transformation; *RM*, regularization multiplier; MTSS, Maximum training sensitivity plus specificity logistic threshold for species presence/absence.

FIGURE S1 Alternative spatiotemporally explicit demographic scenarios used for demographic and genetic simulations in California scrub oak (*Quercus berberidifolia*) and canyon live oak (*Q. chrysolepis*). Local carrying capacities (*K*, colored scale bar) change across the landscape and time periods (from the last glacial maximum to present). Local carrying capacities range from 0 (minimum) to 1 (maximum) and were scaled based on habitat suitabilities estimated from environmental niche models (ENMs) for the two focal taxa and considering different hypothetical interactions (neutral, negative or positive) with other oak species. For detailed model description, see Table 1.

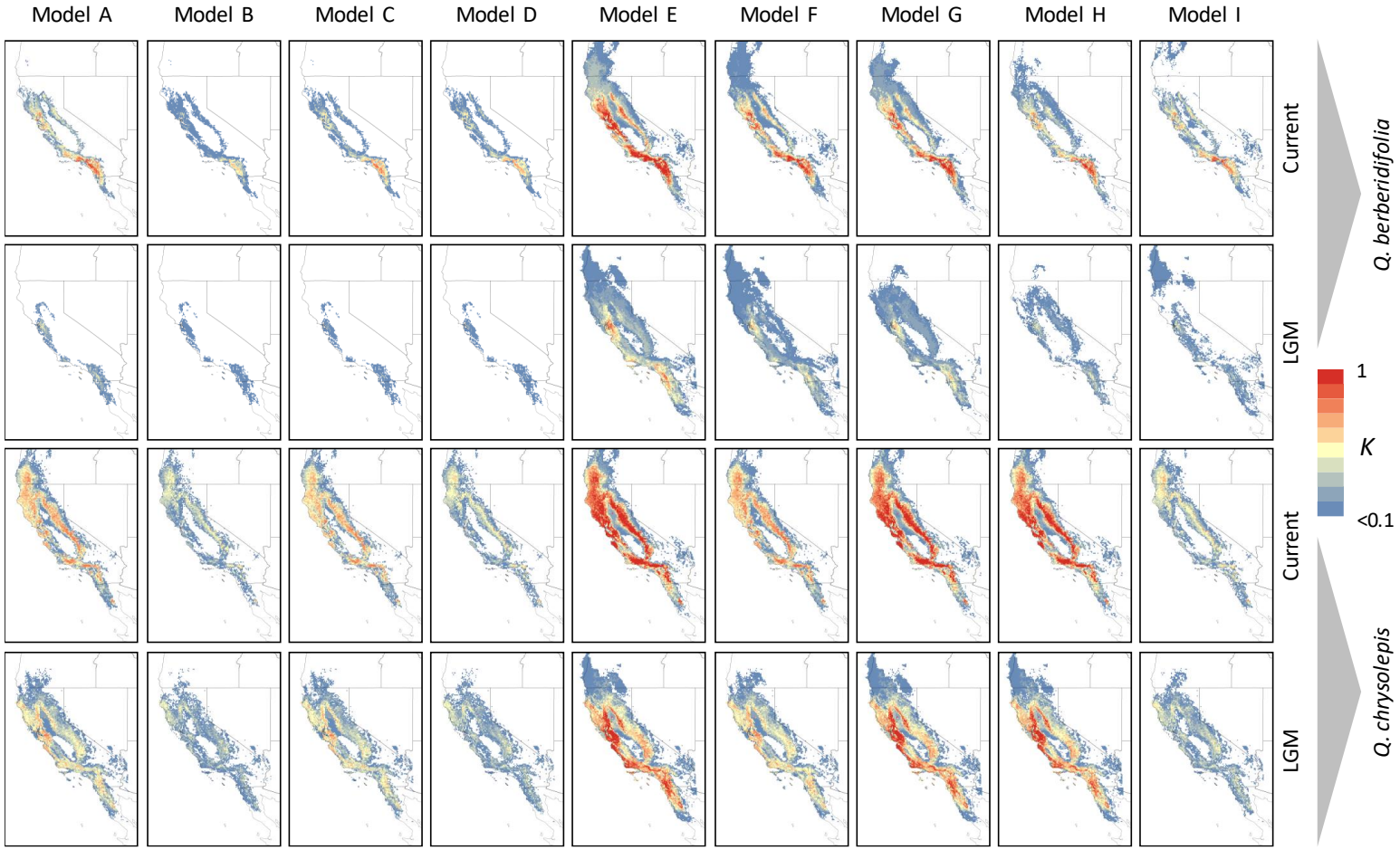


FIGURE S2 Current distribution of Californian oaks (*Quercus* sp.) inferred using environmental niche modeling (ENM). Red color indicates areas predicted to be occupied by the species according to the maximum training sensitivity plus specificity (MTSS) logistic threshold (Table S2). Black dots indicate occurrence records used for ENM. Star colors indicate the taxonomic section of each species (green: *Quercus*; golden: *Protobalanus*; red: *Lobatae*). All maps are available for download at a high resolution at Figshare (https://doi.org/10.6084/m9.figshare.12388781).

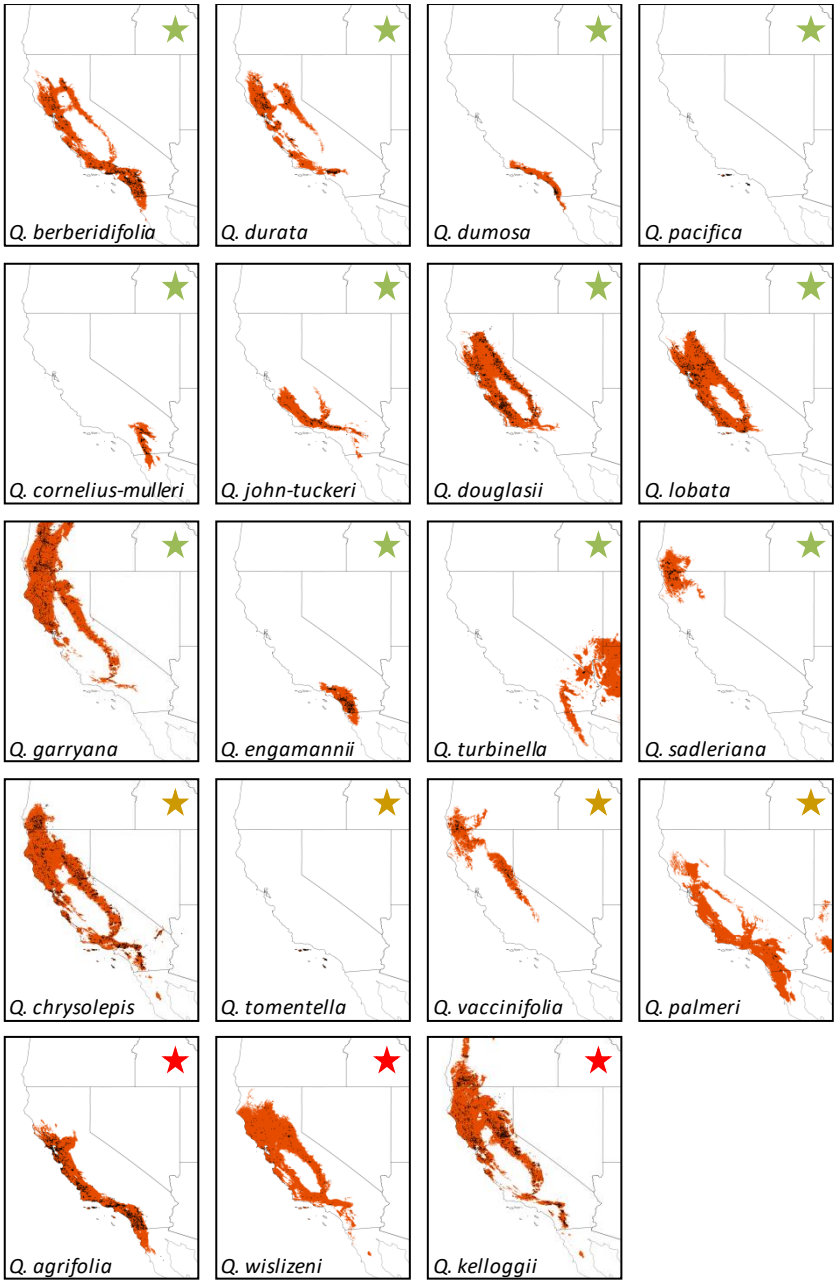


FIGURE S3 Last glacial maximum (LGM) distribution of Californian oaks (*Quercus* sp.) inferred using environmental niche models (ENM) and paleoclimate data simulated under the Community Climate System Model (CCSM4). Blue color indicates areas predicted to be occupied by the species according to the maximum training sensitivity plus specificity (MTSS) logistic threshold (Table S2). Star colors indicate the taxonomic section of each species (green: *Quercus*; golden: *Protobalanus*; red: *Lobatae*). All maps are available for download at a high resolution at Figshare (https://doi.org/10.6084/m9.figshare.12388781).

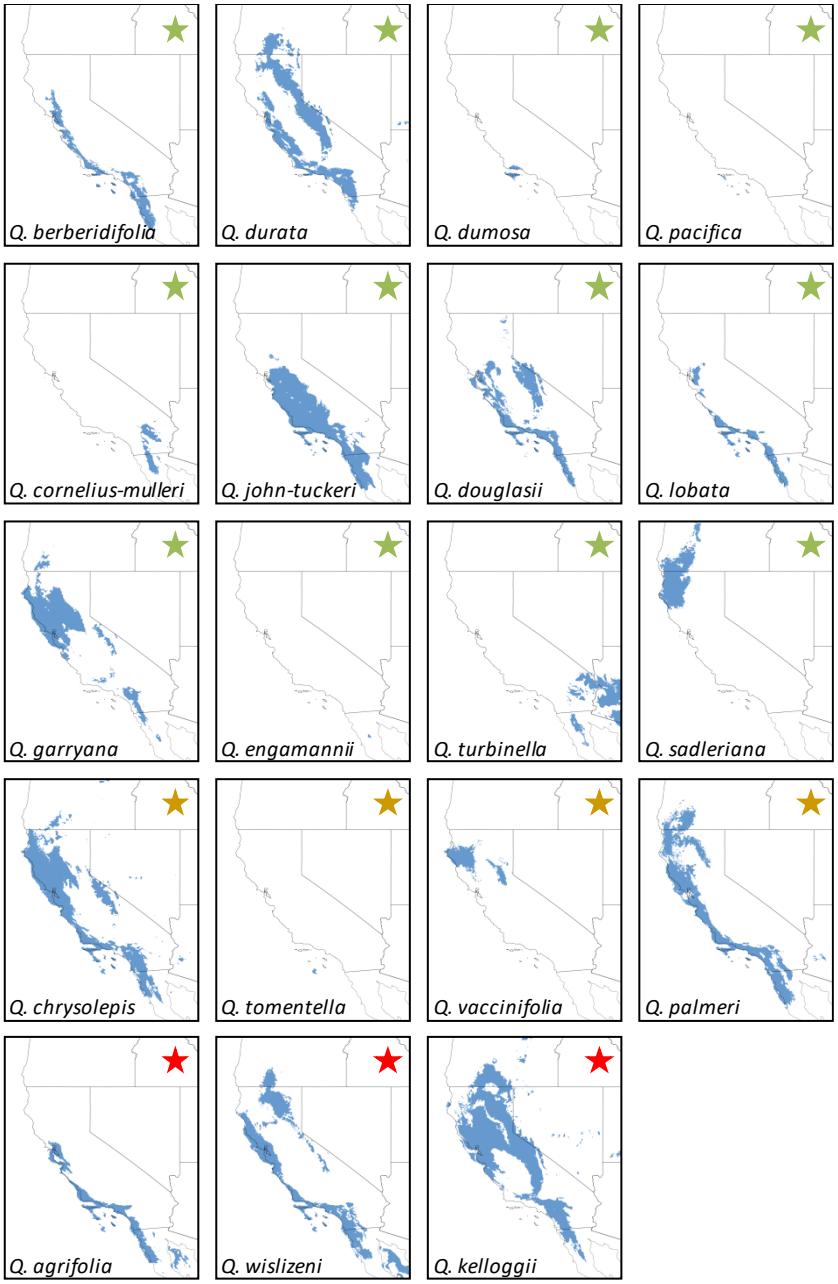
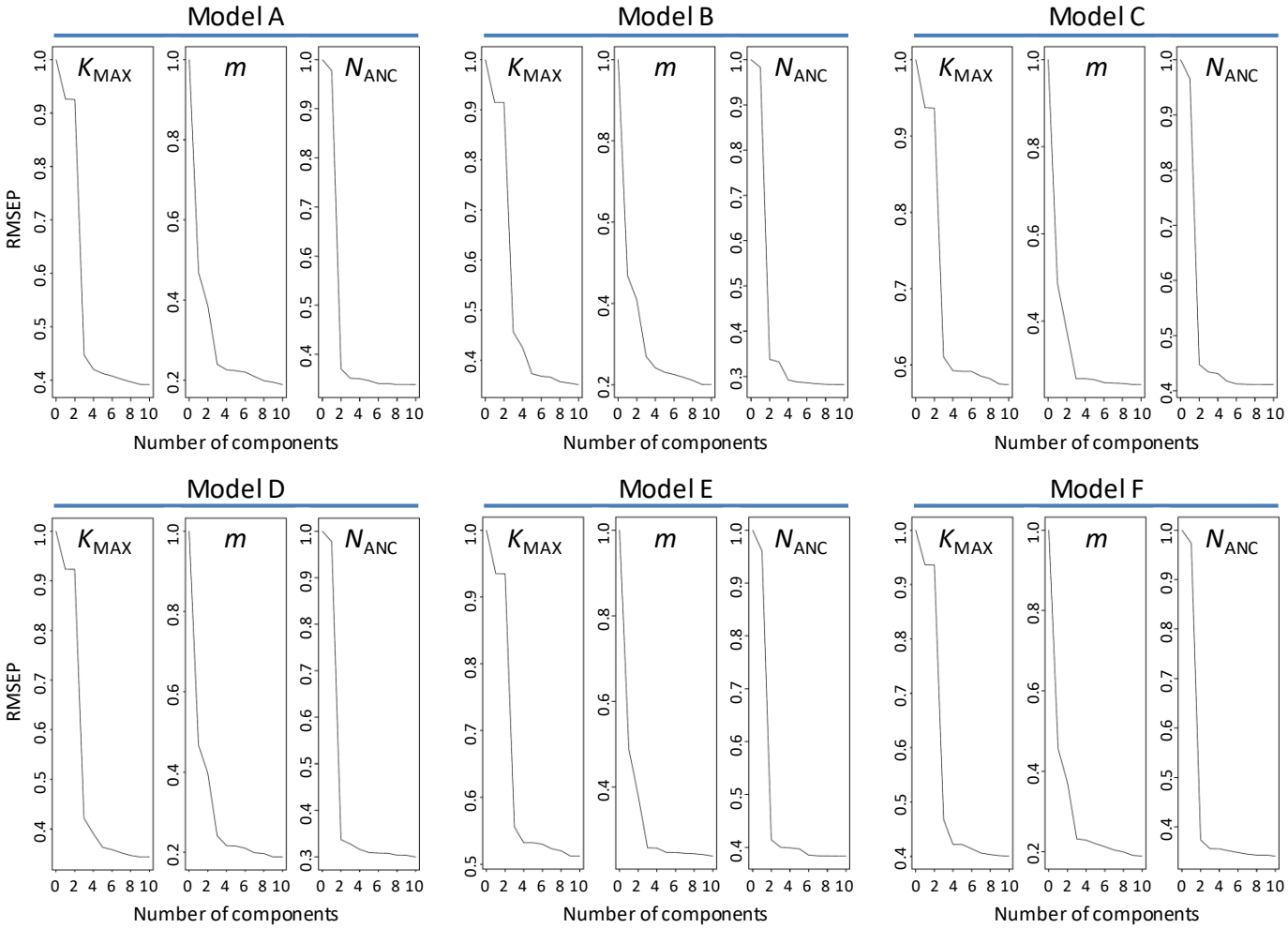


FIGURE S4 Root mean square error (RMSE) of parameter estimates against the number of partial least squares (PLS) components under nine demographic models for California scrub oak (*Quercus berberidifolia*).



continued…

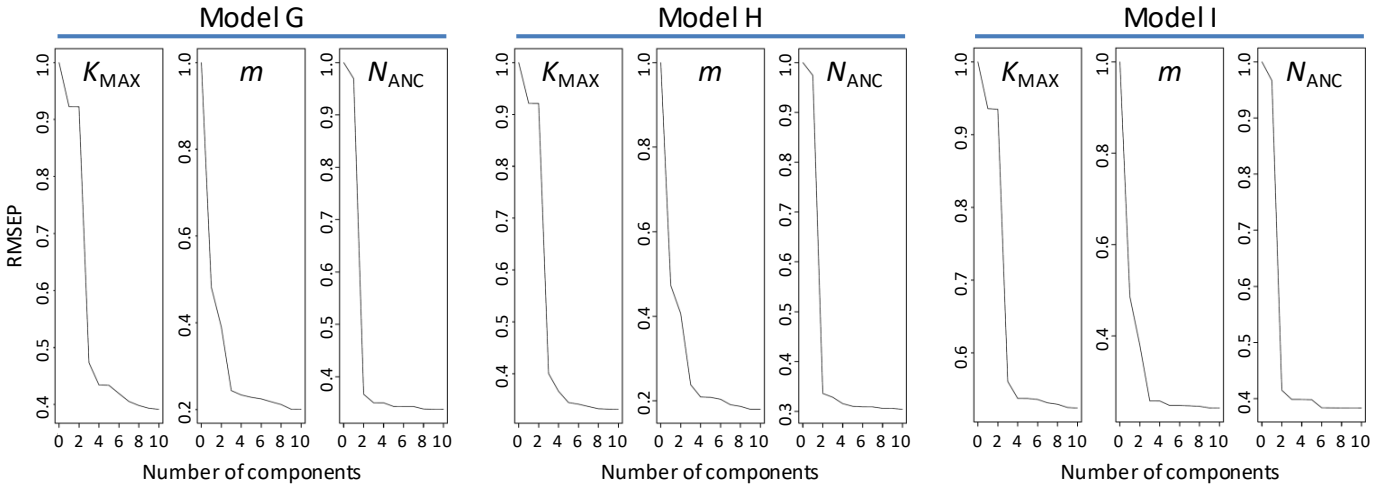
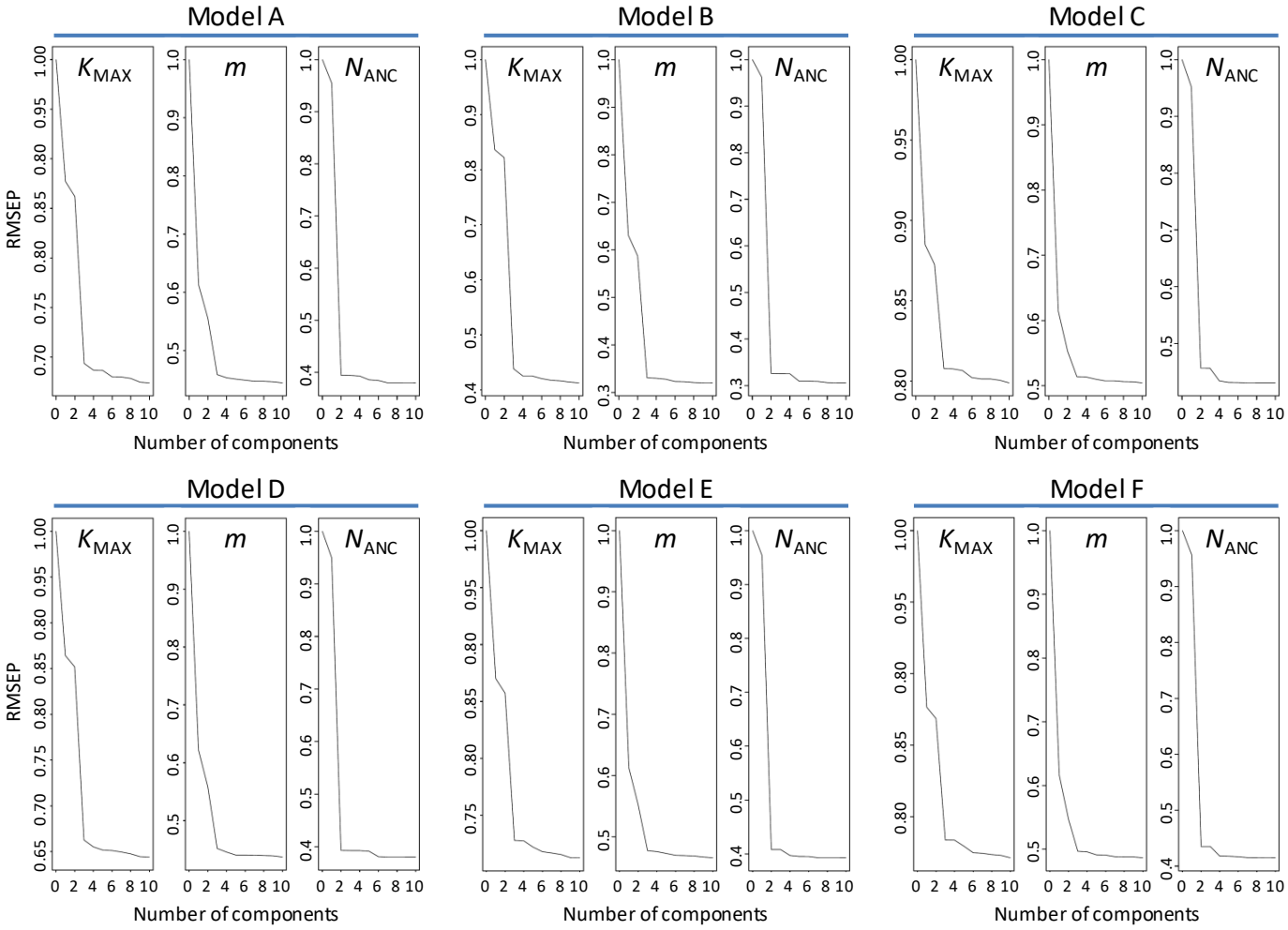

Model G
Model H
Model I
K_{MAX}
m
N_{ANC}
RMSEP
Number of components

FIGURE S5 Root mean square error (RMSE) of parameter estimates against the number of partial least squares (PLS) components under nine demographic models for Canyon live oak (*Quercus chrysolepis*).



continued…

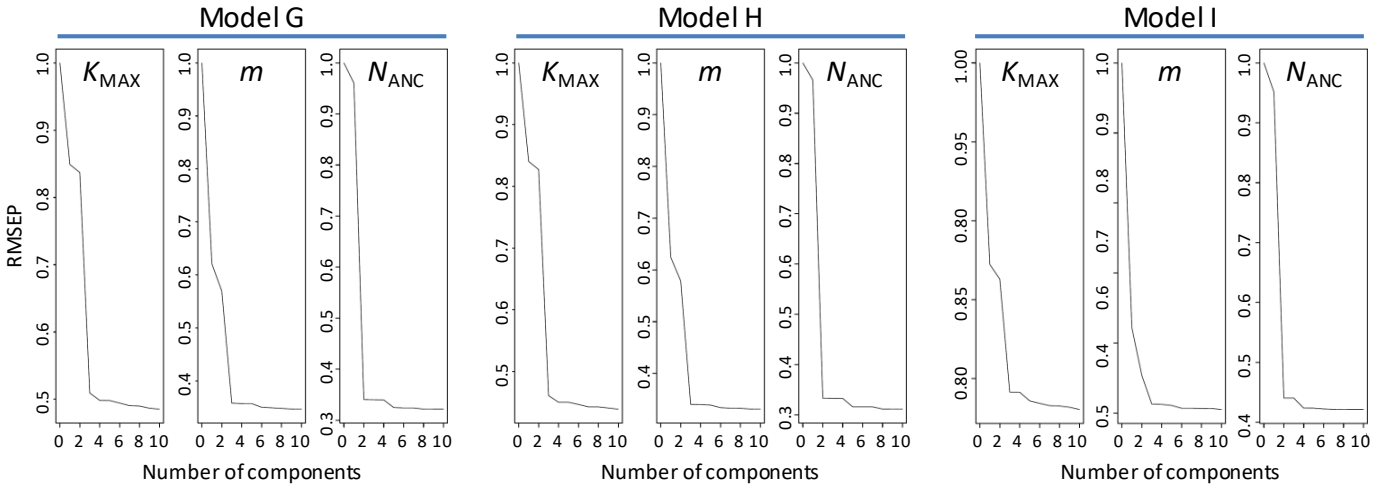

Model G
Model H
Model I
K_{MAX}
m
N_{ANC}
RMSEP
Number of components

FIGURE S6 Number of reads per individual before and after different quality filtering steps in STACKS. The cumulative stacked bars represent the total number of raw reads for each individual. Dark red color represents the reads that were discarded by *process_radtags* due to low quality, adapter contamination or ambiguous barcode. Light red color represents the reads that were discarded by *ustacks* after filtering out repetitive elements and reads that did not comply the criteria required to create a "stack". Green color represents the total number of retained reads used to identify homologous loci. Individuals are sorted by species and populations following the same order and codes presented in Table S1.

FIGURE S7 Results of Bayesian clustering analyses in STRUCTURE for (a) California scrub oak (*Quercus berberidifolia*) and (b) canyon live oak (*Q. chrysolepis*). Plots show mean (±SD) log probability of the data (Ln Pr(X|K)) over 10 runs of STRUCTURE (left *y*-axes, black dots and error bars) for each value of *K* and the magnitude of Δ*K* (right *y*-axes, open dots).

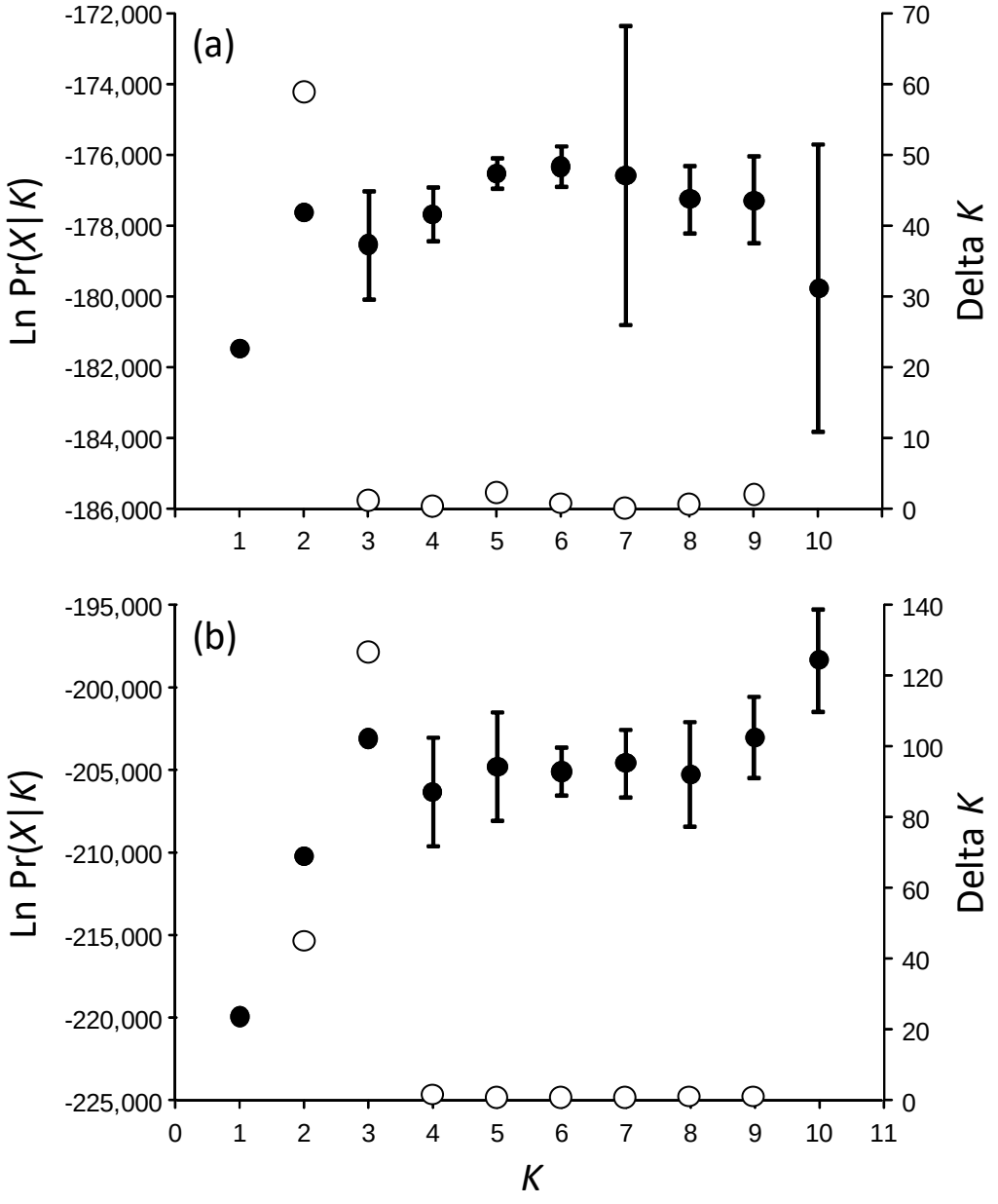


FIGURE S8 Principal component analyses (PCAs) of genetic variation for (a) California scrub oak (*Quercus berberidifolia*) and (b) canyon live oak (*Q. chrysolepis*). Colors indicate the main genetic cluster at which populations were assigned according to STRUCTURE analyses (see Figure 2). Population codes are described in Table S1.

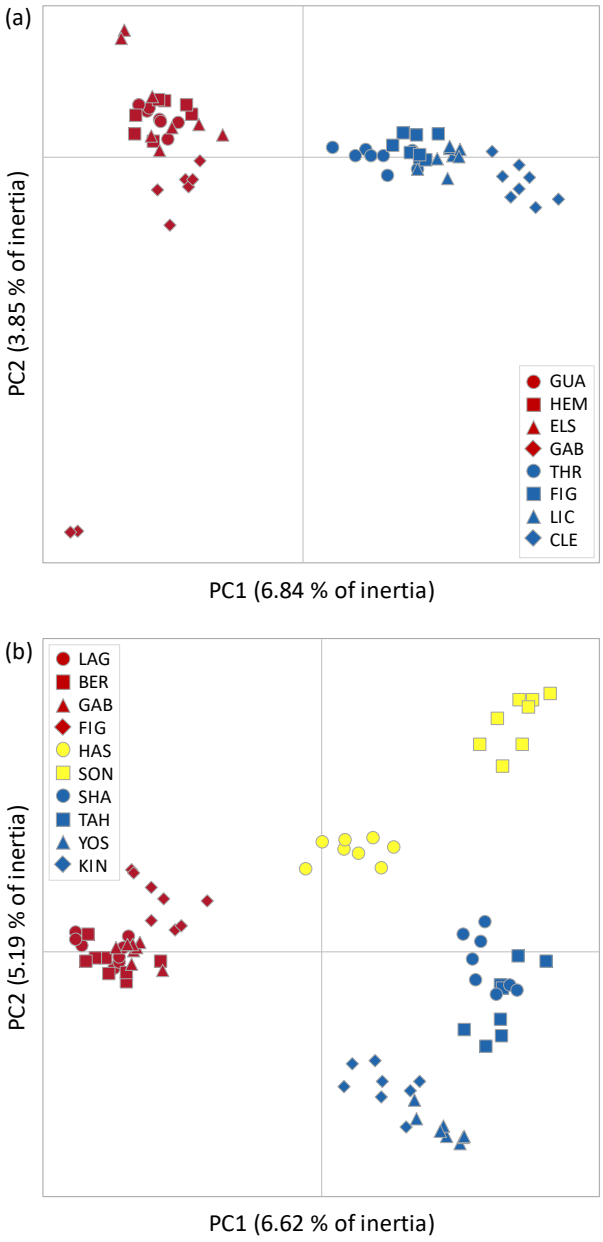
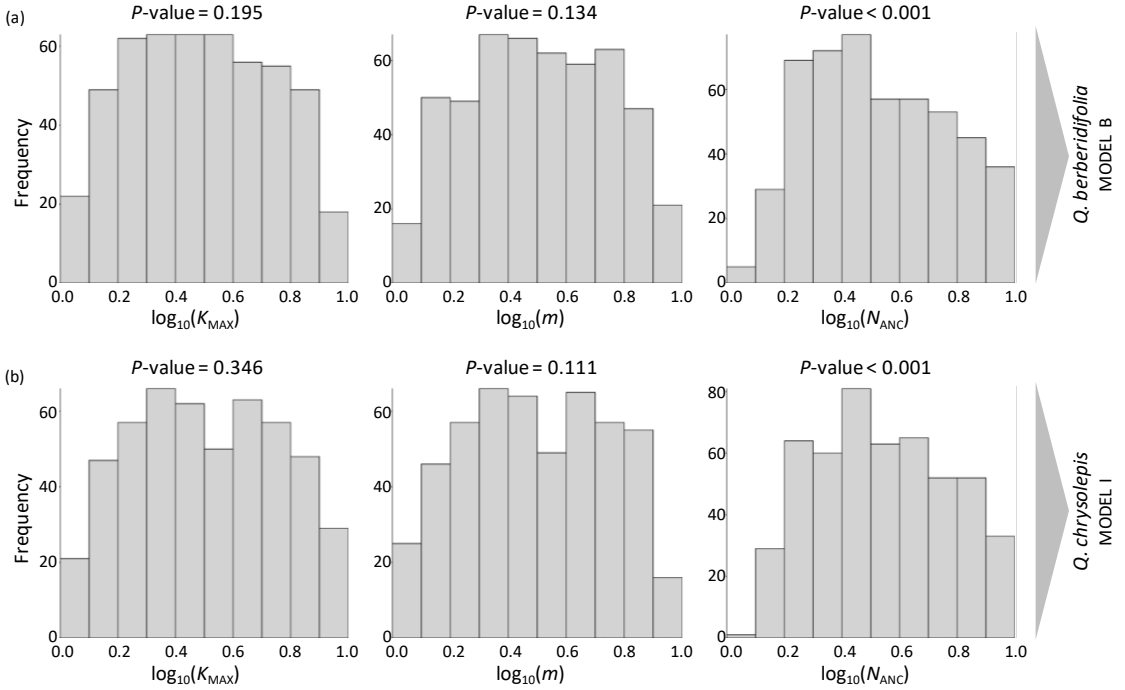


FIGURE S9 Distribution of posterior quantiles of true parameters values used for evaluating potential bias in parameter estimation for the most supported models for (a) California scrub oak (*Quercus berberidifolia*) (Model B) and (b) canyon live oak (*Q. chrysolepis*) (Model I). Analyses are based on 1,000 pseudo-observed datasets (PODs).



REFERENCES

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). STACKS: Building and genotyping loci *de novo* from short-read sequences. *G3-Genes Genomes Genetics, 1*(3), 171-182. doi:10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). STACKS: an analysis tool set for population genomics. *Molecular Ecology, 22*(11), 3124-3140. doi:10.1111/mec.12354

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics, 6*(2), e1000862. doi:10.1371/journal.pgen.1000862

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One, 7*(5), e37135. doi:10.1371/journal.pone.0037135