Genetic Distinctions among Oaks in the University of California, Davis Arboretum: Contributions to Oak Phylogeny

Terence M. Murphy and Daniel Potter
Department of Plant Biology and Department of Plant Sciences
University of California, Davis, CA, USA

Abstract

DNA samples of oaks from the collection at the University of California, Davis Arboretum have been isolated, and partial sequences determined for genes of ribulosebisphosphate carboxylase large subunit and for internal transcribed sequences of ribosomal RNA genes. A comparison of these sequences contributes to an elucidation of the phylogeny of this important plant genus. The agreement between the patterns with *rbcL* and ITS sequences strengthens considerably the conclusion that these molecular data can reveal the actual pattern of evolutionary relationships among the *Quercus* species.

Introduction

Owing to the interest of the late Professor John Tucker and his students, the Arboretum at the Davis campus of the University of California has a particularly large and diverse collection of oak trees. The phylogeny of the oaks has been a subject of interest and some uncertainty for many years. Within the last fifteen years, a few groups (Samuel et al., 1998; Manos et al., 1999; Oh and Manos, 2008) have applied the techniques of molecular phylogenetics to comparisons of oak genes, adding information to the understanding of oak evolution. The present paper extends this information by presenting the sequences of *rbcL*, a chloroplast gene that encodes the large subunit of the photosynthetic enzyme ribulosebisphosphate carboxylase, and the internal transcribed sequences together with 5.8S ribosomal DNA, from UC Davis Arboretum oaks.

Materials and Methods

Samples

Samples of leaves were taken from the University Arboretum at the Davis campus of the University of California. The search facility of the web site of the Arboretum (http://arboretum.ucdavis.edu/collections_search.aspx) lists 163 accessions, including 74 species and subspecies and 22 hybrids. The plants sampled for this report are listed in Table 1. DNA extracts were made from all samples, although not all extracts gave satisfactory DNA sequences.

Table 1. Oak species sampled for DNA. For "Arboretum" information, refer to the web page: http://arboretum.ucdavis.edu/collections_search.aspx. "DCPD" refers to the Davis Center for Plant Diversity, which holds herbarium specimens; for accession information: http://museums.ucdavis.edu/GIS_dataoption_mdb.aspx.

		Arboretum Accession	DCPD Accession	*rbcL* GenBank Accession	ITS GenBank Accession
Quercus ×acutidens		A67.0978	DAV190869	KF683136	
Quercus agrifolia	coast live oak	A64.0713	DAV190999	KF683137	
Quercus arizonica	Arizona white oak	A92.0013	DAV190964		KM200955
Quercus berberidifolia	California scrub oak	A64.1271	DAV190976	KF683138	KM200956
Quercus canariensis	Algerian oak; Mirbeck's oak	A64.1303		KF683140	KM200957
Quercus candicans		A90.0489	DAV190920	KF683139	KM200958
Quercus castaneifolia	Persian oak	A94.0497	DAV190712	KF683141	KM200959
Quercus chrysolepis	canyon live oak	A65.0013	DAV190899	KF683142	KM200961
Quercus crassipes		A68.0361	DAV190892	KF683143	KM200960
Quercus ×deamii		A69.0642	DAV190884	KF683144	KM200962
Quercus diversifolia		A68.0353	DAV190901		
Quercus douglasii	blue oak	A64.0406	DAV190724	KF683145	
Quercus durata	leather oak	A58.0104	DAV190962		
Quercus engelmanii		A65.0011	DAV25499	KF683146	
Quercus faginea	Portuguese oak	A71.0155	DAV190933	KF683147	
Quercus gambelii	Gambel oak	A63.0004	DAV190888	KF683148	
Quercus gravesii		A86.0445	DAV190722	KF683150	KM200965
Quercus greggii		A68.0359	DAV190903	KF683151	
Quercus grisea	gray oak	A63.0002	DAV190736	KF683152	KM200966
Quercus hartwegi		A68.0350		KF683153	
Quercus ×hispanica		A98.0112	DAV190874	KF683154	KM200967
Quercus ibirica		A64.1216		KF683156	
Quercus infectoria	ssp. veneris	A64.1284	DAV190877	KF683157	KM200968
Quercus lobata	valley oak	A33.9041	DAV190958		KM200963
Quercus ×macdonaldii		A74.0008	DAV190973		
Quercus margaretta		A64.0004	DAV190674	KF683158	
Quercus Mexicana		A68.0349	DAV190893	KF683159	KM200970
Quercus mohriana	Mohr's oak	A64.0006	DAV190870	KF683160	KM200971
Quercus muehlenbergii	yellow chestnut oak; chinkapin oak	A63.0009	DAV190714	KF683161	KM200972
Quercus oblongifolia	Mexican blue oak	A64.0075	DAV190726	KF683162	KM200973
Quercus palmeri		A64.1173	DAV190879		
Quercus pilicaulis		A91.0741	DAV190963		
Quercus prinoides	dwarf chinkapin oak	A66.0172	DAV190713	KF683163	KM200974
Quercus pungens		A63.0007	DAV190354	KF683164	KM200975
Quercus rugosa		A65.0838	DAV190001	KF683165	
Quercus serrata	Syn.*: Q. glandulifera*	A64.1306	DAV190919	KF683149	KM200964
Quercus sinuata		A64.0062		KF683166	
Quercus sp., Iran		A96.0684		KF683155	KM200969
Quercus turbinella		A67.1042	DAV190876		
Quercus vaseyana	sandpaper oak	A63.0008	DAV190734	KF683167	KM200976
Quercus wislizeni	Syn.: *Quercus pungens* var. *vaseyana* interior live oak	A36.0031	DAV190956	KF683168	KM200977

Generally, young leaves were sampled in the spring and early summer of 2012 and the spring of 2013. Leaves were collected and frozen at -80°C until DNA extraction.

Leaf samples of 0.05 to 0.1 g were frozen to brittleness in liquid N_2 and then ground in a 1.5-ml plastic centrifuge tube with a plastic pestle turned by a hand drill. CTAB extraction buffer, 300 µl, was added, and the grinding repeated until the slurry was reasonably uniform. (CTAB extraction buffer contains 2% cetyltrimethyammonium bromide, 1.4 M NaCl, 0.1 M trishydroxyaminomethane (Tris)-Cl, and 20 mM ethylenedinitrilotetraacetic acid adjusted to pH 8.) The slurry was extracted with 300 µl of chloroform and centrifuged, and the upper, aqueous phase (approximately 250 µl) was mixed with an equal amount of isopropanol. The mixture was centrifuged, and the pellet was washed with 70% ethanol and dissolved in 50-100 µl of water. Most samples were further purified by adsorption and elution from glass (e.g. Promega Wizard^R, see below).

Polymerase Chain Reactions

Polymerase chain reactions (PCRs) amplified segments of DNA in the sample extracts. Primers were designed to select three segments, a portion of the chloroplast gene for the large subunit of ribulosebisphosphate carboxylase/oxygenase (*rbcL*), the internal transcribed spacers adjacent to, together with, the 5.8S ribosomal gene (ITS), and a 250-base section near the chloroplast *trnF* and *trnL* genes (Figure 1; Table 2).

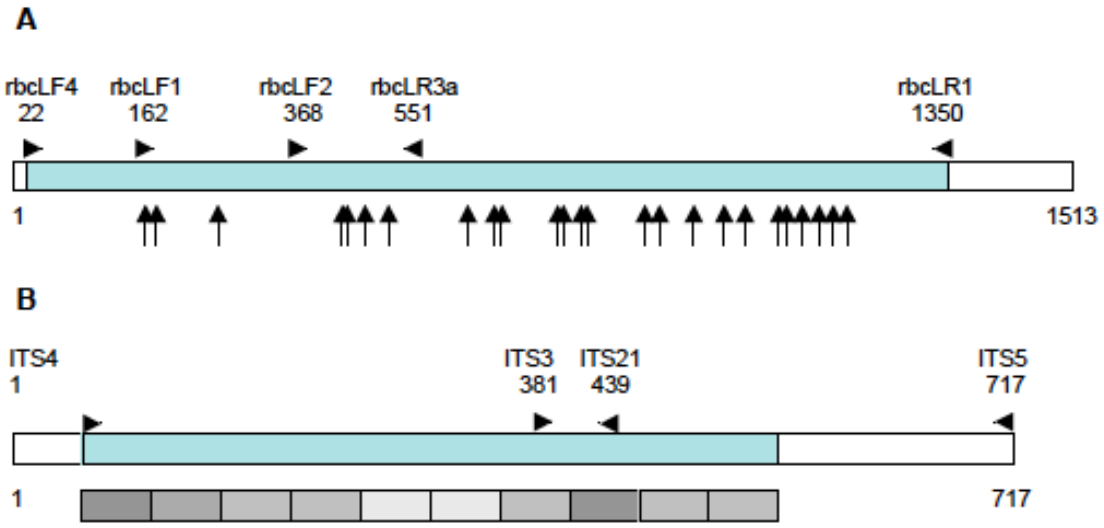


Figure 1. Diagrams of (A) the *rbcL* gene and (B) the ITS sequence, showing the positions of the primers used in amplification and sequence analysis. Shaded regions show the sequences used for the numerical comparisons (Figs. 2-4). In A, arrows point to the positions of base variants identified among the species tested in this work. In B, there were over 80 positions that varied among the species tested. Shading in the lower bar shows the approximate distribution of those variants.

Table 2. Primers used in PCR reactions.

Rubisco		
Forward	rbcLF1	AGTTCCCCCTGAAGAAGCAG
Forward	rbcLF2	TGTTTACTTCCATTGTGGGTAATG
Forward	rbcLF4	ATGTCACCACAAACAGAGACTAA
Reverse	rbcLR1	TTCATTACCCTCACGAGCAAG
Reverse	rbcLR3a	TTCGGTTTAATAGTACAGCCCAAT
ITS		
Forward	ITS3	GCTACGTTCTTCATCGATGC
Forward	ITS4	TCCTCCGCTTATTGATATGC
Reverse	ITS5	GGAAGTAAAAGTCGTAACAAGG
Reverse	ITS21	TATTCAAAACGACTCTCGGCA
TrnF-TrnL		
Forward	trnLF1	AGCTGTTCTAACAAATGGAGTTG
Reverse	trnLF2	GGACTCTATCTTTGTTCTCGTCC
Reverse	trnLF3	TCGACGGATTTTCCTCTTCCTATAAATTTC

Each reaction mixture of 20 µl contained 12.1 µl of water, 4 µl of Green GoTaq buffer (Promega Corporation, Madison, WI, USA), 1.6 µl of dNTPs (2.5 mM of each dNTP), 0.125 µl of Taq DNA polymerase (GoTaq, 5 u/µl, Promega), 0.6 µl of each primer solution (20 µM) and 1 µl of template DNA. Initial PCR conditions were 96^{o}C for 2 min; 35 cycles of 94^{o}C for 30 s, 59^{o}C for 30 s, and 72^{o}C for 1 min; 72^{o}C for 5 min; 4^{o}C hold. Mixtures were separated on 1.5% agarose gels. Bands were cut from the gel and extracted and purified by adsorption and elution from glass filters (Promega Wizard^{R} SV Gel and PCR Cleanup System). Re-amplification of DNA purified from bands used a similar PCR protocol, except the template DNA was diluted (generally 1/10 to 1/100) and only 25 cycles were used for amplification.

Sequence Determination and Analysis

The sequence of each template DNA, using both forward and reverse primers, was determined by the College of Biological Sciences ^{UC}DNA Sequencing Facility (http://dnaseq.ucdavis.edu/). Sequences were aligned and differences identified using Vector NTI Suite 9.

Phylogenetic relationships among *Quercus* species were inferred using nucleotide sequences from internal transcribed spacer 1, the 5.8S ribosomal RNA gene, and internal transcribed spacer 2 (the combination abbreviated ITS herein) for 23 species generated for this study plus 44 sequences published by Manos et al. (1999) and two from Jackson et al. (1999), which we downloaded from GenBank (Table 3). ITS sequences were aligned in ClustalX (Thompson et al. 1997). Bayesian inference was implemented in

MrBayes version 3.2.1 (Ronquist and Huelsenbeck 2003) using the GTR+I+G models and parameters selected based on the Akaike Information Criterion (AIC) with the program jModelTest 2.1.4 (Guindon and Gascuel 2003, Darriba et al. 2012). Two parallel analyses of four Monte Carlo Markov chains each were run for 4 million generations, sampling every 1,000 generations. The first 25% of trees were discarded as burn-in, and the 50% majority-rule consensus tree for the 6,002 trees retained from the two analyses was used to infer phylogenetic relationships and clade support.

Results and Discusson

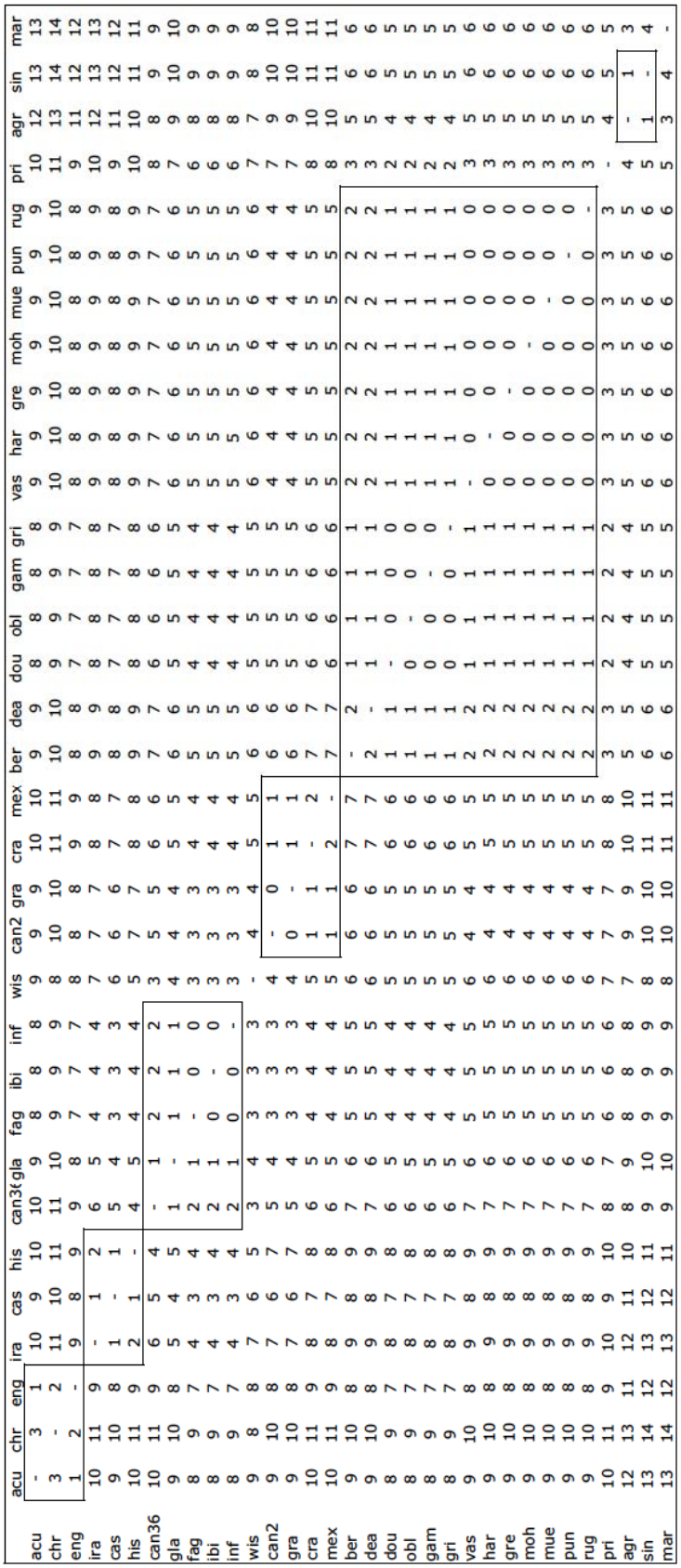
For many of the oak species it was difficult to prepare DNA solutions that did not inhibit the PCR reactions. Leaves collected in the spring provided better templates than ones collected in the summer. Glass purification helped reduce the degree of inhibition. Most samples could be purified using glass spin tubes (Promega Wizard^{R}), but some samples were gelatinous and could only be purified using glass beads, which allow thick solutions of polysaccharides to be washed off. For a few of the species (*Q. durata, Q. lobata, Q. palmeri, Q. turbinella*), it was not possible to obtain PCR products and sequences from the extracted and purified DNA preparations.

rbcL

The primers chosen to determine the sequence of the gene for the large subunit of Rubisco (*rbcL*) gave sequences of 1328 base pairs, representing a large fraction of the gene. Figure 1 shows a diagram of the gene (with length determined from the gene for *Q. suber,* GenBank Accession AB125027.1) and indicates the extent of the sequence amplified by the present primers and the positions of 30 sites that varied among the species that were tested in the present study. Table 1 lists the 33 species from which we were able to determine clear Rubisco sequences. Figure 2 indicates the number of base differences between each pair of species. For one sample, that from a *Quercus* from Iran that was not identified by species, there was a 59-base-pair insertion that was not found in any other sample. For the purposes of comparison in Figure 2, that insertion was counted as one difference.

The data presented in Figure 2 provide a comparison of the *rbcL* sequences of 33 accessions in the UC Davis Arboretum. GenBank also contains the *rbcL* sequences of an additional 16 species found in the Arboretum. References to the GenBank *rbcL* sequences of these additional 16 are given in Appendix 1. Thus *rbcL* sequences are available for 49 of the 74 species and subspecies in the Arboretum. It is unfortunate that four members of the 20 species found in California (Nixon, 2002) refused to give templates for amplification of the *rbcL* gene.

Before the submission of these sequences to GenBank, a search of the GenBank database, using "Quercus AND rbcL" and "Quercus AND carboxylase/oxygenase" produced 103 records involving 47 species and varieties, all different from the ones tested here. Thus the new data reported here have increased by 70% the number of species for which information is available on the *Quercus rbcL* gene.



	acu	chr	eng	ira	cas	his	can36	gla	fag	ibi	inf	wis	can2	gra	cra	mex	ber	dea	dou	obl	gam	gri	vas	har	gre	moh	mue	pun	rug	pri	agr	sin	mar
acu	-	3	1	10	9	10	10	9	8	8	8	9	9	9	10	10	9	9	8	8	8	8	9	9	9	9	9	9	9	10	12	13	13
chr	3	-	2	11	10	11	11	10	9	9	9	8	10	10	11	11	10	10	9	9	9	9	10	10	10	10	10	10	10	11	11	13	14
eng	1	2	-	9	8	9	9	8	7	7	7	8	8	8	9	9	8	8	7	7	7	8	8	8	8	8	8	8	8	9	11	12	12
ira	10	11	9	-	1	2	6	5	4	4	4	7	7	7	8	8	9	9	8	8	8	8	9	9	9	9	9	9	9	10	12	13	13
cas	9	10	8	1	-	1	5	4	3	3	3	6	6	6	7	7	8	8	7	7	7	8	8	8	8	8	8	8	8	9	11	12	12
his	10	11	9	2	1	-	4	5	4	4	4	5	7	7	8	8	9	9	8	8	8	8	9	9	9	9	9	9	9	10	10	11	11
can36	10	11	9	6	5	4	-	1	2	2	2	3	5	5	6	6	7	7	6	6	6	6	7	7	7	7	7	7	7	8	8	9	9
gla	9	10	8	5	4	5	1	-	1	1	1	4	4	4	5	5	6	6	5	5	5	5	6	6	6	6	6	6	6	7	9	10	10
fag	8	9	7	4	3	4	2	1	-	0	0	3	3	3	4	4	5	5	4	4	4	4	5	5	5	5	5	5	5	6	8	9	9
ibi	8	9	7	4	3	4	2	1	0	-	0	3	3	3	4	4	5	5	4	4	4	4	5	5	5	5	5	5	5	6	8	9	9
inf	8	9	7	4	3	4	2	1	0	0	-	3	3	3	4	4	5	5	4	4	4	4	5	5	5	5	5	5	5	6	8	9	9
wis	9	8	8	7	6	5	3	4	3	3	3	-	4	4	5	5	6	6	5	5	5	5	6	6	6	6	6	6	6	7	7	8	8
can2	9	10	8	7	6	7	5	4	3	3	3	4	-	0	1	1	6	6	5	5	5	5	6	6	6	6	6	6	6	7	9	10	10
gra	9	10	8	7	6	7	5	4	3	3	3	4	0	-	1	1	6	6	5	5	5	5	6	6	6	6	6	6	6	7	9	10	10
cra	10	11	9	8	7	8	6	5	4	4	4	5	1	1	-	2	7	7	6	6	6	6	7	7	7	7	7	7	7	8	10	11	11
mex	10	11	9	8	7	8	6	5	4	4	4	5	1	1	2	-	7	7	6	6	6	6	7	7	7	7	7	7	7	8	10	11	11
ber	9	10	8	9	8	9	7	6	5	5	5	6	6	6	7	7	-	2	1	1	1	1	2	2	2	2	2	2	2	3	5	6	6
dea	9	10	8	9	8	9	7	6	5	5	5	6	6	6	7	7	2	-	1	1	1	1	2	2	2	2	2	2	2	3	5	6	6
dou	8	9	7	8	7	8	6	5	4	4	4	5	5	5	6	6	1	1	-	0	0	0	1	1	1	1	1	1	1	2	4	5	5
obl	8	9	7	8	7	8	6	5	4	4	4	5	5	5	6	6	1	1	0	-	0	0	1	1	1	1	1	1	1	2	4	5	5
gam	8	9	7	8	7	8	6	5	4	4	4	5	5	5	6	6	1	1	0	0	-	0	1	1	1	1	1	1	1	2	4	5	5
gri	8	9	7	8	7	8	6	5	4	4	4	5	5	5	6	6	1	1	0	0	0	-	1	1	1	1	1	1	1	2	4	5	5
vas	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	-	0	0	0	0	0	0	3	5	6	6
har	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	-	0	0	0	0	0	3	5	6	6
gre	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	0	-	0	0	0	0	3	5	6	6
moh	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	0	0	-	0	0	0	3	5	6	6
mue	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	0	0	0	-	0	0	3	5	6	6
pun	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	0	0	0	0	-	0	3	5	6	6
rug	9	10	8	9	8	9	7	6	5	5	5	6	4	4	5	5	2	2	1	1	1	1	0	0	0	0	0	0	-	3	5	6	6
pri	10	11	9	10	9	10	8	7	6	6	6	7	7	7	8	8	3	3	2	2	2	2	3	3	3	3	3	3	3	-	4	5	5
agr	12	13	11	12	11	10	8	9	8	8	8	7	9	9	10	10	5	5	4	4	4	4	5	5	5	5	5	5	5	4	-	1	3
sin	13	14	12	13	12	11	9	10	9	9	9	8	10	10	11	11	6	6	5	5	5	5	6	6	6	6	6	6	6	5	1	-	4
mar	13	14	12	13	12	11	9	10	9	9	9	8	10	10	11	11	6	6	5	5	5	5	6	6	6	6	6	6	6	5	3	4	-

Figure 2. Numbers of base sequence differences between the *rbcL* genes of *Quercus* species determined in this work. Abbreviations: acu, *Q.* ×*acutidens*; agr, *Q. agrifolia*; ber, *Q. berberidifolia*; can36, *Q. canariensis*; can2, *Q. candicans*; cas, *Q. castaneifolia*; chr, *Q. chrysolepis*; cra, *Q. crassipes*; dea, *Q.* ×*deamii*; dou, *Q. douglasii*; eng, *Q. engelmanii*; fag, *Q. faginea*; gam, *Q. gambelii*; gla *Q. glandulifera (Q. serrata)*; gra, *Q. gravesii*; gre, *Q. greggii*; gri, *Q. grisea*; har, *Q. hartwegi*; his, *Q.* ×*hispanica*; ibi, *Q. ibirica*; inf, *Q. infectoria veneris*; mar, *Q. margaretta*; mex, *Q. mexicana*; moh, *Q. mohriana*; mue, *Q. muehlenbergii*; obl, *Q. oblongifolia*; pri, *Q. prinoides*; pun, *Q. pungens*; rug, *Q. rugosa*; sin, *Q. sinuata*; ira, *Q. sp.*, Iran; vas, *Q. vaseyana*; wis, *Q. wislizeni*. Boxes point out groups with ≤2 base pair differences.

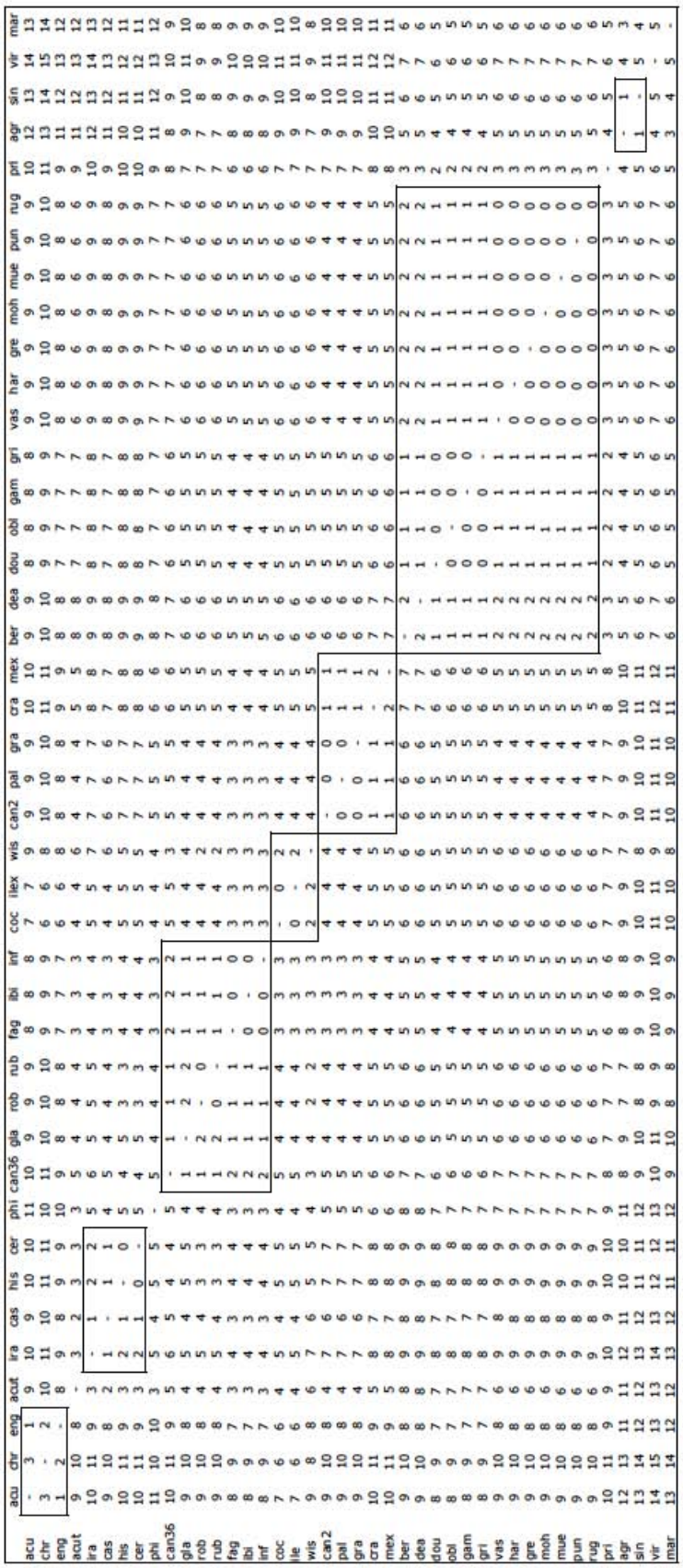


Figure 3. Numbers of sequence differences between the *rbcL* genes of *Quercus* species, including the species noted in Fig. 2 plus the following species: acut, *Q. acutissima*; cer, *Q. cerris*; phi, *Q. phillyraeoides*; rob, *Q. robur*; rub, *Q. rubra*; coc, *Q. coccifera*; ile, *Q. ilex*; pal, *Q. palustris*; vir, *Q. virginiana*. Boxes point out groups with ≤2 base pair differences.

Figure 3 shows a comparison of the sequences used in Figure 2 plus nine more of the longer sequences in the GenBank files, chosen to match as many as possible of the species compared by Manos et al. (1999). What is remarkable is the lack of coincidence between the relationships of the *rbcL* genes and the relationships of the *Quercus* species determined on the basis of the classical and molecular data used by Manos et al. (1999) and later by Oh and Manos (2008). The number of differences among *rbcL* sequences ranged from 0 to 15, with an average of 5.8. There were relatively large base differences between three species grouped in section *Lobatae* (*Q. agrifolia* vs *Q. palustris*, 9 differences; *Q. agrifolia* vs *Q. rubra*, 7). There were also relatively large differences between species grouped in section *Quercus* (*Q. engelmannii* vs *Q. virginiana*, 13; *Q. robur* vs *Q. virginiana*, 9). In contrast, sequences from two species grouped in different sections, *Q. engelmannii* and *Q. chrysolepis*, differed by only one base; similarly, *Q. agrifolia* and *Q. virginiana* differed by only 5 bases; sequences from *Q. berberidifolia* and *Q. douglasii*, again in different sections (Nixon, 2002), did not differ. On the other hand, following the standard taxonomy, the sequences from two species in the "Ilex group", *Q. ilex* and *Q. coccifera*, did not differ at all, and the sequences from two species in the "Cerris group", *Q. cerris, Q. acutissima,* and *Q. phillyraeoides*, differed by only 3-5 bases. Phylogenetic analyses of the *rbcL* sequences generated here (results not shown) provided considerably less resolution than the *ITS* data (see below) and although there was support for some of the same groupings as in the *ITS* analysis (e.g., section Cerris), the placements of several taxa were inconsistent with our *ITS* results and with other phylogenetic studies (Manos et al., 1999; Pearse and Hipp 2009) as well as current infrageneric taxonomy.

ITS

The primers for the ITS sequence defined a section of up to 717 base pairs. Within that section, a segment of 500 base pairs, from base 51 to base 550, was used for comparisons (Fig. 1). Long stretches of poly(G)::poly(C) made sequencing difficult, and even multiple sequencing trials using different primers resulted in some consistently ambiguous sites, although those could represent true heterozygosity. Table 1 lists the 23 species from which we were able to determine ITS sequences. Figure 4 shows the number of clear base differences (+ the number of differences involving ambiguities) between each pair of sequences. In two cases (*Q.* ×*deamii*, *Q. wislizeni*), sequencing with different primers gave results with small differences; both results are shown. *Q. arizonica* gave mixed results: of four individuals with the same accession number, one gave a clear sequence, two showed heterozygosity at several sites, and one did not give clear results.

The data in Fig. 4, comparing ITS sequences determined here, largely support the relationships indicated by the *rbcL* results. Boxes along the diagonal of Fig. 4 were chosen to include, as closely as possible, the species in boxes in Fig. 2, and indeed the numbers in these boxes are relatively low (averaging 6.9, compared to the total collection, which average 14.1). There are also some low numbers that were not indicated in the *rbcL* comparisons. The boxes off the diagonal in Fig. 4, averaging 5.8, indicate a possible relationship between two groups: *Q. canariensis, Q. serrata*, *Q.*

	chr	ira	cas	his	can36	gla-1	gla-2	inf-1	inf-2	dea-1	dea-2	wis-1	wis-2	ari-1	ari-2	can2	gra	cra	mex	ber	obl	gri	vas	moh	pun	lob	pri	mue
chr	-	22+5	21+2	18+3	14+5	13+3	15+3	17+3	17+3	15+7	18+6	23+2	20+2	24+2	24+2	21+6	22+2	19+10	22+5	20+4	17+6	18+3	14+7	16+2	18+3	16+2	15+6	15+3
ira	22+5	-	4+3	7+3	18+6	18+3	18+3	23+4	23+4	20+7	24+5	29+1	25+2	32+1	31+1	27+6	30+0	26+8	28+4	26+4	25+6	26+3	19+10	23+2	22+3	23+2	22+6	22+2
cas	21+2	4+3	-	6+1	17+4	18+1	18+1	23+2	23+2	20+5	23+4	28+0	25+0	31+0	30+0	27+4	29+0	26+7	27+3	27+4	23+5	26+1	19+8	23+0	22+1	23+0	21+5	21+0
his	18+3	7+3	6+1	-	12+5	15+2	15+2	18+3	18+3	15+6	16+5	21+1	20+1	24+1	23+1	20+5	24+1	18+3	22+4	22+5	21+6	23+2	14+9	18+1	17+2	18+1	16+6	14+2
can36	14+5	18+6	17+4	12+5	-	2+5	4+5	8+6	8+6	5+8	11+8	18+4	12+4	18+4	17+4	16+7	16+4	12+12	16+7	9+6	10+8	11+5	4+12	7+4	8+5	5+4	4+8	8+4
gla-1	13+3	18+3	18+1	15+2	2+5	-	3+0	8+3	8+3	5+6	11+5	18+1	12+1	19+1	18+1	16+5	17+0	13+9	17+4	9+5	8+5	9+2	4+9	7+1	8+2	5+1	4+6	8+2
gla-2	15+3	18+3	18+1	15+2	4+5	3+0	-	11+3	11+3	8+6	14+5	21+1	15+1	22+1	21+1	19+5	20+0	16+9	18+4	12+5	9+6	11+2	7+9	10+1	9+2	8+1	6+6	8+2
inf-1	17+3	23+4	23+2	18+3	8+6	8+3	11+3	-	0+0	4+7	9+6	16+2	11+2	17+2	16+2	14+5	16+1	12+9	15+4	11+6	13+6	14+3	3+9	7+2	4+3	8+2	7+7	11+2
inf-2	17+3	23+4	23+2	18+3	8+6	8+3	11+3	0+0	-	4+7	9+6	16+2	11+2	17+2	16+2	14+5	16+1	12+9	15+4	11+6	13+6	14+3	3+9	7+2	4+3	8+2	7+7	11+2
dea-1	15+7	20+7	20+5	15+6	5+8	5+6	8+6	4+7	4+7	-	6+1	12+5	6+5	17+2	15+3	18+7	15+4	11+12	13+7	10+8	10+8	11+5	0+10	2+5	4+6	7+4	6+7	10+5
dea-2	18+6	24+5	23+4	16+5	11+8	11+5	14+5	9+6	9+6	6+1	-	6+4	12+4	10+3	9+3	6+9	11+4	6+11	3+7	16+6	16+7	17+4	6+9	8+4	10+5	13+3	12+6	12+4
wis-1	23+2	29+1	28+0	21+1	18+4	18+1	21+1	16+2	16+2	12+5	6+4	-	6+0	9+0	7+1	4+5	14+0	5+9	7+4	22+3	22+4	23+1	12+9	16+0	17+1	18+0	19+4	18+1
wis-2	20+2	25+2	25+0	20+1	12+4	12+1	15+1	11+2	11+2	6+5	12+4	6+0	-	15+0	13+1	10+4	18+0	10+9	12+4	16+4	16+4	17+1	6+9	10+0	11+1	12+0	12+5	16+1
ari-1	24+2	32+1	31+0	24+1	18+4	19+1	22+1	17+2	17+2	17+2	10+3	9+0	15+0	-	1+1	5+5	12+0	6+8	6+4	23+2	23+4	24+1	16+7	19+0	19+1	19+0	19+5	19+0
ari-2	24+2	31+1	30+0	23+1	17+4	18+1	21+1	16+2	16+2	15+3	9+3	7+1	13+1	1+1	-	3+6	10+1	3+10	4+5	21+3	22+4	23+1	15+7	18+0	18+1	18+0	18+5	18+0
can2	21+6	27+6	27+4	20+5	16+7	16+5	19+5	14+5	14+5	18+7	6+9	4+5	10+4	5+5	3+6	-	9+5	2+14	3+9	19+6	18+6	18+5	12+11	16+4	15+5	15+4	16+8	16+3
gra	22+2	30+0	29+0	24+1	16+4	17+0	20+0	16+1	16+1	15+4	11+4	14+0	18+0	12+0	10+1	9+5	-	6+9	8+4	22+2	23+4	24+1	15+8	17+0	19+1	18+0	19+4	20+0
cra	19+10	26+8	26+7	18+3	12+12	13+9	16+9	12+9	12+9	11+12	6+11	5+9	10+9	6+8	3+10	2+14	6+9	-	4+8	17+11	17+12	18+9	11+14	14+8	14+9	13+8	14+12	14+8
mex	22+5	28+4	27+3	22+4	16+7	17+4	18+4	15+4	15+4	13+7	3+7	7+4	12+4	6+4	4+5	3+9	8+4	4+8	-	20+6	21+7	22+4	13+9	17+3	16+5	17+3	18+7	18+3
ber	20+4	26+4	27+4	22+5	9+6	9+5	12+5	11+6	11+6	10+8	16+6	22+3	16+4	23+2	21+3	19+6	22+2	17+11	20+6	-	11+7	12+4	8+11	13+3	10+5	7+3	6+6	9+5
obl	17+6	25+6	23+5	21+6	10+8	8+5	9+6	13+6	13+6	10+8	16+7	22+4	16+4	23+4	22+4	18+6	23+4	17+12	21+7	11+7	-	0+4	8+12	11+4	10+6	7+4	6+9	8+6
gri	18+3	26+3	26+1	23+2	11+5	9+2	11+2	14+3	14+3	11+5	17+4	23+1	17+1	24+1	23+1	18+5	24+1	18+9	22+4	12+4	0+4	-	9+9	12+1	12+2	8+1	7+7	10+2
vas	14+7	19+10	19+8	14+9	4+12	4+9	7+9	3+9	3+9	0+10	6+9	12+9	6+9	16+7	15+7	12+11	15+8	11+14	13+9	8+11	8+12	9+9	-	0+9	2+10	5+8	4+11	8+9
moh	16+2	23+2	23+0	18+1	7+4	7+1	10+1	7+2	7+2	2+5	8+4	16+0	10+0	19+0	18+0	16+4	17+0	14+8	17+3	13+3	11+4	12+1	0+9	-	7+1	10+0	8+5	13+1
pun	18+3	22+3	22+1	17+2	8+5	8+2	9+2	4+3	4+3	4+6	10+5	17+1	11+1	19+1	18+1	15+5	19+1	14+9	16+5	10+5	10+6	12+2	2+10	7+1	-	7+1	5+5	8+1
lob	16+2	23+2	23+0	18+1	5+4	5+1	8+1	8+2	8+2	7+4	13+3	18+0	12+0	19+0	18+0	15+4	18+0	13+8	17+3	7+3	7+4	8+1	5+8	10+0	7+1	-	2+5	5+1
pri	15+6	22+6	21+5	16+6	4+8	4+6	6+6	7+7	7+7	6+7	12+6	19+4	12+5	19+5	18+5	16+8	19+4	14+12	18+7	6+6	6+9	7+7	4+11	8+5	5+5	2+5	-	4+6
mue	15+3	22+2	21+0	14+2	8+4	8+2	8+2	11+2	11+2	10+5	12+4	18+1	16+1	19+0	18+0	16+3	20+0	14+8	18+3	9+5	8+6	10+2	8+9	13+1	8+1	5+1	4+6	-

Figure 4. Number of sequence differences between the ITS genes of *Quercus* species determined in this work. The two numbers for each entry indicate clear differences + ambiguities, where ambiguities may have occurred because of heterozygosity or because of the difficulty in determining sequences past poly(G) or poly(C) regions through mis-priming in the sequencing reactions. Abbreviations are as given in Figure 2 with the following additions: ari, *Q. arizonica*; lob, *Q. lobata*.

*infectoria veneri*s, and *Q.* x*deamii* with *Q. vaseyana, Q. mohriana, Q. pungens, Q. lobata,* and *Q. prinoides.* Base-sequence similarities in the *ITS* region within the "Ilex group" and the "Cerris group" were also noted by Manos et al. (1999). Samuel et al. (1998) also found ITS identity between *Q. ilex* and *Q. coccifera*.

Phylogenetic analyses (Fig. 5) of ITS sequences determined in this work (Table 1) and by others (Appendix 2) provided support for monophyly of groups corresponding largely to *Quercus* sections *Cerris*, *Lobatae*, *Protobalanus*, and *Quercus*, as was found in previous phylogenetic studies based on ITS sequences (Manos et al. 1999) and AFLP data (Pearse and Hipp 2009). Our results also agree with those past studies with respect to pattern of relationship among the four groups. Sequences from two accessions of *Q. cedrosensis* from Manos et al. (1999) were not resolved within section *Protobalanus* (where the species is classified based on morphology). This result is not surprising, since Manos et al. (1999) also reported anomalous placements of these sequences. Two species classified in section *Quercus*, *Q. canariensis* and *Q. serrata* (synonym of *Q. glandulifera*), were not resolved within the clade with the other members of that section. Those species were not included in the ITS analysis by Manos et al. (1999), but they were resolved within the section *Quercus* clade in the analysis of AFLP data by Pearse and Hipp (2009). We can think of two potential explanations for this discrepancy. First, it is possible that our ITS sequences for these two species are paralogous to those for the other species and include the difference in taxon sampling. Second, we did not have ITS sequences for several of the section *Quercus* species sampled by Pearse and Hipp (2009), and this difference in taxon sampling could explain the difference in phylogenetic resolution between the two studies.

trnL/trnF

The primers chosen to determine the sequence of an intergenic region between *trnL* and *trnF* produced an amplicon of approximately 350 base pairs. Seventeen of the extracts were amplified and sequenced. However, not all amplicons gave clear sequences over the full region; in addition, only four sites showed polymorphisms, providing a maximum sequence difference of three bases between extracts. As a result, the other extracts were not tested. One interesting finding was that the *Quercus* from Iran, noted above as having an insertion in the *rbcL* gene, also had an insertion in the *trnL-trnF* region. This *trnL-trnF* insertion (although not the *rbcL* insertion) was also found in *Q. castaneifolia* (Persian oak). A third species, *Q. faginea*, had an insertion at the same point of the *trnL-trnF* sequence, but its insertion had a different base sequence.

Conclusion

Molecular data such as presented here have contributed to phylogenetic studies, but there are cases where taxonomists question or reject the information (see Nixon, 2002 concerning Manos et al., 2001). It is important to point out that in a gene such as *rbcL*, selection is strongly conservative. Also, base changes that produce amino acid substitutions in the expressed protein may have selective effects. Furthermore, the amino acid substitutions produced by two such base changes could interact positively or

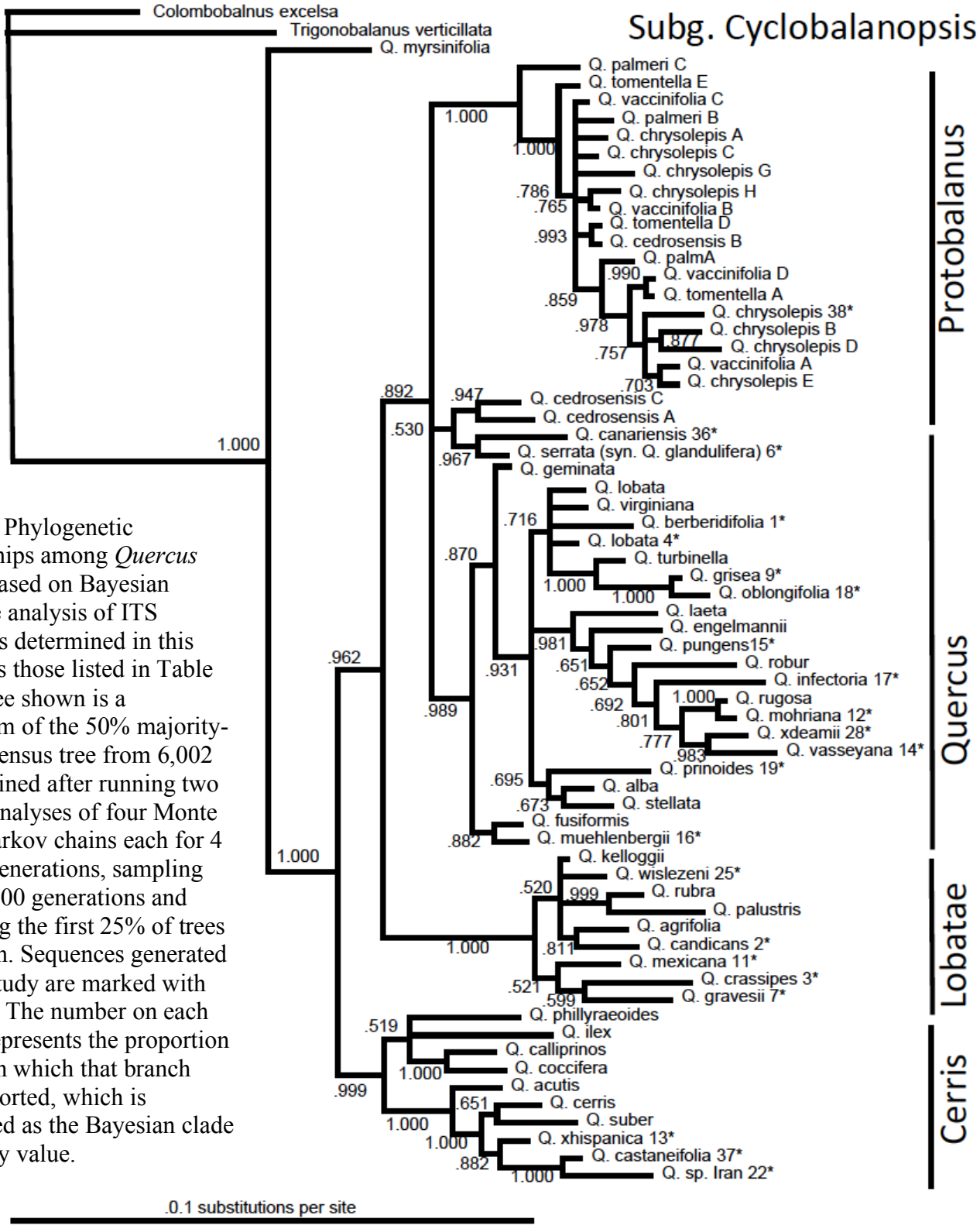


Figure 5. Phylogenetic relationships among *Quercus* species based on Bayesian Inference analysis of ITS sequences determined in this work plus those listed in Table 3. The tree shown is a phylogram of the 50% majority-rule consensus tree from 6,002 trees retained after running two parallel analyses of four Monte Carlo Markov chains each for 4 million generations, sampling every 1,000 generations and discarding the first 25% of trees as burn-in. Sequences generated for this study are marked with asterisks. The number on each branch represents the proportion of trees in which that branch was supported, which is interpreted as the Bayesian clade credibility value.

negatively. The conservative nature of this gene and its limitations for phylogenetic discrimination may be inferred by the shared sequences of groups of species, one of seven species and one of thirteen (Fig. 3). It may be that the sequences of the *rbcL* gene will be found to be particularly sensitive to environmental (external or internal) influences.

Phylogenetic analyses of the more variable non-coding ITS sequences provide support for major groupings and overall relationships within the genus *Quercus*, as was found in a previous study (Manos et al. 1999). While there are potential concerns about the use of ITS in phylogeny reconstruction (e.g., Nixon 2002), especially in a group such as *Quercus* in which hybridization is frequent, the fact that the major groups and patterns of relationship recovered by ITS sequences were also found using AFLP markers (Pearse and Hipp 2009) adds support to the view that these patterns are accurate reflections of phylogenetic relationship. Moreover, the agreement between the patterns with *rbcL* and ITS sequences strengthens considerably the conclusion, earlier advanced by Samuel et al. (1998), that these molecular data can reveal the actual pattern of evolutionary relationships among the *Quercus* species.

References

Darriba, D, Taboada, GL, Doallo, R, and Posada, D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772.

Guindon, S and Gascuel, O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology 52: 696-704.

Jackson, R.B., Moore, L.A., Hoffmann, W.A., Pockman, W.T. and Linder, C.R. 1999. Ecosystem rooting depth determined with caves and DNA. Proc. Natl. Acad. Sci. U.S.A. 96: 11387-11392.

Manos, P. S., Doyle, J. J. and Nixon, K. C. 1999. Phylogeny, biogeography, and processes of molecular differentiation in Quercus subgenus Quercus (Fagaceae). Mol. Phylogenet. Evol. 12: 333-349.

Manos, P. S. and Stanford, A.M. 2001. The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the northern hemisphere. Int. J. Plant Sci. 162 (6 Suppl.): s77-s93.

Manos, P.S., Zhou, Z.-K., and Cannon, C.H. 2001. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. Int. J. Plant Sci. 162: 1361-1379.

Nixon, K.C. 2002. The Oak (*Quercus*) Biodiversity of California and Adjacent Regions. USDA Forest Service Gen. Tech. Rep. PSW-GTR-184. 2002. http://www.fs.fed.us/psw/publications/documents/gtr-184/001_Nixon.pdf

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

Pearse, I.S. and Hipp, A.L. 2009. Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. Proceedings of the National Academy of Sciences USA 106: 18097-18102.

Quercus Portal, https://w3.pierroton.inra.fr/QuercusPortal/index.php?p=PHYLOGENY (accessed 7/23/13)

Ronquist, F. and Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572-1574.

Samuel, R., Bachmair, A., Jobst, J. and Ehrendorfer, F. 1998. ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterrannean, East Asiatic and North American taxa of Quercus (Fagaceae). Plant Syst. Evol. 211: 129-139.

Thompson, J.D., Gibson,T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25:4876-4882.

Xu, L.A. 2004. Diversité de l'ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre Quercus. Thèse Université Henri Poincaré, Nancy, 129 pages.

Appendix 1. Accession numbers for *rbcL* sequences of UCD Arboretum oak species that were not assayed in this project but were given in GenBank. Where multiple GenBank accessions were available, the number for the accession with longest sequence is listed.

Species	Arboretum accession	GenBank accession
Quercus acutissima	A64.0385	AB060578.1
Quercus alba	A64.1174	EU676968.1
Quercus cerris	A64.1304	AB125017.1
Quercus coccifera	A64.1324	AB125018.1
Quercus garryana	A71.0132	HQ184325.1
Quercus ilex	A64.1315	AB125020.1
Quercus ithaburensis	A64.1285	FN675729.1
Quercus macrocarpa	A64.0368	HQ590229.1
Quercus myrsinifolia	A64.0375	AB060572.1
Quercus oleoides	A68.0354	JQ592116.1
Quercus petraea	A93.0319	AB125024.1
Quercus robur	A64.1208	AB125025.1
Quercus suber	A41.0195	AB125027.1
Quercus trojana	A64.0008	FN675725.1
Quercus variabilis	A69.0181	AB060574.1
Quercus virginiana	A64.0012	AF119175.1

Appendix 2. ITS sequence information from GenBank used in the construction of the tree in Fig. 5.

Species	GenBank Accession
Sequences from Manos et al. 1999:	
Colombobalanus excelsa	AF098412
Trigonobalanus verticillata	AF098413
Quercus acutissima	AF098428
Quercus agrifolia	AF098415
Quercus alba	AF098419
Quercus calliprinos	AF098429
Quercus cedrosensis A	AF098449
Quercus cedrosensis B	AF098450
Quercus cedrosensis C	AF098451
Quercus cerris	AF098430
Quercus chrysolepis A	AF098438
Quercus chrysolepis B	AF098439
Quercus chrysolepis C	AF098440
Quercus chrysolepis D	AF098441
Quercus chrysolepis E	AF098442
Quercus chrysolepis F	AF098443
Quercus chrysolepis G	AF098444
Quercus chrysolepis H	AF098445
Quercus coccifera	AF098431
Quercus engelmannii	AF098420
Quercus geminata	AF098426
Quercus ilex	AF098432
Quercus kelloggii	AF098416
Quercus laeta	AF098421
Quercus lobata	AF098422
Quercus myrsinifolia	AF098414
Quercus palmeri A	AF098446
Quercus palmeri B	AF098447
Quercus palmeri C	AF098448
Quercus palustris	AF098417
Quercus phillyraeoides	AF098433
Quercus robur	AF098424
Quercus rubra	AF098418
Quercus rugosa	AF098425
Quercus suber	AF098434
Quercus tomentella A	AF098435
Quercus tomentella D	AF098436
Quercus tomentella E	AF098437
Quercus turbinella	AF098425
Quercus vaccinifolia A	AF098452
Quercus vaccinifolia B	AF098453
Quercus vaccinifolia C	AF098454

Quercus vaccinifolia D	AF098455
Quercus virginiana	AF098427

Sequences from Jackson et al. 1999

Quercus fusiformis	AF174634
Quercus stellata	AF174636