




Complete Chloroplast Genome of Topotype Material of the Coast Live Oak *Quercus agrifolia* Née var. *agrifolia* (Fagaceae) from California

Adam N. Garcia,^a Jennifer Hernandez Ramos,^a Aileen G. Mendoza,^a Asmahan Muhrram,^a Jessica M. Vidauri,^a  Jeffery R. Hughey,^a on behalf of Hartnell College Genomics Group

^aDivision of Mathematics, Science, and Engineering, Hartnell College, Salinas, California, USA

All authors contributed equally to the analysis and writing of this paper.

ABSTRACT Here, we present the chloroplast genome sequence of *Quercus agrifolia* Née, the California live oak, an ecologically important oak species along the coast of California. The genome is 161,283 bp in length, encodes 132 genes, and has a high level of gene synteny to other Fagaceae.

Quercus agrifolia Née, the California live oak or coast live oak, is an evergreen described originally by Luis Née from Monterey, CA (1, 2). The species was said to have sessile axillary fruits and glabrous leaves that were broad, ovate, subcordate, and toothed (1). *Quercus agrifolia* is distributed from northern California to Baja California, Mexico, where it occurs in valleys and slopes in mixed-evergreen forests and woodlands at elevations less than 1,440 meters (3). In addition to the autonym *Q. agrifolia* var. *agrifolia*, two varieties have been proposed, namely, *Q. agrifolia* var. *oxyadenia* (Torr.) J.T. Howell and *Q. agrifolia* var. *frutescens* Engelm. *Q. agrifolia* var. *oxyadenia* is accepted currently while *Q. agrifolia* var. *frutescens* is considered a synonym of *Q. agrifolia* var. *agrifolia* (3). Based on genetic analyses, *Q. agrifolia* is closely related to *Quercus parvula* Greene and *Quercus wislizeni* A. DC. (4–6). More than 30 oak chloroplast genomes have been sequenced (7–9); however, the *Q. agrifolia* genome has not been deciphered. In this study, we assembled and characterized the complete chloroplast genome sequence of *Q. agrifolia* var. *agrifolia* to contribute to the bioinformatics and systematics of this variety and subsection Agrifoliae.

The topotype specimen of *Q. agrifolia* var. *agrifolia* analyzed here was collected from the Presidio in Monterey, CA (36°36'21.9"N 121°53'51.9"W), and deposited at Hartnell College under voucher number HCC 267. The DNA was extracted using the DNeasy blood and tissue kit (Qiagen) following several modifications (10). The binding step during centrifugation was reduced to 4,000 × *g* for 3 minutes, and the DNA was eluted after incubation for 7 minutes in 40 μL Tris-acetate-EDTA (TAE). The 150-bp paired-end PE library was constructed with the NEBNext Ultra II DNA library prep kit (New England BioLabs) and sequenced by Novogene on an Illumina NovaSeq 6000 instrument. The analysis yielded 56,418,120 reads. The genome sequence was assembled by mapping the raw reads onto the reference genome *Quercus coccinea*, GenBank accession number [MN308055](#) (11), using the medium-low sensitivity/fast setting in Geneious Prime 2019.1.3 (Biomatters Limited) resulting in an average coverage of 4,323×. The gaps were closed by iterative mapping using the same settings in Geneious Prime. The annotation was completed using the default settings in GeSeq (12), followed by manual adjustments according to NCBI ORFfinder and Sequin 15.5 (13). Pairwise distances were calculated using the default settings in DIVEIN (14).

The complete chloroplast genome of *Q. agrifolia* var. *agrifolia* is 161,283 bp in length and displays the characteristic flowering plant quadripartite structure (15). The genome contains a

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2022 Garcia et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Jeffery R. Hughey, jhughey@hartnell.edu.

The authors declare no conflict of interest.

Received 6 January 2022

Accepted 15 February 2022

Published 7 March 2022

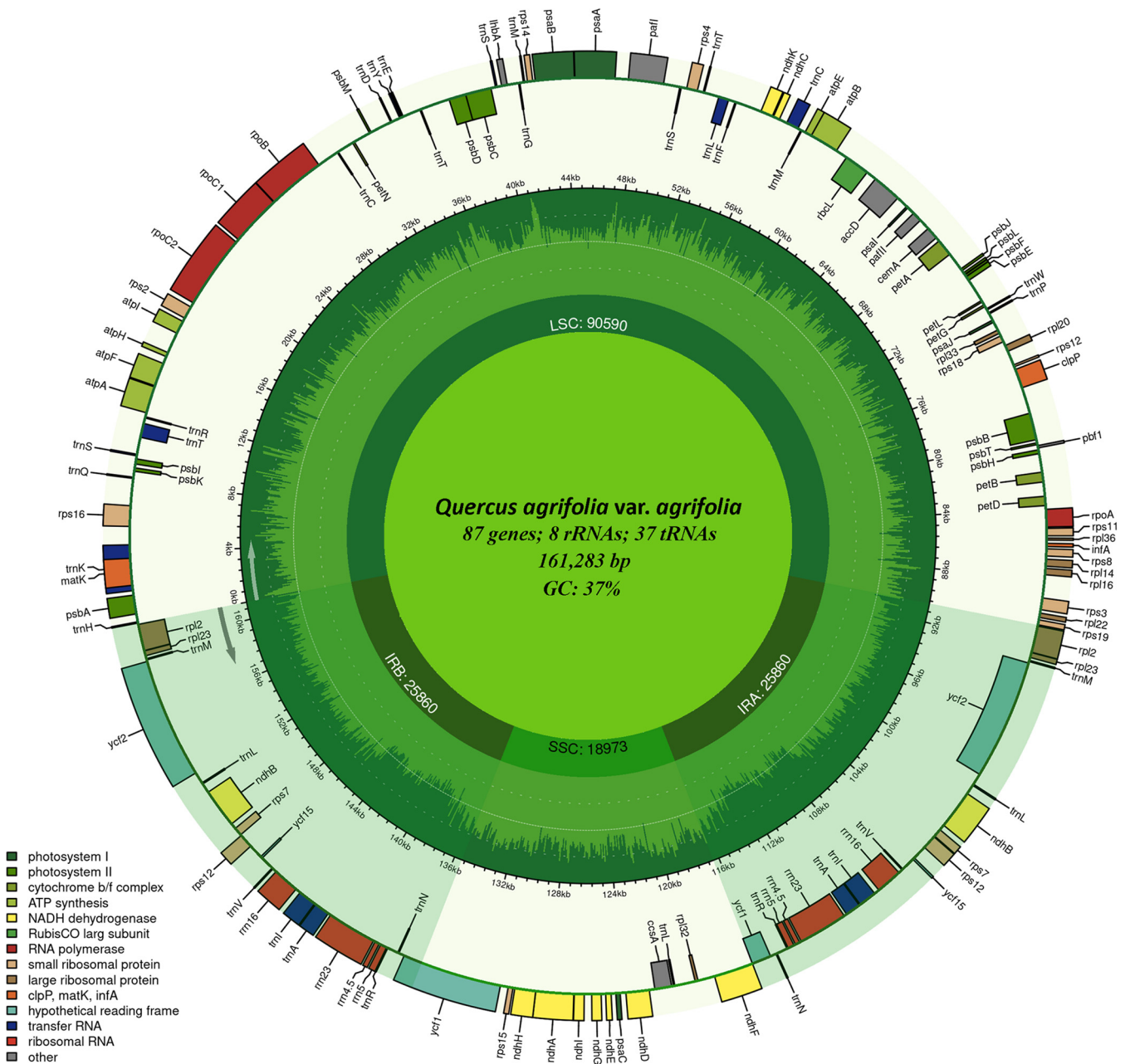


FIG 1 Complete chloroplast genome of *Quercus agrifolia* var. *agrifolia*. The genome was annotated using GeSeq (12), NCBI ORFfinder, and Sequin 15.5 (13) and mapped with Chloroplast (16). The innermost ring identifies the LSC, SSC, and the two inverted repeats. The next ring displays the GC content and direction of transcription, as indicated by the two arrows. The final ring shows the genes. Genes transcribed clockwise are on the inside, while counterclockwise transcriptions are on the outside. The color coding corresponds to genes of different groups as listed in the key in the bottom left.

large single-copy region (LSC), small single-copy region (SSC), and two inverted repeats (IRs) with lengths of 90,590 bp, 18,973 bp, and 25,860 bp, respectively (Fig. 1). The GC content is 37.0%. The genome contains 132 genes, including 87 protein-coding, 37 tRNA, and 8 rRNA genes. Thirteen of the genes contain one intron (*atpF*, *clpP*, *ndhA*, *ndhB*, *rpl2*, *rpoC1*, *rps16*, *trnA*, *trnC*, *trnE*, *trnK*, *trnL*, and *trnT*) and two genes contain two introns (*pafl* and *rps12*). Gene content and organization are identical to other oaks classified in section Lobatae (7, 9, 11). A comparison of the chloroplast genome of *Q. agrifolia* var. *agrifolia* to other Lobatae found it to be 99.79% similar in nucleotide sequence to *Q. rubra*, 99.78% to *Q. palustris*, and 99.77% to *Q. coccinea*.

This chloroplast genome will be useful for examining species-level relationships and *Quercus* systematics, as well as helping to resolve longstanding hypotheses regarding oak hybridization in Lobatae.

Data availability. The complete chloroplast genome sequence of *Quercus agrifolia* var. *agrifolia* is available in GenBank under accession number [OK634019](https://www.ncbi.nlm.nih.gov/nuccore/OK634019). The associated BioProject, SRA, and BioSample numbers are [PRJNA773816](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA773816), [SRX12744819](https://www.ncbi.nlm.nih.gov/sra/SRX12744819), and [SAMN22513939](https://www.ncbi.nlm.nih.gov/biosample/SAMN22513939), respectively. The reference genome for the annotation was *Q. coccinea* (GenBank accession number [MN308055](https://www.ncbi.nlm.nih.gov/nuccore/MN308055)).

ACKNOWLEDGMENT

This research was supported by NSF award number 1832446 to Hartnell College.

REFERENCES

- Née L. 1801. Descripciones de varias especies nuevas de encinos (*Quercus* de Linneo) por D. Luis Née. *An Ci Nat* 3:271–272.
- Greene EL. 1894. Historical notes on some Californian trees.—I. *Erythea* 2: 43–48.
- Baldwin BG, Goldman DH, Keil DJ, Patterson R, Rosatti TJ, Wilken DH (ed). 2012. The Jepson manual: vascular plants of California, 2nd ed. University of California Press, Berkeley, CA.
- Hauser DA, Keuter A, McVay JD, Hipp AL, Manos PS. 2017. The evolution and diversification of the red oaks of the California Floristic Province (*Quercus* section *Lobatae*, series *Agrifoliae*). *Am J Bot* 104:1581–1595. <https://doi.org/10.3732/ajb.1700291>.
- Hermida-Carrera C, Fares MA, Fernández Á, Gil-Pelegrín E, Kapralov MV, Mir A, Molins A, Peguero-Pina JJ, Rocha J, Sancho-Knapik D, Galmés J. 2017. Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. *PLoS One* 12:e0183970. <https://doi.org/10.1371/journal.pone.0183970>.
- Murphy TM, Potter DA. 2017. Genetic Distinctions among oaks in the University of California, Davis Arboretum: contributions to oak phylogeny. *J Int Oak Soc* 28:59–72.
- Alexander LW, Woeste KE. 2014. Pyrosequencing of the northern red oak (*Quercus rubra* L.) chloroplast genome reveals high quality polymorphisms for population management. *Tree Genet Genomes* 10:803–812. <https://doi.org/10.1007/s11295-013-0681-1>.
- Yang Y, Zhu J, Feng L, Zhou T, Bai G, Yang J, Zhao G. 2018. Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: highlighting the effect of codon composition bias in phylogenetic inference. *Front Plant Sci* 9:82. <https://doi.org/10.3389/fpls.2018.00082>.
- Pang X, Liu H, Wu S, Yuan Y, Li H, Dong J, Liu Z, An C, Su Z, Li B. 2019. Species identification of oaks (*Quercus* L., Fagaceae) from gene to genome. *Int J Mol Sci* 20:5940. <https://doi.org/10.3390/ijms20235940>.
- Hughey JR, Gabrielson PW, Maggs CA, Mineur F, Miller KA. 2021. Taxonomic revisions based on genetic analysis of type specimens of *Ulva conglobata*, *U. laetevirens*, *U. pertusa* and *U. spathulata* (Ulvales, Chlorophyta). *Phycological Res* 69:148–153. <https://doi.org/10.1111/pre.12450>.
- Yang X, Yin Y, Feng L, Tang H, Wang F. 2019. The first complete chloroplast genome of *Quercus coccinea* (Scarlet Oak) and its phylogenetic position within Fagaceae. *Mitochondrial DNA B Resour* 4:3634–3635. <https://doi.org/10.1080/23802359.2019.1677189>.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6–W11. <https://doi.org/10.1093/nar/gkx391>.
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW. 2018. GenBank. *Nucleic Acids Res* 46:D41–D47. <https://doi.org/10.1093/nar/gkx1094>.
- Deng W, Maust BS, Nickle DC, Leam GH, Liu Y, Heath L, Kosakovsky Pond SL, Mullins JL. 2010. DIVEIN: a web server to analyze phylogenies, sequence divergence, diversity, and informative sites. *Biotechniques* 48:405–408. <https://doi.org/10.2144/000113370>.
- Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J* 5:2043–2049. <https://doi.org/10.1002/j.1460-2075.1986.tb04464.x>.
- Zheng S, Poccai P, Hyvönen J, Tang J, Amirouf A. 2020. Chloroplot: an online program for the versatile plotting of organelle genomes. *Front Genet* 11:576124. <https://doi.org/10.3389/fgene.2020.576124>.