




# Comparative Analysis of the Chloroplast Genomes of *Quercus* × *morehus* and the Presumptive Parents *Q. wislizeni* and *Q. kelloggii* (Fagaceae) from California

Alejandro Garcia,<sup>a</sup> Althea C. Katada,<sup>a</sup> Alyssa Serrano,<sup>a</sup> Adrea Gonzalez-Karlsson,<sup>a</sup> Angel Carrillo,<sup>a</sup> Angelica Castellanos,<sup>a</sup> Azucena Mendez-Gomez,<sup>a</sup> Carlos J. Flores,<sup>a</sup> Christopher Limon,<sup>a</sup> Cynthia Lopez,<sup>a</sup> Daniela Rosas-Uribe,<sup>a</sup> Dylan J. Hidalgo,<sup>a</sup> Ephraim C. Melgarejo,<sup>a</sup> Erica L. Estamo,<sup>a</sup> Faith Mora,<sup>a</sup> Gabino Guzman,<sup>a</sup> Jason F. Morones,<sup>a</sup>  Jeffery R. Hughey,<sup>a</sup> Jennifer Sanchez-Mendoza,<sup>a</sup> Jimena M. Parra,<sup>a</sup> Joaquin Perez,<sup>a</sup> Joe H. Perez,<sup>a</sup> Joel Viorato Arambula,<sup>a</sup> Juan S. Chavez,<sup>a</sup> Juan R. Figueroa,<sup>a</sup> Juan Rodriguez,<sup>a</sup> Kevin Cardenas,<sup>a</sup> Leslie Trejo,<sup>a</sup> Lizbeth D. Lozano-Ruiz,<sup>a</sup> Loreli Gonzalez,<sup>a</sup> Lorena L. Vargas,<sup>a</sup> Marc Anthony Trujillo,<sup>a</sup> Mariana Rangel,<sup>a</sup> Martin R. Delgado,<sup>a</sup> Mayra A. Ibarra-Moreno,<sup>a</sup> Nancy Chitica Villalobos,<sup>a</sup> Priscila Corona,<sup>a</sup> Quinn Snowden,<sup>a</sup> Roberto Vargas,<sup>a</sup> Robin B. Staretorp,<sup>a</sup> Stephanie Martin,<sup>a</sup> Victor M. Zavala,<sup>a</sup> on behalf of Hartnell College Genomics Group

<sup>a</sup>Division 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 complete chloroplast genomes of *Quercus* × *morehus*, *Q. wislizeni*, and *Q. kelloggii* from California. The genomes are 161,119 to 161,130 bp and encode 132 genes. *Quercus* × *morehus* and *Q. wislizeni* are identical in sequence but differ from *Q. kelloggii* by three indels and eight SNPs.

*Quercus morehus* Kellogg, Abram's oak, was originally proposed from a single specimen from near Clear Lake, CA (1). It was described as a small tree (9.14 m) with black bark, oblong-lanceolate leaves, and oblong nuts. Greene (2) was the first to study *Q. morehus* and concluded it was a hybrid between the interior live oak *Q. wislizeni* A. DC. and the black oak *Q. kelloggii* Newb. Subsequent authors agreed with this hypothesis, including Jepson who itemized six observations supporting the hybrid conclusion (3–7). Many oak chloroplast genomes have been sequenced to date (8–10); however, the genomes of *Quercus* × *morehus*, *Q. wislizeni*, and *Q. kelloggii* have not been analyzed. To contribute to the bioinformatics of *Quercus* × *morehus* and these closely related *Quercus* species, we assembled and characterized the complete chloroplast genomes of the presumptive hybrid and parents.

The leaves of three adjacent specimens were collected in Groveland, California (37°51'22.2N 120°13'36.9W) and deposited at Hartnell College under voucher numbers HCC 268 to 270. The DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) following two modifications: the binding step was centrifuged at 4,000 *g* for 3 min and the DNA was eluted after incubation for 7 min in 40  $\mu$ L TAE (11). The 150 bp PE library was constructed with the NEBNext Ultra II DNA Library Prep kit (New England BioLabs) and sequenced by Novogene on the Illumina NovaSeq 6000. The analysis yielded 40,590,890 (*Quercus* × *morehus*), 17,672,202 (*Q. wislizeni*), and 14,854,920 (*Q. kelloggii*) reads. The adapters and low quality reads were removed using the Trim Adapters and Trim Low Quality default settings with the BBDuk plugin in Geneious Prime 2019.1.3 (Biomatters Limited). The genomes were assembled by mapping reads onto the reference sequence of *Q. agrifolia* Née var. *agrifolia*, GenBank accession number OK634019 (12) using the Medium Sensitivity/Fast setting in Geneious Prime. The mapping coverage for *Quercus* × *morehus* was 4,323 $\times$ , *Q. wislizeni* 1,547 $\times$ , and *Q. kelloggii* 1,885 $\times$ . The gaps were closed by iterative mapping using the same settings in Geneious Prime. The annotation was performed using the default

**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](https://creativecommons.org/licenses/by/4.0/).

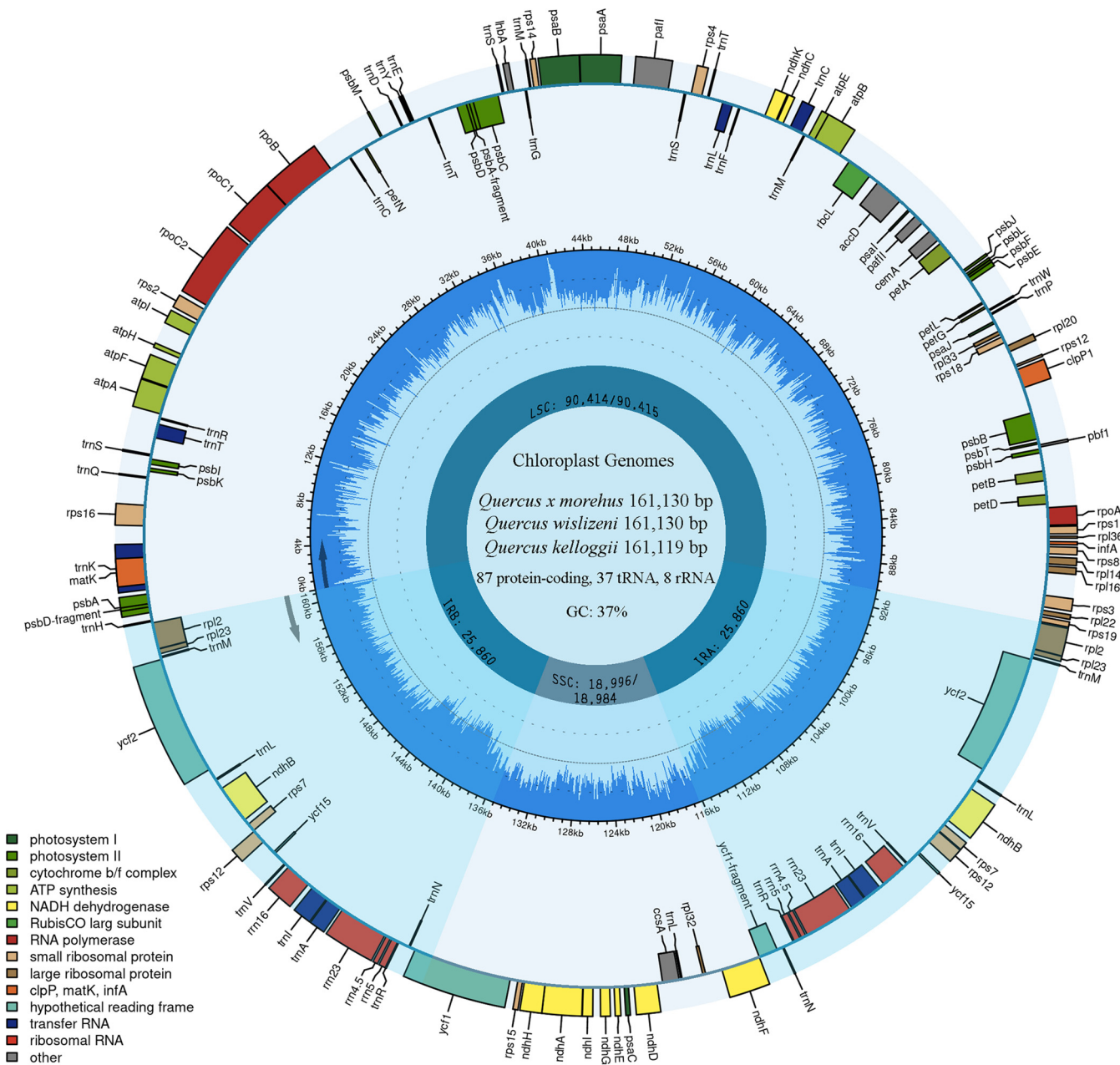
Address correspondence to Jeffery R. Hughey, [jhughey@hartnell.edu](mailto:jhughey@hartnell.edu).

The authors declare no conflict of interest.

**Received** 30 March 2022

**Accepted** 23 May 2022

**Published** 13 June 2022



**FIG 1** Complete chloroplast genomes of *Quercus* × *morehus*, *Q. wislizeni*, and *Q. kelloggii*. The genomes were annotated using GeSeq (13), NCBI ORFfinder and Sequin 15.5 (14), and mapped with CHLOROPLLOT (17). The innermost ring identifies the LSC, SSC, and the two inverted repeats. The numbers before the forward slash correspond to *Quercus* × *morehus* and *Q. wislizeni*, and the numbers after the slash represent *Q. kelloggii*. 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 circle. The color coding corresponds to genes of different groups as listed in the key in the bottom left.

settings in GeSeq (13), followed by manual adjustments according to NCBI ORFfinder and Sequin 15.5 (14).

The complete chloroplast genomes of *Quercus* × *morehus*, *Q. wislizeni*, and *Q. kelloggii* were 161,130, 161,130, and 161,119 bp in length, respectively, and displayed the characteristic flowering plant quadripartite structure (15). Gene content and organization of the three genomes are identical to other oaks classified in section Lobatae (8, 10, 12, 16). The three genomes showed a GC content of 37.0% and contained 132 genes, including 87 protein-coding, 37 tRNA, 8 rRNA genes (Fig. 1). The chloroplast genomes of *Quercus* × *morehus* and *Q. wislizeni* were identical in sequence but differed from *Q. kelloggii* by three indels and eight SNPs (five were located in noncoding and

three in coding regions). Two of the three coding mutations were silent; however, the third altered the stop codon of the *ndhF* gene by 18 bp in *Quercus* × *morehus* and *Q. wislizeni*.

**Data availability.** The complete chloroplast genome sequences of *Quercus* × *morehus*, *Q. wislizeni*, and *Q. kelloggii* are available in GenBank under accession numbers OM541585, OM541583, and OM541584. The Illumina sequencing data for all three specimens are available under BioProject [PRJNA818320](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA818320). The reference genome for the annotation was *Q. agrifolia* var. *agrifolia* (GenBank accession number [OK634019](https://www.ncbi.nlm.nih.gov/nuccore/OK634019)).

## ACKNOWLEDGMENT

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

## REFERENCES

- Kellogg AK. 1863. November 30, 1859: President in the chair. *Proc Calif Acad Sci* 2:36.
- Greene EL. 1889–1890. Illustrations of west American oaks. Drawings by Albert Kellogg, M.D., text by Edward L. Greene. Vols. 1 and 2. Bosqui Engraving and Print Co., San Francisco, California.
- Sargent CS. 1895. The Silva of North America: description of the trees which grow naturally in North America exclusive of Mexico. Houghton, Mifflin and Company, Boston, Massachusetts.
- Jepson WL. 1909. The trees of California. Cunningham, Curtis and Welch, San Francisco, California.
- Jepson WL. 1910. The Silva of California. The University Press, Berkeley, California.
- Munz PA. 1959. A California flora. University of California Press, Berkeley, California.
- 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, California.
- 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>.
- Garcia AN, Hernandez Ramos J, Mendoza AG, Muhram A, Vidauri JM, Hughey JR, Hartnell College Genomics Group. 2022. The complete chloroplast genome of topotype material of the coast live oak *Quercus agrifolia* Née var. *agrifolia* (Fagaceae) from California. *Microbiol Resour Announc* 11. <https://doi.org/10.1128/mra.00004-22>.
- 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>.
- 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>.
- 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>.
- Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. 2020. Chloroplast: an online program for the versatile plotting of organelle genomes. *Front Genet* 11:576124. <https://doi.org/10.3389/fgene.2020.576124>.