



Jul 14, 2023

## Phylogenetic analyses

DOI

[dx.doi.org/10.17504/protocols.io.5jyl8p5j6g2w/v1](https://dx.doi.org/10.17504/protocols.io.5jyl8p5j6g2w/v1)

elizabeth fay<sup>1</sup>, matthew daugherty<sup>1</sup>

<sup>1</sup>Department of Molecular Biology, School of Biological Sciences, University of California, San Diego, La Jolla, CA 92093, USA



Robert Fagiewicz

### Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.5jyl8p5j6g2w/v1>

**Protocol Citation:** elizabeth fay, matthew daugherty 2023. Phylogenetic analyses. **protocols.io**  
<https://dx.doi.org/10.17504/protocols.io.5jyl8p5j6g2w/v1>

**License:** This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Protocol status:** Working

We use this protocol and it's working

**Created:** July 11, 2023

**Last Modified:** July 14, 2023

**Protocol Integer ID:** 84804

**Keywords:** matthew daugherty, original protocol by matthew daugherty, elizabeth fay

**Funders Acknowledgements:**

Aligning Science Across Parkinson's: ASAP

Grant ID: ASAP-000519

## Abstract

Original protocol by Matthew Daugherty and Elizabeth Fay.

## Troubleshooting

- 1 Human LRRK1 (accession NP\_078928.3) and LRRK2 (accession NP\_940980.4) were used as a BLASTp (Altschul et al., 1990) ([http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE\\_TYPE=BlastSearch&LINK\\_LOC=blasthome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)) search query against the Reference Sequence (RefSeq) (<http://www.ncbi.nlm.nih.gov/RefSeq/>) protein database with an e-value cutoff of 1e-20 and a query coverage cutoff of 40%.
- 2 Resulting sequences were aligned using Clustal Omega (Sievers et al., 2011) (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and duplicate and poorly aligning sequences were removed. To reduce the number of nearly identical sequences, sequences with >80% sequence identity were reduced to a single unique sequence using CD-HIT (Fu et al., 2012) (<http://weizhong-lab.ucsd.edu/cd-hit/>) with a 0.8 sequence identity cutoff.
- 3 The resulting 273 sequences (accession numbers and species names found in Supplemental Dataset 1) were realigned with Clustal Omega using two rounds of iteration to optimize the alignment throughout the sequences.
- 4 The resulting alignment is found in Supplemental Dataset 2. To generate maximum likelihood phylogenetic trees of LRRK proteins, IQ-TREE (Nguyen et al., 2015) (<http://www.iqtree.org/>) phylogenies were generated using the "-bb 1000 -alrt 1000" commands for generation of 1000 ultrafast bootstrap (Hoang et al., 2018) and SH-aLRT support values.
- 5 The best substitution model (JTT+F+I+G4) was determined by ModelFinder (Kalyaanamoorthy et al., 2017) using the "-m AUTO" command.
- 6 To confirm that the phylogenetic inferences were not influenced by regions of LRRK proteins that are not well conserved across all family members, the ROC-COR-A region (corresponding to human LRRK1 residues 632-995) and kinase domain region (corresponding to human LRRK1 residues 1229-1534) were extracted from the alignment and concatenated and used as input for IQ-TREE.
- 7 The resulting phylogeny, using the best substitution model (Q.insect+I+G4), shows similarly strong support values in major branches of the phylogeny (Extended Data Fig. 9). Complete phylogenetic trees, with support values, can be found in Supplemental Dataset 3.
- 8 To determine the length of WD40 and COR-B loops, and the aC helix region, well-aligning regions of the alignment, which often corresponded to ordered regions of the LRRK1 and LRRK2 structures, were used as boundaries elements to count the number of intervening residues. Boundary amino acid sequences and residue numbers are shown in Figure 7d.
- 9 For the COR-B loop in cnidarian LRRK4, the automated alignment did not identify the well-conserved WxxGΦxΦ C-terminal boundary element because it is 200+ residues

farther from the N-terminal boundary element in cnidarian LRRK4s than in any other LRRK proteins.

- 10 To measure the loop length, this well-conserved WxxGΦxΦ motif was manually identified in cnidarian LRRK4s and used to count the intervening COR-B “loop” region.
- 11 To determine the presence of basic patches, well-aligning boundary elements flanking each LRRK2 basic patch were identified as described above. 12. For each sequence shown in figure 7g, intervening sequences between the indicated boundary elements were manually searched for any occurrence of three basic residues within a four-residue window.
- 12 To identify the most conserved residues in the first 40 residues of vertebrate LRRK1s residues 1-40 of human LRRK1 were used as a search query against the RefSeq database with an e-value cutoff of 0.05. All resulting 1455 sequences were annotated as vertebrate LRRK1s.
- 13 Sequences were downloaded and aligned using Clustal Omega with two iterations of refinement, and the alignment region corresponding to human residues 1-25 was extracted. Poorly aligning sequences and identical sequences were removed. The remaining sequences were realigned with Clustal Omega with two iterations of refinement. The resulting alignment is shown in Extended Data Figure 7.
- 14 Consensus logos and alignment visualization for basic patch 3 and the COR-B loop region were generated using Geneious Prime 2022.1.1 (<https://www.geneious.com>).

## Protocol references

<https://www.biorxiv.org/content/10.1101/2022.04.26.489605v1>