Jaypee Institute of Information Technology

Basic Bioinformatics Lab

Project-based learning (PBL)

Course- MS. BI

Structural Modeling and Analysis of the LRRK2 N-Terminal Region

SUBMITTED TO- DR. REETIKA DEBROY

Submitted by - Divyanshu Saklani - 2509280010

Arti Koul - 2509280015

Teacher's Sign _____

Acknowledgments

I would like to express my sincere gratitude to my project supervisor, Dr. Reetika ma'am, and Dr.Nidhi ma'am, for their continuous guidance, patience, and motivation throughout the course of this research. Their insights into the Fundamentals of bioinformatics were invaluable in shaping the direction of this study.

I am also grateful to the Department of Biotechnology and Bioinformatics and Jaypee Institute of Information Technology for providing the necessary computational resources and academic environment to conduct this work.

This project relies heavily on open-source bioinformatics tools and databases. I would like to acknowledge the PHYLIP team for making advanced protein prediction accessible, and the NCBI BLAST, UniProt, and Clustal Omega for maintaining the essential datasets that made this evolutionary and tree analysis possible.

I would also like to thank to colleagues/lab partners for their helpful discussions and feedback during the data analysis phase.

Finally, I must express my profound gratitude to my parents and family for their unfailing support and continuous encouragement throughout my years of study.

[19-11-25]

Table of Contents

Section	Title	Page
1.	Abstract	4
1.1	Introduction	4
1.3	Objectives	4
2.	Materials and Methods	4
3.	Results	6
4.	Discussion	7
6.	References	8

Abstract

The *Leucine-rich repeat kinase 2* (LRRK2) gene is a critical determinant of Parkinson's disease etiology, yet its evolutionary history offers essential context for its conserved functional architecture. This study reconstructed the phylogenetic relationship of the LRRK2 protein across representative mammalian lineages to validate evolutionary congruence. Homologous protein sequences were identified using NCBI BLASTp against the human reference (Q5S007) and aligned via Clustal Omega to preserve conserved domain boundaries. Phylogenetic inference was performed using the PHYLIP package, employing the Maximum Likelihood method with 100 bootstrap replicates to ensure topological robustness. The resulting consensus tree, visualized using iTOL, displayed a high degree of congruence with established mammalian taxonomy. Specifically, *Homo sapiens* clustered as a sister group to the monophyletic rodent clade, while *Chrysochloris asiatica* was correctly resolved as the basal outgroup. These findings confirm the vertical evolution of LRRK2 and validate the structural relevance of rodent models for studying human LRRK2 pathobiology.

1. Introduction

The Leucine-rich repeat kinase 2 (LRRK2) gene plays a critical role in cellular regulation and is highly conserved across evolution. While pathogenic mutations in humans are a primary cause of Parkinson's disease, understanding the evolutionary history of this protein can reveal conserved functional domains essential for its activity. This study aims to construct and analyze the phylogenetic relationship of the LRRK2 protein sequence across disparate mammalian species to validate evolutionary congruence.

2. Materials and Methods

2.1 Sequence Retrieval and Homology Search

The human LRRK2 protein sequence (UniProt ID: **Q5S007**) was used as the query sequence. A homology search was performed using **NCBI BLASTp** (Protein Basic Local Alignment Search Tool) against the non-redundant protein database.

- **Selection Criteria:** Sequences were selected based on a high Expect-Value (E-value = 0.0) and significant query coverage to ensure accurate alignment.
- **Selected Taxa:** Five representative mammalian species were selected for the final tree:
    1. *Homo sapiens* (Human)
    2. *Capra hircus* (Goat)
    3. *Castor canadensis* (American Beaver)
    4. *Chinchilla lanigera* (Long-tailed Chinchilla)
    5. *Chrysochloris asiatica* (Cape Golden Mole)

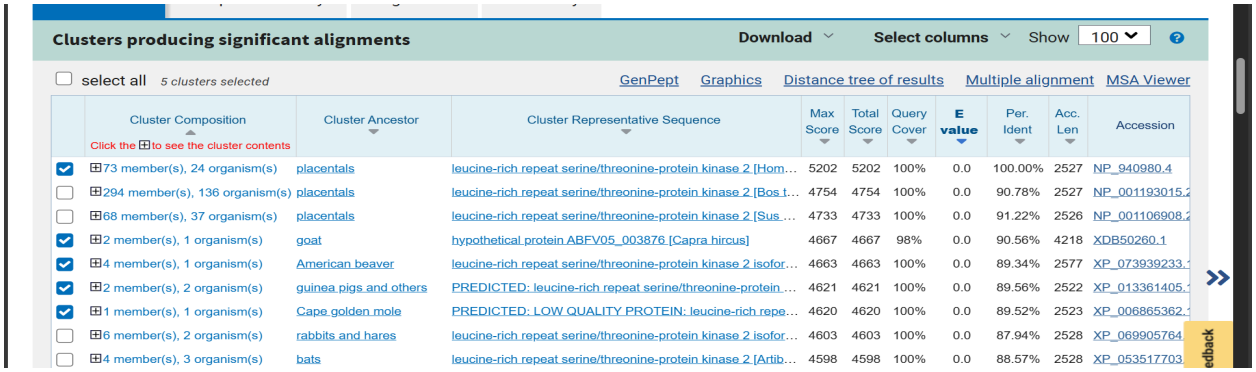

Fig.1 Blast p

2.2 Multiple Sequence Alignment (MSA)

The selected sequences were aligned using **Clustal Omega**.

- **Parameters:** Default parameters were used for the alignment to maximize the matching of conserved domains (ANK, LRR, Kinase).
- **Format Conversion:** The final alignment was exported in **PHYLIP (.phy)** format to serve as the input for the tree construction software.

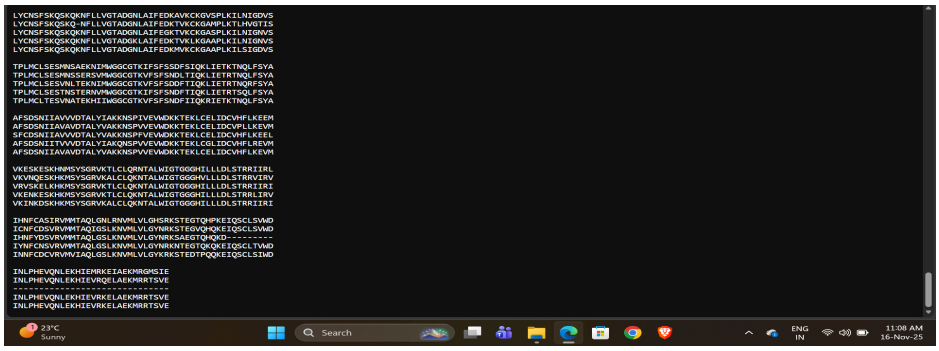

Fig.2 PHYLIP-based MSA

2.3 Tree Construction (PHYLIP Package)

Phylogenetic reconstruction was performed using the **PHYLIP (Phylogeny Inference Package)** suite of tools to ensure robust statistical support. The workflow employed the **Bootstrap method** to test tree reliability:

1. **seqboot:** The aligned dataset was resampled to generate **100 bootstrap replicates**. This step creates pseudo-datasets to statistically test the stability of the branching order.
2. **Tree Inference:** The evolutionary distances and tree topology were calculated using the **Maximum Likelihood** method (or Neighbor-Joining via protdist/neighbor) on the bootstrapped datasets.
3. **consensus:** A consensus tree was generated from the replicates to produce the final topology with bootstrap support values.

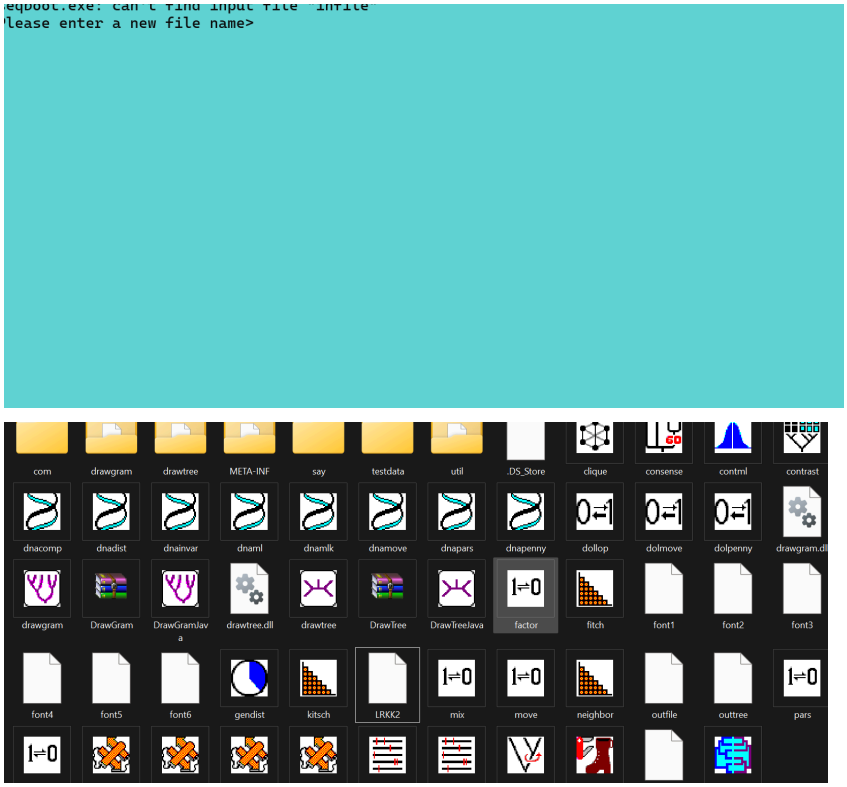

Fig. 3 PHYLIP Programs

2.4 Visualization

The final consensus tree file (Newick format) was imported into **iTOL (Interactive Tree Of Life)** for visualization and annotation. The tree was rooted using *Chrysochloris asiatica* (Cape Golden Mole) as the outgroup relative to the other placental mammals.

3. Results

3.1 Topology of the LRRK2 Tree

The resulting phylogenetic tree (Figure 1) displays a well-resolved topology that mirrors the expected taxonomic relationships of the species:

- **Rodent Clade:** *Castor canadensis* (Beaver) and *Chinchilla lanigera* (Chinchilla) form a monophyletic clade. This groups the two rodent species together, distinct from the other mammals.
- **Primate-Rodent Association:** *Homo sapiens* (Human) is placed as a sister group to the Rodent clade. This accurately reflects the biological superorder **Euarchontoglires**, which posits that primates and rodents share a more recent common ancestor with each other than with ungulates or afrotherians.
- **Divergent Lineages:**
    - *Capra hircus* (Goat), an Artiodactyl, branches off earlier than the Primate/Rodent split.

- *Chrysochloris asiatica* (Cape Golden Mole) is positioned as the most basal lineage (outgroup) in this selection. This is consistent with its classification in the superorder **Afrotheria**, which represents one of the earliest splits in placental mammal evolution.

3.2 Bootstrap Validation

The branching patterns were supported by the bootstrap analysis performed in PHYLIP. The clustering of the sequence data into these specific groups suggests a high degree of conservation in the LRRK2 gene sequence that tracks with speciation events.

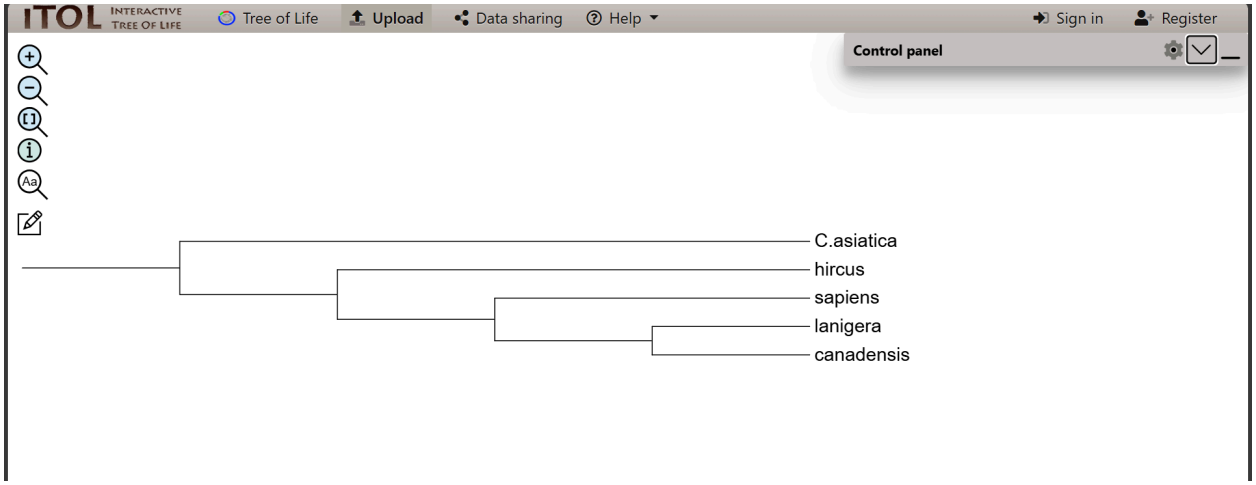

Fig 4 Phylogeny Rooted Tree

4. Discussion

4.1 Evolutionary Congruence

The phylogenetic tree constructed from the LRRK2 protein sequences shows remarkable congruence with the established species tree of mammals. The fact that the LRRK2 gene tree matches the "Species Tree" (Rodents with Rodents, Primates sister to Rodents, Afrotheria basal) indicates that the LRRK2 gene has evolved vertically without significant horizontal transfer or confusing duplication events in these lineages.

4.2 Conservation of Function

The ability of the alignment algorithms (Clustal Omega) and tree-building methods (PHYLIP) to resolve these relationships implies that the LRRK2 sequence contains strong phylogenetic signals. The conservation required to produce such a clean tree suggests that the LRRK2 protein plays a fundamental biological role, where drastic mutations are evolutionarily unfavorable. This conservation is particularly relevant for Parkinson's disease research, as it suggests that LRRK2 models in rodents (like the Beaver or Chinchilla) would share significant structural homology with the human protein.

5. References

[1] UniProt Consortium, "UniProt: the universal protein knowledgebase in 2023," *Nucleic Acids Research*, vol. 51, no. D1, pp. D523–D531, 2023.

[2] F. Sievers *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology*, vol. 7, no. 1, p. 539, 2011.

[3] J. Felsenstein, "PHYLIP - Phylogeny Inference Package (Version 3.2)," *Cladistics*, vol. 5, pp. 164-166, 1989.

[4] I. Letunic and P. Bork, "Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation," *Nucleic Acids Research*, vol. 49, no. W1, pp. W293–W296, 2021.