

Evolutionary Tree Reconstruction of Ebolaviruses

1. Introduction

Ebolaviruses have been a significant focus of research due to their severe impact on human health. This project focuses on reconstructing an evolutionary tree of ebolaviruses to gain insights into their genetic diversity, evolutionary relationships, and transmission patterns. Using genomic sequences from 10 variants retrieved from GenBank, computational methods were applied to identify the origin of the pathogen responsible for the 2014 outbreak in Africa.



2. Objectives

- Reconstruct a phylogenetic tree of ebolaviruses using computational tools.
- Analyze genomic sequences of 10 ebolavirus variants.
- Identify genetic markers and evolutionary patterns among variants.
- Determine the likely origin of the pathogen responsible for the 2014 outbreak.

3. Methodology

3.1 Tools and Software

- **MEGA (Molecular Evolutionary Genetics Analysis):** For sequence visualization and analysis.
- **Clustal Omega:** For multiple sequence alignment.
- **IQ-TREE:** For phylogenetic tree construction using maximum likelihood methods.
- **Python:** For preprocessing and automation of sequence handling.

3.2 Dataset

- Genomic sequences of 10 ebolavirus variants were retrieved from GenBank ([GenBank](#)).
- Sequence lengths ranged between 18,875 and 18,959 base pairs.

3.3 Steps

3.3.1 Sequence Alignment

- Used **Clustal Omega** to perform multiple sequence alignment of the ebolavirus genomic sequences.
- Saved the aligned sequences in a suitable format (e.g., PHYLIP) for further analysis.

3.3.2 Phylogenetic Tree Reconstruction

- Imported aligned sequences into **IQ-TREE**.
- Applied maximum likelihood methods, selected substitution models, and conducted bootstrap analysis to ensure reliability of the tree.

3.3.3 Visualization and Interpretation

- Visualized the phylogenetic tree using **MEGA**.
 - Analyzed the tree to identify evolutionary relationships and genetic diversity among the variants.
 - Interpreted the tree to determine the origin and transmission pathways of the pathogen.
-

4. Results

4.1 Phylogenetic Trees and Observations

Neighbor-Joining Tree

- Constructed using the Neighbor-Joining (NJ) algorithm, which relies on a distance matrix derived from sequence alignment.
- The tree showed the closest relationship between the 2014 outbreak strain (KJ660348) and other Zaire ebolavirus (EBOV) variants, indicating that the outbreak strain belonged to the Zaire species.
- The tree effectively highlighted the genetic divergence between the five species of Ebolavirus, grouping them into distinct clusters.
- Observations revealed significant evolutionary distances between species such as Reston virus (RESTV) and the African variants (EBOV, SUDV, BDBV, TAFV).

UPGMA Tree

- Constructed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), a clustering method assuming a constant molecular clock.
- The tree provided insights into the hierarchical clustering of the Ebolavirus variants, with branch lengths indicating relative divergence times.
- Observations included:
 - The 2014 strain (KJ660348) diverged relatively recently from its closest common ancestor within the Zaire species.

- Branch lengths suggested temporal divergence patterns, with older strains like AF272001 (Yambuku, 1976) showing earlier splits.
- RESTV variants were highly distinct from human-pathogenic species.

Maximum Parsimony Tree

- Constructed using the Maximum Parsimony (MP) method, which minimizes the total number of evolutionary changes.
- The MP tree was unrooted, reflecting ancestral state changes at each node.
- Observations included:
 - Clear characterization of ancestral relationships and character state transitions among nodes.
 - Highlighted key mutations separating the 2014 strain from earlier Zaire variants.
 - Provided insights into specific nucleotide changes and patterns of evolution across species.

4.2 Key Findings

- All three trees consistently identified the Zaire ebolavirus (EBOV) as the likely cause of the 2014 outbreak.
- Each method offered complementary insights:
 - **NJ Tree:** Best for identifying genetic relationships and clustering.
 - **UPGMA Tree:** Useful for exploring temporal divergence and hierarchical clustering.
 - **MP Tree:** Effective for identifying ancestral relationships and character state changes.

5. Discussion

The reconstructed trees provide critical insights into the evolution and transmission dynamics of ebolaviruses. The findings emphasize the role of animal reservoirs, such as bats, in the spread of the virus over large geographical regions. Each tree reconstruction method contributed uniquely to understanding the evolutionary relationships and divergence among the variants. These results are crucial for understanding the mechanisms of zoonotic spillover and for guiding future outbreak management strategies.

6. Conclusion

This project demonstrates the use of computational phylogenetic methods in analyzing ebolavirus evolution. The findings enhance our understanding of genetic diversity, evolutionary relationships, and transmission dynamics, contributing to vaccine development and public health preparedness.

7. References

- [MEGA Documentation](#)
- [Clustal Omega Guide](#)
- [IQ-TREE Manual](#)
- [GenBank Database](#)
- Related academic and research articles on Ebolavirus evolution.