Plotting Phylogenetic Tree Using MOLEBLAST; Technical Guide

Adebule, Adeniyi Philip [1*]

[1*]Department of Microbiology, Ekiti State University, P.M.B. 5363, Ado Ekiti, Nigeria.

*Corresponding author: Adebule, Adeniyi Philip, Email: philbule@yahoo.com

ABSTRACT:

The quest for neighbour sequence of the query for plotting the phylogenetic tree above led to the use of MOLE-BLAST. The construction of phylogenetic tree was demonstrated using three (3) organisms with the accession numbers; NR_074540.1, NR_025685.1 and NR_117706.1 obtained from food waste contaminated soil in Ekiti State, Nigeria and identified as *Bacillus cereus* ATCC strain, *Achromobacter insotitus* strain and *Achromobacter insuavis* strain respectively. These organisms were seen to show some level of relatedness based on the tree diagram. The study guide for the construction phylogenetic tree exposes the easy approach that is needed to utilize the MOLE-BLAST for neighbour search for submitted sequence queries.

Keyword: *Phylogenetic Tree, MOLEBLAST, BLAST, Sequence Analysis, Genetics*

INTRODUCTION:

Mole blast as a neighbor search tool helps taxonomists find the closest database neighbor of submitted query sequences. It can also be termed as a tool that provides taxonomic context for queries by computing the multiple queries sequence alignment [1] along with their top BLAST database hit [2]; discovering their relationship to each other and representing neighborhood search result in a phylogenetic tree. Mole BLAST is a subset of BLAST (Basic local Alignment Search tool), a tool which finds regions of similarity between biological (nucleotide and protein) sequences, compare them to the sequences database and later calculated their statistical significance of matches. BLAST can be used to identify members of gene families by taking into use functional and evolutionary relation between sequences already documented at the database (GENE-BANK) [3].

Mole-BLAST differs from other BLAST tools such as smart BLAST (which finds proteins highly similar to your query); primer-BLAST (Designs primers specific to your PCR templates); Global Align (compare two sequences across their entire span); CD-search which find conserved domains in your sequence, GED (find matches to gene expression profites); igBLAST (search immunoglobutins and t-cell receptor sequences); CDART which find sequences with similar conserved domain architecture; and Bio Assay (which search protein or nucleotide targets in PubChem Bio Assay). MOLE BLAST is a useful tool for Taxonomist, Ecologists, and other scientist to submit sequence to NCBI (National centre for Biotechnology information) to confirm the correct taxonomic annotation of sequences. Also with MOLE-BLAST, assessment of sequence membership in taxonomy groups, discovery of neighbor sequences, visualization of relatedness of type reference specimens to sequences and separation of a large set of sequence into different genes or loci are possible.

MATERIAL AND METHODS

Source and Identity of Organisms

Three (3) organisms with the accession numbers; NR_074540.1, NR_025685.1 and NR_117706.1 were identified as *Bacillus cereus* ATCC strain, *Achromobacter insotitus* strain and *Achromobacter insuavis* strain respectively, were obtained from food waste contaminated soil in Ekiti State, Nigeria.

Open the MOLE-BLAST Interface

In other to access the MOLE-BLAST interface, firstly visit the NCBT website (www.ncbi.nim.nhl.gov), and then click on the BLAST icon situated at the popular resources section (Fig. 1).

On the BLAST page, locate the MOLE-BLAST icon and click to the MOLE-BLAST interface or visit *http://blast.ncbi.nlm.nih.gov/blast/moleblast.cgi* as showcased in Fig. 2.

Entry of sequences queries

Queries can be entered as nucleotide accessions, gis, or FASTA sequences with up to 300 input sequences with up to 5000 bases each. The query sequences to be aligned should be posted in the text area (Fig. 3).
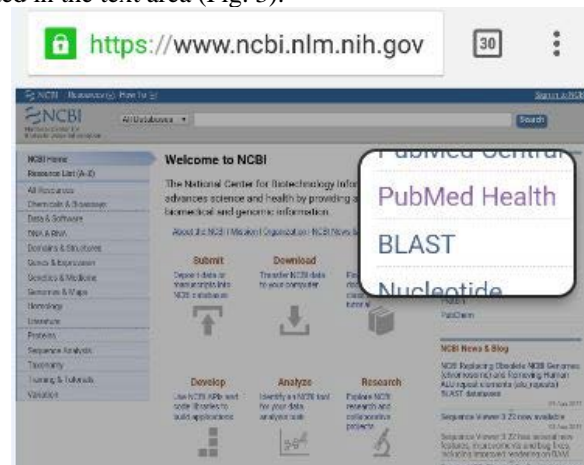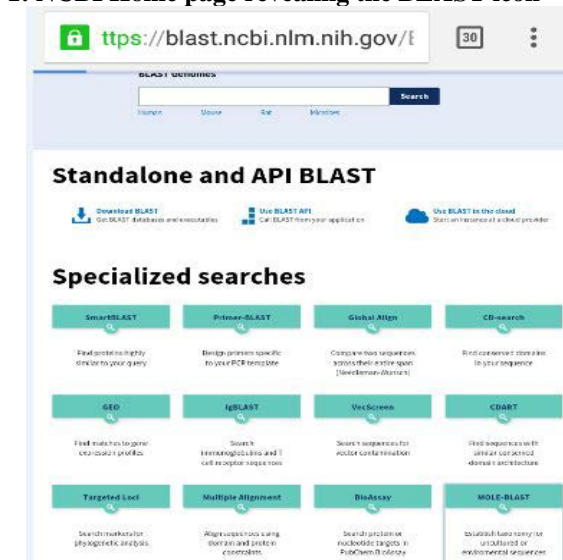

Fig. 1: NCBI Home page revealing the BLAST icon

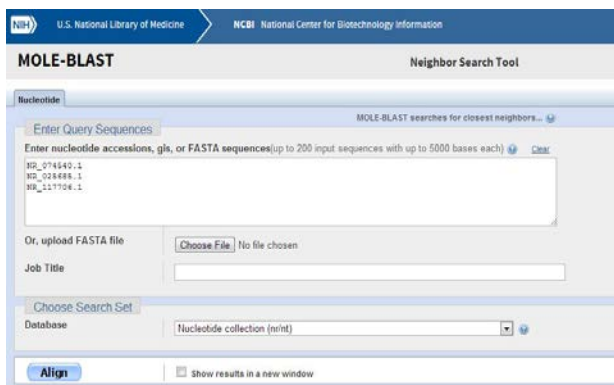**Fig. 2: MOLEBLAST interphase**

Fig. 3: Sequence query entry page

The format of the input is automatically determined. FASTA File is chosen and uploaded in the case of FASTA while accession number can also be imputed. In either case, the same result will be gotten. A title is given to the queries while the search set will be chosen (i.e. nucleotide collection [nr/nt]) as seen in Fig. 3. After adequate entries of queries are made, the ALIGN icon is clicked to query sequences that belong to a cluster along with their top BLAST search whose results are afterwards submitted to Muscle [4] for multiple alignment. This carries out the discovery of neighbour sequence and compares it with the sequences in the data base to draw the phylogenetic tree.

RESULT AND DISCUSSION

The quest for neighbour sequence of the query for plotting the phylogenetic tree above led to the use of MOLE-BLAST**.** In this phylogenetic tree construction guide, Fig. 4 shows the neighbour joining search tools calculating the relatedness of the sequence entry. The result from the calculation is represented in Fig. 5 with phylogenetic tree view either as neighbour joining [5] (5A) or fast minimum evolution [6] (5B). The tree view can be modified using maximum sequence differences (ranging from 0.05 to 0.75) and sequence label [sequence IDo, Taxonomic Name (if available), sequence tittle (if available), Blast name (if available), sequence ID (Blast name), Taxonomic Name (Sequence ID)]. In case of multiple sequences (mostly from difference family), the locus could be more than one (such as locus 1, 2 e. t. c.) (Fig. 6A and B)

The alignment of the neighbour sequences (Fig. 7) discovered at the Gene Bank can be accessed by clicking the "see Alignment" on the phylogenetic tree view page. Better view and modification of the phylogenetic tree view can be accessed by utilizing the different page tool such as the arrow (up, down, right and left), the zoom and refresh key. By clicking the tools icon, other sub-icon is displayed such as download, layout, sort, expand, edit label e. t. c. (Fig. 8). Further click on the layout icon, further give the opportunity to restructure the phylogenetic tree views (Rectangular cladogram, Slant cladogram, Radial tree, circular tree, use distance or mid-point).
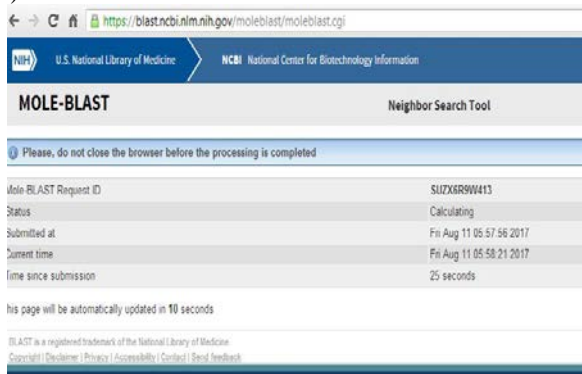

Fig. 4: Neighbor Search Tool

The download icon give room for the easy access of the phylogenetic tree either as ASN text file, ASN binary file, Newick file, NEXUS file or PDF file. For the purpose of this guide, the tree is downloaded as PDF file as shown in Fig.9.
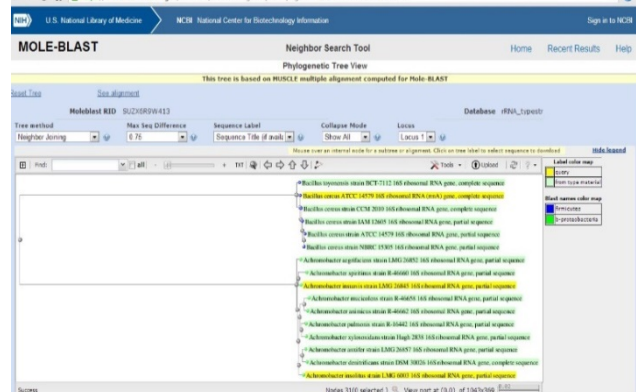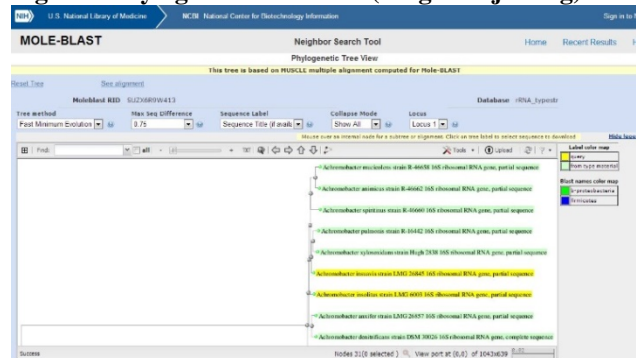

Fig. 5A: Phylogenetic tree view (Neighbor joining)

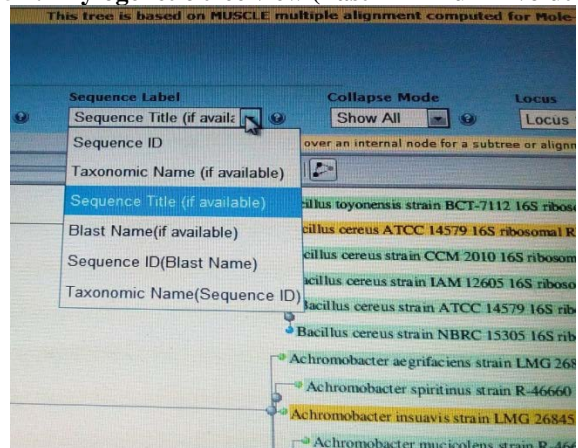**Fig. 5B: Phylogenetic tree view (Fast Minimum Evolution)**
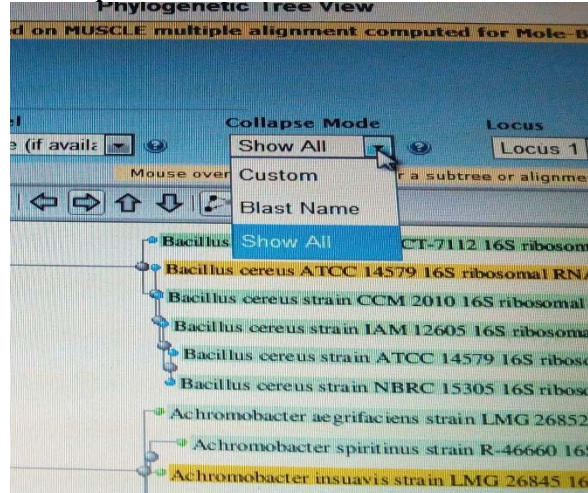

Fig. 6A: Sequence label of trees

Fig. 6B: Collapse Mode
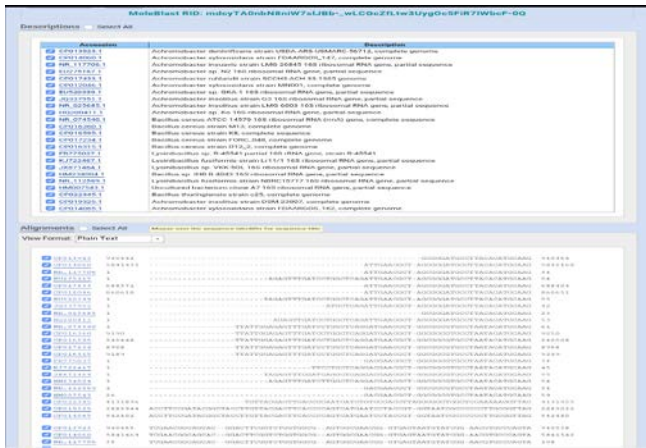
Fig. 7: Alignment of neighbour sequence
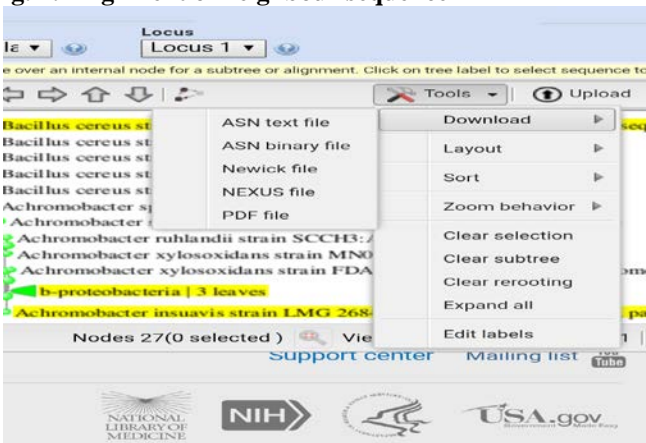
Fig. 8: Tree modification tools page
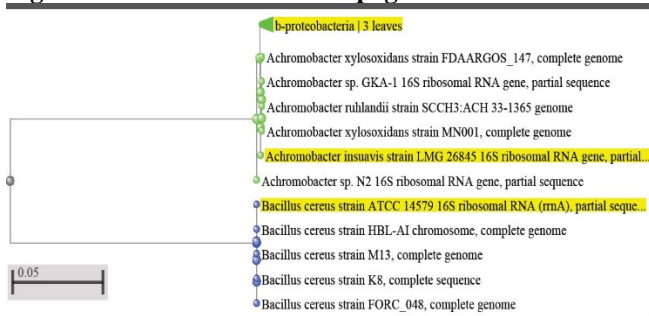

Fig. 9: Sample of phylogenetic tree downloaded in PDF format

CONCLUSION

The study guide for the construction phylogenetic tree exposes the easy approach needed to utilize the MOLE-BLAST for neighbour search for submitted sequence queries. This will further help taxonomist and scientist to verify the correct taxonomic annotation of sequences. In the same biotechnology and genetic student will find it easy to handle MOLE-BLAST search tool to foster better research delivery.

FUTURE FOCUS

- BLAST made easy
- Achieving BLASTING with Mega 7.0.

REFERENCES

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. Journal of Molecular Biology 1990;215:403–410. [PubMed: 2231712]

2. National Center for Biotechnology Information (NCBI), 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research,* 44: 7-19.

3. Wolfe BE, Button JE, Santarelli M and Dutton RJ (2014) Cheese rind communi-ties provide tractable systems for in situ and in vitro studies of microbial diversi-ty, Cell 158:422-433

4. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1997

5. Saitou N and Nei M (1987) The neighbor-joining method: a new method for re-constructing phylogenetic trees. Mol Biol Evol. 4:406-425

6. Desper R and Gascuel O (2004) Theoretical foundation of the balanced mini-mum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. Mol Biol Evol. 21:587-298.