Journal of Phylogenetics & Evolutionary Biology

Khan, J Phylogenetics Evol Biol 2017, 5:2
DOI: 10.4172/2329-9002.1000183

Review Article OMICS International

MEGA - Core of Phylogenetic Analysis in Molecular Evolutionary Genetics

Nida Tabassum Khan*

Department of Biotechnology, Faculty of Life Sciences and Informatics, Balochistan University of Information Technology Engineering and Management Sciences, Quetta, Pakistan

***Corresponding author:** Khan NT, Department of Biotechnology, Faculty of Life Sciences and Informatics, Balochistan University of Information Technology Engineering and Management Sciences, Quetta, Pakistan, Tel: +923368164903; E-mail: nidatabassumkhan@yahoo.com

Receiving date: Jul 27, 2017; **Acceptance date:** Sep 07, 2017; **Publication date:** Sep 14, 2017

Abstract

Since due to tremendous advancement that is being made in sequencing and phylogenetic organisms, wet research techniques single-handedly cannot keep up with the influx of genomic information therefore need of bioinformatics tools such as MEGA is a need. It is user friendly bio-computational software for sequence analysis and phylogenetic tree construction for exploration evolutionary relationships among species or populations.

Keywords: Graphical user interface; Maximum parsimony, Character based methods

Introduction

Molecular evolutionary genetics analysis (MEGA) is a bioinformatics tool used for genome analysis of molecular sequences to measure evolutionary distance for the construction of phylogenies. MEGA software package was designed at the Pennsylvania State University lab under the supervision of Masatoshi Nei along with his postdoctoral fellow Koichiro Tamura and graduate student Sudhir Kumar who were responsible for writing the programming codes for MEGA. This tool was designed to extract valuable information from nucleotide or protein sequence for statistical assessment and biological data mining. Different versions of MEGA were released from time to time with better graphical user interface and enhanced features [1] (Table 1). For the purpose of comparative genome analysis MEGA performs the following:

- Sequence alignments
- Evolutionary distance measurements
- Phylogenetic tree building methods
- Phylogenetic tree evaluation
- Genes/Domains identification
- Selection confirmation
- Implementing sequence statistics

MEGA Layout

MEGA extract valuable information from pairwise or multiple alignments of protein or DNA sequences by employing series of statistical techniques to determine specific physiognomies of nucleotide or proteins for the prediction of evolutionary relationships (Figure 1) [1,2].

Centric design of MEGA

User friendly and context dependent interface: MEGA displays a functional icon that gives users multiple options to choose from, relating to the current input data set that is being analyzed. It enables users to select specific features for computation depending on their designed analysis conditions and purpose of study. Thus making MEGA simple and user friendly (Figure 1).

MEGA Version	Release Date
1	1993
1.1	1994
2	2000
2.1	2001
3	2004
4	2006
4.1	2008
5	2011
5.1	2012
5.2	2013
6	2013
7	2016

Table 1: Major versions of MEGA.

Input data exploration

MEGA comprises of graphic modules for providing facilities like searching, editing, implementing and computing for the query data. It supports different formats (Microsoft Excel, CSV etc) for exporting statistical results for graphical representations and further assessments. In addition functions to select or remove specific genes, domains, and species from primary data set have been possible by means of input data explorers. Besides features like selection of codon positions, codon translations, gaps deletions in alignments etc have been incorporated in MEGA to for better analysis.

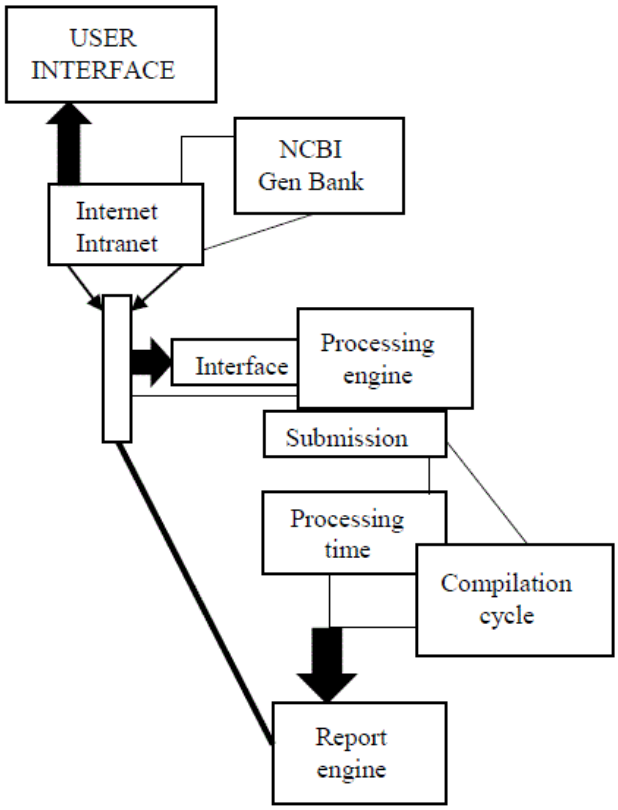

Figure 1: Mega working layout.

Output explorers

MEGA contains many output explorers for graphical representation of the produced results. Such as tree explorer for displaying multiple tree representations, construction of consensus trees, sub-trees compression to identity superior relationships among sequences, printing trees as windows metafiles and TIFF files, transforming trees into newick format, implementation of molecular clock to estimate evolutionary divergence for nodes in a tree etc. Exhibitions of pairwise distances with estimated standard errors is achieved by another output explorer called Distance matrix explorer that also possess sequence or taxa rearrangement facility by drag and drop.

Sequence alignments and data assembly

Sequence data is directly imported into MEGA by alignment explorer that retrieve sequences from databases by means of in-built web browser that directly downloads the selected sequence files in FASTA or other formats. Besides alignment of any user selected sequence region could be edited either inserted or removed from the whole alignment set by means of alignment editor.

Elucidation and transparency results

Diverse computational and statistical analyses could be easily achieved in MEGA through intuitive GUI (Graphical user interface) which enables users to specify their requirements by choosing from multiple options manually i.e. user selection based decision to generate the desired output.

User sessions storage

Attribute of user sessions storage is incorporated in MEGA which allows users to save the current alignment explorer session to a file to which he or she could return later. This saving does not result in loss of alignment visual information and the user can resume the session any time and does not require the need to store data in text rich files [3,4].

Using MEGA for phylogenetic tree construction

Computing evolutionary relationship by means of graphical representation of phylogenies using branching tree like diagram is one of the core applications of MEGA. Phylogenetic tree reveals evolutionary part of a taxa.

Steps to create phylogenetic trees

- Sequences retrieval from databases.
- MSA (multiple sequence alignments) using TCoffee, ClustalW, MUSCLE etc.
- Tree estimation
- Tree evaluation
- Tree representation

Step 1: sequence retrieval

Building sequence alignments involves many steps including acquiring sequences from databanks using browsing facility. Homologous sequences are BLASTed by using either gene name or accession number or a query sequence etc. output containing similar sequence set is generated and displayed on the screen. From this set one can select either all or some sequences i.e. cutting and pasting the sequences from web browser or saving them into files for processing them for sequence alignment. To stream line this process MEGA now includes an integrated web browsing facility to facilitate sequence retrieval from data banks & BLAST search (Figure 2).

Step 2: Sequence alignment

By selecting ADD TO ALIGNMENT option, MEGA exports the sequences automatically to the alignment explorer for producing multiple sequence alignments. Alignment explorer provides two approaches using ClustalW through which data could be displayed as DNA sequence and translated amino acid (Figure 3).

Level 1: construction of data subset containing nearly any combination of sequences including groups, domains, and genes.

Level 2: Inclusion or exclusion of data with missing information or alignment gaps prior to analysis called as complete deletion or pair wise deletion.

Level 3: Data sub setting and transformation in MEGA is accomplished automatically by codon extraction from the selected data subsets and its translation if needed [4].

Step 3: Tree estimation

There are several tree estimation methods given below (Figure 4).

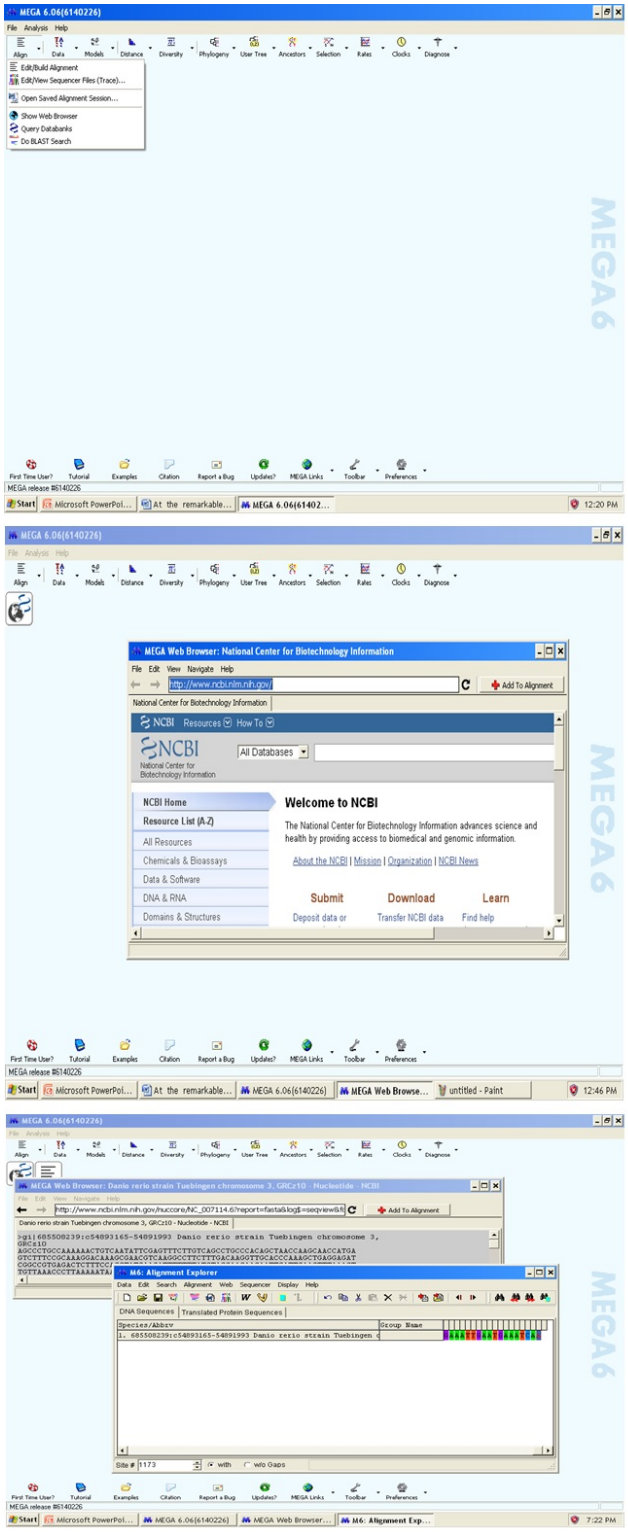

Figure 2: Sequences retrieval in MEGA.

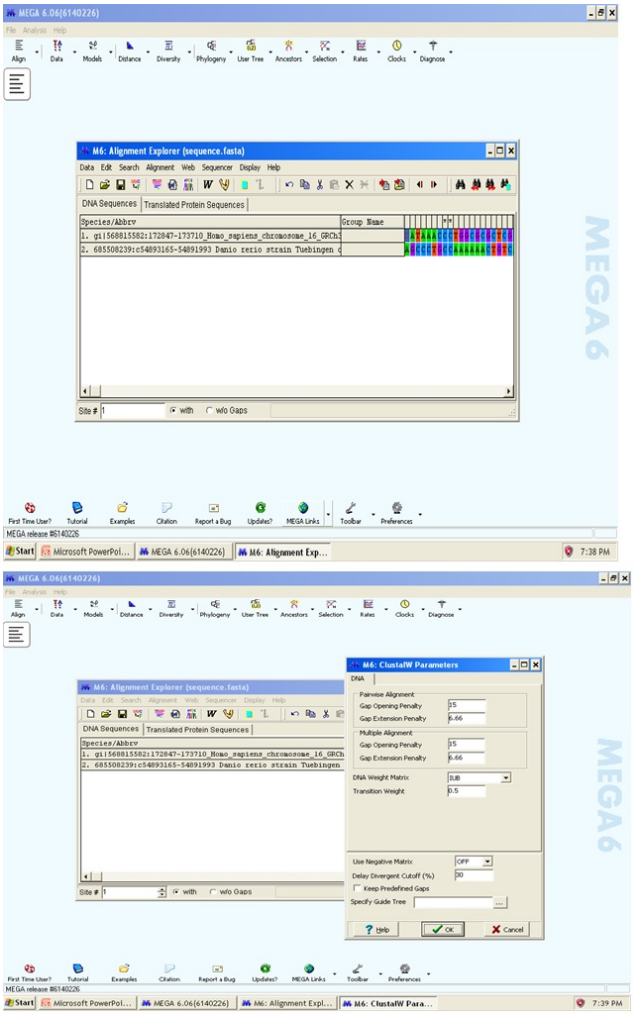

Figure 3: Showing the levels construction of data subset to remove error likely to occur during manual data editing.

Distance based methods such as UPGMA and Neighbor-Joining (NJ) begins with placing all the taxas in a single node and then separates with each repetition. In this way pair of nodes are selected which are grouped at each iteration in order to reduce the overall branch length. However mutation rates are not constant. On the other hand character based methods such as parsimony and probalistic methods selects the tree with the minimum number of alterations as the preferred tree by identifying and estimating the total number of changes at each informative site for each possible tree. Specifically speaking probalistic character based tree estimation methods such as maximum likelihood and bayesian inference finds set of trees with the greatest likelihood by assessing the possibility that a certain evolutionary model (eg. BLOSSUM or PAM matrices) has produced the observed data [5,6].

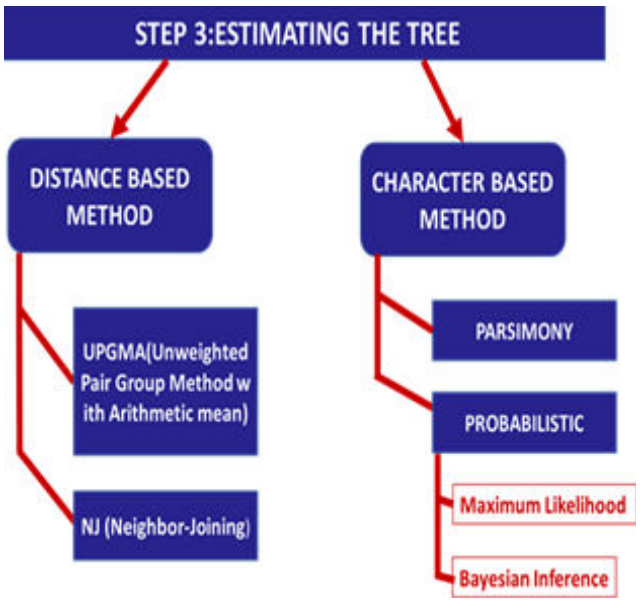

Figure 4: Tree estimation methods in MEGA.

Accuracy				
Bayesian inference				
Maximum likelihood (Decreasing accuracy)				
Maximum parsimony				
Neighbor joining				
Time and Convenience (In terms of time and convenience NJ is found to be the fastest)				
Data set	NeighborJoining	Maximum parsimony	Maximum likelihood	Bayesian inference
Small	1 sec	3 sec	6 sec	-
Small	9 sec	10 min	1 h 34 min	29 min 40 sec
Large	1 sec	22 sec	3 min 29 sec	-
Large	86 sec	10 h 2 min	58 h	6 h 33 min

Table 2: Efficiency of major tree estimating methods.

Comparison of different tree estimating methods

Comparison of different tree estimating methods is mentioned above explaining which method is the most appropriate in terms of accuracy, time and convenience (Table 2) [7].

Step 4: Bootstrapping

It is important to conduct statistical test for evaluating phylogenetic tree. Therefore MEGA runs a statistical re-sampling process called bootstrapping to check trees reliability by measuring the probability of branch recovery if the taxa were sampled again. Its values are typically from 1000 repeated calculations and values >70% is acceptable.

Step 5: Phylogenetic tree exploration

Constructed phylogenetic trees can be visualized in numerous ways such as topologies without branch length e.g. Cladogram or with estimated branch length e.g. Phylogram or in linearized fashioned by means of tree explorer [5,6].

Conclusion

Thus MEGA is user friendly software for sequence alignment and comparative genome analysis. Study of phylogenies by means of phylogenetic tree construction is one of the major applications of this tool.

References

1. Kumar S, Tamura K, Nei M (1994) 'MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers'. Comput Appl Biosci 10: 189-191.
2. Kumar S, Dudley J, Nei M, Tamura K (2008) MEGA: A biologist centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinformatics 9: 299306.
3. Kumar S, Tamura K, Nei M (2004) MEGA3: Molecular Evolutionary Genetics Analysis. Brief Bioinformatics 5: 150-163.
4. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis. Mol Biol Evol 24: 1596-1599.
5. Hall BG (2013) Building phylogenetic trees from molecular data with MEGA. Molecular biology and evolution, 30: 1229-1235.
6. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 28: 2731-2739.
7. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution, 33: 1870-1874.