

MEGA

Molecular Evolutionary
Genetics Analysis Software

Mega: Conceito e Objectivo

- Software para explorar e analisar sequências de proteínas ou ADN
- A sequenciação do genoma está a gerar grandes quantidades de dados, de uma ampla gama de organismos
- Necessidade de programas de fácil utilização, com algoritmos computacionais rápidos e métodos estatísticos úteis

Principais Funcionalidades

Sequence Mining and Alignment

Web data mining and BLASTing
 Importing/editing of Trace files from Sequencers
 Integrated codon/protein Alignment Explorer
 Manual and native-CLUSTAL alignments

Data Exploration and Statistics

Evolutionary Explorer for active sequence data
 One-click translation of coding sequences
 Context-sensitive highlighting of variable and other types of sites
 Export data to other formats (e.g., Nexus, Phylip)
 Estimation of nucleotide and amino acid compositions
 Estimation of relative synonymous codon usage (RSCU)
 * Export statistics in Excel, CSV, and text formats

Data Subsets Supported

Selection of groups, genes, and domains
 Selection of codon positions and individual sites
 Automatic, context-sensitive codon translation
 Handling of gaps (pairwise and complete-deletion)

Exploration of Evolutionary Distances

Display of pairwise distances and errors simultaneously
 Sorting of rows by name, groups, and drag-and-drop
 Computing within and between group distances
 * Summarize sequences involved in invalid distances
 * Export distances in Excel, CSV, and text formats

Data types

DNA and protein sequence alignments
 Pairwise distance matrices
 Unaligned sequences for alignment
 Use predefined or create new genetic code tables

Distances

Separate distances by site degeneracy and codon positions
 Separation of distances into transitions & transversions
 Computation of synonymous and nonsynonymous components
 Estimate differences and ratios between distance components
 Bootstrap and analytical variances of estimates
 Estimate within, between, and net sequence diversity
 Maximum Composite Likelihood (MCL) distance

Models

DNA, codon, and protein models
 No. of differences and p-distance for all types of differences
 Jukes-Cantor and Poisson distances
 Tajima-Nei, Tamura 3-parameter, and Tamura-Nei distances
 Original and modified Nei-Gojobori codon distances
 Original and modified Li-Wu-Lou codon distances
 Equal input, Dayhoff and JTT amino acid distances
 Incorporate base composition differences between lineages
 Incorporate rate variation among sites (Gamma distances)
 * Model selection using Likelihood

Principais Funcionalidades

Exploration of Phylogenies

- Build consensus and condensed trees
- Linearized tree and calibrate clock to estimate divergence times
- Display most parsimonious ancestral states
- Rectangular, radiation, slanted, and circular tree displays
- Flip and swap subtrees, and change position of the root
- Customize size, fonts, and colors of text and lines
- Collapse clusters and associate images with nodes
- Scale-bar display for substitutions and time
- Import/export trees in newick, TIFF and metafile formats

Description of Results and Assumptions

- Figure legend style descriptions for all results
- Context-dependent format for display
- Generates citations/references for all methods employed

Text File Handling

- Edits very large files, permits rectangular block edits/cuts
- Format conversion from popular formats (e.g., Nexus)
- Unlimited Undo, and utilities for reverse complement etc.

Supporting Operating Systems

- Windows Multi-user multi-threaded 32-bit application
- Virtual PC and emulators (Linux, Mac, SunOS)
- Native Linux via WINE application compatibility layer
- HTML context-sensitive help on all platforms

Pattern of substitution

- Homogeneity test for substitution patterns (Disparity Index Test)
- Estimation of global 4x4 pattern of nucleotide substitution (MCL)
- Estimation of global transition/transversion bias (MCL)
- Counts of nucleotide pair frequencies between sequences

Tree Inference

- NJ and UPGMA with random tie-breaking
- ME with Close-Neighbor Interchange (CNI) heuristic search
- MP with Max-mini Branch-and-bound search
- Mini-mini with search factor and CNI heuristic searches for MP
- Fast ordinary least squares branch length estimation
- Average over pathway method for parsimony branch lengths
- Bootstrap tests of the inferred phylogeny (all methods)
- Interior-branch tests for NJ and ME under least-squares
- * Maximum Likelihood for large trees (DNA and protein)

Tests of Selection

- Codon-based tests of selection (Large and small sample)
- Test for sequence pairs, within groups, or average over all
- Tajima's test based on segregating sites and sequence diversity

Relative Rate tests & Molecular clocks

- Tajima's test for DNA and protein data
- Analyze transitions, transversions, or codon positions
- Linearized tree for estimating divergence times (in Tree Explorer)

Evolução Cronológica

- A primeira versão do MEGA foi desenvolvida tendo em conta os recursos disponíveis num computador pessoal no início dos anos 90
- MEGA2 foi designado de forma a aproveitar o aumento de recursos computacionais e fornecer uma interface gráfica
- As versões MEGA3 e MEGA4 forneceram novas funcionalidades e melhoramentos, relativamente às versões anteriores

Evolução Cronológica

B

Sequence Data Explorer
Data Display Highlight Statistics Help

Color C V Pi S 0 2 -4 UUC Phc

Molecular Evolutionary Genetics Analysis
File Data Distances Phylogeny

Click me to activate a data file
Go to the MEGA2 web page
Citing MEGA2 in publications

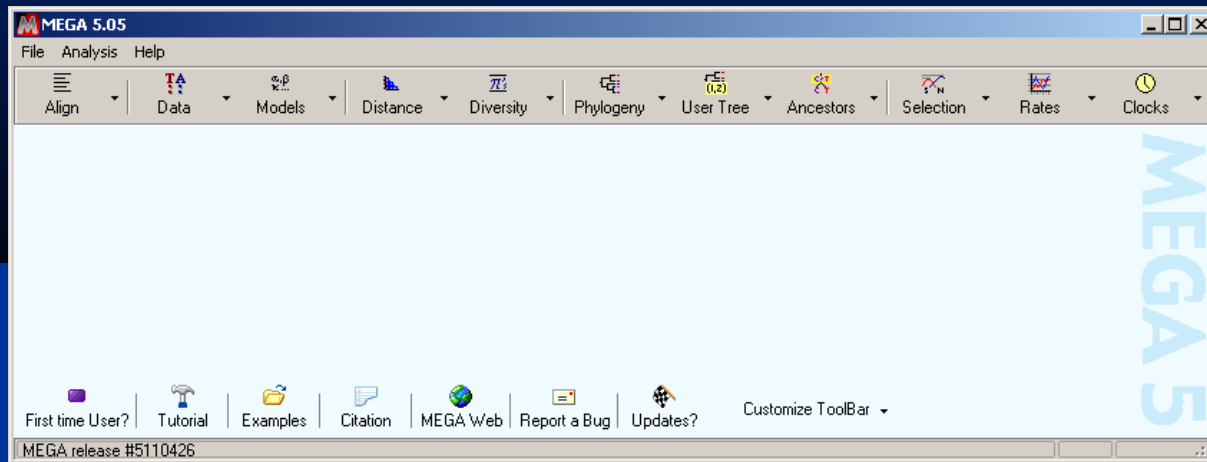
<input checked="" type="checkbox"/> W. Pygmy (1) {African}	C	A	T	A	A	A	A	A	C	G	C	A	A	C	G	C	A	C	A	T	C	A	A	A	C	C	T	G	C	C	C	C	
<input checked="" type="checkbox"/> W. Pygmy (6) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	C	G	C	A	C	A	T	C	A	A	A	C	C	T	A	G	C	C	C
<input checked="" type="checkbox"/> Kung (7) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (9) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (10) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (13) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (17) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (19) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Kung (22) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	C	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Asian (23) {Asian}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	T	C	C	C	
<input checked="" type="checkbox"/> Yoruban (25) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	C	T	C	C	C	
<input checked="" type="checkbox"/> Yoruban (26) {African}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	C	C	C	C	
<input checked="" type="checkbox"/> Asian (29) {Asian}	C	A	T	A	A	A	A	A	A	C	G	C	A	A	T	G	C	A	C	A	T	C	A	A	A	C	C	T	T	C	C	C	

Data File: C:\Program
Title: Adh-sequ

1/633 Variable: 131/633 Data

16:33:11

MEGA5



- Desenvolvido especificamente para reduzir o tempo necessário para tarefas comuns
- Fornece novos métodos estatísticos de análise
- Uma interface mais user-friendly

Conclusão

- O MEGA é uma plataforma com diversas ferramentas úteis para biólogos e bioinformáticos
- Permite aos biólogos executar diversas tarefas, tais como alinhamentos de sequências, análises filogenéticas,...
- A sua interface gráfica facilita a sua utilização e compreensão das ferramentas que possui