Multiple Alignments and Phylogenies

Mark Voorhies

3/31/2011

Review

- Computer “lab notebook”
    - Dates
    - Input/output/parameters

Review

- Computer “lab notebook”
  - Dates
  - Input/output/parameters (screenshots)

Review

- Computer “lab notebook”
    - Dates
    - Input/output/parameters (screenshots)
    - E-mail result synopses (*e.g.*, to PI)

Review

- Computer “lab notebook”
  - Dates
  - Input/output/parameters (screenshots)
  - E-mail result synopses (*e.g.*, to PI)
- BLAST
  - Fast, heuristic local alignment with statistical foundation

Review

- Computer “lab notebook”
  - Dates
  - Input/output/parameters (screenshots)
  - E-mail result synopses (*e.g.*, to PI)
- BLAST
  - Fast, heuristic local alignment with statistical foundation
  - Gosh, we should have saved our BLAST reports!

Comprehending our BLAST results

- We have a bunch of sequences that look similar to our query

Comprehending our BLAST results

- We have a bunch of sequences that look similar to our query
- We infer that they are homologous to each other

Comprehending our BLAST results

- We have a bunch of sequences that look similar to our query
- We infer that they are homologous to each other
- What does that mean, anyway?

Nomenclature

Homologs heritable elements with a common evolutionary origin.

Nomenclature

Homologs heritable elements with a common evolutionary origin.

Orthologs homologs arising from speciation.

Paralogs homologs arising from duplication and divergence within a single genome.

Nomenclature

Homologs heritable elements with a common evolutionary origin.

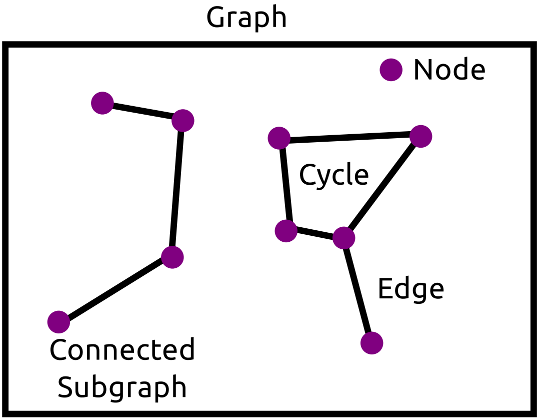
Orthologs homologs arising from speciation.

Paralogs homologs arising from duplication and divergence within a single genome.

Xenologs homologs arising from horizontal transfer.

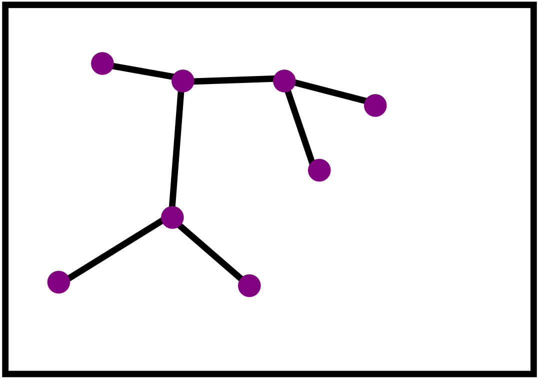
Onologs homologs arising from whole genome duplication.

Nomenclature

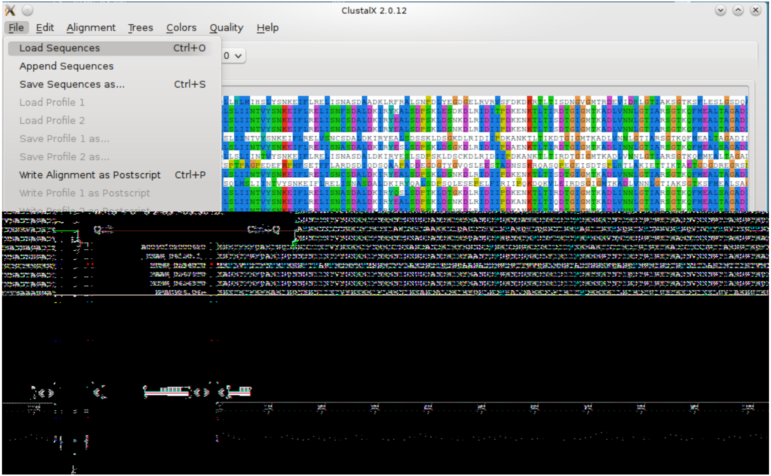


Nomenclature

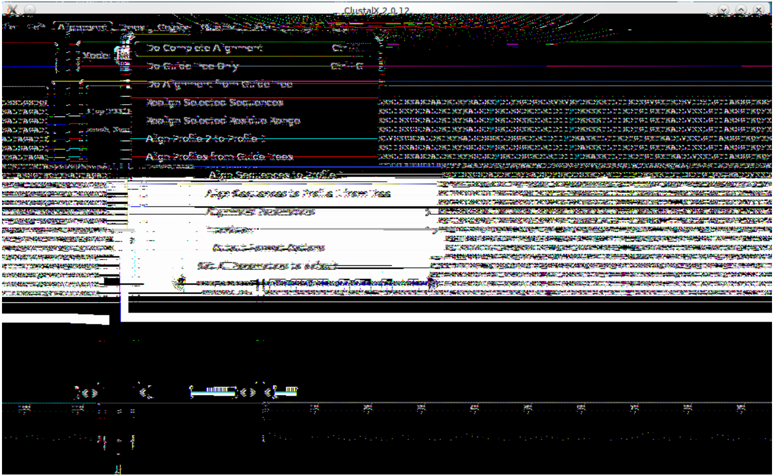
Tree = Connected Graph with no Cycles



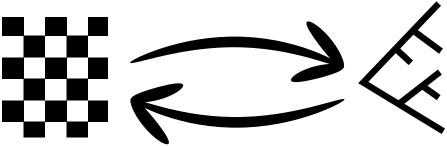
Generating a multiple alignment in CLUSTALX



Generating a multiple alignment in CLUSTALX


ClustalX 2.0.12
Alignment
Do Complete Alignment
Do Guide Tree Only
Do Alignment from Guide Tree
Realign Selected Sequences
Realign Selected Residue Range
Align Profile 2 to Profile 1
Align Profiles from Guide Trees
Align Sequences to Profile 1
Align Sequences to Profile 1 from Tree
Output Format Options

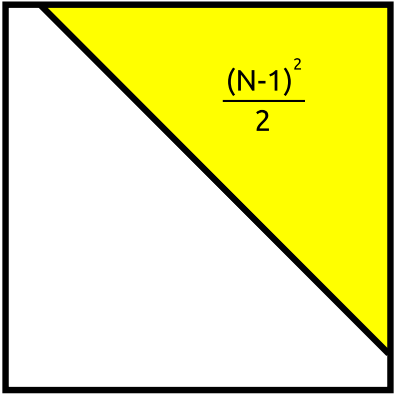
Evolution implies a self-consistent model



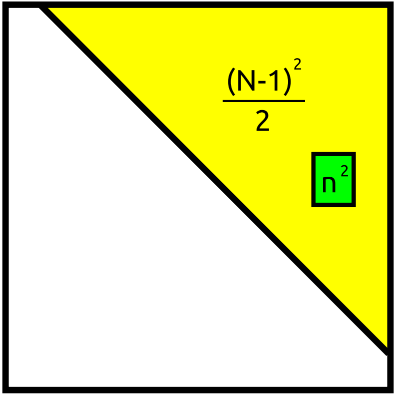
Distances
(Pairwise relationships)

Topology
(Evolutionary history)

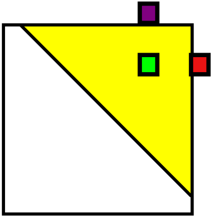
Measure all pairwise distances by dynamic programming

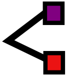


Measure all pairwise distances by dynamic programming

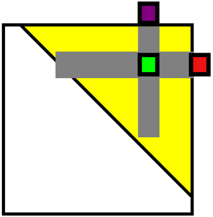


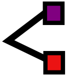
Generate a guide tree by UPGMA



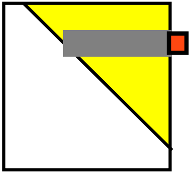


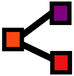
Generate a guide tree by UPGMA



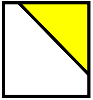


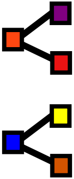
Generate a guide tree by UPGMA



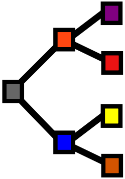


Generate a guide tree by UPGMA

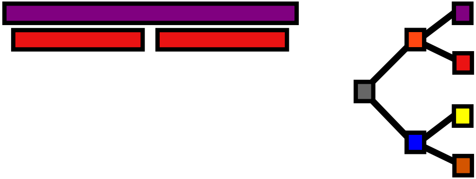




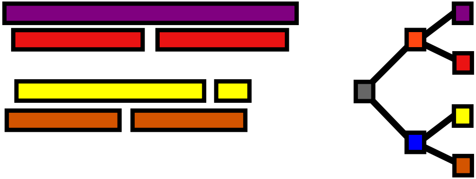
Generate a guide tree by UPGMA



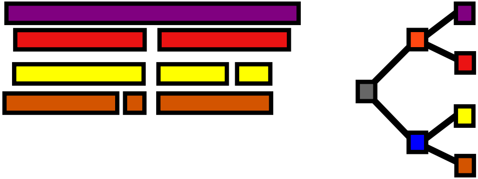
Progressive alignment following the guide tree



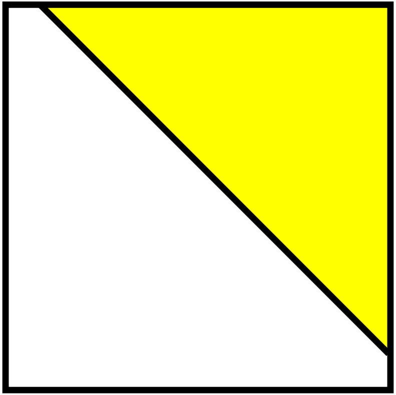
Progressive alignment following the guide tree



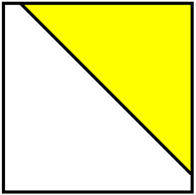
Progressive alignment following the guide tree

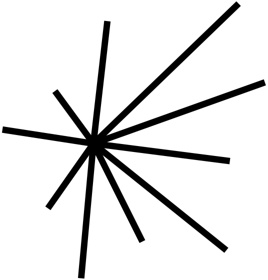


Measure distances directly from the alignment

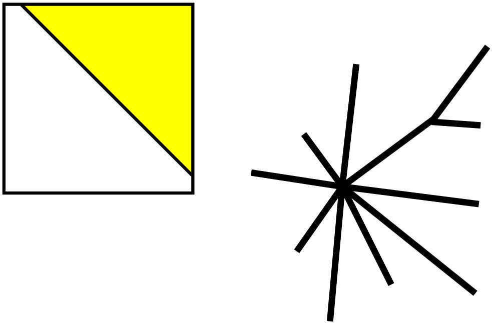


Generate neighbor-joining tree from new distances

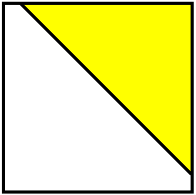


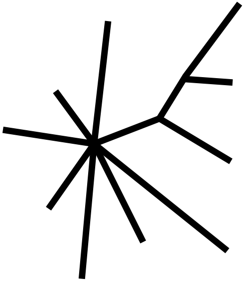


Generate neighbor-joining tree from new distances

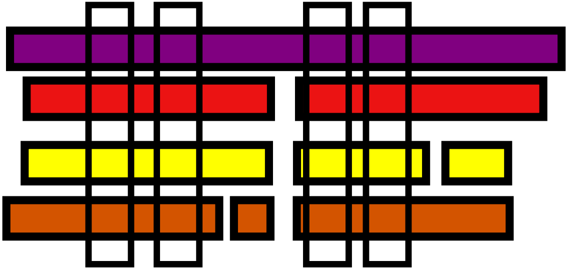


Generate neighbor-joining tree from new distances

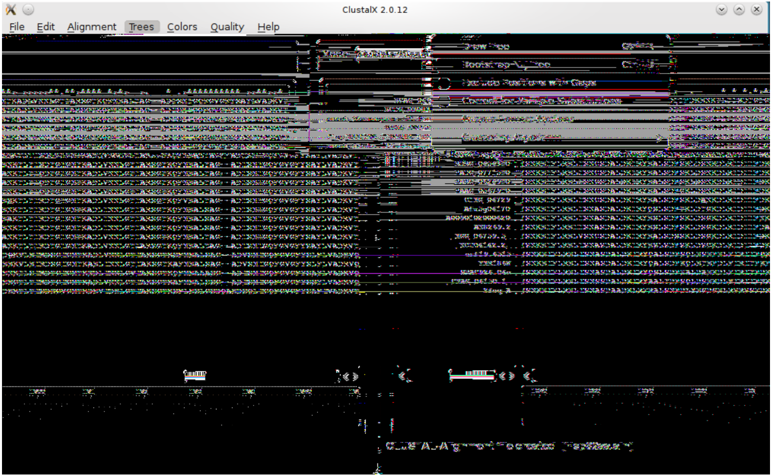




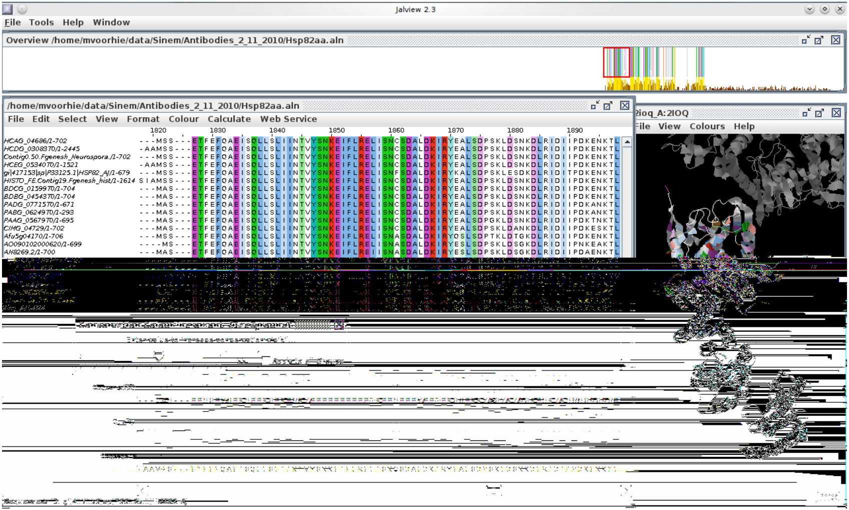
Generate bootstrap values from subsets of the alignment



Generating a neighbor joining tree in CLUSTALX



Viewing the alignment and tree in JALVIEW

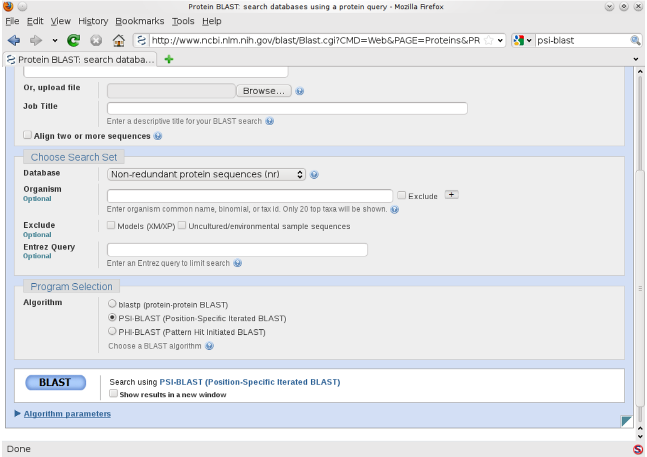

Jalview 2.3
File Tools Help Window
Overview /home/mvoorhie/data/Sinem/Antibodies_2_11_2010/Hsp82aa.aln
/home/mvoorhie/data/Sinem/Antibodies_2_11_2010/Hsp82aa.aln
File Edit Select View Format Colour Calculate Web Service
2ioq_A:2IOQ
File View Colours Help

```
                                         1820       1830       1840       1850       1860       1870       1880       1890
HCAG_04686/1-702                      ---MSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
HCDG_03083T0/1-2445                   -AAMSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
Contig0.50.Fgenesh_Neurospora./1-702  ---MSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
HCEG_05340T0/1-1521                   -AAMSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
gi|417153|sp|P33125.1|HSP82_AJ/1-679  ---MSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNFSDALDKIRYKALSDPSKLESDKDLRIDITPDKENKTL
HISTO_FE.Contig19.Fgenesh_hist/1-1614 SIAMSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
BDCG_01599T0/1-704                    ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
BDBG_04543T0/1-704                    ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKENKTL
PADG_07715T0/1-671                    ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKANKTL
PABG_06249T0/1-293                    ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDSNKDLRIDIIPDKANKTL
PAAG_05679T0/1-695                    ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPGKLDSNKDLRIDIIPDKTNKTL
CIMG_04729/1-702                      ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNCSDALDKIRYEALSDPSKLDANKDLRIDIIPDKESKTL
Afu5g04170/1-706                      ---MSS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNASDALDKIRYQSLSDPTKLDTGKDLRIDIIPDKENKTL
AO090102000620/1-699                  ----MS---ETFEFQAEISQLLSLIINTVYSNKEIFLRELISNASDALDKIRYESLSDPSKLDSGKDLRIDIIPNKEAKTL
AN8269.2/1-700                        ---MAS---ETFEFQAEISQLLSLIINTVYSNKEIFLREIISNASDALDKIRYESLSDPSKLDSGKDLRIDIIPDAENKTL
```

Related tools

- Multiple Alignment
    - T-Coffee
    - MUSCLE
    - COBALT
- Tree building
    - MrBayes (Bayesian MCMC)
    - PhyML (maximum likelihood)

Searching with PSI-BLAST



Protein BLAST: search databases using a protein query - Mozilla Firefox

http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&PAGE=Proteins&PR

Or, upload file — Browse…

Job Title — Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database — Non-redundant protein sequences (nr)

Organism (Optional) — □ Exclude — Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude (Optional) — □ Models (XM/XP) □ Uncultured/environmental sample sequences

Entrez Query (Optional) — Enter an Entrez query to limit search

Program Selection

Algorithm

○ blastp (protein-protein BLAST)

● PSI-BLAST (Position-Specific Iterated BLAST)

○ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

BLAST — Search using PSI-BLAST (Position-Specific Iterated BLAST)

□ Show results in a new window

▶ Algorithm parameters

Done

File Edit View History Bookmarks Tools Help
psi-blast
Protein BLAST: search databa…

Homework

- Play with CLUSTALX, JALVIEW, and PSI-BLAST
- Read PLoS Comp. Biol. 4:e1000069