

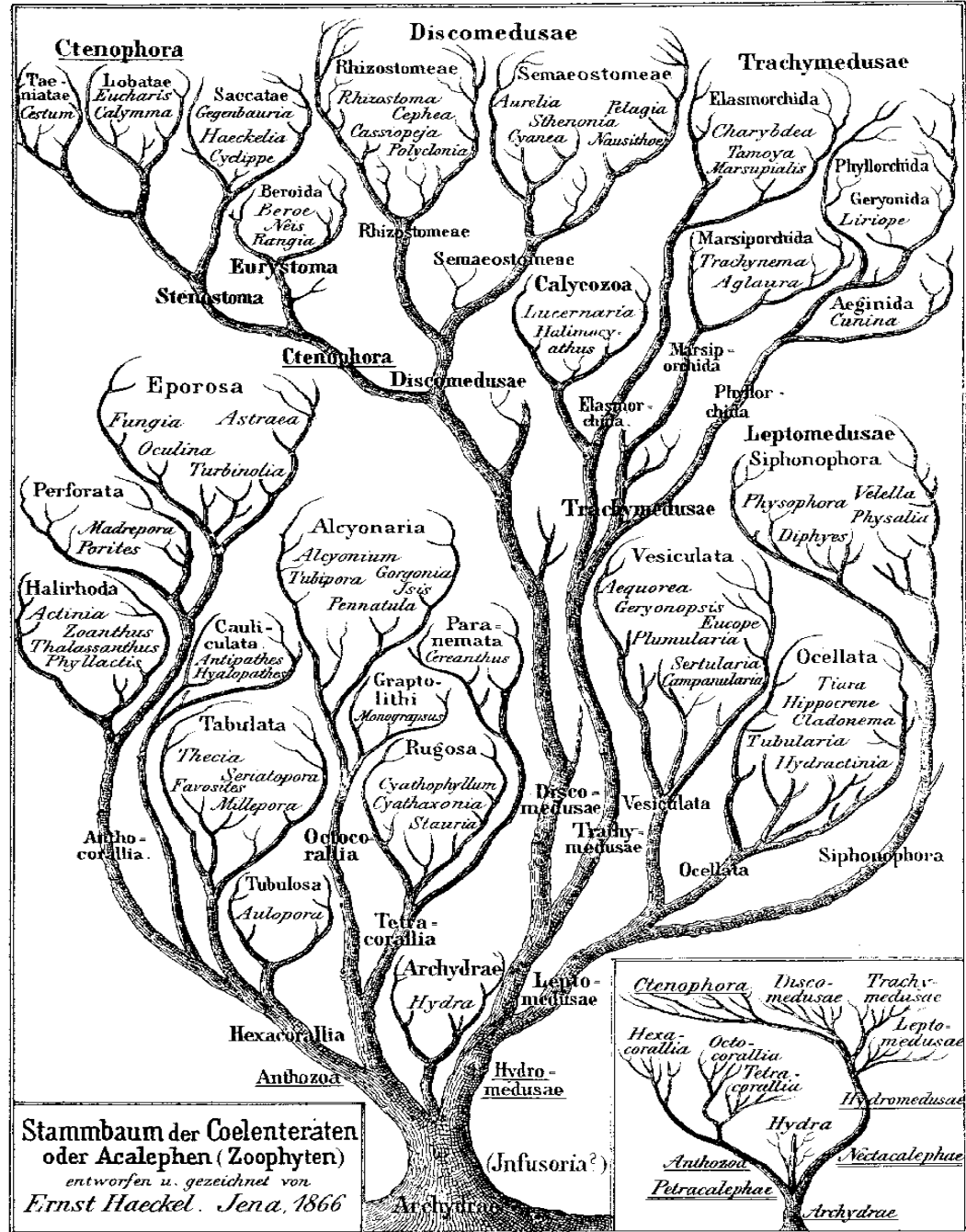
# Phylogenomics

Todd Lowe

BME 110

May 24, 2008

# Haeckel (1866)



Stammbaum der Coelenteraten  
 oder Acalephen (Zoophyten)  
 entworfen u. gezeichnet von  
 Ernst Haeckel. Jena, 1866

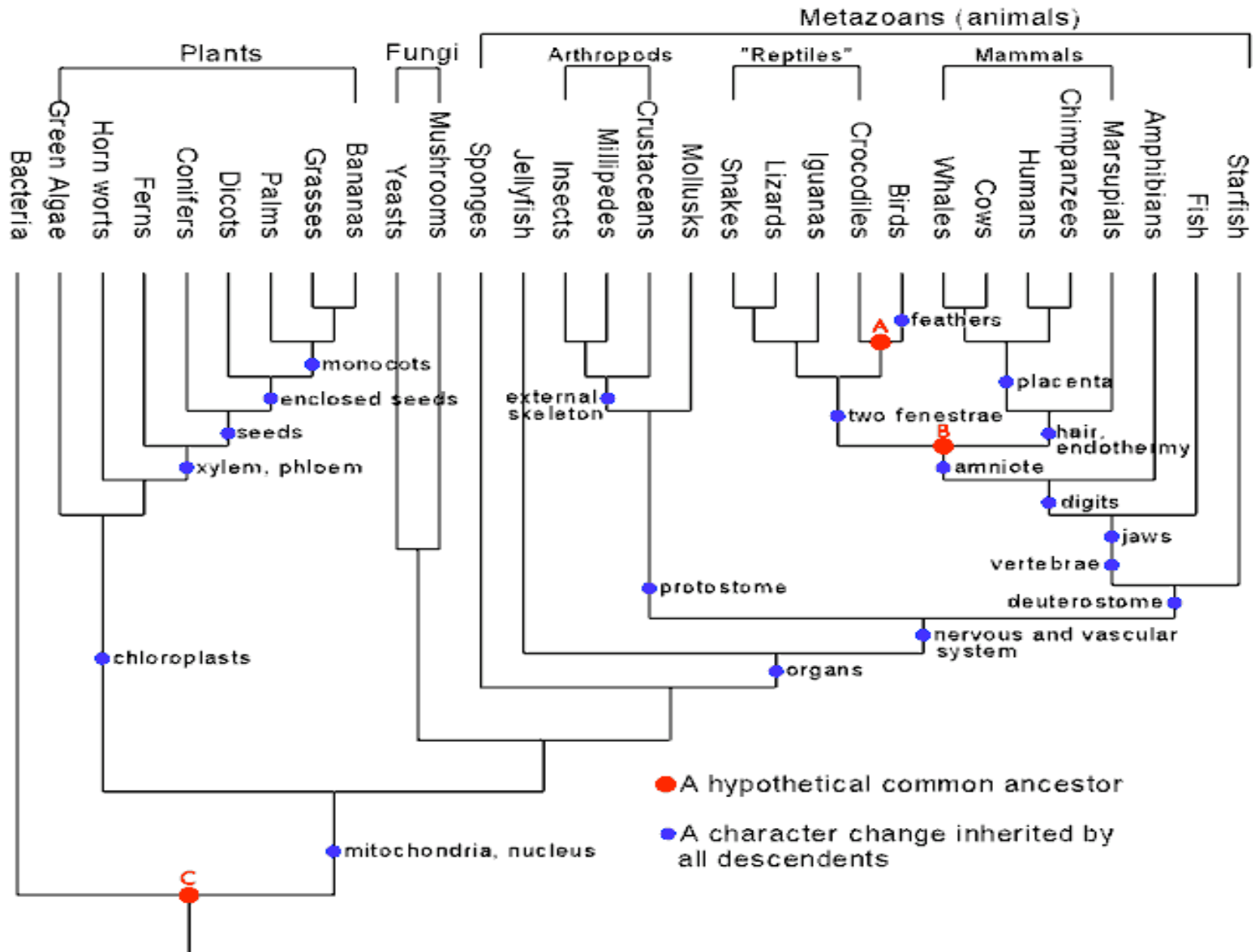
# Phylogenetic Analysis

- Many different characteristics used to determine evolutionary relatedness
  - Morphological
  - Molecular
- Molecular basis can give greatest detail, but complexity of data makes interpretation tricky

# Phylogenetic Trees

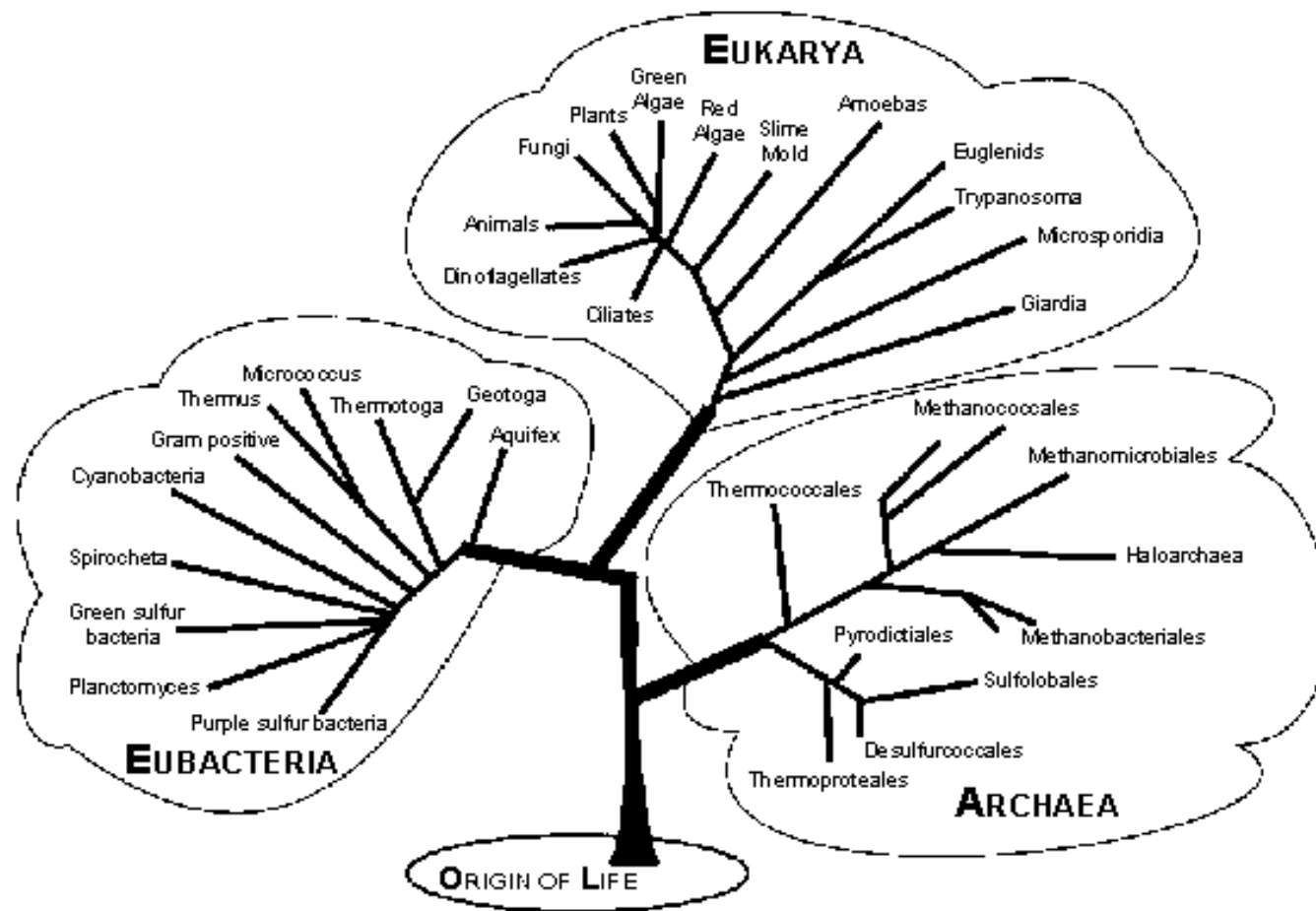
- Traditionally binary
- Can describe
  1. Species
  2. Gene families – to infer function
- May or may not have a “root”
- Length of branches generally indicates evolutionary distance

# Morphology Characters



# Modern Tree of Life

(Molecular characters – ribosomal RNA)



# Molecular Characters

Q9VFS1\_DROME/9-100  
Q9U7C6\_GLOMR/6-100  
Q9VFS0\_DROME/5-98  
Q9U7C5\_GLOMR/5-97  
Q5ZJV8\_CHICK/10-98  
Q6GNZ3\_XENLA/6-94  
CECR1\_BRARE/10-95  
CECR1\_HUMAN/12-102  
CECR1\_PIG/7-97  
Q95WT8\_CULQU/5-95  
Q9NC65\_LUTLO/10-105  
Q9VVK5\_DROME/10-139  
Q26642\_SARPE/10-128  
Q5MIX2\_AEDAL/6-117  
Q8T9T6\_AEDAE/12-123  
Q9VVK4\_DROME/10-113  
Q9V741\_DROME/5-125  
AGSA\_APLCA/12-107  
Q8IQR3\_DROME/54-152  
Q86I77\_DICDI/12-134

SDEEKANSILMNAKR. .AEIAEGLKTPEKYAPAMHFFQGRQYVRQ. SEVFERMI  
GDEEKADQIFLRLKQ. .KELQDGFENVEKNLPGMHYYKAKPLLDE. SKVYQLI  
TEEEKVDDIFMEYKL. .GELAQGFRNAEQNAAALHFFKAKPLIDR. SAIFRFL  
EKEKKVNRIFMRHKH. .NELSVAFNDSSQNAPAMHFFKAKHIIEN. SMVFKFL  
TQEQQLSMKLISLKK. .KEVAAALNT.GQFPPSMHFFRAKSLIEQ. SAVFSIL  
PSEATANQKLMTVKG. .AEFKEAEST.GLFPPSMHFFKARPLIQQ. SHVFSIL  
EREKLLDGKQLKQ. .HDMEAG. . . . .QFPPSMHFFKAKRLIDQ. SPVFNLI  
TKEELANERLMTLKI. .AEMKEAMRT.LIFPPSMHFFQAKHLIER. SQVFNIL  
EEEERANGRLQALKE. .AEMQEAKRT.GIFPPSLHFFQAKGLMEK. SAVFNIL  
EREAQANDIVTRLRS. .KILLEGIANSTGFAPAMHFFQAKPLIES. SPIFRML  
AKEQVVNERLMELKM. .TELKNGLODPAGFIPWNHIFDVLYRINS. SELFHII  
SREIKANETIMKAKL. .KEFDEGLVTPHLFKPSQHIFDVLDGIRN. TDLEKLL  
QKELQANQLIMEAKT. .REYEEGLATPHLFTPSQHLFEVLDDIKQ. SPLFKYI  
ENEQLVNKFIMQMKL. .DEMEKGFNDSYNFIPARHIFEVLDRFGQ. SKVFENVI  
ENEQLVNKFLMRLKL. .EEMVKGFNDSYNFIPARHIFEVLDRFGQ. SKVFKVI  
DRELAANKVIMAVKK. .EEISKGILDPSQFTPGQHIFKTLAEVQK. SPIFNLI  
LPERAANELIMCEKK. .REYKEGIMDPSKFAPGGHIFQVLTKVKK. SPLFQIL  
NEEKVLDDFILAEKR. .KLIDDSRLNQTTEYMPAASFYRSKDFIDT. TFAYKII  
PLEEEANDRLMAIRQVDEEFYNLWRNYHSQPPP. .FLKHLNIID. .TNLYAAL  
SKEQQANTIFLNILO. . . . .NEELSFSNNDPSGVNFFIEKQIIENESTIFKII

# Building a Tree in Jalview

- Class demo starting with reverse gyrase alignment

# Assumptions in Molecular Phylogenetics

- Constant molecular clock (often, probably not true)
- Genetic independence of characters
  - Counter example: RNA secondary structure!!
- No “back” mutations
- Correct alignment!!

# Tree-Building Algorithms

- Neighbor joining (NJ)
- Maximum parsimony (MP)
- Maximum likelihood (ML)

Each has strengths and weaknesses, and often multiple methods are used to find common agreement

# Species Trees

- Want to use the appropriate gradation of clock
- For full Tree of life, very few genes change slowly enough
- Ribosomal RNA is a favorite

# 16S rRNA Partial Alignment

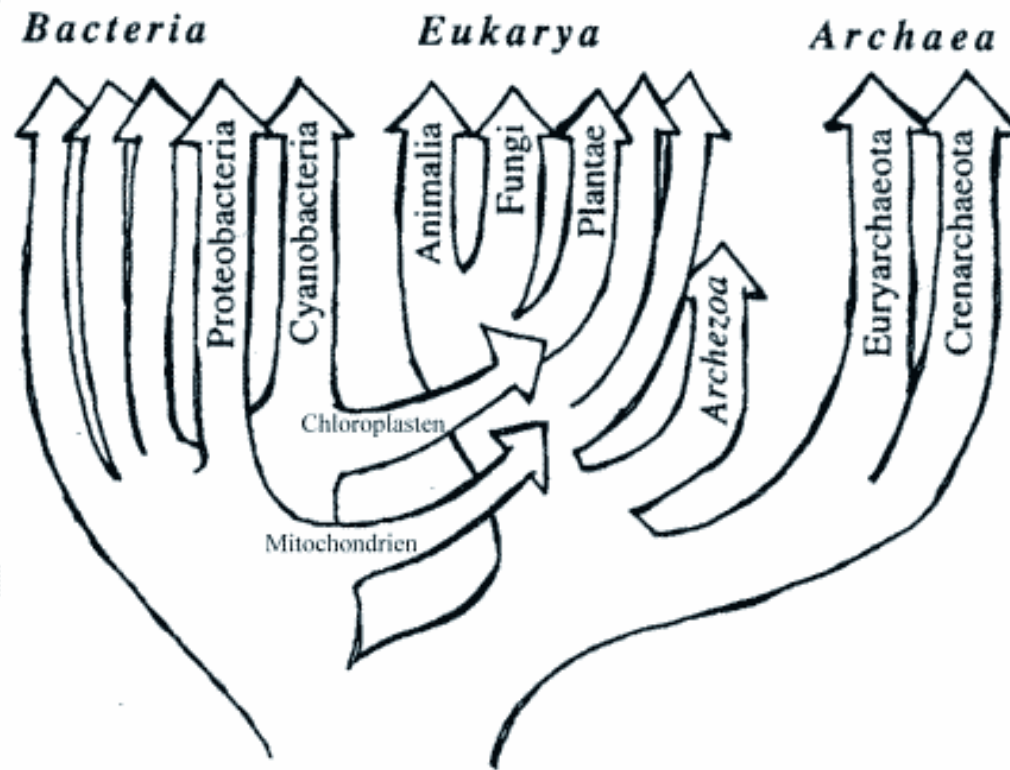
**A)**

<i>Salmonella typhimurium</i> X80681	TTATCCTTTGTTGCCAGCG-ATTAGGTCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGAT
<i>Legionella pneumophila</i> M36023	TNGTCCTTAGTTGCCAGCATGTGATGGTGGGGACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGCCGGGGAT
<i>Escherichia coli</i> AE000406	TTATCCTTTGTTGCCAGCG-GTCCGGCCGGGAACTCAAAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGAT
PCR product (human ganglion)	CTATTGATAGTTACCATC--ATTAAGTTGGGTACTCTATTAAGACTGCCGCTGACAAGGCCGAGGAAGGTGGGGAC
<i>Francisella novicida</i> L26084	CTATTGATAGTTACCATC--ATTAAGTTGGGTACTCTATTGAGACTGCCGCTGACAAGGCCGAGGAAGGTGGGGAC
<i>F. tularensis</i> palearctica L26086	CTATTGATAGTTACCATC--ATTAAGTTGGGTACTCTATTAAGACTGCCGCTGACAAGGCCGAGGAAGGTGGGGAC
<i>F. tularensis</i> tularensis Z21932	CTATTGATAGTTACCATC--ATTAAGTTGGGTACTCTATTGAGACTGCCGCTGACAAGGCCGAGGAAGGTGGGGAC
<i>F. philomiragia</i> Z21933	CTATTGATAGTTACCATC--ATTAAGTTGGGTACTCTATTGAGACTGCCGCTGACAAGGCCGAGGAAGGTGGGGAC
	*    *** *    *    *    *    *    *    **    *    **    *

**B)**

Position	145	162	1273	1290
<i>E. coli</i>	TAACTACTGGAAACGGTA.....	AAGCGGACCTCATAAAGT		
<i>S. typhimurium</i>	TAACTACTGGAAACGGTG.....	AAGCGGACCTCATAAAGT		
<i>L. pneumophila</i>	CAACTTGGGGAAACTNAA.....	GAGCAAATCCTNAAAAGT		
<i>Francisella</i> spp.	TACCAGTTGGAAACGACT.....	GAGCGAAACTCAAAAAGG		
	*    ***	***	*	*
Primer	F11		F5	

# Beware: Horizontal Transfer!



# Gene Trees – Need for Species Trees!

- Interpretable phylogenetic trees generally can be constructed only within sets of orthologs
- A complete list of orthologs also is a prerequisite for any meaningful comparison of genome organization

# Concept Review: Homology

- Even if two sequences are ancestrally derived from same molecule, they may or may not still have the same function
  - Orthologs: homologous genes created by speciation
    - Generally implies function remains the same
  - Paralogs: homologous genes created by a gene duplication event (in same species)
    - Implies function may have changed

# Methods of Comparison

- Must establish orthologous groups to allow genome comparisons
- How do we do this?
- Least to most accurate methods
  - Simple: BLAST – “best hit”
  - BLAST – reciprocal best hit
  - Gene trees

# Non-orthologous gene replacement

- replacement of a gene with same function but independent evolutionary lineage
- Example: two lysyl tRNA-synthetases
  - class II found in most euks & bacteria
  - class I found in archaea & some bacteria

# Basic Procedure

1. BLAST the amino acid translation of the gene.
2. Find bona fide functional proteins found in the BLAST results.
3. Align sequences (clustal).
4. Infer phylogeny (NJ/MP/ML)
5. Make nice looking dendrogram/tree

# Demonstration

- Do alignment using ClustalW
- Upload file
- Do alignment of the unknown gene
- Inspect
- Use the JalView program online to see results and get a dendogram.
- Also use [Phylodendron](#) online program.
  - Choose Phylogram and GIF image.