

Phytopathology in the News

1

Complex factors can drive the emergence and evolution of plant pathogens

NC State source

Researchers have found that the introduction of cucumber crops to North America likely led to the rise of bacterial wilt, which affects zucchini, pumpkin, cucumbers, muskmelon and squash -- costing US farmers tens of millions of dollars each year.

Erwinia tracheiphila is a virulent phytopathogen that infects two genera of cucurbit crop plants, *Cucurbita* spp. (pumpkin and squash) and *Cucumis* spp. (muskmelon and cucumber). One of the unusual ecological traits of this pathogen is that it is limited to temperate eastern North America. Here, we complete the first large-scale sequencing of an *E. tracheiphila* isolate collection. From phylogenomic, comparative genomic, and empirical analyses, we find that introduced *Cucumis* spp. crop plants are driving the diversification of *E. tracheiphila* into multiple lineages. Together, the results from this study show that locally unique biotic (plant population) and abiotic (climate) conditions can drive the evolutionary trajectories of locally endemic pathogens in unexpected ways.

From the news story: “Complex factors can drive the emergence and evolution of plant pathogens,” October 3, 2018.

<https://www.sciencedaily.com/releases/2018/10/181009102511.htm>

Original scientific paper: Lori R. Shapiro, Joseph N. Paulson, Brian J. Arnold, Erin D. Scully, Olga Zhaxybayeva, Naomi E. Pierce, Jorge Rocha, Vanja Klepac-Ceraj, Kristina Holton, Roberto Kolter. **An Introduced Crop Plant Is Driving Diversification of the Virulent Bacterial Pathogen *Erwinia tracheiphila*.** *mBio*, 2018; 9 (5) DOI: [10.1128/mBio.01307-18](https://doi.org/10.1128/mBio.01307-18)

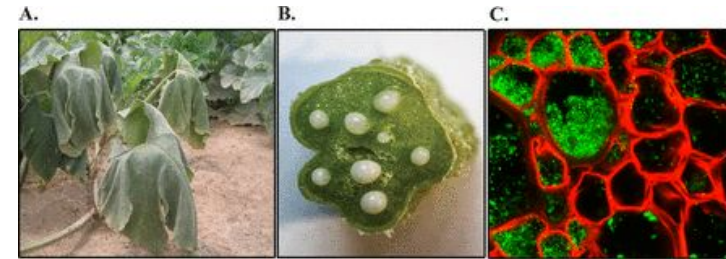
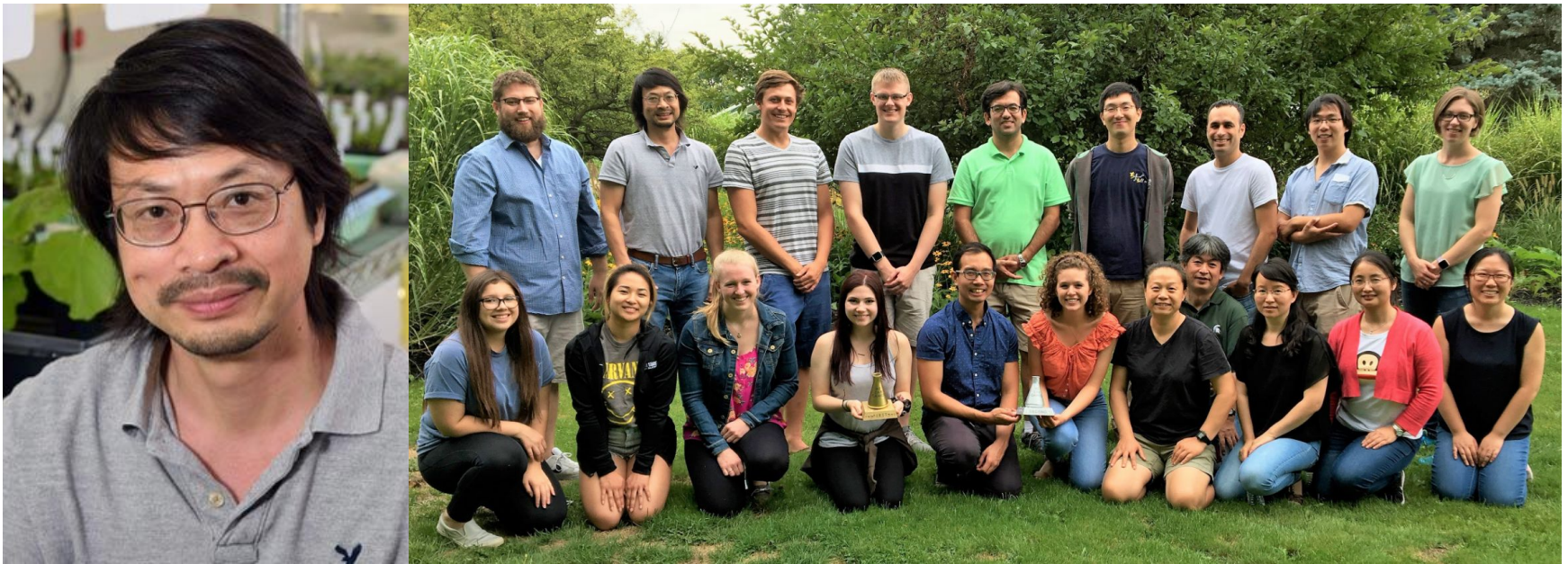


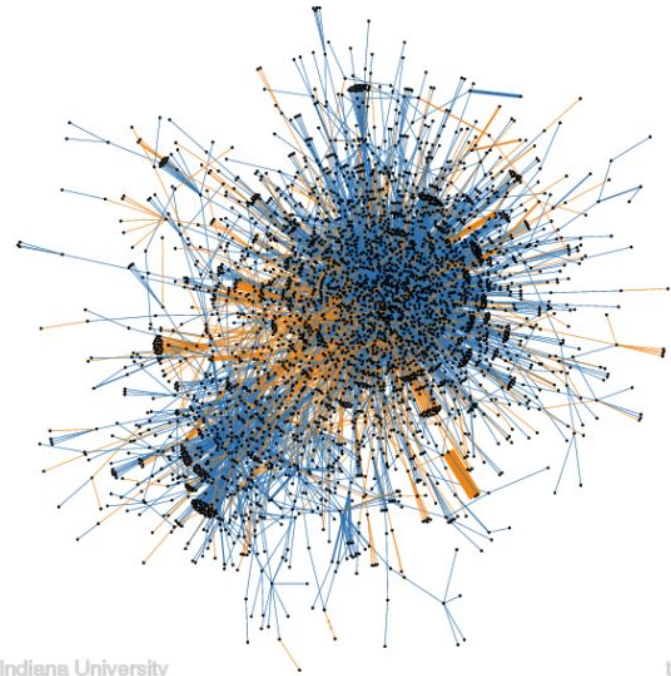
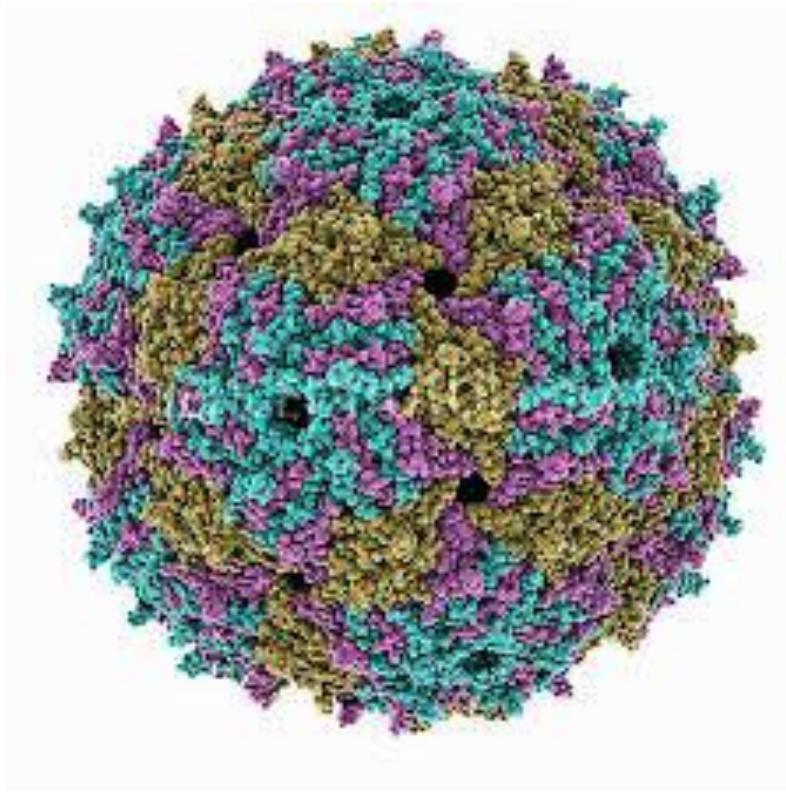
FIG 1

Erwinia tracheiphila infection at the macroscopic and microscopic levels. (A) A vine of a field-infected *Cucurbita pepo* plant shows characteristic systemic wilting symptoms. (B) *E. tracheiphila* can be seen oozing from multiple blocked xylem vessels in a cross-section of a symptomatic cucumber stem. (C) *In planta* confocal microscopy image of *E. tracheiphila* (green) blocking the xylem (red) of a wilting squash plant.

Announcements

- Phytobacteriology in the News! Week 7 [LINK](#)
- Reviews will be returned by tonight!
- Reading for next week (Sheng-Yang He!) [LINK](#)





Copyright 2010 Indiana University

truthy.indiana.edu

Guest lecturer: Dr. Richard Allen White III
Post-doc: Friesen lab

Learning Objectives

1. Phylogenetics review
2. Sequence alignment review
3. Multiple sequence alignment
4. Multiple sequence alignment tools
5. Build a quick ML tree (iqtree)
6. Quick tree building in BLAST

Types of trees

Review

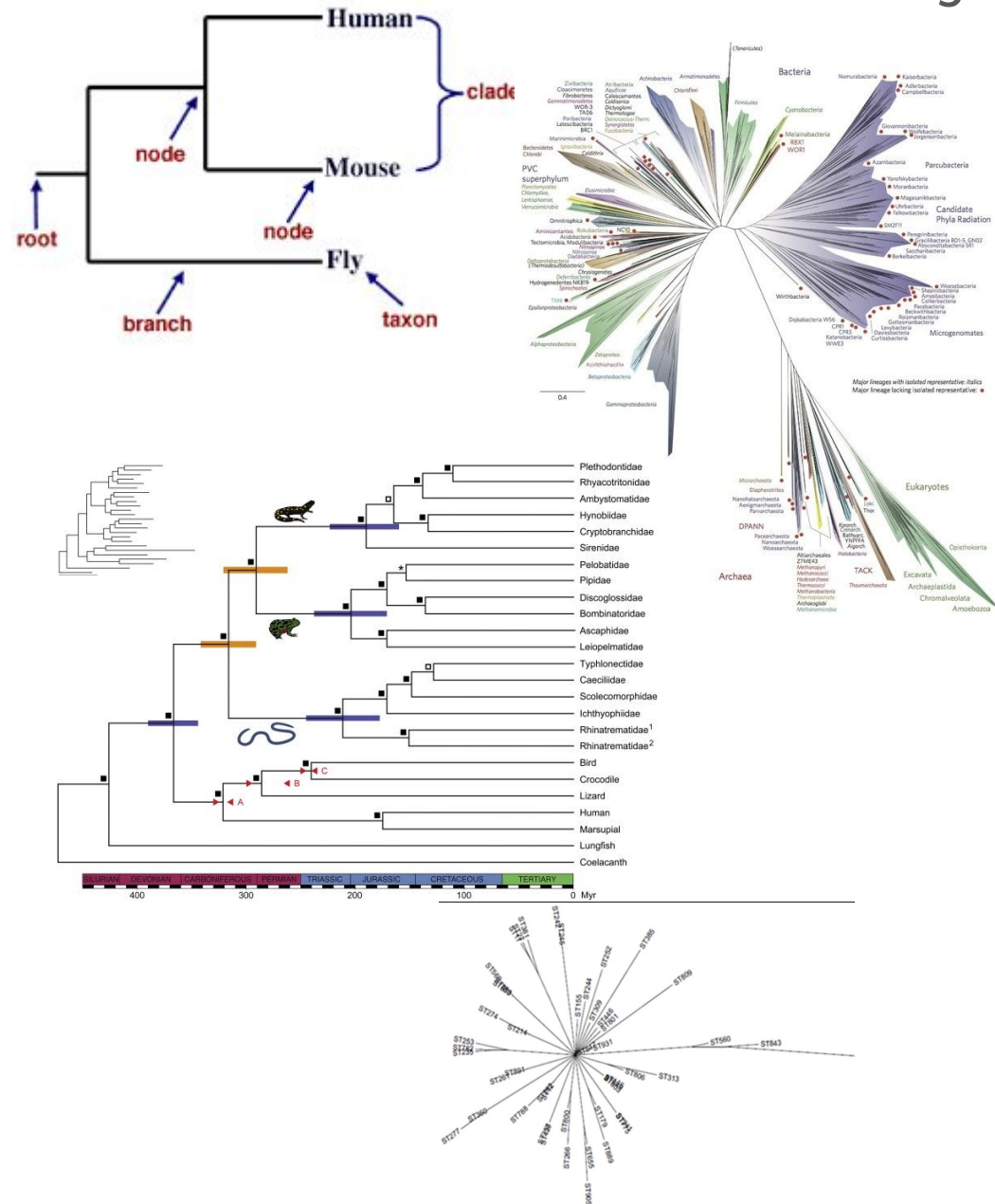
Rooted: ?

Unrooted: ?

Time-calibrated: ?

Star phylogeny: ?

Outgroup: ?



Types of trees

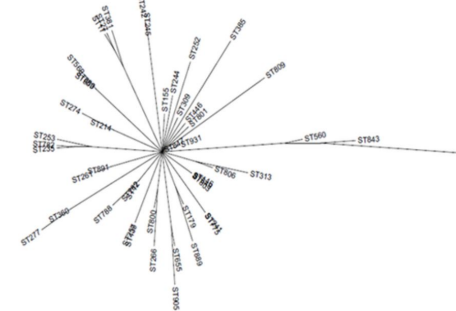
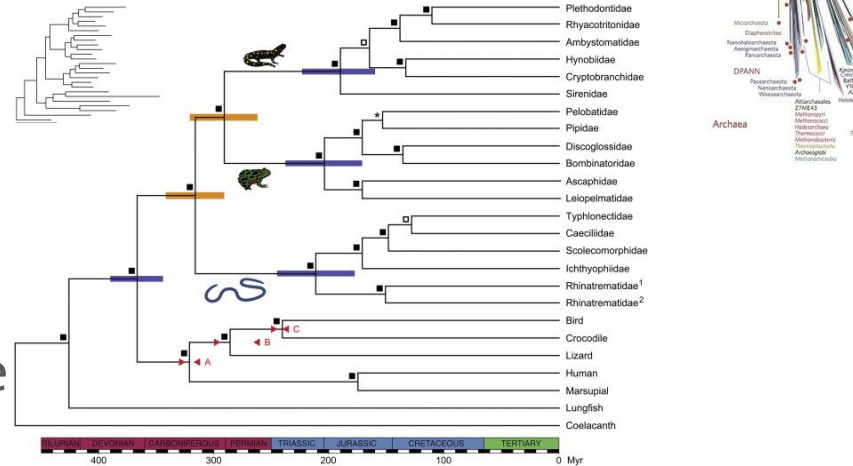
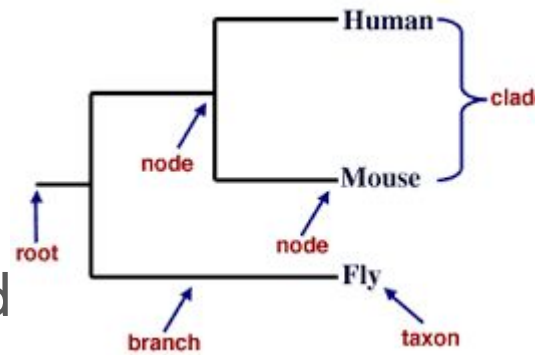
Review

Rooted: the ancestor is depicted and time is oriented on the tree

Unrooted: no ancestor is depicted and thus we don't know which internal nodes are older

Time-calibrated: using the molecular clock or fossils, actual time estimates are included in the tree

Star phylogeny: a phylogeny where all organisms are equally related to one another (e.g., DOI: 10.1371/journal.pone.0044199)



Sequence alignment

Review

Global ?

Example algorithm?

Uses?

Example tool?

Local?

Example algorithm?

Uses?

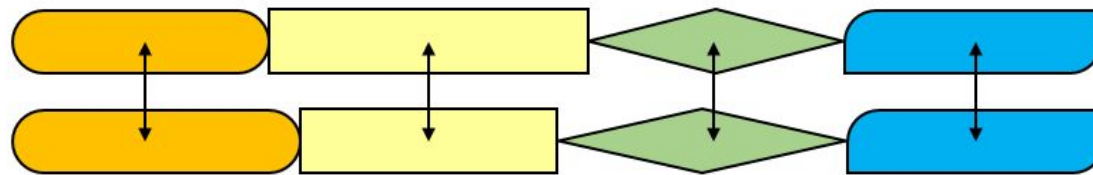
Example tool?

Sequence alignment

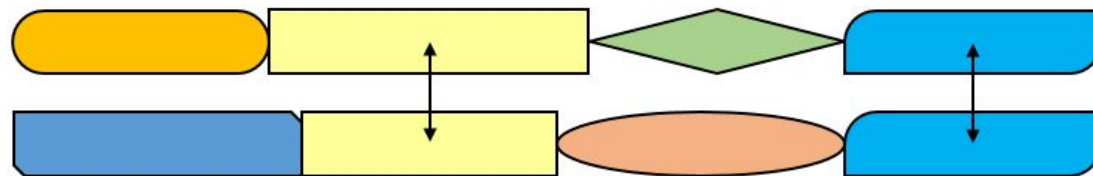
Review

Global: Needleman–Wunsch algorithm, SWORD

Local: Smith-Waterman algorithm, BLAST



Global Alignment



Local Alignment

Multiple sequence alignments

Is an extension of pairwise alignment of 2 or more sequences either protein, DNA or RNA.

Methods for this:

- 1) Progressive alignment
- 2) Iterative methods
- 3) Consensus methods
- 4) Hidden Markov models

Multiple sequence alignments

Download these sequences from dropbox

Link -

<https://www.dropbox.com/sh/nml3oagxo45rpcy/AADE4lHwk9syQSO4oRAdxDKZa?dl=0>

16S_nucleotide.fasta

RecA_protein.fasta

Using unix tell me:

Tell me the number of sequences in each file?

Tell me the number of lines in each file?

Tell me the size of each file?

SUPER BONUS: Create a two new files that lists the headers!

Multiple sequence alignments

11


- 1) **Progressive alignment** - heuristic search known as progressive technique by using clustering (hierarchical or tree method), most widely used, **Example: Clustal (most common)**
- 2) **Iterative methods** - an iterative progressive alignment. IM return to calculated pairwise alignments or initial sub-MSAs incorporating subsets of the query sequence then optimize using objective function finding the highest alignment score, **Example: MUSCLE**
- 3) **Consensus methods** - use multiple methods to find the optimal multiple sequence alignment given multiple different alignments of the same set of sequences. **Example: M-coffee/T-coffee**
- 4) **Hidden Markov models** - are probabilistic models that can assign likelihoods to all possible combinations of gaps, matches, and mismatches, **Example - POA (rarely used)**

Multiple sequence alignments

12

1) Complete a Progressive alignment on protein file using ClustalW

Link - https://npsa-prabi.ibcp.fr/cgi-bin/align_clustalw.pl



PRABI-GERLAND
RHONE-ALPES BIOINFORMATIC POLE GERLAND SITE
Institute of Biology and Protein Chemistry

[Home](#) [Services](#) [Teaching](#) [Publications](#) [Links](#) [Jobs](#) [Contact](#)

CLUSTALW

[Abstract](#) [\[NPS@ help\]](#) [\[Original server\]](#)

Paste a protein sequence databank in Pearson/Fasta format below : [help](#)

```
>ACB15385.1 Rhodospirillum rubrum  
NVEVPTGSLALDMLGIGGLPKGRIVEIYGPSSGKTTIALHVANAQKKGGVAAYI  
DAEHALDPAYAR  
KLGVDTDSIYVSQPDNGEALFIADMLURSGALDVVIDSVAAVPKAEIEGEMGDS  
HYGLQARLMSSQAL  
RKMTGALAAQAGTTAFINQLREKIGVFFGNPETTTGGKALKEYASVRLDIRRIQTLKN  
GDEAVGNRTIRVK  
VVKNMAPPFKSAEFDMLYGEGISREGSVIDMAQQVGVVKKSGSWETIEGDQLG  
QGR
```

All sequence names must be different !

Output width :

CLUSTALW Parameters

Output format :

Output order :

Pairwise alignment type :

Fast pairwise alignment parameters	Slow pairwise alignment parameters
K-tuple (word) size : <input type="button" value="1"/>	Protein weight matrix : <input type="button" value="GONNET"/>
Number of top diagonals : <input type="button" value="5"/>	Gap opening penalty : <input type="text" value="10.0"/>
Window size : <input type="text" value="5"/>	Gap extension penalty : <input type="text" value="0.1"/>
Gap penalty : <input type="text" value="3"/>	
Scoring method : <input type="button" value="Percentage"/>	

Multiple Alignment Parameters :

Weight matrix :

Link - https://npsa-prabi.ibcp.fr/cgi-bin/align_clustalw.pl

Multiple sequence alignments

14

- 2) Complete a Iterative alignment on protein and nucleotide file using MUSCLE
Link - <https://www.ebi.ac.uk/Tools/msa/muscle/>

The screenshot displays the MUSCLE web interface. At the top, there is a navigation bar with links for EMBL-EBI, Services, Research, Training, Industry, and About us. Below this, the word "MUSCLE" is prominently displayed. A secondary navigation bar includes links for Input form, Web services, Help & Documentation, and Bioinformatics Tools FAQ. The main content area is titled "Multiple Sequence Alignment" and provides a brief description of the tool's purpose: "MUSCLE stands for Multiple Sequence Comparison by Log-Expectation. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options." An important note states: "Important note: This tool can align up to 500 sequences or a maximum file size of 1 MB." The interface is divided into two main steps. Step 1, "Enter your input sequences", features a text area where a set of sequences has been pasted. The sequences include a header "D A E F A F U P I T A K K L G V K V E E L T L S Q P U I G E Q A L E I C D I L V R S G G V D M V V Y D S V A A L Y P K A E I E G E M G D S H" followed by several lines of amino acid sequences. Below the text area, there is a button for "Browse..." and a message "No file selected." Step 2, "Set your Parameters", shows a dropdown menu for "OUTPUT FORMAT:" with "ClustalW" selected. At the bottom of the page, a cookie consent banner is visible, stating "This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our Privacy Notice and Terms of Use." and a link to "I agree, dismiss this banner".

EMBL-EBI Services Research Training Industry About us

EMBL-EBI Hinxton

MUSCLE

Input form Web services Help & Documentation Bioinformatics Tools FAQ

Feedback Share

Tools > Multiple Sequence Alignment > MUSCLE

Multiple Sequence Alignment

MUSCLE stands for **M**ultiple **S**equences **C**omparison by **L**og-**E**xpectation. MUSCLE is claimed to achieve both better average accuracy and better speed than [ClustalW2](#) or [T-Coffee](#), depending on the chosen options.

Important note: This tool can align up to 500 sequences or a maximum file size of 1 MB.

STEP 1 - Enter your input sequences

Enter or paste a set of sequences in any supported format:

```
D A E F A F U P I T A K K L G V K V E E L T L S Q P U I G E Q A L E I C D I L V R S G G V D M V V Y D S V A A L Y P K A E I E G E M G D S H
V G L Q A R L M S Q A L R K L T G H I K R T N T L V V F I N Q I R M K I G V V Y G S P E T T T G G N A L K F Y A S V R L D I R R A G Q I K K
G D E I L G N E T K V K V I K N K V A P P F R Q A E F D I L Y G E G I S W E G E L I D I G V K H D I I D K S G A W Y S Y N D T K
>ACB15385.1 Bifidobacterium longum
N V E V I P T G S L A L D M A L G I G G L P K G R I V E I Y G P E S S G K T T L A L H V V A N A Q K G G V A A Y I D A E H A L D P A Y A R
K L G V D T D S L I V S Q P D N G E Q A L E I A D M L I R S G A L D V I V I D S V A A L Y P K A E I E G E M G D S H V G L Q A R L M S Q A L
R K M T G A L A Q A G T T A I F I N Q L R E K I G V F F G N P E T T T G G K A L K F Y A S V R L D I R R I Q T L K N G D E A V G N R T R Y K
V V K N K M A P P F K S A E F D M L Y G E G I S R E G S V I D M A Q Q V G V K K S G S W F T Y E G D Q L G Q G R
```

Or upload a file: No file selected. [Use a example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

STEP 2 - Set your Parameters

OUTPUT FORMAT:

ClustalW

This website requires cookies, and the limited processing of your personal data in order to function. By using the site you are agreeing to this as outlined in our [Privacy Notice](#) and [Terms of Use](#). [I agree, dismiss this banner](#)

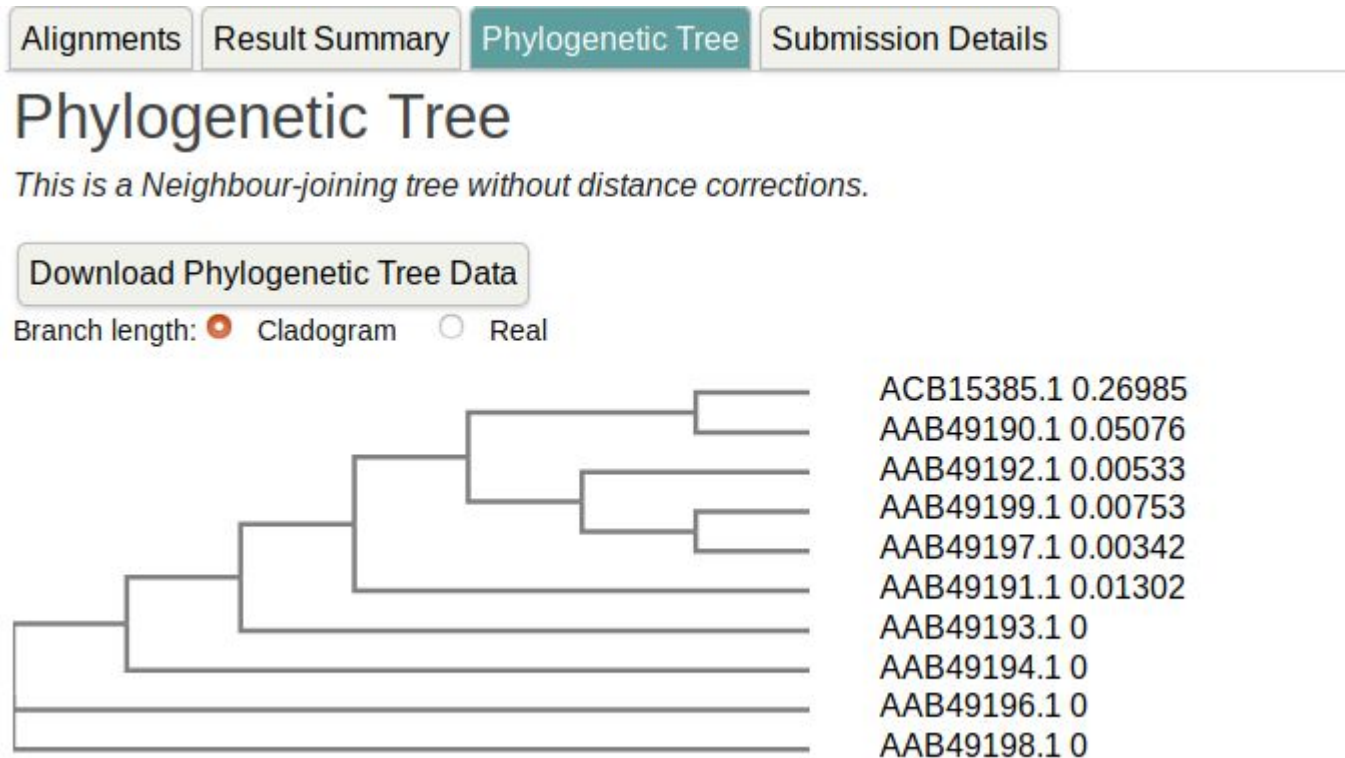
Link - <https://www.ebi.ac.uk/Tools/msa/muscle/>

Multiple sequence alignments

16

- 2) Complete a Iterative alignment on protein file using MUSCLE
Click on simple phylogeny (Tree!!)

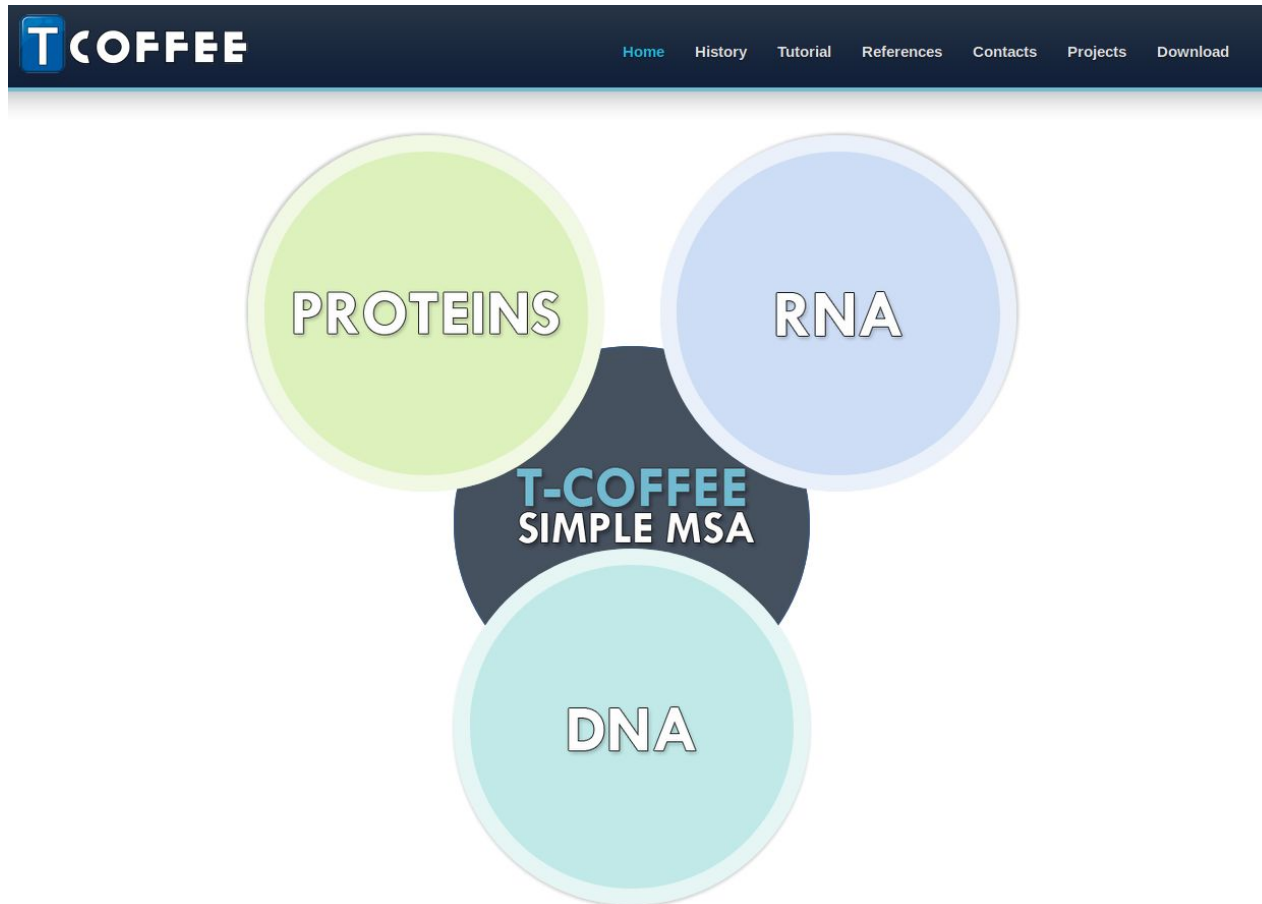
Link - <https://www.ebi.ac.uk/Tools/msa/muscle/>



Multiple sequence alignments

17

- 3) Complete a Consensus alignment on protein or nucleotide file using T-coffee
Link - <http://tcoffee.crg.cat/>



Multiple sequence alignments

18

- 3) Complete a Consensus alignment on protein or nucleotide file using T-coffee
Link - <http://tcoffee.crg.cat/>

TCOFFEE

HomeHistoryTutorialReferencesContactsProjectsDownload

M-Coffee

Aligns DNA, RNA or Proteins by combining the output of popular aligners

Sequences input

Paste or upload your set of sequences in FASTA format

Sequences to align

[Click here to use the sample file](#)

>ACB15385.1 Bifidobacterium longum
NVEVIPTGSLDMLGIGGLPKGRIVEIYGPESSGKTTLALHVVANAQKGGVAAYIDAEHALDPAYAR
KLGVDTSLIVSQPDNGEQALEIADMLIRSGALDIVIDSVAALVPKAEIEGEMGDHVLQARLMSQAL
RKMTGALAAGTTAIFINQLREKIGVFFGNPETTTGGKALKFYASVRLDIRRIQTLKNGDEAVGNRTRVK
VVKNMAPPFKSAEFDMLYGEISREGSVIDMAQQVGVKKSGSWFTYEGDQLGQGR

- OR - [Click here to upload a file](#)

[Show more options](#)

Your email address

SubmitReset

Multiple sequence alignments

19

3) Complete a Consensus alignment on protein or nucleotide file using T-coffee

Link - <http://tcoffee.crg.cat/>

```
T-COFFEE, Version 11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)
Cedric Notredame
SCORE=999
*
  BAD AVG GOOD
*
AAB49191.1 : 99
AAB49192.1 : 99
AAB49193.1 : 99
AAB49194.1 : 99
AAB49196.1 : 99
AAB49197.1 : 99
AAB49198.1 : 99
AAB49199.1 : 99
AAB49190.1 : 100
ACB15385.1 : 99
cons      : 99

AAB49191.1 SIMKMDGS00EENLEVISTGSLGDLALGVGGLPRGRIVEIFGPESSGKTTLCLEAVAQCKKGGVCAF
AAB49192.1 SIMKMDGS00EENLDVISTGSLGVDLALGVGGLPRGRVVEIFGPESSGKTTLCLEAIAQCKNGGICAF
AAB49193.1 AIMKMDGS00EENLEVISTGSLGDLALGVGGLPRGRIVEIFGPESSGKTTLCLEAVAQCKNGGVCFA
AAB49194.1 AIMKMDGS00EENLEVISTGSLGDLALGVGGLPRGRIVEIFGPESSGKTTLCLEAVAQCKNGGVCFA
AAB49196.1 AIMKMDGS00EENLEVISTGSLGDLALGVGGLPRGRIVEIFGPESSGKTTLCLEAVAQCKNGGVCFA
AAB49197.1 SIMKMDGS00EENLDVISTGSLGVDLALGVGGLPRGRVVEIFGPESSGKTTLCLEAIAQCKNGGICAF
AAB49198.1 AIMKMDGS00EENLEVISTGSLGDLALGVGGLPRGRIVEIFGPESSGKTTLCLEAVAQCKNGGVCFA
AAB49199.1 SIMKMDGS00EENLDVISTGSLGVDLALGVGGLPRGRVVEIFGPESSGKTTLCLEAIAQCKNGGICAF
AAB49190.1 SIINMDGSHKQEDLEVISTGSLGDLALGVGGLPRGRVVEIFGPESSGKTTLCLEAIAQCKNGGVCFA
ACB15385.1 -----NVEIPTGSLALDMALGIGGLPKGRIVEIYGPESSGKTTLALHVANAQKKGGVAAV

cons      :.:**.*.***.:*.*.***:***:***:*****.*.:*.:**.*.:*.:

AAB49191.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49192.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49193.1 VDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49194.1 VDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49196.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49197.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49198.1 VDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49199.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
AAB49190.1 IDAEHAFDPVYARKLGKVEEELYSOPDTGEQALEICDTLVRSGGDMVVVDSVAALVPKAEIEGDMGD
ACB15385.1 IDAEHALDPAYARKLGVDTSLVSPDNQGEQALEIADMLIRSGALDVIIVDSVAALVPKAEIEGDMGD

cons      :*****:** *****.:*.*.***:*****.*.*.***:*.*:*****:*****:***

AAB49191.1 SHVGLHARLMSQALRKLTGHIKKTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYSSVRLDIRRTGQ
AAB49192.1 SHVGLQARLMSQALRKLTGHIKRTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYASVRLDIRRTGQ
AAB49193.1 SHVGLQARLMSQALRKLTGHIKKTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYSSVRLDIRRTGS
AAB49194.1 SHVGLQARLMSQALRKLTGHIKKTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYSSVRLDIRRTGS
AAB49196.1 SHVGLQARLMSQALRKLTGHIKKTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYSSVRLDIRRTGS
AAB49197.1 SHVGLQARLMSQALRKLTGHIKRTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYASVRLDIRRTGQ
AAB49198.1 SHVGLQARLMSQALRKLTGHIKKTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYSSVRLDIRRTGS
AAB49199.1 SHVGLQARLMSQALRKLTGHIKRTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYASVRLDIRRTGQ
AAB49190.1 SHVGLQARLMSQALRKLTGHIKRTNTLVVFQINQIRMKIGVMFGSPETTTGGNALKFYASVRLDIRRAGQ
ACB15385.1 SHVGLQARLMSQALRKMTGALAQAQGTATFQINQLREKIGVFVGNPETTTGGKALKFYASVRLDIRRIQT

cons      :*****:*****:**.:*.*.***:*****.*.*.***:*.*:*****:*****:*****

AAB49191.1 IKKGEEVIGNETRVKVIKNKVAPPFROAEFDILYEGGVSWEGELIDIGVKNDIINKSGAWYSYNGAK--
AAB49192.1 IKKGDDVIGNETRVKVIKNKVAPPFROAEFDILYEGGISWEGELIDLGKVDIVEKSGAWYSYNGAK--
AAB49193.1 IKKGEEVLGNETRVKVIKNKVAPPFROAEFDILYEGGISWEGELIDIGVKNDIINKSGAWYSYNGAK--
AAB49194.1 IKKGEEVLGNETRVKVIKNKVAPPFROAEFDILYEGGVSWEGELIDIGVKNDIINKSGAWYSYNGAK--
```

Multiple sequence alignments

20

- 3) Complete a Consensus alignment on protein ONLY T-coffee/M-coffee
Save your alignment file!!

Result files
10 output files - [download them all](#)

Input(s)	Input sequences (3KB)
System	Command line (308 B) Log file (196KB)
Tree	dnd file (287 B) ph file (273 B)
Multiple Alignment	score_html file (11KB) clustalw_aln file (3KB) fasta_aln file (3KB) score_ascii file (3KB) phylip file (3KB)

 [Copy to your Dropbox](#)

Send results
Forward this result to other online tools.

Core/TCS	Evaluates your Alignment indicating the local reliability
ProtoGene	Turning amino acid alignments into bona fide CDS nucleotide alignments
MSA hub	MyHits: a new interactive resource for protein annotation and domain identification
JalView	Open this alignment in the Jalview viewer
ESPrnt	ESPrnt server renders sequence similarities and secondary structure information from aligned sequences

Info

Quick ML tree using iq-tree

21

- 3) Complete a upload your clustalw_aln file or download from dropbox link
Link - <http://iqtree.cibiv.univie.ac.at/>

<https://www.dropbox.com/sh/nml3oaqxo45rpcy/AADE4lHwk9syQSO4oRAdxDKZa?dl=0>

The screenshot displays the IQ-TREE web server interface. At the top, the URL iqtree.cibiv.univie.ac.at is shown. Below the header, a progress bar indicates 'Server load: 4%' and a citation for Trifunopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) is provided. The main navigation bar includes 'Tree Inference', 'Model Selection', and 'Analysis Results'. A tutorial link is available for a quick start. The 'Input Data' section contains an 'Alignment file' input field with a 'Browse...' button and a 'Hide example <' button. Below this, four example alignment files are shown in different formats: PHYLIP, CLUSTAL, FASTA, and NEXUS. The 'Use example alignment' checkbox is checked. The 'Sequence type' is set to 'Auto-detect'. The 'Partition file' is optional. The 'Partition type' is set to 'Edge-linked'. The 'Substitution Model Options' section includes a 'Substitution model' dropdown set to 'Auto', 'FreeRate heterogeneity' checkboxes for '+R', '+G', and '+I', and a '#rate categories' dropdown set to '4'. The 'State frequency' is set to 'Empirical (from data)'. The 'Ascertainment bias correction' is set to 'Yes [+ASC]'. The 'Branch Support Analysis' section includes a 'Bootstrap analysis' dropdown set to 'None', a 'Number of bootstrap alignments' input field set to '1000', a 'Create .ufboot file' checkbox, and a 'Maximum iterations' input field set to '1000'.

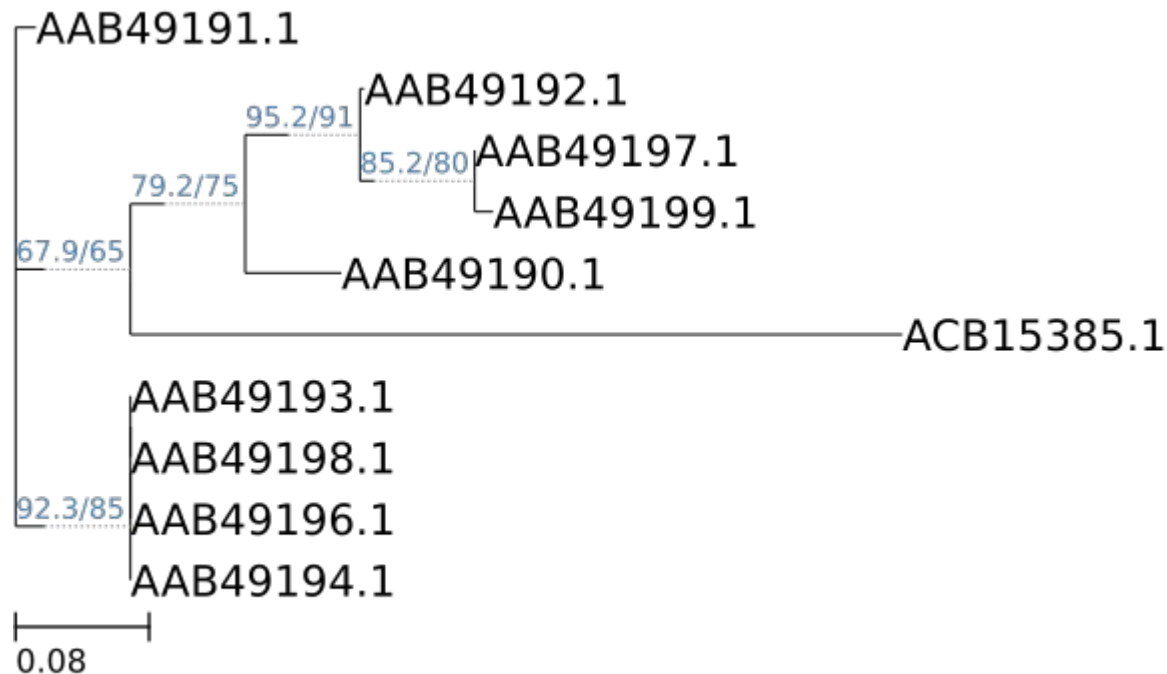
Quick ML tree using iq-tree

22

3) Complete a upload your clustalw_aln file or download from dropbox link

Link - <http://iqtree.cibiv.univie.ac.at/>

<https://www.dropbox.com/sh/nml3oaqxo45rpcy/AADE4lHwk9syQSO4oRAdxDKZa?dl=0>



Build a quick tree in blast

23

With a single sequence from the protein or nucleotide file build a quick tree in blast and save the tree picture.

SUPER BONUS -> build one with multiple sequences