

# Polygenetics Through High Performance Alignment

Dmitry Makarevskiy  
Saeed Keshani

ECE690 SP15

# Overview

- I. MAFFT and Muscle
- II. Alignment MAFFT and MUSCLE
  - 1. Fasta file
  - 2. Submission and Results
  - 3. SeaView
  - 4. Comparison

# I. MAFFT and MUSCLE Alignment

Part 1

# MAFFT Algorithm

- Calculation of the correlation between two amino acid sequences
- We define the correlation  $c(k)$  between two sequences of such vectors as
- $c(k) = c_v(k) + c_p(k)$
- Volume, polarity

# MAFFT Algorithm

- $c_v(k), c_p(k)$
- $V1(n)$  and  $V2(n)$  are the volume component of the  $n$ th site of sequence 1 with the length of  $N$  and that of sequence 2 with the length of  $M$ , respectively
- $c_v(k) \Leftrightarrow V_1^*(m) \cdot V_2(m)$

$$c_v(k) = \sum_{1 \leq n \leq N, 1 \leq n+k \leq M} \hat{v}_1(n) \hat{v}_2(n+k),$$

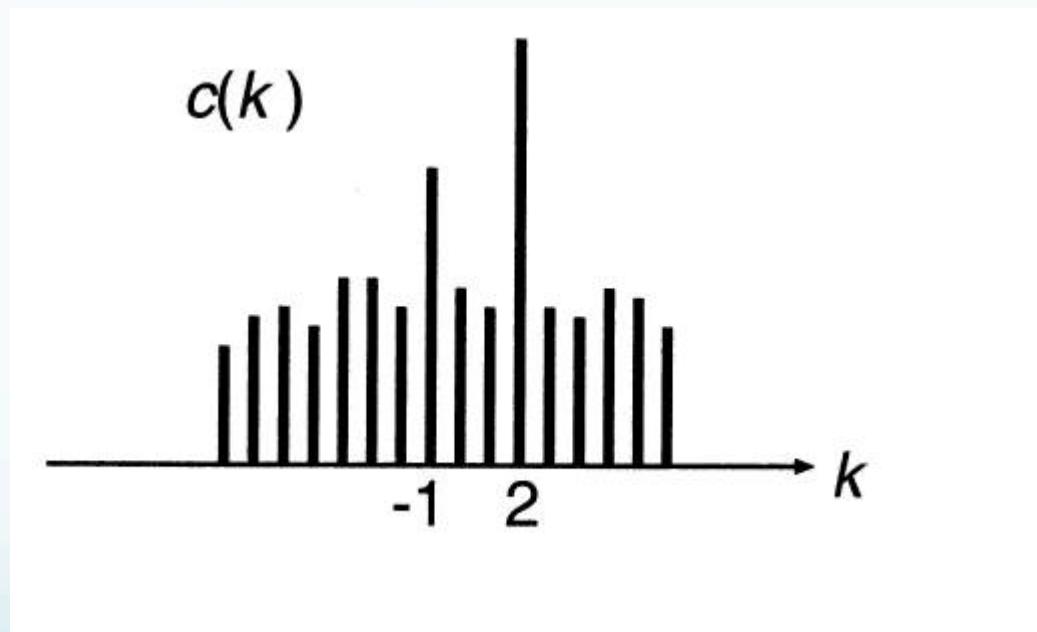
$$c_p(k) = \sum_{1 \leq n \leq N, 1 \leq n+k \leq M} \hat{p}_1(n) \hat{p}_2(n+k)$$

;

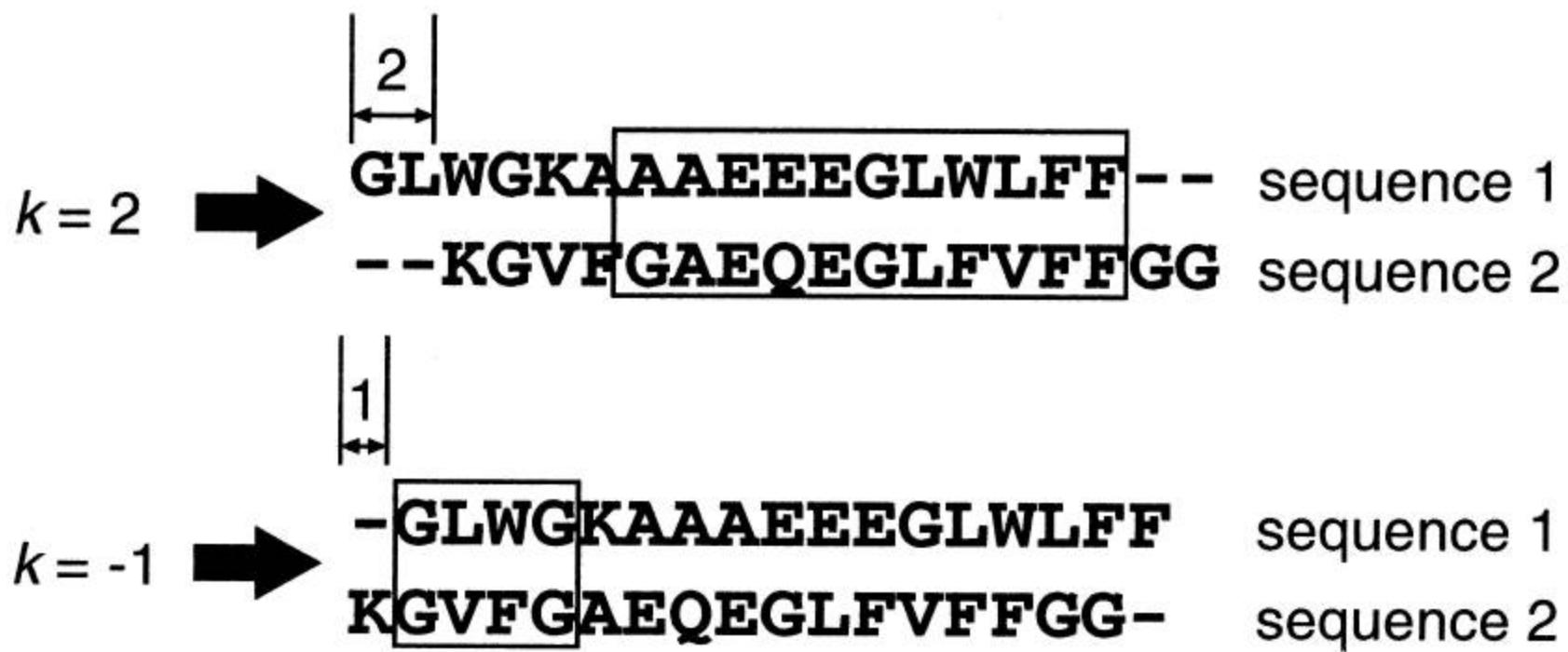
6

# MAFFT Algorithm

- Finding homologous segments

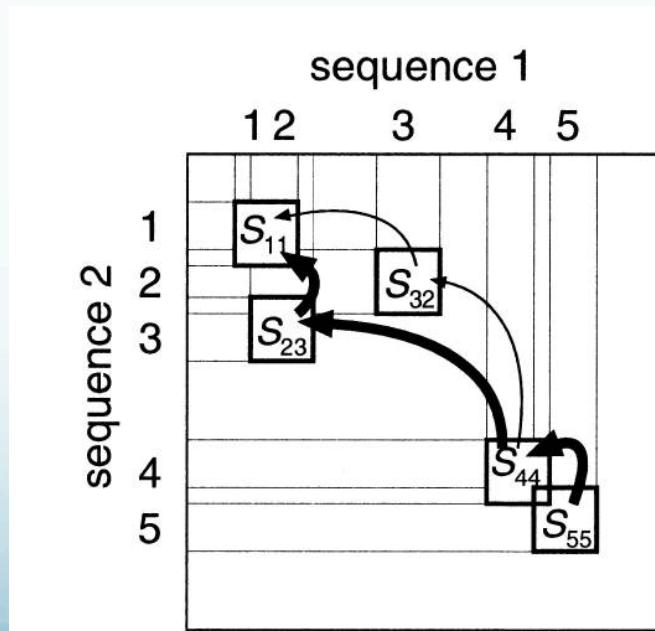


# MAFFT Algorithm



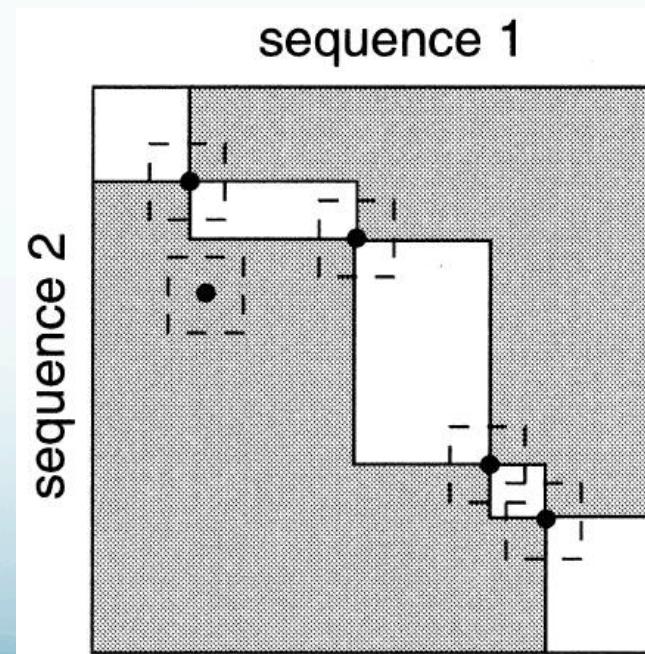
# MAFFT Algorithm

- Dividing a homology matrix
- To obtain an alignment between two sequences, the homologous segments must be arranged consistently in both sequences.



# MAFFT Algorithm

- Overall homology matrix is divided into some sub-matrices at the boundary corresponding to the center of homologous segments
- Shaded area is excluded from the calculation.



# MAFFT Algorithm further reading

- group-to-group alignment
- similarity matrix and gap penalties
- FFT-NS-1, FFT-NS-2, FFT-NS-I
- NW-NS-1/NW-NS-2
- NW-AP-2

# MUSCLE Algorithm

- Biological accuracy and computational complexity
- The basic strategy used by MUSCLE is similar to that used by PRRP and MAFFT.
- MUSCLE has three stages. At the completion of each stage, a multiple alignment is available and the algorithm can be terminated.

# MUSCLE Algorithm

- Stage 1: draft progressive
- The similarity of each pair of sequences is computed, either using k-mer counting or by constructing a global alignment of the pair and determining the fractional identity
- A triangular distance matrix is computed from the pairwise similarities.

# MUSCLE Algorithm

- A tree is constructed from the distance matrix using UPGMA or neighbor-joining
- A root is identified
- A progressive alignment is built

# MUSCLE Algorithm

- Stage 2: improved progressive
- Similarity measure
- A tree is constructed by computing a Kimura distance matrix and applying a clustering method to this matrix.

# MUSCLE Algorithm

- This method applies only to nucleic acids and takes into consideration the fact that transition substitutions (purine-purine or pyrimidine-pyrimidine) often occur much more frequently than transversion substitutions (purine-pyrimidine). Gap positions and ambiguous symbols other than R (purine) and Y (pyrimidine) are not scored.
- $P = \text{transitions} / \text{positions\_scored}$
- $Q = \text{transversions} / \text{positions\_scored}$
- **$\text{distance} = -(1)/(2) \ln[ (1 - 2P - Q) * \sqrt{1 - 2Q} ]$**
- This method gives better distance estimates than the Jukes-Cantor method when the rates of transitional and transversional substitutions are different. However, when the substitution pattern is more complex than this, this method underestimates the true distance for distantly related sequences.

# MUSCLE Algorithm

- Tree comparison
- Progressive alignment

# MUSCLE Algorithm

- Stage 3: refinement
- Choice of bipartition
- Profile extraction
- Re-alignment
- Accept/reject

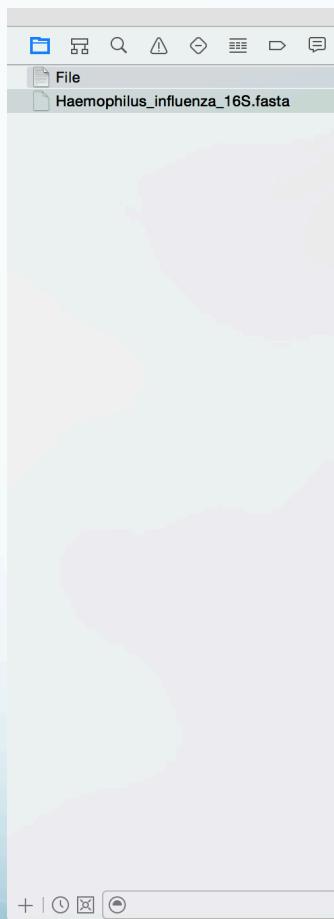
## II. Alignment MAFFT and MUSCLE

Part 2

# Retrieve the fasta file

- Save the files:
- 16S for *Haemophilus influenzae*:  
`/mnt/HA/groups/nsftuesGrp/data/  
Haemophilus_influenzae_16S.fasta`
- GInS for *Haemophilus influenzae*:  
`/mnt/HA/groups/nsftuesGrp/data/  
Haemophilus_influenzae_GInS.fasta`

# Retrieve the fasta file



The screenshot shows a Mac OS X desktop environment with a Finder window open. The window title bar says "File < > | File No Selection". The left sidebar shows two items: "File" and "Haemophilus\_influenza\_16S.fasta". The main pane displays a large block of FASTA sequence data. The data starts with a header line starting with ">Haemophilus\_influenzae\_PittEE\_uid58591\_5" and continues with several more lines of DNA sequence. The sequence consists of a single continuous string of letters representing the nucleotide sequence.

```
>Haemophilus_influenzae_PittEE_uid58591_5
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGGTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittEE_uid58591_6
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_1
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_2
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_3
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_4
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_5
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_PittGG_uid58593_6
GAATTGAAGACTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
GATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGG
>Haemophilus_influenzae_Rd_kw20_uid57771_1
AACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
AGGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_Rd_kw20_uid57771_2
AACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
AGGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_Rd_kw20_uid57771_3
ACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
GGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_Rd_kw20_uid57771_4
ACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
GGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_Rd_kw20_uid57771_5
ACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
GGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_Rd_kw20_uid161923_6
AACTTGAATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
AGGGGATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGG
>Haemophilus_influenzae_R2866_uid161923_1
AATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGGGG
ATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGGGG
>Haemophilus_influenzae_R2866_uid161923_2
AATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGGGG
ATAACGACGGAAACTGTCGCTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGGGG
>Haemophilus_influenzae_R2866_uid161923_3
AATTGAAGAGTTGATCATGGCTCAGATTGAAACGCTGGCGCAGGCTTAACACATGCAAGTCGAACGGTAGCAGGAGAAAGCTTCTTCTGCTGACGAGTGGCGGACGGGTGAGTAATGCTTGGAAATCTGGCTATGGAGGGGG
```

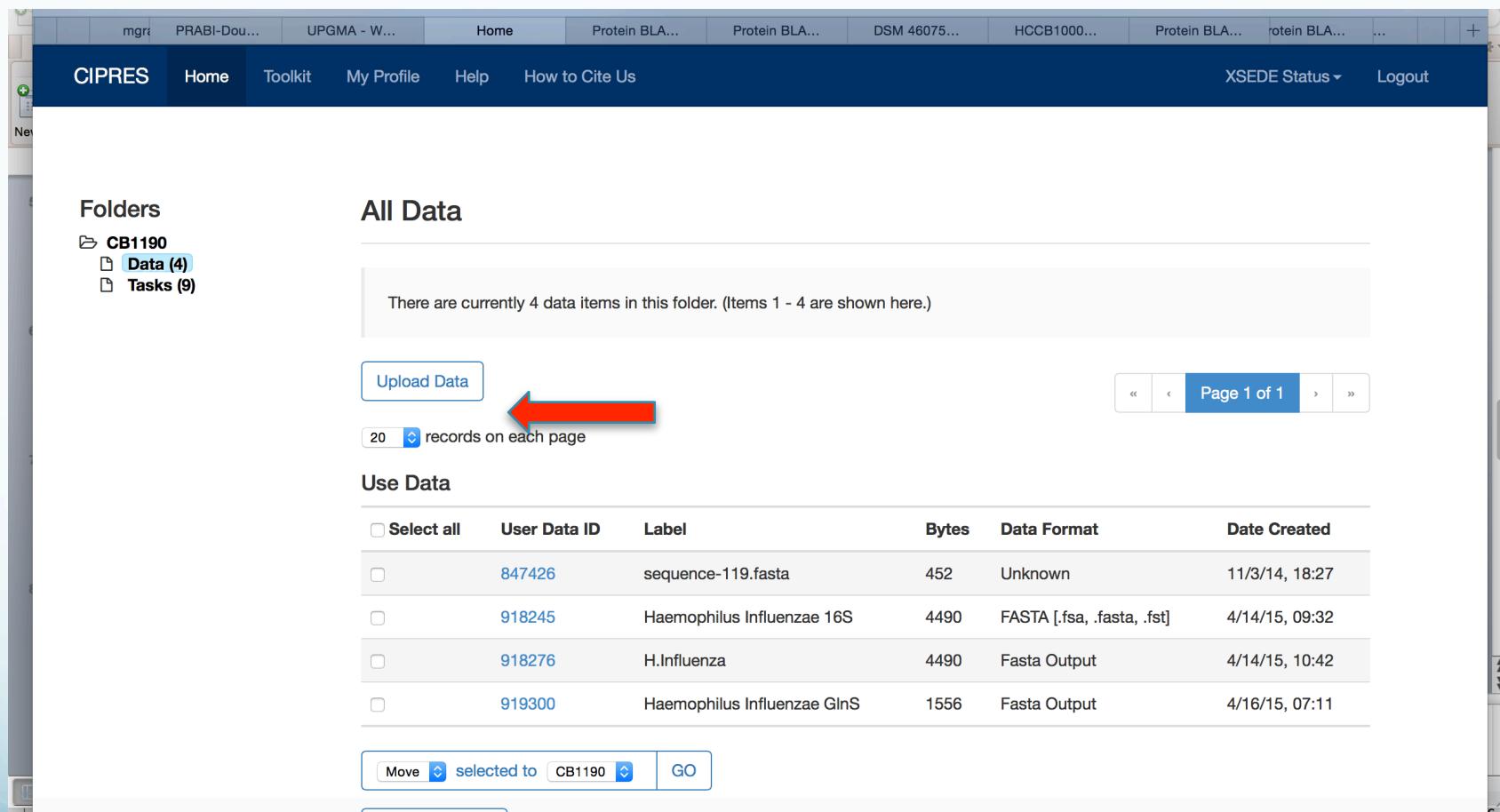
# Submission and Results

- Open an account and create a profile in **CIPRES**  
<https://www.phylo.org/portal2/home.action>

# Submission and Results

The screenshot shows the CIPRES Science Gateway interface. At the top, there is a navigation bar with various links like 'Home', 'Protein BLA...', 'DSM 46075...', etc. Below the navigation bar is the CIPRES logo and the XSEDE logo. The main content area has a sidebar on the left labeled 'Folders' containing a tree icon and a folder named 'CB1190' which contains 'Data (4)' and 'Tasks (9)'. The main panel displays a welcome message: 'Welcome saeedkeshani to the CIPRES Science Gateway V 3.3. If you would like to view a demo, click [here](#).'. Below this message is a 'Create New Folder' button. A red arrow points from the text 'Current Folder Details' down to the 'Create New Folder' button. The 'Current Folder Details' section shows the label 'CB1190' and three buttons: 'Create Subfolder', 'Edit Folder', and 'Delete Folder'.

# Submission and Results



The screenshot shows the CIPRES web interface. The top navigation bar includes links for mgmt, PRABI-Dou..., UPGMA - W..., Home, Protein BLA..., Protein BLA..., DSM 46075..., HCCB1000..., Protein BLA..., Protein BLA..., ..., and +. Below the navigation bar, the CIPRES logo is on the left, followed by Home, Toolkit, My Profile, Help, How to Cite Us, XSEDE Status, and Logout.

The main content area has two sections: "Folders" and "All Data". The "Folders" section shows a single folder named CB1190 with sub-items Data (4) and Tasks (9). The "All Data" section displays a message: "There are currently 4 data items in this folder. (Items 1 - 4 are shown here.)". Below this, there is an "Upload Data" button, a dropdown menu set to "20" records on each page, and a page navigation bar showing "Page 1 of 1". A red arrow points to the "Upload Data" button.

**All Data**

There are currently 4 data items in this folder. (Items 1 - 4 are shown here.)

Upload Data

20  Page 1 of 1

**Use Data**

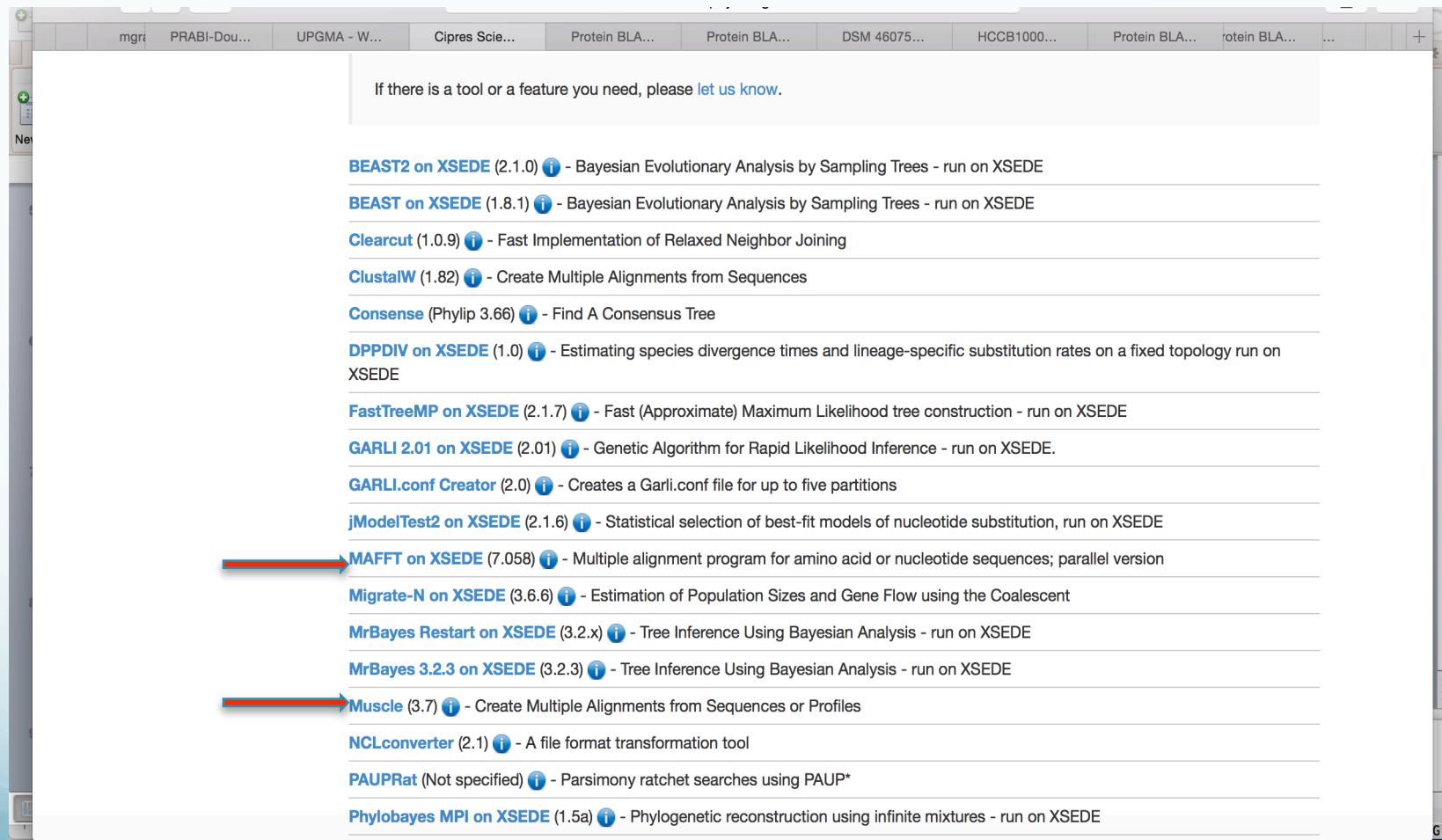
<input type="checkbox"/> Select all	User Data ID	Label	Bytes	Data Format	Date Created
<input type="checkbox"/>	847426	sequence-119.fasta	452	Unknown	11/3/14, 18:27
<input type="checkbox"/>	918245	Haemophilus Influenzae 16S	4490	FASTA [.fsa, .fasta, .fst]	4/14/15, 09:32
<input type="checkbox"/>	918276	H.Influenza	4490	Fasta Output	4/14/15, 10:42
<input type="checkbox"/>	919300	Haemophilus Influenzae GlnS	1556	Fasta Output	4/16/15, 07:11

Move  CB1190

# Submission and Results

The screenshot shows a web-based application for managing and submitting tasks. At the top, there is a navigation bar with several tabs: 'mgra', 'PRABI-Dou...', 'UPGMA - W...', 'Home', 'Protein BLA...', 'Protein BLA...', 'DSM 46075...', 'HCCB1000...', 'Protein BLA...', 'Protein BLA...', and '...'. Below the navigation bar, there is a sidebar titled 'Folders' containing a single folder named 'CB1190' which contains 'Data (4)' and 'Tasks (9)'. The main content area is titled 'Create new task' and includes tabs for 'Task Summary' (which is selected), 'Select Data', 'Select Tool', and 'Set Parameters'. A message in the center says 'You may edit your task using the tabs above.' and 'Current CPU Hr Usage: 1 Explain this?'. Below these tabs are four input fields: 'Description' (with a 'Description' placeholder), 'Input' (with a 'Select Input Data' button), 'Tool' (with a 'Select Tool' button), and 'Input Parameters' (with a 'Set Parameters' button). At the bottom of the form are three buttons: 'Save Task' (green), 'Save and Run Task' (green), and 'Discard Task' (blue). A note at the bottom states: 'Saved tasks can be run later from the task list' and 'XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.'

# Submission and Results



If there is a tool or a feature you need, please [let us know](#).

**BEAST2 on XSEDE** (2.1.0) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

**BEAST on XSEDE** (1.8.1) ⓘ - Bayesian Evolutionary Analysis by Sampling Trees - run on XSEDE

**Clearcut** (1.0.9) ⓘ - Fast Implementation of Relaxed Neighbor Joining

**ClustalW** (1.82) ⓘ - Create Multiple Alignments from Sequences

**Consense (Phylib 3.66)** ⓘ - Find A Consensus Tree

**DPPDIV on XSEDE** (1.0) ⓘ - Estimating species divergence times and lineage-specific substitution rates on a fixed topology run on XSEDE

**FastTreeMP on XSEDE** (2.1.7) ⓘ - Fast (Approximate) Maximum Likelihood tree construction - run on XSEDE

**GARLI 2.01 on XSEDE** (2.01) ⓘ - Genetic Algorithm for Rapid Likelihood Inference - run on XSEDE.

**GARLI.conf Creator** (2.0) ⓘ - Creates a Garli.conf file for up to five partitions

**jModelTest2 on XSEDE** (2.1.6) ⓘ - Statistical selection of best-fit models of nucleotide substitution, run on XSEDE

**MAFFT on XSEDE** (7.058) ⓘ - Multiple alignment program for amino acid or nucleotide sequences; parallel version

**Migrate-N on XSEDE** (3.6.6) ⓘ - Estimation of Population Sizes and Gene Flow using the Coalescent

**MrBayes Restart on XSEDE** (3.2.x) ⓘ - Tree Inference Using Bayesian Analysis - run on XSEDE

**MrBayes 3.2.3 on XSEDE** (3.2.3) ⓘ - Tree Inference Using Bayesian Analysis - run on XSEDE

**Muscle** (3.7) ⓘ - Create Multiple Alignments from Sequences or Profiles

**NCLconverter** (2.1) ⓘ - A file format transformation tool

**PAUPRat** (Not specified) ⓘ - Parsimony ratchet searches using PAUP\*

**Phylobayes MPI on XSEDE** (1.5a) ⓘ - Phylogenetic reconstruction using infinite mixtures - run on XSEDE

# Submission and Results

- MAFFT on XSEDE:

Extreme Science and Engineering Discovery Environment (XSEDE) –

- Muscle:

# Submission and Results

The screenshot shows a software interface for sequence alignment. At the top, there is a navigation bar with various tabs: PRA, UPGMA - W..., Muscle, Protein BLA..., Protein BLA..., DSM 46075..., HCCB1000..., Protein BLA..., Protein BLA..., p://ftp.ncbi..., ..., +. Below the navigation bar, the main window displays the 'Output Options (select as many as you like)' section. This section contains a list of output formats with corresponding checkboxes:

- Specify output formats
- Fasta output
- Clustal output (.aln)
- Strict Clustal output (.aln; has the Clustal header)
- HTML output (.html)
- GCG output (.msf)
- Interleaved Phylip output
- Sequential Phylip output (.phy)

At the bottom of this section are three buttons: Save Parameters, Reset, and Cancel.

Below the output options, there is a link to 'Advanced Help'. Under 'Advanced Help', there are two sections: 'Maximum Hours to Run (click here for help setting this correctly)' and 'Preserve the order of the input sequences'. The 'Maximum Hours to Run' section provides instructions for estimating run time and notes that shorter times often result in faster completion. The 'Preserve the order of the input sequences' section explains that Muscle re-orders sequences for easier visual evaluation and describes the ?stable option for maintaining input order.

**Output Options (select as many as you like)**

Specify output formats   
Fasta output   
Clustal output (.aln)   
Strict Clustal output (.aln; has the Clustal header)   
HTML output (.html)   
GCG output (.msf)   
Interleaved Phylip output   
Sequential Phylip output (.phy)

[Save Parameters](#) [Reset](#) [Cancel](#)

[Advanced Help](#)

**Maximum Hours to Run (click here for help setting this correctly)**

Estimate the maximum time your job will need to run (up to 71.75 hrs). Your job will be killed if it doesn't finish within the time you specify, however jobs with shorter maximum run times are often scheduled sooner than longer jobs.

**Preserve the order of the input sequences**

By default, Muscle re-arranges sequences so that similar sequences are adjacent in the output file. (This is done by ordering sequences according to a prefix traversal of the guide tree). This makes the alignment easier to evaluate by eye. If you want the sequences to be output in the same order as the input file, you can use the ?stable option.

**Write a Matrix Score File**

Write a score file. This contains one line for each column in the alignment. Each line contains the letters in the column followed by the average BLOSUM62 score over pairs of letters in the column.

# Submission and Results

The screenshot shows a user interface for managing biological tasks. At the top, there is a navigation bar with various links: PRA, UPGMA - W..., Cipres Scie..., Protein BLA..., Protein BLA..., DSM 46075..., HCCB1000..., Protein BLA..., Protein BLA..., p://ftp.ncbi..., ..., +. Below the navigation bar, a green notification box displays the message "28 parameters successfully set to current task." A red arrow points upwards from the bottom of the interface towards this message.

**Folders**

- CB1190
  - Data (4)
  - Tasks (9)

**Create new task**

You may edit your task using the tabs above.  
Current CPU Hr Usage: 1 [Explain this?](#)

**Description**

**Input** [1 Inputs Set](#)

**Tool** [Muscle](#) Click for more info

**Input Parameters** [29 Parameters Set](#)

[Save Task](#) [Save and Run Task](#) [Discard Task](#)

Saved tasks can be run later from the task list  
XSEDE tasks are limited to 168 hours. Non-XSEDE tasks are limited to 72 hours.

# Submission and Results

The screenshot shows a web-based bioinformatics tool interface. At the top, there is a navigation bar with several tabs: PRA, UPGMA - W..., Cipres Scie..., Protein BLA..., Protein BLA..., DSM 46075..., HCCB1000..., Protein BLA..., Protein BLA..., p://ftp.ncbi..., and ... . The tabs Protein BLA... and Protein BLA... are underlined, indicating they are active.

Below the navigation bar, there is a search bar with the placeholder "Search" and a dropdown menu showing "Show 20 records on each page". To the right of the search bar is a blue button labeled "Page 1 of 1".

The main content area is a table with the following columns: Label, Tool, Input, Parameters, Date Created, and Action. The table lists eight entries:

<input type="checkbox"/>	Label	Tool	Input	Parameters	Date Created	Action
<input type="checkbox"/>	Clone Haemophilus Influenzae GlnS. MAFFT	MAFFT on XSEDE	<a href="#">View (1)</a>	<a href="#">View (16)</a>	4/16/15, 07:14	<a href="#">View Output</a>
<input type="checkbox"/>	Clone Haemophilus Influenzae GlnS. Muscle	Muscle	<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/16/15, 07:12	<a href="#">View Output</a>
<input type="checkbox"/>	Clone H.Influenza	Muscle	<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 10:42	<a href="#">View Output</a>
<input type="checkbox"/>	Clone Muscle H.Influenzae	Muscle	<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 10:33	<a href="#">View Output</a>
<input type="checkbox"/>	Clone Haemophilus Influenzae 16S	MAFFT on XSEDE	<a href="#">View (1)</a>	<a href="#">View (16)</a>	4/14/15, 09:46	<a href="#">View Output</a>
<input type="checkbox"/>	Clone Haemophilus Influenzae 16S	Muscle	<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 09:32	<a href="#">View Output</a>
<input type="checkbox"/>	Clone SEQ1	ClustalW	<a href="#">View (1)</a>	<a href="#">View (43)</a>	11/4/14, 16:25	<a href="#">View Output</a>
<input type="checkbox"/>	Clone Tree	BEAST2	<a href="#">View (1)</a>	<a href="#">View (5)</a>	11/3/14,	<a href="#">View Output</a>

# SeaView

- Visualization tool,...  
<http://doua.prabi.fr/software/seaview>



**seaview**

# SeaView

The screenshot shows a web browser window with the PRABI-Doua website loaded. The top navigation bar includes links like Hae, Cipres Scie..., mgraist\_tuto..., PRABI-Doua..., UPGMA - W..., Cipres Scie..., Protein BLA..., Protein BLA..., DSM 46075..., ICCB1000..., ..., +, and a search bar with the number 3. The main content area features the PRABI-Doua logo and the text "Pôle Rhône-Alpes de Bioinformatique Site Doua". Below this, a section titled "SeaView" discusses "Version 4.5.4" with a list of new features. To the right, there are social sharing icons for Google+ (15), Facebook (Recommend 18), and Twitter (2). A sidebar on the left contains links for Presentation, Online Services, Databases, Software and packages (highlighted in red), Miscellaneous, Legacy, and Partners. A small envelope icon is also present in the sidebar.

## SeaView

### Version 4.5.4

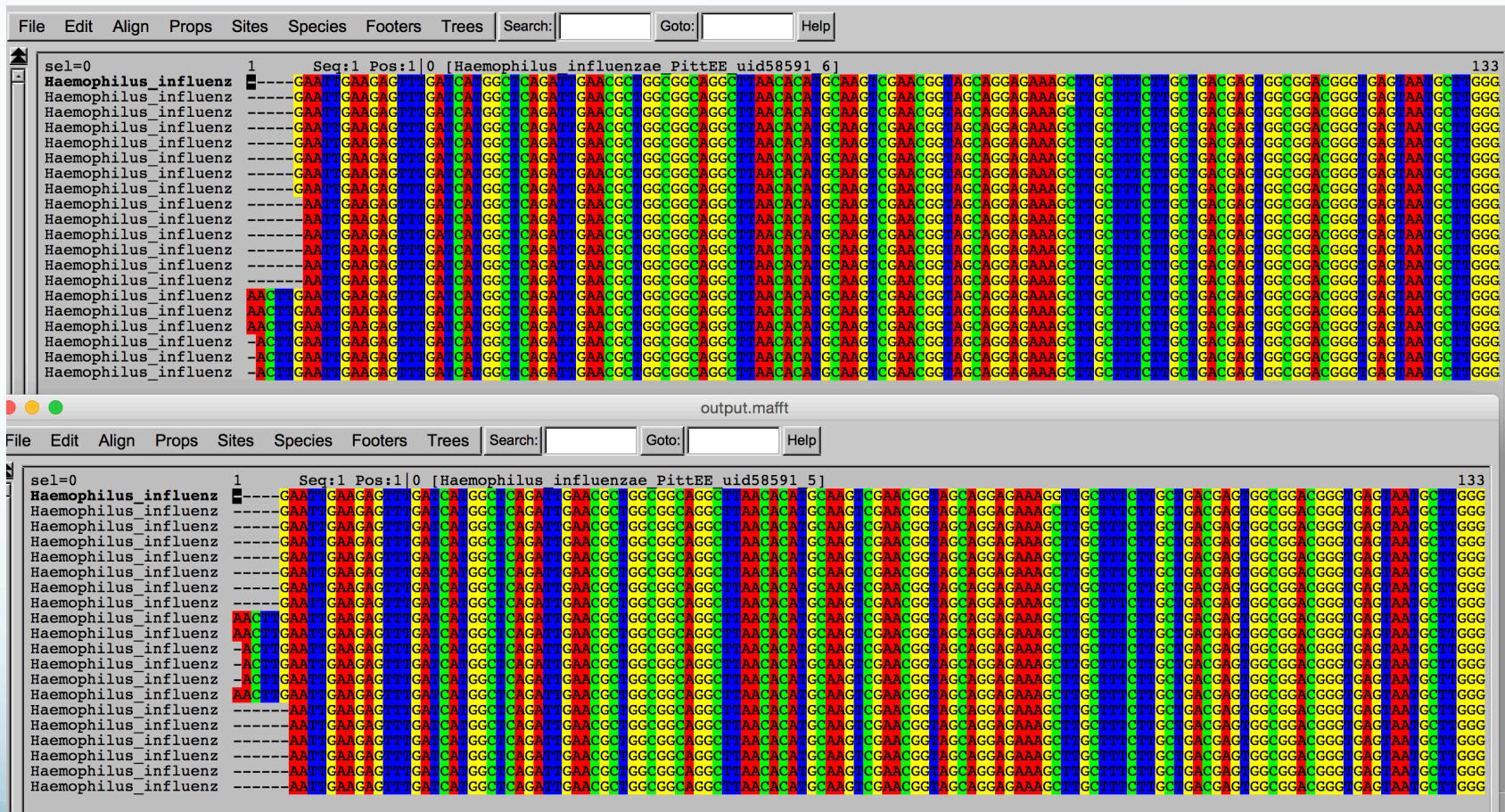
NEW: 64-bit version for the MS Windows platform  
NEW: multiple-tree windows  
NEW: seaview uses **PHYLIP v3.696** to compute parsimony trees  
NEW: seaview can be run without GUI using a command line  
NEW: seaview drives the **PhyML v3.1** program to compute maximum likelihood phylogenetic trees.  
NEW: seaview drives the **Gblocks** program to select blocks of conserved sites.  
NEW: seaview drives the **Clustal Ω** program to perform multiple sequence alignment.

SeaView is a multiplatform, graphical user interface for multiple sequence alignment and molecular phylogeny.

- SeaView reads and writes various file formats ([NEXUS](#), MSF, CLUSTAL, FASTA, PHYLIP, [MASE](#), Newick) of DNA and protein sequences and of phylogenetic trees.
- SeaView drives programs [muscle](#) or [Clustal Omega](#) for multiple sequence alignment, and also allows to use any external alignment algorithm able to read and write FASTA-formatted files.
- Seaview drives the [Gblocks](#) program to select blocks of evolutionarily conserved sites.
- SeaView computes phylogenetic trees by
  - parsimony, using PHYLIP's [dnaps/prot�aps](#) algorithm,
  - distance, with [NJ](#) or [BioNJ](#) algorithms on a variety of evolutionary distances,
  - maximum likelihood, driving program [PhyML](#) 3.1.
- SeaView prints and draws phylogenetic trees on screen, SVG, PDF or PostScript files.
- SeaView allows to download sequences from EMBL/GenBank/UniProt using the Internet.

Screen shots of the main [alignment](#) and [tree](#) windows. Dialog window to perform [Maximum-Likelihood](#) tree-building.  
On-line [help](#) document.Old [seaview version 3.2](#)

# Comparison



# Comparison

The screenshot shows a web browser window with two main content areas. On the left, a table titled "MAFFT on XSEDE - Parameters" displays various command-line parameters and their values. On the right, a log table from "phylo.org" lists completed jobs, each with a "View" link, a date and time, and a "View Status" button. A red arrow points to the date "4/16/15, 07:14" for the second job in the log.

MAFFT on XSEDE - Parameters			
Simple Parameters			
auto_analysis_	true		
datatype_	0		
fmodel_	0		
memSave_	0		
noScore_	1		
opPenaltyGroupToGroup_	1.53		
outputFormat_	0		
outputGuideTree_	0		
outputOrder_	--inputorder		
runtime_	2		
usePartTree_	0		
use_add_	false		
use_addfrag_	false		
use_addprof_	false		
use_merge_	false		
use_seed_	false		

put	Parameters	Date Created	Action
<a href="#">View (1)</a>	<a href="#">View (16)</a>	4/16/15, 07:14	<a href="#">View Status</a>
<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/16/15, 07:14	<a href="#">View Status</a>
<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 10:42	<a href="#">View Output</a>
<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 10:33	<a href="#">View Output</a>
<a href="#">View (1)</a>	<a href="#">View (16)</a>	4/14/15, 09:46	<a href="#">View Output</a>
<a href="#">View (1)</a>	<a href="#">View (28)</a>	4/14/15, 09:32	<a href="#">View Output</a>
<a href="#">View (1)</a>	<a href="#">View (43)</a>	11/4/14, 16:25	<a href="#">View Output</a>
<a href="#">View (1)</a>	<a href="#">View (5)</a>	11/3/14, 18:41	<a href="#">View Output</a>

# Works Cited

- Katoh, Kazutaka et al. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30.14 (2002): 3059–3066. Print.
- Edgar, Robert C. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32.5 (2004): 1792–1797. PMC. Web. 19 Apr. 2015.

**Thanks**

**Questions?**