
Tutorial #3

— MUSCLE and MAFFT Alignments —
Using Cipres

What is a sequence alignment

- A sequence alignment is extremely important when comparing multiple versions of the same gene.
- Only by lining up the nucleotides will true differences show themselves
- In the tutorial we will walk through the MUSCLE and MAFFT alignment algorithms


Cipres

<http://www.phylo.org/>



CIPRES

Cyberinfrastructure for
Phylogenetic Research



[Home](#) [About](#) [Help](#) [Links](#) [News](#)

[Home »](#)

Submissions are working normally

[Codes](#)
[Requirements](#)
[Limitations](#)
[Architecture](#)
[Known Issues](#)
[Usage Statistics](#)
[User Locations](#)
[Survey Results](#)
[Publications](#)

The CIPRES Science Gateway V. 3.3

The CIPRES Science Gateway [V. 3.3](#) is a public resource for inference of large phylogenetic trees. It is designed to provide all researchers with access to NSF [XSEDE](#)'s large computational resources through a simple browser interface. **You can now also access these same capabilities programatically with the [CIPRES REST API](#).**

High Performance Parallel Codes for Large Tree Inference and Sequence Alignment on XSEDE:
[RAxML](#); [MrBayes](#); [BEAST](#); [BEAST2](#); [GARLI](#); [MAFFT](#); [DPPDIV](#); [FastTree](#); [jModelTest2](#); [PAUP](#); [ParallelStructure](#); and [Migrate-N](#). *If you need access to [PhyloBayes](#), please inquire.*

Serial Codes for Tree Inference:
[PAUP*](#) (Inference by Parsimony); [Poy](#) (Alignment and Inference);

Serial Codes for Sequence Alignment:
[ClustalW](#); [Contraalign](#); [MUSCLE](#); [PROBCONS](#); [PROBALIGN](#)

[▶ Use the CIPRES Science Gateway](#)



Join the [CIPRES Google Group](#) for questions and problems.

Creating an account



Missing results?

Send us the [job handle](#), and we may be able to help.

More Information

[About Us](#)
[Requirements](#)
[Known Issues](#)
[Usage Statistics](#)
[User Locations](#)
[Enabled Publications](#)

The CIPRES Science Gateway now offers BEAST2 and PhyloBayes MPI, along with RAXML, MrBayes and other codes.

First Time Users: Please review the [XSEDE Primer](#) and our [Fair Use Policy](#).

CIPRES Login:

Username

CountMelon

Password

.....

Login

Reset

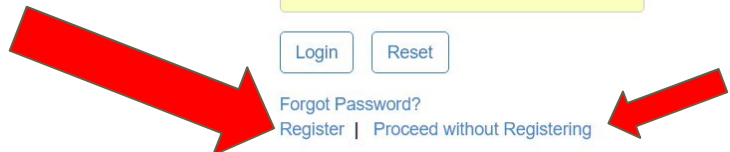
[Forgot Password?](#)

[Register](#) | [Proceed without Registering](#)

iPlant users login here:

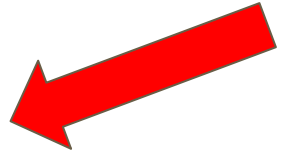


[What is this?](#)



Grab the data from proteus

/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_16S.fasta



/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_GlnS.fasta

Creating a new data storage folder

The screenshot displays the CIPRES Science Gateway interface. The top navigation bar includes the CIPRES logo, navigation links (Home, Toolkit, My Profile, Help, How to Cite Us), and user status (XSEDE Status, Logout). The main content area is divided into a left sidebar and a central form.

Left Sidebar:

- Folders
- Total Storage: 700 KB
- Tutorial #3
 - Data (2)
 - Tasks (4)

Enter Folder Detail Form:

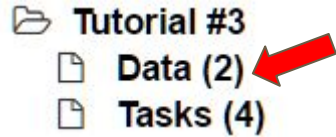
This form contains two input fields: "Label" and "Description". Two large red arrows point to these fields, indicating where to enter information. Below the fields are "Save" and "Cancel" buttons.

Current Folder Details Callout Box:

This box provides a summary of the current folder's information:

Current Folder Details	
Label	Tutorial #3
Description	tutorial #3 for Dr Rosen's class
<div><button>Create Subfolder</button><button>Edit Folder</button><button>Delete Folder</button></div>	

Uploading data



Folders

Total Storage: 700 KB

- test
 - Data (0)
 - Tasks (0)
- Tutorial #3
 - Data (2)
 - Tasks (4)

All Data

There are currently 2 data items in this folder. Items 1 - 2 are shown here.)

Upload Data

« < Page 1 of 1 > »

20 records on each page

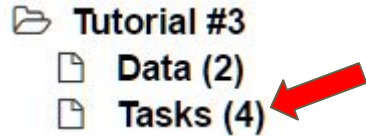
Use Data

<input type="checkbox"/> Select all	User Data ID	Label	Bytes	Data Format	Date Created
<input type="checkbox"/>	1385520	Haemophilus_influenzae_16S.fasta	66688	Unknown	2/6/17, 13:34
<input type="checkbox"/>	1385521	Haemophilus_influenzae_GlnS.fasta	11711	Unknown	2/6/17, 13:34

Move selected to Tutorial #3 GO

Delete Selected

Creating a task



Folders
Total Storage: 700 KB

- test
 - Data (0)
 - Tasks (0)
- Tutorial #3
 - Data (2)
 - Tasks (4)

Tasks

Refresh Tasks

Current CPU Hr Usage: 2 [Explain this?](#)

There are currently 4 tasks in this tab. (1/4 are shown here.)

Create New Task

Show 20 records on each page

Page 1 of 1

<input type="checkbox"/>	Select All	Label	Tool	Input	Parameters	Date Created	Action
<input type="checkbox"/>	Clone	MUSCLE alignment of H influenzae GlnS	Muscle	View (1)	View (28)	2/6/17, 13:51	View Output
<input type="checkbox"/>	Clone	MAFFT of H. influenzae GlnS	MAFFT on XSEDE	View (1)	View (22)	2/6/17, 13:46	View Output

Setting the task parameters

The screenshot displays the CIPRES Science Gateway interface. At the top is a dark blue header with the CIPRES logo and navigation links: CIPRES, Home, Toolkit, My Profile, Help, and How to Cite Us. On the left, a 'Folders' sidebar shows a tree structure: test (Data (0), Tasks (0)), Tutorial #3 (Data (2), Tasks (4)). The main content area is titled 'Create new task' and features three tabs: Task Summary, Select Data, Select Tool, and Set Parameters. Three red arrows point to these tabs. Below the tabs, a message states: 'You may edit your task using the tabs above. Current CPU Hr Usage: 2 Explain this?'. The 'Description' field has a red arrow pointing to it. Below the description are three buttons: 'Select Input Data', 'Select Tool', and 'Set Parameters'. At the bottom are three buttons: 'Save Task', 'Save and Run Task', and 'Discard Task'.

CIPRES SCIENCE GATEWAY

CIPRES Home Toolkit My Profile Help How to Cite Us

Folders

Total Storage: 700 KB

- test
 - Data (0)
 - Tasks (0)
- Tutorial #3
 - Data (2)
 - Tasks (4)

Create new task

Task Summary Select Data Select Tool Set Parameters

You may edit your task using the tabs above.
Current CPU Hr Usage: 2 [Explain this?](#)

Description

Input Select Input Data

Tool Select Tool

Input Parameters Set Parameters

Save Task Save and Run Task Discard Task

- 1) Select your data
- 2) Select your tool
- 3) Review the input parameters
- 4) Make sure you press set parameters at the bottom
- 5) Enter a task description

If you don't enter the description as the final step it will not recognize it and throw an error

View your results

Show

20

 records on each page

«

<

Page 1 of 1

>

»

☐

Select All

Label

Tool

Input

Parameters

Date Created

Action

☐

Select all

Tool Output

File Name

File Size (Bytes)

☐

PROCESS_OUTPUT

STDOUT

0

View

Download

☐

STDERR

2071

View

Download

☐

all_txt

stdout.txt

0

View

Download

☐

stderr.txt

2071

View

Download

☐

term.txt

331

View

Download

☐

start.txt

40

View

Download

☐

logfile.txt

324

View

Download

☐

done.txt

49

View

Download

☐

fasta_output

infile.fasta

11711

View

Download

☐

output.fasta

11900

View

Download

☐

jobinfo

_JOBINFO.TXT

495

View

Download

☐

logfileout

logfile.txt

324

View

Download

Muscle

View (1)

View (28)

2/6/17, 13:51

View Output

MAFFT on XSEDE

View (1)

View (22)

2/6/17, 13:46

View Output

MAFFT on XSEDE

View (1)

View (22)

2/6/17, 13:39

View Output

Muscle

View (1)

View (28)

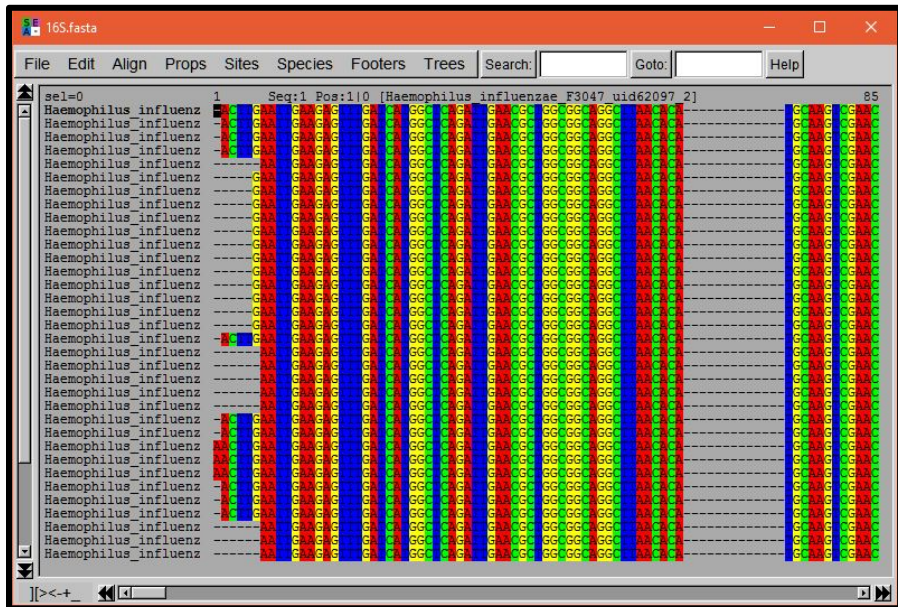
2/6/17, 13:35

View Output

</

Visualize

MUSCLE



MAFFT



An Overview of Alignment Algorithms

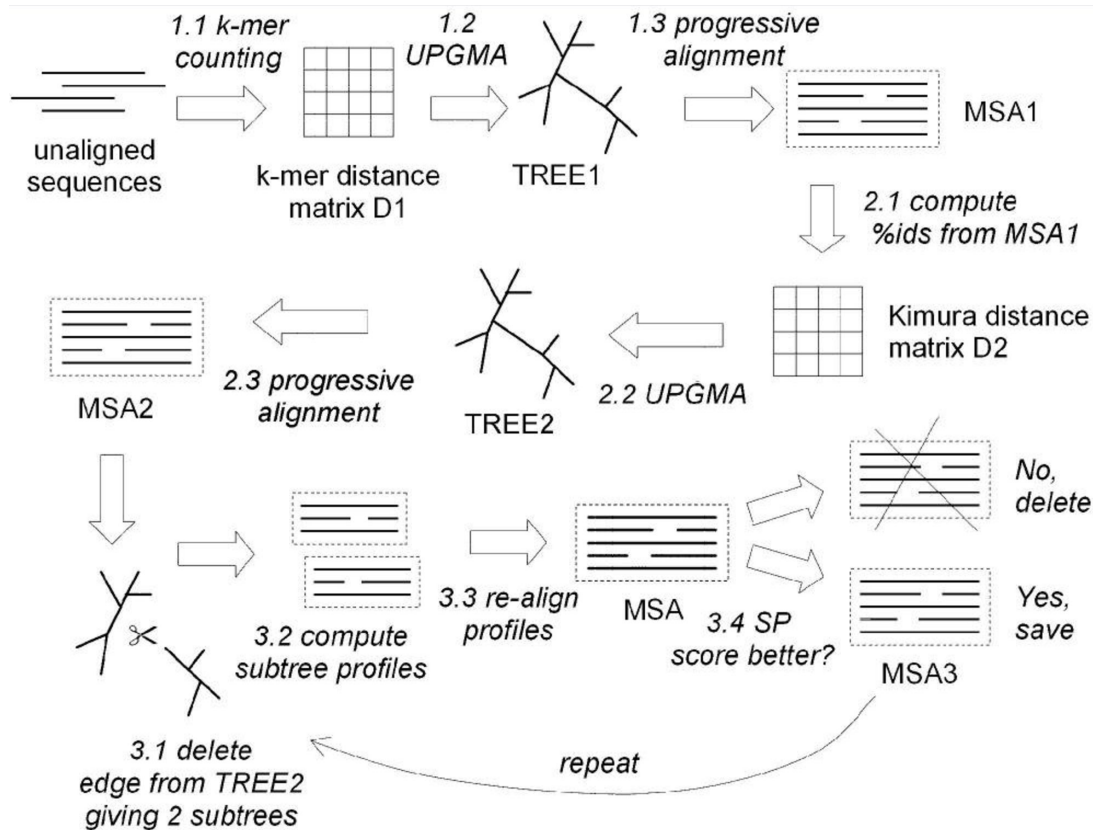
- Global Alignment
 - Usually used when sequences are generally very similar
- Local Alignment
 - Usually used when there is a need to identify similar regions in somewhat dissimilar sequences
- Variety of methods
 - Iterative methods based on scoring
 - Motif finding
 - Hidden Markov models
 - Distance matrix

MUSCLE Algorithm

- Draft Progressive

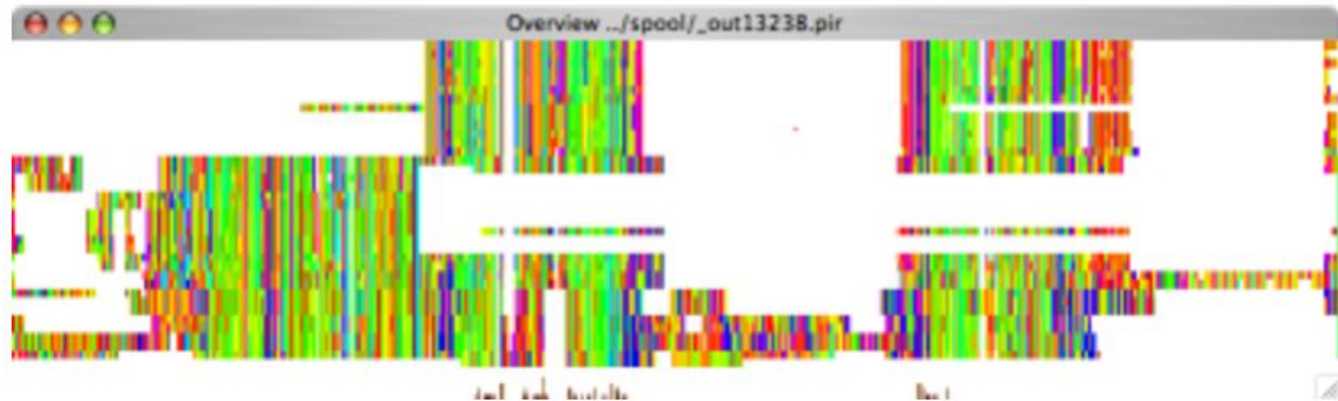
- Improved Progressive

- Refinement



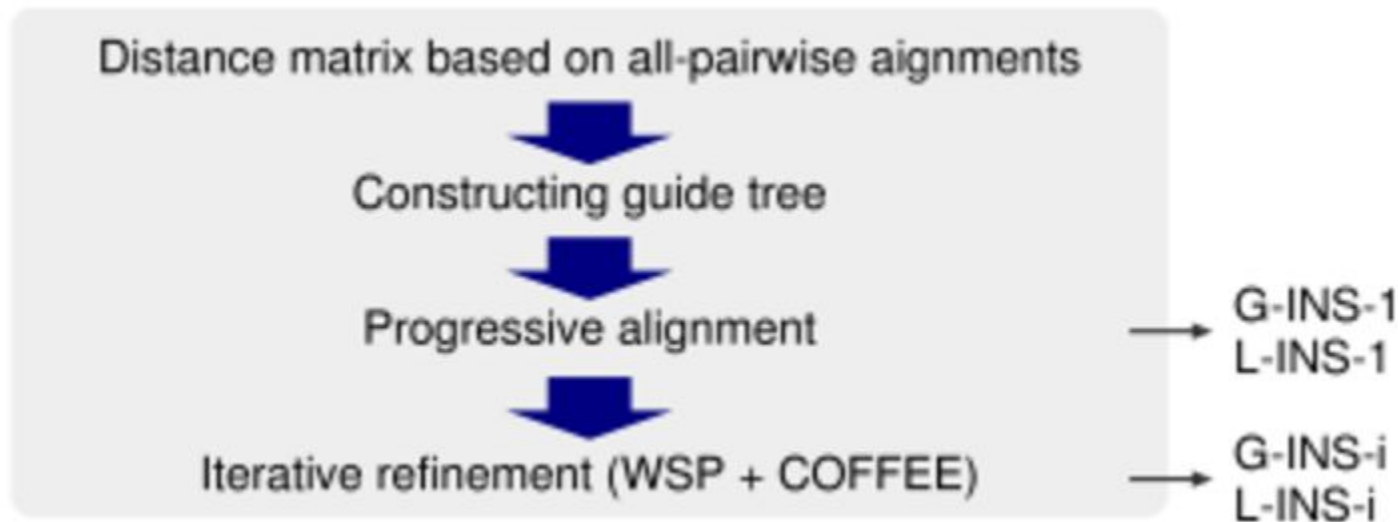
MAFFT Algorithm

- Two main algorithms
 - L-INS-i
 - FFT-NS-2
- In MAFFT, the scoring system was designed to allow for large gaps which means that it's especially effective on alignments that gap regions, as seen in both our results and the image below



MAFFT Algorithm (L-INS-i)

- Follows a similar structure to MUSCLE
- For it's pairwise alignments, uses three different types of algoirthms



Algorithm Comparison

	MUSCLE	MAFFT
16S (42 sequences)	4m 10s	34s
GlnS (7 sequences)	50s	1m 20s
All_16S (7476 sequences)	1h 3m 21s	2m 18s