Tutorial #3

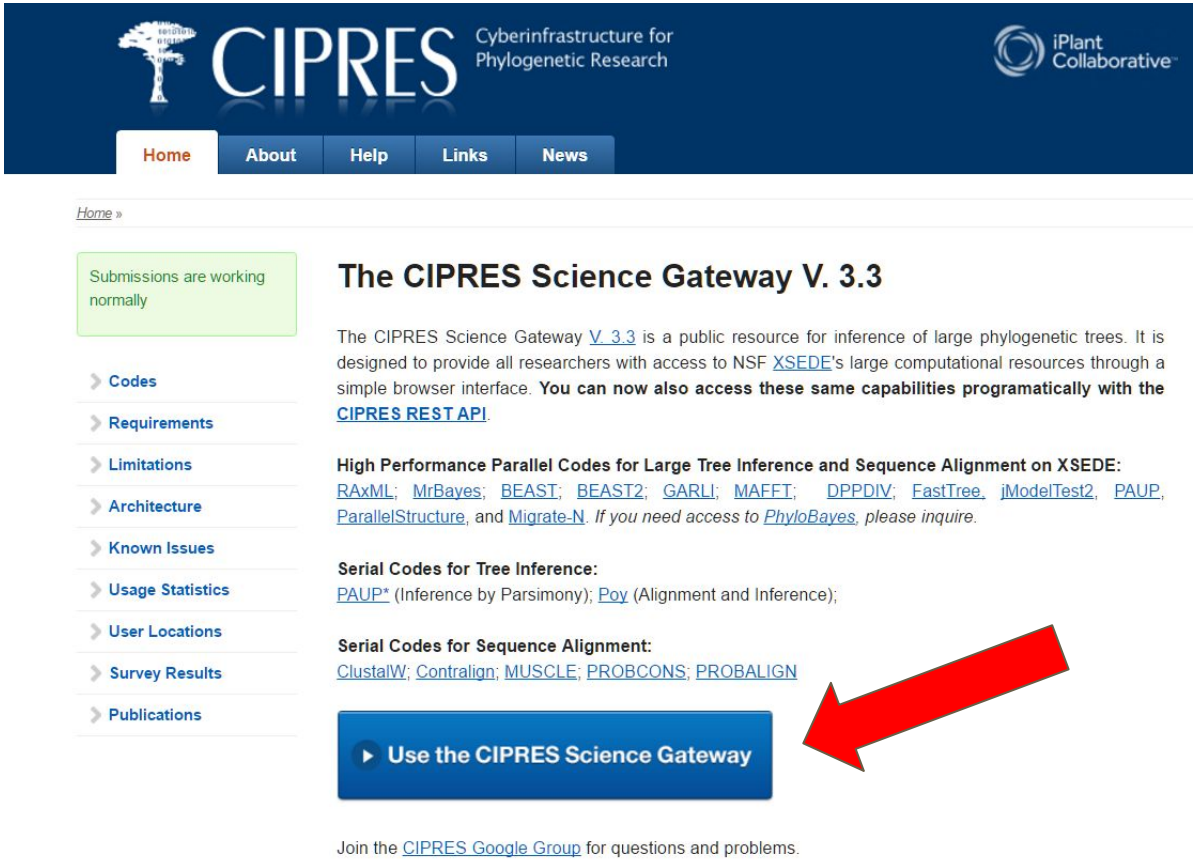
MUSCLE and MAFFT Alignments
Using Cipres

What is a sequence alignment

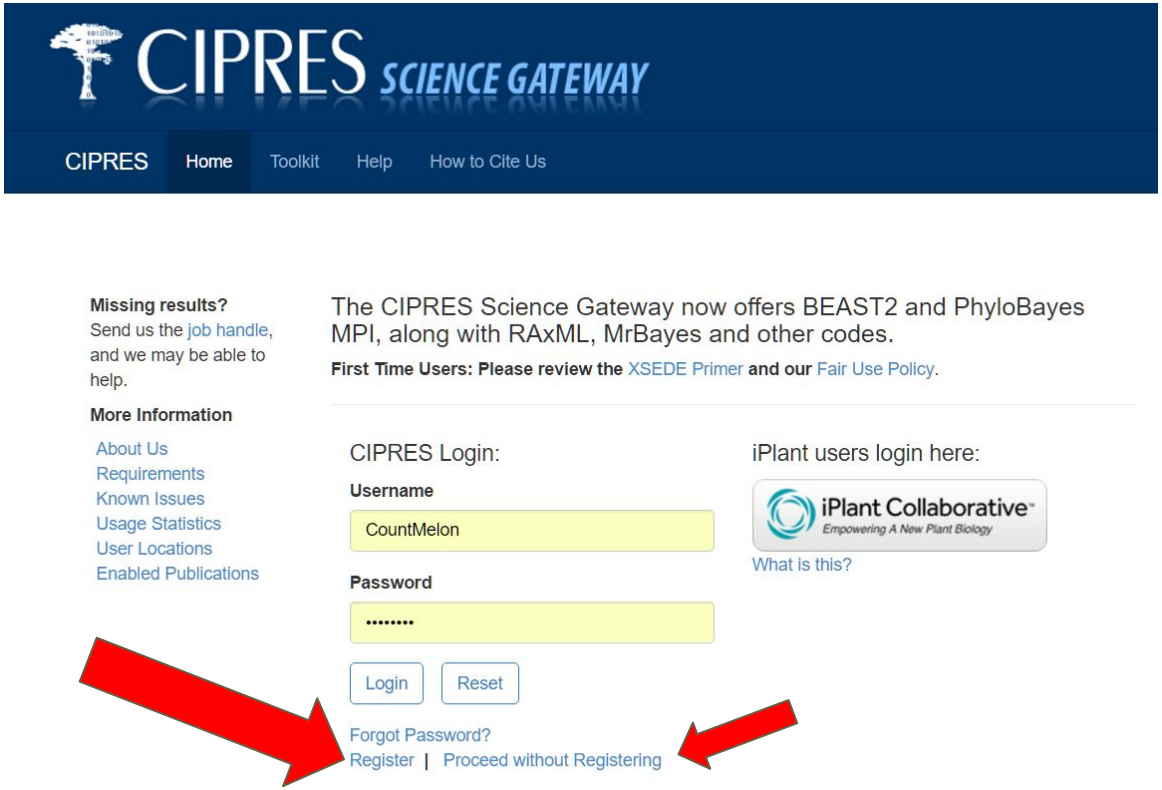
- A sequence alignment is extremely important when comparing multiple versions of the same gene.
- Only by lining up the nucleotides will true differences show themselves
- In the tutorial we will walk through the MUSCLE and MAFFT alignment algorithms

Cipres

http://www.phylo.org/

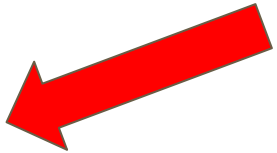


Creating an account



Grab the data from proteus

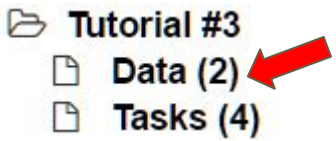
/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_16S.fasta

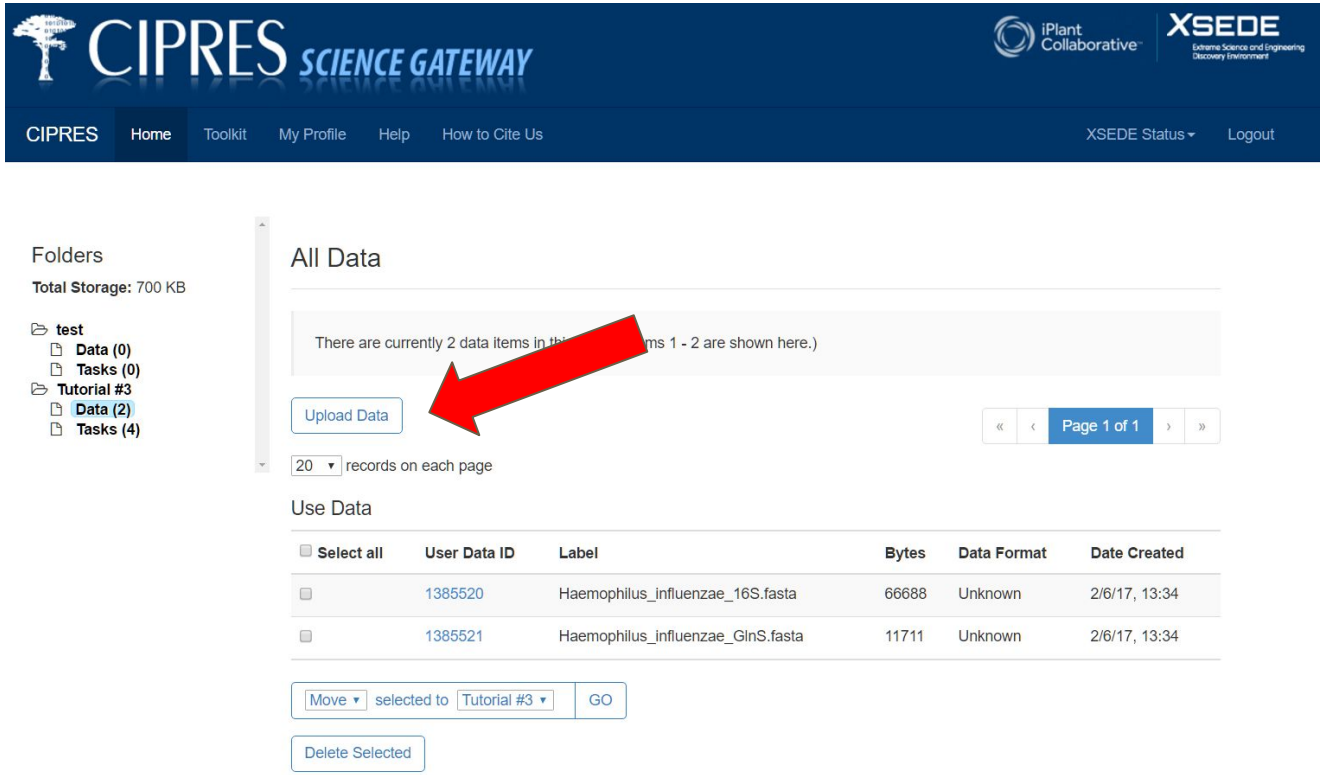


/mnt/HA/groups/nsftuesGrp/data/Haemophilus_influenzae_GlnS.fasta

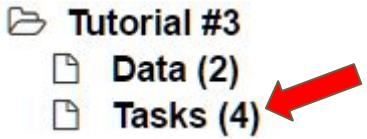
Creating a new data storage folder

Uploading data





Creating a task



CIPRES SCIENCE GATEWAY

iPlant Collaborative | XSEDE Extreme Science and Engineering Discovery Environment

CIPRES | Home | Toolkit | My Profile | Help | How to Cite Us | XSEDE Status | Logout

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. (It[illegible]re shown here.)

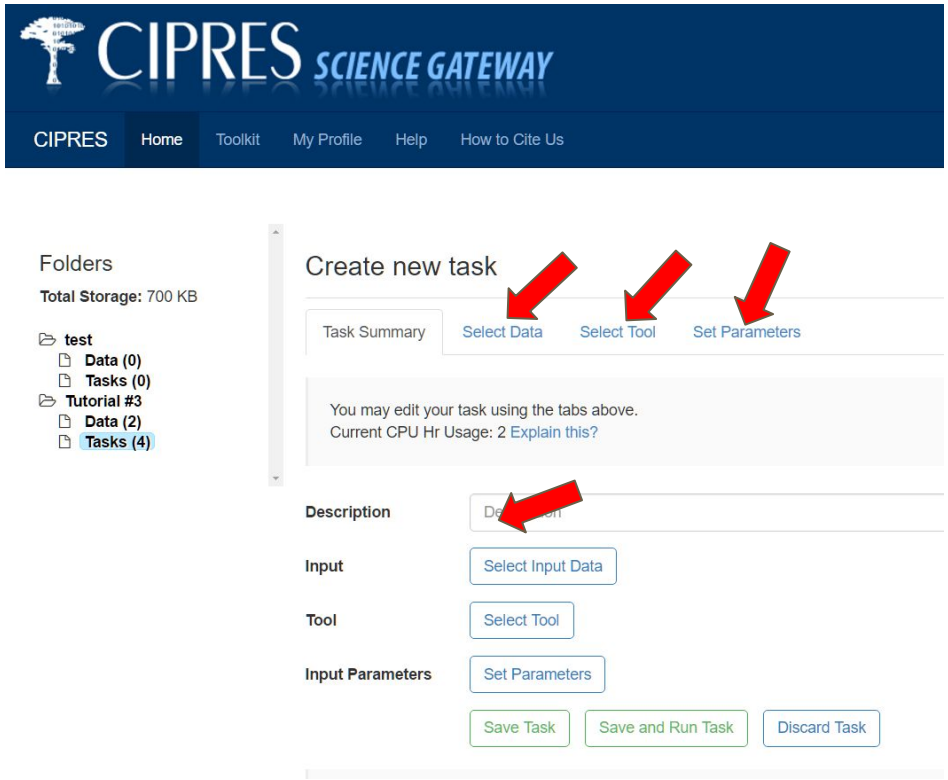
Create New Task

Show 20 records on each page

« ‹ Page 1 of 1 › »

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

Setting the task parameters



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

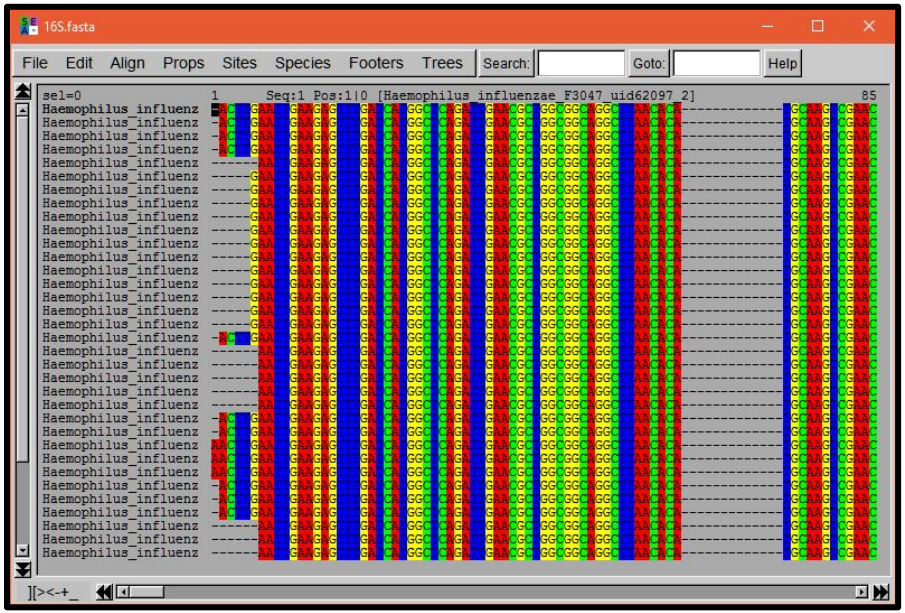
« ‹ Page 1 of 1 › »

Select All	Label	Tool	Input	Parameters	Date Created	Action
		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

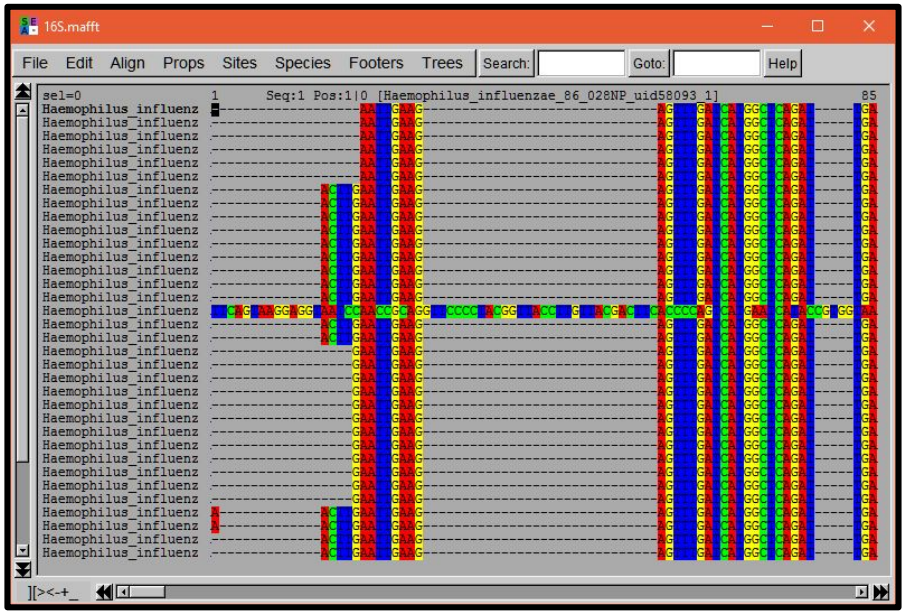
Select all	Tool Output	File Name	File Size (Bytes)		
	PROCESS_OUTPUT	STDOUT	0	View	Download
		STDERR	2071	View	Download
	all_txt	stdout.txt	0	View	Download
		stderr.txt	2071	View	Download
		term.txt	331	View	Download
		start.txt	40	View	Download
		logfile.txt	324	View	Download
		done.txt	49	View	Download
	fasta_output	infile.fasta	11711	View	Download
		output.fasta	11900	View	Download
	jobinfo	_JOBINFO.TXT	495	View	Download
	logfileout	logfile.txt	324	View	Download

Visualize

MUSCLE



MAFFT

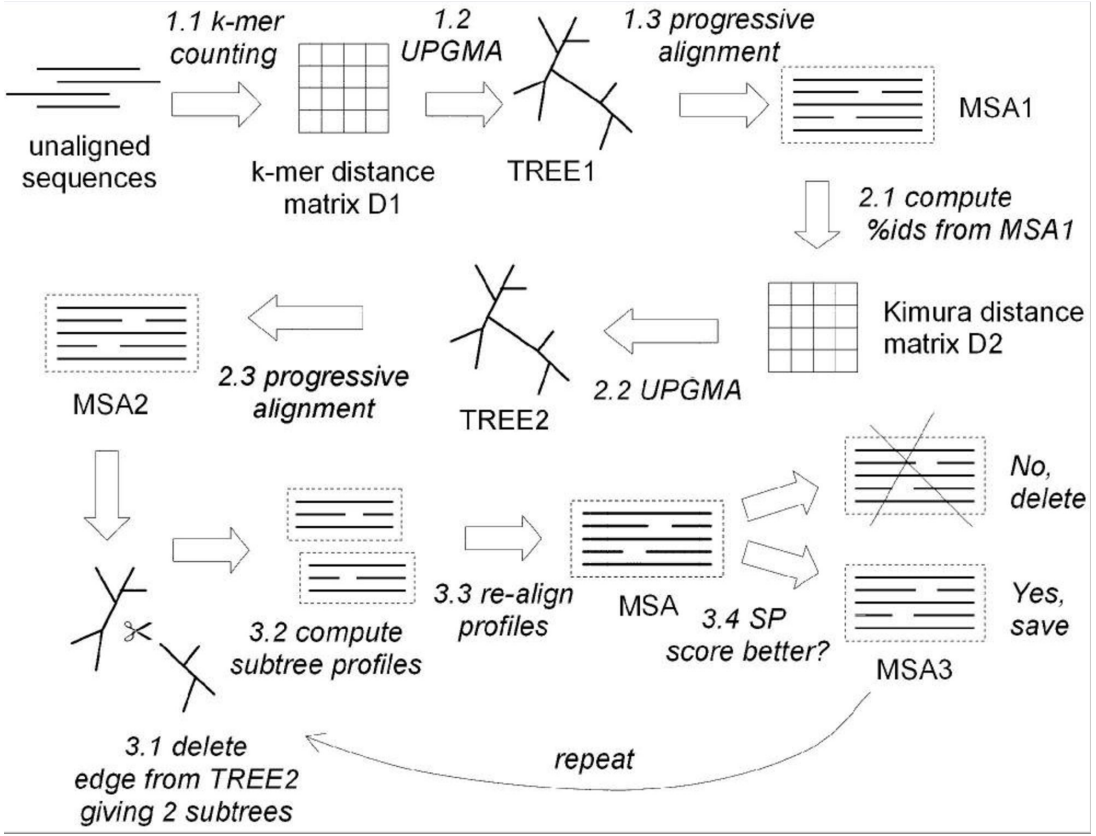


An Overview of Alignment Algorithms

- 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

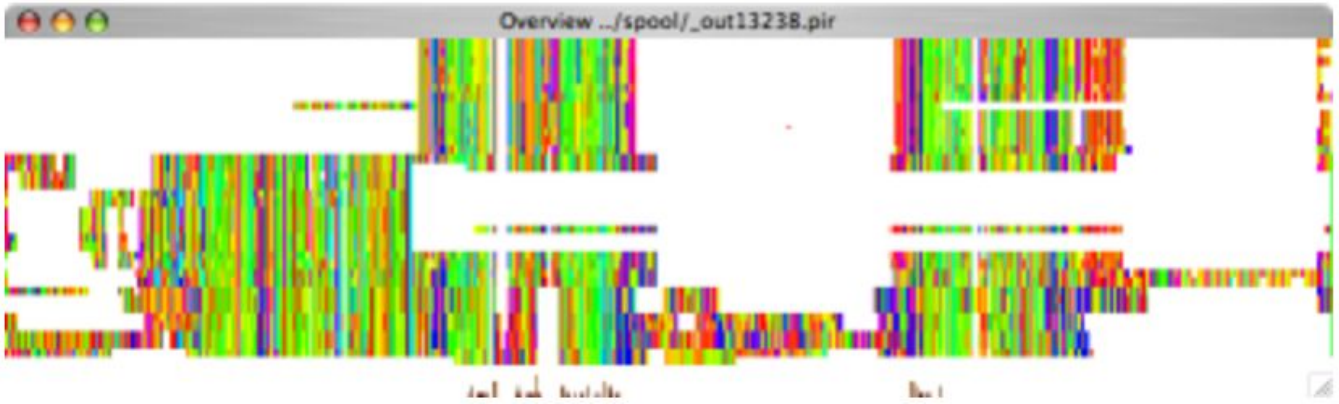
MUSCLE Algorithm

- Draft Progressive
- Improved Progressive
- Refinement



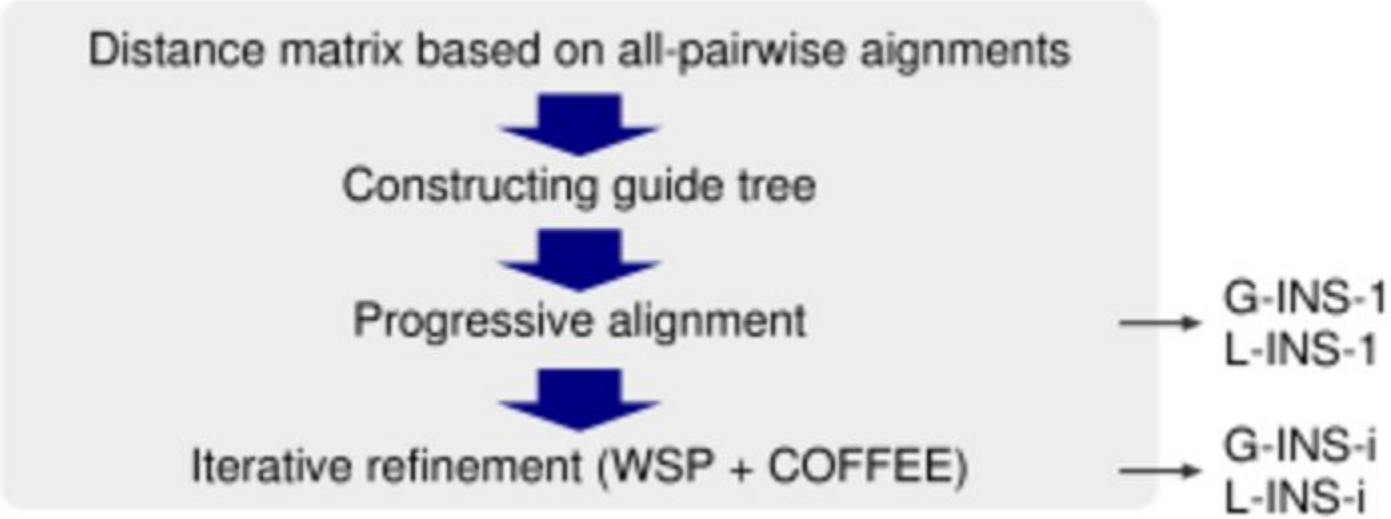
MAFFT Algorithm

- 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