

Exercise Objective: Evaluation of multiple sequence alignments generated by different algorithms. Each sequence per species represents a consensus output of previous read mapping and assembly steps.

IMPORTANT: The Alignment_Exercise contains 3 folders. The Infiles folder and Outfiles folder contain reference files for programs. Do not save any results to these folders. All of your work will be saved in Student_files.

Morbilliviruses and host receptors: A case study with Nectin4 (poliovirus receptor-related protein 4 PVRL4) and Canine Distemper Virus

Background:

There are seven known Morbilliviruses that infect Carnivores, Cetaceans, Primates, ungulates, small ruminants:

Feline morbilliviruses

Canine distemper virus

Phocine morbillivirus

Rinderpest virus

Small ruminant morbillivirus

Measles morbillivirus

CDV is known to access cell entry into the host through host genes SLAM and NECTIN4 (PVRL4). It is hypothesized that NECTIN4 is linked with neurological forms of canine distemper. We will conduct a comparative genomic analysis of NECTIN4 in mammals to determine structure, function and evolution. Our study will focus only on the coding regions (CDS) of the gene.

We have searched RefSeq and NCBI for CDS (exons) of Nectin4 from Primates, Artiodactyls, Carnivores and Cetaceans and downloaded representative sequences that are full-length CDS of NECTIN4.

PART1. GENEIOUS Importing/Editing Sequence

What makes a good alignment? One major issue for alignment methods are insertion/deletion events. We will compare the performance of 3 major algorithms, MUSCLE, CLUSTAL Omega, and MAFFT using GENEIOUS.

First create a new subfolder and import sequences into GENEIOUS

Step 1.1: Open GENEIOUS PRIME in GDWAPPS

Step 1.2: Create [GDW_Nectin4](#) Folder in Local Folder

File-New-Folder

[GDW_Nectin4](#)

Step 1.3: Import Raw Sequences

File-Import-Files-Autodetect format

[Alignment_Exercise/Infiles/32Nectin4_Data.fasta](#)

Step 1.4: Align Using MUSCLE

Tools-Align/Assemble/Multiple Align

Alignment

Geneious Alignment **MUSCLE Alignment** Clustal Omega Realign Region

Translation Align Consensus Align

Muscle 3.8.425 by Robert C. Edgar

☒ Re-align sequences ☐ Refine existing alignment

Maximum number of Iterations: 8

☒ Optimize profile-dependent parameters ?

Use custom MUSCLE executable: Change...

Maximum number of trees to build: 1

Sequence Order: ☒ Group sequences by similarity ☐ Preserve sequence order

Optimization: ☐ Diagonal ☒ Anchor

Keep tree from: ☐ Iteration 1 ☐ Iteration 2

Distance measure: Iteration 1: kmer4_6 Subsequent: pctid_kimura

Clustering method: Iteration 1 & 2: Neighbor joini... Subsequent: Neighbor joini...

Tree Rooting method: Iteration 1 & 2: pseudo Subsequent: pseudo

Sequence weighting scheme: Iteration 1 & 2: CLUSTALW Subsequent: CLUSTALW

Terminal gaps: Half penalty Objective score: spm

Anchor spacing: 32 Gap open score: -1

Diagonals: Min length: 24 Margin: 5

Minimum column anchor scores: Min best: 90 Min smoothed: 90

Hydrophobicity: Multiplier: 1.2 Window size: 5

Maximum Memory (MB): 500

☐ Quit and return best alignment found if not finished after 1 hours

Extra Options:

^ Fewer Options Cancel OK

Caption

Step 1.5: View Results. Click on upper right Expand Button of the alignment panel to enlarge display. Use command+ to zoom into alignment. Click 'Allow Editing Button'.

Step 1.6: Cleaning up regions. This is CDS only yet submitted with additional regions for 5' regions for Ma's Night monkey, and 3' regions for leopard and Cat1. To clean up, set cursor to position 393 for Ma's night monkey.

Use shift-left arrow to select entire segment.

Hit Delete key.

Go to 3' end. To edit fragments place cursor at consensus line at position 1554.

Use shift right arrow to select segment.

Hit Delete.

Normally we can export this file, but not for this exercise.

Part 2: MUSCLE and NECTIN Redux

Step 2.1: Import trimmed file arranged by phylogenetic lineage.

Import-Files-Autodetect format

[Alignment_Exercise/Infiles/Nectin32Taxa.fasta](#)

Step 2.2: Align Using MUSCLE-be sure to click 'preserve input order' as our sequences are arranged by lineage.

Tools-Align/Assemble/Multiple Align

Alignment

Geneious Alignment **MUSCLE Alignment** Clustal Omega Realign Region

Translation Align Consensus Align

[Muscle 3.8.425 by Robert C. Edgar](#)

☒ Re-align sequences ☐ Refine existing alignment

Maximum number of iterations: 8

☒ Optimize profile-dependent parameters ?

☐ Use custom MUSCLE executable: Change...

Maximum number of trees to build: 1

Sequence Order: ☐ Group sequences by similarity ☒ Preserve sequence order

Optimization: ☐ Diagonal ☒ Anchor

Keep tree from: ☐ Iteration 1 ☐ Iteration 2

Distance measure: Iteration 1: kmer4_6 Subsequent: pctid_kimura

Clustering method: Iteration 1 & 2: Neighbor joini... Subsequent: Neighbor joini...

Tree Rooting method: Iteration 1 & 2: pseudo Subsequent: pseudo

Sequence weighting scheme: Iteration 1 & 2: CLUSTALW Subsequent: CLUSTALW

Terminal gaps: Half penalty Objective score: spm

Anchor spacing: 32 Gap open score: -1

Diagonals: Min length: 24 Margin: 5

Minimum column anchor scores: Min best: 90 Min smoothed: 90

Hydrophobicity: Multiplier: 1.2 Window size: 5

Maximum Memory (MB): 500

☐ Quit and return best alignment found if not finished after 1 hours

Extra Options:

Fewer Options Cancel OK

Step 2.3: View Results. Click on Expand button to alignment panel enlarge display. Use command+ to zoom into alignment. Click 'Allow Editing Button'.

Step 2.4: Locate Insertion/deletions

Positions 43-53

Positions 1006-1008

Position 1241-1252

Step 2.5: Testing for Biological integrity-i.e. intact codons.

Translate Alignment using Translation Button on side panel
The amino acid letter is centered under the second position of each codon.

Question: Do the alignment gaps keep codon triplets together in insertion/deletion?

Step 2.6: Export file

File-Export-document-Fasta_Sequence/Alignment (*.fasta)

[Alignment Exercise/Student_files/Nectin32TaxaMuscle.fasta](#)

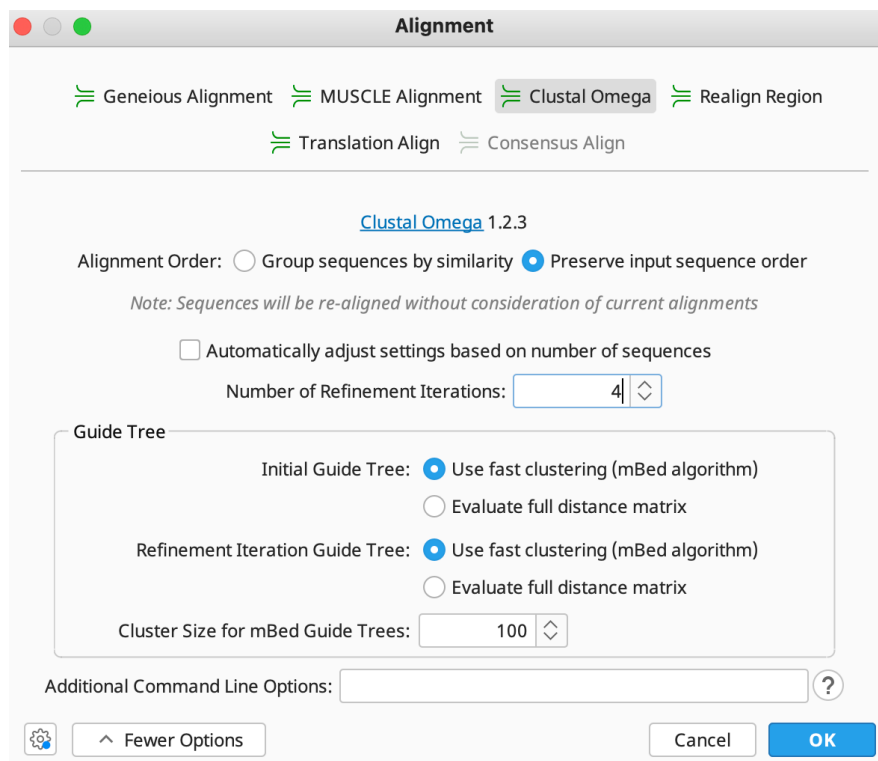
Part 3: CLUSTAL Omega and Nectin4

Step 3.1: Select unaligned file from top panel list

[Alignment_Exercise/Infiles/Nectin32Taxa.fasta](#)

Step 3.2: Align Using Clustal Omega

Tools-Align/Assemble/Multiple Align



Step 3.3: View Results. Click on Expand button to alignment panel enlarge display. Use command+ to zoom into alignment. Click 'Allow Editing Button'.

Step 3.4: Locate Insertion/deletions

Positions 44-57

Positions 1008-1010

Position 1241-1252

Step 3.5: Testing for Biological integrity-i.e. intact codons.

Translate Alignment using Translation Button on side panel

The amino acid letter is centered under the second position of each codon.

Question: Do the alignment gaps keep codon triplets together in insertion/deletion?

Step 3.6: Export file

File-Export-document-Fasta_Sequence/Alignment (*.fasta)

Save into

[Alignment Exercise/Student_files/Nectin32TaxaClustal.fasta](#)

Part 4: MAFFT and Nectin4

MAFFT is fast, powerful with high scalability capable of handling very large data sets. MAFFT has different functions depending on the types of sequences to be aligned. Here are some commands for different cases.

Step 4.1: Open terminal and go to the Infiles folder

```
cd Desktop/Alignment_Exercise/Infiles
```

Step 4.2: Type `mafft`. An interactive menu appears to help you select options.

Alternatively, you can use command line here to specify the options.

Step 4.3: If in the Desktop/Alignment_Exercise/Infiles folder, the complete `mafft` command is

```
mafft --thread -1 --globalpair --maxiterate 16 Nectin32Taxa.fasta >
/Users/gdw/Desktop/Alignment_Exercise/Student_files/Nectin4Mafft.fasta
```

Step 4.4: Open output in **GENEIOUS**.

Import-Files-Autodetect format

[Desktop/Alignment_Exercise/Infiles/Nectin4Mafft.fasta](#)

Step 4.5 Locate Indels within the alignment.

Position 43-53

Position 106-108

Position 1241-1252

Translate Alignment using Translation Button on side panel

The amino acid letter is centered under the second position of each codon.

Question: Do the alignment gaps keep codon triplets together in insertion/deletion?

Part 5: Re-Alignment of Selected Regions with GENEIOUS.

5.1 Import Nectin4Mafft.fasta from Student_files

5.2 Tools-Align/Assemble/Multiple Align

Geneious Alignment MUSCLE Alignment Clustal Omega Realign Region

Translation Align Consensus Align

Realign bases 40 to 60

Region alignment options: MUSCLE Alignment

[Muscle 3.8.425 by Robert C. Edgar](#)

☒ Re-align sequences ☐ Refine existing alignment

Maximum number of iterations: 8

☒ Optimize profile-dependent parameters ?

☐ Use custom MUSCLE executable: Change...

Maximum number of trees to build: 1

Sequence Order: ☐ Group sequences by similarity ☒ Preserve sequence order

Optimization: ☐ Diagonal ☒ Anchor

Keep tree from: ☐ Iteration 1 ☐ Iteration 2

Distance measure: Iteration 1: kmer4_6 Subsequent: pctid_kimura

Clustering method: Iteration 1 & 2: Neighbor joini... Subsequent: Neighbor joini...

Tree Rooting method: Iteration 1 & 2: pseudo Subsequent: pseudo

Sequence weighting scheme: Iteration 1 & 2: CLUSTALW Subsequent: CLUSTALW

Terminal gaps: Half penalty Objective score: spm

Anchor spacing: 32 Gap open score: -1

Diagonals: Min length: 24 Margin: 5

Minimum column anchor scores: Min best: 90 Min smoothed: 90

Hydrophobicity: Multiplier: 1.2 Window size: 5

Maximum Memory (MB): 500

☐ Quit and return best alignment found if not finished after 1 hours

Extra Options:

Caption

What happened to INDEL region?

Repeat RE-ALIGNMENT for 3rd INDEL region ~1241.

5.4 Repeat Re-Alignment for MUSCLE and CLUSTAL Alignments.

Summary & Discussion.

Extra Exercise:

Part 6: Other Tools in Alignment Programs

Profile to Profile alignment programs. GENEIOUS will do this easily by selecting the two datasets from panel.

6.1 Import separately 2 files into GENEIOUS:

Alignment_Exercise/Infiles/ Nectin4Carnivores.fasta

Alignment_Exercise/Infiles/ Nectin4Primates.fasta

6.2 Select both aligned files and Align using either MUSCLE or CLUSTAL.

To do this in MAFFT:

6.3: Return to terminal and Infiles folder

```
cd /Users/gdw/Desktop/Alignment_Exercise/Infiles
```

6.4: MAFFT also allows you to add sequences to existing alignments. We can add the carnivores to the alignment using the following commands:

```
mafft --add Nectin4Carnivores.fasta --reorder Nectin4PrimatesMafft.out >  
/Users/gdw/Desktop/Alignment_Exercise/Student_files/Nectin4PrimatesCarnivoresMafft.out
```

6.5: Open Nectin4PrimatesCarnivoresMafft.out in Aliview using commands from Part 1.