

Multiple sequence alignment in R

Philip Woods

Introduction

- Multiple sequence alignment directly from R
- Provided by the msa external library
- IMPORTANT: If first time using script, must install the msa, Biostrings, & seqinr packages ‘install.packages(c(Biostrings, “msa”, “seqinr”))’

```
library(Biostrings)
```

```
## Warning: package 'Biostrings' was built under R version 4.3.3
```

```
## Loading required package: BiocGenerics
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      anyDuplicated, aperm, append, as.data.frame, basename, cbind,  
##      colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,  
##      get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,  
##      match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,  
##      Position, rank, rbind, Reduce, rownames, sapply, setdiff, sort,  
##      table, tapply, union, unique, unsplit, which.max, which.min
```

```
## Loading required package: S4Vectors
```

```
## Loading required package: stats4
```

```
##
```

```
## Attaching package: 'S4Vectors'
```

```
## The following object is masked from 'package:utils':
```

```
##
```

```
##      findMatches
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

## The following object is masked from 'package:grDevices':
##
##     windows

## Loading required package: XVector

## Loading required package: GenomeInfoDb

## Warning: package 'GenomeInfoDb' was built under R version 4.3.3

##
## Attaching package: 'Biostrings'

## The following object is masked from 'package:base':
##
##     strsplit

library(msa)
```

Read in Data and Convert to an Amino Acid StringSet

- Must convert sequences to AAStringSet (or DNAStringSet).
- 'seqtype': the nature of the sequences: 'DNA' or 'AA'
- 'as.string': if 'TRUE' sequences are returned as a string instead of a vector of single characters
- 'forceDNAtolower': whether sequences with 'seqtype == DNA' should be returned as lower case letters
- 'set.attributes': whether sequence attributes should be set

```
# sequences = seqinr::read.fasta(file = file.choose(),
#                               seqtype = "AA",
#                               as.string = T)

sequences = readAAStringSet(file = file.choose())

sequences = AAStringSet(sequences)
```

Perform alignment using ClustalW

- Has access to 3 aligners: 'ClustalW', 'ClustalOmega', or 'MUSCLE'.

```
myAlignment = msa(sequences, "ClustalW")
```

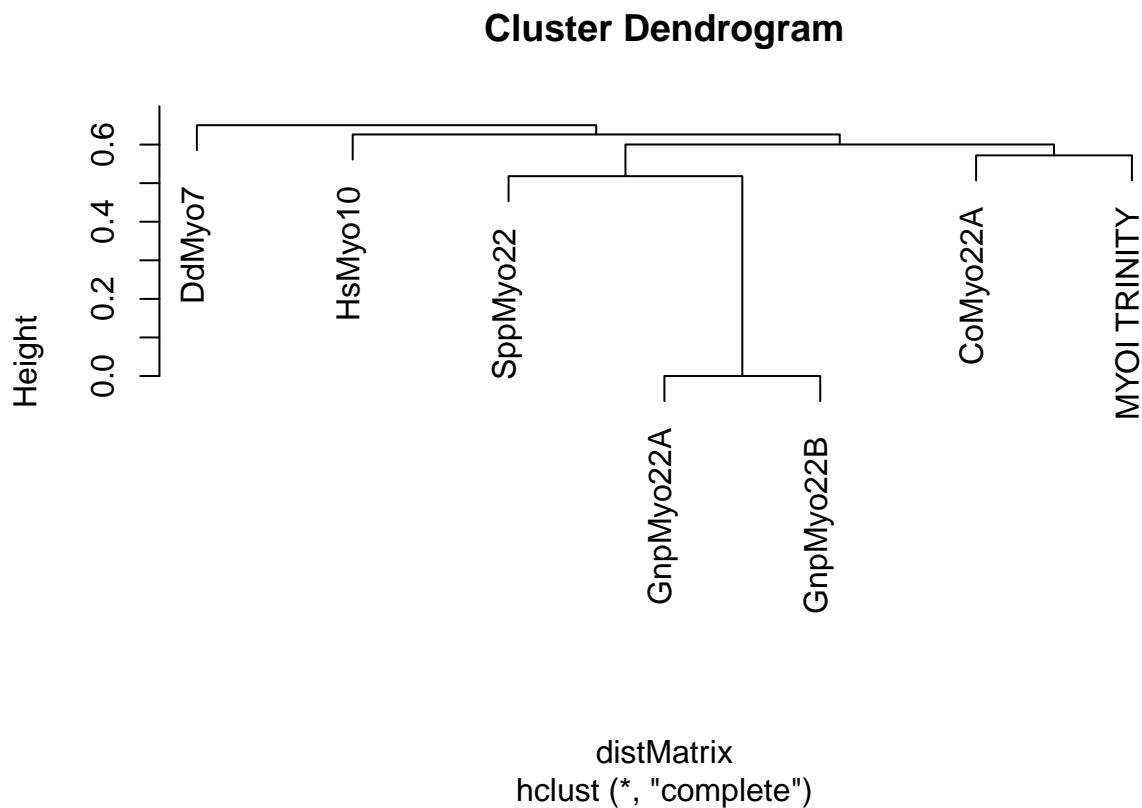
```
## use default substitution matrix
```

Compute distances between sequences using 'seqinr' library

```
# Convert the alignment format to seqinr using msaConvert
alignment = msaConvert(myAlignment, "seqinr::alignment")

# Compute the distance between the alignments
distMatrix = seqinr::dist.alignment(alignment, "similarity")

# Cluster the distance matrix using hclust
clustering = hclust(distMatrix)
plot(clustering)
```

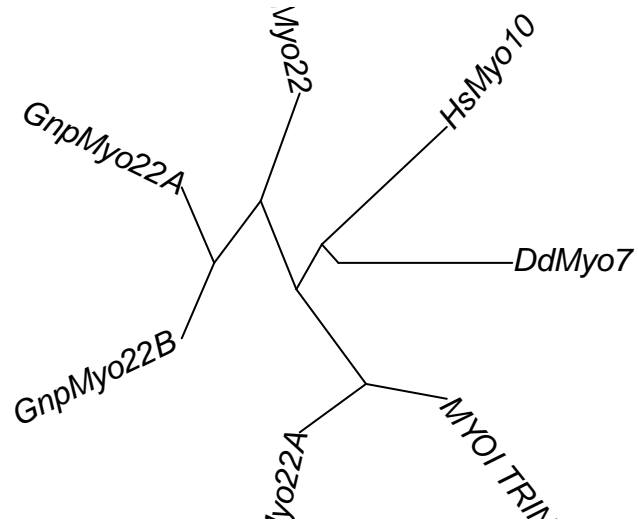


```
# Making Dendrogram look better. Requires 'ape()' package
```

```
# Transform the clustering as a dendrogram object
dendrogram = as.dendrogram(clustering)
```

```
# Transform into phylo object
phylo tree = ape::as.phylo(clustering)
```

```
# plot as radial. Can spice things up, like adding color via. 'tip.color = c("red", "blue", etc....)'
plot(phylo tree, type = "radial")
```



```
# View Alignment Results
```

```
myAlignment
```

```
## CLUSTAL 2.1
##
## Call:
##   msa(sequences, "ClustalW")
##
## MsaAAMultipleAlignment with 7 rows and 2748 columns
##   aln                                     names
## [1] -----...SEPKPAKAEPKESEPIAESEEKAAD CoMyo22A
## [2] -----...----- MYOI TRINITY
## [3] -----...----- GnpMyo22A
## [4] -----...----- GnpMyo22B
## [5] -----...----- SppMyo22
## [6] MDNFFTEGTRVWLRENGQHFPSTVNS...----- HsMyo10
## [7] -----...----- DdMyo7
## Con -----...----- Consensus
```