Sequence alignment

Marc A.T. Teunis

2019-05-05

R & Bioconductor for Bioinformatics - May 2018

## Introduction

Learn:

* How to perform multiple alignment
* How to build graphs for alignment
* How to build a fylogenetic tree
* How and why to build DNAStringset and AAStringSet objects
* Appply sequence alignment in {DECIPHER} package on a new dataset

## Packages

library(msa)  
library(tidyverse)  
library(biomaRt)  
library(annotate)  
library(Biostrings)  
library(seqinr)  
# install.packages('msaR')  
library(DECIPHER)

## Vignettes

browseVignettes("msaR")  
browseVignettes("DECIPHER")  
system.file("tex", "texshade.sty", package="msa")

## Theory

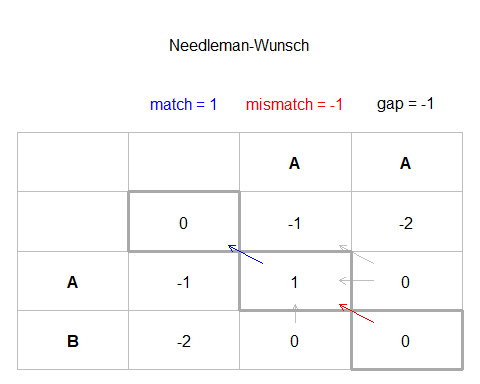
Sequence alignment is important tool to investigate structural, evolutionary and mechanistical simularities in (biological) strings.

One algorithm used to score alignment of strings is the Needleman-Wunsch algorithm. We will explain how this works below. First let’s create a visual representation of the algorithm score matrix

### Needlman-Wunsch alignment score matrix for two sequences

1. seq1 = AB
2. seq2 = AA

source(file = file.path(root, "code", "plot\_needleman\_wunsch.R"))  
  
seq1 <- c("AB")  
seq2 <- c("AA")  
  
  
plotNeedlemanWunsch(seq1 = seq1,  
 seq2 = seq2,  
 1, 1, 1)



### How do we calculate the overall score?

### How does the alignment work?

## Data

Remember the data with the homolog sequences for Caspase-3, -4 and -5 from Human, Orang Utan and Mouse? We will use it here to learn how to use Bioconductor to do multiple sequence alignments.

example\_caspase <- read\_rds(path = file.path(root, "data", "biomaRt\_homology\_example\_aa.rds"))  
example\_caspase\_dna <- read\_rds(path = file.path(root, "data", "biomaRt\_homology\_example\_dna.rds"))

## Filter for Caspase-4

For example sake we use the sequences for Caspase-4 only. To select the right caspases, we need to convert to lower case for all the symbol names

example\_caspase$external\_gene\_name <- tolower(example\_caspase$external\_gene\_name)   
levels(as.factor(example\_caspase$external\_gene\_name))

## [1] "casp3" "casp4" "casp5"

casp4 <- example\_caspase %>%  
 filter(external\_gene\_name == "casp4")  
casp4

## # A tibble: 7 x 4  
## amino\_acid\_sequence ensembl\_gene\_id external\_gene\_na~ species  
## <chr> <chr> <chr> <chr>   
## 1 MADAMKKKHSKVGEMLLQTFFSVDPGSHH~ ENSMUSG00000033~ casp4 M. mus~  
## 2 MAENKHPDKPLKVLEQLGKEVLTEYLEKL~ ENSMUSG00000033~ casp4 M. mus~  
## 3 MAENKHPDKPLKVLEQLGKEVLTEYLEKL~ ENSMUSG00000033~ casp4 M. mus~  
## 4 MADSMQEKQRMAGQMLLQTFFNIDQISPN~ ENSG00000196954 casp4 H. sap~  
## 5 MAEGNHRKKPLKVLESLGKDFLTGVLDNL~ ENSG00000196954 casp4 H. sap~  
## 6 MAEGNHRKKPLKVLESLGKDFLTGVLDNL~ ENSG00000196954 casp4 H. sap~  
## 7 MAEGNHRKKPLKMLESLGKDFLTGVLDNL~ ENSPPYG00000003~ casp4 P. abe~

## AAStringSet

caspase\_aa <- AAStringSet(casp4$amino\_acid\_sequence)  
caspase\_aa

## A AAStringSet instance of length 7  
## width seq  
## [1] 93 MADAMKKKHSKVGEMLLQTFFSVDPGSHHGE...QEIYPIKEANGRTRKALIICNTEFKHLSLRY  
## [2] 374 MAENKHPDKPLKVLEQLGKEVLTEYLEKLVQ...QSFEKASIHSQMPTIDRATLTRYFYLFPGN\*  
## [3] 119 MAENKHPDKPLKVLEQLGKEVLTEYLEKLVQ...LPNKGGQWPYTKGSYHMQYRVQTSLTEVWG\*  
## [4] 322 MADSMQEKQRMAGQMLLQTFFNIDQISPNKK...QSFETPRAKAQMPTIERLSMTRYFYLFPGN\*  
## [5] 378 MAEGNHRKKPLKVLESLGKDFLTGVLDNLVE...QSFETPRAKAQMPTIERLSMTRYFYLFPGN\*  
## [6] 147 MAEGNHRKKPLKVLESLGKDFLTGVLDNLVE...LCKERAEEIYPIKERNNRTRLALIICNTEFD  
## [7] 378 MAEGNHRKKPLKMLESLGKDFLTGVLDNLVE...QSFETPRAKAQMPTIERLSMTRYFYLFPGN\*

## Set names to AAStringSet sequences

names(caspase\_aa) <- paste(casp4$species, casp4$ensembl\_gene\_id, sep = "\_")  
#duplicated(casp5$coding[1:42])  
caspase\_aa

## A AAStringSet instance of length 7  
## width seq names   
## [1] 93 MADAMKKKHSKVGEMLLQTFF...RTRKALIICNTEFKHLSLRY M. musculus\_ENSMU...  
## [2] 374 MAENKHPDKPLKVLEQLGKEV...MPTIDRATLTRYFYLFPGN\* M. musculus\_ENSMU...  
## [3] 119 MAENKHPDKPLKVLEQLGKEV...KGSYHMQYRVQTSLTEVWG\* M. musculus\_ENSMU...  
## [4] 322 MADSMQEKQRMAGQMLLQTFF...MPTIERLSMTRYFYLFPGN\* H. sapiens\_ENSG00...  
## [5] 378 MAEGNHRKKPLKVLESLGKDF...MPTIERLSMTRYFYLFPGN\* H. sapiens\_ENSG00...  
## [6] 147 MAEGNHRKKPLKVLESLGKDF...IKERNNRTRLALIICNTEFD H. sapiens\_ENSG00...  
## [7] 378 MAEGNHRKKPLKMLESLGKDF...MPTIERLSMTRYFYLFPGN\* P. abelii\_ENSPPYG...

# Aligment of sequences, using the {msa} package

## Using the {msa} package

caspase\_align <- msa::msa(caspase\_aa)

## use default substitution matrix

caspase\_align

## CLUSTAL 2.1   
##   
## Call:  
## msa::msa(caspase\_aa)  
##   
## MsaAAMultipleAlignment with 7 rows and 430 columns  
## aln names  
## [1] MAEGNHRKKPLKVLESLGKDFLT...----------------------- H. sapiens\_ENSG00...  
## [2] MAEGNHRKKPLKMLESLGKDFLT...KAQMPTIERLSMTRYFYLFPGN- P. abelii\_ENSPPYG...  
## [3] -----------------------...KAQMPTIERLSMTRYFYLFPGN- H. sapiens\_ENSG00...  
## [4] MAEGNHRKKPLKVLESLGKDFLT...KAQMPTIERLSMTRYFYLFPGN- H. sapiens\_ENSG00...  
## [5] -----------------------...HSQMPTIDRATLTRYFYLFPGN- M. musculus\_ENSMU...  
## [6] -----------------------...----------------------- M. musculus\_ENSMU...  
## [7] -----------------------...----------------------- M. musculus\_ENSMU...  
## Con -----------------------...??QMPTI?R???TRYFYLFPGN- Consensus

## Print alignment

print(caspase\_align, show="complete")

##   
## MsaAAMultipleAlignment with 7 rows and 430 columns  
## aln (1..49) names  
## [1] MAEGNHRKKPLKVLESLGKDFLTGVLDNLVEQNVLNWKEEEKKKYYDAK H. sapiens\_ENSG00...  
## [2] MAEGNHRKKPLKMLESLGKDFLTGVLDNLVEQNVLNWKEEEKKKYYDAK P. abelii\_ENSPPYG...  
## [3] ------------------------------------------------- H. sapiens\_ENSG00...  
## [4] MAEGNHRKKPLKVLESLGKDFLTGVLDNLVEQNVLNWKEEEKKKYYDAK H. sapiens\_ENSG00...  
## [5] ------------------------------------------------- M. musculus\_ENSMU...  
## [6] ------------------------------------------------- M. musculus\_ENSMU...  
## [7] ------------------------------------------------- M. musculus\_ENSMU...  
## Con ------------------------------------------------- Consensus   
##   
## aln (50..98) names  
## [1] TEDKVRVMADSMQ-EKQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPP H. sapiens\_ENSG00...  
## [2] TEDKVRVMADSIQ-EKQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPP P. abelii\_ENSPPYG...  
## [3] -------MADSMQ-EKQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPP H. sapiens\_ENSG00...  
## [4] TEDKVRVMADSMQ-EKQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPP H. sapiens\_ENSG00...  
## [5] -------MAENKHPDKPLKVLEQLGKEVLTEYLEKLVQSNVLKLKEEDK M. musculus\_ENSMU...  
## [6] -------MAENKHPDKPLKVLEQLGKEVLTEYLEKLVQSNVLKLKEEDK M. musculus\_ENSMU...  
## [7] -------MADAMK-KKHSKVGEMLLQTFFSVDPGSHHGEANLEMEEPE- M. musculus\_ENSMU...  
## Con -------MADSMQ-EKQRMAGQMLLQTFFNIDQISPNKKAHPNMEAGPP Consensus   
##   
## aln (99..147) names  
## [1] ESGESTDALKLCP------HEEFLRLCKERAEEIYPIKERNNRTRLALI H. sapiens\_ENSG00...  
## [2] ESGESTDALKLCP------HEEFLRLCKERAEEIYPIKERNNRTRLALI P. abelii\_ENSPPYG...  
## [3] ESGESTDALKLCP------HEEFLRLCKERAEEIYPIKERNNRTRLALI H. sapiens\_ENSG00...  
## [4] ESGESTDALKLCP------HEEFLRLCKERAEEIYPIKERNNRTRLALI H. sapiens\_ENSG00...  
## [5] QKFNNAERSDKRWVFVDAMKKKHSKVGEMLLQTFFSVDPGSHHGEANLE M. musculus\_ENSMU...  
## [6] QKFNNAERSDKRWVFVDAMKKKHSKVGEMLLQTFFSVDPGSHHG----- M. musculus\_ENSMU...  
## [7] ESLNTLKLCSP---------EEFTRLCREKTQEIYPIKEANGRT----- M. musculus\_ENSMU...  
## Con ESGESTDALKLCP------HEEFLRLCKERAEEIYPIKERNNRTRLALI Consensus   
##   
## aln (148..196) names  
## [1] ICNTEFD------------------------------------------ H. sapiens\_ENSG00...  
## [2] ICNTEFDHLPPRNGADFDITG---------------------------- P. abelii\_ENSPPYG...  
## [3] ICNTEFDHLPPRNGADFDITG---------------------------- H. sapiens\_ENSG00...  
## [4] ICNTEFDHLPPRNGADFDITG---------------------------- H. sapiens\_ENSG00...  
## [5] MEEPEESLNTLKLCSPEEFTRLCREKTQEIYPIKEANGRTRKALIICNT M. musculus\_ENSMU...  
## [6] ------------------------------------------------- M. musculus\_ENSMU...  
## [7] ------------------------------------------------- M. musculus\_ENSMU...  
## Con ICNTEFD????????????T?---------------------------- Consensus   
##   
## aln (197..245) names  
## [1] ------------------------------------------------- H. sapiens\_ENSG00...  
## [2] -----------------MKELLEGLDYSVDVEENLTARDMESALRAFAA P. abelii\_ENSPPYG...  
## [3] -----------------MKELLEGLDYSVDVEENLTARDMESALRAFAT H. sapiens\_ENSG00...  
## [4] -----------------MKELLEGLDYSVDVEENLTARDMESALRAFAT H. sapiens\_ENSG00...  
## [5] EFKHLSLRYGANFDIIGMKGLLEDLGYDVVVKEELTAEGMESEMKDFAA M. musculus\_ENSMU...  
## [6] ----------------------EDL------------------------ M. musculus\_ENSMU...  
## [7] ------------------------------------------------- M. musculus\_ENSMU...  
## Con -----------------MK?LLE?L?Y?V?V?E?LTA??MES????FA? Consensus   
##   
## aln (246..294) names  
## [1] ------------------------------------------------- H. sapiens\_ENSG00...  
## [2] RPEHKSSDSTFLVLMSHGILEGICGTVHDEKKPDVLLYDTIFQIFNNRN P. abelii\_ENSPPYG...  
## [3] RPEHKSSDSTFLVLMSHGILEGICGTVHDEKKPDVLLYDTIFQIFNNRN H. sapiens\_ENSG00...  
## [4] RPEHKSSDSTFLVLMSHGILEGICGTVHDEKKPDVLLYDTIFQIFNNRN H. sapiens\_ENSG00...  
## [5] LSEHQTSDSTFLVLMSHGTLHGICGTMHSEKTPDVLQYDTIYQIFNNCH M. musculus\_ENSMU...  
## [6] ------------------------------------------------- M. musculus\_ENSMU...  
## [7] ------------------------------------------------- M. musculus\_ENSMU...  
## Con ??EH??SDSTFLVLMSHG?L?GICGT?H?EK?PDVL?YDTI?QIFNN?? Consensus   
##   
## aln (295..343) names  
## [1] ------------------------------------------------- H. sapiens\_ENSG00...  
## [2] CLSLKDKPKVIIVQACRGANRGELWVRDSPASLEVASSQSPENLEEDAV P. abelii\_ENSPPYG...  
## [3] CLSLKDKPKVIIVQACRGANRGELWVRDSPASLEVASSQSSENLEEDAV H. sapiens\_ENSG00...  
## [4] CLSLKDKPKVIIVQACRGANRGELWVRDSPASLEVASSQSSENLEEDAV H. sapiens\_ENSG00...  
## [5] CPGLRDKPKVIIVQACRGGNSGEMWIRESSKPQLCRGVDLPRNMEADAV M. musculus\_ENSMU...  
## [6] -P-----------------NKGGQWP----------------------- M. musculus\_ENSMU...  
## [7] -------------------RKALII------------------------ M. musculus\_ENSMU...  
## Con C??L?DKPKVIIVQACRG?N?GE?W?R?S?????????????N?E?DAV Consensus   
##   
## aln (344..392) names  
## [1] ------------------------------------------------- H. sapiens\_ENSG00...  
## [2] YKTHVEKDFIAFCSSTPHNVSWRDSTMGSIFITQLITCFQKYSWCCHLE P. abelii\_ENSPPYG...  
## [3] YKTHVEKDFIAFCSSTPHNVSWRDSTMGSIFITQLITCFQKYSWCCHLE H. sapiens\_ENSG00...  
## [4] YKTHVEKDFIAFCSSTPHNVSWRDSTMGSIFITQLITCFQKYSWCCHLE H. sapiens\_ENSG00...  
## [5] KLSHVEKDFIAFYSTTPHHLSYRDKTGGSYFITRLISCFRKHACSCHLF M. musculus\_ENSMU...  
## [6] ------------YTKGSYHMQYRVQTS----LTEVWG------------ M. musculus\_ENSMU...  
## [7] ------------CNTEFKHLSLRY------------------------- M. musculus\_ENSMU...  
## Con ???HVEKDFIAFCS?TPH??S?RD?T?GS?FIT?LI?CF?K????CHL? Consensus   
##   
## aln (393..430) names  
## [1] -------------------------------------- H. sapiens\_ENSG00...  
## [2] EVFRKVQQSFETPRAKAQMPTIERLSMTRYFYLFPGN- P. abelii\_ENSPPYG...  
## [3] EVFRKVQQSFETPRAKAQMPTIERLSMTRYFYLFPGN- H. sapiens\_ENSG00...  
## [4] EVFRKVQQSFETPRAKAQMPTIERLSMTRYFYLFPGN- H. sapiens\_ENSG00...  
## [5] DIFLKVQQSFEKASIHSQMPTIDRATLTRYFYLFPGN- M. musculus\_ENSMU...  
## [6] -------------------------------------- M. musculus\_ENSMU...  
## [7] -------------------------------------- M. musculus\_ENSMU...  
## Con ??F?KVQQSFE??????QMPTI?R???TRYFYLFPGN- Consensus

## Pretty print aligment

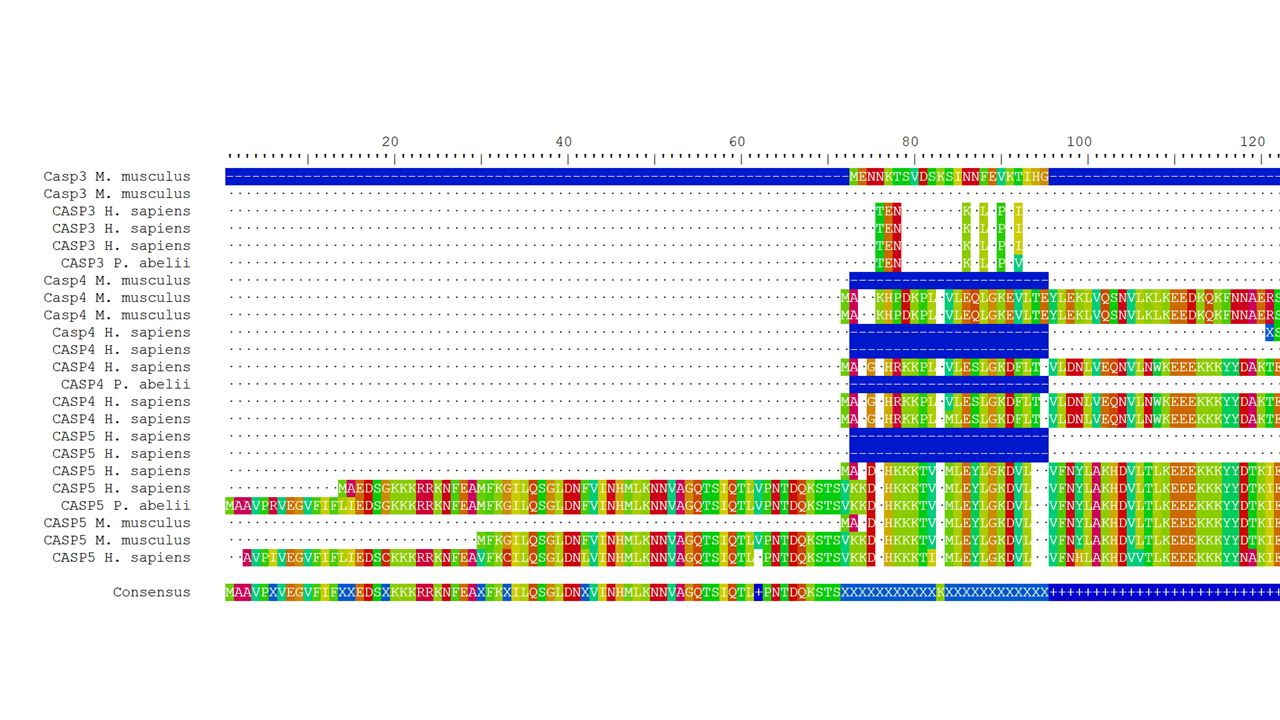
## does not work on my pc  
msaPrettyPrint(caspase\_align, output="", y=c(10, 30), showNames="none",  
 showLogo="none", askForOverwrite=FALSE, verbose=FALSE)

# Alignment using the {DECIPHER} package

## Amino acids

AA <- AlignSeqs(caspase\_aa, processors = 8) # align the translation  
BrowseSeqs(AA, highlight=1) # view the alignment

## Alignment in browser



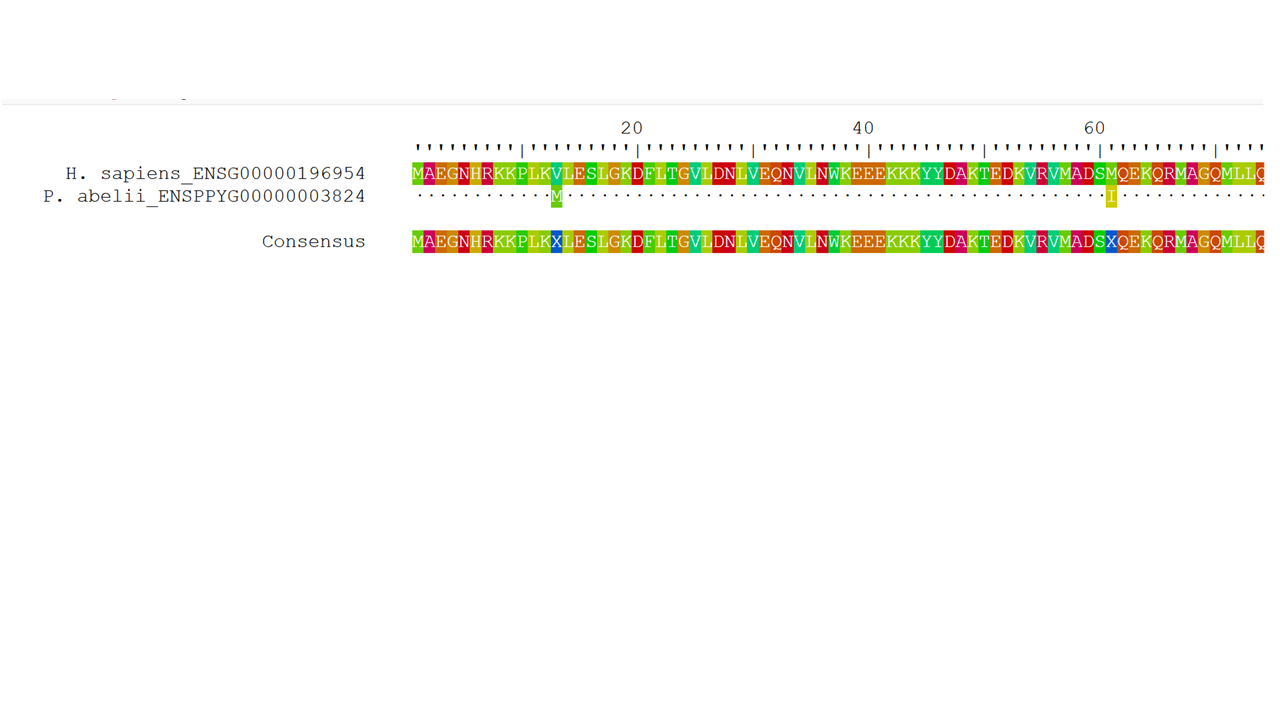
## Subsetting an alignment

Subsetting an alignment works the same as for ranges. First we select the sequences that we want subset, using normal indexing with [], than we use the function Biostrings::subseq()

Let’s look at the sequence 5 for Human and 7 for Orang Utan only

caspase\_aa  
AA\_subs <- Biostrings::subseq(caspase\_aa[c(5, 7),], start = 1, end = 158)  
  
AA\_subs <- AlignSeqs(AA\_subs, processors = 8) # align the translation  
BrowseSeqs(AA\_subs, highlight=1) # view the alignment

## Alignment in browser

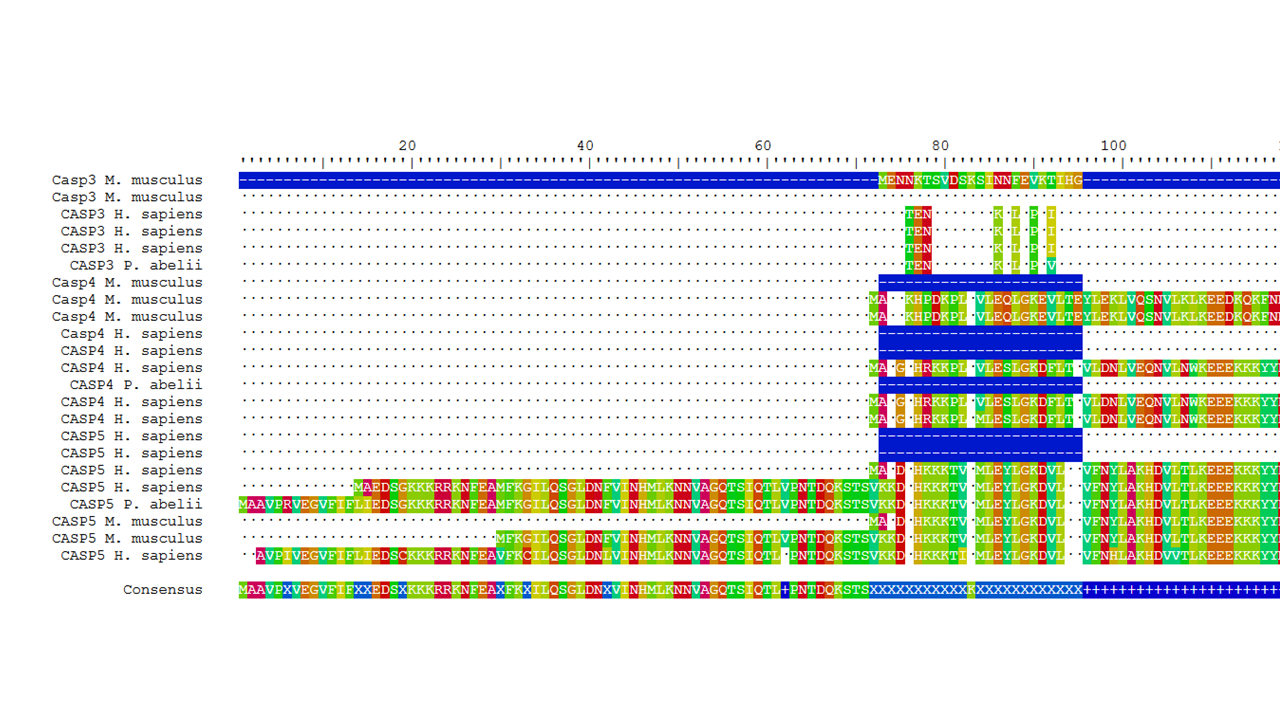


## Aligning Amino acid sequences from DNA sequences

To align a translation from DNA use the AlignTranslation() function from {DECIPHER}

dna\_strings <- DNAStringSet(example\_caspase\_dna$coding)  
names(dna\_strings) <- paste(example\_caspase\_dna$external\_gene\_name,   
 example\_caspase$species)  
  
AA <- AlignTranslation(dna\_strings, type="AAStringSet") # align the translation  
BrowseSeqs(AA, highlight=1) # view the alignment

## Alignment in browser



## Caspase-4 DNA

Let’s look at the DNA alignments for the caspase-4 human and oran-utan subset above

casp4\_dna <- example\_caspase\_dna %>%  
 dplyr::mutate(external\_gene\_name = tolower(external\_gene\_name)) %>%  
 dplyr::filter(species == "H. sapiens" | species == "P. abelii") %>%  
 dplyr::filter(external\_gene\_name == "casp4") %>%  
 print()

## # A tibble: 5 x 4  
## coding ensembl\_gene\_id external\_gene\_n~ species   
## <chr> <chr> <chr> <chr>   
## 1 ATGGCAGACTCTATGCAAGAGAAGCAACG~ ENSG00000196954 casp4 H. sapi~  
## 2 ATGGCAGAAGGCAACCACAGAAAAAAGCC~ ENSG00000196954 casp4 H. sapi~  
## 3 NACAACGTGTCCTGGAGAGACAGCACAAT~ ENSG00000196954 casp4 H. sapi~  
## 4 ATGGCAGAAGGCAACCACAGAAAAAAGCC~ ENSG00000196954 casp4 H. sapi~  
## 5 ATGGCAGAAGGCAACCACAGAAAAAAACC~ ENSPPYG00000003~ casp4 P. abel~

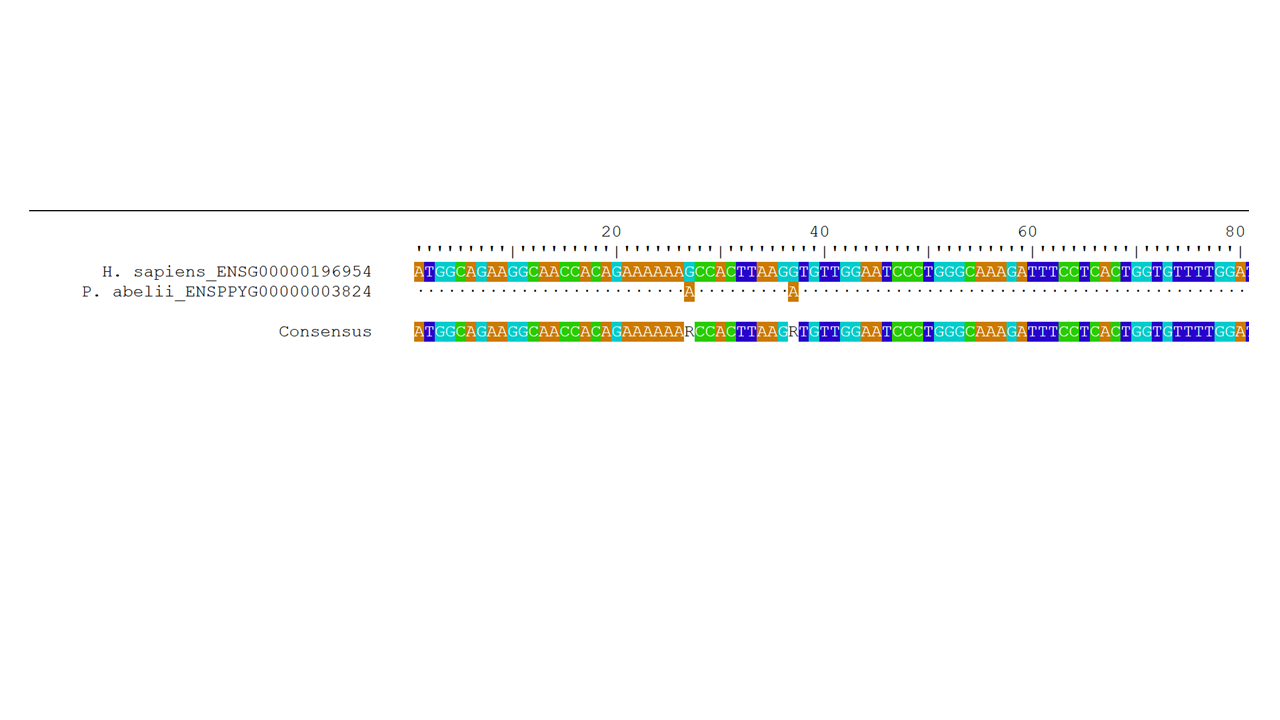
dna\_subset <- DNAStringSet(casp4\_dna$coding)  
names(dna\_subset) <- paste(casp4\_dna$species, casp4\_dna$ensembl\_gene\_id, sep = "\_")

## Aligning Human and Orang Utan sequences

casp\_dna\_subset <- AlignSeqs(dna\_subset[c(4,5),], processors = 8)  
BrowseSeqs(casp\_dna\_subset, highlight=1) # view the alignment

How do the DNA and Amino acid aligment relate?

## Alignment in browser



## Inspecting alignment Quality

* Look at the alignment
* BrowseSeqs()

BrowseSeqs(caspase\_aa[c(5, 7),], highlight = 0)

## Fylogenetic alignment (Evolutionary relationships)

We create a fylogeny for Caspase-4, using the DNA from Human, Orang Utan and Mouse. Below we filter the DNA dataset for the relevant data

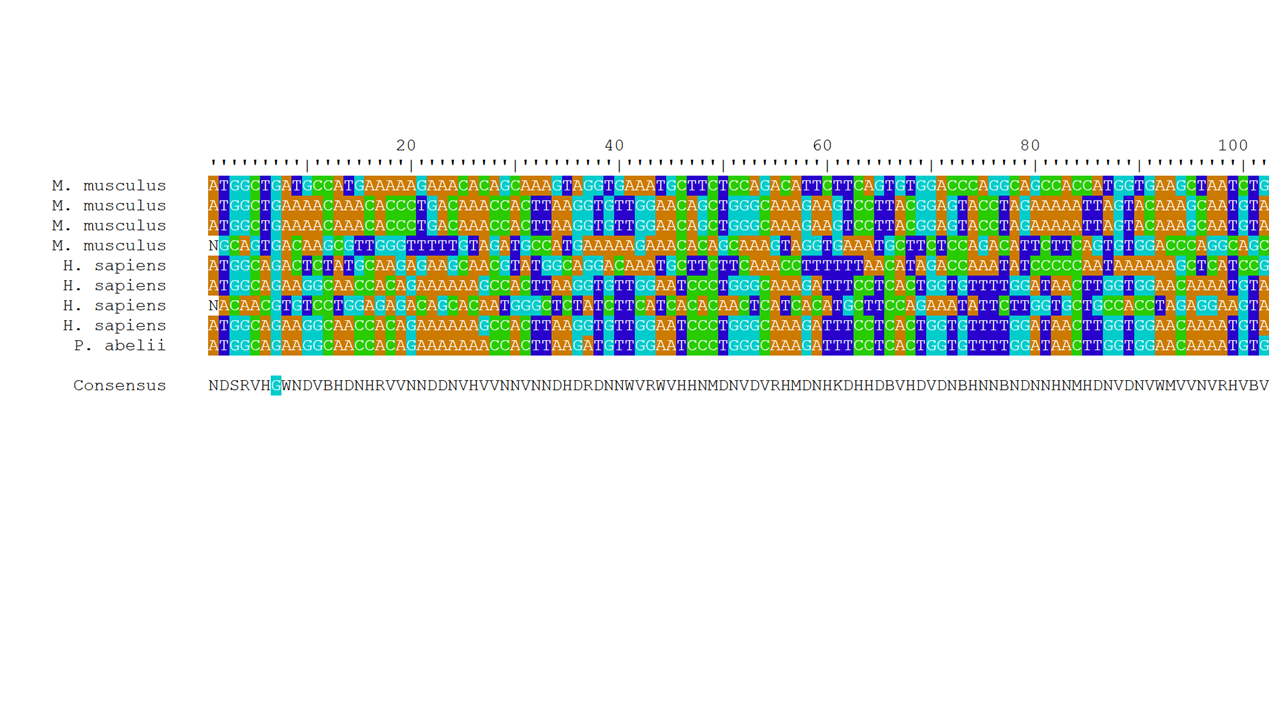
mouse\_human\_orang\_dna <- example\_caspase\_dna %>%  
 dplyr::mutate(external\_gene\_name = tolower(external\_gene\_name)) %>%  
 dplyr::filter(external\_gene\_name == "casp4") %>%  
 print

## # A tibble: 9 x 4  
## coding ensembl\_gene\_id external\_gene\_n~ species   
## <chr> <chr> <chr> <chr>   
## 1 ATGGCTGATGCCATGAAAAAGAAACACA~ ENSMUSG000000335~ casp4 M. musc~  
## 2 ATGGCTGAAAACAAACACCCTGACAAAC~ ENSMUSG000000335~ casp4 M. musc~  
## 3 ATGGCTGAAAACAAACACCCTGACAAAC~ ENSMUSG000000335~ casp4 M. musc~  
## 4 NGCAGTGACAAGCGTTGGGTTTTTGTAG~ ENSMUSG000000335~ casp4 M. musc~  
## 5 ATGGCAGACTCTATGCAAGAGAAGCAAC~ ENSG00000196954 casp4 H. sapi~  
## 6 ATGGCAGAAGGCAACCACAGAAAAAAGC~ ENSG00000196954 casp4 H. sapi~  
## 7 NACAACGTGTCCTGGAGAGACAGCACAA~ ENSG00000196954 casp4 H. sapi~  
## 8 ATGGCAGAAGGCAACCACAGAAAAAAGC~ ENSG00000196954 casp4 H. sapi~  
## 9 ATGGCAGAAGGCAACCACAGAAAAAAAC~ ENSPPYG000000038~ casp4 P. abel~

## Prepare alignment for fylogeny

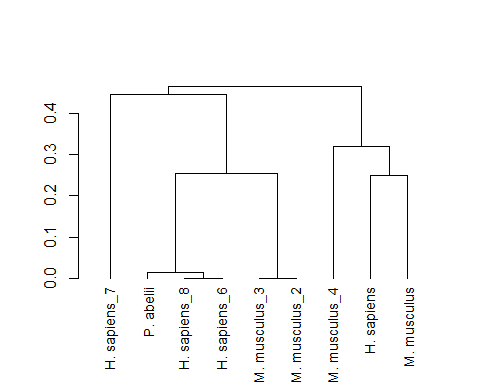
mouse\_human\_orang\_dna\_string\_set <- DNAStringSet(mouse\_human\_orang\_dna$coding,   
 start = 1, end = 210)  
names(mouse\_human\_orang\_dna\_string\_set) <- mouse\_human\_orang\_dna$species  
  
alignment\_aa <- AlignTranslation(mouse\_human\_orang\_dna\_string\_set,   
 type = "AAStringSet", processors = 8)  
staggered <- StaggerAlignment(alignment\_aa)  
BrowseSeqs(staggered)

## Alignment in browser



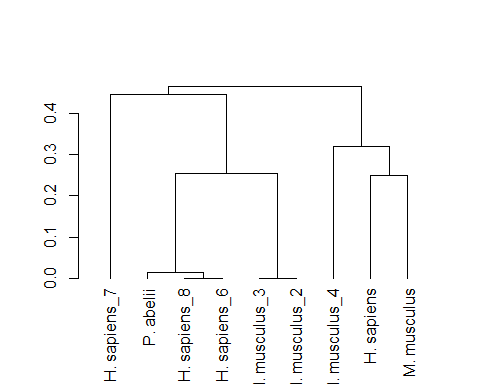
## Fylogenetic tree from distance matrix

distance\_matrix <- DistanceMatrix(staggered)  
dendrogram <- IdClusters(distance\_matrix, showPlot = TRUE, type = 'both')



## Accessing the dendrogram plot

plot(dendrogram[[2]])



## Save the plot

png(filename = "./images/dendrogram.png")  
plot(plot[[2]])  
dev.off()

NEEDS WORK

## How could we come to a single sequence per species?

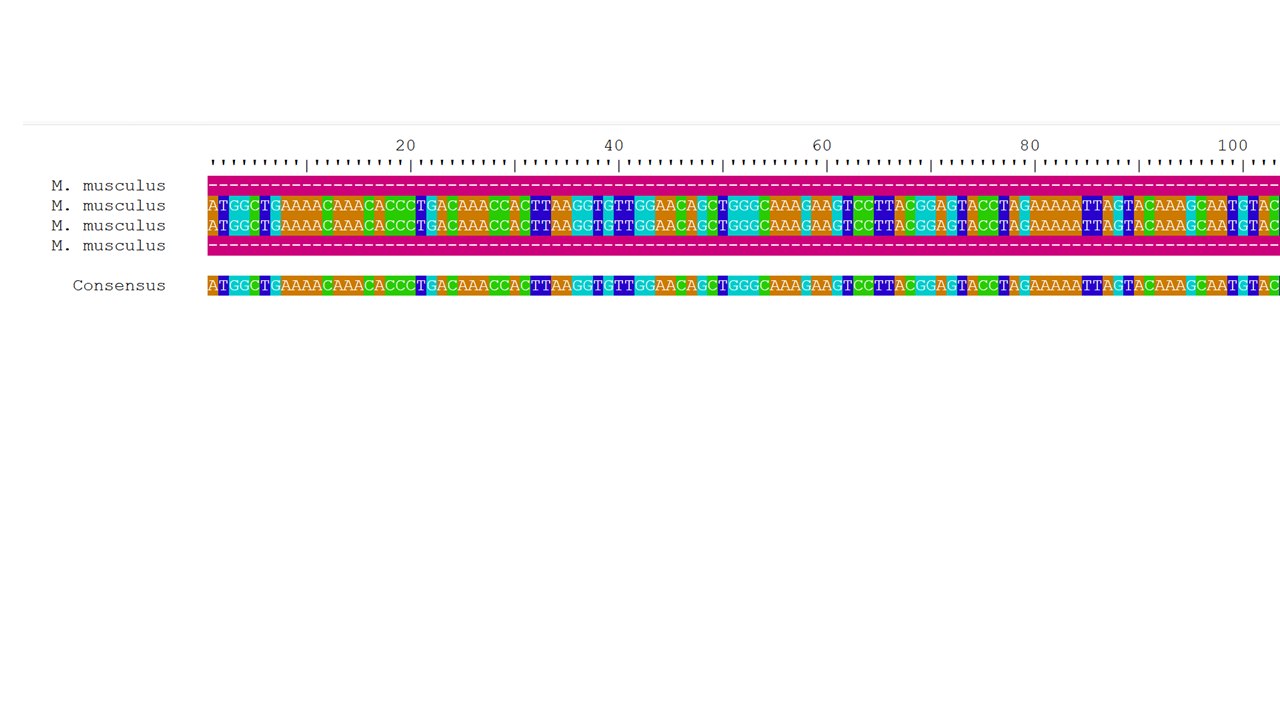
* Align sequences, isolate consensus
* Align Human sequences, isolate consensus
* Align human mouse consensus to Orang Utan sequence,

## Towards a single sequence for each species

### Mouse

staggered  
  
# mouse  
mouse\_only <- AlignTranslation(mouse\_human\_orang\_dna\_string\_set[c(1:4),],  
 type = "AAStringSet")  
  
mouse\_stagger <- StaggerAlignment(mouse\_only)  
  
mouse\_consensus <- ConsensusSequence(mouse\_stagger)  
BrowseSeqs(mouse\_stagger)

## Alignment in browser



## Human

human\_only <- AlignTranslation(mouse\_human\_orang\_dna\_string\_set[c(5:8)],  
 type = "AAStringSet")  
human\_stagger <- StaggerAlignment(human\_only)  
  
  
# CorrectFrameshifts(human\_only)  
human\_consensus <- ConsensusSequence(human\_stagger)   
BrowseSeqs(human\_stagger)

## Alignment in browser



# orang utan

orang\_only <- mouse\_human\_orang\_dna\_string\_set$`P. abelii` %>%   
 Biostrings::translate() %>%   
 AAStringSet() %>%  
 print()

## prepare data for alignment

all\_aa <- c(human\_consensus, mouse\_consensus, orang\_only)  
  
names(all\_aa) <- c("mouse", "human", "orang utan")  
  
equal\_length <- AAStringSet(all\_aa, start = 1, end = 8)  
  
#all\_aa\_aligned <- AlignSeqs(equal\_length, type = "AAStringSet", processors = 8)  
  
#staggered <- StaggerAlignment(all\_aa\_aligned)  
#BrowseSeqs(staggered)

## Build new tree

distance\_matrix <- DistanceMatrix(all\_aa\_aligned)  
dendrogram <- IdClusters(distance\_matrix, showPlot = TRUE, type = 'both')

## Dendrogram of single sequence per species

Why has the distance between Human and Orang Utan increased?

dendrogram[[2]] %>% plot()

## Aligning RNA sequences

16S ribosomal rRNA is used to characterize bacteria. Here we show an example with the build-in dataset from {DECIPHER}

# database containing 16S ribosomal RNA sequences  
db <- system.file("extdata", "Bacteria\_175seqs.sqlite", package="DECIPHER")  
rna <- SearchDB(db, remove="all", type="RNAStringSet")

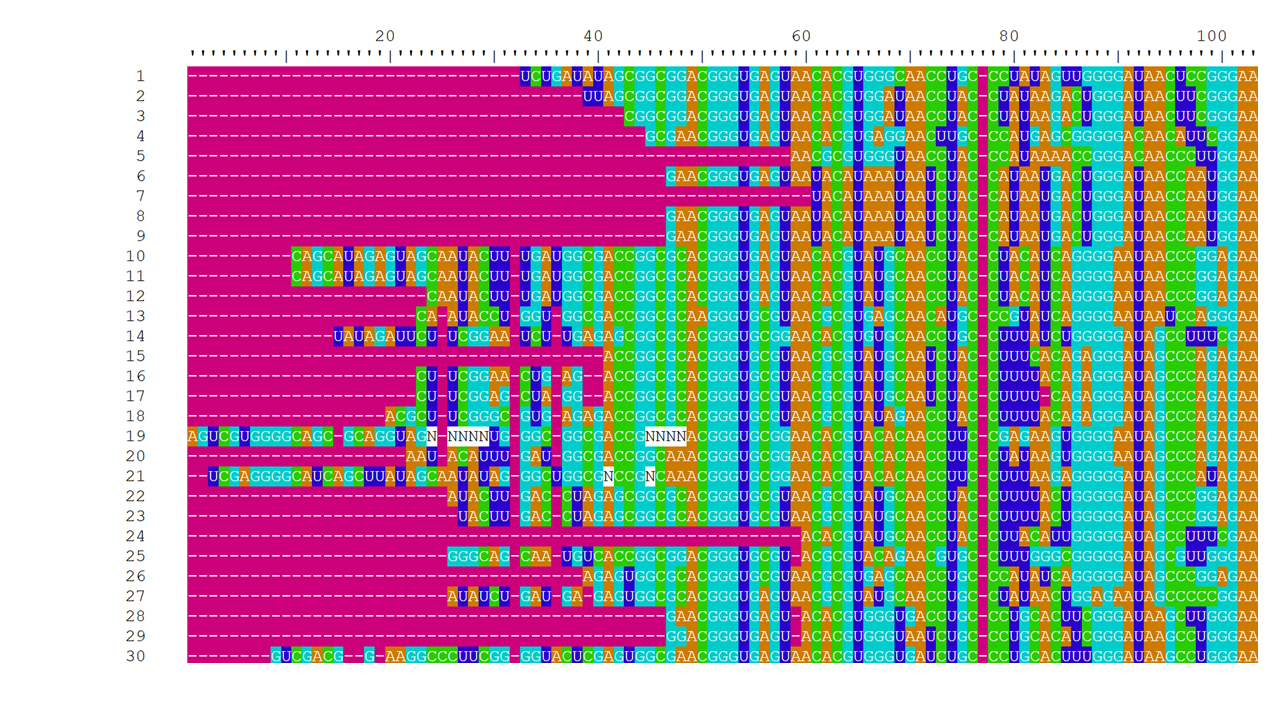
## Search Expression:  
## select row\_names, sequence from \_Seqs where row\_names in (select  
## row\_names from Seqs)  
##   
## RNAStringSet of length: 175  
## Time difference of 0.03 secs

alignedRNA <- AlignSeqs(rna, processors = 8, verbose = FALSE) # align with RNA secondary structure

## Viewing the alignment

BrowseSeqs(alignedRNA)

## Alignment in browser



## Bacterial fylogenetic tree

distance\_matrix <- DistanceMatrix(StaggerAlignment(alignedRNA), verbose = FALSE)  
bacteria <- IdClusters(distance\_matrix, showPlot = TRUE, type = 'dendrogram')

## Bacterial dendrogram (fylogenetic tree)

bacteria[[2]] %>% plot()

