

# The journey of Multiple Sequence Alignment

Diving in the Notum

Constantinos Yeles (Konstantinos Geles)

First version: Tue 26/07/2021, Latest update: Tue 27/Jul/2021

## Contents

<b>Blast alignment</b>	<b>1</b>
<b>Import results into R</b>	<b>2</b>
print an example of the results	2
get the 10 top hits identifiers	4
import fasta db	4
<b>Multiple Sequence Alignment</b>	<b>4</b>
print the MSA	4

## Blast alignment

First we need to download the fasta sequence and the database from dropbox? Maybe can be uploaded later to zenodo (probably I will have more context to add later).

We will use local blast run on docker as the link wasn't working.

```
# get the docker and run
docker pull ncbi/blast
docker run --rm -ti -v $(pwd):/home/my_data ncbi/blast
# move inside /home dir of the container
cd /home

# create a folder for the output of the blast database and unzip the fasta sequences
mkdir my_data/some_species_blast_db
gunzip my_data/oma_db.gz

# make a blast database
makeblastdb -dbtype "prot" -in "my_data/oma_db" -input_type "fasta" -out "my_data/some_species_blast_db"
```

```
# blast the sequence of Notum
blastp -query "my_data/dm_notum.fasta" -db "my_data/some_species_blast_db/blast_oma.db" -out "my_data/
```

## Import results into R

print an example of the results

```
notum_blst %>% arrange(evalue)
```

```
## GRanges object with 197 ranges and 26 metadata columns:
##               seqnames      ranges strand |      query_id query_length
##               <Rle> <IRanges> <Rle> | <character>      <numeric>
##    [1]                Dm|Notum      1-671    + |      Dm|Notum          671
##    [2]      Bf|XP_002599183.1    86-411    + |      Dm|Notum          671
##    [3]                La|919081898    67-400    + |      Dm|Notum          671
##    [4]                Sk|269784925    68-450    + |      Dm|Notum          671
##    [5]                Ct|ELT91091.1    48-379    + |      Dm|Notum          671
##    ...                ...            ...    ... |      ...            ...
##   [193]      sb|3460350|      526-652    + |      Dm|Notum          671
##   [194]      Ta|XP_002109594.1     32-146    + |      Dm|Notum          671
##   [195]      Cg|762136134  9401-9532    + |      Dm|Notum          671
##   [196]      d_gig|g110.t1     408-437    + |      Dm|Notum          671
##   [197] a_aur|scaffold226.g1.. 7906-8083    + |      Dm|Notum          671
##               subject_id subject_length  q_start    q_end    s_start
##               <character>      <numeric> <numeric> <numeric> <numeric>
##    [1]                Dm|Notum          671         1      671         1
##    [2]      Bf|XP_002599183.1          487        88      412        86
##    [3]                La|919081898          559        88      418        67
##    [4]                Sk|269784925          508        88      470        68
##    [5]                Ct|ELT91091.1          462        88      418        48
##    ...                ...            ...    ...    ...    ...
##   [193]      sb|3460350|          3301       435      564       526
##   [194]      Ta|XP_002109594.1          1016       412      538        32
##   [195]      Cg|762136134          9959       443      579      9401
##   [196]      d_gig|g110.t1           467       598      627       408
##   [197] a_aur|scaffold226.g1..          8310       417      593      7906
##               s_end      query_seq      subject_seq      evalue
##               <numeric>      <character>      <character> <numeric>
##    [1]          671 MAVEQIDKMAAKAGEATNKW.. MAVEQIDKMAAKAGEATNKW.. 0.00e+00
##    [2]          411 LKRANLANTSITCNDGSHAG.. MKLHKLRLNTSVTCNDGSPAG.. 4.55e-121
##    [3]          400 LKRANLANTSITCNDGSHAG.. LKRHFLLTNRVTVCNDGSPAG.. 1.01e-119
##    [4]          450 LKRANLANTSITCNDGSHAG.. MKLRYLENTTVTCNDGSPAG.. 2.01e-113
##    [5]          379 LKRANLANTSITCNDGSHAG.. MKRHFIRNPSVTCNDGSKAG.. 7.05e-108
##    ...                ...            ...    ...    ...
##   [193]          652 EHANNQRHQRHRQLRQKH.. EQRTENREQRTENREQRTEN.. 5.2
##   [194]          146 STRSRRHDKLKRSTEPSTAV.. SERRRNHDDDKGSSRHARKE.. 6.5
##   [195]         9532 QRHRQLRQKHNVAQSGG.. ERKRKMLEKYKQLEEELEAE.. 6.9
##   [196]          437 CGLRLLERCSWPQCNHSCPT.. CKSALIDDCSQPLCNLCPLA.. 7.2
##   [197]         8083 RHDKLKRSTEPSTAVSHEPH.. RKRREKKKGEEVEHGSDKEE.. 8.4
```

```

##          bit_score      score alignment_length percent_identity identical
##          <numeric> <numeric>          <numeric>          <numeric> <numeric>
##      [1]         1400         3625             671          100.000         671
##      [2]          373          957             333           53.453         178
##      [3]          372          954             339           51.327         174
##      [4]          353          907             396           44.697         177
##      [5]          338          866             338           48.817         165
##      ...           ...           ...           ...           ...           ...
##     [193]         35.0           79             130           23.846          31
##     [194]         34.7           78             127           29.921          38
##     [195]         34.7           78             141           27.660          39
##     [196]         34.3           77              30           46.667          14
##     [197]         34.3           77             182           19.231          35
##          mismatches positives percent_positives query_sbjct_frames query_frame
##          <numeric> <numeric>          <numeric>          <character> <numeric>
##      [1]           0         671           100.00             1/1           1
##      [2]          140         231           69.37             1/1           1
##      [3]          152         238           70.21             1/1           1
##      [4]          193         244           61.62             1/1           1
##      [5]          160         224           66.27             1/1           1
##      ...           ...           ...           ...           ...           ...
##     [193]          96          57           43.85             1/1           1
##     [194]          77          62           48.82             1/1           1
##     [195]          89          70           49.65             1/1           1
##     [196]          16          17           56.67             1/1           1
##     [197]         138          74           40.66             1/1           1
##          sbjct_frame subject_strand percent_query_coverage_per_subject
##          <numeric>      <character>          <numeric>
##      [1]           1             N/A             100
##      [2]           1             N/A             59
##      [3]           1             N/A             60
##      [4]           1             N/A             66
##      [5]           1             N/A             63
##      ...           ...           ...           ...
##     [193]           1             N/A             23
##     [194]           1             N/A             19
##     [195]           1             N/A             20
##     [196]           1             N/A             62
##     [197]           1             N/A             29
##          percent_query_coverage_per_hsp percent_query_coverage_per_uniq_subject
##          <numeric>          <character>
##      [1]           100             N/A
##      [2]           48             N/A
##      [3]           49             N/A
##      [4]           57             N/A
##      [5]           49             N/A
##      ...           ...           ...
##     [193]           19             N/A
##     [194]           19             N/A
##     [195]           20             N/A
##     [196]            4             N/A
##     [197]           26             N/A
## -----
## seqinfo: 110 sequences from an unspecified genome; no seqlengths

```

## get the 10 top hits identifiers

```
top10_seqs <- notum_blst %>% arrange(evalue) %>% head(11) %>% .$subject_id
```

## import fasta db

import the sequences and filter them for the 10 best hits from the previous results in order to make the multiple sequence alignments

```
suppressPackageStartupMessages({
  library(Biostrings)
})
fasta_db <- readAAStringSet("oma_db")

# I didn't consider that the names will be cut in the db creation thus I have to manipulate them to work
names_fst_db <- fasta_db %>% names %>% str_remove(" .+")

# filter for the top 10
fasta_db_fil <- fasta_db[names_fst_db %in% top10_seqs]
```

## Multiple Sequence Alignment

I found two packages in R implementing various algorithms. One is msa and the other DECIPHER for now I will try msa

```
suppressPackageStartupMessages({
  library(msa)
})
myFirstAlignment <- msa(fasta_db_fil)
```

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

## print the MSA

```
print(myFirstAlignment, show="complete")
```

```
##
## MsaAMultipleAlignment with 11 rows and 744 columns
##      aln (1..53)                                names
## [1] -----MLFLAVVLLVLVREGSEVFGRSVGGHLSDE Bf|XP_002599183.1...
## [2] -----MGRGVRVLLLLSLLHCAGGSEGRKTWRR--RGQQPPP Hs|NOTUM gi|76799...
## [3] -----MLLFLVVAVTFCAKMTSSMNSHRG--NGADSNE Sk|269784925 ref|...
## [4] -----MVLAA-----SNH Ct|ELT91091.1 ELT...
## [5] -----MVTVTCLPRTSGH-----TDD Pa|g11818.t1
## [6] -----MWLSFIVGAILTTGTVTFPVDQRA-----NTD La|919081898 ref|...
```

```

## [7] -----MTKIYFIVTFWVCAIYDASSMAR-----RLD Cg|762161189 ref|...
## [8] MAVEQIDKMAAKAGEATNKWIKPQQPLLTLTLLLATFSQLPAVC-----SS Dm|Notum NP_73009...
## [9] -----MKVVVLMLLLAIELGCCQHAQ-----NE Xb|g3677.t1
## [10] -----MLQHTLAVIVLTVLQSSR-----AD Ep|XP_020912206.1...
## [11] -----MDTLWSLLALLVLCISFSKCA-----EE Nv|NVE7485
## Con -----????????????????????-----??? Consensus
##
##      aln (54..106)                                names
## [1] FAIDVTKLADSEALMTQTLGRTQRRKSRRKRPNFERKMKDLAEA-----VHAC Bf|XP_002599183.1...
## [2] PPRTEAAPAAGQPVESFPLDFTAVEGN---MDSFMAQVKSQAQS-----LYPC Hs|NOTUM gi|76799...
## [3] CVLDLQMSDDPE-----RSMTQIKYLASA-----LNLK Sk|269784925 ref|...
## [4] FPLDLAELDAIGI-----NAEDLLVHLKHLADG-----NYVC Ct|ELT91091.1 ELT...
## [5] IILDIDELSKLG-----DTDKILEAIRKLAGS-----LHGC Palg11818.t1
## [6] IVLHLEELSKYG-----DPAKVIKMIQDLAGG-----LHGC La|919081898 ref|...
## [7] MPLDVNKLESLG-----NTDTLMKSLKDIAQK-----SHNC Cg|762161189 ref|...
## [8] SILDAASLQEKD-----PLRDTSMNMIQRNYMVMHSAS Dm|Notum NP_73009...
## [9] FILDVANVDPE-----LLIEEIKNLAQS---LHSC- Xb|g3677.t1
## [10] KITDLQMGIKMLA-----QGLQNCQRGLNS-----Ep|XP_020912206.1...
## [11] SVMDLQAQLRYMA-----QQLQACEGDLPSL-----Nv|NVE7485
## Con ??LD???L????-----???L???K?LA??-----???C Consensus
##
##      aln (107..159)                                names
## [1] GSMGF-TDMKLHKLRLNTSVTCNDGSPAG-----YYLRRSHGSKRWLLFLEG-- Bf|XP_002599183.1...
## [2] SAQQLNEDLRLHLLNTSVTCNDGSPAG-----YYLKESRGSRRWLLFLEG-- Hs|NOTUM gi|76799...
## [3] GFVPELPKMKRLRYLENTTVTCNDGSPAG-----YYLYPSNGSTRWLIFLEG-- Sk|269784925 ref|...
## [4] G-PRQMSEMKRHFIRNPSVTCNDGSKAG-----YYLRRSPSSSRWIVFLEG-- Ct|ELT91091.1 ELT...
## [5] G-VKGVPNLSRVNLTR-SVSCNDGSLAG-----YYIRKSFGSTKWIIFLEG-- Palg11818.t1
## [6] G-IKEIPRLKRHFLTNRTVTCNDGSPAG-----YYIRQSYGSKRWIVYLEG-- La|919081898 ref|...
## [7] G-VKEIPKLRHRLNTSVTCNDGSRAG-----YYLRKSHGSKKWIIFLEG-- Cg|762161189 ref|...
## [8] GSGDHSRSLKRANLANTSITCNDGSHAG-----FYLRKHPSSKKWIVLLEG-- Dm|Notum NP_73009...
## [9] GDADVPR-LRLQWLSDPDVRCDGSPAGNADSGVQPVILPASDVFTGFIHRTR Xb|g3677.t1
## [10] ----DVFDMLHYL-DVNTTCNDGTRAG-----YYLKTSYGSKRWIVYLEG-- Ep|XP_020912206.1...
## [11] ----SAFDLKLHYLTDVNTTCNDGSPAG-----YYLKESPKSKRWLVYLEG-- Nv|NVE7485
## Con G-?????LKLH?L?N?SVTCNDGSPAG-----YYLR?S?GSKRWI?FLEG-- Consensus
##
##      aln (160..212)                                names
## [1] ----GWYCFDQASCRNRWANMPRN---LMSSKGWPRDKKNGILSPDPEENPY Bf|XP_002599183.1...
## [2] ----GWYCFNRENCDSRYDTMR---LMSSRDWPRTGTGTGILSSQPEENPY Hs|NOTUM gi|76799...
## [3] ----GWYCFDDSCQSRWESMRG---LMSSTRWTPEKAGSGLLSPDPEENPN Sk|269784925 ref|...
## [4] ----GWMCFDQGCQGRWINTP---HLMSSGHWAETRKGDGILSWNPDENPF Ct|ELT91091.1 ELT...
## [5] ----GGYCFDGATCHSRWINMQNVERNLMSRRKQWETRIGTGILSWDPEENPS Palg11818.t1
## [6] ----GGYCFDQLSCHGRWLSR---KLMSSNEWPETKEGTGILSWHPEENPY La|919081898 ref|...
## [7] ----GWYCFDRFSCELRWSSKMR---KYMTSNGWPEYKTGTGILSWDPKENPY Cg|762161189 ref|...
## [8] ----GWHCFDVRSCRSRWMRLRH---LMTSSQWPETRDVGILSPHPEENPY Dm|Notum NP_73009...
## [9] AVWSGWYCYDSNSCMVRWRMRMS---FMTSESWSYNRTGSGILSGNPEENPD Xb|g3677.t1
## [10] ----GWFACTNRQSCNRRANSRMR---NLMTSRHWRNITIGTGMSSSPTENPN Ep|XP_020912206.1...
## [11] ----GWFCYNQMSCNIRANSQMR---YLMTSKNWSKTKRGSGMLSPQPEENPN Nv|NVE7485
## Con ----GWYCFD??SC??RW????---?LMSS??W??T??G?GILS??PEENP? Consensus
##
##      aln (213..265)                                names
## [1] WWNANTVYVPYCSDVWSGMSPRHDKDD----FAFMGALILQEVLRLDPLGL Bf|XP_002599183.1...
## [2] WWNANMVFIPYCSDVWSGASSKSEKNE----YAFMGALIIQEVVRELLGRGL Hs|NOTUM gi|76799...
## [3] WWNANKVFIPYCSDVWSGT-ARADQGG----YAFMGALILQEVIRELIPQGL Sk|269784925 ref|...
## [4] IHDGNLVYVPYCSDSWSGTYKAQAKGE----FSFMGSLILQEVIRDLVEHHD Ct|ELT91091.1 ELT...

```

```

## [5] FFTANVVLVPYCSDSAWSGTAKARSERE---FSFMGSQILAEVVKDLLANQG Pa|g11818.t1
## [6] FYHANIVVVPYCSDSWSGHKMASGKGE---FSFLGSLIVQEVIKDLLNMSD La|919081898 ref|...
## [7] YFNANIVVVPYCSDSWTGTSLRNGDG-----YAFLGSYIIIEVIRDLIPRGL Cg|762161189 ref|...
## [8] WHNANHVLIPYCSDSWSGTRTEPDTSDRENSWRFMGALILRQVIAELIPVGL Dm|Notum NP_73009...
## [9] WWNANTVLIPYCSDLTWGR---SDAGDNDLPFSFHGSKILRSVIRELLPMGL Xb|g3677.t1
## [10] WWNANHVLIPYCSDAWTGNASRHETGEK---FSFLGSRIIEKVIEDILPQGL Ep|XP_020912206.1...
## [11] WWNANHVLIPYCSDAWSGNASRHETGEK---FSFLGARILEKVIEDLLPRGL Nv|NVE7485
## Con WWNAN?V?IPYCSD?WSG???????GE---FSFMGSLIL?EVIRDLLP?GL Consensus
##
##      aln (266..318)                                names
## [1] --KNSKTLTLLSGSSAGGTGVILNLDRAEFLRREGS-SVQVQGVADSGWFLDN Bf|XP_002599183.1...
## [2] --SGAKVLLLAGSSAGGTGVLLNVDRVAEQLEKLGYPAIQVRGLADSGWFLDN Hs|NOTUM gi|76799...
## [3] --LVANKILLAGSSAGGTGVLLNLDYVSDMLSAAGS-NAVVRGICDSGWFLDT Sk|269784925 ref|...
## [4] -MESASKLYLAGSSAGGTGVLLNLDVAAQMSLLAP-RVEVRGIADSGWFLDN Ct|ELT91091.1 ELT...
## [5] -MHKGKLLLTGSSAGGLGVLLNVDKVAAQVKKVVS-TMDVRGIADSGWFMDN Pa|g11818.t1
## [6] GLSDGNKLYLAGSSAGGTGVLLNLDVADLVHQAP-NIEVRGIADSGWFLDN La|919081898 ref|...
## [7] --ARGKKFLTGTSGAGTGVLMNLDRIADLVKSLAP-RVEVRGIADSGWFLDI Cg|762161189 ref|...
## [8] GRVPGGELMLVGSSAGGMVMLNLDRIIRDFLVNEKKLQITVRGVSDSGWFLDR Dm|Notum NP_73009...
## [9] --YDADTLLLAGSSAGGLGVMLNLDIARELG-ELGASVDVRGLIDSGWYLDN Xb|g3677.t1
## [10] --YRAKHLTLAGSSAGAIGVLLNVDRVAAKLRLSGF-KVNVRLVDSGWYLDN Ep|XP_020912206.1...
## [11] --YNAKHLTLAGSSAGGIGVILNLDRISTKLHAMGF-AVEVRGLADSGWYLS Nv|NVE7485
## Con --??AK?LLLAGSSAGGTGVLLNLDRA??L????-?V?VRG?ADSGWFLDN Consensus
##
##      aln (319..371)                                names
## [1] KQYMPTECTE-TLSCAPTEAIRRGIQWNGQVPERCARQYSKDEQWRCFFGYR Bf|XP_002599183.1...
## [2] KQYRHTDCVD-TITCAPTEAIRRGIYWNGVPERCRRQFQEGEEWNCFFGYK Hs|NOTUM gi|76799...
## [3] VQHRAQPCTN-TLSCAPSEVIKRGIKLWSGQVPARCSEEYSYNDQWKCFGYR Sk|269784925 ref|...
## [4] KQYEHVKCSE-VHSCAPTEAVMRGFKLWHAIEVPDKCRGQYPDDQHWRCFFGYR Ct|ELT91091.1 ELT...
## [5] KQYRDGDCSD-ARTCAPTEGIQRGMQLWKANIPEGCRKQYPNE-EWKCFGYR Pa|g11818.t1
## [6] EPFKVAACTEYSHICSPTEAIKRGIFHWGGKVPQACKESFPGNQWQSCYFGYN La|919081898 ref|...
## [7] PQFNEKTCTE-PLSCSPTTGIIKGFQWRGRVPEACKNEYPDQEQWRCYFGYR Cg|762161189 ref|...
## [8] EPTYPAAVAS-----NEAVRQGWKLWQGLLPEECTKSYPTTEP-WRCYGYR Dm|Notum NP_73009...
## [9] PQYNPAECVEVFSCP-PREVVMDGIQIWDVMPVECMARFVDEYWRCCFSHE Xb|g3677.t1
## [10] ITRKPG-CTQG---SCPAKTIKEGMRVWKGVPDACAAYPLQE-WKCYFGNI Ep|XP_020912206.1...
## [11] RPFESS-CPPGVKECGPVKTIKEGMMYWRGIVPENCTKENLLQP-WMCYFGET Nv|NVE7485
## Con ?QY????C??-???C?PTE?I?RG???W?G?VPE?C???YP???W?CFFGYR Consensus
##
##      aln (372..424)                                names
## [1] AYPTLQAPLFVIQWLFDEAQMVMNV-----GTPVDKEQWNYIHNL Bf|XP_002599183.1...
## [2] VYPTLRCPVFVQWLFDEAQLTVDNVHLT-----GQPVQEGRLRYIQNL Hs|NOTUM gi|76799...
## [3] IYPTLQTPVFIQWLYDEAQLVVGMT-----GPPAKLEHWNMQQL Sk|269784925 ref|...
## [4] IYSTLKTVPFVQHLFDEAQITVNNVGPP-----VKKAQWQYIHNI Ct|ELT91091.1 ELT...
## [5] LYPFIKAPVFIQWLFDEAQIAAENVGTP-----LKKEQWQYIHNM Pa|g11818.t1
## [6] IYPTLNTPVFIQWLYFDEAQITADHVGTP-----VNKEQWHYIHNL La|919081898 ref|...
## [7] MYPTLKTVPFVIQWLYFDEAQILANNLINQNKLMNYNGSNELLSKEQWEYLYKL Cg|762161189 ref|...
## [8] LYPTLKTPLFVQWLFDEAQMVRDVGAP-----VTPQQWNYIHEM Dm|Notum NP_73009...
## [9] IYQTLSTPVFVQWLYDEAQMMLLDNVGIP-----VSTSHWRYIQR I Xb|g3677.t1
## [10] VYPTLNSSTFIFQWLYDETQLALDGSGLPK-----SLKGVTEQEQVFIVML Ep|XP_020912206.1...
## [11] VYPTITAPLFIFQWLYDEAQLALDGSIQPR-----GIQTIDLKQIKTIFKI Nv|NVE7485
## Con ?YPTL?TPVFV?QWLFDEAQ???DNVG?P-----???V???EQW?YI???L Consensus
##
##      aln (425..477)                                names
## [1] GVDLRKTLTN--VTGVFAPACLAHTLITKSDWMT-VQMKGVSPLNALHCWEQS Bf|XP_002599183.1...
## [2] GRELRHTLKD--VPASFAPACLSHEIIRSHWTD-VQVKGTSPLRALHCWD RS Hs|NOTUM gi|76799...

```

```

## [3] GRELRHSLKN--VSAVFAPACYSHKVIDKMQLN-VHVKGISLPNALECWLS Sk|269784925 ref|...
## [4] GQDMKRTLTN--VSALFAPACLSHIVLARSEWQR-VSISGVTLPALKCWEDG Ct|ELT91091.1 ELT...
## [5] GKSIKQSLQN--VSAVFASACLSHTTLTLPDWHK-MTIKGVTLTDALKCWTS Palg11818.t1
## [6] GNDIRNTLEN--VSATFAPACYSHKLLTKKNWQN-LKISGHSLPQALRCWEES La|919081898 ref|...
## [7] GEKVKQTLDN--VTAVFAPACLSHEVLLKKDWHK-ISVKGISLSQSIYCWENS Cg|762161189 ref|...
## [8] GGALRSSLDN--VSAVFAPACISGHVLFKRDWVN-IKIDDISLPSALRCWEHS Dm|Notum NP_73009...
## [9] GRQLRESLSN--VSAVYAPACISHTYITKPEVFKSADVNGVTLPAIECWMSA Xb|g3677.t1
## [10] GIQLKHSRLRQNNVHSVFAPACLYHTVLTDSNWLK-IKIDKYTLDDALTCWMNS Ep|XP_020912206.1...
## [11] GRKIRESLKRARVRHVFPACISHTILTHSSWLN-IRLKGASLNDILTCWYHT Nv|NVE7485
## Con G??LR?SL?N--VSAVFAPACLSH??LTK??W??-???KG?SLP?AL?CW??S Consensus
##
##      aln (478..530)                                names
## [1] -----Bf|XP_002599183.1...
## [2] -----Hs|NOTUM gi|76799...
## [3] NEESNH-----Sk|269784925 ref|...
## [4] R-----Ct|ELT91091.1 ELT...
## [5] D-----Palg11818.t1
## [6] T-----La|919081898 ref|...
## [7] FEGEKC-----Cg|762161189 ref|...
## [8] TRSRHDKLKRSTEPSTAVSHPEHANNQRHQRHRLQRQKHNNVAQSGGQQR Dm|Notum NP_73009...
## [9] DQYP-----ADVTRYRRRRS-----Xb|g3677.t1
## [10] AG-----Ep|XP_020912206.1...
## [11] DG-----Nv|NVE7485
## Con ??-----Consensus
##
##      aln (531..583)                                names
## [1] -----LLWGKVEPQEPT-----Bf|XP_002599183.1...
## [2] -----LHDSHKASKTPLKG-----Hs|NOTUM gi|76799...
## [3] -----SLRNEDAAPEDSDDRQMIHVHPASTQGRKQR Sk|269784925 ref|...
## [4] -----QETNHY-----E Ct|ELT91091.1 ELT...
## [5] -----HEPNSYYPSTLIPSIRSNNENKLLGDEAVPR Palg11818.t1
## [6] -----SEANHYDQKDASARENSEGDGTNLQGDDPR La|919081898 ref|...
## [7] -----ISATNGSTSVKLIRPYGDAPVKARNQNKKKSQ Cg|762161189 ref|...
## [8] KHNHLSKEEREERKRLRQEQRRKQRRRQQQKKANGGQEHNRKKDNSPKSS Dm|Notum NP_73009...
## [9] -----PSARRR-----GGRVRRD-----PVGX Xb|g3677.t1
## [10] -----HNQGSKRHRN-RIRRGK-----Ep|XP_020912206.1...
## [11] -----HEDGHGKHGHGKHGHGKHGPE-----Nv|NVE7485
## Con -----????????????????????-----Consensus
##
##      aln (584..636)                                names
## [1] -----Bf|XP_002599183.1...
## [2] -----Hs|NOTUM gi|76799...
## [3] QESDEYETHAH-----Sk|269784925 ref|...
## [4] VTHQKDR-----Ct|ELT91091.1 ELT...
## [5] VKKNKNKKKGRKSKKKKGNKRRRKERRRRNRKRIRSERRGTKNSQSNEP Palg11818.t1
## [6] SNKRRHKKNK-----KGRRRGRKGRKRDKNRERNREDRQR La|919081898 ref|...
## [7] KKKKKNKRKRD-----KNKRRRRNRNRKNKNKNKNKNKRKKN Cg|762161189 ref|...
## [8] NGNDQRKQRRRQQLTAERQEQRRRKAQQQMKMQREQPAAGVFLEASAPQ Dm|Notum NP_73009...
## [9] SDNDQ-----TTWDESNIQLMMDADTLI Xb|g3677.t1
## [10] -----Ep|XP_020912206.1...
## [11] -----Nv|NVE7485
## Con ??????-Consensus
##
##      aln (637..689)                                names

```

```

## [1] -----CVSQLVDSCPWPH-CNPTCPKLRNPFT Bf|XP_002599183.1...
## [2] -----CPVHLVDSCPWPH-CNPSCPTVRDQFT Hs|NOTUM gi|76799...
## [3] -----TVHQSRCCHHVIDNAPCPQ-CNPMCPKLVNPFT Sk|269784925 ref|...
## [4] -----CSHHRMDHCAWPQ-CNFSCP KIRNPFT Ct|ELT91091.1 ELT...
## [5] PRRERSAFIEYINPGIEEHRLLYHPTCRHHLADTCEWPQ-CNQSCP KLRNPFS Pa|g11818.t1
## [6] SQRSLDFLQSDNIHAQHHRQHHSNQCVHHLIDNCGWPQ-CNLSCP KLTNPYT La|919081898 ref|...
## [7] RKKKRKGKGNRKKNRQRSKSNVSNCSHHLIEQCPLPQWCNSFCPTFRNQYN Cg|762161189 ref|...
## [8] KTRSSNNASAGTKSKKRHRVPRVPEKCGRLRLLERCSWPQ-CNHSCP TLTNPMT Dm|Notum NP_73009...
## [9] TSHSETSAVEAT-----CTQH AIDQCNWPQ-CNPACPNLKNPFT Xb|g3677.t1
## [10] -----CRMHLVDRCHHPQ-CNPTCPNPRDPYT Ep|XP_020912206.1...
## [11] -----NQHHWSTHQVDHCLYIQ-CNPTCPIPRNPFT Nv|NVE7485
## Con -----????C??HL?D?C?WPQ-CNP?CP?LRNPFT Consensus
##
##      aln (690..742) names
## [1] GEEMDFIPLMM-ELVGIDMNAIAEQMGMDPDDLVRMLTS----- Bf|XP_002599183.1...
## [2] GQEMNVAQFLM-HMG-FDMQTVAQPQGLEPSELLGMLSNGS----- Hs|NOTUM gi|76799...
## [3] DEEMEFLPFLK-LMG-FDLGSLAKLLGLDAQTMAMMEG----- Sk|269784925 ref|...
## [4] GEEMDFIDLLM-QFG-LDLSSIAEALGMDLPTLQAMDHDVLLRLLVKNSR--- Ct|ELT91091.1 ELT...
## [5] GEEIDFIDLLM-QFG-LDMSSIAEALGKDLKTLQMDHAKLIKLLMEQE---- Pa|g11818.t1
## [6] GEEIDFIDLLI-QFG-IDLSSIANALNMDLATMQAMDHETLLKILTQQQ---- La|919081898 ref|...
## [7] GQEIDFFKVFA-ASG-MDSKSMAELLGVNTSTVNAMNANTISTLLLDTRQKR Cg|762161189 ref|...
## [8] GEEMRFLELLT-AFG-LDIEAVAAALGVMHTLNNMERTELVNMLTQQAN--- Dm|Notum NP_73009...
## [9] GEEIELVQLLT-QLG-FDLGSLARQIGVSPMMLATIEPHLILQMLLST----- Xb|g3677.t1
## [10] GQPYGSLTRPGLVIPGVPISE---LNNPPSQVDRGFQ----- Ep|XP_020912206.1...
## [11] GKPY-PMKLPG-HLPGIPLSRIDSFMNIPSNLNRD----- Nv|NVE7485
## Con GEE??F??LL?-??G-?D?S??A??LG?D??TL??M???????L???----- Consensus
##
##      aln (743..744) names
## [1] -- Bf|XP_002599183.1...
## [2] -- Hs|NOTUM gi|76799...
## [3] -- Sk|269784925 ref|...
## [4] -- Ct|ELT91091.1 ELT...
## [5] -- Pa|g11818.t1
## [6] -- La|919081898 ref|...
## [7] PQ Cg|762161189 ref|...
## [8] -- Dm|Notum NP_73009...
## [9] -- Xb|g3677.t1
## [10] -- Ep|XP_020912206.1...
## [11] -- Nv|NVE7485
## Con -- Consensus

```