

Übungsblatt 4

Multiple Sequence Alignment

Aufgabe 4.1

Teilaufgabe 4.1.1

S1 A A T C G
S2 A A C G
S3 C G
S4 A A

m	+1
mm	-1
g	-2

• Paarweise Alignment

$$\left[\begin{array}{l} S1 \text{ A A T C G} \\ S2 \text{ A A - C G} \end{array} \right] \left. \begin{array}{l} p = 0.2 \\ \text{score} = 1+1-2+1+1 = 2 \end{array} \right\}$$

$$\left[\begin{array}{l} S1 \text{ A A T C G} \\ S3 \text{ - - - C G} \end{array} \right] \left. \begin{array}{l} p = 0.6 \\ \text{score} = -2-2-2+1+1 = -4 \end{array} \right\}$$

$$\left[\begin{array}{l} S1 \text{ A A T C G} \\ S4 \text{ A A - - -} \end{array} \right] \left. \begin{array}{l} p = 0.6 \\ \text{score} = +1+1-2-2-2 = -4 \end{array} \right\}$$

$$\left[\begin{array}{l} S2 \text{ A A C G} \\ S3 \text{ - - C G} \end{array} \right] \left. \begin{array}{l} p = 0.5 \\ \text{score} = -2-2+1+1 = -2 \end{array} \right\}$$

$$\left[\begin{array}{l} S2 \text{ A A C G} \\ S4 \text{ A A - -} \end{array} \right] \left. \begin{array}{l} p = 0.5 \\ \text{score} = -2 \end{array} \right\}$$

$$\left[\begin{array}{l} S3 \text{ C G} \\ S4 \text{ A A} \end{array} \right] \left. \begin{array}{l} p = 1 \\ \text{score} = -1 \end{array} \right\}$$

• p-Distanzmatrix

	S1	S2	S3	S4
S1	-	0.2	0.6	0.6
S2		-	0.5	0.5
S3			-	1
S4				-

→ S1S2 $\hat{=}$ Cluster X

$$d(X, S3) = \frac{1}{2.1} (0.2 + 0.5) = 0.55$$

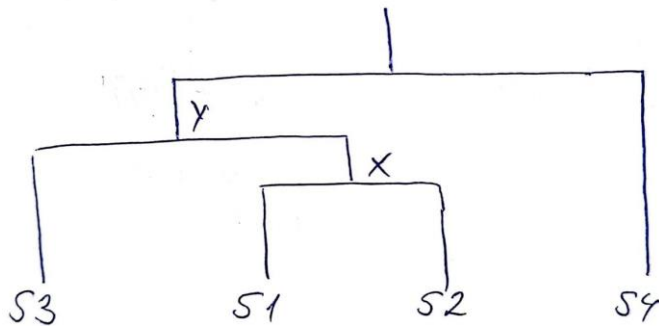
$$d(X, S4) = \frac{1}{2.1} (0.6 + 0.5) = 0.55$$

	X	S3	S4
X	-	0.55	0.55
S3		-	1
S4			-

→ X S3 \triangleq cluster Y

	Y	S4
Y	-	1.13
S4		-

$$d(Y, S4) = \frac{1}{3 \times 1} (0.2 + 0.6 + 0.6 + 0.5 + 0.5 + 1) = 3.4/3 \approx 1.13$$



Teilaufgabe 4.1.2

A A T C G
 A A - C G
 - - - C G
 A A - - -

Konsens Alignment: A A T C G

Teilaufgabe 4.1.3

Sum-of-Pairs-Score: S1 A A T C G
 S2 A A - C G
 S3 - - - C G
 S4 A A - - -

$$(S1, S2) + (S1, S3) + (S1, S4) + (S2, S3) + (S2, S4) + (S3, S4) = 2 - 4 - 4 - 2 - 2 - 1 = -13$$

Aufgabe 4.2

CLUSTAL O(1.2.4) multiple sequence alignment

```
sp|P02112|HBB_CHICK   MVHWTAEEKQLITGLWGKVNVAECGAEALARLLVYPWTQRRFFASFGNLSSPTAILGNPM   60
sp|P02070|HBB_BOVIN  --MLTAEKAAVTAFWVGKVKVDEVGGEALGRLLVYPWTQRRFFESFGDLSTADAVMNNPK   58
sp|P02067|HBB_PIG    MVHLSAEEKAEAVLGLWGKVNVDDEVGGEALGRLLVYPWTQRRFFESFGDLSTADAVMGNPK   60
sp|P68871|HBB_HUMAN   MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRRFFESFGDLSTPDAMGNPK   60
sp|P68873|HBB_PANTR   MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRRFFESFGDLSTPDAMGNPK   60
sp|P68872|HBB_PANPA   MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRRFFESFGDLSTPDAMGNPK   60
sp|P02024|HBB_GORGO   MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVYPWTQRRFFESFGDLSTPDAMGNPK   60
      .*** ..***** * * * * * .***** .***. *..**

sp|P02112|HBB_CHICK   VRAHGKKVLTSFGDAVKNLNLIKNTFSQLSELHCDKLHVDPENFRLLGDILIVLAHFS   120
sp|P02070|HBB_BOVIN   VKAHGKKVLDSFSNGMKHLDDLKGTFAALSELHCDKLHVDPENFKLLGNVLVVLARNFG   118
sp|P02067|HBB_PIG     VKAHGKKVLQSFSDGLKHLNLIKGTFAKLSELHCDQLHVDPENFRLLGNVIVVLARRLG   120
sp|P68871|HBB_HUMAN   VKAHGKKVLGAFSDGLAHLNLIKGTFAATLSELHCDKLHVDPENFRLLGNLVCVLAHHFG   120
sp|P68873|HBB_PANTR   VKAHGKKVLGAFSDGLAHLNLIKGTFAATLSELHCDKLHVDPENFRLLGNLVCVLAHHFG   120
sp|P68872|HBB_PANPA   VKAHGKKVLGAFSDGLAHLNLIKGTFAATLSELHCDKLHVDPENFRLLGNLVCVLAHHFG   120
sp|P02024|HBB_GORGO   VKAHGKKVLGAFSDGLAHLNLIKGTFAATLSELHCDKLHVDPENFKLLGNLVCVLAHHFG   120
      *.***** .* .. .**.* ** .*****.*****.***.... *** ..

sp|P02112|HBB_CHICK   KDFTEPCQAAWQKLVRVVAHALARKYH   147
sp|P02070|HBB_BOVIN   KEFTPVLQADFQKVAVAGVANALAHRYH   145
sp|P02067|HBB_PIG     HDFNPNVQAAFQKVAVAGVANALAHKYH   147
sp|P68871|HBB_HUMAN   KEFTPPVQAAYQKVAVAGVANALAHKYH   147
sp|P68873|HBB_PANTR   KEFTPPVQAAYQKVAVAGVANALAHKYH   147
sp|P68872|HBB_PANPA   KEFTPPVQAAYQKVAVAGVANALAHKYH   147
sp|P02024|HBB_GORGO   KEFTPPVQAAYQKVAVAGVANALAHKYH   147
      :.* * **.* **.***.*
```

Phylogenetischer Baum (Cladogram)



- Die Aminosäuresequenz des Hemoglobins subunit beta der Rinde und des Schweins sind am nächsten verwandt im Kontext der 7 Spezies, die ich aligned habe. Am verwandtesten zum Cluster Rind&Schwein ist das Bankivahuhn. Nach diesem Prinzip folgen die Gorilla, der Mensch, der Schimpanse und der Bonobo.

Reference sequence (1): sp|P02112|HBB_CHICK
Identities normalised by aligned length.
Colored by: identity

	cov	pid	
1 sp P02112 HBB_CHICK	100.0%	100.0%	MVHWT EEKQLITGLWCKNVAECEGAALRLLLIYPWTORFFSEGNSSFTAILGPNMVRHAHKKK/LTSFQDAVKND
2 sp P02070 HBB_BOVIN	98.6%	63.4%	--MLT EEEKAAVTAFNCKKVDEVGGEALGRLLVIYPWTORFFESFGDLSTADAVMNPKVKAHKK/VLDSFSNGMKHL
3 sp P02067 HBB_PIG	100.0%	66.7%	MVHLSAEEKEAVLGLWCKNVDEVGGEALGRLLVIYPWTORFFESFGDLSNADAVMCNPVKVAHKK/VLDSFDGLKHLD
4 sp P68871 HBB_HUMAN	100.0%	69.4%	MVHLTPEEKSAVTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGDLSTPDAMVCNPKVKAHKK/VLGFASDGLAHL
5 sp P68873 HBB_PANTR	100.0%	69.4%	MVHLTPEEKSAVTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGDLSTPDAMVCNPKVKAHKK/VLGFASDGLAHL
6 sp P68872 HBB_PANPA	100.0%	69.4%	MVHLTPEEKSAVTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGDLSTPDAMVCNPKVKAHKK/VLGFASDGLAHL
7 sp P02024 HBB_GORG	100.0%	68.7%	MVHLTPEEKSAVTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGDLSTPDAMVCNPKVKAHKK/VLGFASDGLAHL
consensus/100%			..hhosEEKthlhuHWCKVPSESdualRLLLIYPWTORFFESFGS SsssaAlhsNPVVAHKK/LtuFusuhtpL
consensus/90%			..hhosEEKthlhuHWCKVPSESdualRLLLIYPWTORFFESFGS SsssaAlhsNPVVAHKK/LtuFusuhtpL
consensus/80%			MVHTSEEKAVLTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGD SosDAVMCNPKVKAHKK/VLsuFSDGLtHL
consensus/70%			MVHTSEEKAVLTALWKCKNVDEVGGEALGRLLVIYPWTORFFESFGD STPDAMVCNPKVKAHKK/VLsuFSDGLtHL

	cov	pid	
1 sp P02112 HBB_CHICK	100.0%	100.0%	NLKNTFSQ SE HODK H D ENFR L LGSLILTVLA AHFSKDFTECQA AAWQKLV R V AHALARKYH
2 sp P02070 HBB_BOVIN	98.6%	63.4%	DLKGTFAAL SE HODK H D ENFR K LGNLVLVLAARFGKEFTPVLQDFQKV VAGVANALAHRYH
3 sp P02067 HBB_PIG	100.0%	66.7%	NLKGTFAKL SE HODK H D ENFR L LGNVIVVLAARRLGDFNPVQAAPQKV VAGVANALAHKYH
4 sp P68871 HBB_HUMAN	100.0%	69.4%	NLKGTFAAT SE HODK H D ENFR L LGNLVLVCLVAHFHGKEFTPVPQAAYQKV VAGVANALAHKYH
5 sp P68873 HBB_PANTR	100.0%	69.4%	NLKGTFAAT SE HODK H D ENFR L LGNLVLVCLVAHFHGKEFTPVPQAAYQKV VAGVANALAHKYH
6 sp P68872 HBB_PANPA	100.0%	69.4%	NLKGTFAAT SE HODK H D ENFR L LGNLVLVCLVAHFHGKEFTPVPQAAYQKV VAGVANALAHKYH
7 sp P02024 HBB_GORG	100.0%	68.7%	NLKGTFAAT SE HODK H D ENFR L LGNLVLVCLVAHFHGKEFTPVPQAAYQKV VAGVANALAHKYH
consensus/100%			sLKSTFut SE HODp H D ENF+ LGslllhVLAtpHu+-FSP.hQAsaOKLVtsvAPA++YH
consensus/90%			sLKSTFut SE HODp H D ENF+ LGslllhVLAtpHu+-FSP.hQAsaOKLVtsvAPA++YH
consensus/80%			NLKGTFAp SE HODK H D ENF+ LGNLVLvsVLA++FGK-FTPslOaaqKVVAGVANALAHKYH
consensus/70%			NLKTFas SE HODK H D ENFR LGNLVLvsVLA+HFGKEFTpsVOAaaqKVVAGVANALAHKYH

MVHLS¹EEK²NAV³TALGK⁴VN⁵DEVGGEALGR⁶LLVVP⁷TQRF⁸FE⁹SD¹⁰STPD¹¹AVMGN¹²EKKV¹³AK¹⁴KK¹⁵VLS¹⁶u¹⁷FSD¹⁸GI¹⁹TH²⁰LD²¹
 NIK²²CTFAS²³LSE²⁴HC²⁵DKL²⁶HYD²⁷ENFRLLCN²⁸VLVSVLA²⁹+HEG³⁰KEFT³¹PS³²VQAA³³OK³⁴VVAG³⁵VALAH³⁶KY³⁷

Aufgabe 4.3

```
> setwd("~/Universität/12. Semester/Bioinformatik/R/Übungen/Übung 4")
> aligned_seq <- read.delim("human_ACTG_isoforms.txt")
> ali_mat <- as.matrix(aligned_seq)
> View(ali_mat)
```

```
1 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISVGRPRHQVGMVGMDQKDSVVGDEAQSKRGILTLYKPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMTQIMFETFTNPAMYVAIQVLSLYASGR.....TTGIVMDSGDGVTHT
2 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISVGSWAWA...
3 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
4 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
5 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
6 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
7 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
8 MEEEEIALVIDNGSGMCKAGFAGDDAPRAVFPISGRPRHQV...
```

```
> install.packages("stringr")
```

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

<https://cran.rstudio.com/bin/windows/Rtools/>

Installing package into 'C:/Users/raliz/AppData/Local/R/win-library/4.2'

(as 'lib' is unspecified)

```
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/stringr_1.4.0.zip'
```

Content type 'application/zip' length 212170 bytes (207 KB)

downloaded 207 KB

```
package 'stringr' successfully unpacked and MD5 sums checked
```

The downloaded binary packages are in

C:\Users\raliz\AppData\Local\Temp\RtmpKESbte\downloaded_packages

```
> library(stringr)
```

```
> str_extract_all(ali_mat, boundary("character"))
```

[[1]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "A" "S" "W" "W" "A" "W" "A" "R" "R" "T" "-" "-" "-" "-" "P" "T" "W" "A"
"T" "R" "P" "R" "A" "S" "V" "A" "S" "-" "-" "-"

[67] "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[100] "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[133] "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[[2]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "R" "P" "R" "H" "Q" "-" "-" "-" "-" "-" "-" "-" "-" "-" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"

[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"

[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "N" "R" "E" "K" "M" "T"
"Q" "I" "M" "F" "E" "T" "F" "N" "T" "P" "A" "M"

[133] "Y" "V" "A" "I" "Q" "A" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "G" "V" "T" "H" "T"

[[3]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"

[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"

[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "N" "R" "E" "K" "M" "T"
"Q" "I" "M" "F" "E" "T" "F" "N" "T" "P" "A" "M"

[133] "Y" "V" "A" "I" "Q" "A" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "G" "V" "T" "H" "T"

[[4]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"

[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"

[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "N" "R" "E" "K" "M" "T"
"Q" "I" "M" "F" "E" "T" "F" "N" "T" "P" "A" "M"

[133] "Y" "V" "A" "I" "Q" "A" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "G" "V" "T" "H" "T"

[[5]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"

[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"

[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "N" "R" "E" "K" "M" "T"
"Q" "I" "M" "F" "E" "T" "F" "N" "T" "P" "A" "M"

[133] "Y" "V" "A" "I" "Q" "A" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "G" "V" "T" "H" "T"

[[6]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"

[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"

[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[133] "-" "-" "-" "-" "-" "-" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-"

[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "G" "V" "T" "H" "T"

[[7]]

[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"

```
[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"
```

```
[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"
```

```
[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "N" "R" "E" "K" "M" "T"
"Q" "A" "L" "F" "L" "S" "R" "H" "F" "L" "P" "E"
```

```
[133] "A" "-" "-" "-" "-" "S" "R" "L" "C" "L" "R" "P" "S" "T" "P" "R" "P" "C" "T" "W" "P"
"S" "R" "P" "C" "C" "P" "S" "T" "P" "L" "G" "A"
```

```
[166] "P" "L" "A" "L" "S" "W" "T" "L" "E" "T" "G" "S" "P" "T" "R"
```

```
[[8]]
```

```
[1] "M" "E" "E" "E" "I" "A" "A" "L" "V" "I" "D" "N" "G" "S" "G" "M" "C" "K" "A" "G" "F"
"A" "G" "D" "D" "A" "P" "R" "A" "V" "F" "P" "S"
```

```
[34] "I" "V" "G" "R" "P" "R" "H" "Q" "G" "V" "M" "V" "G" "M" "G" "Q" "K" "D" "S" "Y" "V"
"G" "D" "E" "A" "Q" "S" "K" "R" "G" "I" "L" "T"
```

```
[67] "L" "K" "Y" "P" "I" "E" "H" "G" "I" "V" "T" "N" "W" "D" "D" "M" "E" "K" "I" "W" "H"
"H" "T" "F" "Y" "N" "E" "L" "R" "V" "A" "P" "E"
```

```
[100] "E" "H" "P" "V" "L" "L" "T" "E" "A" "P" "L" "N" "P" "K" "A" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "
```

```
[133] "-" "-" "-" "-" "-" "-" "V" "L" "S" "L" "Y" "A" "S" "G" "R" "-" "-" "-" "-" "-" "-"
"-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "-" "
```

```
[166] "T" "T" "G" "I" "V" "M" "D" "S" "G" "D" "-" "-" "-" "-" "
```

```
> consensus <- function(ali_mat, method = c( "majority", "threshold", "IUPAC",
"profile"), threshold = 0.70,
+
+           warn.non.IUPAC = FALSE, type = c("protein")){
+
+   if(inherits(ali_mat, "alignment")) ali_mat <- as.matrix(ali_mat)
+   if(!is.matrix(ali_mat)) stop("matrix or alignment object expected")
+   if(storage.mode(ali_mat) != "character") stop("matrix of characters expected")
+
+   method <- match.arg(method)
+
+   if(method == "IUPAC"){
+     type <- match.arg(type)
+     res <- apply(ali_mat, 2, bma, warn.non.IUPAC = warn.non.IUPAC, type = type)
+     names(res) <- NULL
+   }
+ }
```



```

+       return(res)
+   }
+
+   if(method == "majority"){
+       majority <- function(x) names(which.max(table(x)))
+       res <- apply(ali_mat, 2, majority)
+       names(res) <- NULL
+       return(res)
+   }
+
+   if(method == "profile"){
+       obsvalue <- levels(factor(matali))
+       nrow <- length(obsvalue)
+       row.names(ali_mat)<-NULL
+       res <- apply(ali_mat, 2, function(x) table(factor(x, levels = obsvalue)))
+       return(res)
+   }
+
+   if(method == "threshold"){
+       profile <- consensus(ali_mat = , method = "profile")
+       profile.rf <- apply(profile, 2, function(x) x/sum(x))
+       res <- rownames(profile.rf)[apply(profile.rf, 2, which.max)]
+       res <- ifelse(apply(profile.rf, 2, max) >= threshold, res, NA)
+       names(res) <- NULL
+       return(res)
+   }
+ }
+
>
> con <- consensus
> con(ali_mat)

[1]
"MEEEEIAALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHH
TFYNELRVAPEEHPVLLTEAPLNPKANREKMTQIMFETFNTPAMYVAIQAVLSLYASGR-----
TTGIVMDSGDGVTH"

```