

Find a Gene

```
library(bio3d)
```

```
alignment <- read.fasta("aln-fasta.txt")
alignment
```

```
[Truncated_Name:1]Human_PRL      1      .      .      .      .      50
[Truncated_Name:2]Lesser_hed     MNIKGSPWK-GSL--LLLLVSNLLLCQSVAPLPICPGGAARCQVTLRDLF
[Truncated_Name:3]Sus_scrofa     MANKGSSPK-GSLLLLLLLLSNLLLCQGVASLPICPRGAVNCQVSLRDLF
[Truncated_Name:4]Sea_otter      MDRNGSSQK-GSLLLLLLLLVSNLFLCKSVASLPICPSGAVNCQVSLRDLF
[Truncated_Name:5]Northern_f     MDDKGWSVK-GSLLPLLLLVSDDLCCQGVASLPICPTGAVNCQVSLRDLF
[Truncated_Name:6]Grey_seal      MDNKGWSLKAGSLLPLLLLVSLLLLGQSVDSLPICTGAVNCQVSLRDLF
[Truncated_Name:7]Giant_pand     MDNKGWSLK-GSLLPLLLLVSLLLLGQSVASLPICPTGAVNCQVSLRDLF
* ^*  * ***  ****^* * *  *  *****  **  ***^*****
1      .      .      .      .      50

51      .      .      .      .      100
[Truncated_Name:1]Human_PRL      DRAVVLSHYIHNLSSEMFSEFDKRYTHGRGFITKAINSCHTSSLATPEDK
[Truncated_Name:2]Lesser_hed     DRAVILSHYIHNLSSEMFNEFDRRYAQGRGFITKAINSCHTSSLATPEDK
[Truncated_Name:3]Sus_scrofa     DRAVILSHYIHNLSSEMFNEFDKRYAQGRGFITKAINSCHTSSLSTPEDK
[Truncated_Name:4]Sea_otter      DRAVILSHYIHNLSSEMFNEFDRRYAQGRGFITKAINSCHTSSLSTPEDK
[Truncated_Name:5]Northern_f     DRAVILSHYIHNLSSEMFNEFDKRYAQGRGFITKAINSCHTSSLSTPEDK
[Truncated_Name:6]Grey_seal      DRAVILSHYIHNLSSEMFNEFDKRYAQGRGFITKAINSCHTSSLSTPEDK
[Truncated_Name:7]Giant_pand     DRAVILSHYIHNLSSEMFNEFDKRYAQGRGFITKAINSCHTSSLSTPEDK
****^***** **^** *****
51      .      .      .      .      100

101     .      .      .      .      150
[Truncated_Name:1]Human_PRL      EQAQQMNQKDFLSLIVSILRSWNEPLYHLVTEVRGMQEAPAILSKAVEI
[Truncated_Name:2]Lesser_hed     EQAQQIHEDLLNLVVKVLRSWNDPLYHLVSEVRGMEEAPDAILSRAIEI
```

```

[Truncated_Name:3]Sus_scrofa  EQAQQIHHEILLNLILRVLRSWNDPLYHLVTEVRGMQEAPDAILSRAIEI
[Truncated_Name:4]Sea_otter   EQAQQIHHEILLNLILRVLRSWNDPLYHLVTEVRGMEEAPDSILSRAIEI
[Truncated_Name:5]Northern_f  EQAQQIHHEILLNLILRVLRSWNDPLYHLVTEVRGMQEAPDSILSRAIEI
[Truncated_Name:6]Grey_seal   EQAQQIHHEILLNLILRVLRSWNDPLYHLVTEVRGMQEAPDSILSRAIEI
[Truncated_Name:7]Giant_pand  EQAQQIHHEILLNLILRVLRSWNDPLYHLVTEVRGMQEAPDSILSRAIEI
                                *****~ * *^^ ^*****^*****^***** ***^ ***^*^**
101                             . . . . . 150

                                151 . . . . . 200
[Truncated_Name:1]Human_PRL  EEQTKRLLLEGMEILVSVQVHPETKENEIYPVWSGLPSLQMADEESRLSAYY
[Truncated_Name:2]Lesser_hed EEQNKRLLEGMEKIVGQVHPGVKENEVYSVWSGLPSLQMADEDSRLFAFY
[Truncated_Name:3]Sus_scrofa EEQNKRLLEGMEKIVGQVHPGIKENEVYSVWSGLPSLQMADEDTRLFAFY
[Truncated_Name:4]Sea_otter  EEQNRRLLEGMEKIVGQVHPGVRENEVYSVWSGLPSLQMADEDSRLFAFY
[Truncated_Name:5]Northern_f EEQNRRLLEGMEKIVGQVHPGVKENEVYSVWSGLPSLQMADEDTRLFAFY
[Truncated_Name:6]Grey_seal  EEQNRRLLEGMEKIVGQVHPGIKENEVYSVWSGLPSLQMADEDTRLFAFY
[Truncated_Name:7]Giant_pand EEQNRRLLEGMEKIVGQVHPGVRENEVYSVWSGLPSLQMADEDTRLFAFY
                                *** ^***** ** **** ^***^* *****^~** *^*
151                             . . . . . 200

                                201 . . . . . 230
[Truncated_Name:1]Human_PRL  NLLHCLRRDSHKIDNYLKLLKCRIIHNNNC
[Truncated_Name:2]Lesser_hed NLLHCLRRDSHKIDNYLKLLKCRIVYDSNC
[Truncated_Name:3]Sus_scrofa NLLHCLRRDSHKIDNYLKLLKCRIIYDSNC
[Truncated_Name:4]Sea_otter  NLLHCLRRDSHKIDNYLKLLKCRIVYDSNC
[Truncated_Name:5]Northern_f NLLHCLRRDSHKIDNYLKLLKCRIVYDSNC
[Truncated_Name:6]Grey_seal  NLLHCLRRDSHKIDNYLKLLKCRIVYDSNC
[Truncated_Name:7]Giant_pand NLLHCLRRDSHKIDNYLKLLKCRIVYDSNC
                                *****~ **
201                             . . . . . 230

```

Call:

```
read.fasta(file = "aln-fasta.txt")
```

Class:

```
fasta
```

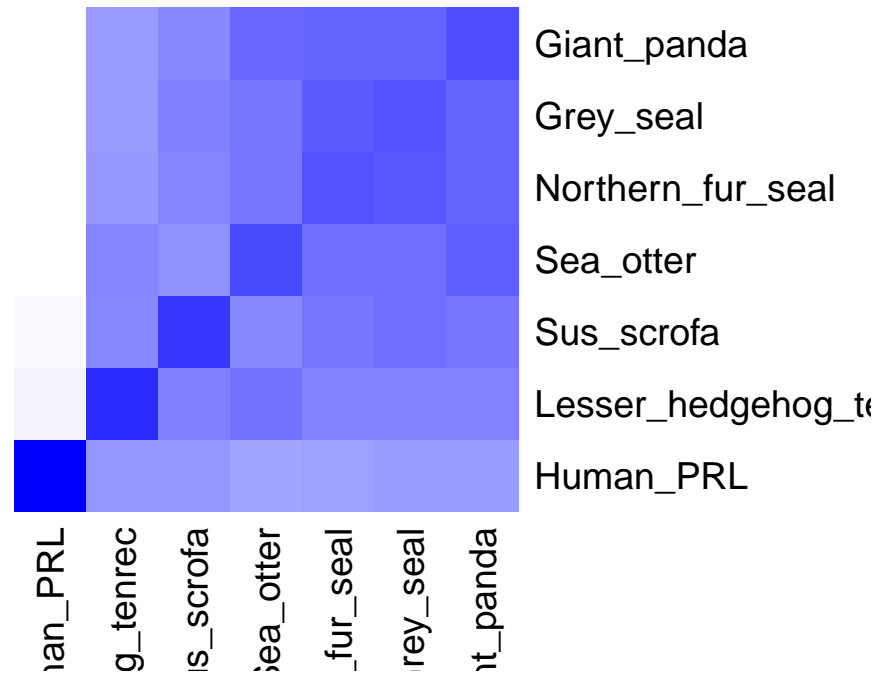
Alignment dimensions:

```
7 sequence rows; 230 position columns (227 non-gap, 3 gap)
```

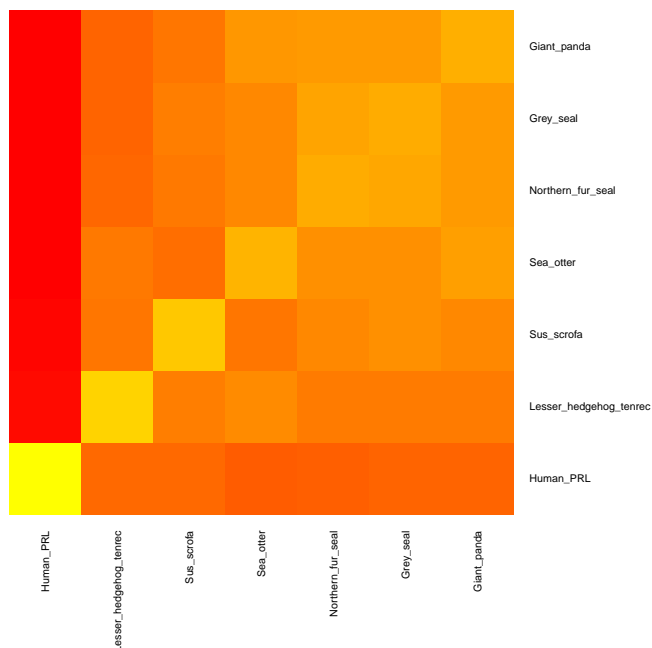
```
+ attr: id, ali, call
```

```
identity_matrix <- seqidentity(alignment)
```

```
heatmap(identity_matrix, Rowv = NA, Colv = NA, col=colorRampPalette(c("white", "blue"))(100))
```



```
heatmap(identity_matrix, Rowv = NA, Colv = NA, col = colorRampPalette(c("red", "yellow"))(100),
        labRow = alignment$Name, labCol = alignment$Name,
        cexRow = 0.4, cexCol = 0.4, # Reduce the label font size if needed
        las = 2)
```



```
sus <- read.fasta("Sus scrofa.txt")
```

```
blast_result <- blast.pdb(sus)
```

Searching ... please wait (updates every 5 seconds) RID = 7AAW5JXC016

Reporting 16 hits

```
top_hits <- head(unique(blast_result$hit),4)
top_hits
```

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_9427	1RW5_A	81.407	199	37	0	31
2	Query_9427	3EW3_A	78.788	198	42	0	32
3	Query_9427	2Q98_A	81.383	188	35	0	42
4	Query_9427	3D48_P	81.283	187	35	0	43

	q.end	s.start	s.end	evaluate	bitscore	positives	mlog.evaluate	pdb.id	acc
1	229	1	199	9.32e-125	352	91.96	285.5910	1RW5_A	1RW5_A
2	229	6	203	1.04e-118	337	89.90	271.6658	3EW3_A	3EW3_A
3	229	4	191	2.72e-116	330	92.55	266.0992	2Q98_A	2Q98_A

```
4    229          2    188 1.74e-115      328      92.51      264.2434 3D48_P 3D48_P
```

```
Cons_seq <- consensus(alignment)
print(Cons_seq$seq)
```

```
[1] "M" "D" "-" "K" "G" "-" "S" "-" "K" "-" "G" "S" "L" "L" "-" "L" "L" "L"
[19] "L" "V" "S" "-" "L" "L" "L" "C" "Q" "S" "V" "A" "S" "L" "P" "I" "C" "P"
[37] "-" "G" "A" "V" "N" "C" "Q" "V" "S" "L" "R" "D" "L" "F" "D" "R" "A" "V"
[55] "I" "L" "S" "H" "Y" "I" "H" "N" "L" "S" "S" "E" "M" "F" "N" "E" "F" "D"
[73] "K" "R" "Y" "A" "Q" "G" "R" "G" "F" "I" "T" "K" "A" "I" "N" "S" "C" "H"
[91] "T" "S" "S" "L" "S" "T" "P" "E" "D" "K" "E" "Q" "A" "Q" "Q" "I" "H" "H"
[109] "E" "D" "L" "L" "N" "L" "I" "L" "R" "V" "L" "R" "S" "W" "N" "D" "P" "L"
[127] "Y" "H" "L" "V" "T" "E" "V" "R" "G" "M" "Q" "E" "A" "P" "D" "-" "I" "L"
[145] "S" "R" "A" "I" "E" "I" "E" "E" "Q" "N" "-" "R" "L" "L" "E" "G" "M" "E"
[163] "K" "I" "V" "G" "Q" "V" "H" "P" "G" "-" "K" "E" "N" "E" "V" "Y" "S" "V"
[181] "W" "S" "G" "L" "P" "S" "L" "Q" "M" "A" "D" "E" "D" "-" "R" "L" "F" "A"
[199] "F" "Y" "N" "L" "L" "H" "C" "L" "R" "R" "D" "S" "H" "K" "I" "D" "N" "Y"
[217] "L" "K" "L" "L" "K" "C" "R" "I" "V" "Y" "D" "S" "N" "C"
```

```
BlastResult <- blast.pdb(Cons_seq$seq)
```

```
Searching ... please wait (updates every 5 seconds) RID = 7AAWCC53013
```

```
.
```

```
Reporting 14 hits
```

```
TopHits <- head(unique(BlastResult$hit),14)
TopHits
```

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens	q.start
1	Query_71831	1RW5_A	80.402	199	34	5	26
2	Query_71831	3EW3_A	78.000	200	35	6	27
3	Query_71831	2Q98_A	80.319	188	33	4	36
4	Query_71831	3D48_P	80.214	187	33	4	37
5	Query_71831	3MZG_A	80.541	185	32	4	39
6	Query_71831	3N06_A	80.000	185	33	4	39
7	Query_71831	3NOP_A	80.000	185	33	4	39
8	Query_71831	3NCB_A	80.000	185	33	4	39
9	Query_71831	1F6F_A	44.845	194	98	6	27

10	Query_71831	1HGU_A	26.562		192	123	6	37
11	Query_71831	1Z7C_A	27.225		191	117	7	37
12	Query_71831	1MMO_G	28.889		45	32	0	56
13	Query_71831	1FYZ_E	28.889		45	32	0	56
14	Query_71831	1XMF_E	28.889		45	32	0	56

	q.end	s.start	s.end	evaluate	bitscore	positives	mlog.evaluate	pdb.id	acc
1	219	1	199	2.51e-115	328.0	90.95	263.877003	1RW5_A	1RW5_A
2	219	6	203	3.69e-110	315.0	88.50	251.978734	3EW3_A	3EW3_A
3	219	4	191	1.12e-108	311.0	91.49	248.565861	2Q98_A	2Q98_A
4	219	2	188	8.22e-108	308.0	91.44	246.572620	3D48_P	3D48_P
5	219	2	186	1.06e-106	306.0	91.35	244.015751	3MZG_A	3MZG_A
6	219	2	186	1.84e-105	303.0	90.81	241.161669	3N06_A	3N06_A
7	219	2	186	1.84e-105	303.0	90.81	241.161669	3NOP_A	3NOP_A
8	219	2	186	1.84e-105	303.0	90.81	241.161669	3NCB_A	3NCB_A
9	214	5	195	4.89e-51	165.0	64.95	115.844647	1F6F_A	1F6F_A
10	217	4	188	3.01e-13	67.0	51.56	28.831666	1HGU_A	1HGU_A
11	214	4	185	2.65e-11	61.6	51.31	24.353876	1Z7C_A	1Z7C_A
12	100	64	108	6.20e+00	28.9	53.33	-1.824549	1MMO_G	1MMO_G
13	100	67	111	7.20e+00	28.5	53.33	-1.974081	1FYZ_E	1FYZ_E
14	100	66	110	7.40e+00	28.5	53.33	-2.001480	1XMF_E	1XMF_E

```
pdb.annotate("3EW3_A")
```

	structureId	chainId	macromoleculeType	chainLength	experimentalTechnique
3EW3_A	3EW3	A	Protein	203	X-ray

	resolution	scopDomain	pfam	ligandId
3EW3_A	3.8	<NA> Somatotropin hormone family (Hormone_1)		<NA>

	ligandName	source
3EW3_A	<NA>	Homo sapiens

3EW3_A the 1:2 complex between a Nterminal elongated prolactin and the extra cellular domain

	citation	rObserved	rFree	rWork
3EW3_A	Broutin, I., et al. J Biol Chem (2010)	0.256	0.3237	0.2521

	spaceGroup
3EW3_A	P 43 21 2