Class 11

Jesus (A17597539)

#PART 1

pdb|1AKE|A

121

We need some packages for today's class. These include bio3d and msa

The msa package id from the BioConductor. These packages focus on the genomics type work and are managaed by the BiocManager package.

Install install.packages("BiocManager") and then BiocManager::install("msa") all entered in the R "brain" console

```
library(bio3d)
  aa <- get.seq("1ake_A")</pre>
Warning in get.seq("lake_A"): Removing existing file: seqs.fasta
Fetching... Please wait. Done.
  aa
pdb | 1AKE | A
             \tt MRIILLGAPGAGKGTQAQFIMEKYGIPQISTGDMLRAAVKSGSELGKQAKDIMDAGKLVT
pdb | 1AKE | A
              DELVIALVKERIAQEDCRNGFLLDGFPRTIPQADAMKEAGINVDYVLEFDVPDELIVDRI
            121
```

60

60

120

120

180

180

VGRRVHAPSGRVYHVKFNPPKVEGKDDVTGEELTTRKDDQEETVRKRLVEYHQMTAPLIG

```
181
                                                   214
              YYSKEAEAGNTKYAKVDGTKPVAEVRADLEKILG
pdb|1AKE|A
            181
                                                   214
Call:
  read.fasta(file = outfile)
Class:
  fasta
Alignment dimensions:
  1 sequence rows; 214 position columns (214 non-gap, 0 gap)
+ attr: id, ali, call
Now I can search the PDB database for related sequences:
  #b <- blast.pdb(aa)</pre>
  #hits <- plot(b)</pre>
  #attributes(b)
  #head(b$hit.tbl)
These are the related structures in the PDB database that we found via a BLAST search...
```

```
hits <- NULL
hits$pdb.id <- c('1AKE_A','6S36_A','6RZE_A','3HPR_A','1E4V_A','5EJE_A','1E4Y_A','3X2S_A','
```

Side-note: Lets annotate these structures (in other words find out what they are, what species they are from, stuff about the experiment they were solved in etc.)

For this we can use the pdb.annotate()

```
anno <- pdb.annotate(hits$pdb.id)</pre>
attributes(anno)
```

\$names

```
[1] "structureId"
                             "chainId"
                                                      "macromoleculeType"
 [4] "chainLength"
                             "experimentalTechnique" "resolution"
 [7] "scopDomain"
                             "pfam"
                                                      "ligandId"
[10] "ligandName"
                             "source"
                                                      "structureTitle"
[13] "citation"
                             "rObserved"
                                                      "rFree"
[16] "rWork"
                             "spaceGroup"
$class
[1] "data.frame"
```

\$row.names

- [1] "1AKE_A" "6S36_A" "6RZE_A" "3HPR_A" "1E4V_A" "5EJE_A" "1E4Y_A" "3X2S_A"
- [9] "6HAP_A" "6HAM_A" "4K46_A" "3GMT_A" "4PZL_A"

head(anno)

	structureId	chainId	macromo	leculeType	chainLen	øth exi	perime	ental	Technique
1AKE_A	1AKE	A	maor omo.	Protein		214	, , , , , , , , ,	J11 0 W.1	X-ray
6S36_A		A		Protein		214			X-ray
6RZE_A		A		Protein		214			X-ray
3HPR_A		A		Protein		214			X-ray
1E4V_A		A		Protein		214			X-ray
5EJE A	5EJE	A		Protein		214			X-ray
_	resolution	sco	pDomain						pfam
1AKE_A	2.00 A		-	Adenylate	kinase,	active	site	lid	-
6S36_A	1.60	v	<na></na>	Adenylate	kinase,	active	site	lid	(ADK_lid)
6RZE_A	1.69		<na></na>	Adenylate	kinase,	active	site	lid	(ADK_lid)
3HPR_A	2.00		<na></na>	Adenylate	kinase,	active	site	lid	(ADK_lid)
1E4V_A	1.85 A	denylate	kinase	Adenylate	kinase,	active	site	lid	(ADK_lid)
5EJE_A	1.90		<na></na>	Adenylate	kinase,	active	site	lid	(ADK_lid)
	ligandId ligandName							lame	
1AKE_A		AP5		BIS(ADENOSINE)-5'-PI	ENTAPI	HOSPE	IATE
6S36_A	CL (3), NA, MG	(2)	CHLORID	E ION (3),	SODIUM IO	N,MAGNI	ESIUM	ION	(2)
6RZE_A	NA (3),CL	. (2)		SOD	IUM ION (3),CHL	ORIDE	ION	(2)
3HPR_A		AP5		BIS(ADENOSINE)-5'-PI	ENTAPI	HOSPE	IATE
1E4V_A		AP5		BIS(ADENOSINE)-5'-PI	ENTAPI	HOSPE	IATE
5EJE_A	AP	5,CO BIS	(ADENOS	INE)-5'-PEI	NTAPHOSPH.	ATE,CO	BALT ((II)	ION
source									
1AKE_A	Escherichia coli								
6S36_A	S_A Escherichia coli								
6RZE_A	RZE_A Escherichia coli								

```
3HPR_A
                        Escherichia coli K-12
1E4V_A
                             Escherichia coli
5EJE_A Escherichia coli 0139:H28 str. E24377A
1AKE A STRUCTURE OF THE COMPLEX BETWEEN ADENYLATE KINASE FROM ESCHERICHIA COLI AND THE INHIB
6S36 A
6RZE A
3HPR_A
1E4V_A
5EJE_A
                                                                                          Crys
                                                     citation rObserved rFree
1AKE_A
                      Muller, C.W., et al. J Mol Biol (1992)
                                                                 0.1960
                       Rogne, P., et al. Biochemistry (2019)
6S36_A
                                                                 0.1632 0.2356
                       Rogne, P., et al. Biochemistry (2019)
6RZE A
                                                                 0.1865 0.2350
3HPR_A Schrank, T.P., et al. Proc Natl Acad Sci U S A (2009)
                                                                 0.2100 0.2432
1E4V_A
                       Muller, C.W., et al. Proteins (1993)
                                                                 0.1960
                                                                            NA
5EJE_A Kovermann, M., et al. Proc Natl Acad Sci U S A (2017)
                                                                 0.1889 0.2358
        rWork spaceGroup
1AKE_A 0.1960 P 21 2 21
6S36_A 0.1594
                C 1 2 1
6RZE_A 0.1819
                 C 1 2 1
3HPR_A 0.2062 P 21 21 2
1E4V_A 0.1960 P 21 2 21
5EJE_A 0.1863 P 21 2 21
Now we can download all these structures for further analysis with the get.pdb() function
  # Download releated PDB files
  files <- get.pdb(hits$pdb.id, path="pdbs", split=TRUE, gzip=TRUE)</pre>
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/1AKE.pdb.gz exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6S36.pdb.gz exists. Skipping download
Warning in get.pdb(hits$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):
pdbs/6RZE.pdb.gz exists. Skipping download
```

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE):

pdbs/3HPR.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/1E4V.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/5EJE.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/1E4Y.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/3X2S.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/6HAP.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/6HAM.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/4K46.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/3GMT.pdb.gz exists. Skipping download

Warning in get.pdb(hits\$pdb.id, path = "pdbs", split = TRUE, gzip = TRUE): pdbs/4PZL.pdb.gz exists. Skipping download



١	=======================================		46%
			54%
	=======================================		62%
	=======================================		69%
-			
-	=======================================		77%
١			
١	=======================================		85%
١			
١			92%
ĺ			
ĺ		:	100%

Now we have all these related structures we can align and Superpose...

```
# Align releated PDBs
pdbs <- pdbaln(files, fit = TRUE, exefile="msa")</pre>
```

```
Reading PDB files:
pdbs/split_chain/1AKE_A.pdb
pdbs/split_chain/6S36_A.pdb
pdbs/split_chain/6RZE_A.pdb
pdbs/split_chain/3HPR_A.pdb
pdbs/split_chain/1E4V_A.pdb
pdbs/split_chain/5EJE_A.pdb
pdbs/split_chain/1E4Y_A.pdb
pdbs/split_chain/3X2S_A.pdb
pdbs/split_chain/6HAP_A.pdb
pdbs/split_chain/6HAM_A.pdb
pdbs/split_chain/4K46_A.pdb
pdbs/split_chain/3GMT_A.pdb
pdbs/split_chain/4PZL_A.pdb
   PDB has ALT records, taking A only, rm.alt=TRUE
     PDB has ALT records, taking A only, rm.alt=TRUE
       PDB has ALT records, taking A only, rm.alt=TRUE
```

. PDB has ALT records, taking A only, rm.alt=TRUE \dots

Extracting sequences

pdb/seq: 1 name: pdbs/split_chain/1AKE_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 2 name: pdbs/split_chain/6S36_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 3 name: pdbs/split_chain/6RZE_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 4 name: pdbs/split_chain/3HPR_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 5 name: pdbs/split_chain/1E4V_A.pdb pdb/seq: 6 name: pdbs/split_chain/5EJE_A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 7 name: pdbs/split_chain/1E4Y_A.pdb pdb/seq: 8 name: pdbs/split_chain/3X2S_A.pdb pdb/seq: 9 name: pdbs/split_chain/6HAP_A.pdb pdb/seq: 10 name: pdbs/split chain/6HAM A.pdb PDB has ALT records, taking A only, rm.alt=TRUE pdb/seq: 11 name: pdbs/split chain/4K46 A.pdb PDB has ALT records, taking A only, rm.alt=TRUE name: pdbs/split_chain/3GMT_A.pdb pdb/seq: 12 pdb/seq: 13 name: pdbs/split_chain/4PZL_A.pdb

pdbs

[Truncated_Name:1]1AKE_A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated Name:2]6S36 A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:3]6RZE_A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated Name:4]3HPR A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:5]1E4V_A.pdb ----MRIILLGAPVAGKGTQAQFIMEKYGIPQIS [Truncated_Name: 6] 5EJE_A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:7]1E4Y_A.pdb -----MRIILLGALVAGKGTQAQFIMEKYGIPQIS [Truncated_Name:8]3X2S_A.pdb -----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:9]6HAP_A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:10]6HAM_A.pdb ----MRIILLGAPGAGKGTQAQFIMEKYGIPQIS [Truncated_Name:11]4K46_A.pdb -----MRIILLGAPGAGKGTQAQFIMAKFGIPQIS [Truncated_Name: 12] 3GMT_A.pdb -----MRLILLGAPGAGKGTQANFIKEKFGIPQIS

TENLYFQ	SNAMRIILLGA	PGAGKGTQAI	KIIEG	(KYNIA	HIS
	^**	*****	*	*^ *	**
1	•				40
41		•			80
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDACKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDCGKI	LVTDE	ELVIAL	VKE
TGDMLRA	.AVKSGSELGKQ	AKDIMDAGKI	LVTDE	ELVIAL	VRE
TGDMLRA	AIKSGSELGKQ	AKDIMDAGKI	LVTDE	CIIIAL	VKE
TGDMLRA	AIKAGTELGKQ	AKSVIDAGQI	LVSDI	OIILGL	VKE
TGDMLRA	AVKAGTPLGVE	AKTYMDEGKI	LVPDS	SLIIGL	VKE
TGDMIRE	TIKSGSALGQE	LKKVLDAGEI	LVSDE	EFIIKI	VKD
****^*	^* *^ **	* ^* >	** *	^^ ^	*^^
41	•				80
81	•				120
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKE	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDO	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQEDS	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RICQEDS	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RICQEDS	RNGFLLDGFPR	TIPQADAMKI	EAGIN	IVDYVL	EFD
RIAQDDC	AKGFLLDGFPR	TIPQADGLKI	EVGVV	VDYVI	EFD
RLKEADO	ANGYLFDGFPR	TIAQADAMKI	EAGVA	IDYVL	EID
RISKNDO	NNGFLLDGVPR	TIPQAQELD	KLGVN	IDYIV	EVD
*^ *					
81	•				120
121	•				160
VPDELIV	DRIVGRRVHAP	SGRVYHVKFI	IPPKV	EGKDD	VTG
VPDELIV	DKIVGRRVHAP	SGRVYHVKFI	IPPKV	EGKDD	VTG
VPDELIV	DAIVGRRVHAP	SGRVYHVKFI	IPPKV	EGKDD	VTG
VPDFI TV	DRIVCRRVHAP	SCRVYHVKFI	IPPKI	ובכאטט	CTC
	1 41 TGDMLRA T	******* 1 41 TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAIKSGSELGKQ TGDMLRAAVKSGSELGKQ TGDMLRAAVESELGKQ TGDMLRAAVES TGDMLRAAVKSGSELGKQ TGDMLRAAVES TGDMLRAAVES TGDMLRAAVES	******* ******************************	****** ****** ** 1	1 41 TGDMLRAAVKSGSELGKQAKDIMDAGKLVTDELVIAL TGDMLRAAIKAGTELGKQAKSVIDAGQLVSDDIILGL TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGL TGDMLRAAVKAGTPLGVEAKTYMDEGKLVPDSLIIGL TGDMLRETIKSGSALGQELKKVLDAGELVSDEFIIKI ****^

[Truncated_Name:5]1E4V_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated_Name:6]5EJE_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated_Name:7]1E4Y_A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated Name:8]3X2S A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated Name:9]6HAP A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated Name:10]6HAM A.pdb VPDELIVDRIVGRRVHAPSGRVYHVKFNPPKVEGKDDVTG [Truncated Name:11]4K46 A.pdb VADSVIVERMAGRRAHLASGRTYHNVYNPPKVEGKDDVTG [Truncated_Name:12]3GMT_A.pdb VPFSEIIERMSGRRTHPASGRTYHVKFNPPKVEGKDDVTG [Truncated Name:13]4PZL A.pdb VADNLLIERITGRRIHPASGRTYHTKFNPPKVADKDDVTG ^^^ ^ *** * *** ** ^**** *** ** 121 160 161 200 [Truncated_Name:1]1AKE_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:2]6S36_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:3]6RZE_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:4]3HPR_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:5]1E4V_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:6]5EJE_A.pdb EELTTRKDDQEECVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated Name:7]1E4Y A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated Name:8]3X2S A.pdb EELTTRKDDQEETVRKRLCEYHQMTAPLIGYYSKEAEAGN [Truncated Name:9]6HAP A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:10]6HAM_A.pdb EELTTRKDDQEETVRKRLVEYHQMTAPLIGYYSKEAEAGN [Truncated_Name:11]4K46_A.pdb EDLVIREDDKEETVLARLGVYHNQTAPLIAYYGKEAEAGN [Truncated_Name: 12] 3GMT_A.pdb EPLVQRDDDKEETVKKRLDVYEAQTKPLITYYGDWARRGA [Truncated_Name: 13] 4PZL_A.pdb EPLITRTDDNEDTVKQRLSVYHAQTAKLIDFYRNFSSTNT * * * * * * * * * ** ^* 161 200 201 227 [Truncated_Name:1]1AKE_A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated_Name:2]6S36_A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated_Name:3]6RZE_A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated Name: 4] 3HPR A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated Name:5]1E4V A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated Name: 6] 5EJE A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated Name:7]1E4Y A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated Name:8]3X2S A.pdb T--KYAKVDGTKPVAEVRADLEKILG-[Truncated_Name:9]6HAP_A.pdb T--KYAKVDGTKPVCEVRADLEKILG-[Truncated_Name:10]6HAM_A.pdb T--KYAKVDGTKPVCEVRADLEKILG-[Truncated_Name:11]4K46_A.pdb T--QYLKFDGTKAVAEVSAELEKALA-[Truncated_Name: 12] 3GMT_A.pdb E----YRKISG-

KIPKYIKINGDQAVEKVSQDIFDQLNK

[Truncated_Name: 13] 4PZL_A.pdb

Call:
 pdbaln(files = files, fit = TRUE, exefile = "msa")

Class:
 pdbs, fasta

Alignment dimensions:
 13 sequence rows; 227 position columns (204 non-gap, 23 gap)

+ attr: xyz, resno, b, chain, id, ali, resid, sse, call

Vector containing PDB codes for figure axis ids <- basename.pdb(pdbs\$id)

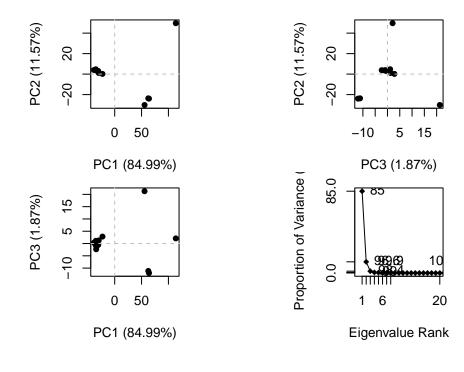
Draw schematic alignment #plot(pdbs, labels=ids)

##Principal Component Analysis

Perform PCA

pc.xray <- pca(pdbs)</pre>

plot(pc.xray)



```
# Visualize first principal component
pc1 <- mktrj(pc.xray, pc=1, file="pc_1.pdb")</pre>
```

And a nicer PC plot

```
# Calculate RMSD
rd <- rmsd(pdbs)</pre>
```

Warning in rmsd(pdbs): No indices provided, using the 204 non NA positions

```
# Structure-based clustering
hc.rd <- hclust(dist(rd))
grps.rd <- cutree(hc.rd, k=3)

plot(pc.xray, 1:2, col="grey50", bg=grps.rd, pch=21, cex=1)</pre>
```