

Find a Gene Project: Final

Yoonjin Lim (PID: A16850635)

Table of contents

Question 7: Creating Heatmap

```
library(bio3d)

pth <- "QuerySequence_93902/"
pdb.files <- list.files(path=pth, full.names = TRUE, pattern = ".pdb")
```

Align and superpose all these models.

```
file.exists(pdb.files)
```

```
[1] TRUE TRUE TRUE TRUE TRUE
```

```
pdbbs <- pdbaln(pdb.files, fit=TRUE, exefile="msa")
```

Reading PDB files:

```
QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb
QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb
QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_004_alphafold2_ptm_model_1_seed_000.pdb
QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
.....
```

Extracting sequences

```
pdb/seq: 1    name: QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
pdb/seq: 2    name: QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb
```

```

pdb/seq: 3   name: QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_003_alphafold2_pt
pdb/seq: 4   name: QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_004_alphafold2_pt
pdb/seq: 5   name: QuerySequence_93902//QuerySequence_93902_unrelaxed_rank_005_alphafold2_pt

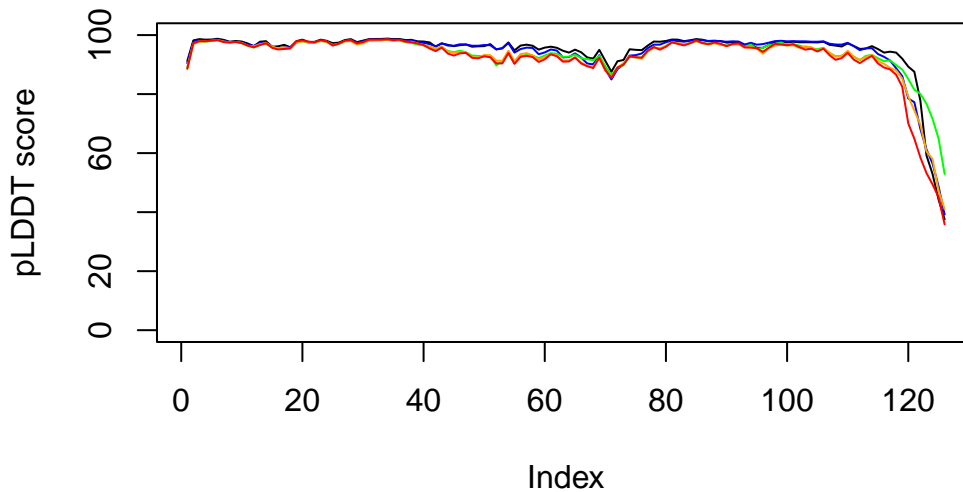
```

```
#view.pdbs(pdbs)
```

```

plot (pdbs$b[1,], typ = "l", ylim=c(0,100), ylab="pLDDT score")
lines (pdbs$b[2,], typ = "l", col="blue")
lines (pdbs$b[3,], typ = "l", col="green")
lines (pdbs$b[4,], typ = "l", col="orange")
lines (pdbs$b[5,], typ = "l", col="red")

```



```

fasta <- bio3d::read.fasta("muscle-I20250525-034425-0597-74261255-p1m.fa")
# fasta <- read.fasta(system.file("muscle-I20250525-034425-0597-74261255-p1m.fa", package="b

```

```
matrix <- seqidentity(fasta$ali)
```

```

# pheatmap(matrix,
#           cluster_rows = TRUE,
#           cluster_cols = TRUE,
#           main = "Sequence Identity Heatmap",

```

```
#      fontsize = 7,
#      fontsize_number = 7,
#      cellwidth = 15,
#      cellheight = 15)
```

Question 8:

```
library(bio3d)
```

```
blast <- blast.pdb(fasta)
```

Warning in blast.pdb(fasta): Multiple sequences detected - using only the first sequence in input object

```
Searching ... please wait (updates every 5 seconds) RID = 37M8NBVF016
.....
Reporting 181 hits
```

```
attributes(blast)
```

```
$names
[1] "hit.tbl" "raw"      "url"
```

```
$class
[1] "blast"
```

```
head(blast$hit.tbl,10)
```

	queryid	subjectids	identity	alignmentlength	mismatches	gapopens
1	Query_5745927	3RGK_A	98.4	125	2	0
2	Query_5745927	1MWC_A	93.6	125	8	0
3	Query_5745927	1MYH_A	92.8	125	9	0
4	Query_5745927	1MYI_A	92.8	125	9	0
5	Query_5745927	1MYJ_A	92.8	125	9	0
6	Query_5745927	1M6C_A	92.8	125	9	0
7	Query_5745927	1MNJ_A	92.0	125	10	0
8	Query_5745927	1MNK_A	92.0	125	10	0
9	Query_5745927	5YCG_A	90.4	125	12	0
10	Query_5745927	1MNH_A	92.0	125	10	0

	q.start	q.end	s.start	s.end	evalue	bitscore	positives	mlog.evalue	pdb.id
1	1	125	27	151	6.05e-86	248	99.2	196.2223	3RGK_A
2	1	125	27	151	1.76e-81	237	96.0	185.9441	1MWC_A
3	1	125	27	151	3.26e-81	236	96.0	185.3277	1MYH_A
4	1	125	27	151	6.36e-81	235	95.2	184.6594	1MYI_A
5	1	125	27	151	7.75e-81	235	95.2	184.4617	1MYJ_A
6	1	125	27	151	2.22e-80	234	95.2	183.4093	1M6C_A
7	1	125	27	151	8.65e-80	233	95.2	182.0492	1MNJ_A
8	1	125	27	151	1.93e-79	231	94.4	181.2467	1MNK_A
9	1	125	31	155	2.23e-79	231	95.2	181.1022	5YCG_A
10	1	125	27	151	4.06e-79	231	94.4	180.5030	1MNH_A

```

acc
1 3RGK_A
2 1MWC_A
3 1MYH_A
4 1MYI_A
5 1MYJ_A
6 1M6C_A
7 1MNJ_A
8 1MNK_A
9 5YCG_A
10 1MNH_A

```

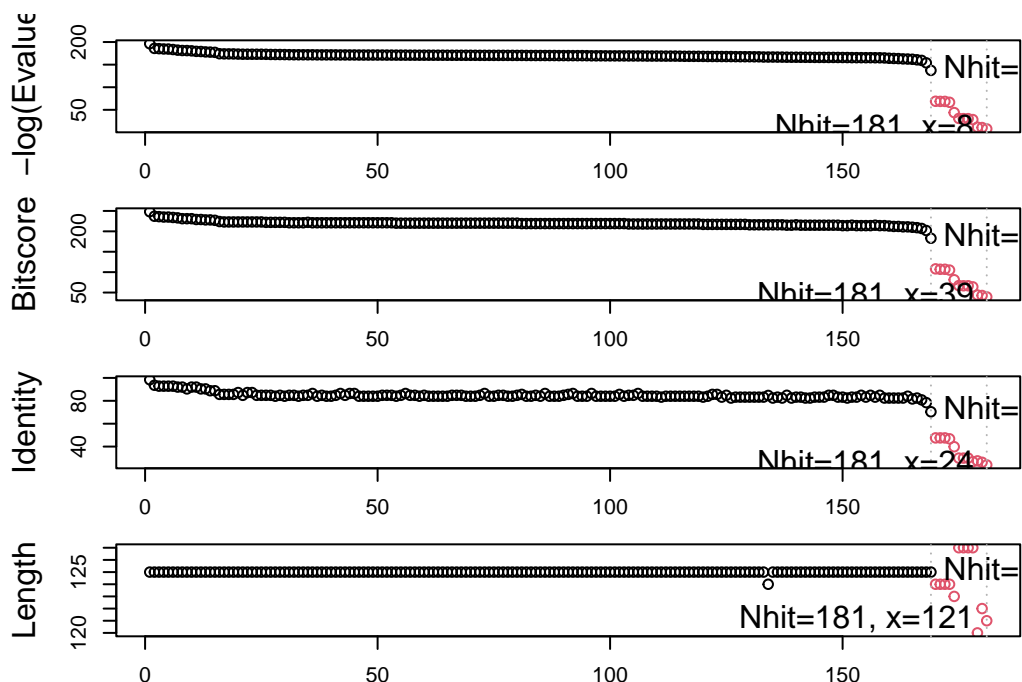
```
hits<- plot(blast)
```

```

* Possible cutoff values:    137 7
    Yielding Nhits:        169 181

* Chosen cutoff value of:    137
    Yielding Nhits:        169

```



hits\$ pdb.id

```
[1] "3RGK_A" "1MWC_A" "1MYH_A" "1MYI_A" "1MYJ_A" "1M6C_A" "1MNJ_A" "1MNK_A"
[9] "5YCG_A" "1MNH_A" "1MNI_A" "7DDT_A" "7DDR_A" "5YCH_A" "5YCI_A" "1A6K_A"
[17] "104M_A" "1UFP_A" "7CEZ_A" "4DC7_A" "5XKV_A" "5YL3_A" "1AZI_A" "5B85_A"
[25] "4QAU_A" "1A6G_A" "4OF9_A" "5XKW_A" "109M_A" "1CIO_A" "1J3F_A" "4OOD_A"
[33] "1MLM_A" "5B84_A" "8KFH_A" "1NZ2_A" "2OHA_A" "4H07_A" "2SPO_A" "1C09_A"
[41] "7XCF_A" "1RSE_A" "3089_A" "1YMA_A" "1HRM_A" "2BLH_A" "1D01_A" "1CH3_A"
[49] "1LTW_A" "1CPW_A" "2OH8_A" "4PQ6_A" "4PQC_A" "1MLQ_A" "4IT8_A" "6LTL_A"
[57] "3A2G_A" "5YZF_A" "6JP1_A" "1WVP_A" "1MLF_A" "1CP5_A" "1CH2_A" "1JDO_A"
[65] "2OHB_A" "3SDN_A" "7XC9_A" "7XCQ_A" "2OH9_A" "1MGN_A" "1C08_A" "8XBJ_A"
[73] "1XCH_A" "1MTI_A" "1CIK_A" "50J9_A" "1DTM_A" "4LPI_A" "6KRC_A" "1IRC_A"
[81] "1NZ3_A" "5UTD_A" "106M_A" "5ZEO_A" "101M_A" "1BJE_A" "2MGF_A" "1CH9_A"
[89] "1CPO_A" "6BMG_A" "3WYO_B" "5Z7E_A" "5ZZF_A" "1MCY_A" "3NML_A" "3HC9_A"
[97] "5KD1_A" "1CH1_A" "1MOC_A" "1DTI_A" "102M_A" "5Z7F_A" "1CH7_A" "3RJ6_A"
[105] "4TWV_A" "3HEN_A" "7DDU_A" "1MOB_A" "1CH5_A" "2MGJ_A" "5UTB_A" "7VW4_A"
[113] "2MGC_A" "7EHX_A" "8QBA_A" "4PQB_A" "3MNO_A" "4FWZ_A" "6G5T_A" "1H1X_A"
[121] "5HLQ_A" "7V5P_A" "3RJN_B" "1LUE_A" "1MBS_A" "2E2Y_A" "4FWX_A" "2W6Y_A"
[129] "1MLU_A" "8EKO_A" "1016_A" "6M8F_A" "1FCS_A" "1EMY_A" "7VUC_A" "8KFJ_A"
[137] "4TYX_A" "4NS2_A" "10BM_A" "50JA_A" "3H57_A" "1DXC_A" "7SLH_A" "6Z4R_A"
[145] "10FJ_A" "10FK_A" "7DGM_A" "7DGJ_A" "3K9Z_A" "6Z4T_A" "2W6X_A" "6G5A_A"
[153] "7KYR_A" "6LS8_A" "6F17_A" "7V5Q_A" "8KFI_A" "7V5R_A" "7L3U_A" "7SPE_A"
```

[161] "6D45_A" "50JB_A" "3M38_A" "3WYO_A" "7SLI_A" "7YLK_A" "7SPG_A" "50JC_A"
[169] "1LHS_A"

Question 9: