Class 11

Josefina O'Toole

Here we analyze our AlphaFold structure prediction models. The input directory/folder comes from the ColabFold server:

```r
# Change this for YOUR results dir name
results_dir <- "FindAGene_e2383"
```

```r
# File names for all PDB models
pdb_files <- list.files(path=results_dir,
                        pattern="*.pdb",
                        full.names = TRUE)

# Print our PDB file names
basename(pdb_files)
```

```
[1] "test_e2383_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "test_e2383_unrelaxed_rank_002_alphafold2_ptm_model_1_seed_000.pdb"
[3] "test_e2383_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb"
[4] "test_e2383_unrelaxed_rank_004_alphafold2_ptm_model_4_seed_000.pdb"
[5] "test_e2383_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

I will use the Bio3D package for analysis

```r
library(bio3d)
```

Align and Superpose

```r
# Read all data from Models
#  and superpose/fit coords
pdbs <- pdbaln(pdb_files, fit=TRUE, exefile="msa")
```

```
Reading PDB files:
FindAGene_e2383/test_e2383_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
FindAGene_e2383/test_e2383_unrelaxed_rank_002_alphafold2_ptm_model_1_seed_000.pdb
FindAGene_e2383/test_e2383_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_000.pdb
FindAGene_e2383/test_e2383_unrelaxed_rank_004_alphafold2_ptm_model_4_seed_000.pdb
FindAGene_e2383/test_e2383_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
.....

Extracting sequences

pdb/seq: 1    name: FindAGene_e2383/test_e2383_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_
pdb/seq: 2    name: FindAGene_e2383/test_e2383_unrelaxed_rank_002_alphafold2_ptm_model_1_seed_
pdb/seq: 3    name: FindAGene_e2383/test_e2383_unrelaxed_rank_003_alphafold2_ptm_model_3_seed_
pdb/seq: 4    name: FindAGene_e2383/test_e2383_unrelaxed_rank_004_alphafold2_ptm_model_4_seed_
pdb/seq: 5    name: FindAGene_e2383/test_e2383_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_
```

pdbs

```
                                    1         .         .         .         .        50
[Truncated_Name:1]test_e2383      MSPHPTALLGLVLCLAQTIHTQEEPLPRPSISAEPGTVIPPGRPVTIVCR
[Truncated_Name:2]test_e2383      MSPHPTALLGLVLCLAQTIHTQEEPLPRPSISAEPGTVIPPGRPVTIVCR
[Truncated_Name:3]test_e2383      MSPHPTALLGLVLCLAQTIHTQEEPLPRPSISAEPGTVIPPGRPVTIVCR
[Truncated_Name:4]test_e2383      MSPHPTALLGLVLCLAQTIHTQEEPLPRPSISAEPGTVIPPGRPVTIVCR
[Truncated_Name:5]test_e2383      MSPHPTALLGLVLCLAQTIHTQEEPLPRPSISAEPGTVIPPGRPVTIVCR
                                  **************************************************
                                    1         .         .         .         .        50

                                   51         .         .         .         .       100
[Truncated_Name:1]test_e2383       GPVGVDQFRLEREDRSKFNDTKDVSQASPSESEARFRIDSVSEGNTRRYC
[Truncated_Name:2]test_e2383       GPVGVDQFRLEREDRSKFNDTKDVSQASPSESEARFRIDSVSEGNTRRYC
[Truncated_Name:3]test_e2383       GPVGVDQFRLEREDRSKFNDTKDVSQASPSESEARFRIDSVSEGNTRRYC
[Truncated_Name:4]test_e2383       GPVGVDQFRLEREDRSKFNDTKDVSQASPSESEARFRIDSVSEGNTRRYC
[Truncated_Name:5]test_e2383       GPVGVDQFRLEREDRSKFNDTKDVSQASPSESEARFRIDSVSEGNTRRYC
                                   **************************************************
                                   51         .         .         .         .       100

                                  101         .         .         .         .       150
[Truncated_Name:1]test_e2383       CLYRKSAEWSERSDYLDLVVKETSGDTDSPVTEPDSSAGPTQRPSDNSHN
[Truncated_Name:2]test_e2383       CLYRKSAEWSERSDYLDLVVKETSGDTDSPVTEPDSSAGPTQRPSDNSHN
[Truncated_Name:3]test_e2383       CLYRKSAEWSERSDYLDLVVKETSGDTDSPVTEPDSSAGPTQRPSDNSHN
[Truncated_Name:4]test_e2383       CLYRKSAEWSERSDYLDLVVKETSGDTDSPVTEPDSSAGPTQRPSDNSHN
[Truncated_Name:5]test_e2383       CLYRKSAEWSERSDYLDLVVKETSGDTDSPVTEPDSSAGPTQRPSDNSHN
```

```
                           **************************************************
                           101           .           .           .           .         150


                           151           .           .           .           .         200
[Truncated_Name:1]test_e2383    EHAPASQGLSAEHLYILIGVSVVFLFCLLLLVLFFLHRQNQMKQGPPRSK
[Truncated_Name:2]test_e2383    EHAPASQGLSAEHLYILIGVSVVFLFCLLLLVLFFLHRQNQMKQGPPRSK
[Truncated_Name:3]test_e2383    EHAPASQGLSAEHLYILIGVSVVFLFCLLLLVLFFLHRQNQMKQGPPRSK
[Truncated_Name:4]test_e2383    EHAPASQGLSAEHLYILIGVSVVFLFCLLLLVLFFLHRQNQMKQGPPRSK
[Truncated_Name:5]test_e2383    EHAPASQGLSAEHLYILIGVSVVFLFCLLLLVLFFLHRQNQMKQGPPRSK
                           **************************************************
                           151           .           .           .           .         200


                           201           .           .           .           .         250
[Truncated_Name:1]test_e2383    DEEQKLQQRPDLAVDVLERTADKATVNGLPEKATINGLPEKDREMDTSAP
[Truncated_Name:2]test_e2383    DEEQKLQQRPDLAVDVLERTADKATVNGLPEKATINGLPEKDREMDTSAP
[Truncated_Name:3]test_e2383    DEEQKLQQRPDLAVDVLERTADKATVNGLPEKATINGLPEKDREMDTSAP
[Truncated_Name:4]test_e2383    DEEQKLQQRPDLAVDVLERTADKATVNGLPEKATINGLPEKDREMDTSAP
[Truncated_Name:5]test_e2383    DEEQKLQQRPDLAVDVLERTADKATVNGLPEKATINGLPEKDREMDTSAP
                           **************************************************
                           201           .           .           .           .         250


                           251           .           .           .           .         296
[Truncated_Name:1]test_e2383    AAGDPQEVTYAQLDHWALTRRTAQAVSPQSTEPMAESSTYAAVARH
[Truncated_Name:2]test_e2383    AAGDPQEVTYAQLDHWALTRRTAQAVSPQSTEPMAESSTYAAVARH
[Truncated_Name:3]test_e2383    AAGDPQEVTYAQLDHWALTRRTAQAVSPQSTEPMAESSTYAAVARH
[Truncated_Name:4]test_e2383    AAGDPQEVTYAQLDHWALTRRTAQAVSPQSTEPMAESSTYAAVARH
[Truncated_Name:5]test_e2383    AAGDPQEVTYAQLDHWALTRRTAQAVSPQSTEPMAESSTYAAVARH
                           *********************************************
                           251           .           .           .           .         296


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

Class:
  pdbs, fasta

Alignment dimensions:
  5 sequence rows; 296 position columns (296 non-gap, 0 gap)

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

RMSD

```
rd <- rmsd(pdbs, fit=T)
```

```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 296 non NA positions
```

```
range(rd)
```

```
[1]   0.000 30.793
```

```
library(pheatmap)

colnames(rd) <- paste0("m",1:5)
rownames(rd) <- paste0("m",1:5)
pheatmap(rd)
```