Class 11

Isabel Philip (A16855684)

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

```
# Change this for YOUR results dir name
results_dir <- "HIVPr_94b5b"
```

```
# 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] "HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

I will use the Bio3D pacakge for analysis

```
library(bio3d)
```

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

```
Reading PDB files:
HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb
HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb
HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb
HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb
HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb
.....

Extracting sequences

pdb/seq: 1    name: HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
pdb/seq: 2    name: HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
pdb/seq: 3    name: HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
pdb/seq: 4    name: HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
pdb/seq: 5    name: HIVPr_94b5b/HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
```

pdbs

```
                                      1        .         .         .         .          50
[Truncated_Name:1]HIVPr_94b5     PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:2]HIVPr_94b5     PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:3]HIVPr_94b5     PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:4]HIVPr_94b5     PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
[Truncated_Name:5]HIVPr_94b5     PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGI
                                 *************************************************
                                      1        .         .         .         .          50


                                      51       .         .         .         .          99
[Truncated_Name:1]HIVPr_94b5      GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:2]HIVPr_94b5      GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:3]HIVPr_94b5      GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:4]HIVPr_94b5      GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
[Truncated_Name:5]HIVPr_94b5      GGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF
                                 *************************************************
                                      51       .         .         .         .          99

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

Class:
  pdbs, fasta
```
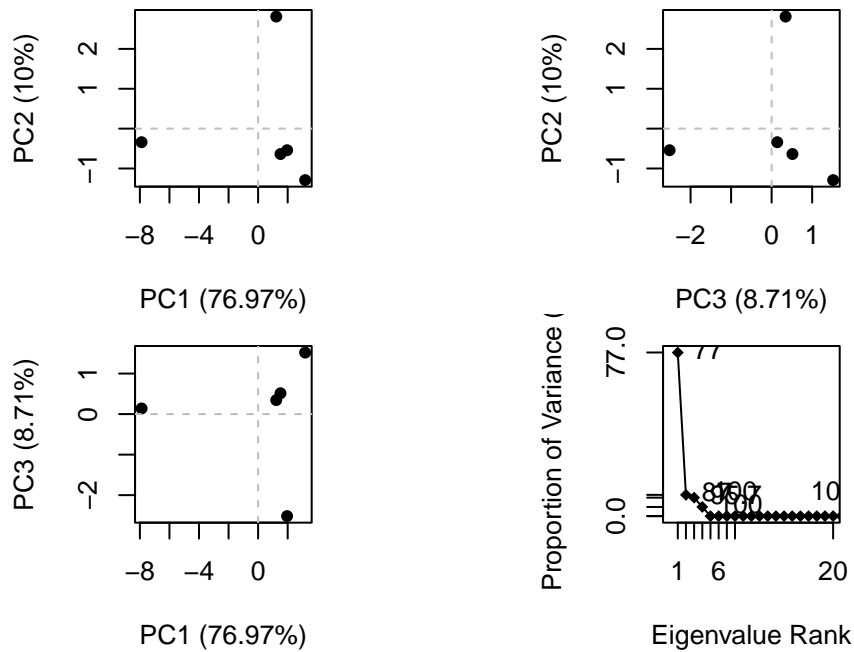
```
Alignment dimensions:
  5 sequence rows; 99 position columns (99 non-gap, 0 gap)

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

A quick

```
pc <- pca(pdbs)
plot(pc)
```


##RMSD analysis RMSD is a common measure of structural distance used in structural biology.

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

```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 99 non NA positions
```

```
rd
```

```
                                                  HIVPr_94b5b_unrelaxed_rank_00
HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
```

```
HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        HIVPr_94b5b_unrelaxed_rank_002
HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        HIVPr_94b5b_unrelaxed_rank_003
HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        HIVPr_94b5b_unrelaxed_rank_004
HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
                                                                        HIVPr_94b5b_unrelaxed_rank_005
HIVPr_94b5b_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000
HIVPr_94b5b_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000
HIVPr_94b5b_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000
HIVPr_94b5b_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000
HIVPr_94b5b_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000
```

```r
library(pheatmap)

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