AlphaFold Analysis

Dora Deng (A17445600)

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

```
results_dir <- "hivpr_monomer_94b5b_1"
```

```
# 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_monomer_94b5b_1_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_000.pdb"
[2] "hivpr_monomer_94b5b_1_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_000.pdb"
[3] "hivpr_monomer_94b5b_1_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_000.pdb"
[4] "hivpr_monomer_94b5b_1_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_000.pdb"
[5] "hivpr_monomer_94b5b_1_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_000.pdb"
```

I will use the Bio3D package 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_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_001_alphafold2_ptm_model_5_seed_00
hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_002_alphafold2_ptm_model_4_seed_00
```

```
hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_003_alphafold2_ptm_model_1_seed_0
hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_004_alphafold2_ptm_model_3_seed_0
hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_005_alphafold2_ptm_model_2_seed_0
.....

Extracting sequences

pdb/seq: 1   name: hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_001_alphafold2
pdb/seq: 2   name: hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_002_alphafold2
pdb/seq: 3   name: hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_003_alphafold2
pdb/seq: 4   name: hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_004_alphafold2
pdb/seq: 5   name: hivpr_monomer_94b5b_1/hivpr_monomer_94b5b_1_unrelaxed_rank_005_alphafold2
```

RMSD analysis

RMSD is a common measure of stuctural distance used in structual biology

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

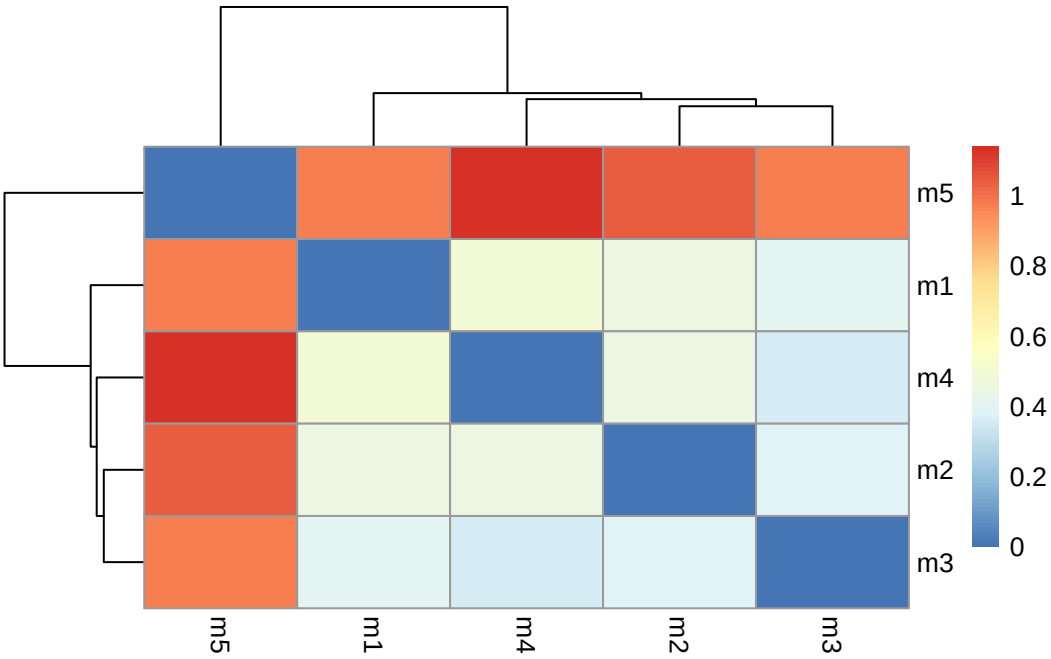
```
Warning in rmsd(pdbs, fit = T): No indices provided, using the 99 non NA positions
```

```
range(rd)
```

```
[1] 0.000 1.139
```

```
library(pheatmap)

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


m5
m1
m4
m2
m3
1
0.8
0.6
0.4
0.2
0
m5
m1
m4
m2
m3