

Introduction to Biocomputing Tutorial

Week 11

Debartolo 319

Review

- MUSCLE alignment

```
./muscle -in <a fasta file with sequences to be aligned> -out <a file name for the aligned sequences, in fasta format, to be saved>
```

- Profile HMM

```
./hmmbuild <a file name to store a profile HMM in> <the file containing a fasta formatted sequence alignment>
```

```
./hmmsearch <a file containing a profile HMM> <a fasta file containing sequences to be searched using the HMM>
```

If we want to get hmmsearch output in a tabular formatted file we add the `--tblout` option:

```
./hmmsearch --tblout <file name to store output> <a file containing a profile HMM> <a fasta file containing sequences to be searched using the HMM>
```

Review

- For loop in Unix

```
for file in *.fasta
do
    echo $file
    cat $file | grep ">" | wc -l
done
```

Review

- Regex in Python

```
infile = open("indivIDs.txt", "r")
outfile1 = open("out1.fasta ", "w")
outfile2 = open("out2.fasta", "w")
```

```
for line in infile:
    line = line.strip()

    if re.search("([ATGC]{8})(AATTC)([ATGC]+)", line):
        outfile.write("> " + line + "\n")

infile.close()
```

Announcements

- Office hour Monday 3-5pm, Galvin 266