

Assignment 4

December 15, 2020

```
[8]: def fasta_data(name):  
    from Bio import Entrez  
    Entrez.email = 'zw2946@nyu.edu'  
    handle = Entrez.esearch(db='protein', term=name, retmax=200) #search based on  
    ↪on disease name:COVID-19 and return 200 seqs  
    record = Entrez.read(handle)  
    id_list = record['IdList']  
    efetch = Entrez.efetch(db='protein', id = id_list, rettype = 'fasta',  
    ↪retmode = 'text') #fetch based on idlist  
    file = '/Users/zunqiuwang/Desktop/%s.fasta' % name #write to file  
    out = open(file, 'w')  
    out.write(efetch.read())
```

```
[9]: def MSA(fasta_data, name):  
    from Bio.Align.Applications import ClustalOmegaCommandline #using  
    ↪ClustalOmega to save command into a file  
    in_file = '/Users/zunqiuwang/Desktop/%s.fasta' % name  
    out_file = "/Users/zunqiuwang/Desktop/%s_aligned.aln" % name  
    clustalomega_cline = ClustalOmegaCommandline(infile=in_file,  
    ↪outfile=out_file, verbose=False, auto=False)  
    return(str(clustalomega_cline) + ' --resno' + ' --outfmt=clu')
```

```
[10]: def runningMSA(clustalomega_cline):  
    import os  
    os.system(clustalomega_cline)
```

```
[11]: runningMSA(MSA(fasta_data('COVID_19'),'COVID_19'))
```

```
[ ]:
```