

Python for Bioinformatics

Working with DNA Data



UNIVERSITY OF
LIVERPOOL



The University
Of Sheffield.

Map

DNA Sequence Data:

Biopython & SeqIO

The FASTQ Format





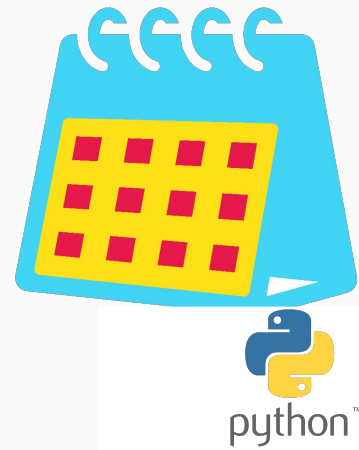
Calendar

Morning 1: Foundations of Python

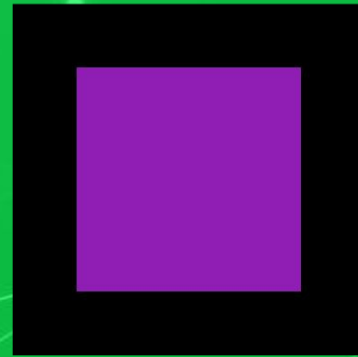
Afternoon 1: Data Structures, Conditions, & Loops

Morning 2: **Reading & Manipulating DNA Data**

Afternoon 2: Reading and Writing Files

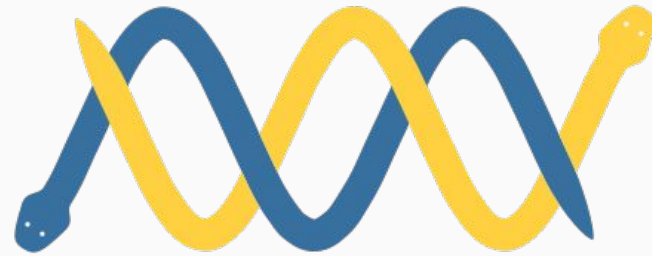


DNA Data Files



Biopython

Tools for working
with DNA data in
Python



biopython

<https://biopython.org/>

The FastQ Format



```
@MISEQ:42:000000000-A8DD3:1:1101:15046:1468  
TTCACATCCTGACCATTGAGTTGAGCAAATAGTTCTTCAGT  
+  
CCFFFFFFGHHHHJIIJJJIIJJJIIJJJIIIIJJGFIIIJEDD
```

SeqIO.parse

Use and options



```
seqIO.parse(sequence_file, format)
```

Format options: "fastq", "fasta"

SeqIO.parse

Use and options



```
seqIO.parse(sequence_file, format)
```

```
seqIO.parse(my_seq_data.fastq, "fastq")
```


SeqIO.parse

Read a sequence file
&

“Parse” the file into
constituent parts

print(id)

```
@MISEQ:42:000000000-A8DD3:1:1101:15046:1468
```

print(seq)

```
TTCACATCCTGACCATTGAGTTGAGCAAATAGTTCTTCAGT
```

Modify Reads

Components can
then be modified



```
print(record.seq) # full sequence
```

```
GGGGGGGGGGGATCATCTATATCA
```

Modify Reads

Components can
then be modified



```
print(record.seq) # full sequence
```

```
GGGGGGGGGGATCATCTATATCA
```

Modify Reads

Components can
then be modified



```
print(record.seq) # full sequence
```

```
GGGGGGGGGGGATCATCTATATCA
```

```
print(seq.id[9:]) # trim 10 bases
```

```
ATCATCTATATCA
```

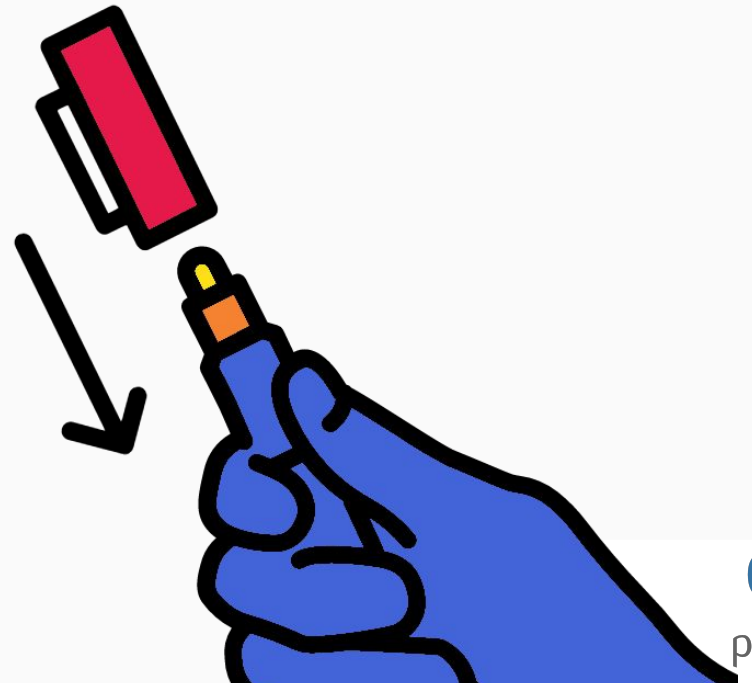
Recap

BioPython

FastQ Format

SeqIO

Modification



Questions?

