**Introduction to Python**

**Day Four Cheatsheet: Biopython**

**Sequence file input/output**

For file input/output, load the SeqIO or/and AlignIO modules from BioPython:

from Bio import SeqIO

from Bio import AlignIO

|  |  |  |
| --- | --- | --- |
| Function | What it does | Usage syntax |
| SeqIO.read() | Read a single sequence from a file | SeqIO.read("filename", "format") |
| SeqIO.parse() | Read mutliple sequences from a file | list(SeqIO.parse("filename", "format")) |
| AlignIO.read() | Read a single alignment from a file | AlignIO.read("filename", "format") |
| AlignIO.parse() | Read mutliple alignments from a file | list(AlignIO.parse("filename", "format")) |
| SeqIO.write() | Write sequences to a file | SeqIO.write(seq\_record(s), "filename", "format") |
| SeqIO.convert() | Convert file of sequence(s) to a different format | SeqIO.convert("infile", "informat", "outfile", "outformat") |
| AlignIO.convert() | Convert file of alignment(s) to a different format | Align.convert("infile", "informat", "outfile", "outformat") |

**BioPython Seq objects**

**To use and/or manipulate Seq objects, load these modules:**

from Bio.Seq import Seq

from Bio.Alphabet import \* # usually not needed, except to access/change .alphabet

**Defining a Seq object**

my\_seq = Seq("ACGT") # no alphabet, defaults to Alphabet()

my\_seq = Seq("ACGT", alphabet = IUPACAmbiguousDNA)

**Useful Seq object attributes**

my\_seq.seq # returns the sequence (you can re-cast to str)

my\_seq.alphabet # returns the alphabet

**Useful Seq object methods**

my\_seq.transcribe()

my\_seq.back\_transcribe()

my\_seq.translate()

my\_seq.complement()

my\_seq.reverse\_complement()

my\_seq.tomutable () # allows you to change the Seq.seq attribute

**BioPython SeqRecord objects**

**To use and/or manipulate Seq objects, load these modules:**

from Bio.Seq import Seq

from Bio.Alphabet import \* # usually not needed, except to access/change .alphabet

from Bio.SeqRecord import SeqRecord

**Defining a SeqRecord object**

my\_seq = Seq("ACGT") # SeqRecords are defined with Seq objects

my\_seqrecord = SeqRecord( my\_sec, id = "my\_id")

**Useful SeqRecord object attributes**

my\_seqrecord.seq

my\_seqrecord.id

my\_seqrecord.description # Useful for GenBank sequence records

my\_seqrecord.annotations # Useful for GenBank sequence records

my\_seqrecord.features # Useful for GenBank sequence records

**Example: Saving sequences to a file**

from Bio import SeqIO

from Bio.Seq import Seq

from Bio.SeqRecord import SeqRecord

# Sequences we want to save

sequences = {"id1": "ACGTACGT", "id2": "GCTAGCTA", "id3": "TACCGAT"}

# Empty list to store SeqRecord objects

records = []

# Populate records list

for seq\_id in sequences:

seq\_rec = SeqRecord( Seq(sequences[seq\_id]), id = seq\_id )

records.append( seq\_rec )

# Save records to file

SeqIO.write(records, "outputfile.fasta", "fasta")