

Introduction to Programming (Python) (IPP)

Using Biopython I

\$Date: 2004/11/18 16:08:45 \$

Biopython organisation

- Basically, Biopython is a collection of Python modules, useful for doing bioinformatics.
- Let's look at (a copy of) `http://www.biopython.org/docs/api/public/trees.html` for an overview of what's there,
- and how it's organised.

Packages and modules

This works:

```
>>> import Bio.Fasta.FastaAlign
```

because

1. The file `~/lib/python/Bio/Fasta/FastaAlign.py` exists;
2. and because the directory `~/lib/python` is in my Python module search path.

`Bio` and `Bio.Fasta` are *packages*.

`Bio.Fasta.FastaAlign` is a *module*.

IPP-11

More on packages

- Packages are basically just directories which must contain a file called `__init__.py`.
- `__init__.py` can be empty, but if it isn't and it defines Python objects (e.g. classes) then these can be imported just as if the package were a normal module.
- For example, `Bio/Fasta/__init__.py` includes a definition of a class called `Record`, so this works:

```
>>> from Bio.Fasta import Record
```

IPP-11

Getting FASTA sequences into your program

- Let's look at `get_first_fasta_seq.py`
- You (and your program) don't need to understand what a `Fasta.Iterator` object is,
- just that it has a `next()` method which returns the FASTA sequence (by default, as a simple string)

Using an iterator in a 'for' loop

- We can use an iterator object to construct a 'for' loop,
- to iterate over all sequences in the file.
- Let's look at `get_all_fasta_seqs.py`

Parsing en passant

- It would be nicer to have each record as a `SeqRecord` object, rather than a mere string.
- We can do this by supplying a `Fasta.SequenceParser` object when we construct the iterator.
- Let's look at `parse_fasta_seqs.py`, `parse_fasta_seqs2.py` and `parse_fasta_seqs3.py`

SeqRecord objects

- As Schuerer & Letondal say in the *Python course in Bioinformatics*: “SeqRecord is central in Biopython”.
- There’s a link to this book on the module web page.
- Chapter 11 has lots of useful diagrams, let’s have a look at them.
- So a SeqRecord object contains a Seq object.

How to use Biopython

The required skills are:

1. Having the patience to actually read the documentation (and often the source code).
2. Understanding essential OO concepts, and their implementation in Python.
3. General programming skills

Styles of using Biopython

- Cookbook approach: Follow the instructions in Chapter 3 of the Biopython tutorial to get some 'recipes'.
- Cookbook approach: Grab a script from somewhere, run it, tailor it to your needs.
- Bottom up: Build you program from scratch.

Biopython example scripts

- Looking at examples is a good way to learn.
- Here's some I grabbed from the Pasteur Institute.
- Merci bien!