Genome 559
Intro to Statistical and Computational Genomics

Lecture 20b:
Biopython
Larry Ruzzo

Biopython and Blast

Can run Blast

Either locally or over net

Save results

Parse and analyze results

A sample problem:

How good is Blast at finding tRNAs in Mj?

Exercise

```
from Bio.Blast import NCBIWWW
from Bio.Blast import NCBIXML
import os

if(not os.path.exists("trnablast.xml")):
    query = "GGGGCCGTGGGGTAGCCTGGATATCCTGTGCGC...CCA"
    eq = "Methanocaldococcus jannaschii[Organism]"
    res_handle = NCBIWWW.qblast(
            "blastn", "nr", query, entrez_query = eq)
    svfl = open("trnablast.xml", "w")
    svfl.write(res_handle.read())
    svfl.close()
    res_handle.close()

resultHandle = open("trnablast.xml", "r")
blastRecord = NCBIXML.read(resultHandle)
print blastRecord.alignments[0].hsps[0]

# Find data: score, Evalue, align len, start coord
```

How would I use Biopython?

Biopython is not a program itself; it's a collection of tools for Python bioinformatics programing

When doing bioinformatics, keep Biopython in mind

Browse the documentation;become familiar with its capabilities

Use help(), type(), dir() & other built-in features to explore

You might prefer it to writing your own code for:

- Defining and handling sequences and alignments

- Parsing database formats

- Interfacing with databases

You don't have to use it all! Pick out one or two elements to learn first

Code re-use

If someone has written solid code that does
what you need, use it

Don't "re-invent the wheel" unless you're doing
it as a learning project

Python excels as a "glue language" which can
stick together other peoples' programs,
functions, classes, etc.

Python – What next?

Read

    scour the python/biopython web sites

    look at other people's programs

    look at bits in the standard libraries (yes, some will be over your head, but it gets better...)

    use google

Write

    programming takes practice - keep it up.

    small project in your lab? *automated workflow?* display your data on a pretty web page? redo early HW using tools learned later?

    keep statistics for your soccer team?

Other tools?

these are more complex, but might pay off

Again, wikipedia is often a good starting place, to get a general idea of what it is/whether it might be useful to you

HTML

plistlib & XML

SQL and data bases

"User Interfaces" e.g., tkinter

Bioinformatics - What next?

work more of the Biopython tutorial
focus on stuff you might use in the lab, to start
next journal club, *don't* just skim the methods
even follow up with referenced papers
wikipedia or other books: "Python for Bioinformatics"
text books. Durbin Eddy Mitchison & Krogh especially
recommended for the probabilistic modeling aspect.
Ewans & Grant (we'll add a few refs to the web)

send me an email if you find something you like!