

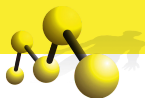
Review. Future

Luis Pedro Coelho

Programming for Scientists

September 24, 2012





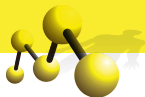
Course Content: Python

- Basic types: int, float, list, dict, set
- Control flow: for, while, if, elif, else ...
- Defining types: class, `__init__`, ...
- Errors (Exceptions): try, except, raise, ...
- Modules & Standard Library: import



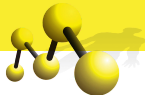
Memory & Numeric Representations

- It's bits all the way down
- Binary representation of signed & unsigned integers
- Floating point numbers
- When handling a lot of data, think of memory usage



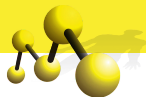
Parsing files

- Files are just Bytes (sequence of small numbers)
- It is all in how you interpret them
- There are standard character assignments for text files
- ASCII (English only), Latin-15 (used in Portugal), UTF-8 (usable for everything).



Open Source

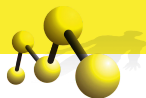
- Free as in beer, free as in speech (gratis/freedom distinction)
- Copyleft vs. liberal licenses
- It is not about price



Testing

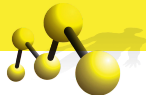
- Testing is good and you should do it

What Was Not in The Course

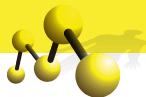


Missing

- Some more advanced programming details
- Version Control
- Unix & Shell & Interacting with Other Programmes
- More specific tools

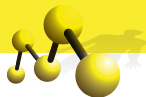


Next Generation Sequencing



- Third-party libraries are hit & miss
- HTSeq is good
- <http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>

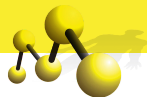
Examples of HTSeq Usage



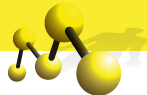
```
maxq = 0
for seq in HTSeq.FastqReader('hw-HeLa.fq.gz', 'solexa'):
    maxq = max(maxq, seq.qual.max())

print 'Maximum quality found was {0}'.format(maxq)
```

Continuing...

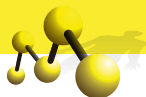


```
import HTSeq
allqs = np.array([seq.qual
                  for seq in
                    HTSeq.FastqReader('hw-HeLa.fq.gz', 'solexa')
```



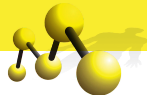
- Now, you want to align the sequences.
- Just use `bowtie2`

Calling External Programmes

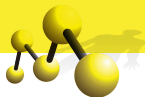


```
from os import system

system('bowtie2 -x hg19reference -U {0} -S {1}' \
      .format(ifilename, ofilename))
```

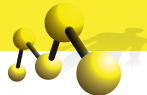


- Parses GFF/GTF files
- Basis of htseq-count



- How to Think Like a Computer Scientist <http://interactivepython.org/courselib/static/thinkcspy/index.html>
- Stack Overflow
for asking questions
- Python.org & numpy.scipy.org
for documentation

Any Other Questions...



Email me: luis@luispedro.org