

Computing for Medicine: Phase 3, Seminar 3 Project

Jennifer Campbell

Associate Professor, Teaching Stream

campbell@cs.toronto.edu



Reminder: Project Deadlines

- Original plan: project due 4 weeks after each session
- New plan:
 - No restriction on completing a project within a particular time period after each session.
 - Instead, there are three due dates by which you must submit any one of the projects:
 - December 20, 2016 6:00pm
 - January 17, 2016 6:00pm
 - February 28, 2017 6:00pm

Seminar 3 Project

- The project handout is posted:
 - <http://c4m.cdf.toronto.edu/cohort1/phase3/>
- Two approaches for doing your work:
 - Use the Computer Science Teaching Labs computing network.
 - Use your personal computer.
- Python3 packages to install:
 - Biopython
 - <http://biopython.org/wiki/Download>
 - `pip3 install biopython`

GLOB

Python's glob module

- <https://docs.python.org/3/library/glob.html>
- Used to find files whose names match a given pattern.
- Symbols used:
 - * (matches zero or more characters)
 - ? (matches exactly one character)
 - [] (matches one character contained within the brackets)

Demo

- Example directory contains the following files:
 - a.txt, apple.txt, b.jpg, banana.txt, carrot.txt, carrot.jpg

```
>>> glob.glob('* .txt')
['a.txt', 'apple.txt', 'banana.txt', 'carrot.txt']
>>> glob.glob('* .jpg')
['b.jpg', 'carrot.jpg']
>>> glob.glob('? .txt')
['a.txt']
>>> glob.glob('? .*')
['a.txt', 'b.jpg']
```

Demo (continued)

```
>>> glob.glob('a*')
```

```
['a.txt', 'apple.txt']
```

```
>>> glob.glob('*a*')
```

```
['a.txt', 'apple.txt', 'banana.txt',  
'carrot.jpg', 'carrot.txt']
```

```
>>> glob.glob('[ab].*')
```

```
['a.txt', 'b.jpg']
```

```
>>> glob.glob('[bc]*.txt')
```

```
['banana.txt', 'carrot.txt']
```

BIOPYTHON

Python's `biopython` module

- The starter code in `calculate_consensus.py` uses the `biopython` module.
- To complete this project, you should read the starter code and aim to understand what that `biopython` code is doing.
- You will need to model part of your solution to `find_mutations.py` on the starter code provided in `calculate_consensus.py`.

UPCOMING SEMINARS

Seminar 4: Dr. Chris McIntosh

- Tuesday November 29 2016 6-8pm
- Location: DCS Innovation Lab
- Topic: Computer Vision
- <https://scholar.google.ca/citations?user=PHQGLfoAAAAJ&hl=en>

FEEDBACK

Phase 3, Seminar 3 Survey

- You will receive an email with the subject “C4M: Phase 3, Seminar 3 Feedback Survey”.
- Please complete that survey now:
 - <https://www.surveymonkey.com/r/C4MSeminar3>