Testing, PyCogent scripting and work with Greengenes

Daniel McDonald – Knight Lab
7.19.11
Overview

• Discuss testing and why testing is important in software development
• Talk about the tax2tree project in collaboration with Greengenes
• Develop a command line script
Testing

- Software is complex and bugs can be sneaky
- Humans are error-prone, computers are stupid
- There are many types of testing
  - White box
  - Grey box
  - Black box
- There are many types of bugs
  - Arithmetic
  - Logic
  - Performance
- Publications based on untested code is dangerous
Can you spot the bugs?
Testing

• Software is complex and bugs can be sneaky
• Humans are error-prone, computers are obedient and stupid
• There are many types of testing
  – White box
  – Grey box
  – Black box
• There are many types of bugs
  – Arithmetic
  – Logic
  – Performance
• Publications based on untested code is dangerous
Testing Cont’d

• **Unit-testing**
  – A unit is a small well defined block of code
  – Tests verify the correctness of these blocks
  – Mocking is encouraged

• **Integration testing**
  – The whole is greater than the sum of the parts
  – Verify whole workflows

• **Regression testing**
  – Verify components as software evolves
  – Automated testing
• “Imperfect tests, run frequently, are much better than perfect tests that are never written at all.”
  
  – Martin Fowler (http://www.martinfowler.com/articles/continuousIntegration.html#N100F5)
PyCogent Scripting

• Use the PyCogent template script
• Add support for querying NCBI
• Write the sequences to a file
• Align the sequences and write to a file
• Reverse complement the alignment and write to a file
• Produce a consensus sequence for the alignment and print it to the screen
• Build a tree from the reverse complemented alignment and write it to a file