An in silico first semester freshman laboratory as an introduction to bioinformatics

Randy Bennett, Vince Buonaccorsi and Jill Keeney
Department of Biology, Juniata College, Huntingdon, PA 16652

Abstract

We have developed a laboratory module to introduce freshman biology majors to the basic tools of bioinformatics, including BLAST, protein structure viewing, and multiple sequence alignments.

The module focuses on the human superoxide dismutase (SOD1) gene, mutations in which can cause an inherited form of Amyotrophic Lateral Sclerosis (ALS).

The module is writing intensive. The student grade is based on the generation of figures and figure legends and the content of the abstract, results and discussion sections of a scientific manuscript.

A rough draft is graded and returned, allowing students to make corrections and improvements before turning in the final version.

It is hoped that an early experience in writing expectations will improve student writing in subsequent courses.

Introduction

Bioinformatics tools are rapidly becoming common place in many fields of biological study. It is important for students to learn of these tools early in their undergraduate career.

Bioinformatics tools are not only useful as research tools. As educators we should find these tools useful in both teaching the process of doing science and in teaching many concepts of biology. We have developed a 4 day lab sequence for our freshman biology laboratory. This lab module serves three purposes:

- An introduction to basic bioinformatics tools
- An introduction to science writing
- An introduction to basic concepts in gene and protein structure and regulation

The module is very intensive for the instructors. The module is writing intensive and requires extensive one-on-one attention to overcome tendency to be overly descriptive and non-analytical in their writing.

Day One

- Introduce network
- Focus on central dogma and protein primary structure
- Teach amino acid properties and open reading frame
- Hypothesize effects of mutation based on only primary structure information
- Introduce BLAST and CLUSTAL, do CLUSTAL alignment
- Introduce Manuscript template
- Have student obtain cDNA sequence file
- Using ORFind (NCBI tools), determine likely ORF
- Using BLAST determine ORF identity (Human Sod1)
- Identify wild type and two mutant sequence files that have protein structure data files (for later use, day 2)
- Use BLINK/COMMONTREE, find 5 non-human SOD1 sequences (one must be invertebrate)
- Using ClustalW, align sequences
- Observe conserved and variable regions. Where do identified human mutations map?

Students end day one by producing two figures, one showing cDNA sequence with translation and positions of two mutations marked, and the second showing the multi-sequence alignment produced by ClustalW.

Day Two

- Cover basics of X-ray crystallography.
- Introduce Cn3D and VAST tools.
- Learn fine scale structure and functional biochemistry of SOD1
- Have students obtain protein structure files for wild type and one mutant SOD1.
- Using tools in Cn3D, explore secondary and tertiary protein structure and discover relationships of amino acid properties and protein structure.
- A useful website for SOD1 is found at http://www.nottingham.ac.uk/biochemcourses/students/sod1/index.htm

Students produce a figure annotating the mutant SOD1 structure.

Day Three

- Have students align wild-type and mutant protein structures using VAST
- Using Cn3D, view the structural alignments; using knowledge gained in Day 2, explore and determine effect of mutation on protein structure.
- Build hypothesis as to why specific mutation effects SOD1 function and causes disease.
- Students construct a figure annotating important findings

Day Four

- Review manuscript template and requirements
- Template has been downloaded from the Journal of Virology, students are provided with "instructions for authors".
- Students put together figures, figure legends, write results and discussion, and abstract.
- Students must turn in rough draft of paper before leaving the lab. (for us this is usually a Thursday or Friday)
- Faculty member reviews and makes extensive comments on manuscript and returns by Monday.
- Final Drafts are due by Friday of that week.

Assessment

- 70% of student grade is based on manuscript
- 20% rough draft, 50% final draft
- To encourage students to push hard in rough draft, a students grade on the final draft is limited to 12 pts higher than rough draft grade.
- Remaining 30% is based on completing daily assignments and participation (attitude and attendance).
- We run approximate 150-180 students through this module in teams of two. Grading (and in particular review of rough draft) is very time intensive. Grading schedule is tight.
- Day 4 is an intense day, few students leave lab early. Students often need extensive one-on-one attention to overcome tendency to be overly descriptive and non-analytical in their writing.

- Student assessment of course:
  - This past year we moved lab module to Spring from Fall
  - Effect: more students felt prepared for material than in the previous year. (43% vs. 30%). Our 1st semester Biology course contains elements of molecular biology.
  - Most students (79%) did not feel the module helped their writing skills, yet most (69%) felt the module improved the abilities in scientific writing style. (approx. 65% felt they entered the course with limited knowledge of science writing).
  - Most students do not consider figures and figure legends to be part of science writing.
  - Most students felt the rewrite was a valuable experience.
  - The above results (except point one) were largely similar between students who had just completed course and those one year removed (and still in the Biology major).
  - For students one year removed, few appeared to be using the tools available at NCBI, including Pubmed, and instead seemed to revert to other databases for obtaining articles in biology.
  - Students one year removed reported that exposure to the module helped them with respect to topics covered in Cell Biology and Genetics (43% agreed to this statement, 30% were neutral, remaining disagreed)

Course Logistics

- Course runs for 4 lab periods. Each day has specific goals. At the end of each day, students prepare a figure and legend and a rough Results paragraph related to the day's work.
- Our campus is highly 'connected'; students are able to access and share network drives on which they can run programs such as CLUSTAL and Cn3D.
- To prevent loss of time due to network outages, sample 'results' from searches are available if needed. We avoid using them if at all possible because it eliminates the decision making process inherent in designing and interpreting search strategies and results.
- Project is initiated by students accessing a sequence file as if the DNA sequence was being sent from a DNA sequencing facility. The students are told that is a cDNA sequence for a gene that is linked to a human disease. It is their job to find out what the gene is.

Table 1. Basic Knowledge Developed in Lab

<table>
<thead>
<tr>
<th>Bioinformatics</th>
<th>Molecular Biology</th>
<th>Science Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLAST</td>
<td>Gene Structure</td>
<td>Abstract</td>
</tr>
<tr>
<td>CLUSTAL</td>
<td>RNA processing</td>
<td>Figures and Legends</td>
</tr>
<tr>
<td>Protein</td>
<td>cDNA vs.</td>
<td>Results vs. Discussion</td>
</tr>
<tr>
<td>Structural</td>
<td>Genomic DNA</td>
<td>Titles</td>
</tr>
<tr>
<td>databases</td>
<td>ORFs</td>
<td></td>
</tr>
<tr>
<td>Pubmed</td>
<td>Protein Structure</td>
<td></td>
</tr>
<tr>
<td>CN3D/VAST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This work was supported by NSF grant MCB-0722774 to J.B.K. and a grant from the William J. von Liebig Foundation to Juniata College