JOURNAL OF MICROBIOLOGY & BIOLOGY EDUCATION, December 2014, p. 292-294
DOI: http://dx.doi.org/10.1128/jmbe.v15i2.638

JMBE Journal of Microbiology & Biology Education

Supplemental Materials

for

Undergraduate Bioinformatics Workshops Provide Perceived Skills

Robin Herlands Cresiski
Department of Physical and Life Sciences, Nevada State College, Henderson, NV 89002

Table of Contents

(Total pages 11)

Corresponding author. Mailing address: Nevada State College, 1125 Nevada State Dr., Henderson NV 89002. Phone: 702-992-2651. Fax: 702-992-2501. E-mail: Robin.Cresiski@nsc.edu.

Appendix 1: Pre-workshop survey.

---

1. How would you define bioinformatics?

2. How would you rate yourself in terms of your computer proficiency and skill?
    a. Very strong
    b. Strong
    c. Comfortable with computers for basic use (word processing, Facebook)
    d. Timid, I prefer not to use computers
    e. Uncomfortable/weak, I fear computers

3. Which choice below reflects your attitude towards bioinformatics and its usefulness?
    a. Very positive
    b. Positive
    c. Neutral
    d. Negative
    e. Very Negative

4. How much would you say you currently know the *purpose* of bioinformatics tools
    a. I know a lot
    b. I know some
    c. I know a little
    d. I know almost nothing

5. How much would you say you currently know about *how to use* bioinformatics tools?
    a. I know a lot
    b. I know some
    c. I know a little
    d. I know almost nothing

6. I think bioinformatics tools will be:
    a. Easy to use
    b. Will be somewhat easy to use with some introduction (1-3 hours)
    c. Will be usable after several (4-6) hours of training
    d. Will be usable after many days of training
    e. Will be very difficult even with extensive training

7. What best describes your perception of how much you will use or rely on bioinformatics in your own future career?
    a. I will use bioinformatics actively in my career
    b. I will use bioinformatics on occasion in my career
    c. I will probably not use bioinformatics in my career
    d. It is very unlikely I will use bioinformatics after this class

Please write in your hopeful career here _____

Appendix 2: Workshop session 1.

Bioinformatics Workshop I Worksheet
Using the Basic Local Alignment Search Tool (BLAST)

by Richard Tillett,
Nevada Center for Bioinformatics, University of Nevada, Reno,
Guest Speaker for Nevada State College Bioinformatics Workshop I

Outline of Guest Lecture:

1. What is BLAST?
    a. BLAST (Basic Local Alignment Search Tool) is an online tool used to identify similarities and differences between a query nucleotide/protein sequence with sequences stored in a database. For a given query sequence, BLAST reports the aligned regions of similarity.
2. Why use BLAST? What questions does BLAST answer?

- "What does this gene do?"
    - You obtained DNA sequence from an experiment and need to know more about it
    - You want to compare a novel sequence to known sequences to hypothesize gene function
    - You want to compare a partial sequence to whole genome sequences
- "An interesting gene was just reported - is a homologous gene present in my favorite species?"

3. Specialized BLASTs
    a. Make gene-specific primers for PCR (Primer-BLAST)
    b. Screen a sequence for vectors (VecScreen)
    c. Compare two sequences (bl2seq)

Hands-on Activity: Using BLAST to answer "What does this gene do?"

1. Obtain our example sequence
    a) Go to http://www.ncbi.nlm.nih.gov/
    b) In the "all databases" text box type "VvCBF4" and click search. This is a gene from Vitis vinifera, the plant that produces grapes for wine!
    c) On the Cquery page that results click "1" under Nucleotide Sequences to get to the 1 record of this sequence
    d) Click the word FASTA in the upper left
    e) Select and copy the nucleotide sequence in addition to theheader line ">". If something went wrong, go here http://www.ncbi.nlm.nih.gov/nuccore/DQ497624.1?report=fasta&log$=seqview&format=text and select all, then copy

2. Performing BLAST at NCBI
    a) Go to http://www.ncbi.nlm.nih.gov/
    b) On the right, under Popular Resources, click BLAST to go to the BLAST home page http://blast.ncbi.nlm.nih.gov/
    c) In the center of the page, under Basic BLAST, choose the program "nucleotide blast." If you have any difficulty, go here: http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome
    d) Paste in the sequence of the VvCBF4 gene you have copied and click BLAST button (don't change any of these settings for now)
    e) Results page group discussion: What is interesting about these results? What species are these sequences from? How much similarity is there between sequences? What part of this gene seems most conserved?

3. Modifying BLAST parameters
    a) Let's change *the type of nucleotide blast* used
        i) At the top left, click "edit and resubmit"
        ii) In Program selection, select "Somewhat similar sequences (blastn)" and click BLAST button
        iii) Results: more dissimilar species!

b) Try these other options to see how your results change (restrict “Organism” like Zea Mays, search only the NCBI genomes “Database”, or choose the “blastp” tab to compare nucleotide sequences to proteins).

Appendix 3: Workshop session 2.

Name: _____ Date: _____

Bioinformatics Workshop II Worksheet

Today, you will be given a nucleotide sequence found in real human DNA that is associated with a genetic disease when mutated. Your task is to compare the sequence you are given with known genes in the National Center for Biological Information (NCBI) website, using their Basic Logical Alignment Search Tool (BLAST) program[1]. Where text is bolded, you are required to enter some information or an answer.

Directions:

Obtain a sequence from your instructor. **Do you notice anything interesting about this sequence? Are there any patterns or repeats? Please describe:**

1) Go to the NCBI website, found at http://www.ncbi.nlm.nih.gov/, and click on the word "BLAST," located on the list of Popular Resources on the top right of the page.
2) Under the section "Basic Blast", click on the link for "nucleotide blast."
3) Enter the unidentified nucleotide sequence your instructor provided into the large white box at the top of the screen.
4) Make sure that under "Choose Search Set" that "Others" is selected. In "Program selection" you want to select Highly Similar Sequences for now. Now click on the "BLAST!" button.
5) On the webpage of your BLAST results, you should see a Graphic Summary section with the horizontal colored lines. Carefully position your mouse on the start of the first colored line. You should see a phrase appear in the white box (above "color key for alignment scores") to indicate an accession number of a sequence, a genus species name, and the name of the gene associated with the nucleotide sequence.
6) Slowly continue moving the mouse down each of the next 5-10 successive colored lines, taking note what appears in the white box. **What seems to be the name of your gene?** *Make sure you look in the first several sequence titles to see if there is any disease association in the titles associated with your gene.

If this is a disease you have never heard of, search for it in a Genetics Home Reference (http://ghr.nlm.nih.gov/) to learn briefly about it. We will discuss these in class after the workshop.

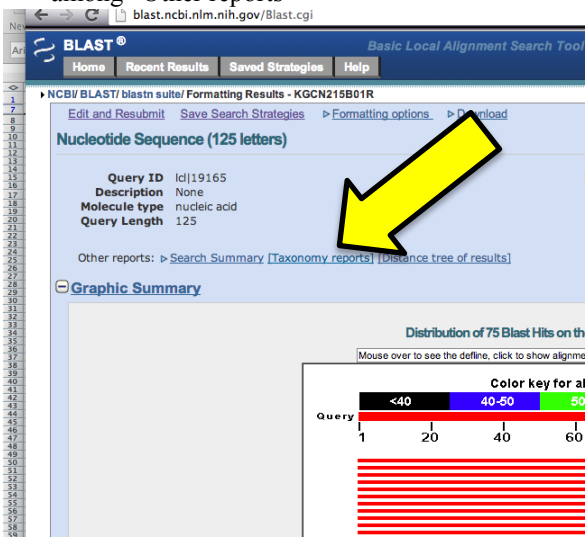
7) On your BLAST results page, scroll down below the graphic of colored lines to the "Descriptions" portion. Since we used Megablast, which looks for Highly Similar Sequences, many of these may be *Homo sapiens* (human) genes. **Do you see any genes that are not *Homo sapiens*? Circle Yes or No.**
   If Yes – **list 3-4 other *Genus species* names that seem to match your sequence, and list their Max Ident %** (this is the percent similarity between the query and subject sequences over the length of the coverage area). Once you've listed them, **Google the genus species name to find out the common name of the organism.**

Genus species	Max Ident	common name
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

[1] These activities are adapted from web-published exercises by the East Bay Bioinformatics Education Project (www.ebbep.com/docs/**bioinformatics**/intro**bioinformatics**.pdf).

What do you notice about the organisms besides *Homo sapiens* that share high sequence similarity to this gene?

8) Scroll to the top of the page. Above the "Graphic Summary", click on "Taxonomy Reports" which is listed among "Other reports"



Here in the lineage report, you can see a variety of organisms that matched your genetic sequence closely! Besides the ones listed above in step 7**, what additional organisms do you see? (common name use is sufficient here):**

What does this tell you about the evolutionary heritage of this gene?

Why might it be advantageous to know if a gene related to human disease is present in other species?

9) On your BLAST results page, look in the Description section. To the far right of each sequence is an accession number. **Write down 2 accession numbers for 2 sequences from <u>distinct organisms</u>** (Note: do not use accession number that has an XM or NM or any other one that has initials followed by a space – only choose accession numbers where there is no space between the capital letters and the number).:

_____ _____

10) Click on the DNA shaped symbol in the upper left-hand corner to return to the NCBI home. Click on "BLAST" again from the "Popular Resources." Scroll down to the "Specialized BLAST" section and click on "Align" two (or more) sequences using BLAST.

11) Add the accession numbers you wrote down in the two query boxes. Be sure to select "Somewhat similar sequences" in the "Program Selection." Click "BLAST."

12) You should arrive at a page similar to the one seem previously but only with one colored line in the Graphic Summary. Scroll down to the "Alignments" section.

13) Remember that you are being shown one sequence (the "Query" Sequence aligned on top of another sequence, the "Subject" sequence.). Where the sequences are identical, a vertical line runs from the Query nucleotide to the Subject nucleotide (this helps build the Identities score). Where sequences are not identical, there is no vertical line. Where one sequence has several nucleotides that are missing compared to the other sequence,

horizontal hashmarks ("------") will appear, indicating a deletion or insertion mutation and this helps build the Gaps score.

Record your Identities _____%, and Gaps _____%

Do you see any places that are insertions or deletions? Circle on: Yes or No
Briefly describe (ex: "a 10 nucleotide deletion seems to be present in the howler monkey compared to the human")

Here, you are comparing sequence data from two unique species, but we could similarly have run a query comparing two human sequences from different individuals. **Why might it be useful to identify the exact nucleotides that are distinct from one species to the next? From one human individual to the next?**

14) Finally, go back to the NCBI homepage by clicking on the DNA symbol at top left or going to http://www.ncbi.nlm.nih.gov/
15) Choose "SNP" from the "Popular Resources" link at the top right, which will bring you to http://www.ncbi.nlm.nih.gov/snp/. Note: Once on this site, clicking on "Overview of dbSNP" provides good background on SNPs if a review is needed.
16) To the right, there is a list of links under the title "Access Data." Choose "Web Search." This takes you to the SNP Advanced Search Builder page. Under builder, enter the name of your ***disease*** (not the gene name) in the top query box (you can leave it on 'All Fields."). Click "Search."

17) The number of entries found on this page indicates how many SNPs are affiliated with this disease. **Enter your number of results here _____.**
The results page will show you sequences where a SNP is present. The black nucleotides are the part of the sequence flanking the SNP, and the SNP will be highlighted in red. A red "[A/G]" symbolizes that some individuals will have an A nucleotide at that location, while others have a G.
What is the significance of this information?

18) If you were a researcher interested in treating this disease, how would you possibly use this information about existing SNPs to direct your studies? Propose what research you would do next based on SNP information. (Note: this may be assigned as a homework question to do outside of the actual workshop due to time restraints).

Appendix 4: Unidentified nucleotide sequences and instructor key.

These sequences were obtained and used with permission from East Bay Biotechnology Education Project, http://www.ebbep.com/docs/bioinformatics/introbioinformatics.pdf.

Nucleotide Sequences

Sequence A:

ATGGCGACCCTGGAAAAGCTGATGAAGGCCTTCGAGTCCCTCAAGT

CCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC

AGCAGCAGCAGCAGAGCAGCAGCAACAGCCGCC

Sequence B:

ATGGCGGGTCTGACGGCGGCGGCCCCGCGGCCCGGAGTCCTCCTG

CTCCTGCTGTCCATCCTCCACCCCTCTCGGCCTGGAGGCGTCCCTG

GGGCC ATTCCTGGTGGAGTTCCTGGAGGAGTCTT

Sequence C:

ATGCTCACATTC ATGGCCTCTGACAGCGAGGAAGAAGTGTGTGATG

AGCGGACGTCCCTAATGTCGGCCGAGAGCCCCACGCCGCGCTCCTG

CCAGGAGGGCAGGCA GGGCCCAGAGGATGGAG

Sequence D:

ATGTTTTATACAGGTGTAGCCTGTAAGAG ATG AAGCCTGGTATTTA

TAG AAATTG ACTTATTTTATTCTCATATACAGTCATAATTTTCC

ATATGCC AGAAAAGTTGAATAGTATCAG ATTCCAAATCT

Sequence E:

ATGCGTCGAGGGCGTCTGCTGGAGATCGCCCTGGG ATTTACCGTGCT

TTTAGCGTCCTACACG AGCCATGGGGCGG ACGCC A ATTTGG AGGC

TGGGAACGTGA AGG AAACCAG AGCCAGTCGGGCC

Sequence F:

ATGCCGCCCA AAACCCCCCGAAAA ACGGCCGCCACCGCCGCCGCTGC

CGCCGCGGAACCCCCGGCACCGCCGCCGCCGCCCCCTCCTG AGGAG

G ACCC AGAGCAGG ACAGCGGCCCGGAGGAC

Sequence G:

ATGTTGTGACA TATCC ATCTACTGTAGTTAAGATATTCAGTAC

TTTGTTTTTCATAAGCATGTA ATTGATCATATTTCTGCCAAGGATGT

GCCTTCAACTTTATA ATTATAGTGTTGTAA A A T A T T T G

Sequence H:

ATGCCATCTTCCTTG ATGTTGGAGGTACCTGCTCTGGCAG ATTTCA

ACCGGGCTTGG AC AG AACTTACCG ACTGGCTTTCTCTGCTTGATC

A AGTTATAA A ATCACAG AGGGTG ATGGTGGGTG ACCTT

Teacher Sequence Key:

Sequence A: Huntington's Disease

Sequence B: William's Syndrome

Sequence C: Alzheimer's

Sequence D: Cystic Fibrosis

Sequence E: Marfan Syndrome

Sequence F: Retinoblastoma

Sequence G: Human Menkes Disease

Sequence H: Muscular Dystrophy

Appendix 5: Post-workshop survey.

---

1. How would you define bioinformatics?

2. How would you rate yourself in terms of your computer proficiency and skill?
    a. Very strong
    b. Strong
    c. Comfortable with computers for basic use (word processing, Facebook)
    d. Timid, I prefer not to use computers
    e. Uncomfortable/weak, I fear computers

3. Which choice below reflects your attitude towards bioinformatics and its usefulness?
    a. Very positive
    b. Positive
    c. Neutral
    d. Negative
    e. Very Negative

4. How much would you say you currently know the *purpose* of bioinformatics tools
    a. I know a lot
    b. I know some
    c. I know a little
    d. I know almost nothing

5. How much would you say you currently know about *how to use* bioinformatics tools?
    a. I know a lot
    b. I know some
    c. I know a little
    d. I know almost nothing

6. I think bioinformatics tools used in the workshop were:
    a. Easy to use
    b. Will be somewhat easy to use with some introduction (1-3 hours)
    c. Will be usable after several (4-6) hours of training
    d. Will be usable after many days of training
    e. Will be very difficult even with extensive training

7. What best describes your perception of how much you will use or rely on bioinformatics in your own future career?
    a. I will use bioinformatics actively in my career
    b. I will use bioinformatics on occasion in my career
    c. I will probably not use bioinformatics in my career
    d. It is very unlikely I will use bioinformatics after this class

Please write in your hopeful career here _____

Appendix 6: Grading rubric for bioinformatics definition.

	Point Value
Mention of computers, computation, information systems, or software	1
Mention of genes, genomes, proteins, sequences, or "biological" information	1
Mention of data/information storage	1
Mention of data analysis generally or specifically (sequence comparisons, SNP identification, homologous sequence identification, etc.).	1