Tested Studies for Laboratory Teaching
Proceedings of the Association for Biology Laboratory Education
Volume 38, Article 4, 2017

Collaboration between Educators and Deep Ocean Scientists to Create Research-Based Labs: Focus on Bioinformatics

Angela Gee and Marissa Pantastico-Caldas

Los Angeles Trade Tech College, Science Department, 400 W. Washington Blvd, Los Angeles CA 90015 USA
(**geeal@lattc.edu**)

Many undergraduate students in introductory biochemistry courses find it challenging to understand how different levels of protein structure relate to each other. To address this problem, we introduced an inquiry-based laboratory exercise in which students are challenged to explain how the effects of mutations on different levels of protein structure lead to changes in protein function and ultimately to genetically-inheritable diseases. The implementation of this exercise in a large, second-year undergraduate, introductory biochemistry course led to a high level of student satisfaction and a more integrated view of biochemistry and genetics.

Keywords: bioinformatics, metabolism, research-based lab, translation, proteins

Introduction

Students in community college are often not exposed to scientific research. This exercise is part of a collaborative nationwide project to expose students to deep ocean research. In this research-based lab, students use bioinformatics and databases to analyze similarities and differences in amino acid sequences for cytochrome C across species. Two species of bacteria and four species of eukaryotes (humans, chimpanzees, bottle-nose dolphins and honey bees) are examined to understand their ancestry and cytochrome C function. Bioinformatics is a valuable, well-used tool by scientists so it is critical that students are exposed to this technique.

Target Audience and Timeline

This laboratory exercise is designed for General Biology (cell and molecular biology) for Majors. It is appropriate for entry level biology majors, and has been modified for an upper level Human Physiology course. It has been piloted for students at Community College. The exercise is designed for one three hour session, although it is completed in three sections so it can be modified for shorter length sessions ranging from one to two hours.

Collaboration with Deep Ocean Scientists

A unique collaborative team involved with the Center for Dark Energy Biosphere Investigations (C-DEBI) funded by the National Science Foundation created a resource of teaching tools for community college instructors, one of which is this laboratory exercise. C-DEBI is a group of institutions doing groundbreaking research on the deep ocean biosphere. Scientists, instructors, educators and web designers nationwide have collaboratively developed and implemented ways to bring this research to curricula in the sciences with a specific focus on community colleges that do not typically have access to research. Toolkits were created which are stand-alone labs and activities that can be integrated into relevant curriculum. All materials are open source and posted at http://www.coexploration.org/C-DEBI

For this particular exercise, research scientists provided background and sequence data from a project investigating metabolic processes in deep ocean bacteria. In addition, scientists shared the bioinformatics techniques used to analyze the data, which are implemented in this exercise. This collaboration provided students with an understanding of diversity of life, scientific research and how it is relevant to them.

© 2017 by Angela Gee and Marissa Pantastico-Caldas

Objectives

Students completing this exercise will be able to:

- Understand and apply bioinformatics with the programs BLAST and Seaview.
- Compare and contrast amino acid sequences of cytochrome C in humans, other animals and bacteria.
- Relate amino acid sequences of cytochrome C from various species to genome evolution and protein function.

Preparation and Materials

Students should have a general knowledge of the process of translation, proteins and metabolism. In addition, basic computer literacy is needed. No prior set-up is required by the instructor. Each group of two students should have access to one computer with internet connection along with the exercise documents provided. The two bioinformatics programs used are open source. BLAST (Basic Local Alignment Search Tool; **http://blast.ncbi.nlm.nih.gov/Blast.cgi**) is an online database and Clustal Omega (**http://www.ebi.ac.uk/Tools/msa/clustalo/**) is an online sequence alignment program. All materials including supplements are available online at **http://www.coexploration.org/C-DEBI/toolkits_biology.html.**

Student Outline
Amino Acid Sequence Analysis of Cytochrome C in Bacteria and Eukarya Using Bioinformatics

Introduction

All life forms undergo metabolic processes to obtain energy. The way organisms metabolize substances, however, vary depending upon their ancestry and the environmental conditions they live in. For example, many eukaryotes, including humans, acquire energy from sugars mainly through aerobic cellular respiration. Bacteria, on the other hand, can metabolize through various mechanisms such as nitrogen fixation, photosynthesis, aerobic and anaerobic respiration. They can use different sources for energy such as single carbons, nitrogen and sulfur.

A molecule found common to bacteria and eukaryotes is the protein cytochrome C. It aids in metabolism by helping transfer electrons from one electron carrier to another. Cytochrome C consists of about 100 amino acids. In different species, the amino acid sequences for cytochrome C are similar but not identical.

In this lab, we will use bioinformatics to analyze similarities and differences in amino acid sequences for cytochrome C in two species of bacteria and four species of eukaryotes (humans, chimpanzees, bottle-nose dolphins and honey bees) to understand their ancestry and cytochrome C function.

Objectives

1. Understand and apply bioinformatics with the programs BLAST and Clustal Omega.
2. Compare and contrast amino acid sequences of cytochrome C in humans, other animals and bacteria.
3. Relate amino acid sequences of cytochrome C from various species to genome evolution and protein function.

Hypotheses

1. Organisms that are more closely related will share more common amino acid sequences in the cytochrome C protein.
2. Organisms with similar types of metabolism will share more common amino acid sequences in the cytochrome C protein.

Materials and Methods

Materials (See Bioinformatics Reference Sheet)

- Computer with internet connection
- BLAST (Basic Local Alignment Search Tool) website: http://blast.ncbi.nlm.nih.gov/Blast.cgi
- Clustal Omega website: http://www.ebi.ac.uk/Tools/msa/clustalo/
- Amino acid sequences for subjects
- Description of subjects

Methods

Exercise 1: Use BLAST to Identify Accuracy of Given Amino Acid Sequences

Fill in Table 1. Video instructions: http://youtu.be/yv7Tbo3XyOw

1. Use the BLAST website: http://blast.ncbi.nlm.nih.gov/Blast.cgi
2. Under “Basic BLAST”, choose the “protein blast” program to run because we are using amino acid sequences.

3. Enter your amino acid sequence for search.
   In the "enter query sequence" box on the upper left corner, copy and paste the amino acid sequence for the organism you want to test. For example, for the *Unknown*, enter:
   QTAYAGDAANGMRVFNTYCSDCHSITEGKNKIGPSLWNVVGRKPASISDFNYSDAMRKNDIIWTEDRISTYITN
   PQGLLPGVKMAFPGLKDPQKCADVIQFLSQQH
   Scroll down to the bottom of the page and click "BLAST". Wait for your results to show.
   Note: Results show the species and protein with the best match at the top of list.
4. Fill in information in Table 1 using the first match shown under the "description" tab.
   a. Fill in the species identified located in [brackets].
   b. The protein is listed before the species.
   c. For "fraction of AA sequences matched" and "identity value".
      i. click on the first match.
      ii. look at the first set of data on the top of page. Find the 4th column labeled "identities". Here is the fraction of AA matched (e.g. 105/105) and percentage for the identity value (e.g. 100%).
5. Repeat for the remaining organisms. Check for species match, protein, fraction of AA match and identity match.

Exercise 2: Use BLAST to Compare Amino Acid Sequences for Cytochrome C Between Species

Fill in Table 2. Video instructions: http://youtu.be/xY_FOkSfhu0

1. Use the BLAST website: http://blast.ncbi.nlm.nih.gov/Blast.cgi
2. In your protein BLAST search page under "enter query search", click the box titled "Align two or more sequences." A second search box will appear underneath.
3. Enter the sequence of *Homo sapiens* in the second box titled "Subject". Enter all the sequences of the organisms that you want to compare it to in the first "Query" box. Be sure to hit return after each sequence when entering multiple organisms.
4. Click "BLAST".
5. Scroll down and find the "identities" match percentage between *Homo sapiens* and each other organism. Enter in the appropriate box in Table 2. Note: The two sequences are aligned next to each other. The differences in amino acids are identified with a blank or a "+" sign.
6. Repeat for other organisms as necessary.

Exercise 3: Use Clustal Omega to Visually Align Cytochrome C Amino Acid Sequences for All 5 Organisms

Answer questions based on alignment. Video instructions: https://youtu.be/H07g6X-RzRo

Answer questions based on alignment.

1. Use Clustal Omega website: http://www.ebi.ac.uk/Tools/msa/clustalo/
2. Enter the 5 sequences to be aligned.
   a. Go to Appendix 1. Amino Acid Sequences for Species
   b. Highlight ALL the text below the heading including the species name for the 5 species.
   c. Copy (CTRL+C) and paste (CTRL+V) into the "Step 1: Enter your input sequences". Make sure the drop down menu is on "protein" since we are aligning amino acids for proteins.
   d. For "Step 2: Set your parameters" select on the drop down menu "Clustal w/numbers". This gives the number of amino acids.
3. Align the sequences
   a. For "Step 3: Submit your job" click the submit button and wait.
   b. When the sequences appear aligned, click the "show colors" box.
   c. Note: the "*" denotes amino acids that are identical in all the species at that position. Same color represents amino acids that are similar.
4. Answer corresponding questions in the results section.

Results and Guide Questions

Exercise 1: Use BLAST to Identify Accuracy of Given Amino Acid Sequences

Table 1. Protein Identification for Different Species

	Species	Protein	Fraction of AA sequences matched	Identity match (%)
Bacteria	1.			
	2. *Sphingomonas*			
	3. *Caldithrix abyssi*			
Eukarya	4. *Homo sapiens*			
	5. *Pan troglodytes*			
	6. Apis mellifera			

Based on Species #1

1. Was the protein or organism identified first? What does this imply about genome evolution?
2. Do your given sequences correctly identify the protein and organism? Explain using the data in table.

Exercise 2: Use BLAST to Compare Amino Acid Sequences for Cytochrome C between Species

Fill in the table below with the "identity values" for each pair of organisms.

Table 2. Amino Acid Identity Table for Cytochrome C.

Cytochrome C	*Homo Sapiens*	*Pan troglodytes*	*Apis mellifera*	*Sphingo-monas*	*Caldithrix abyssi*
Homo Sapiens	-				
Pan troglodytes (Clint)		-			
Apis mellifera			-		
Sphingomonas				-	
Caldithrix abyssi					-

1. What is the identity value that compares the two bacterial species?

2. What are the identity values for three pairwise comparisons among the 4 eukaryotic species?

   _____ _____ _____

   average of the identity values = _____

3. Contrast the identity matches between bacteria and the average for eukaryotes. _____ vs _____

4. What do these results imply about the evolution (diversification) of bacterial and eukaryotic genomes?

5. List in order the species that have cytochrome C that is the most similar to least similar to humans. What does this result imply about our evolutionary relationships with bacteria and other eukaryotes?

6. What amino acid(s) are different between *Homo sapiens* and *Pan troglodytes*? Do you think this difference has any effect on cytochrome C function? Briefly justify your answer.

Exercise 3: Use Clustal Omega to Visually Align Cytochrome C Amino Acid Sequences for All Five Organisms

Comparison of Cytochrome C sequence in five species using Clustal Omega. Use your alignment of cytochrome C sequences in Clustal Omega to answer the following questions.

1. Which domain of life looks more diverse, bacteria or eukarya? Briefly explain your answer by pointing out patterns from the data.

2. Which species has the sequence that is the MOST different from the other species? Suggest a reason (based on protein function) why this species' cytochrome C sequence might differ the most.

3. a. List the name of amino acids that were identical in position in the cytochrome C sequence across all species investigated. The part of the genome that codes for these amino acids are referred to as "conserved regions" of the gene.

   b. What might be the function of these conserved regions for the cytochrome C protein?

Discussion and Conclusion

Evaluate the two hypotheses based on patterns derived from the data.

Bioinformatics Reference Sheet

1. Amino Acid Sequences for Species

>Unknown
QTAYAGDAANGMRVFNTYCSDCHSITEGKNKIGPSLWNVVGRKPASISDFNYSDAMRKNDIIWTEDRISTYITNPQG
LLPGVKMAFPGLKDPQKCADVIQFLSQQH
>Sphingomonas_echinoides
LAAYTGDAKKGETDFITCKTCHAIEAGVNKIGPSLHGVVGRKAGSIPGFTYSTANKNSGITWTEEKLFQYLENPQRV
VPGTKMTFAGWPTDPQKRADVIAYLKSNS
>Caldithrix_abyssi
ENSSFGEGMSLAELGAKLYKSKACVTCHSVDGSPLVGPTFKGVFGHTVKLNDGSSVKADENYLRESILKPQAKVVE
GFQPVMPTYQSILKPREVDALIEYIKSLGE
>Homo_sapiens
MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKK
YIPGTKMIFVGIKKKEERADLIAYLKKATNE
>Pan_troglodytes
MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQAAGYSYTVANKNKGITWGEDTLMEYLENPKK
HIPETKMIFVGIKKKEERADLIAFLKKATNE
>Apis_mellifera
MGIPAGDPEKGKKIFVQKCAQCHTIESGGKHKVGPNLYGVYGRKTGQAPGYSYTDANKGKGITWNKETLFEYLEN
PKKYIPGTKMVFAGLKKPQERADLIAYIEQASK

2. Amino Acid Abbreviation Codes

Amino Acid	3-Letter Code	1-Letter Code
Alanine	Ala	A
Cysteine	Cys	C
Aspartic acid or aspartate	Asp	D
Glutamic acid or glutamate	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y

3. Description of Species

Bacteria: (ocean dwellers)

1. ***Unknown*:** Bacteria from sediments that live off of single carbon compounds such as methanol and methylamine
2. ***Sphingomonas***: Aerobic bacteria that is really versatile. It can "eat" almost anything!
3. ***Caldithrix abyssi***: Anaerobic isolated from a hydrothermal vent where temperature can rise up to 400ºC. Name literally means "cauldron (hot) threads of the abyss".

Eukarya:

1. ***Homo sapiens:*** Healthy human individual.
2. ***Pan troglodytes***: Chimpanzee.
3. ***Apis mellifera:*** European honey bee.

4. Description of Programs and Databases

BLAST (Basic Local Alignment Search Tool; http://blast.ncbi.nlm.nih.gov/Blast.cgi) is a program that compares DNA and amino acid sequences to a known database of sequences of organisms. It calculates the significance of matches. BLAST can be uses to test for genetic diseases. BLAST can also be used to infer functional and evolutionary relationships between sequences. BLAST is maintained by the National Center for Biotechnology Information (NCBI) which is a subdivision of the National Institutes of Health (NIH). Research groups worldwide contribute to the database.

Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) is a program that allows visualization of the alignment of multiple sequences. It can also organize phylogenic trees based on evolutionary relationships. It is maintained by the institution European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) in the Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom. Different research groups contribute to its database.

Materials

Each pair of students requires one computer or tablet with internet connection and laboratory handouts.

Notes for the Instructor

The following materials are posted on http://www.coexploration.org/C-DEBI/under the link http://www.coexploration.org/C-DEBI/toolkits_biology.html

- PPT (for instructor): Introduction, results and conclusion of the activity
- Student Activity: handout for student
- Student Activity with answers (for instructor)
- Appendix: handout for student needed to complete activity
- Videos (embedded in PPT):
    - C-DEBI introduction: https://www.youtube.com/watch?v=wiYzGL4iTY8&feature=youtu.be
    - Exercise 1 demonstration: http://youtu.be/yv7Tbo3XyOw
    - Exercise 2 demonstration: http://youtu.be/xY_FOkSfhu0
    - Exercise 3 demonstration (Seaview): http://youtu.be/-XfF8sJaZ-c

Note: Clustal Omega is an online sequence alignment program that is used as an alternative to the Seaview program in exercise 3. Clustal Omega is more practical to use as it requires no previous set up and downloading of programs. The College of Exploration website above uses the Seaview program.

Activity Outline

1. Instructor introduces the activity (20 min)
    a. Use PPT as a guideline to introduce metabolism and bioinformatics.
    b. Video introduction of C-DEBI.
2. Students complete activity (60 – 120 min)
    a. Divide students into groups of two per computer.
    b. Have students follow the appropriate YouTube video as they complete each of the three exercises. It is best if they watch a part of the video, pause it, complete that task, then continue to the next part of the video.
    c. Students complete the three exercises in the activity. Note: To reduce time, you may group two pairs of students together for a larger group of four. Both pairs of students complete Exercise 1. Then one pair completes Exercise 2. The other pair completes Exercise 3. All four students then explain their exercise to each other. All four students complete conclusion together.
3. Instructor reviews activity with the students (20 min)
    a. Use PPT to review results and discussion.

Expansion of this Lab

Some instructors may be interested in expanding this laboratory to address species other than mammals and bacteria. Here are cytochrome C sequences for species from plants and fungi:

>Barley Wheat *(Hordeum vulgare)*: plant
TEFKAGSAKKGATLFKTRCELCHTVEKGGPHKVGP
NLHGIFGRHSGQAQGSYTDANIKKNVLWDENNMS
EYLTNPKKYIPGTKMAFGGLKKEKDRNDLITYLKK
ACE

>Baker's yeast *(Saccharomyces cerevisiae)*: fungi
TEFKAGSAKKGATLFKTRCLQCHTVEKGGPHKVGP
NLHGIFGRHSGQAEGYSYTDANIKKNVLWDENNM
SEYLTNPKKYIPGTKMAFGGLKKEKDRNDLITYLK
KACE

The exercises can be expanded to inquire about the evolutionary relationships between animals, fungi, plants and bacteria. Students can compare these amino acid sequences for cytochrome C. For example, students can be asked, based on cytochrome C sequences, are fungi more closely related to humans or bacteria? Yeast have an identity value of 64% to humans, but only 45% and 30% to the respective bacteria. Similarly, are plants more closely related to humans or bacteria? Barley wheat has an identity value of 62% to humans but only 45% and 30% to the respective bacteria. Yeast and barley wheat, however, have a 96% identity value.

In Clustal Omega, after you get the results, you can click on the "phylogenetic tree" and a phylogeny will appear. This feature is a possible expansion of the lab if you want to gear it more towards evolutionary relationships.

To make this lab more inquiry based for advanced classes, I did not provide the hypotheses for the students. Instead, after I gave them the introduction and previewed the three exercises, I made the students develop one hypothesis before starting the lab.

Acknowledgments

We would like to thank C-DEBI (Center for Dark Energy Biosphere Investigations) and NSF for funding and support, Stephanie Schroeder, and the entire collaborative team involved in the C-DEBI Toolkit

Project. The team includes John Kirkpatrick who provided the research data, Lynn Whitley, Peter Tuddenham, Tina Bishop, researchers and community college instructors.

About the Authors

Angela Gee earned B.A.s in both Molecular and Cell Biology and Psychology at U.C. Berkeley and her Ph.D. in Neurobiology and Behavior at Columbia University. Angela was a Postdoctoral Fellow at Columbia where she developed and taught Frontiers of Science. Her research investigated neural mechanisms that underlie how primates pay attention to the visual world. After, Angela was an Assistant Professor of Biology at LaGuardia Community College in Queens, New York. She then worked at UCLA to promote diversity and outreach in science research. Since 2012, Angela has been an Instructor at Los Angeles Trade Tech College where she teaches courses and develops curriculum in biology. Angela's goal is to help a diverse group of students appreciate science and succeed at community college and beyond.

Marissa Pantastico-Caldas earned B.S. in Biology at the University of the Philippines in Los Banos and her Ph.D. in Ecology and Evolutionary Biology at the University of Arizona. She was also a Postdoctoral Fellow at the Australian National University where she conducted field experiments to test the "Storage Effect Theory" as a Mechanism for Species Coexistence. Marissa has been an Instructor at Los Angeles Trade Tech College since 1996 where she is now a full professor teaching courses and helping develop curriculum in biology.

Mission, Review Process & Disclaimer

The Association for Biology Laboratory Education (ABLE) was founded in 1979 to promote information exchange among university and college educators actively concerned with teaching biology in a laboratory setting. The focus of ABLE is to improve the undergraduate biology laboratory experience by promoting the development and dissemination of interesting, innovative, and reliable laboratory exercises. For more information about ABLE, please visit http://www.ableweb.org/

Papers published in *Tested Studies for Laboratory Teaching: Peer-Reviewed Proceedings of the Conference of the Association for Biology Laboratory Education* are evaluated and selected by a committee prior to presentation at the conference, peer-reviewed by participants at the conference, and edited by members of the ABLE Editorial Board.

Citing This Article

Gee A., Pantastico-Caldas M. 2017. Collaboration between Educators and Deep Ocean Scientists to Create Research-Based Labs: Focus on Bioinformatics. Article 4 In: McMahon K, editor. Tested studies for laboratory teaching. Volume 38. Proceedings of the 38th Conference of the Association for Biology Laboratory Education (ABLE). **http://www.ableweb.org/volumes/vol-38/?art=4**

Compilation © 2017 by the Association for Biology Laboratory Education, ISBN 1-890444-17-0. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the copyright owner. ABLE strongly encourages individuals to use the exercises in this proceedings volume in their teaching program. If this exercise is used solely at one's own institution with no intent for profit, it is excluded from the preceding copyright restriction, unless otherwise noted on the copyright notice of the individual chapter in this volume. Proper credit to this publication must be included in your laboratory outline for each use; a sample citation is given above.