Rebuiding a Minimal Synthetic Cell

(from William and Mary iGEM; contact: igem@wm.edu or mssaha@wm.edu)

Motivation

The idea behind this activity is that the minimal synthetic cell built by the J. Craig Venter institute is a fascinating example of synthetic biology. The concept of a minimal cell is very interesting because it prompts thoughts about what genes are and are not essential for life. However, asking all students to build a synthetic cell from the ground up is quite difficult, especially if the students have limited molecular biology knowledge. Rather, using a top down approach allows a unique opportunity for a student to become intimately familiar with one specific gene, deepening their understanding of genes required for life to function, and demonstrating the power of bioinformatics tools to identify the function of a protein from a DNA or amino acid sequence.

Summary

NCBI BLAST is an incredibly powerful and easily accessed bioinformatics tool. Using NCBI BLAST within the classroom presents an opportunity to expose biology students to the potential of bioinformatics, while also deepening their understanding of molecular biology. However, if students were simply assigned an amino acid sequence to research, interest in the activity would be reduced, as there would be no ultimate goal present in the assignment. However, synthetic biology presents a unique opportunity to combine bioinformatics with a fascinating end goal : that of a minimal synthetic cell. In their 2016 paper *Design and synthesis of a minimal bacterial genome*, Hutchison et. al., detail how they were able to iteratively whittle down the genome of Mycoplasma mycoides (used to create their first iteration of a minimal cell, *JCV-syn1.0*). Through a comprehensive spreadsheet provided by the authors every protein present within *JCV-syn1.0* has: its amino acid sequence listed, a description of how essential it is, and whether or not it was kept or deleted from the genome of the final minimal synthetic cell *JCV-syn3.0*. By running a BLAST on an assigned amino acid sequence, students will gain (potentially their first) experience using bioinformatics tools to identify the function of a given protein. This will increase their awareness of how powerful large databases are for current, and future biologists. Furthermore, students will gain a deeper understanding of the power and use of synthetic biology, an increased understanding of one the use and function of a particular protein, and an increased breadth of molecular biology knowledge by seeing which classes of proteins are and are not essential to life. Finally, students will gain a valuable collaborative experience with their peers. While not an overly difficult tool to use navigating and utilizing BLAST will provide students a challenge that they are able to overcome. Since collaboration will be encouraged, students can consult each other on using BLASt, and discuss why or why not they think their protein is or isn't essential.

Learning Goals

Learning goals of this activity are as follows:

- A deeper understanding of one, or a few particular proteins.
  - By investigating whether or not a protein is essential, a student will have to become familiar with the function of a given protein. Furthermore, they will have to apply their knowledge of the central dogma and requirements for life. With the synthesis of this knowledge, and the investigation of this protein, the students will gain valuable experience having to think critically about biology.
- A broader understanding of the proteins essential for life
  - By seeing the spreadsheet collaboratively filled by themselves, and through the guided discussion from the instructor, the students will better understand what exactly a microorganism needs to be alive. While the specific proteins within each class won't be memorized, students will be able to recognize that for example, proteins associated with protein translation are generally essential or quasi essential, while many non-glucose catabolic proteins are not essential.
- Greater familiarity with using, and knowledge of the power of bioinformatics tools
  - Using NCBI BLAST to identify an unknown amino acid sequence for the first time is a very rewarding experience. By going from a completely unknown string of letters to identify the function of a protein, students will be able to recognize the power of bioinformatics. This serves many purposes, from deepening a students understanding of molecular biology, to informing students to an incredibly powerful resource, this activity will increase student's comfort with using BLAST as a bioinformatics tool.

Procedure

Using the spreadsheet from the paper *Design and synthesis of a minimal bacterial genome*, Hutchison et. al., 2016., pick a gene or a few genes for each student that has been categorized as having a known function (not putative function). Since the spreadsheet only offers amino acid sequences, students will be provided the AA sequences in lieu of the DNA sequences for ease of use. AA sequences will be provided to students through any convenient means (i.e. Amino acid sequences can be assigned to students in a spreadsheet, or each student can get a random sequence through some online random generator). After being assigned an amino acid sequence, the students use NCBI protein BLAST to identify what protein amino acid sequence they were given codes for. Using any available resources online (but preferably academic journals) and through collaboration with their peers, the students will identify what the protein they were assigned does within the bacterial cell. Most likely, TAs and instructors will need to walk around the lecture hall and help answer questions that the students may have, and help guide them to find sources for what their protein does. After obtaining enough information to identify what their protein does, the students will decide whether they think their protein is essential (required for life) (e), quasi essential (slows growth significantly if removed) (ie), quasi nonessential (slows growth somewhat if removed) (in), or non essential (doesn't slow growth down if the bacteria are given appropriate growth conditions) (n). After classifying their proteins, students will submit their amino acid sequence, protein essentiality classification, and reasoning behind why they believe this into a google form. This google form will then populate a google sheet. The students will then be granted read-only access to the spreadsheet, and can view how close their guess was to the actual result. A final question can ask the student how their guess compares to the result that they found. They can analyze why they were or weren't correct. Given sufficient time, the instructor can look over the results, and see what types of proteins students correct were able to assign essential status to, and what types of proteins the students had trouble with. The instructor can then focus on discussing what the role of these proteins are. For large lectures, the instructor can ask students to talk to their neighbors. If the results were fairly consistent for all classes of protein, then the discussion can revolve around analyzing why wide varieties of protein classes are or are not essential.

This activity can be applied to a broad range of classrooms, with students of varying age and molecular biology knowledge. For younger students who do not yet have a deep understanding of molecular biology, students can be assigned to groups so that several students can collaborate on deciding whether or not a particular protein is essential. The instructor(s) can then visit each group and offer guidance to the groups. For older and more experienced students, they can be tasked with quickly identifying whether or not a particular protein is essential, and then be asked to try to design a synthetic cell from the ground up. Depending on the level of experience and time, the requirements for the assignment can be more lenient, where the students may list broad categories of required genes. Likewise, a more rigorous activity could ask students to classify required functions of genes within each of the required categories (e.g. a necessary category may be DNA synthesis and repair, and specific functions can be a DNA polymerase, a proofreading protein, a DNA ligase, etc.).