Incorporating Genomics in Undergraduate Biology: an Avenue for Inquiry-based Research

Alice Saliba-Tarun, Ph.D. Physical and Life Sciences

Alfred State SUNY College of Technology

Abstract

The science of Genomics and Bioinformatics have revolutionized the field of biological research in the 21st century. The availability of bioinformatics online resources and genomic databases make it possible for teachers to incorporate coursework and activities that allow students to conduct research problems in genomics using bioinformatics tools. I recently developed a project based-learning course on genomics that combined lectures on the science and ethical, legal and social implications (ELSI) of genomics with hands-on lab activities on bioinformatics and DNA analysis. Using bioinformatics tools available at the Microbial Genome Annotation Network (http://www.geni-act.org/), the students gained foundational bioinformatics skills and experience first had the concepts of gene and genome structure by annotating genes from the bacterium *Glaciecola psychrophila*, a novel psychrophilic bacteria isolated from an arctic glacier. The students carried out a DNA Barcoding project to catalog and identify native trees growing on campus. Lastly, the students isolated their genomic DNA, submitted it for sequencing and analyzed the sequence to determine their maternal genetic ancestry. Future plans for the course are to incorporate analysis of the personal genome of students who are interested in having their genome analyzes and to provide a follow-up independent research course for students to do functional genomics

1. Gene Annotation using Geni-Act

Students learned to use bioinformatics tools to annotate genes from the bacterium *Glaciecola psychrophila* strain 170is, a gram negative, aerobic, psychrophilic, pigmented, motile bacterium that thrives in seawater isolated from Antartica (Zhang et al., 2006). Online bioinformatics tools in geni-act (geni-act.com) uses a modular approach to allow students to perform the DNA and protein analyses shown below. Each student was assigned to manually annotate two genes, including a gene annotated as hypothetical.

Modules of the GENI-ACT (http://www.geni-act.org/) were used to complete *Glaciecola psychrophila* 170 genome annotation. The modules are described below:

Modules	Activities	Questions Investigated
Module 1- Basic Information Module	DNA Coordinates and Sequence, Protein Sequence	What is the sequence of my gene and protein? Where is it located in the genome?
Module 2- Sequence-Based Similarity Data	Blast, CDD, T-Coffee, WebLogo	Is my sequence similar to other sequences in Genbank?
Module 3- Cellular Localization Data	Gram Stain, TMHMM, SignalP, PSORT, Phobius	Is my protein in the cytoplasm, secreted or embedded in the membrane?
Module 4- Alternative Open Reading Frame	IMG Sequence Viewer For Alternate ORF Search	Has the amino acid sequence of my protein been called correctly by the computer?
Module 5- Structure-Based Evidence	TIGRfam, Pfam, PDB	Are there functional domains in my protein?
Module 6- Enzymatic Function	KEGG, MetaCyc, E.C. Number,	In what process does my protein take part?
Module 7- Gene Duplication/ Gene Degradation	Paralog, Pseudogene	Are there other forms of my gene in the bacterium? Is my gene functional?
Module 8- Evidence for Horizontal Gene Transfer	Phylogenetic Tree,	Has my gene co-evolved with other genes in the genome?
Module 9- RNA	RFAM	Does my gene encode a functional RNA?

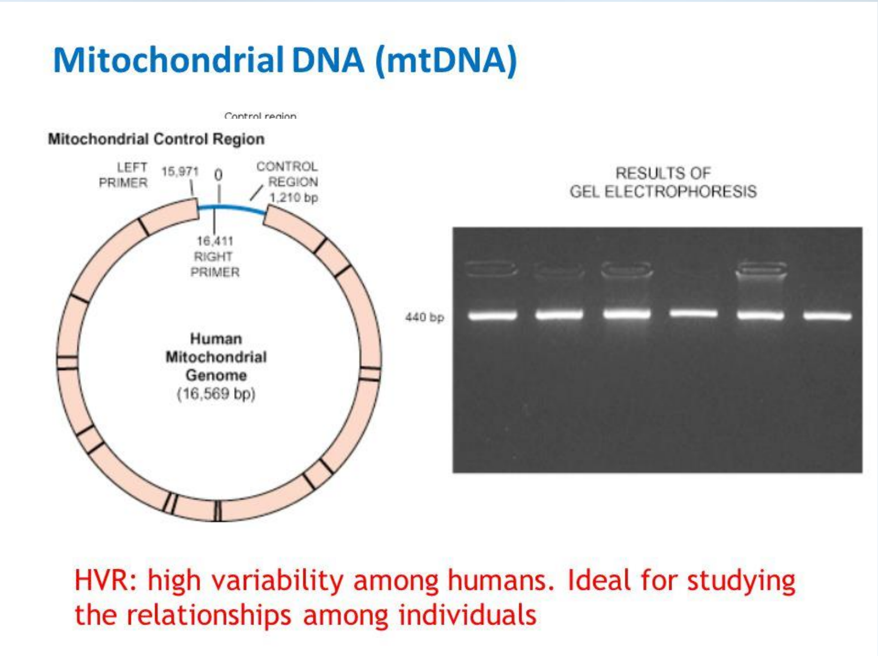
Course Implementation, Logistics

- Students met for a two hour lecture and a three hour lab each week in Fall 2015
- Lecture topics and lab activities were selected to teach students concepts and to be able to use and apply online tools in genomics and bioinformatics.

Week	Lecture Topics	Lab
1	Overview of Genomics and the course	Overview of projects; DNA Barcoding; Mitochondrial DNA sequencing; Gene annotation
2	Introduction to Genomes and their evolution	Bioinformatics and gene annotation:
3	Mapping, sequencing, annotation and databases	BLAST; Sequence alignment
4	Comparative Genomics	ORF prediction; protein structure prediction
5	Genome Evolution and Phylogenetics	Phylogenetics; multiple sequence alignment
6	Prokaryotic Genomes and metagenomics	DNA Barcoding Sample Prep
7	Eukaryotic Genomes	Bioinformatic tools for the laboratory
8	Human Genomics	Mitochondrial DNA analysis
9	Ethical, legal and social issues in personal genomics	ELSI; Case studies on Personalized Genomics
10	Ancestry and gene association studies	Metabolic Pathways; Gene Ontology
11	ENCODE Cancer Genomics	SNP, GWAS and HapMap
12	Transcriptomics	Mitochondrial DNA sequence analysis
13	Proteomics	Finalize gene annotation and class project
14	Systems Biology and metabolomics	Poster Presentation

Assessment: 2 exams, worksheets on bioinformatics exercises; geni-act notebook and poster presentation of gene annotation data; Website created presenting the class project on DNA barcoding

2. Mitochondrial DNA & Ancestry



http://www.geneticorigins.org/mito/mitoframeset.htm

Mitochondrial DNA (mtDNA) is passed down almost unchanged from a mother to her children. Thus sequencing of mtDNA, particularly in the hypervariable control region, provides genetic information revealing maternal genetic ancestry. Students isolated their genomic DNA from buccal mouthwash, used PCR to amplify the D-loop region and submitted the DNA for sequencing. The DNA sequences were analyzed and compared to a human mitochondrial database.

Acknowledgements

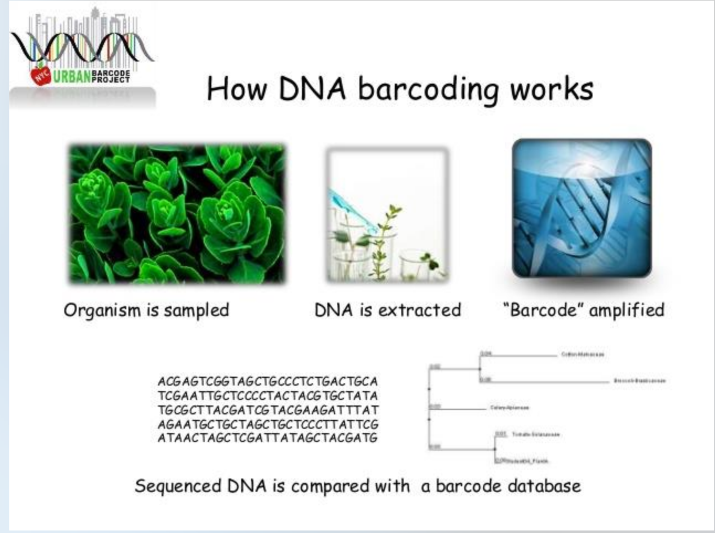
BIOL 5003 students
Dr. Steve Jacobi
Ms. Amy Warner

Dr. Stephen Koury
University of Buffalo

Dr. Lori Scott, Augustana College

microbial genome annotation network mgan

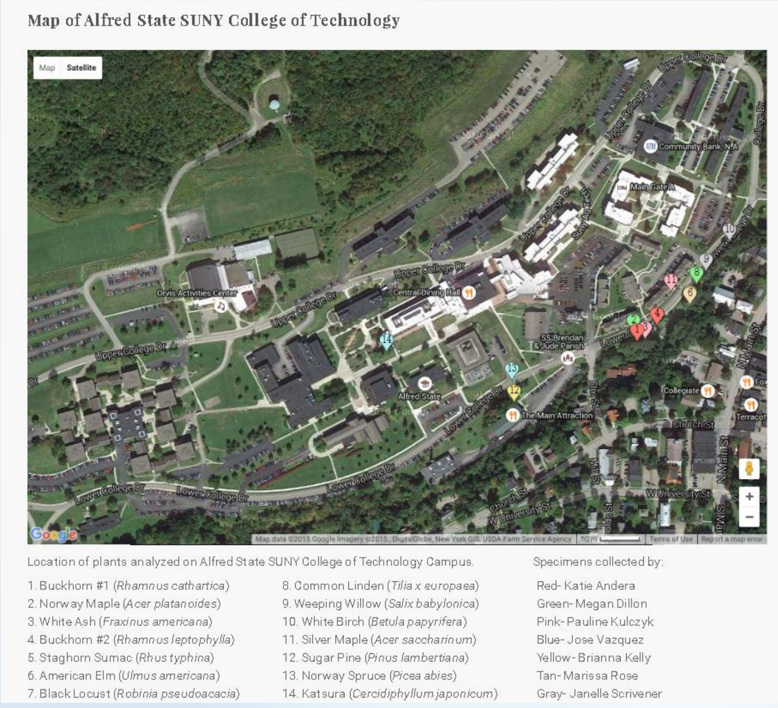
3. DNA Barcoding



https://www.dnalc.org/websites/dnabarcoding101.html

DNA Barcoding works by comparing short sequences of the genome that are highly variable among different species. To identify plant species, a region of the *rbcL* gene found in the chloroplast is amplified and sequenced.

The class chose to work on a project to use DNA barcoding to identify the species of native trees growing around the Alfred State campus. Leaf samples from 14 different trees on campus were collected. Genomic DNA from these leaves were isolated, amplified and sequenced.The sequences were matched to the Genbank database using BLAST.



Map of Alfred State College and the location of trees sampled

Results

1. 14 genes of *G. psychrophila* were manually annotated and compared to the automated gene annotation. Online geni-act notebooks for each gene were submitted to the geni-act website. Students presented posters on their results at the Alfred State Student Showcase.
2. The students created a website to present their DNA barcode project (http://alfredtrees.weebly.com/). The website was recognized for an award in civic engagement at the Alfred State Student Showcase.
3. Students were very enthusiastic in learning about their maternal genetic ancestry from mitochondrial DNA sequencing.

Conclusions

Using individual and group bioinformatics projects, students experienced first-hand the process of scientific inquiry. Genome annotation reinforced basic biological concepts such as gene structure, protein structure and function, gene homology and sequence similarity. Mitochondrial DNA sequencing enabled students to learn genomic tools used in determining gene ancestry and evolution. A future goal is for students to use what they have learned to conduct their own independent research and to include the analysis of the student's own DNA (personalized genomics)

References

Scott, L and Ghrist, A. 2012. An Authentic Undergraduate Research Experience that Combines Genome Annotation and Functional Genomics. Tested Studies for Laboratory Teaching. Proceedings of the Association for Biology Laboratory Education Vol. 33, 322-325, 2012

Zhang, D.-C., Yu, Y., Chen, B., Wang, H.-X., Liu, H.-C., Dong, X.-Z., & Zhou, P.-J. (2006). *Glaciecola psychrophila sp. nov.*, a novel psychrophilic bacterium isolated from the Arctic. International Journal of Systematic and Evolutionary Microbiology, 56(Pt 12), 2867–2869.