10-9-2019

Molecular Phylogeny Implemented in an Introductory Plant Classification Course

Chao Cai
Purdue University, caic@purdue.edu

Jo Ann Banks
Purdue University

Molecular Phylogeny Implemented in an Introductory Plant Classification Course

Chao Cai[1], Jody Banks[2]

[1]Purdue University Libraries and School of Information Studies; [2]Department of Botany and Plant Pathology, Purdue University

Abstract

Plant classification is one of the core components in undergraduate programs related to plant sciences. Traditionally plant classification courses primarily introduce morphology-based taxonomy because of practical needs in the field. However, the publication of new plant classification systems by Angiosperm Phylogeny Group (APG) using molecular phylogeny methods leads to the trends of using molecular evidence (DNA barcode) for plant identification. In our introductory plant classification course, we included a two-week module (lectures and labs) to introduce key concepts and fundamental skills in molecular phylogeny. Week 1 included concepts of evolutionary tree thinking, data mining in NCBI using BLAST search, and phylogenetic tree building. Week 2 introduced concepts of DNA sequencing and barcoding for plant identification. Student selected their own plants to sequence the DNA barcodes, which were then used in the final exam for practice and summative assessments. One challenge we are constantly dealing with is the increasing difficulty in finding diverse sequence using BLAST because of the fast-growing number of angiosperm genomes sequenced.

Course Settings

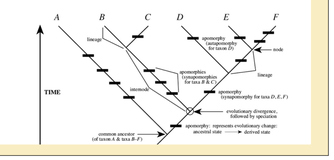
- Purdue University is the land-grant university of Indiana, and has an undergraduate enrollment > 32,000 students.
- The Department of Botany and Plant Pathology is housed in the College of Agriculture.
- *BTNY305: Fundamentals of Plant Classification* is one of the required **core curricula** for the "**Plant Science**" major with the Department of Botany and Plant Pathology, and is also required for majors with **forestry** and **horticulture** departments.
- The course is capped at 22 students per semester due to the lab and field trip capacity.
- The student pool consists of sophomore, junior and senior students.
- The course is offered only in fall semester to adapt the flowering season of most Indiana floral.

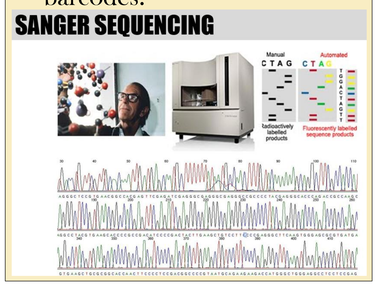
Learning Objectives

- Students will be able to find orthologs of a given gene using BLAST search, perform a multiple sequence alignment using MAFFT and build a maximum likelihood phylogenetic tree using the IQ-TREE web server.
- Students will be able to identify plants using DNA barcodes and assess the evolutionary relationships of plant species using phylogenetic trees built with DNA barcodes.

Design of the Module

	Lecture	Lab
Week 1	**Introduction to phylogenetics** • Phylogeny infers evolution; • How to interpret a phylogenetic tree; • Using DNA or protein sequence to build trees. Divergence: A split from one lineage into two, which may lead to speciation Autapomorphy: apomorphy for a single lineage Synapomorphy: apomorphy for two or more lineages	**BLAST search and tree building** • Search genes by similarity ; • Multiple sequence alignment using MAFFT; • Build phylogenetic trees; • Identify orthologs and species using tress.
Week 2	**DNA sequencing and barcodes** • Sanger sequencing; • Next-generation sequencing; • DNA barcoding and common barcodes; • Phylogeny using DNA barcodes.	**Genomic DNA extraction and PCR lab** • Students chose species ; • Genomic DNA extraction; • Amplify the rbcL (Rubisco) using PCR; • Sequence PCR products.
Final Exam	**Sample Final Exam Questions, Fall 2016** 1. A) Identify the species and family from which each of the 17 DNA sequences were derived (the sequences are provided on the last pages of this exam). These are the result of your lab experiment. Replace whatever comes after the > with the species and family name. B) Using these sequences, construct a Neighbor Joining Tree with bootstrap values. When using MEGA, ignore or say no anything referring to coding sequence. C) Is your tree consistent with what we know about angiosperm phylogeny (compare to your poster)? 2. *Flowering Locus T* (*FT*) is a gene that regulates flowering time in angiosperms. You want to test the hypothesis that this gene is present in angiosperms but not other non-seed/non-flowering plants. In other words, is this gene an apomorphy of the angiosperms? Construct a tree (using amino acid sequences) that would test your hypothesis. Based on this tree, was your hypothesis correct?	**Sample Final Exam Questions, Fall 2017** **1.** FLOWERING LOCUS T, or FT, is a very interesting protein that regulates flowering time in Arabidopsis. Use FT and Arabidopsis to find it's sequence. **A)** Paste the amino acid sequence of the Arabidopsis FT protein (it is 175 amino acids) below. **B)** You want to know: is this gene present only in flowering plants or is it present in all land plants, including those that never evolved flowers? You will answer this by making a phylogenetic tree of FT proteins. ******Go to the blastp page and select what you think are the appropriate land plant taxa in the Database Organism part of the web page. (Hint: avoid crop plants if you can because you will get many hits of the same gene.) **C)** What does this tree tell you about FT proteins (and their evolution) in land plants? **4.** Below are the results of your PCR sequencing. **1)** What gene did you sequence (use blastx of the first gene to find out)? **2)** Using the same sequence (AD1) do a blastx and blastn. From what organisms are the top hits from the two blast searches? **3)** Why didn't blastx and blastn give the same result? **4)** What species and family is associated with each sequence (there are 24)? Change the label to include this information. For example, the first line of the sequence >AD1 should be changed to >Genus species Family (whatever the species and family is). **5)** Using IQ-TREE, make a Maximum Likelihood tree and include bootstrap values. Download, save and print your tree. **6)** Does this tree jive with what we know about the evolutionary relationships of angiosperms (the poster I gave you at the beginning of the class will be useful in answering this question)?





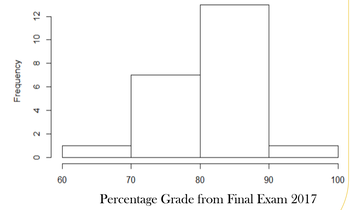
Blast searches and making phylogenetic trees 1
December 2017

Exercise 1
1) Go to http://www.ncbi.nlm.nih.gov/
2) In search field enter: Dynein heavy chain 1 AND homo and click search
3) Click on: Protein
4) Find and click on: **dynein heavy chain 1**, axonemal [**Homo** sapiens] (it's the 5th on the list)
How large is this protein? (Click FASTA to get only the amino acid sequence; click NM 015512.4 then FASTA to get the nucleotide sequence).
5) On the upper right hand side of the screen click under Analyze this sequence click: Run BLAST; BLAST is a program that will identify all other proteins in the NCBI database that are similar to your query (ie, dynein). Note that this is a blastp search. Other blast options (blastn, blastx, tblastn or tblastx) will depend on your query gene/protein and the database you wish to search. Blastp uses a protein as a query and searches for similar proteins in the protein database. In this example, your query sequence is NP_056327. You could also paste the amino acid sequence as a FASTA file into this box. What is FASTA?
6) Under organism, type then click Viridiplantae and on the next line type then click Homo sapiens. By doing this we are only searching for plant and human genes (not insect, worm, other mammals, etc.). We are limiting the number of taxa because it would take too long to include sequences from all organisms.

PCR protocol
_____ 1) You will do PCR on each of your 3 samples. Label the lid of your 3 PCR tubes.
_____ 2) Prepare a 10-fold dilution of each sample if necessary (see spreadsheet). In a new, labeled 1.5mL tube, add 90µL of water and 10µL of your DNA sample and mix.
_____ 3) You will prepare a 20µL PCR reaction for each sample. Add the following to each tube:
_____ 12.0 µL PCR MasterMix (contains the Taq polymerase, buffer, all 4 dNTPs & primers)
_____ 6.0 µL sterile water
_____ 2.0 µL 10X diluted plant DNA sample
Total volume = 20.0 µL
The thermocycler has been setup with the following cycle program:
95˚C - 2 minutes
(95˚C - 30 seconds, 52˚C - 30 seconds, 72˚C - 45 seconds) - repeated 19 times
72˚C - 3 minutes
4˚C - hold indefinitely
The rbcLa-F forward primer is 5' ATGTCACCACAAACAGAGACTAAAGC 3'
The rbcLa-R reverse primer is 5' GTAAAATCAAGTCCACCRCG 3'
R=A or G. This means that the reverse primer consists of two different primers; one with an A in the 18th position, and one with a G in the 18th position.

Assessment

- Low-stakes formative assessment
  - Lab participation and completion
  - Ensure student engagement
- High-stakes summative assessment
  - Final exam
  - Students worked on their own data



What We Learned from the Implementation

- Autonomy boosts student intrinsic motivation
  - Having student choose their own plant species to work on enhanced engagement in the lab;
  - Students have better grades for the final exam question using their own data.
- Consider introducing phylogeny **earlier** in the course
  - The idea of "common ancestor" should be emphasized more frequently;
  - Provide students more opportunities to interpret phylogenetic trees.

Challenge

- Too many genome sequences from model species and crop plants to obtain diversity from a simple BLAST search;
- We ask students to exclude Brassicaceae and Poaceae from their search.

Computational Tools

- BLAST:
  https://blast.ncbi.nlm.nih.gov/Blast.cgi
  Reference: https://doi.org/10.1016/S0022-2836(05)80360-2
- MAFFT server:
  https://mafft.cbrc.jp/alignment/server/
  Reference: https://doi.org/10.1093/nar/gkf436
- IQ-TREE server:
  http://www.iqtree.org/
  https://www.hiv.lanl.gov/content/sequence/IQTREE/iqtree.html
  Reference: https://doi.org/10.1093/nar/gkw256

You can get access to the poster using this QR code.
For more information, please contact Chao Cai: caic@purdue.edu