Testing the Highland Hybrid Hypothesis – Phylogenetic and phylogenomic assessment of putative hybrid frogs

Josiah Harold Townsend[1],* and Kayla D. Weinfurther[1]

Affiliations:

[1]Department of Biology, Indiana University of Pennsylvania, Indiana, Pennsylvania 15705, USA.

*Correspondence to: JHT, josiah.townsend@iup.edu

Type of Manuscript: *CourseSource* Lesson Manuscript

Title and Description of Primary Image: Subadult female (UF 166643) and (UF 166637) of putative hybrid frogs from Reserva Biológica Cerro Uyuca, 1,640 m elevation, Departamento de Francisco Morazán, Honduras. Photo by Jason M. Butler.

Scientific Teaching Context

Learning Goal(s)

- *Students will understand the utility of different genomes in evolutionary analyses.*
- *Students will demonstrate the ability to carry out bioinformatical analyses across multiple platforms.*
- *Students will be able to apply a variety of methods to addressing evolutionary problems.*

Learning Objective(s)

- *Students will be able to explain the pros and cons of using DNA sequence data from the mitochondrial genome.*
- *Students will be able to carry out all the steps of a phylogenetic analysis using DNA sequence fragments (alignment, substitution model selection, and phylogeny estimation).*
- *Students will demonstrate proficiency in command line program usage, including conducting a BLAST search with a whole genome.*
- *Students will evaluate a genome-wide SNP dataset and compare their results to the mtDNA analysis using the same samples.*
- *Students will be able to design a research methodology for evaluating an evolutionary problem using multiple genomes and analytical approaches.*

1   Introduction

2   We are privileged to be part of the renaissance era of systematic biology, driven by the rapid progression of

3   technological approaches to understanding and cataloging biodiversity. With the frequency that new methods

4   are being developed, it is a challenge for researchers to continually incorporate and integrate the latest

5   approaches. Tasked with providing and evolutionary basis for understanding the diversity of life through

6   elucidating the relationships and classification of organisms, the modern systematic biologist is availed a

7   robust toolbox for revealing cryptic diversity and revisiting confusing or unresolved taxa.

8   One outstanding taxonomic problem involves populations of leopard frogs (Ranidae: *Rana*: *Pantherana*) that

9   occur in the Lenca Highlands of southwestern Honduras. In the only taxonomic assessment of these

10   populations, McCranie and Wilson (2002) used comparative external morphology and opted to consider

11   these populations to be hybrids between two species found in the adjacent Caribbean and Pacific lowlands, *R.*

12   *brownorum* and *R. forreri*, based on morphological traits inferred to be intermediate between those

13   populations. No genetic evidence has been provided to support or refute the hybrid hypothesis.

14   Using samples from all three OTU's (*R. brownorum*, *R. forreri*, and *R. brownorum* X *forreri*), students will

15   utilize two sets of data and methodological approaches to assess the validity of thy hypothesis that the

16   highland populations represent hybrids between the two lowland forms. Students will use the program

17   MEGA6 to analyze data from two widely used mitochondrial loci, 16S and COI. Students will then gain

18   experience working in UNIX, the command line operating system that provides the platform for programs

19   used in the analysis of large genomic datasets. This will allow students to analyze a set of genome-wide single

20   nucleotide polymorphisms (SNPs) from all three populations.

21   *Intended Audience*

22   This module is intended for upper division biology majors and biology graduate students enrolled in BIOL 451/551

23   Evolutionary Biology.

24   *Required Learning Time*

25   The module is intended to be taught over the course of four weeks (one 3 hour session per week).

26   *Pre-requisite student knowledge*

27   Students should have a foundational understanding of the principles of genetics, typically having completed a

28   prerequisite course in genetics (BIOL 202 Genetics and Development). Prior to this module, students should have

29   been provided a basic introduction to working with DNA sequence data, including alignment, generation of

30   summary statistics, selection of best-fit models for nucleotide substitution, and phylogenetic analysis.

31   Scientific Teaching Themes

32   Explain how the Lesson relates to the Scientific Teaching Themes of:

33   *Active Learning*

34   *Tutorials* – Students will be engaged through instructor-led tutorials in a lab setting, which
35   allows the instructor to monitor the progress of students while answering questions and guiding
36   students through troublesome areas.

37   *Independent In-class Work* – For the mtDNA phylogenetic project, students will rely on previous
38   instruction and work independently to design a short study and carry out the analyses. This will
39   take place during the class, allowing students to ask questions and troubleshoot issues with the
40   instructor.

41   *Group Discussion* – results will be discussed in an open class forum, to promote expression of
42   diverse interpretations of results and viewpoints on how the methods or analyses could be
43   improved.

44   *Assessment*

45   *mtDNA Phylogeny Results* – Students will provide submit their alignments to the instructor, and
46   printed or digital copies of modeltest results, summary statistics, and two phylogenies. These
47   results will be accompanied by a one-page summary of the results and the student's
48   interpretation of the phylogenies.

49   *Command Line BLAST Search Results* – Students will be assessed based on successful completion
50   of the genome BLAST search; they will complete a short worksheet where they provide and
51   interpret results from the results output file.

52   *Results Discussion Session* – After completion of analyses of the SNP dataset, students will write a
53   brief 1-2 page summary of the results and their interpretation, which will be turned in during the
54   4th lab of the module. We will then spend ~1 hour in an open discussion of the results.

55   *Project Design Practicum* – This module will culminate in an in-class, open book, written
56   practicum in which students are given a choice of outstanding evolutionary questions and will
57   have the 3 hour period to design a methodological approach for attacking their question. The
58   methods should include utilization of multiple genomes and analytical approaches, and should
59   include relevant references, a summary of available data,

60

Lesson Plan

61

62   Provide a detailed description of the Lesson that is sufficiently complete and detailed to enable another
63   teacher to replicate it. You may think of it as if you were explaining to a colleague how to teach your class for
64   you next week, or preparing a detailed methods protocol for a lab technique. For clarity, you may need/want

65 to include subsections such as: pre-class preparation and in-class script. Add subheadings as needed,

66 following the embedded styles in Microsoft word:

67 Lab 1 – Evaluating hybridization using mitochondrial gene fragments

68 Pre-class preparation

69 Students will have already completed a two-lab module that covers the following topics: finding

70 homologous sequences for a target sequence using BLASTN; aligning a set of homologous sequences;

71 outgroup selection; evaluating an aligned dataset using summary statistics; selecting a best-fit model of

72 nucleotide substitution for their alignment; and conducting phylogenetic analysis using neighbor-joining and

73 maximum likelihood methods.

74 In class activities

75 Students will be introduced to the study system and problem, and will be given DNA sequences from

76 to mtDNA loci (16S and COI) to form their basis of their analysis. Students will work independently to design

77 a methodological approach for analyzing the data using the methods covered in the previous two weeks.

78 Lab 2 – Taking Command - working with genomic data in Unix

79 Pre-class preparation

80 For homework, students will complete a short tutorial and associated exercises to familiarize

81 themselves with working in UNIX.

82 In class activities

83 Students will complete a series of short, instructor-led tutorials in carrying out basic command-line

84 driven operations using simple genomic datasets. Students will then receive an "unknown" genome, and will

85 carry out a command line BLAST search to identify homologous genomes.

86 Lab 3 – SNP'ing away at the Hybrid Hypothesis

87 Pre-class preparation

88 For homework, students will read a short paper on genome-wide SNP scans using ddRADSeq

89 methods, and will complete a short introductory tutorial on the program PyRAD.

90 In class activities

91 Students will be provided a two subsets of demultiplexed ddRADSeq data, and will evaluate the

92 datasets at both population and phylogenetic levels. One dataset will be trimmed down to a size that will

93 facilitate successful analytical runs during the class, in order to familiarize students with the methods and

94 interpretation of results. The second dataset will be larger, and students will complete analyses outside of the

95 classroom.

Lab 4 – Discussion of SNP results and Lab Practicum

Pre-class preparation

Students will complete their SNP analysis and associated write-up prior to class.

In class activities

The first 30 – 60 minutes of class will be devoted to discussion of the SNP analysis and results. The remainder of the time will be devoted to the Project Design Practicum (outlined above under assessment).