The Cow Diversity Project: introducing undergraduates to molecular population genetics.

Richard M. Kliman *, Dept. of Biological Sciences, Cedar Crest College
André Walther, Dept. of Biological Sciences, Cedar Crest College
Debra Walther, Biology Dept., Muhlenberg College

* corresponding author:
email: rmkliman@cedarcrest.edu
telephone: 610-606-4666 ext 3501
mailing address:
Dept. of Biological Sciences
Cedar Crest College
100 College Drive
Allentown, PA 18104

Overview.

The Cow Diversity Projects allows students to use common molecular genetic tools to explore genetic variation within and between species. The Cow Diversity Project was developed with the following learning goals in mind: (1) to reinforce understanding of molecular genetics principles and methods; (2) to introduce principles and methods in molecular population genetics; (3) to introduce the method of linear regression; and (4) to guide students through the process of writing a formal, manuscript-style research report.

Ultimately, the multi-week exercise leads to analysis of DNA sequence **diversity** (variation within species) and **divergence** (variation between species) at two mitochondrial loci: the D-loop and the *ND3* gene. While students may have heard these terms before, it is rare that they have actually worked with data that show how individuals vary within and between species. They also do not likely have intuition for the extent of genetic similarity among individuals, even between individuals of different (but closely related) species. This project allows students to "see" divergence and diversity in a concrete way. Examining the genetic variation, or the lack of genetic variation, in a population can also lead into discussions about the implications of genetic diversity, or lack thereof, in populations.

Students begin with uncooked beef samples that they have individually obtained from a source away from campus (often close to their hometown). They extract and purify genomic DNA, amplify the two loci by PCR, sequence the loci (either in-house or using an external service), and analyze sequences that meet suitability criteria. Both diversity and divergence are estimated by average pairwise measures of variation, the rationale being that alternative measures would not be intuitive with their limited background.

The exercise can be completed in four 3-hour lab sessions. Details are provided in the accompanying notes for instructors, but this is the general schedule.

Week 1 (previous evening): set up overnight Proteinase K digest of beef sample
Week 1 (in lab): genomic DNA purification, digestion of λ-phage DNA with *Hin*dIII
Week 2: PCR, assessment of genomic DNA
Week 3: PCR purification, assessment of PCR, DNA sequencing reactions
Week 4: Analysis of DNA sequence variation

* Between Week 3 and Week 4, the DNA sequences must be curated and aligned. Students are provided with multiple sequence alignments for analysis in Week 4.

Context.

The exercise has been used for more than ten years in college sophomore-level *Genetics* or *Cell and Molecular Biology* courses required for all bioscience majors at Cedar Crest College. Although our first-year/second-year core curriculum has changed, students have been introduced to fundamental principles of classical genetics, population genetics (Hardy-Weinberg equilibrium), and molecular genetics (central dogma) in both lecture and lab in prerequisite courses. Students performed PCR in the first-year curriculum.

The main difference in higher-level preparation has been temporal, not qualitative. In 2004-2008, and starting again in 2016, our students took *Genetics* in the second year. This course covered concepts in classical, population, and molecular genetics in a single semester. In 2009-2015, genetics material was split between a pair of courses: *Ecology, Evolution and Genetics* in the fall and *Cell and Molecular Biology* in the spring; the exercise was used in the latter. Thus, while population genetics material would be fresher when the exercise was used in *Genetics*, the lecture material necessary for the lab to make sense was essentially the same. Higher-level preparation in principles of evolution occurred in the second-year fall course (*Ecology, Evolution and Genetics* paired with *Cell and Molecular Biology*, and *Animal Ecology, Development and Evolution* paired with *Genetics*).

The exercise was introduced in Spring 2016 for the first time in the 200-level *Genetics* course offered as an elective to bioscience students at Muhlenberg College. Students enrolling in this course are in their second semester of sophomore year or later and would have completed a three-semester *Principles of Biology* sequence of core courses (the latter two semesters include an accompanying lab component). Students have been introduced to fundamental principles of classical genetics, population genetics (Hardy-Weinberg equilibrium), and molecular genetics (central dogma) in both lecture and lab. Higher-level preparation in these topics and in principles of evolution in the *Genetics* course was helpful in preparing students for this lab exercise. Students performed PCR several times in the *Principles of Biology III* lab curriculum and covered principles of DNA sequencing and analysis.

Ideally, students will have been introduced to the following principles and techniques prior to the exercise:

Principles

- Central dogma as it relates to the exercise (DNA → mRNA → polypeptide)
- Introduction to noncoding DNA
- Mitochondrial structure/function (for context of the loci)
- Mechanisms of DNA replication
- PCR (and its relationship to DNA replication)
- Restriction enzyme function
- Point mutations (including synonymous vs. nonsynonymous categories)
- Descriptive statistics (including variance)
- Basic principles of population ecology
- Phylogenies of closely related taxa

Techniques

- Familiarity with standard equipment for molecular biology (*i.e.*, micropipettors, microcentrifuges, balances, vortex mixers)
- Agarose gel electrophoresis (including sample loading)
- Manual graphing (specifically, *x*-*y* plots)

While prior experience performing PCR is beneficial, it is probably not essential. However, this should be a consideration in time management during the laboratory session. Also, second-year bioscience students at Cedar Crest College begin to write sections of, and ultimately complete, lab reports in the fall. Given that we have students write formal reports as part of this exercise, prior experience writing reports should be a consideration. We do provide detailed guidance regarding the structure of the report, and students have been introduced to primary literature; however, we do not expect that students will be able to fully comprehend primary research articles. We have experimented with having students do literature searches and cite relevant articles, but have concluded that they were not really gaining much from this; in recent years, we have provided students with two articles that are clearly relevant.

Rationale for selection of the organism and loci

Cows were selected for this project for two reasons. First, tissue samples are easily obtained, and a geographically diverse sample collection is possible with good planning. We recommend that students obtain samples during a break, when they are likely to disperse. Prior to leaving for a break, students are provided with the materials and instructions for sample collection and storage.

The second reason for using cows is that they are not humans. While using their own DNA may increase student engagement, there are ethical considerations. At a most fundamental level is genetic privacy. Unless a student must know something about her/his genotype for reasons that are of the student's choosing (*e.g.*, choosing to participate in forensic genotyping projects, which require that one know one's genotype to exclude contamination), there is no compelling reason for students to learn anything about their genotypes from participation in an academic exercise. While various markers (*e.g*, forensic microsatellite loci) may currently have no known relationship to phenotype, this may

change in the future; markers may turn out to be statistically associated with phenotypic character states. We recognize that not all will agree, but we choose to eschew exercises that reveal personal genetic information to students.

A third reason for using cows is that mitochondrial DNA sequences for a close evolutionary outgroup, *Bison bison*, are available. This allows for comparison of diversity within cows to divergence between cows and bison.

Mitochondrial loci were selected for technical reasons. First, they are usually effectively monoploid, being maternally inherited and generally invariant within an individual. Nuclear genes may be heterozygous, which complicates DNA sequencing, especially if there is length heterozygosity. Second, mitochondrial loci are at high copy number relative to nuclear loci. Thus, even poorly stored tissue samples usually yield template DNA of sufficient quality and quantity for PCR. Third, the *ND3* gene is short (345 bp) and close to its adjacent genes; consequently, it can be sequenced in its entirety in both directions using PCR (and sequencing) primers that correspond to conserved protein-coding regions of adjacent genes. [The primers for the D-loop were selected by aligning a broad geographic sample of cow sequences and identifying regions that were well-conserved among individuals.]

It is expected (and generally revealed) that the noncoding D-loop has a higher level of diversity and divergence than the protein-coding *ND3* locus. Almost all of the variation in *ND3* is synonymous, and there are numerous fixed differences from bison. In our experience, there are 18 fixed differences, only one of which leads to an amino acid substitution. As expected, divergence is generally at least an order of magnitude greater than diversity.