American Museum of Natural History

Science Research Mentoring Program

GENETICS

This course introduces students to the structure and function of the genome, techniques in molecular biology used in Museum laboratories, and to the generation, transmission, and study of genetic variation in the context of evolution and species identification.

Session One: DNA Biochemistry

LEARNING OBJECTIVES

Students should understand that DNA is common to all living organisms, know the principles of base pairing, identify and define the 5′ and 3′ ends of a DNA molecule, and name the structural features of DNA that are pertinent to its study in the laboratory.

KEY TOPICS

- What is DNA?
- Nucleotide structure
- Hydrogen bonding, pyrimidines vs. purines, and base pairing
- Antiparallel structure of DNA

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
15 minutes	**Assessment and Icebreaker**	Have all students introduce themselves, and provide one fact about DNA.
20 minutes	**Tour of Sackler Lab**	Discuss what we do with DNA at the museum: purify, amplify, sequence.
15 minutes	**DNA Structure Review: Part 1**	Divided into groups, students list everything they know about the structure of DNA. Ask them to organize from small to large. Give magnetic DNA models to help brainstorm.
50 minutes	**Discussion: DNA Structure Review**	Map student lists on whiteboard, from the smallest structures to the largest. Include these key points, even if not generated by students: elements in DNA, nucleotide structure, biochemistry of base pairing and strand formation, hydrogen bonds in base pairing vs. covalent bonds in backbone, antiparallel structure, chromatin.
20 minutes	**Introduction to the Use of DNA in Museum Research**	Discussion of cladogram on 4th floor – what do trees represent, and how does this relate to DNA? What does genetic information contribute to the study of evolutionary relationships?

Session One: DNA Biochemistry (continued)

MATERIALS

- Magnetic molecular models of DNA
- Whiteboard markers

PREPARATION

None required.

EXHIBITION HALLS

- Fourth Floor: Dinosaur Tree

AUDIO-VISUAL NEEDS

None required.

HOMEWORK

Optional reading: Graham, E. 2007. DNA Reviews: Ancient DNA. Forensic Sci Med Pathol, 3:221

http://www.springerlink.com/content/k6v11122m42550w2/

Session One: DNA Biochemistry

Activity: DNA Structure Review

OVERVIEW

This activity assesses students' prior knowledge of DNA structure and biochemistry, and uses whiteboard mapping and DNA models to organize facts about DNA structure from the atomic to the cellular levels. It emphasizes aspects of DNA structure that are important for common laboratory techniques such as polymerase chain reaction.

TIME FRAME

approximately one hour

MATERIALS

- molecular models of nucleotides, preferably magnetic (demonstrating hydrogen bonding and sugar-phosphate bonding). If necessary, the activity can be done without models.

PROCEDURE

1. In pairs or small groups, have students brainstorm a list of everything they know about DNA structure. Ask students to begin with the smallest unit of DNA structure or components that they can come up with (the elements found in DNA). Then ask them to come up with the next largest, and so on. Remind them to focus on what DNA looks like at the molecular level, not what it does. Then ask them to choose the single most important fact about the function of DNA (hereditary material).
2. Map these student-generated facts about DNA structure on a whiteboard, from smallest to largest. At each step, use the magnetic models to illustrate, and ask students to draw the structures or molecules wherever applicable.

   The following should be touched on:
   - Elements in DNA (C, H, O, N, P)
   - Structure of a single nucleotide (sugar, phosphate, base)
   - Connecting nucleotides to each other (Covalent bonds in the sugar-phosphate backbone; Hydrogen bonds between base pairs)
   - Base pairing
   - Purines vs pyrimidines (What is the difference in basic structure? Why do these always pair with each other?)
   - G-C vs A-T pairs (number of hydrogen bonds)
   - Antiparallel structure (5', 3')
   - DNA in the cell (chromatin, histones, and chromosomes)

Session Two: The Dogma of Molecular Biology

LEARNING OBJECTIVES

Students should be able to describe the distinguishing characteristics of a DNA molecule, an RNA molecule, and a protein molecule. They should also be able to explain that DNA codes for RNA, which is translated into protein, and that proteins are the major molecular players in cellular structure and function.

KEY TOPICS

- DNA chemistry vs RNA chemistry
- Protein function and biochemistry
- Amino acid code & degeneracy
- Transcription and translation

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Review of the Dogma**	Lecture covering basics of RNA structure, transcription, translation, and functions of proteins.
1 hour 15 minutes	**Modeling Transcription and Translation**	In groups, students build candy & toothpick models of transcription and translation.
15 minutes	**Model Sharing**	Students explain their models to each other.

MATERIALS

- toothpicks
- candy (marshmallows, M&Ms, sour patch kids, Twizzlers, etc)
- large paper to cover tables
- bowls
- markers

HOMEWORK

Optional reading: Henikoff, S. 2002 Beyond the Central Dogma. Bioinformatics, 18:223
http://www.scienceboard.net/community/perspectives.81.html

Session Two: The Dogma of Molecular Biology

Activity: Modeling Transcription and Translation

OVERVIEW

This is a hands-on review of the process of RNA transcription and protein translation. Students use candy to build relatively detailed models of these processes, and explain their models to each other.

TIME FRAME

one hour, 15 minutes

MATERIALS

- craft paper rolls
- assorted candies in a variety of colors, shapes, and sizes (such as marshmallows, M&Ms, gummi anything, Twizzlers)
- toothpicks
- markers
- bowls for candy

PROCEDURE

1. Divide students into small groups. Cover tables with craft paper and provide students with colored markers. Explain that they will be working in groups to make a model that illustrates the following process: transcription and translation of a four-amino-acid peptide. The model can be made of any materials, and will be used to explain these processes to someone else. Tell them to make sure to distinguish between DNA and RNA molecules, to include start and stop codons, and to label parts as necessary.
2. Let students work fairly independently as time allows.
3. Bring students together as a class and have each group explain their model to the class.

Session Three: Heredity and Mendelian Genetics

LEARNING OBJECTIVES

Students should be able to explain the relationship and difference between alleles, genotypes, phenotypes, and traits; construct and analyze a punnet square.

KEY TOPICS

- Mendelian inheritance
- Homozygous vs heterozygous
- Dominant, recessive, codominant
- Pedigrees
- Monogenic vs polygenic traits

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Lecture**	Covering the relationships between genotype, phenotype, traits, gene, allele, locus. Basic Mendelian genetics (Punnet square construction, dominant, recessive, and co-dominant alleles).
1 hour 10 minutes	**Activity: Phenotype, Genotype, and Paternity**	Students type a family's "blood samples" and determine the phenotypes and genotypes of siblings and parents
20 minutes	**Discussion**	Beyond Mendel: monogenic vs polygenic traits, epigenetics

MATERIALS

- Blood typing and pedigree worksheets
- "blood" and "antigen" solutions aliquoted
- micropipetters
- pipet tips
- parafilm

Session Three: Heredity and Mendelian Genetics (continued)

PREPARATION

Aliquot solutions for each group.

HOMEWORK

Optional reading: Crow, JF. 1993. Felix Bernstein and the First Human Marker Locus. Genetics 133:4

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1205297/?tool=pmcentrez

Session Three: Heredity and Mendelian Genetics

Activity: Phenotype, Genotype, and Paternity

OVERVIEW

This activity illustrates the relationship between genotype and phenotype by using the principles of Mendelian inheritance to identify a family pedigree. Using a synthetic blood typing kit, students determine the phenotypes and potential genotypes of a group of siblings and their potential parents, then reconstruct the siblings' pedigrees and actual genotypes.

TIME FRAME

one hour, 10 minutes

MATERIALS

- worksheets
- micropipetters
- pipet tips
- Parafilm or blood typing trays
- Carolina synthetic blood Whose Baby? Kit. Label aliquots of the "blood" as follows (70 ul each), enough for one per group:

Siblings:	Mother: A	Potential Fathers:
1 – A		1 – AB
2 – AB		2 – O
3 – O		3 – B
4 – A		
5 – B		

- Aliquot approximately 500 ul of the A and B "antisera" per group

PROCEDURE

1. Ask students what they know about blood type. How many blood types are there? Use this knowledge about to discuss the blood alleles A, B, and O, and have the students list all the potential genotypes for this locus. Have the students rely on prior knowledge to explain whether each allele is dominant, recessive, or codominant. Then ask them to list the phenotypes associated with each genotype. There are 6 possible genotypes (AA, AO, BB, BO, AB, and OO) but only 4 possible phenotypes (A, B, AB, and O).

Session Three: Activity: Phenotype, Genotype, and Paternity (continued)

2. Briefly discuss the use of antiserum to determine blood type. Ask students to explain whether this type of test describes genotype or phenotype, and when phenotype would or would not make it possible to predict paternity.
3. Students identify the phenotype of each sample by pipetting antiserum and blood samples onto parafilm and checking for precipitation. Next, ask them to answer the questions on the worksheet. Note that this activity assumes that students are familiar with the use of a micropipetter. If not, allow time for them to practice, or use the dropper bottles provided by the kit.

Session Three: Heredity and Mendelian Genetics: WORKSHEET

Blood Type: Phenotype, Genotype, and Paternity

INTRODUCTION

Blood type is determined by co-dominant alleles with a Mendelian pattern of inheritance. There are three alleles for human blood type: A, B, and O. A and B are co-dominant, while O is recessive. A and B are genes involved in the production of proteins that are present on the surface of red blood cells, and can cross-react with antibodies, or immune proteins. The O allele is a deletion mutation, and means that the blood cell protein can't be detected using antibodies.

How many possible genotypes are there for blood type? How many possible phenotypes are there for blood type?

Blood type is tested by using antibodies that react with either the A or B-modified proteins. If anti-A antibody is added to either type A or AB blood, an agglutination reaction will occur and the blood cells will clump up in visible grains. If anti-A antibody is added to type B or O blood, no agglutination reaction will occur, and the blood remains smooth.

Prior to the development of DNA fingerprinting technology, blood typing could be used to attempt to determine paternity. Consider the following scenario: A family has quintuplets, who were conceived using in vitro fertilization. The mother suspects that there was a mix-up at the clinic, and that her husband is not their biological father.

Your job is to investigate which of three possible candidates is the father of these quintuplets. You will be given blood samples from each of the quintuplets, the mother, and three potential fathers, as well as anti-A and anti-B antibodies. You will test each blood sample to answer the questions on the back of the page.

PROCEDURE TO TEST BLOOD SAMPLES

1. Pipet 30 ul of each blood sample onto parafilm, or into each of two wells of a blood typing slide (30 ul per well). If you are using parafilm, use a Sharpie to label one side "A" and one side "B".
2. Add 30 ul of anti-A (blue) to the well/parafilm labeled A.
3. Add 30 ul of anti-B (yellow) to the well/parafilm labeled B.
4. Using the pipet tip, gently stir the blood and anti-serum drops for 30 seconds.
5. Carefully examine the films of liquid left behind. If a film remains uniform in appearance, there is no agglutination. If the sample appears granular, agglutination has occurred.
6. Wipe the blood typing slide with a wet paper towel and dry between each use.

WORKSHEET: Blood Type: Phenotype, Genotype, and Paternity (continued)

YOUR OBSERVATIONS

1. What is the phenotype of each individual?

2. What is the genotype of each parent?

3. Who is the father?

4. What is the genotype of each quintuplet?

5. The mother's husband is potential father #2. What is the probability that their next child will have blood type B? What is the probability that their next child will have blood type O? Show your work.

Session Four: What is in the Genome?

LEARNING OBJECTIVES

Students should be able to explain the difference between coding and noncoding DNA, know that coding DNA makes up only a small percentage of the eukaryotic genome, and be able to name three DNA elements that define a coding region.

KEY TOPICS

- Coding vs noncoding DNA
- What characterizes a coding region – polyadenylation, Kozak sequences, ORFs
- Functional non-coding DNA – promoters, telomeres, introns
- Repetitive sequences – transposons, microsatellites, telomeres, duplications
- Identification of individuals using microsatellites – applications to pop/con bio

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Lecture**	Covering definitions of coding and non-coding DNA, genes, features of coding regions
30 minutes	**Activity: Gene Finding**	Students use online tools to identify start/stop sites, ORFs, Kozak sequences, polyA signal sequences, etc. in an unknown sequence. Adapted from DNA Learning Center, http://www.dnai.org/c/index.html?m=1,1
1 hour	**Lecture**	Covering functions of noncoding DNA, repetitive sequences, gene duplications, use of microsatellites in pop/con bio and paternity testing

MATERIALS

- Laptops w/sequence files and links to online tools
- Paper copies of sequence for annotation

PREPARATION

Load sequence files

HOMEWORK

Optional reading: VIB (the Flanders Institute for Biotechnology). "Saved By Junk DNA: Vital Role In The Evolution Of Human Genome." ScienceDaily, 30 May 2009. Web. 15 Jul. 2011.

Session Five: Mutations and the Mutation Rate

LEARNING OBJECTIVES

Students should understand the various sources of mutations, their potential consequences, and the effect of selection on the heritability of mutations and the mutation rate.

KEY TOPICS

- Mechanisms of mutation generation
- Synonymous vs nonsynonymous mutations
- Heritability of mutations
- Mutation rates

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Lecture**	Mutations and how they occur
45 minutes	**Worksheet: Mutation**	Students complete individual worksheets covering sense, missense, nonsense, frameshift mutations
45 minutes	**Discussion**	What affects the mutation rate? Compare b/w species, types of DNA (coding vs. noncoding). Why is this important when scientists use DNA to infer phylogeny?

MATERIALS

- mutation worksheets, found here: http://heredity.wikispaces.com/file/view/mutations_worksheet.pdf

HOMEWORK

Optional reading: Marris, E. Molecular clock tied to fossil record. Nature News, Oct 11, 2004. http://www.nature.com/news/2004/041011/full/news041011-2.html

Session Six: DNA Sequence Databases

LEARNING OBJECTIVES

Students should be able to use Genbank to find and download nucleic acid or protein sequences, explain the difference between mitochondrial and nuclear DNA, and the difference between introns and exons.

KEY TOPICS

- Using Genbank and other sequence databases
- Independent project: collecting gene sequences for alignment

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Lecture and Demonstration**	Introducing students to the Genbank database with sample searches. Important databases: Protein, Nucleotide, Gene. Students can follow along on laptops.
30 minutes	**Activity: Genbank Game**	Students compete to find facts using Genbank
1 hour	**Independent Project: Downloading Sequences**	Students work in pairs to decide on a question/ hypothesis for their independent project, and locate and download DNA sequences of appropriate species/genes.

MATERIALS

- Laptops with internet, worksheets for independent project

Session Six: DNA Sequence Databases

Activity: Genbank Game

OVERVIEW

Students practice retrieving information from NCBI's Genbank in a gameshow-style contest.

TIME FRAME

30 minutes

MATERIALS

- Computers with internet (one per group)
- Projector and Powerpoint presentation with questions

PROCEDURE

(Assumes you have reviewed the Nucleotide, Protein, and Gene databases on Entrez.)

1. Assign students to internet-connected computers. (Group size will depend on how many computers are available; students in large groups can take turns searching for answers.) Explain that they will be doing searches for information on Genbank in head-to-head competition – the first group to find a correct answer wins a point. Offer an appropriate prize. (For example, let the winning group choose get the first choice of taxa for the independent project.)
2. Have the students load the Genbank main page: http://www.ncbi.nlm.nih.gov/genbank/
3. Remind students that they need to read the questions carefully in advance: will they need to look for a protein sequence? An mRNA sequence? Which will be the appropriate database to search? Will they need to add any additional information besides the species and the gene name, such as "mRNA"? Remind them that they may need to use the scientific name of the organism in their search to get the best results.
4. Project slides with the following questions. When a group arrives at a correct answer, tally the point and move on to the next slide. If necessary, do the first few questions as a group, making sure to ask students what database they think they should search, and why.

   NOTE: As the database is updated, you may want to doublecheck the answers to these questions.

Session Six: Activity: Genbank Game (continued)

- From what species is the gene with accession number GU452716?
  Answer: Triticum aestivum

- How many base pairs are transcribed from the chimpanzee dopamine receptor D2 gene?
  Answer: 1322

- How many exons does the human actin alpha 1 gene have?
  Answer: 6

- How many amino acids does the cytochrome c oxidase subunit 1 have in Aspidosiphon elegans?
  Answer: 216

- For how many different species is there a record of a sonic hedgehog gene or gene homolog?
  Answer: 16

- What is the first amino acid in the Bos Taurus alpha-synuclein protein?
  Answer: M

- How many introns does the human Lrp10 gene have?
  Answer: 6

- What chromosome is the rat Mdm2 gene on?
  Answer: 7

- What are the last three amino acids of the Danio rerio Wnt2 protein?
  Answer: TQS

- What is the 300th nucleotide in the transcript of human neuropeptide Y?
  Answer: A

Session Six: DNA Sequence Databases

Activity: Independent Tree Building Project

Day 1: Choosing a question and downloading sequences

OVERVIEW

This independent project leads students through the construction of a phylogenetic tree, from downloading sequence data to aligning sequences to tree building. Students will compare two types of DNA to build trees of relationships of the same set of taxa, and then consider which tree is more "accurate" and why. Students will work in pairs, and present their work in the form of a scientific poster.

This activity is step one of the independent project: students choose a set of taxa to focus on from a given list, choose 2 types of DNA to compare, generate a hypothesis, and download and format the DNA sequences.

TIME FRAME

1 hour

MATERIALS

- Computers with internet access
- Copies of reference papers for each set of taxa
- Sequence downloading worksheet

PROCEDURE

(Assumes you have reviewed the Nucleotide, Protein, and Gene databases on Entrez.)

1. Divide students into pairs. Give them the independent project overview handout, and have them select a) a question, i.e., which types of DNA they will compare in this exercise, and b) a taxa to focus on. Provide the appropriate reference paper for each taxa. Have each team choose a number of species in the paper, and locate the gene IDs they'll need to search for in order to download the sequences. Important: students will need to choose an outgroup for their tree and download the outgroup sequences as well. Appropriate outgroups can be found in the reference papers.
2. Before downloading sequences, have each pair come up with a hypothesis as to which tree they build will be more accurate. Have students explain what is meant by "accurate" in this case (i.e., closest to the tree in the reference paper with the more complete dataset), and share their hypotheses and reasoning.

Session Six: Activity: Independent Tree Building Project (continued)

3. Give students sequence downloading and formatting worksheets to follow. Each group will need to download either 1 or 2 sequences per species. If they are comparing nuclear introns or exons and mtDNA, they will need to download 2 separate sequences. If they are comparing longer to shorter DNA sequences, they will only need to download one sequence.
4. Make sure that students are formatting their sequences according to the instructions, and that they are properly saved for the next session's alignment activity.

REFERENCES

(students may need information provided in supplements, so be sure to include these):

Rodents: Montgelard, C et al. 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evolutionary Biology 8:321

Turtles: Naro-Maciel, E et al. 2008. Evolutionary relationships of marine turtles: A molecular phylogeny based on mitochondrial and nuclear genes. Molecular Phylogenetics and Evolution, 49:659

Lemurs: Yoder, A and Yang, Z. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. Molecular Ecology, 13:757

Plasmodium: Perkins, S and Scholl, J. 2002. A molecular phylogeny of malarial parasites recovered from Cytochrome b gene sequences. The Journal of Parasitology 88:972

Session Six: DNA Sequence Databases: WORKSHEET

Genetics Independent Project

INTRODUCTION

Working with a partner, construct phylogenetic trees (trees depicting evolutionary relationships) using different portions of genomic DNA (including mitochondrial DNA, nuclear DNA exons, and nuclear DNA introns) or using different lengths of genes. This exercise will help you understand the different mutational rates associated with diverse loci and to conceptualize how different evolutionary relationships may be recovered using different portions or lengths of genomic DNA.

PROJECT OUTLINE

Data collection for student projects will begin with this session. You and your partner will work together over the next three sessions to construct phylogenetic trees. You will then create a poster together, and all results will be shared during Session 10. By the end of this assignment, you will be able to construct a tree of evolutionary relationships for its animal group, and to analyze the results in a phylogenetic context. This is definitely beyond the scope of any "regular" high school class, so you should be very proud of yourselves!

Here's a brief outline of what you will accomplish during each session:

Session 6: Download Genbank DNA sequences and fix file formatting

Session 7: Alignments of sequences and divergence estimates

Session 8: Phylogenetic tree building

Session 9: Creation of Posters

Session 10: Poster Presentation/Discussion Session

The data you will collect is provided below. During Session 6, we will ask you each to pick your favorite question, and you will then choose among the following taxa and loci for completion of your project.

WORKSHEET: Genetics Independent Project (continued)

DATA COLLECTION

QUESTION	TAXA	LOCI
Nuclear Exon (nuExon vs. Mitochondrial (mtDNA)	1. Rodents (Montgelard et al., 2008)	vWF (nuExon), CYTB (mtDNA)
	2. Turtles (Naro-Maciel et al., 2008)	C-MOS (nuExon), 16S (mtDNA)
	3. Lemurs (Yoder and Yang, 2004)	IRBP (nuExon), COII (mtDNA)
Mitochondrial (mtDNA) vs. Nuclear Intron (nuInt)	1. Rodents (Montgelard et al., 2008)	CYTB (mtDNA), SPTBN (nuInt)
	2. Turtles (Naro-Maciel et al., 2008)	16S (mtDNA), R35 (nuInt)
	3. Lemurs (Yoder and Yang, 2004)	COII (mtDNA), Transthyretin (nuInt)
300 bp DNA vs. all available DNA	1. Rodents (Montgelard et al., 2008)	IRBP (nuExon)
	2. Turtles (Naro-Maciel et al., 2008)	RAG-1 (nuExon)
	3. Lemurs (Yoder and Yang, 2004)	CYT-B (mtDNA)
	4. Plasmodium (Perkins + Schall, 2002)	CYT-B (mtDNA)

Session Six: DNA Sequence Databases: WORKSHEET

Downloading Sequences and File Formatting

PROCEDURE

1. Ingroup/Outgroup Selection: Decide which taxa/individuals you would like to include in your analysis (4-5 in the "ingroup"). The decision may depend upon on whether the project is focused on recovering relationships between genera, or between species, or between individuals, etc. You may want to also choose based on the geographic distribution of taxa, in order to relate how geography may influence the shape of the tree (also known as topology). You will also need one outgroup (to root the tree). The papers should list appropriate outgroup taxa.

2. Accession Numbers: Find the accession numbers for these taxa/individuals in the paper. Make sure you choose the accession numbers that correspond to the loci. (Correct loci are provided.) Accession numbers typically look something like: "HQ130531."

3. Genbank and Fasta Files: Find your sequences in Genbank (search nucleotides) at **http://www.ncbi.nlm.nih.gov/genbank/**

4. Click "FASTA" on the upper right hand side. Copy and paste this sequence to a textedit (MACS) or notepad (PCs) document. Go to "Format" "Make Plain Text". Add a ">name" above the sequence. Keep this name short (less than 10 characters, if possible), but informative enough that you can cross-reference the new names to the names in the paper. Repeat this until you have finished copying and pasting your sequences into the correct files.

5. You should have separate files for each locus! Save each of these files as "YOURGENE.TXT". For example, if you were working with ND2 and RAG1, you would have two separate files: "ND2.txt" and "RAG1.txt". These files should have sequences for the same taxa/individuals.

6. Everyone's FASTA files should look exactly like this. Don't forget to have an outgroup! Most groups will have two files – one for each locus. However, groups using different lengths of one gene will only have one file (for now).

Session Seven: Sequence Alignments and Sequence Divergence

LEARNING OBJECTIVES

Students should: understand the reason for doing sequence alignments, be able to use BLAST and Muscle to do alignments, and be able to estimate sequence divergence by hand.

KEY TOPICS

- Sequence alignments – why we need to do them
- Gene homology
- Software for alignments

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
20 minutes	**Lecture and Discussion**	What is an alignment and why do we need to do it? Discuss an image of a gene alignment and its features (matches, gaps, mismatches).
1 hour	**Worksheet: Alignment and Discussion**	Students work individually to align DNA sequences by hand and calculate percent similarity. Share percent similarities, and discuss why everyone got different results. (There is no right way to do an alignment.) Discuss the tradeoffs between mismatches and gaps.
40 minutes	**Independent Project: Alignments**	Using worksheet instructions, students use MUSCLE online to align the sets of sequences downloaded in the last class. They should save alignments; each group should have two.

MATERIALS

- Laptops
- Worksheets for alignment and independent projects

Session Seven: Sequence Alignments and Sequence Divergence: WORKSHEET

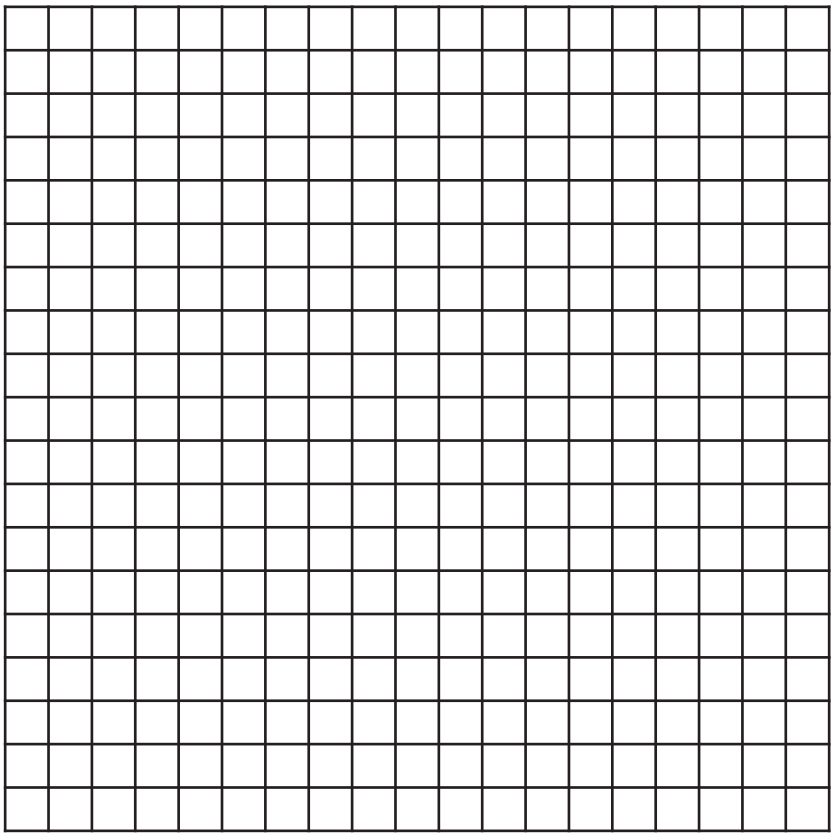
Alignments by Hand

PROCEDURE

Use the directions and grid below to align the following sequences to the best of your their ability.

1. **Label each sequence with its identifying number.**
2. **Use dashes to represent gaps in the sequence.**
3. **The longest sequence should not have any gaps.**
4. **Mismatches are allowed, but not preferable. Changes between (C and T) and (G and A) are preferred over others. Why do you think that is?**
5. **The goal is to maximize similarity, i.e., as many similar letters and as few gaps in each column as possible.**

   1 - GAATCTGCTC
   2 - GAATCCTGCCC
   3 - GAAATCTCTC
   4 - GAATCCTTGCTC
   5 - GATCCGCTC
6. **Calculate the overall percent similarity between these sequences. (Percent similarity consists of the number of sites with identical bases/total length of longest sequence.)**



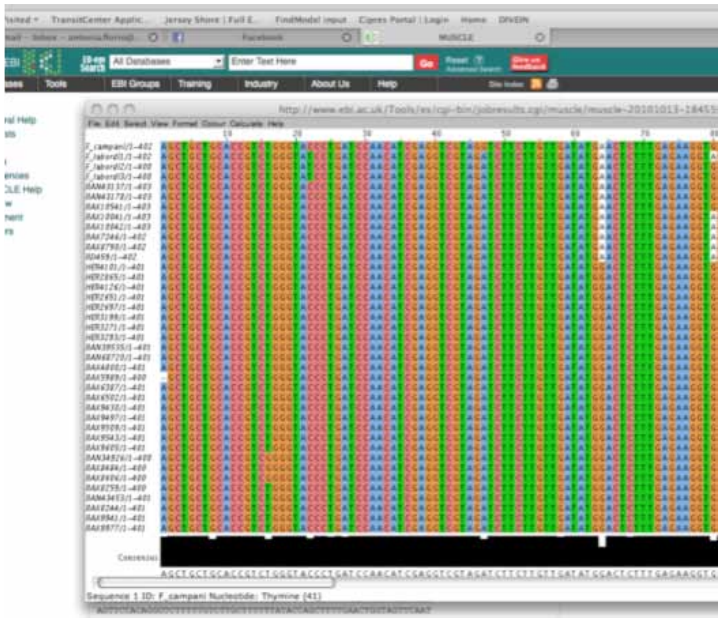
Session Seven: Sequence Alignments and Sequence Divergence: WORKSHEET

Independent Project Sequence Alignments

PROCEDURE

1. Check FASTA files from last week
   — No spaces in names
   — Make sure you have an outgroup
   — Separate files for each locus

2. Multiple Sequence Alignment
   — We will use MUSCLE
   http://www.ebi.ac.uk/Tools/muscle/index.html
   — Upload your FASTA files and then press "SUBMIT"
   — Once you get your output, go to "Result Summary", and then click "Start Jalview"
   — You should see now be able to visualize your multiple sequence alignment



WORKSHEET: Independent Project Sequence Alignments (continued)

PROCEDURE

3. FASTA File Formatting
   — Follow these steps to get your alignment into a new FASTA file
   — File > Outbox to textbox > FASTA > Copy to Notepad or TextEdit
   — Save as "yourgene_aligned.txt" (ex. ND2_aligned.txt)
   — Fix filenames: Get rid of text following the original taxa/individual name you assigned last week. For example, for the alignment above, I would need to delete every "/1-401"
   — Correct FASTA alignment files should look similar to this:

   Notice: The file is named 16S_aligned.txt, and there are no numbers behind the taxa/individual names. Also, there are now some –'s in your file. Ask students why those now occur?

4. Depends on your group! For groups looking at Questions 1 and 2 (nuclear exons/mitochondrial DNA or mitochondrial DNA/nuclear introns)
   — Simply repeat the above steps with your second locus
   — For groups looking at Question 3 (the different lengths exercise with one locus)
   — Do alignment of whole sequences and save it as "gene_alignwhole.txt"
   — Delete from either the beginning or end of the alignment so you only have about 300 bp remaining
   — Do this while in Jalview: highlight all of the sequences and simply press "DELETE" on your keyboard
   — Repeat Step Three, and now save this new alignment as "gene_alignshort.txt". You should now have two FASTA aligned documents like everyone else in the class!

Session Eight: Tree Building

LEARNING OBJECTIVES

Students should be able to use free software to construct a tree, to explain the reason that outgroups are used in tree construction, and to understand that the genes chosen for tree construction can affect outcome.

KEY TOPICS

- Tree building with molecular data
- Effect of outgroups
- Programs for building trees
- Analysis of trees

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
15 minutes	**Review of Trees**	Review major features of trees in discussion.
30 minutes	**Worksheet: Tree Building**	Students build trees by hand with sequence data. Discuss the correct tree.
45 minutes	**Independent Project: Tree Building**	Students build trees online with downloaded sequence data.

MATERIALS

- Laptops with internet connection
- Tree-building worksheets
- Independent project tree-building worksheet

Session Eight: Tree Building: WORKSHEET

Tree Building by Hand

PROCEDURE

1. Construct a phylogenetic tree of the species below from the following sequence data:

Sequence of KITH Gene	
Species Scott (Outgroup):	CCGTAGTATACG
Species Kevin:	CCGTAGTAAACG
Species Bruce:	CCCTACTAAGCC
Species Dave:	CCCTAGTAAACG

2. How many characters are in this data? How many are informative? How many are uninformative?

3. What would be the effect of changing the outgroup from Scott to Dave? If you have time, construct this tree.

Session Eight: Tree Building: WORKSHEET

Tree Building for Independent Project

PROCEDURE

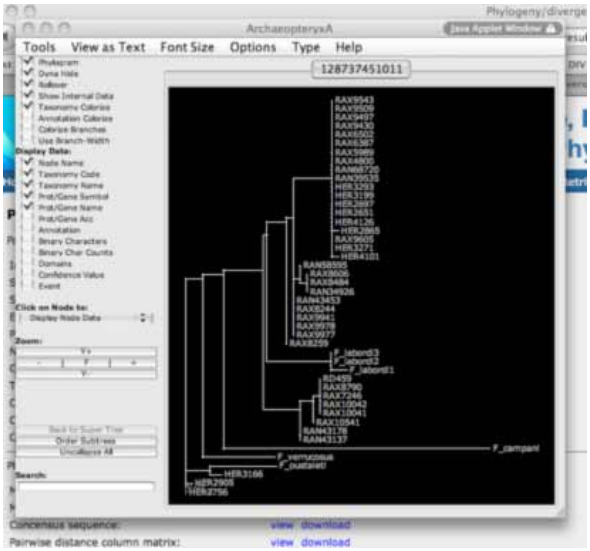
1. Check your alignment files. Make sure your aligned files look exactly like this one. Only the species name should be included after the >. Your sequences should have some – between your DNA bases; these signify gaps in the alignment.

2. Fill in the necessary information to create your phylogenetic tree. Go to this website: http://indra.mullins.microbiol.washington.edu/DIVEIN/index.html and choose the option: Phylogeny/Divergence/Diversity
   — Data Input: Upload your alignment file, make sure to choose the option "FASTA" (because you still have a FASTA file), and choose your sequence data type as "DNA"
   — Substitution Model: Make sure to choose the option "GTR", other options can remain as default
   — Tree Searching: For type of tree improvement, choose "Best of NNI and SPR", Optimise tree as "topology and branch lengths"
   — Divergence/Diversity measurements: Choose to calculate divergence from "Consensus" and Calculate divergence/diversity based on "Pairwise distance"
   — Enter your email address. Your results will be emailed to this address, so make sure you type it in correctly. Then repeat this process for your second loci.

   ***REMEMBER: Each group should have two alignment files because you are comparing the topologies obtained from two different loci. Therefore, you will have to submit two separate requests, one for each loci.

WORKSHEET: Tree Building for Independent Project (continued)

3. Retrieve your phylogenetic tree. Check your email to retrieve your results. The email should say "Please click here to receive your results."
   — Choose the option MLE tree and click "view." A black window with your phylogenetic tree should (hopefully) appear.
   — Remember that we did not assign an outgroup when we sent in our request. Therefore, we will have to now define your outgroup. Under "click on node to," choose the option "root/reroot" and then click the node connected to your outgroup taxon. Play around with the other options (for example, the option "order subtrees" might make it easier to see the relationships in your phylogenetic tree).
   — This is your phylogenetic tree for one of your loci. Complete the same task for the other loci, and start drawing conclusions about your hypotheses. Are the trees the same? Do they show the same relationships? How do your trees compare to the trees in the original papers?
   — Save and/or copy your trees. Check your alignment files. Make sure your aligned files look exactly like this one. Only the species name should be included after the >. Your sequences should have some – between your DNA bases; these signify gaps in the alignment.



4. Estimate divergence. The website offers a few ways to calculate divergence estimates (how different your sequences are from one another). In STEP ONE, we chose to calculate the divergence from consensus (the consensus sequence of all your sequences). This is not ideal, but it is one way to visualize how much your sequences diverge. Click "view" for both the "Pairwise based divergence from Consensus" and "Pairwise based diversity." Concentrate on the mean values – do these differ much between loci? Do these values differ between groups?

Session Nine: Poster Preparation

LEARNING OBJECTIVES

Students should be able to: name the elements of a scientific poster, and explain how their study was designed and whether or not their hypothesis was supported.

KEY TOPICS

- Posters as science communication
- Analysis and explanation of evolutionary trees

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
15 minutes	**Discussion: Posters**	Using a scientific poster as an example, discuss the components of scientific communication (Intro/ background, Methods, Results, Discussion)
1 hour and 40 minutes	**Poster Preparation**	Working in pairs, students make posters to present their independent project.

MATERIALS

- Markers
- Colored paper
- Butcher paper
- Tape
- Glue
- Printer (optional)

Session Ten: Poster Session

LEARNING OBJECTIVES

Students should be able to present scientific data to their peers orally and visually, and answer questions about their work.

KEY TOPICS

- Scientific communication
- Effect of gene selection on tree building

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
1 hour and 15 minutes	**Poster Presentations**	Students orally present the results of their tree building exercises
45 minutes	**Class Discussion**	Students discuss findings of class overall. What makes a gene good for tree building?

MATERIALS

- Student posters
- Tape

Session Eleven: Caviar Lab: Species Identification with Molecular Methods

LEARNING OBJECTIVES

Students should understand the purpose of using molecular information to identify species, the value of RFLP fingerprinting for species identification, the principles of gel electrophoresis, and the use and limitations of COI barcoding.

KEY TOPICS

- Why we need to identify species with molecular data
- Gel electrophoresis
- DNA barcoding – COI and its limitations

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
20 minutes	**Lecture and Discussion**	Covers principles of single-locus DNA barcoding, and its use in identifying samples of known or unknown provenance.
1 hour and 20 minutes	**Activity: Caviar Gel Electrophoresis Lab**	Students use gel electrophoresis to analyze DNA from an unidentified caviar sample to identify the species it belongs to. Introduction to RFLPvs. DNA sequencing, a few minutes of practice with micropipetting, loading of samples and ladder onto gels, running and photographing gels, and class discussion of results. Lab is adapted from AMNH Wildlife Forensics: http://www.amnh.org/education/offering.php?print=true&audience=school_groups&id=430
20 minutes	**Wrap-Up Discussion**	Discussion of single-gene barcoding and COI as the standard. What characteristics should this gene have in terms of speed of molecular evolution? Does it work across all taxa?

MATERIALS

- Gel electrophoresis apparatus
- DNA samples
- Loading dye
- Agarose gels
- TAE buffer
- Molecular marker
- Micropipetters
- Pipet tips
- UV light box

PREPARATION

Cast gels, make TAE buffer, aliquot DNA samples and laders

Session Twelve: Genome Sequencing, the Human Genome Project, and Bioethics

LEARNING OBJECTIVES

Students will be able to name one characteristic that can be inferred from the analysis of an individual human genome, and one that cannot. They will also be able to explain how an animal is cloned, and how it does or does not differ from the animal from which it was cloned.

KEY TOPICS

- Genome sequencing
- Ethics of genetic privacy
- Genetic engineering
- Cloning

CLASS OUTLINE

TIME	TOPIC	DESCRIPTION
30 minutes	**Video: This American Life: "If by Chance We Meet Again"**	Students watch a "This American Life" segment – first segment of Season One, Episode 101.
30 minutes	**Discussion of Video**	Students discuss ethics of cloning as presented in the film. What are the reasons for and against cloning animals, including humans?
15 minutes	**Video: NOVA Science Now: "Public Genomes"**	Students watch a "NOVA" segment
45 minutes	**Discussion of Video**	Students discuss science behind "personal genomics" as presented in video. Have students thoroughly explain genetic risk. Would they have the test shown in the video? Why or why not?

MATERIALS

- Computer with internet for playing NOVA video and This American Life videos