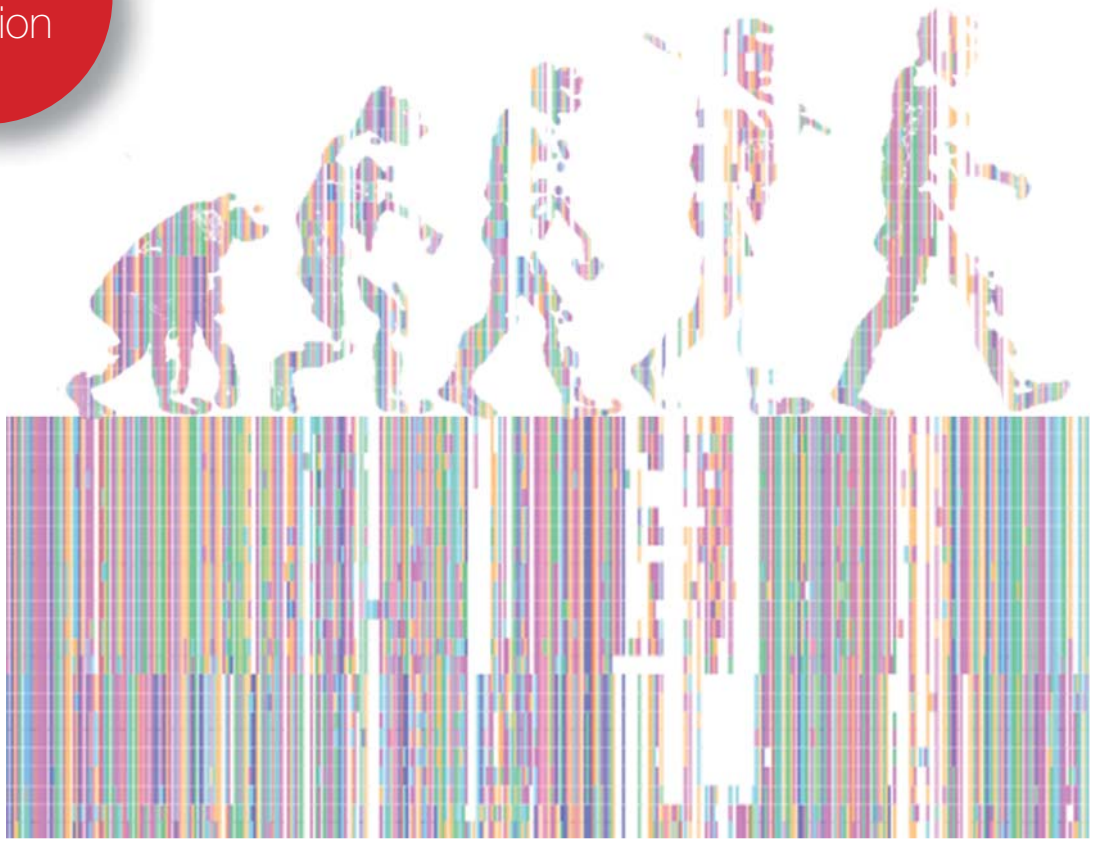
A First Look at the Code of Life

Cleopatra Kozlowski

An introductory course to concepts in Bioinformatics

English version



ELLS – European Learning Laboratory for the Life Sciences

EMBL

Cleopatra Kozlowski

A First Look at the Code of Life

Teacher Handout

Preface

This course contains four lessons on Bioinformatics aimed at high-school students. Each lesson builds on the previous one.

I have tried to design activities that aim to inspire students who have only a general background in biology. Since the course is aimed at a non-specialized audience, I draw parallels with other disciplines to make the ideas more accessible. I have also designed manual activities, which are done on paper not with a computer, since it is only too easy to let a computer do all the work, without understanding the background principles. Most of the activities are group activities for 3-4 students.

This document is for use by the instructor (and contains questions and answers). There is also a student handout.

This document was prepared at the EMBL for the European Learning Laboratory for the Life Sciences (ELLS) project. Please see:

http://www.embl.org/training/ells/index.html

Table of contents

1 Gene Finding

Title: Gene Finding
Time: 60 min + 10 min if optional computer activity
Material: Pencil, Paper
Useful website: http://www.expasy.org/tools/dna.html
Aim: Understand DNA-> RNA -> protein.

1.1 Introduction

- DNA carries the 'genetic material'. All DNA in one organism is identical (except the germline). DNA is inherited by an organisms' offspring.
- Only ~1 % of DNA in humans codes for proteins. But for this course, we are interested in that DNA!
- DNA-> RNA -> protein
- DNA is double stranded, RNA is single stranded.
- DNA and RNA are read from '5' to 3' direction.
- A-T, C-G form base pairs
- DNA has sense and antisense strands.
- mRNA is transcribed from the antisense strand.
- U is substituted for T in mRNA
- Proteins are translated from the mRNA sequence.
- Proteins are made up of amino acids
- During translation, 3 nucleotides code for a single amino acid.
- Concept of reading frame=> which sets of 3 do you choose?
- Methionine is at the beginning of a protein.
- There is a stop codon at the end of a protein.

1.2 The 20 Amino Acids and their Symbols

Name	Abbreviations	
Alanine	ala	a
Arginine	arg	r
Asparagine	asn	n
Aspartic acid	asp	d
Cysteine	cys	c
Glutamine	gln	q
Glutamic acid	glu	e
Glycine	gly	g
Histidine	his	h
Isoleucine	ile	i
Leucine	leu	l
Lysine	lys	k
Methionine	met	m
Phenylalanine	phe	f
Proline	pro	p
Serine	ser	s
Threonine	thr	t
Tryptophan	trp	w
Tyrosine	tyr	y
Valine	val	v

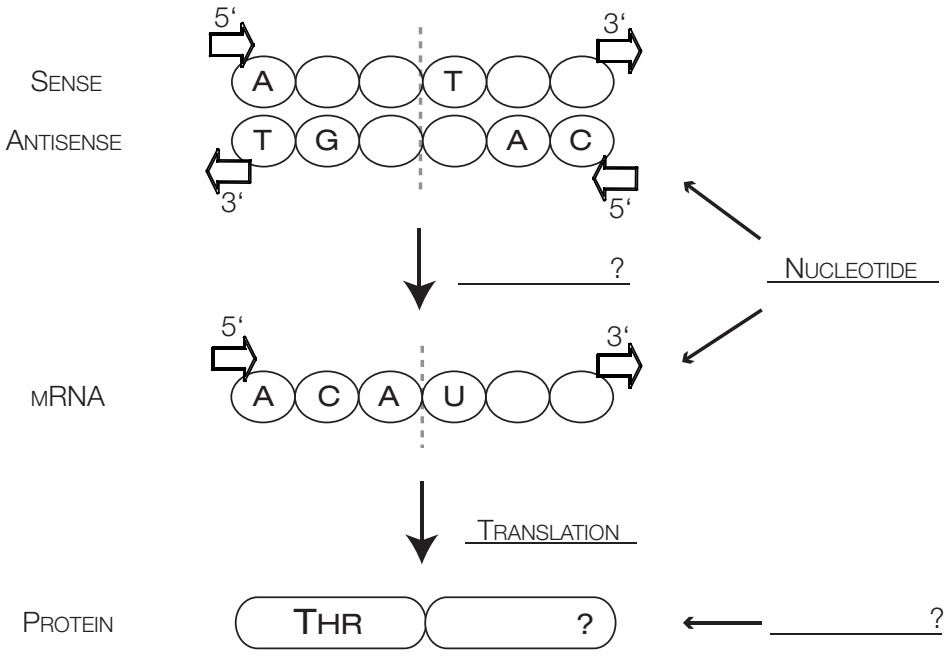
1.3 Table of the Genetic Code

In the presence of zinc ions, insulin forms hexamers (groups of 6 molecules, Fig. 1.9), resulting in a torus-like (or "doughnut") shape. Insulin is stored in β-cells and secreted in the bloodstream as a hexamer. However, the active form is a monomer.

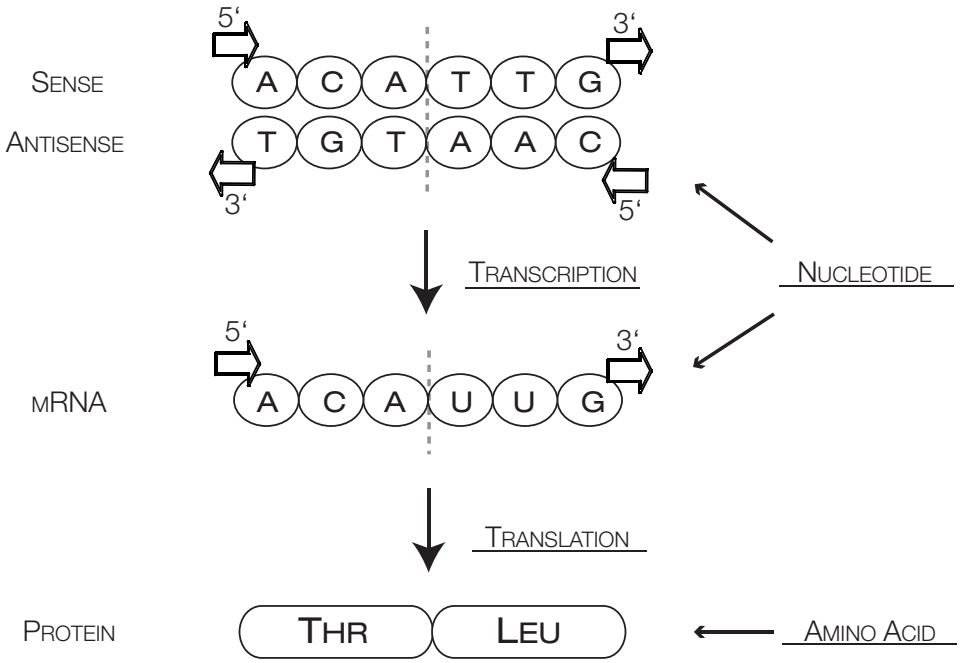
1st Base	2nd Base				3rd Base
	U	C	A	G	
U	UUUP Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine	U
	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine	C
	UUA Leucine	UCA Serine	UAA Stop	UGA Stop	A
	UUG Leucine	UCG Serine	UAG Stop	UGG Tryptophan	G
C	CUU Leucine	CCU Proline	CAU Histidine	CGU Arginine	U
	CUC Leucine	CCC Proline	CAC Histidine	CGC Arginine	C
	CUA Leucine	CCA Proline	CAA Glutamine	CGA Arginine	A
	CUG Leucine	CCG Proline	CAG Glutamine	CGG Arginine	G
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	U
	AUC Isoleucine	ACC Threonine	AAC Asparagine	AGC Arginine	C
	AUA Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine	A
	AUG Methionine	ACG Threonine	AAG Lysine	AGG Arginine	G
G	GUU Valine	GCU Alanine	GAU Aspartic Acid	GGU Glycine	U
	GUC Valine	GCC Alanine	GAC Aspartic Acid	GGC Glycine	C
	GUA Valine	GCA Alanine	GAA Glutamic Acid	GGA Glycine	A
	GUG Valine	GCG Alanine	GAG Glutamic Acid	GGG Glycine	G

1.4 Exercise

Fill in the missing parts of the diagram.



1.5 Exercise Solution



1.6 Gene Finding Game

Find the following short peptides in the DNA sequence given below.

Students may work in groups, and the first group to find all the proteins can have a little prize. A representative can write the answers on the blackboard, with the number of the nucleotide in the DNA sequence that the proteins are in, for example “230-240”.

- Met Ser Ile Leu Leu Tyr Stop
- Met Ser Ile Leu Val Glu Arg Stop
- Met Cys Arg Thr Stop

1-50	TGGTCCTGCA GTCCTCTCCT GGCGCCCCGG GGGCGAGCGG ATGTCGATTC
51-100	TCGTGGAAAG ATAGTCCCGC TGCCTGCGGG CGGAGGGACC GTGCTGACCA
101-150	AGATGTACCC GCGCGGCAAC CACTGGGCGG TGGGGCACTT AATGGGGAAA
151-200	AAGAGCACAG GGGAGTCTTC TTCTGTATGT TCTGAGAGAG GGAGCCTGAA
201-250	GCAGCAGCTG AGAGATGTGT CGAACGTGAA GTACATCAGG TGGGAAGAAG
251-300	CTGCAAGGAA TTTGCTGGGT CTCATAGAAG CTAAAGGAGA ACAGAAACCA
301-350	CCAGCCACCT CAACCCAAGA TGTCGATTCT ACTTTATTAA GCCCTGGGCA
351-400	ATCAGCAGCC TTCGTGGGAT ATGTCAGAGG ATAGCAGCCA ACTGAATAGC

1.7 Gene Finding Game Solution

1-50 TGGTCCTGCA GTCCTCTCCT GGCGCCCCGG GGGCGAGCGG **ATGTCGATTC**

51-100 **TCGTGGAAAG ATAG**TCCCGC TGCCTGCGGG CGGAGGGACC GTGCTGACCA

101-150 AGATGTACCC GCGCGGCAAC CACTGGGCGG TGGGGCACTT AATGGGGAAA

151-200 AAGAGCACAG GGGAGTCTTC TTCTGTATGT TCTGAGAGAG GGAGCCTGAA

201-250 GCAGCAGCTG AGAG**ATGTGC CGAACGTGA**A GTACATCAGG TGGGAAGAAG

251-300 CTGCAAGGAA TTTGCTGGGT CTCATAGAAG CTAAAGGAGA ACAGAAACCA

301-350 CCAGCCACCT CAACCCAAG**A TGTCGATTCT ACTTTATTAA** GCCCTGGGCA

351-400 ATCAGCAGCC TTCGTGGGAT ATGTCAGAGG ATAGCAGCCA ACTGAATAGC

ATG TCG ATT CTC GTG GAA AGA TAG

Met Ser Ile Leu Val Glu Arg Stop

(DNA region: 41 – 64)

ATG TGC GCA ACG TGA

Met Cys Arg Thr Stop

(DNA region: 215 – 229)

ATG TCG ATT CTA CTT TAT TAA

Met Ser Ile Leu Leu Tyr Stop

(DNA region: 320 – 340)

1.8 Computer Activity

Go to website:

http://www.expasy.org/tools/dna.html

Paste the DNA sequence into the dialog box. You will find 6 different ways the DNA can be 'translated' depending on the direction the sequence is read (if it's read from 5'->3' or 3'->5', and the frame.

Note you have to use the single letter code now. The proteins, in single letter code, spell "SILVER" "CAT" and "SILLY". They are in the 5'->3' direction, in frame 2.

This activity aims to give an idea of the difficulty of finding proteins in the DNA, and how computers can make the task much easier.

1.9 Questions

1. Which amino acid does AGG code for?

Answer: arginine (arg)

2. Which codons code for Serine?

Answer: UCU, UCC, UCA, UCG, AGU, AGC

3. How many ways can 'Cys Arg Thr' be coded?

*Answer: 2*6*4 = 48 ways*

4. What is the 'antisense' strand of DNA?

Answer: The strand from which mRNA is transcribed. The 'sense' strand corresponds to the mRNA product, with the T's replaced with U's.

5. If you look at the table of codons, you see that the last nucleotide in the codon often doesn't matter, for what amino acid is coded. What consequences may this have?

Answer: Redundancy gives flexibility. Mutations in the last codon => silent mutations. To be discussed in Lesson 2.

2 Mutations

Title: Point Mutations and their Consequences
Time: 40 min
Material: nucleic acid + amino acid cards
Aim: Understand the effects of mutations – insertions, deletions, substitutions, frame shifts, and silent mutations.

2.1 Introduction

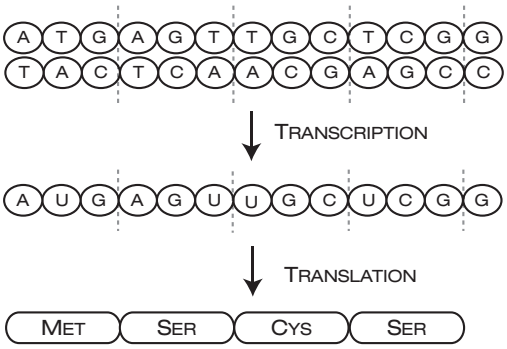
Point Mutations: An error in a single site in the DNA.

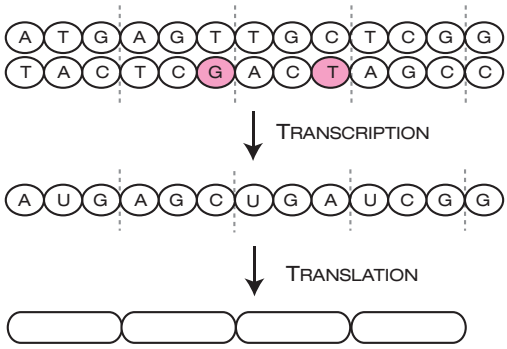
May be:

- substitution: a nucleotide is replaced,
- insertion or deletion: cause a frame shift
- A point mutation at the last place of the codon often does not make a difference, due to the redundancy of the genetic code => a ‚Silent' mutation.
- Mutations can create new stop codons, to give truncated proteins.

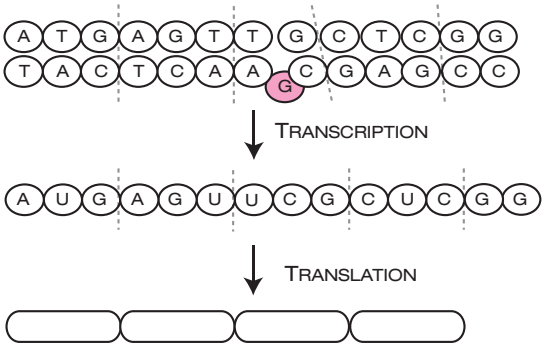
2.2 Exercise

Fill in the blanks in the diagram below.

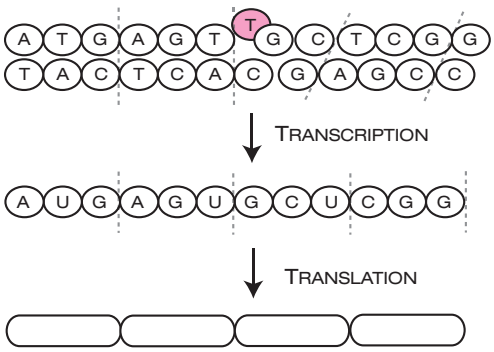




WHAT KIND OF MUTATION?
RESULT:

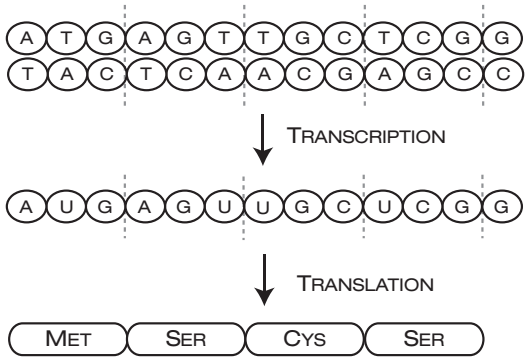


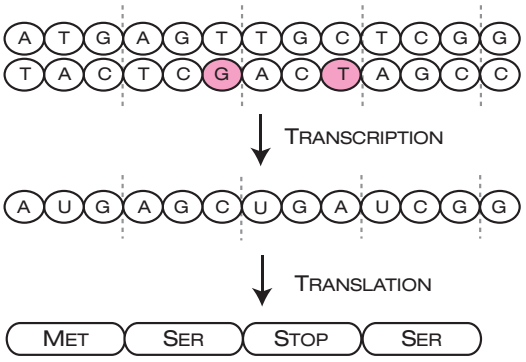
WHAT KIND OF MUTATION?
RESULT:



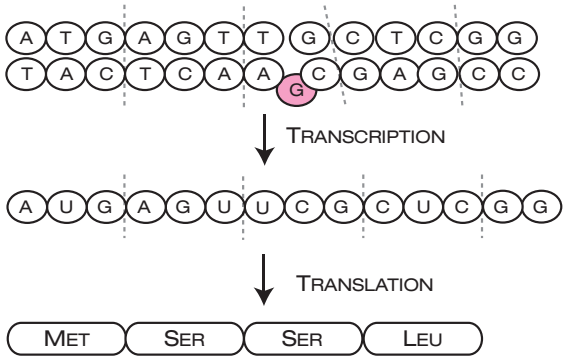
WHAT KIND OF MUTATION?
RESULT:

2.3 Exercise Solution

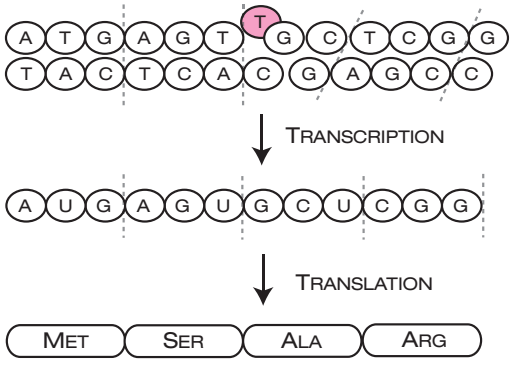




SUBSTITUTION
2 MUTATIONS:
1 SILENT
1 CREATED A STOP CODON



INSERTION
FRAMESHIFT



DELETION
FRAMESHIFT

2.4 Mutation Game

This is a fun way to learn about the consequences of mutations.

Cut out the “Nucleic Acid Cards” and “Amino Acid Cards” in Appendix I and II.

Rules:

Play in groups of 3-4.

1. Cut out the nucleic acid and amino acid cards from the student handout. Shuffle the nucleic acid cards, and spread 15 on the table. For example:

AGUAGAUCGAAUACU

This is a sequence of mRNA.

2. Place the start card, AUG, in front of the sequence, to define the frame.

AUG AGUAGAUCGAAUACU

Leave the rest of the nucleic acids in a pile to the side.

3. Now distribute 5 amino acids to each player. They can be open – other players should know which amino acids you have.

4. The first player takes any single nucleic acid from the pile. Now put the nucleic acid somewhere in the sequence. You can either create a point mutation, by replacing a nucleotide, or create a frame shift. You may also create a deletion, without using the nucleotide from the pile. Silent mutations are allowed, but the nucleotide sequence MUST be mutated in some way.

5. After creating the mutation, place your amino acids next to a codon that codes for that amino acid. The numbers on the cards (+1, +2 etc) indicate how many points you make by placing the amino acid card. You should place as many amino acids as you can.

6. Then record the total score (the sum of the numbers on the amino acid cards) next to you name, in the supplied table. Correctly record what kind of mutation you made: Substitution, Deletion, Insertion, and specify if you have also made a frame shift, or a silent mutation. If the record is incorrect, you are penalized by one point.

7. The next player can ignore the amino acids that have been put down so far, and start again, but with the mutated sequence.

8. The game continues for 2 rounds. The person with the highest score wins.

Example:

(Demonstrate this example to the class.)

The idea is to make mutation in such a way, as to maximize the number of amino acids that you can put down. For example, suppose you have a Tyrosine (TAT, TAC) and a Lysine (AAA, AAG). With the above example,

AUG AGUAGAUCGAATACU

you cannot put either of the cards. Notice that although there is the sequence TAC, it is out of frame, so Tyrosine cannot be placed.

However, you can make an insertion, using an A card from the nucleotide pile:

AUG AGUAAGAUCGAATACU

Now you can place both his amino acids, properly in frame.

AUG AGUAAGAUCGAATACU

Since Tyrosine and Lysine are each worth 2 points, you have made 4 points. (The rarer amino acids are worth more points). You must correctly record in the table that you have made an 'insertion, which caused a frame shift'.

Note:

Remember that if a start or a stop codon if created, you must all respect the rules of translation: No amino acids may be placed above a start codon, and none can be placed downstream of a stop codon. However, players may remove the start or stop codon by creating mutations.

The teacher may join one of the groups playing the game, and record the sequence at each step of the mutation. It may be interesting to look at the change of the nucleotide sequence, and the corresponding changes in the amino acid sequence.

2.5 Questions

1. What kind of mutations will have the greatest effect on the organism?

Answer: Truncations and frame shifts (that often also result in truncations).

2. How does the redundancy of the genetic code reduce the effects of mutations?

Answer: About one third pf substitutions are silent.

3 Phylogenetic Tree

Title: Building a phylogenetic tree
Time: 90 min
Materials: pencil and paper
Aim: Mutations explain how genes gradually change over time. Learn how this can be used to deduce evolutionary relationships between organisms.
Links:
http://www.ncbi.nlm.nih.gov/About/primer/phylo.html
http://www.genebee.msu.su/services/phtree_reduced.html
http://stripe.colorado.edu/~am/GeneticDistance.html
http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/T/Taxonomy.html#PhylogeneticTrees

3.1 Introduction

Think about how you would classify diverse animals. Traditionally, physical differences between organisms were used to deduce evolutionary relationships between organisms. For example, whether an organism has a 'backbone', or if it has wings. This may cause problems – for example, birds, bats, and insects all have wings, but are they closely related? How do you put a number on how recently in the past the organisms diverged? (10 min)

If you assume that mutations in the DNA occur randomly at a certain rate, that they are passed on to the organisms' offspring, and that all organisms have a common ancestor, you can use the differences in homologous sequences to measure how long it has been, since the organisms diverged.

In other words, the longer the time since two species have diverged from a common ancestor, the more different their DNA sequences will be.

Homologous sequences are defined to be those sequences in 2 organisms that have a common origin. In reality we don't really have a proof that any 2 sequences are homologous (we were not there to watch the DNA changing over time!) but if they are sufficiently similar, we often assume that they are 'homologues'. To know how similar 2 sequences are, you need to align them correctly (but this is not part of this course).

Note that different regions of the DNA –coding and non-coding regions – evolve at different speeds. In general, coding regions evolve more slowly, because a mutation in a protein is generally more 'costly' to the organism – it is less likely to survive and leave offspring. We shall discuss this more later.

To the concept of homology, you can use the example of philology – the study of the evolution of languages. In fact, there are many parallels between the methods used to study evolution of language and organisms.

Using the differences between fragments of DNA sequences, is a bit like comparing a word that means the same thing in different languages, to compare how closely they are related.

3.2 List of 'Cat' in Indo-European languages

Armenian: gatz

Basque: catua

Dutch: kat

English: cat

Estonian: kass

Finnish: kissa

French: chat

German: katze

Icelandic: kottur

Italian: gattor

Norwegian: katt

Polish: kot

Portuguese: gato

Russian: kots

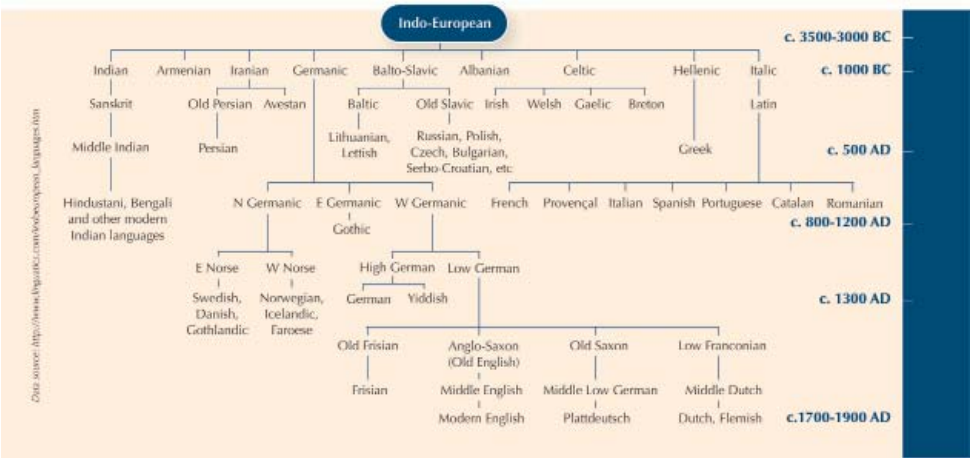
Spanish: gato

Swedish: katt

http://www.geocities.com/Heartland/Estates/2466/cat.html

You can see that Italian, Spanish, and Portuguese are almost the same: gatto, gato, and gato. Swedish and Norwegian are both 'katt' but you see that Finnish is different, 'kissa', even though it is a Scandinavian country. Finnish is more similar to Estonian, kass. In fact, the two languages are closely related. So you can learn a little bit about language relationships by studying how the words have changed over time.

3.3 Indo-European Language Tree



3.4 Drawing a phylogenetic tree of 5 Primates

Draw a phylogenetic tree, using 5 (fictional) homologous sequences from Primates.

Note that we cannot deduce any real estimates of genetic distance, as the sequences are made up by me! In any case, we would need far longer sequences to obtain a meaningful phylogenetic tree.

We will use the following sequences:

n neanderthal	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
h human	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC
c chimp	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC
g gorilla	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC
o orangutan	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

3.5 Measuring Evolutionary Distance

Count the number of differences between two sequences, and record it in the table given. The number of different nucleotides between two sequences/ total number of nucleotides in each sequence is the percentage 'distance' between the two sequences.

This is easy to do if you compare each sequence side by side.

n TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC

h TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC

differences: 3

n TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC

c TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

differences: 4

n TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC

g TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

differences: 11

n TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC

o ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

differences: 16

h TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC

c TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

differences: 5

h TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC

g TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

differences: 12

h TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC

o ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

differences: 17

c TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

g TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

differences: 11

c TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

o ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

differences: 14

g TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

o ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

differences: 14

3.6 Table of Evolutionary Distance

Table 1	Neanderthal	Human	Chimp	Gorilla	Orangutan
Neanderthal	0	3	4	11	16
Human	3	0	5	12	17
Chimp	4	5	0	11	14
Gorilla	11	12	11	0	14
Orangutan	16	17	14	14	0

(number of nucleotides = 46)

Now group the two of the most similar (least different) species together: The neanderthal and human.

Record in the 'table of evolutionary distance' that the nucleotide difference between human and neanderthal is 3, so the '% distance' is 3/46 = 0.065.

	Differences	% Difference
Neanderthal and Human	3	3/46=0.065
Neanderthal/Human and Chimpanzee		

Update the table by taking the distance from the 'average sequence' of the neanderthal and human. This 'average sequence' is assumed to be the 'ancestor' of the neanderthal and human. Therefore, we are now calculating the evolutionary distance between this ancestor, and all other primates in the group.

The differences between human and chimp is 5, the difference between neanderthal and chimp is 4. So the average distance of human/neanderthal and chimp is 4.5.

Table 2	**Neanderthal/ Human**	**Chimp**	**Gorilla**	**Orangutan**
Neanderthal/ Human	0	(4+5)/2=4.5	(11+12)/2=11.5	(16+17)/2=16.5
Chimp	(4+5)/2=4.5	0	11	14
Gorilla	(11+12)/2=11.5	11	0	14
Orangutan	(16+17)/2=16.5	14	14	0

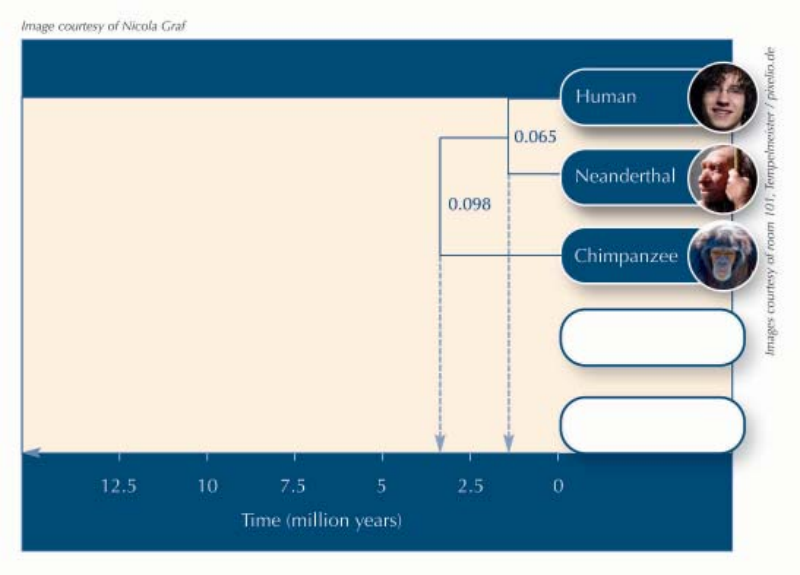
The next closest group to the neanderthal/human branch is the chimp, 4.5, compared to the gorilla (11.5) and orangutan (16.5). Record this in the table of evolutionary distance: 4.5/46 = 0.98.

You may now start drawing the evolutionary tree. Connect human and neanderthal, with a line. The branch length should correspond to how long it takes for humans and neanderthals to diverge from the common ancestor.

Let us assume that it would take 20 million years for all the nucleotides in this particular DNA region, to change 100%. Then for the DNA sequence to change by 0.065 = 6.5%, it would take 0.065*20 million = 1.3 million years. So make the branch length correspond to this time, using the time line.

To calculate how long ago the ancestor of Chimpanzees have diverged from the ancestor of humans, you must add the branch lengths. (0.065 + 0.098)* 20 million = 0.163*20 million = 3.3 million years ago.

3.7 Incomplete Phylogenetic Tree



Recalculate the entries as below.

Table 3	Neanderthal/Human/Chimp	Gorilla	Orangutan
Neanderthal/Human/Chimp	0	(11.5+11)/2=11.25	(16.5+14)/2=15.25
Gorilla	(11.5+11)/2=11.25	0	14
Orangutan	(16.5+14)/2=15.25	14	0

Record 11.25 in the table of evolutionary distance.

The next similar organism to the neanderthal/human/chimp is the gorilla.

Table 4	Neanderthal/Human/Chimp/Gorilla	Orangutan
Neanderthal/Human/Chimp/Gorilla	0	(13.75+14)/2=14.625
Orangutan	(13.75+14)/2=14.625	0

Finally, record the distance of the orangutan from all other organisms.

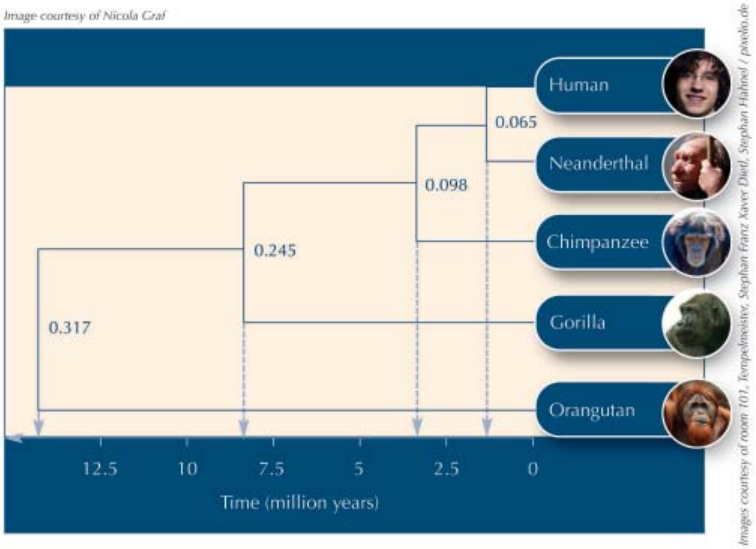
3.8 Table2 of Evolutionary Distance

	Differences	% Difference
Neanderthal and Human	3	3/46=0.065
Neanderthal/Human and Chimp	4.5	9.8
Neanderthal/Human/Chimp and Gorilla	11.25	24.5
Neanderthal/Human/Chimp/Gorilla and Orangutan	14.625	31.7

Now use this table to finish the phylogenetic tree, as shown below. Each number at the node represents the genetic distance between the organism, and the group that it is connected to.

If you know how long it takes on for a single mutation to occur in the sequence you are using to compare, you can calculate how long it has been, since the organisms diverged from the same ancestor.

3.9 Complete Phylogenetic Tree



3.10 Questions

1. How many years ago did gorillas and humans diverge from a common ancestor? How about orangutans and humans?

Gorilla: (0.065 + 0.098+0.245) 20 million = 0.408*20 million = 8.16 million years ago*

*Orangutan: (0.065+0.098+0.245+0.302) *20 million = 0.710* 20 million = 14.2 million years ago.*

2. Why may trees built with different regions of the DNA look different?

Answer: Different parts of the genome diverge at different speeds.

It may help to think about the language example again. Consider the word 'hello'.

Spanish: Hola

French: Bonjour

Italian: Buon giorno

Polish: Czesc

Russian: Zdravstvuite

The polish and Russian, which are very similar languages, are now more different. Similarly, some parts of the genome change at different speeds.

An extreme example of this is a region in the human genome, called the 'polymorphic region'. This DNA in this region changes so fast, that it is used for forensic science and paternity testing. Instead of an evolutionary tree, you can build a family tree by comparing DNA sequences in this region!

3. What regions of DNA should you use to compare organisms that are closely related?

Answer: Those regions that diverge quickly = normally 'non-essential'.

4. What kind of genes should you use to compare organisms which are evolutionarily distant from each other?

Answer: Essential genes that code for important proteins. These regions change very slowly, because any mutation is likely to have a negative effect on the survival rate of the organism, so that it is less likely to have offspring that inherit the change. One essential gene that can be used to compare distant organisms is cytochrome C.

5. What should you do if you are comparing two sequences, but one of them has gaps, due to insertions/ deletions?

Answer: Gaps can be given 'penalty' values when calculating genetic distances, but aligning gapped sequences is a whole course in itself!

6. Can you think of reasons why this method of simply comparing the number of differences between the nucleotides cannot work, if you are comparing organisms that are very different? Think about the fact that we are assuming it takes 20 million years for every nucleotide in a sequence to mutate.

Answer: The model also breaks down if the sequences have diverged long enough for on average, because more than one mutation to have occurred in the same site. The more distant the organisms, the more you will underestimate their distance.

7. Can you think of other reasons why it may not be so good to use this method to calculate genetic distances? What simplifications have we made?

Answers: For example: It also does not take account of the fact that purines (A or G) are more likely to replace purines, pyrimidines for pyrimidines (T or C). Sometimes genes are duplicated, so that the selective pressure on one copy is removed (discussed in the next lesson). There are events that can change the rate of mutations (for example selective pressure from the environment).

8. Can you think of reasons why if you are studying more distant organisms, it is better to compare amino acid sequences than DNA sequences?

Answer: Looking at the amino acid sequences gives a functional meaning to the mutations, that DNA sequences do not. For example, silent mutations in the DNA will make no difference to the amino acids; it follows that the third codon in the DNA is more likely to be mutated than the others!!

4 Mobile DNA

Title: Mobile DNA
Time: 40 min
Material: pencil, paper, dice
Aim: Why point mutations don't explain everything. Mobile DNA

4.1 Introduction

Transposons are mobile DNA, that can cut or copy themselves from the genome, and paste themselves into a different region. They can replicate themselves using the cell's DNA replication material.

How can transposons be useful to evolution? During rare events, they can excise themselves incorrectly, carrying with them a useful, neighboring gene. They can also interrupt other genes, for example by causing insertions and frame shifts

4.2 Transposon Game

Take a cookbook recipe in your language, of 36 words, and fit it into a 6*6 table.

Suppose you have a transposon, a mobile element of the sentence which can be a common word, such as "and". This transposon can hop to another location in the sentence, copying a neighboring word with it, to a new location in the sentence.

Roll the dice twice. The first number tells you which line, the second number tells you which word in the line you transpose to (this is a replicative transposition – you copy the transposon and also the neighboring word). If the sum of the numbers is even, choose the word on the left. If the sum is odd, choose the word on the right, to transpose together with the word 'and'.

Try this 4 times, for each 'and' in the sentence. Here is an example...

Delicious	and (1)	Fun	Christmas	Cake	Recipe
Preheat	the	oven	to	300	Degrees.
Cream	together	the	butter	and (2)	sugar.
Add	Eggs,	Flour	and (3)	spices	mixing
well.	Fill	buttered	pan	with	batter,
and (4)	bake	for	about	1	hour.

After transposition:

Deli-cious	and (1)	Fun	Christ-mas	Cake	and (2)	sugar	Recipe		
Pre-heat	the	oven	to	300	De-grees				
Cream	batter	and (4)	to-gether	the	butter	and (2)	sugar		
Add	Eggs,	Flour	and (3)	spices	mixing				
well	Fill	Flour	and (3)	but-tered	pan	and (1)	fun	with	batter
and (4)	bake	for	about	1	hour.				

4.3 Questions

1. Have the 'transposons' destroyed the meaning of sentences, or added something new?

Answer: It depends, but probably there will be some parts of the recipe that still make sense.

2. Do you think you could get a new kind of cake this way? What is the parallel to evolution?

Answer: This is one way organisms can evolve, much more quickly than by accumulating point mutations. Also, if there are two copies of a gene, one gene can start mutating and acquire a new function, without selective pressure, as there is a spare copy of the gene that carries out the original gene function.

3. What kind of problems could transposition cause when comparing sequences?

Answer: If there is more than one copy of similar genes in the organisms you are comparing, you will not know which you should compare to construct the tree. It is important to know that there are many difficulties in comparing sequences.

Appendix I: Amino Acid Cards

Cut along the black lines to use the amino acid cards for Lesson 2.

Ala Alanine GCU, GCC, GCA, GCG **+1**	**Arg** Arginine CGU, CGC, CGA, CGG, AGA, AGG **+1**
Cys Cysteine UGU, UGC **+3**	**Glu** Glutamic Acid GAA, GAG **+2**
Ile Isoleucine AUU, AUC, AUA **+1**	**Leu** Leucine UUG, UUA, CUU, CUC, CUA, CUG +1
Asn Asparagine AAU, AAC **+2**	**Ser** Serine UCU, UCC, UCA, UCG, AGU, AGC **+1**
Gly Glycine GGU, GGC, GGA, GGG **+1**	**Val** Valine GUU, GUC, GUA, GUG **+1**
Lys Lysine AAA, AAG **+2**	**Met** Methionine (start codon) AUG **+3**

Phe Phenylalanine UUU, UUC **+2**	**Pro** Proline CCU, CCC, CCA, CCG **+1**
Trp Tryptophan UGG **+3**	**Tyr** Tyrosine UAU, UAC **+2**
Asp Aspartic Acid GAU, GAC **+2**	**His** Histidine CAU, CAC **+2**
End Terminator UAA, UAG, UGA **+1**	**Thr** Threonine ACU, ACC, ACA, ACG **+1**
A U G **(Start)**	

Appendix II: Nucleic Acid Cards

Cut along the black lines to use the amino acid cards for Lesson 2.

A	G	U	C	A	G
U	C	A	G	U	C
A	G	U	C	A	G
U	C	A	G	U	C
A	G	U	C	A	G
U	C	A	G	U	C
A	G	U	C	A	G
U	C	A	G	U	C
A	G	U	C	A	G
U	C	A	G	U	C

Acknowledgements

For suggestions offered during the preparation of this activity, and for critical reading of the draft, I thank Alexandra Manaia and Julia Willingale-Theune.

The cover image by Petra Riedinger;

The Indo-European language tree and phylogenetic trees by Nicola Graf;

Layout design by Nicola Graf;

Edited by Corinne Kox.

ELLS employs creative commons copyrights to protect material produced for ELLS LLABs which will subsequently be used by teachers and other institutions. The copyright symbols also appear on the ELLS TeachingBASE website and in the downloadable pdfs/docs/ppts.

Attribution Non-commercial Share Alike

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms. Others can download and redistribute your work just like the by-nc-nd license, but they can also translate, make remixes, and produce new stories based on your work. All new work based on yours will carry the same license, so any derivatives will also be non-commercial in nature.

Furthermore, the author of the derivative work may not imply that the derivative work is endorsed or approved by the author of the original work.

Explanation of the copyright symbols

 to Share

 to Remix

 Attribution

 Noncommercial

 Share alike

For further details, see http://creativecommons.org