BIOL312

Practical 9

Bioinformatics with pen and paper: Building a phylogenetic tree

This exercise is modified from the online version developed for educational use by Cleopatra Kozlowski, obtained from The European Learning Laboratory for the Life Sciences (ELLS), issue 17. You can access online from http://www.scienceinschool.org/2010/issue17/bioinformatics.

Introduction

Think about how you would classify diverse animals. Traditionally, physical differences between organisms were used to deduce evolutionary relationships between them, for example, whether an organism has a backbone, or if it has wings. This may cause problems, however. For example, birds, bats and insects all have wings, but are they closely related? How do you measure how recently the organisms diverged from a common ancestor?

We know from DNA sequencing studies that DNA mutations occur randomly at a very slow rate and are passed from parents to offspring. Thus, if you assume 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 diverged from a common ancestor, the more different their DNA sequences will be.

Homologous sequences are defined as those sequences in two organisms that have a common origin. In reality we don't really have proof that any two 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 two sequences are, you need to *align* them correctly (but this is not part of this activity).

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 that causes a change in a protein is generally more costly to the organism – it is less likely to survive and leave offspring.

Constructing a phylogenetic tree of primates

In this activity, we will construct a phylogenetic tree using five homologous DNA sequences from primates. Because the sequences have been made up, we cannot deduce any real estimates of genetic distance; to create a meaningful phylogenetic tree from real data would require far longer sequences. Nonetheless, the fictional sequences (in Table 2) have been chosen to give a reasonably accurate picture of primate relationships. Image courtesy of Stephan Franz Xaver Dietl / pixelio.de

Table 2: Five DNA sequence from primates

Primate	Sequence
Neanderthal (n)	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
Human (h)	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC
Chimpanzee (c)	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC
Gorilla (g)	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC
Orangutan (o)	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

1. Count the number of differences between each pair of sequences, and record it in Table 4. This is easy to do if you compare each sequence side by side. For example, Neanderthals and humans differ at three nucleotides in the sequence (Table 3a).

Table 3a: A comparison of Neanderthal and human sequences

n	TGGTCCTGC**A**GTCCTCTCCTGGCGCCC**C**GGGCGCGAGCGG**T**TGTCC
h	TGGTCCTGC**T**GTCCTCTCCTGGCGCCC**T**GGGCGCGAGCGG**A**TGTCC

Differences: 3

Table 3b: A comparison of Neanderthal and chimpanzee sequences

n	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
c	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

Differences: 4

Table 3c: A comparison of Neanderthal and gorilla sequences

n	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
g	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

Differences: 11

Table 3d: A comparison of Neanderthal and orangutan sequences

n	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
o	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differences: 16

Table 3e: A comparison of human and chimpanzee sequences

h	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC
c	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC

Differences: 5

Table 3f: A comparison of human and gorilla sequences

h	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC
g	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

Differences: 12

Table 3g: A comparison of human and orangutan sequences

h	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC
o	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differences: 17

Table 3h: A comparison of chimpanzee and gorilla sequences

c	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC
g	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

Differences: 11

Table 3i: A comparison of chimpanzee and orangutan sequences

c	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC
o	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differences: 14

Table 3j: A comparison of gorilla and orangutan sequences

g	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC
o	ACAACCTGCACTCCTATTCTGCCGAGCCGGGCGCGTGGCAAAGTCC

Differences: 14

Table 4: Sequence differences between primates

	Neanderthal	Human	Chimpanzee	Gorilla	Orangutan
Neanderthal	0	3	4	11	16
Human	3	0	5	12	17
Chimpanzee	4	5	0	11	14
Gorilla	11	12	11	0	14
Orangutan	16	17	14	14	0

The number of nucleotide differences between two sequences divided by the total number of nucleotides in each sequence (in this case, 46) gives the proportional distance between the two sequences.

> 2. Consider the two species with the most similar sequences: Neanderthal and human. In Table 5, record the number of nucleotide differences (3) and the proportional difference (3/46 = 0.065).

Table 5: Evolutionary distances between primate ancestors and primates

	Differences	Proportional difference
Neanderthal and human	3	3/46 = 0.065
Neanderthal/human and chimpanzee	4.5	4.5/46= 0.097
Neanderthal/human/chimpanzee and gorilla	11.25	11.25/46= 0.244
Neanderthal/human/chimpanzee/gorilla and orangutan	14.625	14.625/46= 0.317

The 'average sequence' of two species is assumed to be their ancestor. In this exercise, we do not directly calculate the average sequence of, for example, Neanderthals and humans, but the evolutionary distance between the Neanderthal / human ancestor, and all other primates in the group. **To fill the rest of the table 5 first complete table 6a-6c.**

3. Calculate the distance between the average sequence of the Neanderthals and humans, and the other primate species and enter the data in Table 6a.

Example: There are four differences between Neanderthal, and chimpanzee and five differences between human and chimpanzee. Thus the average distance between Neanderthal / human and chimpanzee is 4.5.

Table 6a: Sequence differences between the Neanderthal/human ancestor and other primates

	Neanderthal/human	Chimpanzee	Gorilla	Orangutan
Neanderthal /human	0	(4+5)/2 = 4.5	(11+12)/2= 11.5	(16+17)/2= 16.5
Chimpanzee	(4+5)/2 = 4.5	0	11	14
Gorilla	11.5	11	0	14
Orangutan	16.5	14	14	0

Table 6b: Sequence differences between the Neanderthal/human/chimpanzee ancestor and other primates

To calculate the sequence differences between the Neanderthal/human/chimpanzee ancestor and gorillas, average the sequence differences between the Neanderthal/human ancestor and the gorilla (11.5, see Table 6a) and between chimpanzees and gorillas (11, see Table 4).

	Neanderthal/human/chimpanzee	Gorilla	Orangutan
Neanderthal /human/chimpanzee	0	(11.5+11)/2=11.25	(16.5+14)/2= 15.25
Gorilla	11.25	0	14
Orangutan	15.25	14	0

Table 6c: Sequence differences between the Neanderthal/human/chimpanzee/gorilla ancestor and other primates

	Neanderthal/human/chimpanzee /gorilla	Orangutan
Neanderthal /human/chimpanzee /gorilla	0	(15.25+14)/2=14.625
Orangutan	14.625	0

4. As before, these distances can be turned into proportional differences by dividing by the number of nucleotides in each sequence (46). Calculate the proportional distances between the average sequence of the Neanderthals / humans, and the other primate species. Enter the figures in Table 5.

For chimpanzees, the proportional distance from the Neanderthal / human ancestor is 4.5/46 = 0.98. Using Table 5, you can begin to construct the evolutionary tree.

> 5. Connect Neanderthals and humans with a line. The branch length should correspond to how long it took for humans and Neanderthals to diverge from their common ancestor.
>
> Let us assume that it would take 20 million years for every single nucleotide in this particular DNA sequence to change. Thus for the DNA sequence to change by 0.065, it would take 0.065*20 million = 1.3 million years. The branch should, therefore, measure 1.3 million years on the time scale (see Figure 2).
>
> 6. To calculate how long ago the ancestor of chimpanzees diverged from the ancestor of humans (the branch length), add up the proportional differences in Table 5.
>
> Remember that the proportional distance between the Neanertha / human ancestor and the chimpanzee was 0.98. Thus the time since chimpanzees, humans and Neanderthals diverged from a common ancestor is:
>
> (0.065 + 0.098) * 20 million
>
> = 0.163 * 20 million
>
> = 3.3 million years ago.

(0.24+0.065+0.098)*20 = 0.403*20 = 8 million years ago

(0.24+0.065+0.098+0.32)*20= 0.72*20= 14,4 million years ago

7. Continue the calculations. Repeat steps 3 to 6 to calculate how long ago the Neanderthal / human / chimpanzee ancestor diverged from the gorilla and from the orangutan. Then calculate how long ago the Neanderthal / human / chimpanzee / gorilla ancestor diverged from the orangutan. Enter the results in Table 5.

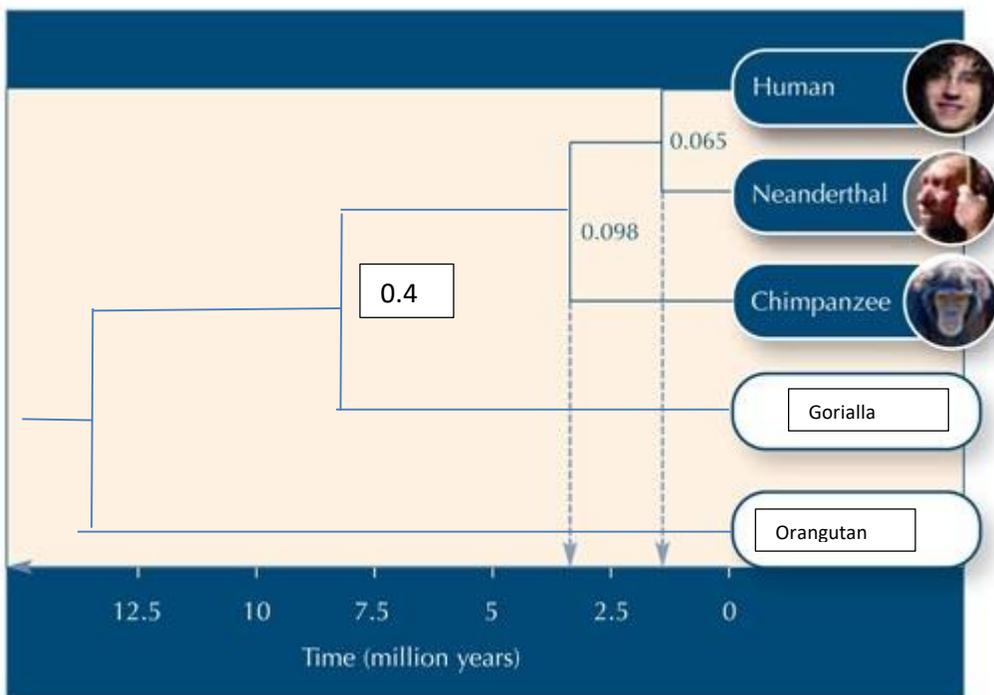**Figure 2: Incomplete phylogenetic tree**
Image courtesy of Nicola Graf