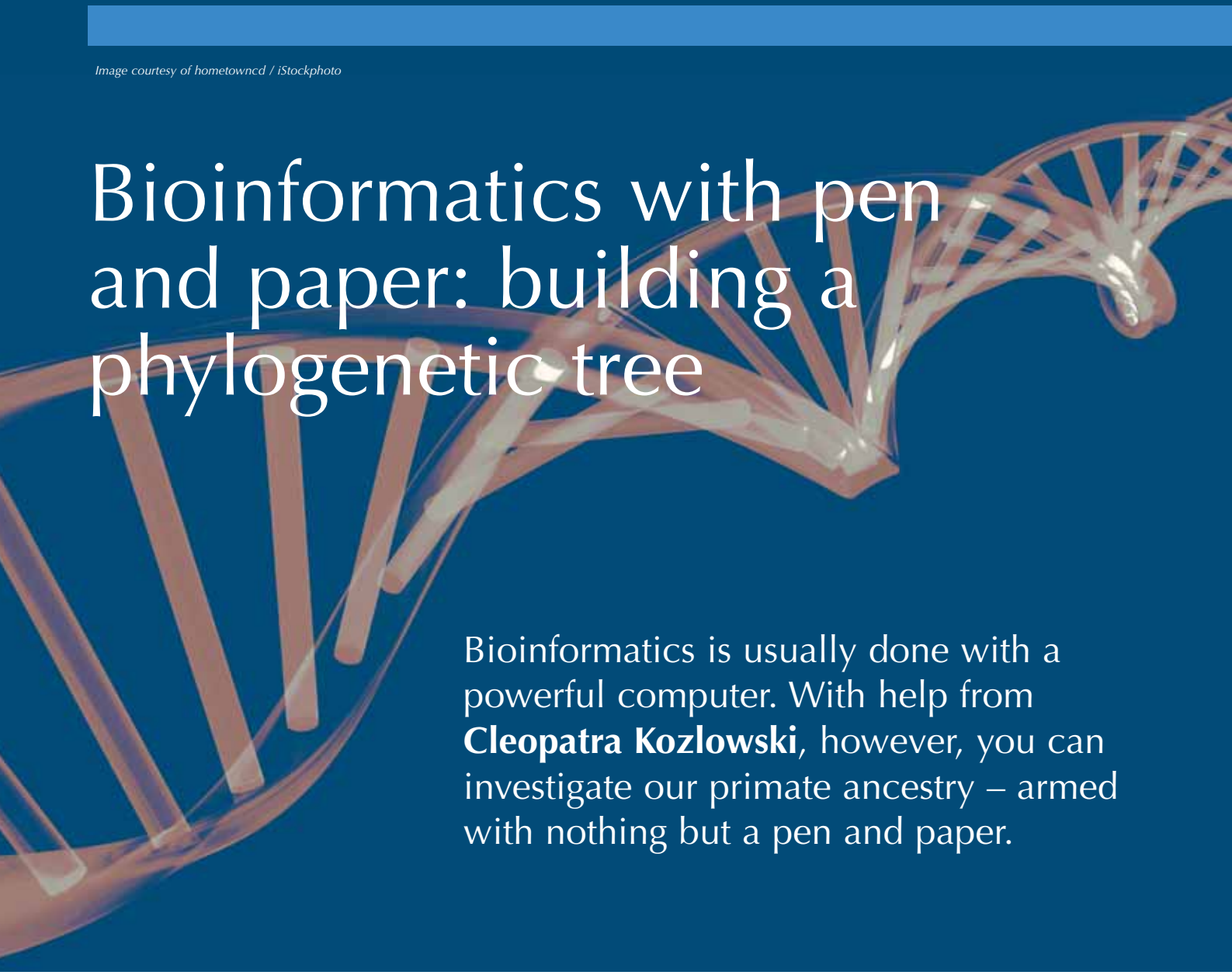


Image courtesy of hometowncd / iStockphoto

Bioinformatics with pen and paper: building a phylogenetic tree

Bioinformatics is usually done with a powerful computer. With help from **Cleopatra Kozlowski**, however, you can investigate our primate ancestry – armed with nothing but a pen and paper.

As a result of recent technological advances, it is relatively quick and easy to determine a DNA or protein sequence. These sequences by themselves, of course, tell us very little: GAATCCA, for example. We need to know what those sequences mean. Which proteins are encoded by that DNA sequence; does the sequence indeed encode a protein at all? What effect does a small change in the DNA sequence have on the structure of the encoded protein? What function does that protein have in the cell? And, of course, what can our DNA sequence tell us about our evolutionary history?

These and other important biological questions can be tackled with bioinformatics: essentially, by com-

REVIEW

When we think of bioinformatics we probably imagine huge computers and sequencing machines, but the methods of this new science can be presented by means of simple classroom activities to be carried out with pencil and paper, as Cleopatra Kozlowski does in this article.

The author challenges us with the building of the family tree of humans and other primates on the basis of the genetic differences between short (fake) DNA sequences. The proposed activity can be profitably (and enjoyably) exploited in secondary schools to address some tricky biology topics such as the use of molecular clocks in the study of evolution.

The article is aimed at science teachers, who will find useful comprehension exercises at the end of the text; students can also use the questions to deepen their understanding of the topic. The quoted web references provide further information and resources.

Giulia Realdon, Italy

paring DNA or protein sequences – for example, by comparing newly discovered sequences with sequences for which we already have a lot of information (perhaps they have a similar function?) or comparing similar sequences in different species.

Bioinformatics is, of course, normally done with the aid of a powerful computer. However, it is all too easy to let a computer do all the work without understanding the underlying principles involved. For this reason, these activities are designed to be done on paper, to get the students to understand *how* bioinformatic analysis works.

This article includes one of a group of four activities. The two introductory activities ('Gene finding' and 'Mutations') and the concluding activity ('Mobile DNA') can be downloaded from the website of the European Learning Laboratory for the Life Sciences (ELLS)[w1]. All the tables required for students to complete this activity, together with the step-by-step procedure and answers to the comprehension questions, can be downloaded from the *Science in School* website[w2].

Constructing a phylogenetic tree

The accumulation of mutations causes DNA sequences to change over generations. The following activity demonstrates how this can be used to deduce evolutionary relationships between organisms. It takes about 90 min and requires nothing but a pen and the tables, which can be downloaded from the *Science in School* website[w2].

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

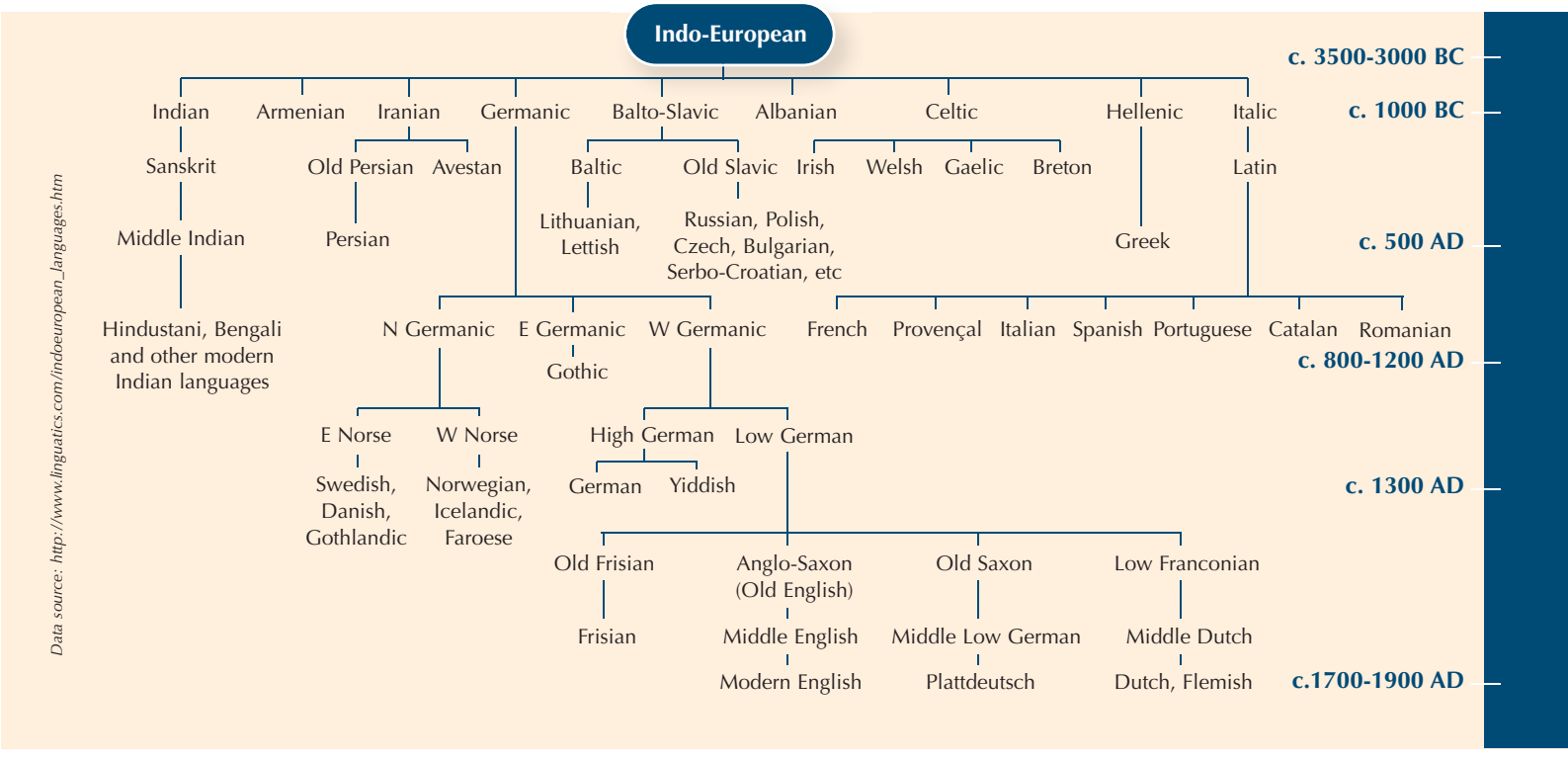


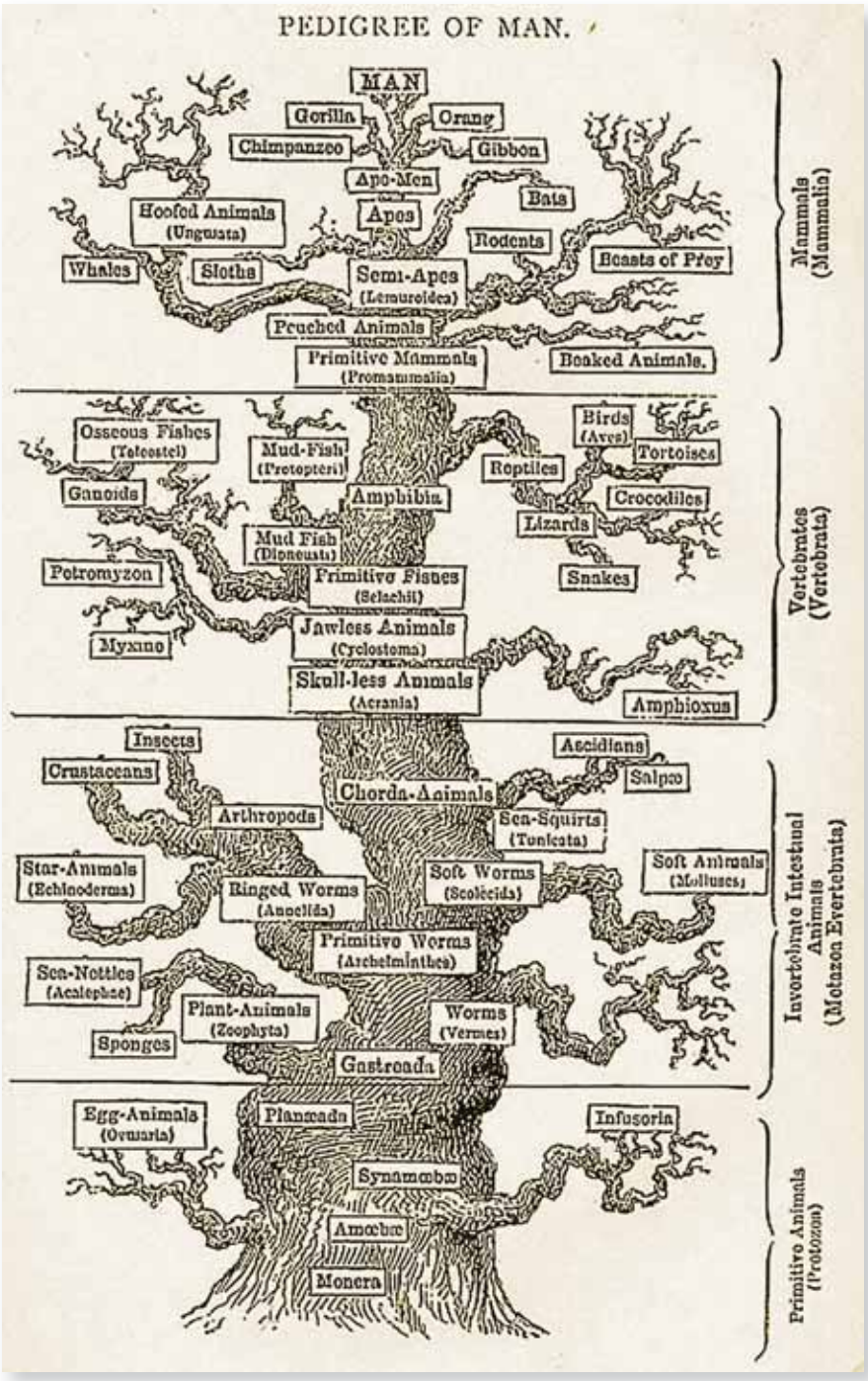
Figure 1: The Indo-European language tree. Note that although Indian, Germanic, Romance and many other European languages belong to this family, Finnish, Estonian and Hungarian do not: they belong to the Uralic language group

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. This is discussed in the 'Mobile DNA' activity.

To illustrate 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 see how closely they are related.



Haeckel's tree of life from *The Evolution of Man* (1879)

Public domain image; image source: Wikimedia Commons

Table 1: List of 'cat' in Indo-European languages

Armenian	gatz
Basque	katu
Dutch	kat
English	cat
Estonian	kass
Finnish	kissa
Icelandic	kottur
Italian	gatto
Norwegian	katt
Polish	kot
Portuguese	gato
Russian	kot
Spanish	gato
Swedish	katt

You can see that the words for 'cat' in Italian, Spanish and Portuguese are almost the same: gatto, gato and gato. In both Swedish and Norwegian, the word is 'katt' but you see that in Finnish it is different: 'kissa'. Although, like Sweden and Norway, Finland is a Nordic country, the Finnish word for 'cat' is more similar to the Estonian word, '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.

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.

Note: all the tables required for students to complete this activity can be downloaded from the *Science in School* website[w2].

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) whereas chimpanzees and gorillas differ at 11 points (Table 3b).

Comparison tables for all the pairs of species, and the completed table of sequence differences (Table 4), can be downloaded from the *Science in School* website[w2].

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).

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.

Table 2: Five DNA sequences from primates

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

Table 3a: A comparison of Neanderthal and human sequences

Neanderthal	TGGTCCTGCAGTCCTCTCCTGGCGCCCCGGGCGCGAGCGGTTGTCC
Human	TGGTCCTGCTGTCCTCTCCTGGCGCCCTGGGCGCGAGCGGATGTCC

Table 3b: A comparison of chimpanzee and gorilla sequences

Chimpanzee	TGATCCTGCAGTCCTCTTCTGGCGCCCTGGGCGCGTGCGGTTGTCC
Gorilla	TGGACCTGCAGTCATCTTCTGCCCGCCCGAGCGCTTGCCGATGTCC

Table 4: Sequence differences between primates

	Neanderthal	Human	Chimpanzee	Gorilla	Orangutan
Neanderthal	0	3			
Human	3	0			
Chimpanzee			0	11	
Gorilla			11	0	
Orangutan					0

Table 5: Evolutionary distances between primate ancestors and primates

	Differences	Proportional difference
Neanderthal and human	3	3/46 = 0.065
Neanderthal / human and chimpanzee		
Neanderthal / human / chimpanzee and gorilla		
Neanderthal / human / chimpanzee / gorilla and orangutan		

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	
Chimpanzee	(4+5)/2 = 4.5	0		
Gorilla	(11+12)/2=11.5		0	
Orangutan				0

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.

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.

There are 11 differences between Neanderthal and gorilla, and 12 differences between human and gorilla. Thus the average distance between Neanderthal/human and gorilla is 11.5.

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 Neanderthal / human ancestor and the chimpanzee was 0.98. Thus the time since chimpanzees, humans and Neanderthals diverged from a common ancestor is:

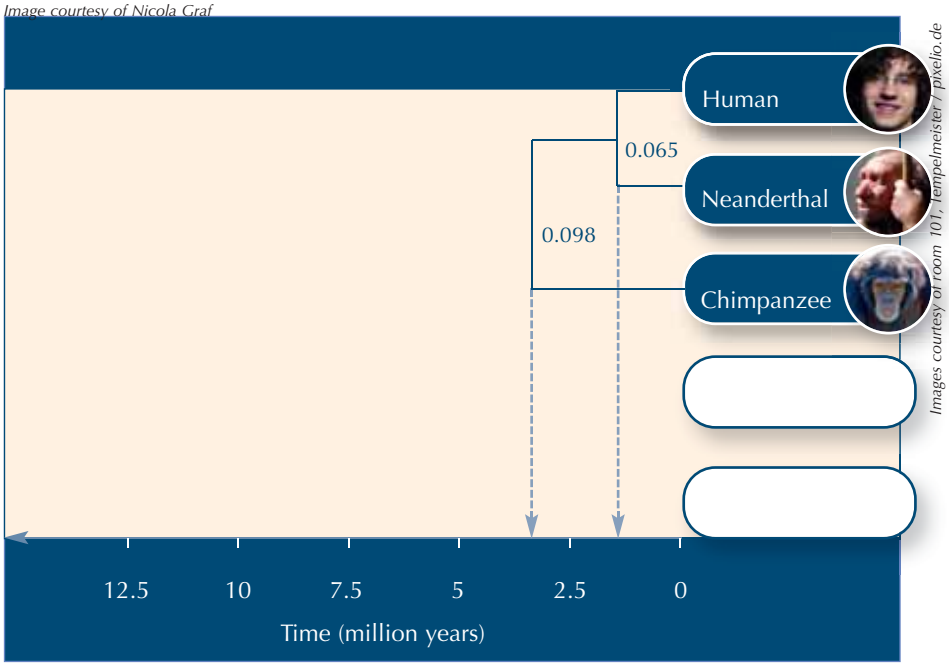
   (0.065 + 0.098) * 20 million
   = 0.163 * 20 million
   = 3.3 million years ago.

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.

If you need help, you can download the step-by-step procedure from the *Science in School* website.

8. Use the completed Table 5 to finish the phylogenetic tree, as shown on page 33.

Image courtesy of Nicola Graf



Images courtesy of room 101, lempelmeister / pixelio.de

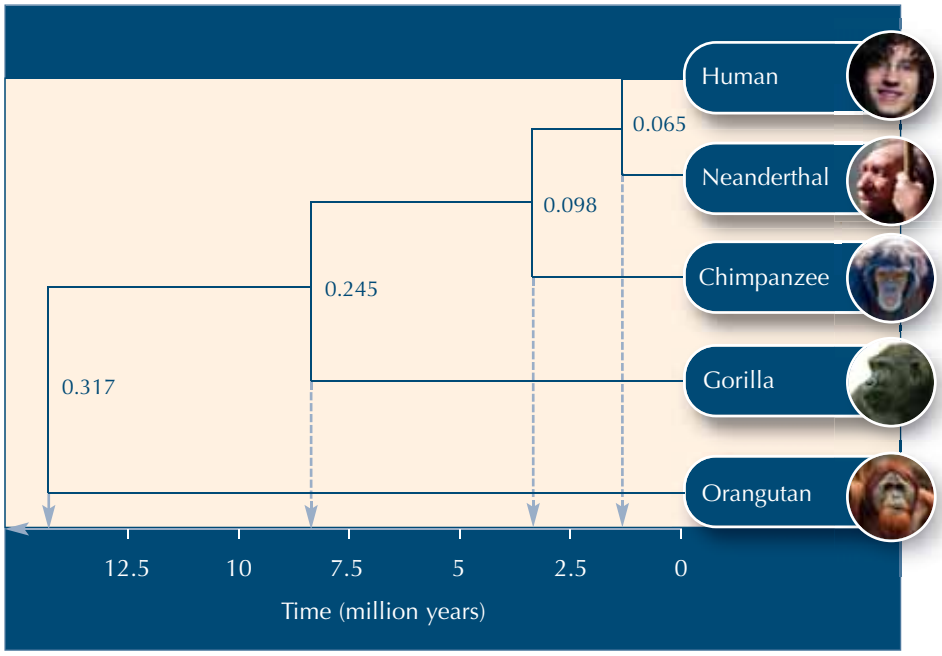
Figure 2: Incomplete phylogenetic tree

Questions

Below are some questions you could use to test your students' understanding of the activity. Answers can be downloaded from the *Science in School* website[w2].

1. In your phylogenetic tree, how many years ago did gorillas and humans diverge from a common ancestor? What about orangutans and humans?
2. Can you find out if these and the other estimates in your tree are correct?
3. Why may phylogenetic trees constructed using different regions of the DNA look different?
4. What regions of DNA should you use to compare organisms that are closely related?
5. What kind of genes should you use to compare organisms that are evolutionarily distant from each other?
6. What should you do if you are comparing two sequences, but one of them has gaps due to deletions (or insertions in the other sequence)?
7. Can you think of reasons why this method of simply comparing the number of differences between the nucleotides may not work if you are comparing organisms that are very different? Remember that we are assuming it takes 20 million years for every nucleotide in a sequence to mutate.

Image courtesy of Nicola Graf



Images courtesy of room 101, Tempelmeister, Stephan Franz Xaver Dietl, Stephan Hahnel / pixelio.de

Figure 3: Complete phylogenetic tree

8. Can you think of other reasons why it may not be so good to use this method to calculate evolutionary distances? What simplifications have we made?
9. Can you think of reasons why if you are studying more distant organisms, it is better to compare amino acid sequences than DNA sequences?
10. In this exercise, we have concentrated on working out *when* the five primate species diverged from each other (the scale of the tree). Often, however, we do not even know *the order* in which the species diverged from one another (the shape of the tree). How do we know, for example, that humans and chimpanzees are more closely related than gorillas and chimpanzees are? If the latter were true, how would the sequence differences (Table 4) differ?

Acknowledgement

This activity was developed in a special collaboration between the European Learning Laboratory for the Life Sciences (ELLS)[w1] and the European Molecular Biology Laboratory's E-STAR Fellows to develop teaching resources for schools. Cleopatra Kozlowski was supported by an E-STAR fellowship funded by the European Commission's Framework Programme 6 Marie Curie Host Fellowship for Early Stage Research Training, under contract number MEST-CT-2004-504640.

Web references

w1 – The European Learning Laboratory for the Life Sciences (ELLS) is an education facility which brings secondary-school teachers into the research lab for a unique hands-on encounter with state-of-the-art molecular biology techniques. ELLS also gives scientists a chance to work with teachers, helping to bridge the widening gap between research and schools. The activity described in this article was designed as a teaching resource for ELLS' professional development programme for European teachers. For more information about ELLS, see: www.embl.org/ells

w2 – Download all the tables required for students to complete this activity, together with the step-by-step procedure and answers to the comprehension questions, from the *Science in School* website: www.scienceinschool.org/2010/issue17/bioinformatics#resources

Resources

The website of the US National Center for Biotechnology Information (NCBI) offers an introduction to phylogenetics. See: www.ncbi.nlm.nih.gov/About/primer/phylo.html

To learn more about using protein sequences to establish phylogenetic trees, see: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages or use the direct link: http://tinyurl.com/2wqp7nq

To learn about how a group of scientists recreated the new tree of life, tracing the course of evolution, see:

Hodge R (2006) A new tree of life. *Science in School* **2**: 17-19. www.scienceinschool.org/2006/issue2/tree

The Interactive Tree Of Life is an online tool for the display and manipulation of phylogenetic trees. To learn more, see: http://itol.embl.de

To browse other evolution-related articles in *Science in School*, see: www.scienceinschool.org/evolution

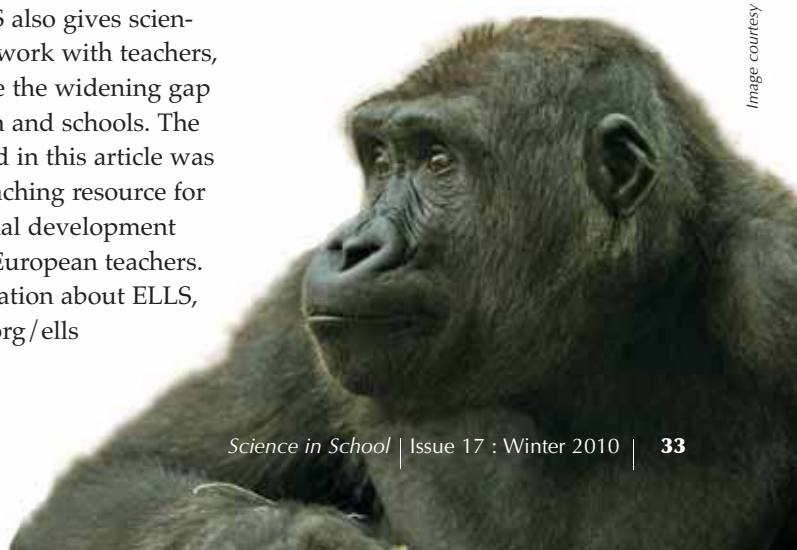


Image courtesy of Stephan Franz Xaver Dietl / pixelio.de