1. Introduction



1.1. Which of these 9 unrooted trees has a different topology than the others?



1.2. If the tree is not rooted, can we infer the order of events for the appearance of the red color and the flower shape?

2. Kungas

Mesopotamian tablets and bas-reliefs indicate that humans were already using equines for travel and warfare 4,500 years ago, while domestic horses did not appear in the Fertile Crescent until 500 years later. Some 4,500-year-old cuneiform clay tablets mention "kungas", prestigious equids of

high market value. In the 4,500-year-old princely burial complex of Umm el-Marra (northern Syria), animal bones have been found that could be these famous "kungas".



Figure 2.1. Equidae tomb in Umm-El-Marra (Syria).

DNA, although degraded, was extracted from the bones and partially sequenced. A 324bp fragment of a mitochondrial gene was obtained from 2 individuals and compared to available equine sequences.



Figure 2.2. Median-joining network obtained from the mitochondrial sequences of 278 individuals. The position of the sequences obtained from the Umm el-Marra samples is indicated by pink stars. Syrian Hemippus: subspecies of hemiones, wild Syrian donkeys that disappeared at the beginning of the 20th century, from which genomic sequences could be recovered. Asses: domestic donkeys (E. a. asinus and E. a. Somaliensis). Horses: horses (E. f. caballus and E. f. Przewalskii).

2.1. What can you infer from Figure 2.2?

A 168bp fragment of a gene carried by the Y chromosome was obtained from the same 2 individuals and compared to available equine sequences.



Figure 2.3. Median-joining network obtained from Y chromosome sequences. The position of the sequences obtained from the two samples from Umm el-Marra is shown in pink. The numbers indicate the number of mutations separating each sequence. Syrian hemippus: wild Syrian donkey.

2.2. What can you infer from Figure 2.3?

To find out more, the researchers tried to sequence as many nuclear DNA fragments as possible from the bone that contained the best quality DNA. They were able to sequence several thousand small fragments, but never the same fragment from the same region twice (which would have allowed them to find heterozygous sites). The sequences obtained were aligned with those of a hemione (wild Syrian donkey) and a domestic donkey. Of the 4738 variable sites or SNPs (Single Nucleotide Polymorphism) that can distinguish hemiones from domestic asses, half correspond to the sequence of the hemione and the other half to the sequence of the domestic donkey (Figure 9.2.4).



Figure 2.4. Proportion of SNPs detected in the sample that are identical to hemione or donkey.

2.3. What can you infer from Figure 2.4? Is this consistent with previous results?

Crosses between hemiones and domestic donkeys appear to be difficult. In Russia in the 1940s an experiment was done where female domestic donkeys were crossed with male hemiones, Turkmen kulans (*Equus hemionus kulan*) and they produced sterile offspring. Hemiones were probably very difficult to keep in pens.

2.4. Propose a scenario to explain the origin of kungas among wealthy Mesopotamian landowners.



Figure 2.5. Hemione hunting and capture.

2.5. Suggest an explanation related to the way kungas were produced that might explain why kungas were supplanted by domestic horses.



3. Amylase and starch digestion

Figure 3.1. Boxes represent all haploid amylase gene copies found in representative species or clades. Lighter colored boxes represent the variation in copy numbers found in at least two individuals of a given species or in reference genomes of at least two species within a clade. The numbers in parentheses following clade names indicate the number of species used for estimating the gene copy numbers. Amylase is an enzyme that catalyses the hydrolysis of starch into sugars. Amylase is present in the saliva of humans and some other mammals, where it begins the chemical process of digestion. The pancreas and salivary gland make amylase (alpha-amylase) to hydrolyse dietary starch into disaccharides and trisaccharides which are converted by other enzymes to glucose to supply the body with energy. Ancestrally, amylase was a pancreatic enzyme in mammals. Plants and some bacteria also produce amylase.

3.1. Based on the tree in Figure 3.1, how many copies of the amylase gene were probably present in the ancestor of Metatheria?

- 3.2. What type of mutation can lead to the gain of a new gene?
- **3.3. And to the loss of a gene?**

3.4. What phenotype is expected from an increase in the number of copies of amylase genes?



Figure 3.2. (a) Phylogenetic tree of amylase amino acid sequences. Bootstrap values are shown next to the major branch splits. (b) In human and mouse, all the amylase genes are located at the same locus. Depiction of the transposons linked with amylase copies in mouse and human genomes. Small boxes symbolize the positions of mobile elements, HERVE_a-int LTR for humans (blue) and L1Md_T for mouse (purple). The dotted arrows indicate the likely origin of derived gene duplicates.

3.5. Based on the phylogenetic tree of Fig. 3.2a, what can you conclude about the amylase genes of pigs and dogs?

3.6. Based on the phylogenetic tree of Fig. 3.2a, infer the most probable scenario for the evolution of the amylase genes in mice and rats.

In humans the gene AMY2A is deleted in some individuals.

	Number of individuals	Number of individuals	Frequency of the
Region	tested	harboring a AMY2A deletion	AMY2A deletion
Africa	36	0	0
West Asia/Caucasus	110	5	0.04
South Asia	58	1	0.02
Southwest Europe	46	5	0.11
Northeast Europe	158	16	0.1
Central Asia	48	4	0.08
Volga Ural region	44	8	0.18
West Siberia	36	10	0.28
South Siberia	68	6	0.09
Central Siberia	54	3	0.06
Northeast Siberia	46	24	0.52
South America	56	2	0.04
East Asia	52	1	0.02
Island Southeast Asia	90	1	0.01
Papua New Guinea	6	0	0

3.7. In which geographical region can we find the highest proportion of individuals carrying the AMY2A deletion? Propose two alternative hypotheses to explain why this region has the highest proportion (one related to the small population size and one related to dietary habits)?

4. Evolution of lice

Humans can be infected by three types of lice: *Pediculus humanus capitis* (head louse), *Pediculus humanus corporis* (body louse) and *Pthirus pubis* (crab louse). Primate species close to humans are infected by only one species of lice each, which is specific to each host species.

Pediculus humanus capitis and *Pediculus humanus corporis* are very similar morphologically and can still hybridize in the laboratory although they live in different parts of the human body and have slightly different DNA sequences. Body lice probably evolved from the ancestors of head lice when humans began to wear clothing, which is necessary for body lice to lay eggs.

Approximately 0.36 synonymous mutations per kilobase are detected between the DNA sequences of *Pediculus humanus capitis* and *Pediculus humanus corporis*, whereas 30 synonymous mutations per kilobase are detected between the DNA sequences of *Pediculus humanus capitis* and *Pediculus schaeffi*, the chimpanzee louse.

4.1 Knowing that chimpanzees diverged from humans about 5.5 million years ago, estimate the date of appearance of clothing in humans.

Nucleotide diversity (π) is the average number of nucleotide differences per site between two DNA sequences for all possible pairs in the sample population. It is a measure of genetic diversity. Greater nucleotide diversity is found in African head lice than in non-African head lice.

4.2. How to explain this result?



Figure 4.3. Phylogeny based on nuclear genes for mammals (right) and for lice (left). Numbers indicate estimated divergence times in millions of years from DNA sequences.

4.3. What can you infer from Figure 4.3 about the origin of crabs?

5. Reconstruction of infectious transmission chains

The SARS-CoV-2 coronavirus, responsible for COVID-19 disease, has an RNA genome of 30,000 nucleotides. This virus accumulates on average 2 mutations per month, sometimes at an accelerating rate, as with the appearance of delta or omicron variants. On January 11-12, 2022 in Hong Kong, a saleswoman (patient 1) and a customer (patient 2) of a pet store started to have COVID-19 symptoms. Analysis showed that they were both infected with the delta variant, although it was not circulating in the region (zero-covid strategy). Comparison of the viral sequences of the 2 patients revealed 5 nucleotide substitutions, at positions spaced along the genome.

5.1. Can you imagine human-to-human transmission from patient 1 to patient 2?

Samples were then taken from the animals sold in the store: rabbits, guinea pigs and hamsters. Of all the samples, about ten hamsters tested positive for the SARS-CoV-2 coronavirus. The animals present in the warehouse supplying this pet shop (rabbits, guinea pigs, hamsters, chinchillas, mice) were then also tested. Only one hamster in the warehouse tested positive. All positive animals came from the Netherlands. The complete genome sequence of SARS-CoV-2 was then obtained for each positive sample and a phylogenetic tree was made with all sequences.



Figure 5.1. Phylogenetic tree of the SARS-CoV-2 sequences obtained in this study (colored, top) and the closest SARS-CoV-2 sequences available in the databases.

5.2. What can you infer from Figure 5.1 about patients 1 and 2? 5.3. What can you deduce from Figure 5.1 regarding hamsters?

Patient 3 is the husband of patient 2. He has never been to the pet store and started developing COVID-19 symptoms on January 17, 2022.

5.4. What can you infer from Figure 5.1 about patient 3?