Practicals Biology – Virginie Courtier-Orgogozo 2021

I. Amylase evolution

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.

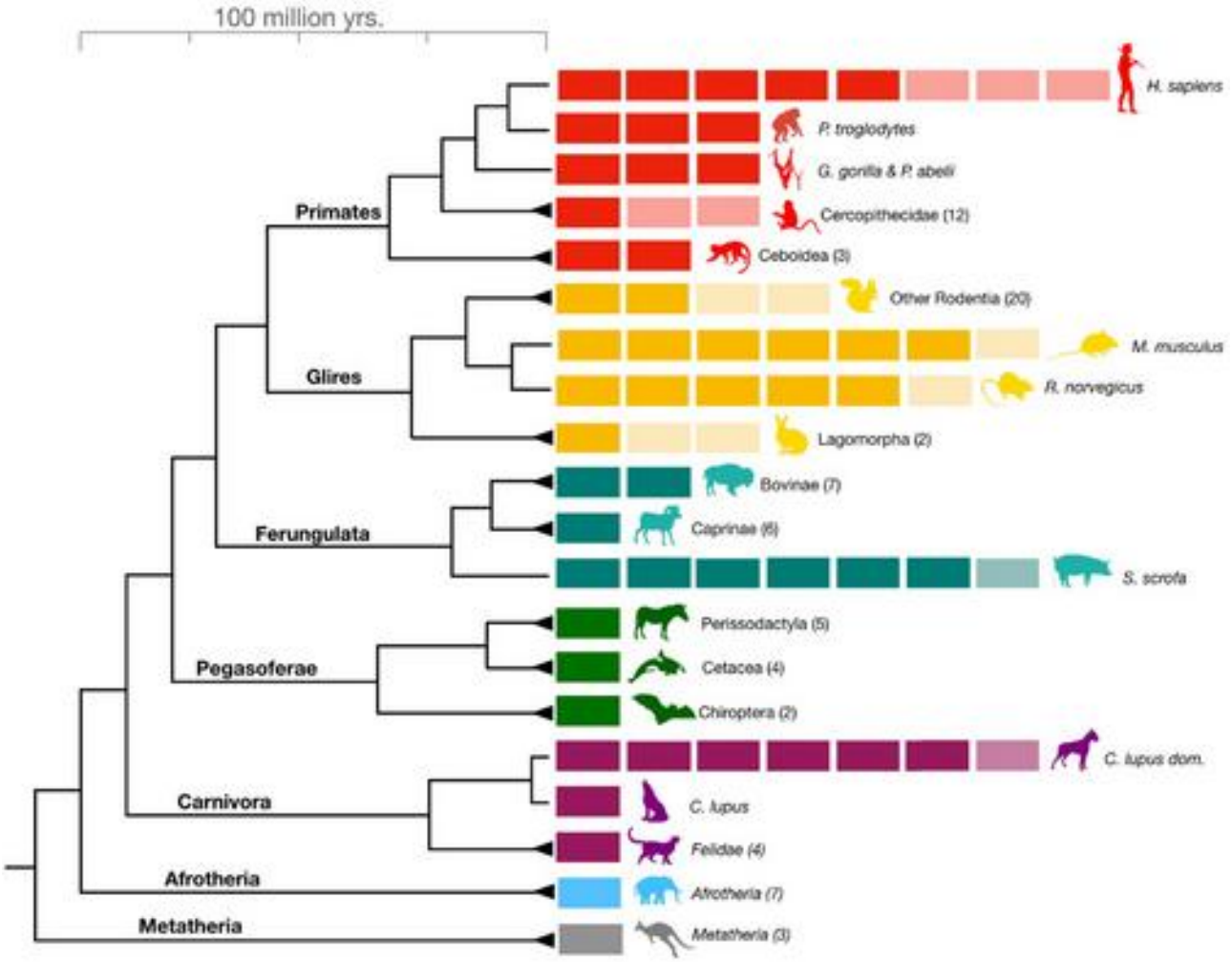


Fig. 1. Boxes represent all haploid amylase gene copies found in representative species or clades. Lighter colored boxes represent the variation in copy numbers found in at least two individuals of a given species or in reference genomes of at least two species within a clade. The numbers in parentheses following clade names indicate the number of species used for estimating the gene copy numbers.

a. Based on the tree in Figure 1, how many copies of the amylase gene were probably present in the ancestor of Metatheria?
b. What type of mutation can lead to the gain of a new gene?
c. And to the loss of a gene?
d. What phenotype is expected from an increase in the number of copies of amylase genes?

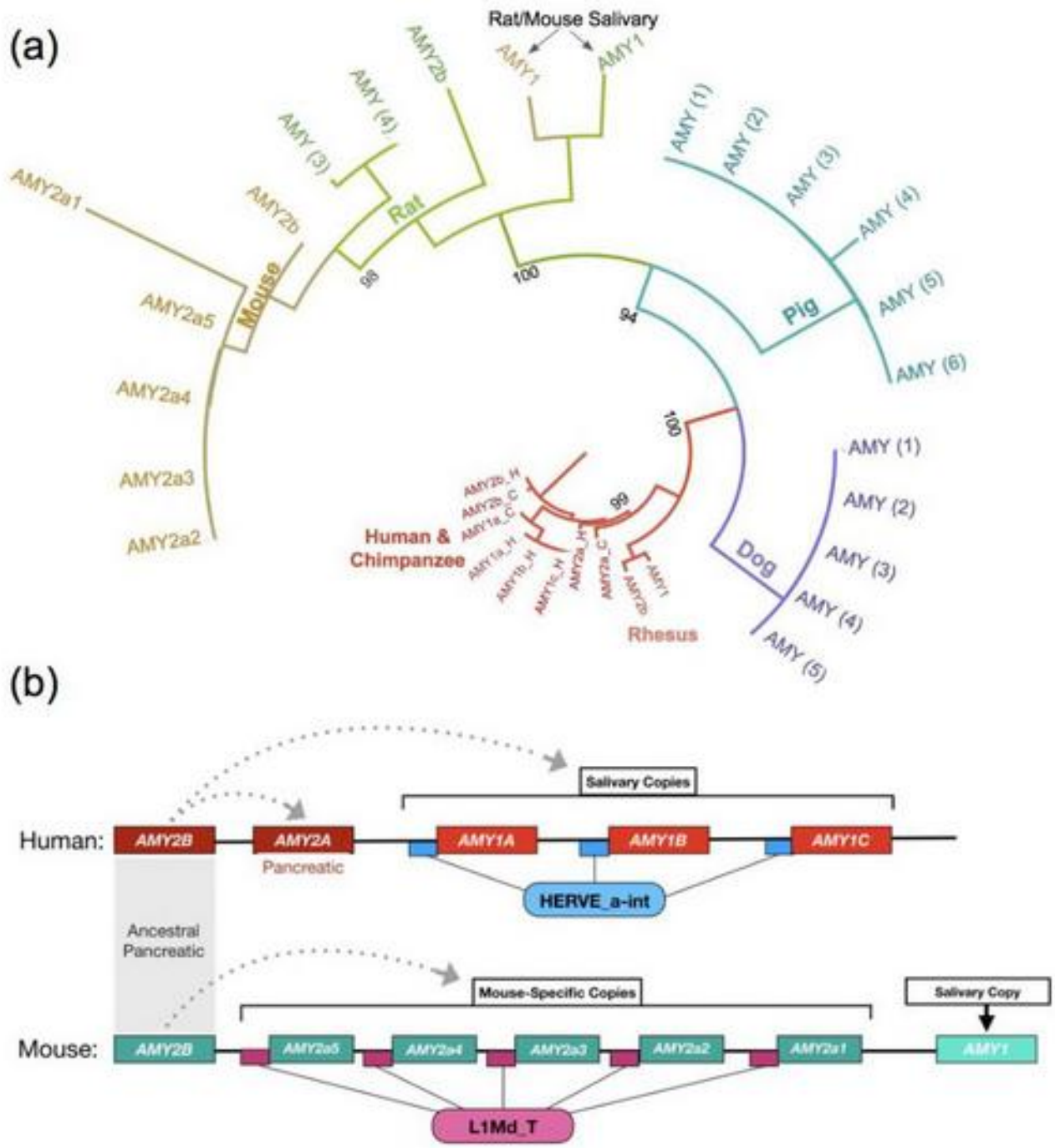


Fig. 2. (a) Phylogenetic tree of amylase amino acid sequences. Bootstrap values are shown next to the major branch splits. (b) In human and mouse, all the amylase genes are located at the same locus. Depiction of the transposons linked with amylase copies in mouse and human genomes. Small boxes symbolize 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.

e. Based on the phylogenetic tree of Fig. 2a, what can you conclude about the amylase genes of pigs and dogs?
f. Based on the phylogenetic tree of Fig. 2a, infer the most probable scenario for the evolution of the amylase genes in mice and rats.

Transposable elements are DNA sequences that can copy themselves to other positions within a genome. Fig. 2b shows that transposable element HERVE_a-int is found in three copies at the *amylase* locus in humans, each time next to an amylase coding region.

g. Provide a possible evolutionary scenario to explain the presence of these three copies of HERVE_a-int in this region of the human chromosome.

Another transposable element is found at the amylase locus in mice, L1Md_T (Fig. 2b). In rat, there is yet another transposable element present near the amylase genes and in pigs and dogs different ones also. The presence of different sets of transposable elements in each species is still not fully understood.

In humans the gene *AMY2A* is deleted in some individuals.

Region	Number of individuals tested	Number of individuals harboring a AMY2A deletion	Frequency of the 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

h. In which geographical region can we find the highest proportion of individuals carrying the AMY2A deletion?

i. 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)?

Further reading

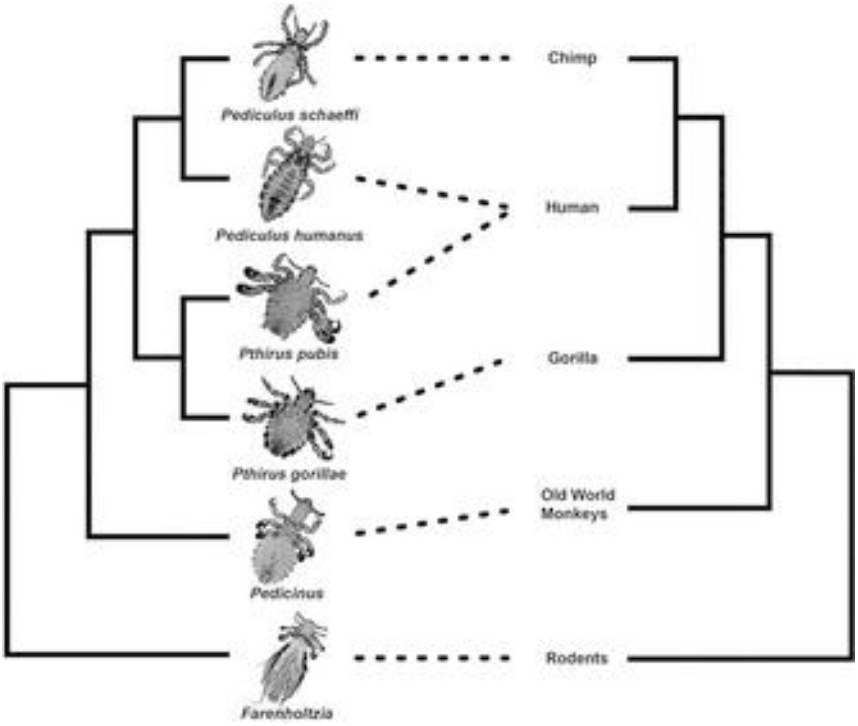
Pajic, P., Pavlidis, P., Dean, K., Neznanova, L., Romano, R. A., Garneau, D., ... & Gokcumen, O. (2019). Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife*, *8*, e44628.
https://elifesciences.org/articles/44628

Inchley, C. E., Larbey, C. D., Shwan, N. A., Pagani, L., Saag, L., Antão, T., ... & Kivisild, T. (2016). Selective sweep on human amylase genes postdates the split with Neanderthals. *Scientific reports*, *6*(1), 1-10.
https://www.nature.com/articles/srep37198

II. Evolution of lice

Humans can be infected by three types of lice: *Pediculus humanus capitis* (human head louse), *Pediculus humanus corporis* (human body louse) and *Pthirus pubis* (crab louse). Our ape relatives are infected by only one type of louse species.
Pediculus humanus capitis and *Pediculus humanus corporis* are very similar morphologically and can still hybridize in the laboratory although they live on different places on the human body and possess slightly different DNA sequences. DNA comparison suggest that the most recent common ancestor of *Pediculus humanus capitis* and *Pediculus humanus corporis* lived less than 70,000 years ago.
a. What can you infer about the evolution of clothes in humans?

Below is a phylogeny based on both nuclear and mitochondrial genes for mammals (right) and for lice (left).
b. What can you infer? (at least two scenarios are possible)



d. Use Timetree (http://www.timetree.org/**) to add time estimates on relevant nodes and find which scenario is most probable.**

III. Evolution of fruitfly glue

Juste before entering metamorphosis, Drosophila larvae produce a glue to stick themselves on natural substrates (wood, leaves, stones, bear bottles, etc.). Then the animal (the pupa) remains attached to the substrate for several days until the end of metamorphosis where the adult emerges from the pupal case. The glue is made of several proteins named Sgs1 to Sgs8 (for Salivary gland secretory protein). How these proteins confer adhesiveness is unknown.

We designed an assay to quantify adhesiveness of single pupae. Fig. 1 shows adhesion strength in various strains of *Drosophila melanogaster* (named A5, B4, A6, etc.) originating from different parts of the world.

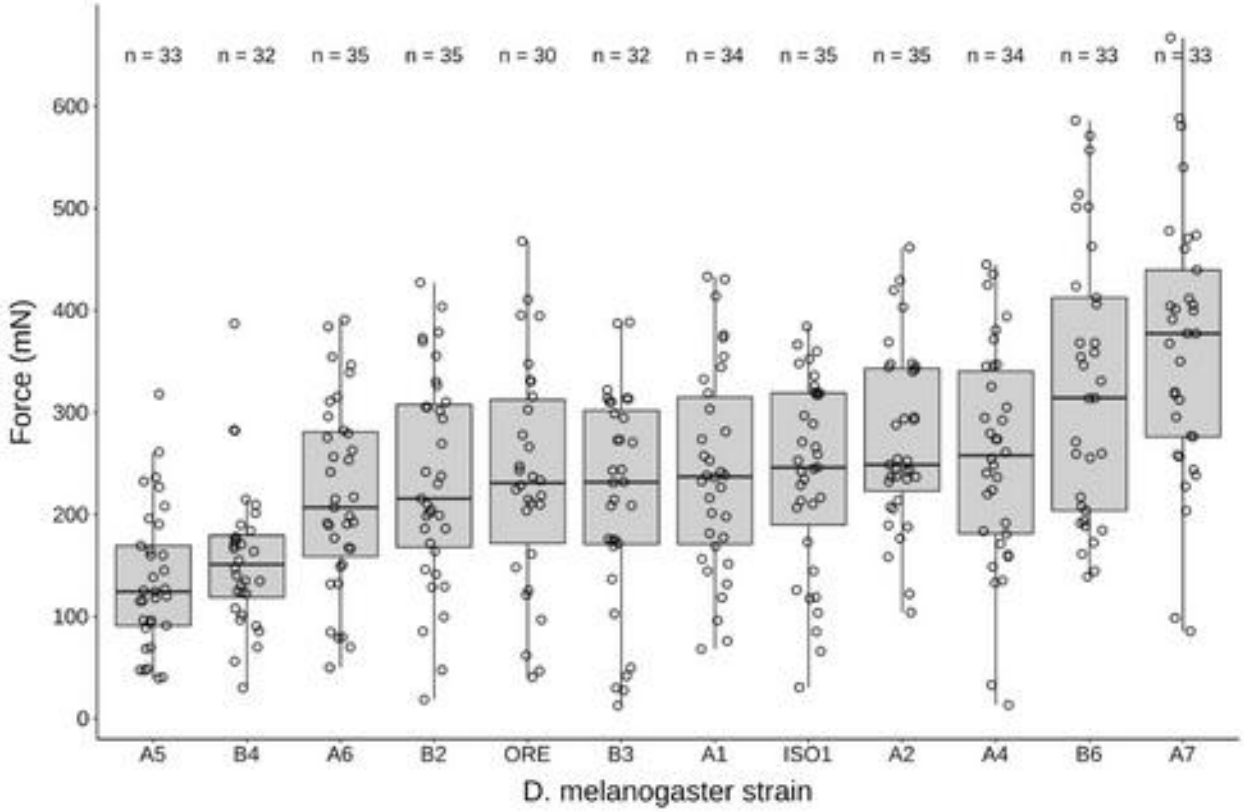


Figure 1. Adhesion strength of DSPR lines. Force indicates the force required to detach a pupa naturally attached to a glass slide. Each dot corresponds to a single pupa and n indicates the total number of pupae tested for each strain. Ends of the boxes define the first and third quartiles. The black horizontal line represents the median. The vertical line on the top of the box extends to the largest value no further than 1.5 * IQR from the upper hinge of the box. The vertical line on the bottom of the box extends to the smallest value at most 1.5 * IQR of the hinge. (IQR: inter-quartile range is the distance between the first and the third quartiles). Data beyond the end of these lines are "outlying" points.

All the experiments were performed at 25°C.

a. What are the mechanisms that can explain the high variability in adhesion values for a given strain?

Strain A7 appears to stick more than the other strains, and strain A5 appears to stick less. So we decided to investigate whether mutation(s) in some of the *Sgs* genes could explain their weaker and stronger adhesion, respectively. Complete sequences for all the Sgs genes were found in all strains.
Two candidate mutations were found: one in *Sgs5* gene (chromosome 3) for strain A7 and one in *Sgs8* gene (chromosome 3) for strain A5. Below are alignments of parts of the relevant *Sgs* sequences with strain iso-1, which is considered here as the "wild-type" normal strain.

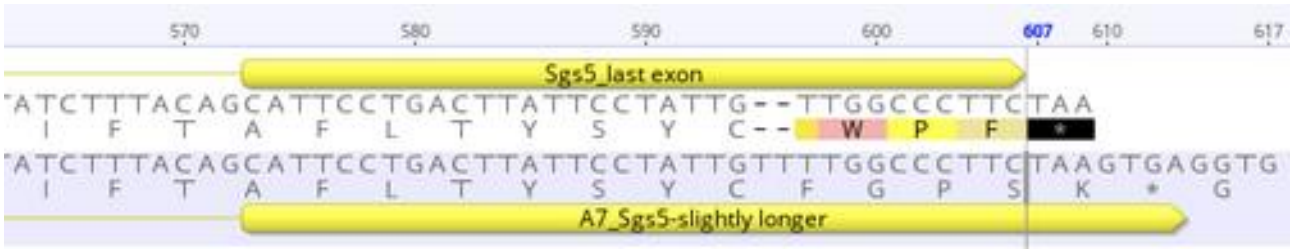


Fig. 2. Alignment of partial sequences of *Sgs5* for strain iso-1 (top) and strain A7 (bottom). The normal amino acid sequence of SGS5 ends with CWPF.

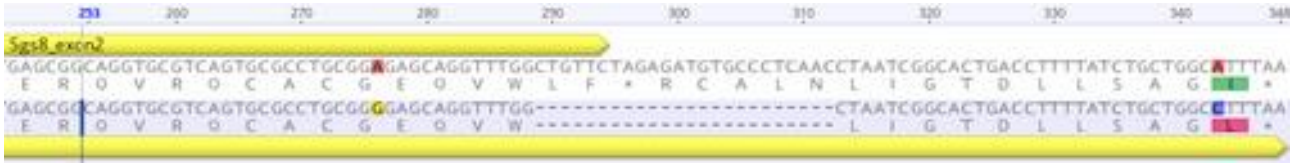


Fig. 3. Alignment of partial sequences of *Sgs8* for strain iso-1 (top) and strain A5 (bottom). The normal amino acid sequence of SGS8 ends with VWLF.

b. Describe the two candidate mutations and their effects on glue proteins.

We designed PCR primers so that they amplify a 162-bp fragment of the *Sgs8* gene from iso-1 strain and a 138-bp fragment from A5 strain. The two primers are Sgs8F (ATGGAGGGTCTTGAGCGGCA) and Sgs8R (GGAGTTTGGGTCGTGCAGGA).

c. Use OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) to calculate the hybridation temperature to be used for the PCR reaction.

We decided to test whether the candidate mutation in *Sgs8* is responsible for the reduction in adhesion. One solution is to make a CRISPR mutation in the iso-1 stock, but this experiment takes several months and costs about 1500 euros. So, in a first step, we decided to cross strain A5 with strain iso-1, collect the F1 hybrid progeny flies and cross them together to obtain F2 individuals. We measured adhesion of several F1 pupae and 150 F2 pupae, extracted their DNA and genotyped each individual at the *Sgs8* locus by PCR with primers Sgs8F and Sgs8R.

500bp-

300bp-

200bp-

100bp-

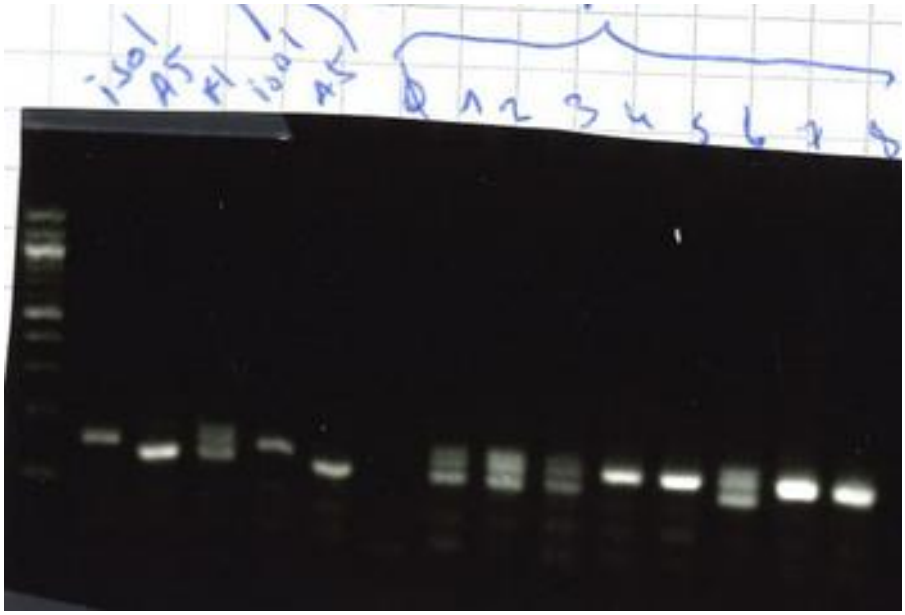


Fig. 4. Electrophoresis gel of the PCR fragments. Lines are in the following order: molecular weight size marker, mix of five iso-1 individuals, mix of five A5 individuals, mix of five F1 hybrid individuals from A5 x iso-1 cross, mix of five other iso-1 individuals, mix of five other A5 individuals, no DNA, individual F2 progeny pupae named from #1 to #8.

d. Explain the results of this electrophoresis gel.

e. Why was a PCR reaction performed in absence of DNA? (Note: You don't need to explain the presence of a third band in the F1.)

MW F1 A5 F1 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

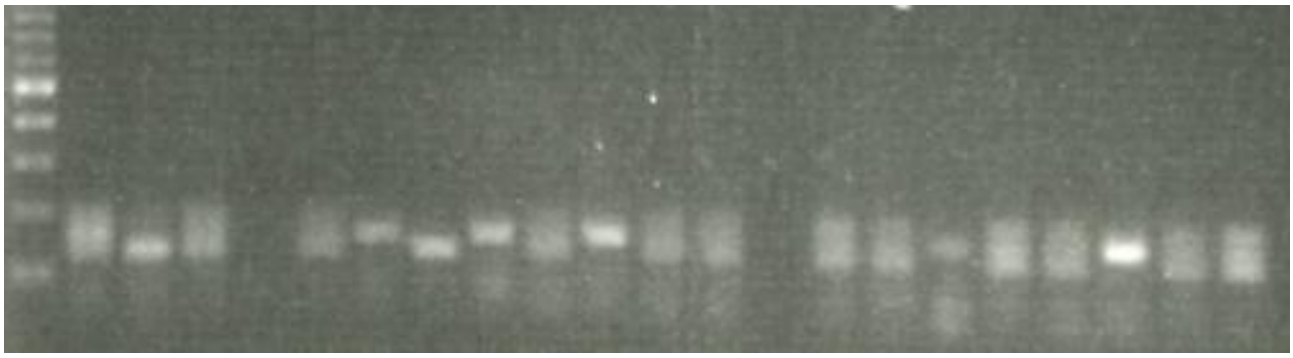


Fig. 5. Electrophoresis gel of other PCR fragments. F2 progeny individuals #9 to #24 are shown.

e. What is the genotype of individuals 1 to 24?

IV. SARS-CoV-2 evolution

NCBI GenBank is a large repository of nucleic sequences hosted in the USA. Every day the sequence data is synchronized between the European database EMBL-EBI/ENA, the Japanese database DDBJ/SRA and GenBank.

The reference sequence

a. Go to GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and search for the reference sequence of SARS-CoV-2.
b. What is the length of this sequence?
c. What information can you find in Genbank about the patient from which this sequence was obtained?
d. In which publication was this sequence published?
e. Each sequence in GenBank has an accession number. What is the accession number for the SARS-CoV-2 reference sequence?
Note that the current accession number ends with .2. This means that there has been an update of the sequence in GenBank.
Go to https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.1?report=girevhist to compare the two versions.
f. When was the second version of the sequence published on NCBI?
g. Is there a difference in the reference authors list?
h. Save the sequence as a FASTA file and examine the text file.

NCBI Taxonomy

NCBI Taxonomy is a curated classification and nomenclature for all of the organisms in the public sequence databases.
https://www.ncbi.nlm.nih.gov/taxonomy
i. What is the taxon ID for SARS-CoV-2?
j. Among the Orthocoronaviridae, how many groups are defined besides the alphacoronaviruses and betacoronaviruses?

Finding related sequences

BLAST (Basic Local Alignment Search Tool) finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.
https://blast.ncbi.nlm.nih.gov/Blast.cgi

k. Use nucleotide BLAST to find other coronaviruses related to SARS-CoV-2.
Hint: you have to exclude an organism.
The first sequences found by BLAST that are most closely related are synthetic constructs and sequences whose taxon field was incorrectly entered.
l. Based on BLAST, what is the name of the virus that is most closely related to SARS-CoV-2 (outside of humans)?
m. What is the percentage identity indicated between this virus and SARS-CoV-2?
n. On which animal host species was this virus collected?
o. By which research team?
p. When and where was this virus sampled?

q. Is *Rhinolophus malayanus* part of the Microchiroptera?
r. Go to the LinkOut "DNA barcoding : Rhinolophus malayanus" http://www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=12183 and search for more information about this species.