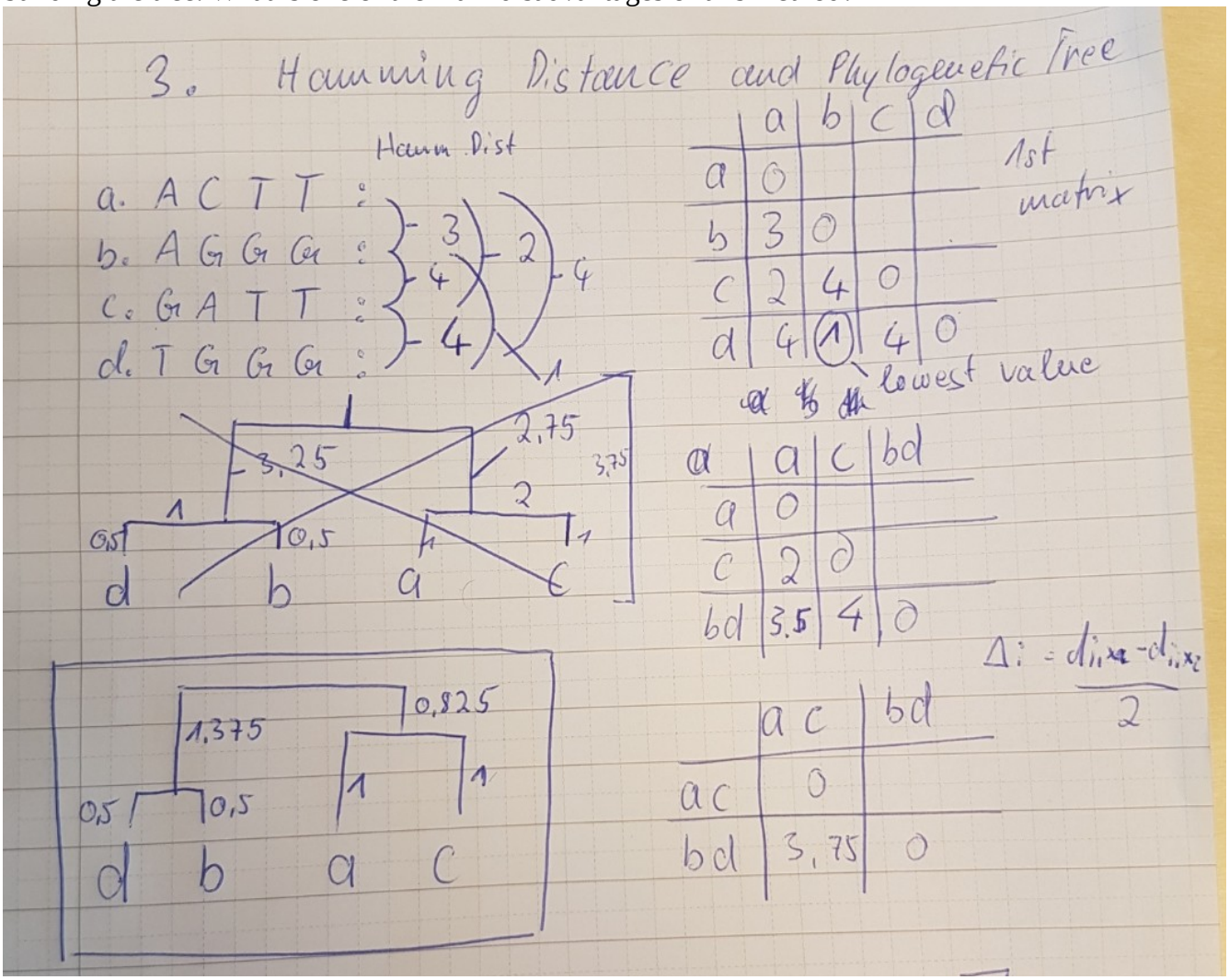
1. What kind of information does a phylogenetic tree illustrate?

The phylogenetic tree shows the evolutionary relationships between different biological species. This relationship can be called phylogeny and is referring to the similarities and differences in their physical or genetic characteristics.

2. What is the difference, in terms of the information displayed, between rooted and unrooted trees?

In contrast to rooted trees, unrooted trees only show the relatedness of the nodes and do not show ancestal roots. Rooted trees additionally can illustrate time estimation for evolutionary relationships.

3. Use the UPGMA method together with the Hamming distance to build the phylogenetic tree for the following DNA sequences: ACTT, AGGG, GATT, TGGG. Show every step you performed when building the tree. What is one of the main disadvantages of this method?



So first one calculates the Hamming distance between the different sequences. Given those one builds a matrix, where the different hamming distances between sequences are present. Then, one takes the lowest value, which is 1 between d and b, connect these two, meaning that sequence d and b are most closely related. In the next step, one produces a second matrix (below the first matrix) with the formula given at the left side of the image. For example, if one wants to calculate the distance between a and bd it is: distance=(a to b, which is 3, plus a to d, which is 4) divided by 2 equals 3.5. In this second step,

the smallest value is 2 between a and c, so one have to connect those two. Their arms have each a value of 1. In the third step, one calculates the distance between bd and ac, which is, using the given function, 3,75. By dividing this by 2 and substracting the already given values for the previous arms, one then can calculate the value of the two missing arms.
The main disadvantage of this method is that it is not possible to look at local high-variation regions appearing in other connected subtrees.

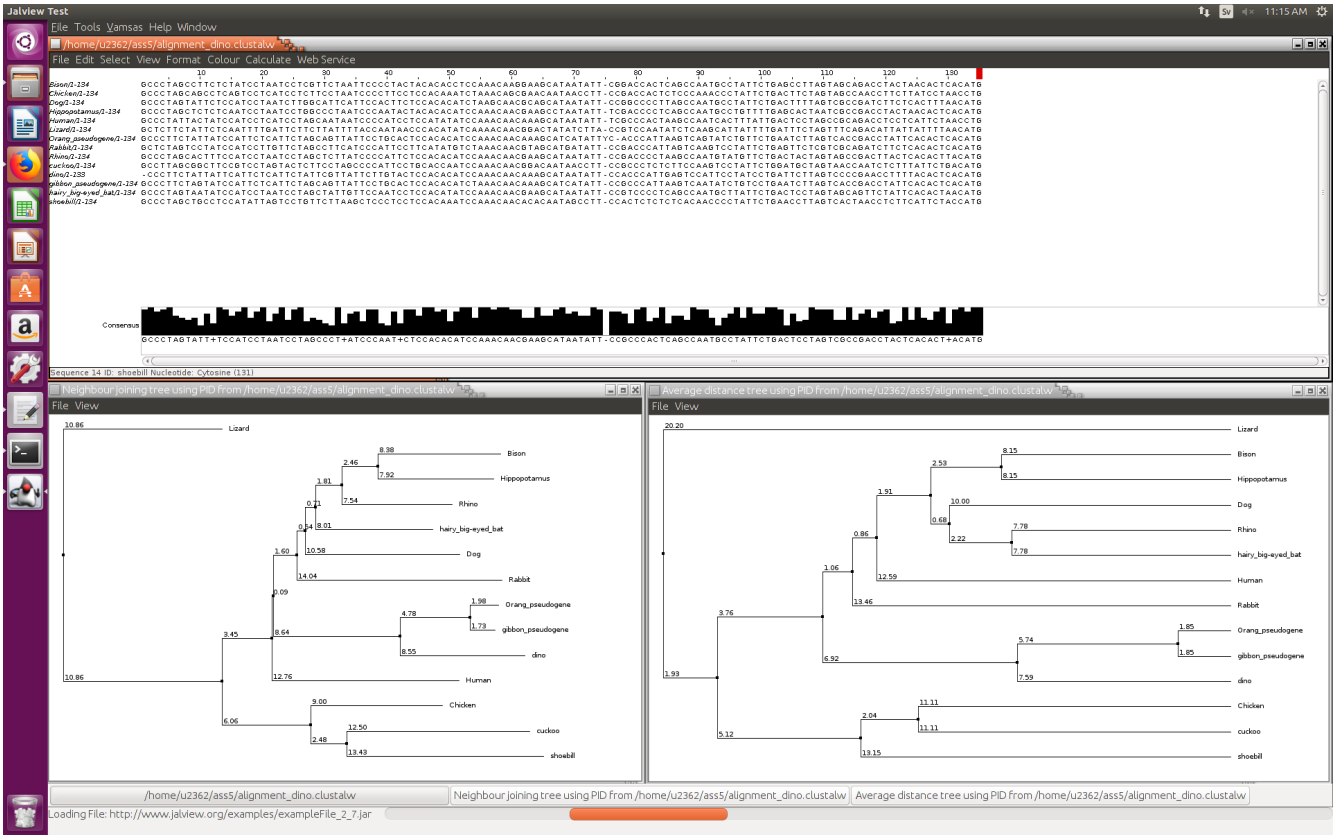
4. Which mouse has the largest evolutionary distance to the common ancestor (tree root)?

The grey mouse has with the value "10" the biggest distance between the sequences.

5. In terms of ordinary time, which mouse diverged first, the grey one or the purple one?

In the given image, there is no molecular clock given. The distance does not say anything about whether any of the mice diverged first.

6. Use one of the methods shown in the previous lab to align the sequences and construct an evolutionary tree. Create and submit a print screen image of the multiple sequence alignment and the phylogenetic tree.



7. Which three species sequences are the most related to the dinosaur sequence according to your phylogenetic tree?

Human, Orang_pseudogene, gibbon_pseudogene

8. What can you conclude from your observations, in terms of biology and evolution?

The phylogenetic tree shows, that the found sample was probably not a dino or the result is scientifically not significant, because then it should have been more related to the e.g. birds in the phylogenic tree. It is very strange, that the “dino” sample is closely related to primates or even mammals.
Possibly, the sample was decontaminated by human DNA.

9. What do you think is a protein domain?

Protein domains are different units inside a protein, which are classified as units because of different structures or functions or because they can fold independently (which is the reason for conservation).

10. What do you think is a protein fold?

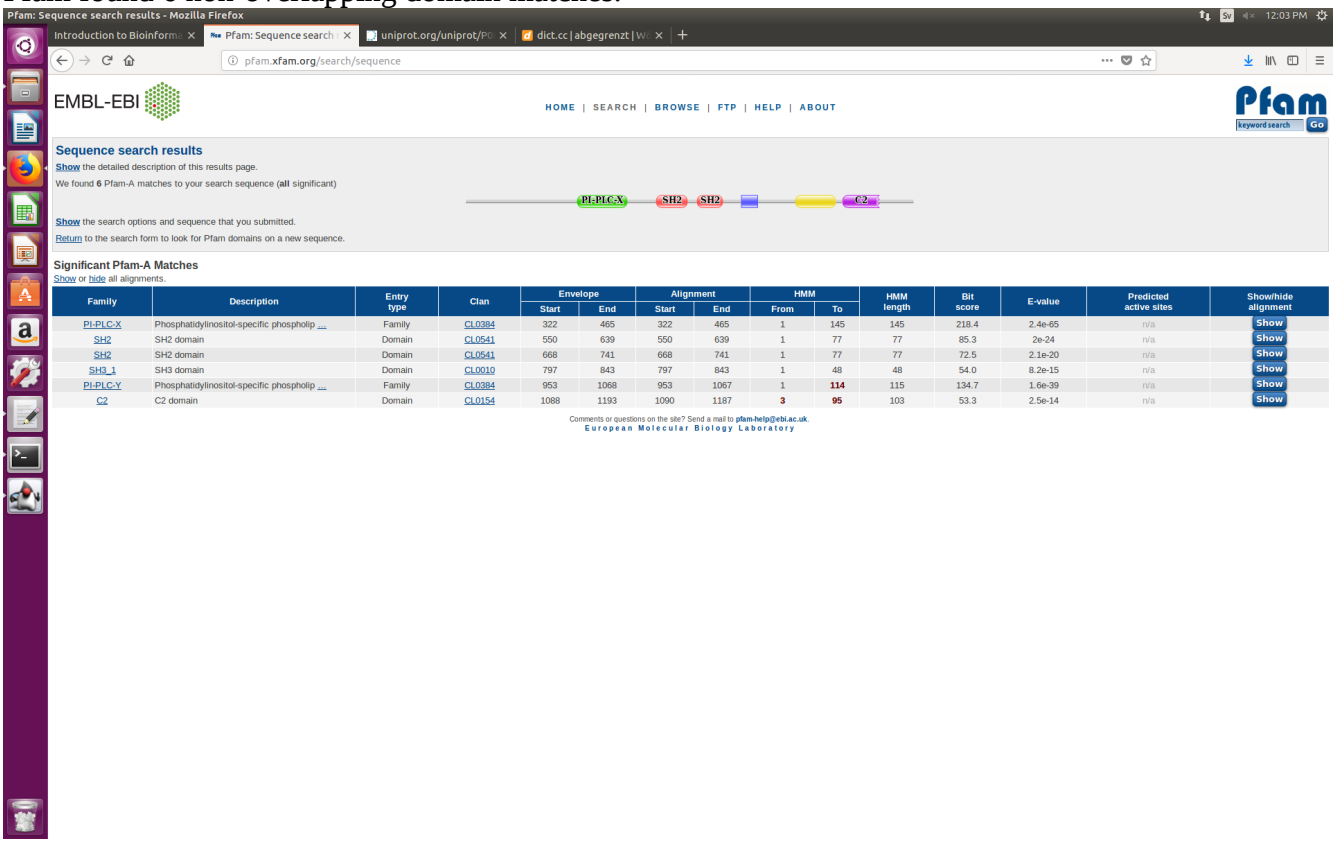
The protein fold describes how elements of secondary structure are located to each other in space.

11. What procedure does Pfam use to create sequence families and to add new members to existing families? What do you think is the role of HMM profiles in this process?

To create a new family PFAM is performing a high-quality seed alignment of a representative subset of sequences. Then, a profile hidden Markov model is produced by using HMMER. Taken the HMM profile, one then can search sequences from databases and compare them. Given a previously set threshold, all hits obtained are then classified is members of the new sequence family. The role of the HMM is that of a profile, with which members of the given sequence family can be identified/classified.

12. Find the entry "PLCG1_BOVIN" in the Uniprot (SwissProt) database. Get its peptide sequence and search it using the Pfam Search utility. How many non-overlapping Pfam-A domains does it have? Include a screen shot of the Pfam predictions for this protein in your report.

Pfam found 6 non-overlapping domain matches.



Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
PI-PLC-X	Phosphatidylinositol-specific phospholip ...	Family	CL0384	322	465	322	465	1	145	145	218.4	2.4e-65	n/a	Show
SH2	SH2 domain	Domain	CL0541	550	639	550	639	1	77	77	85.3	2e-24	n/a	Show
SH2	SH2 domain	Domain	CL0541	668	741	668	741	1	77	77	72.5	2.1e-20	n/a	Show
SH3_1	SH3 domain	Domain	CL0010	797	843	797	843	1	48	48	54.0	8.2e-15	n/a	Show
PI-PLC-Y	Phosphatidylinositol-specific phospholip ...	Family	CL0384	953	1068	953	1067	1	114	115	134.7	1.6e-39	n/a	Show
C2	C2 domain	Domain	CL0154	1088	1193	1090	1187	3	95	103	53.3	2.5e-14	n/a	Show