**Bioinformatic assignment: phylogenetic analysis**

1. Newick tree formats: Write the Newick notation of the following tree:

A black line drawing of a couple of rectangular objects

Description automatically generated

1. UPGMA vs Neighbor-Joining (NJ): We performed an MSA of all Caspases (paralogs and orthologs) for you which you can find in the file (in Fasta Format).
   1. From the MSA, generate a phylogenetic tree using UPGMA. Paste a picture of the tree. Where is the root in this tree? What is the relationship between the orthologs and the paralogs according to the tree, does it make sense?
   2. From the MSA, generate a phylogenetic tree using NJ. Paste a picture of the tree. Does the tree resemble the tree in (a)? Why yes/no? Generate a mid-point rooted tree from the NJ-tree and paste it here. How does it compare to the UPGMA tree.

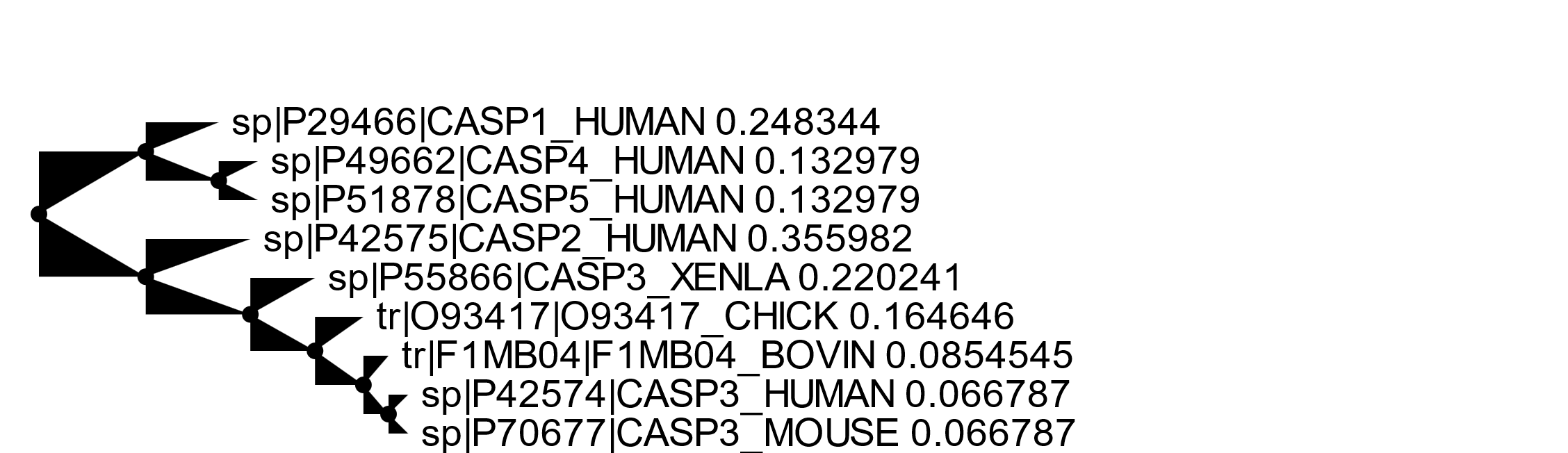
**A diagram of a person and animals

Description automatically generatedtreedyn is problematic, itols.embl.de**

1. Outgroups and re-rooting. We performed a sequence alignment of P53 from multiple species and used NJ to generate a corresponding phylogenetic tree (Newick format).
   1. Compare the generated P53 tree with automatic mid-point rooting to the known species tree (right). Is it the same or is it different?
   2. Where is the root in the P53 tree? Based on the species tree, which P53 could be used as an outgroup?
   3. Reroot the P53 using the outgroup. Paste a picture of the re-rooted tree. Does it make more sense now?

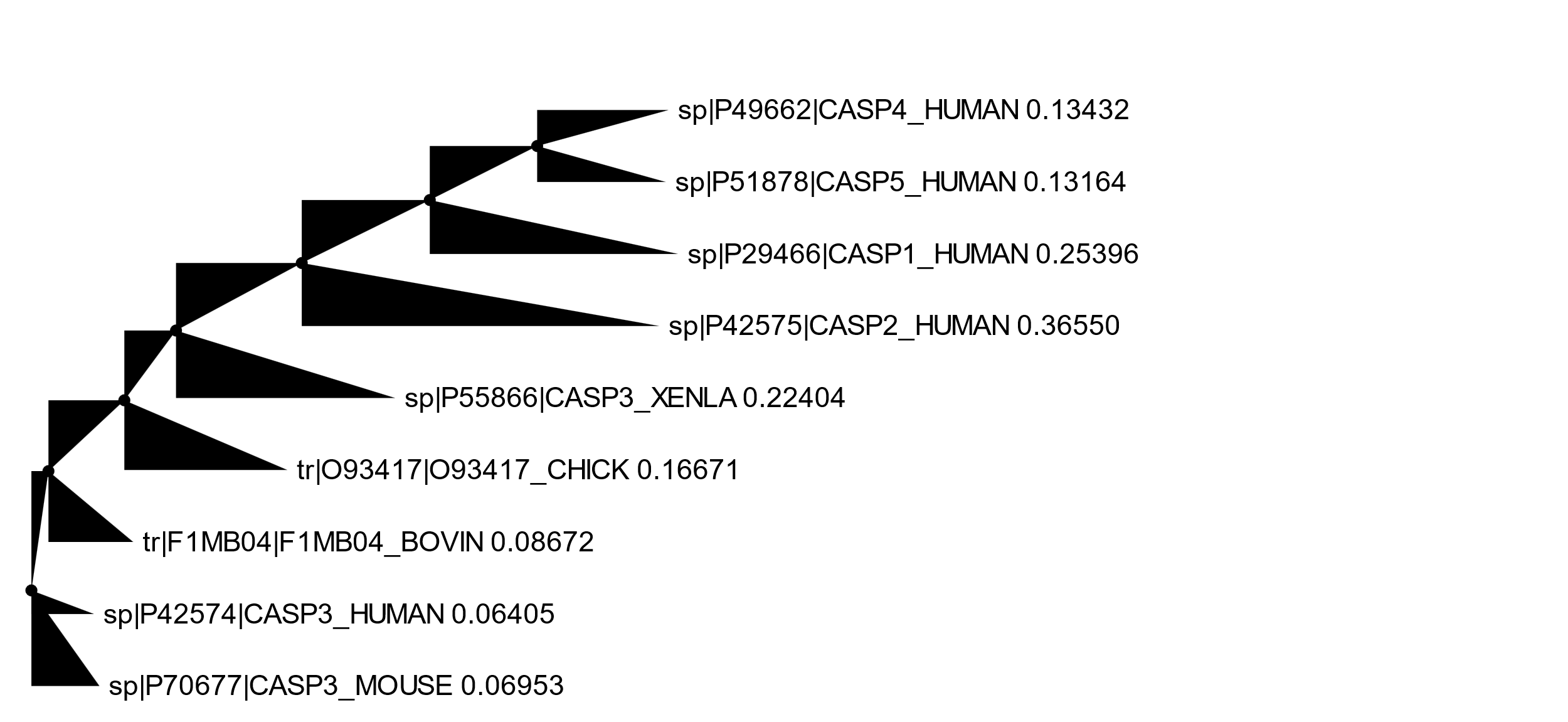
A diagram of a number of objects

Description automatically generated with medium confidence

1. To understand the phylogeny of different organisms its often makes sense to compare multiple proteins. We extracted the ortholog sequences of two proteins (PGK1 and FKH2) from 9 yeast species (see phylogeny on the right).
   1. Perform MSA (MUSCLE or Clustalw) for the first protein PGK1 and construct a phylogenetic tree using NJ. Look at the tree in TreeDyn.
   2. Which protein would you set to be the out group? Paste apicture of the rerooted tree.
   3. Perform the same steps for the FKH2.
   4. Which one better reflects the species phylogeny? Looking at the two alignments, why might this be?
2. (G,((C,B),(N,M))) - ((Chimpanzee,Bonobo), (Neanderthals,ModernHuman)),Gorilla);
3. 

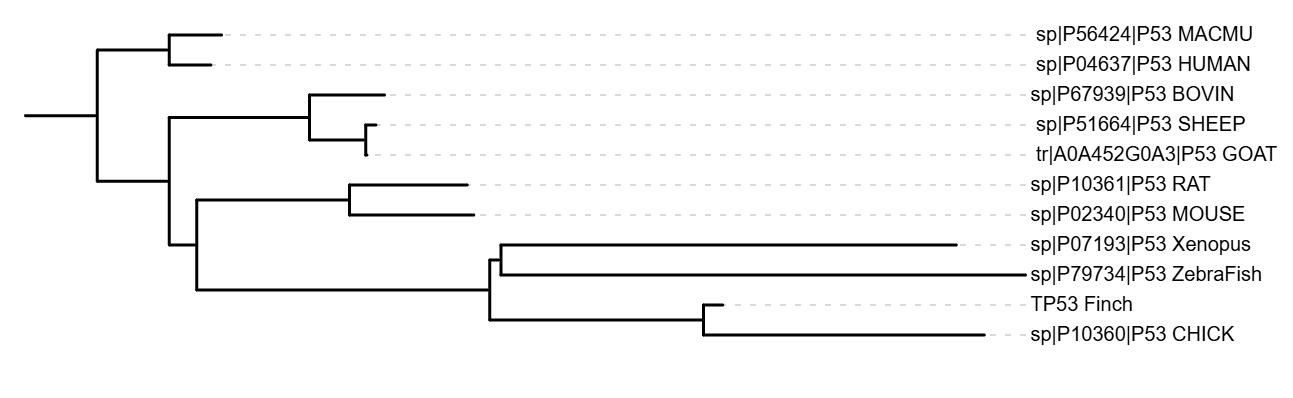
UPGMA

* The root divides three paralogs on the left branch and in the right one there are some paral and orthol.
* The human-chicken-mouse doesn’t make sense since human Is also in the top branch



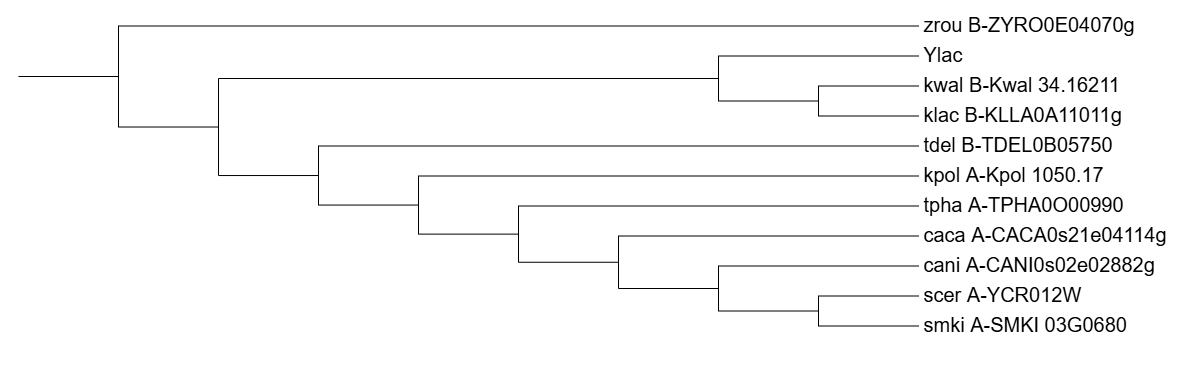
NJ

* It does resemble the upgma tree in having the human casp3 branched at the same point as the mouse casp3 but in this tree all the other human casps are “together” branching from one another, unlike in the upgma tree

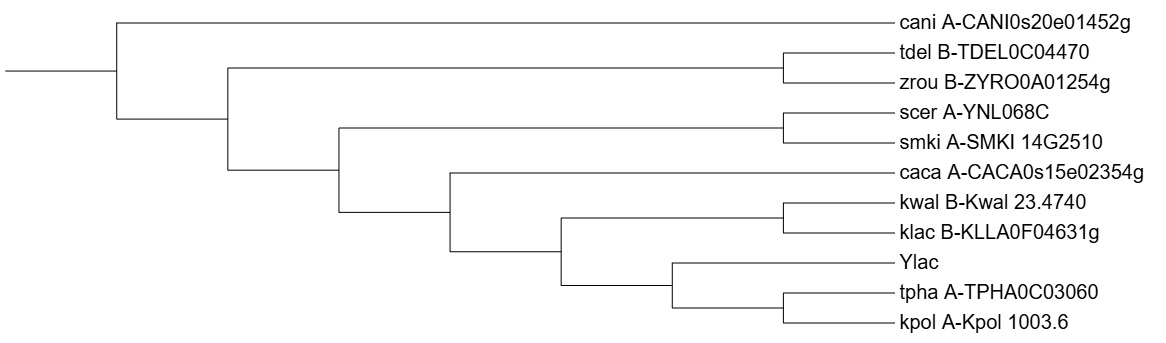
1. –
   1. It is similar, in the p53 tree human is closer to macmu, sheep to goat then to bovine, mouse to rat ecc.
   2. Root has 3 branches, one to human and macmu, one to bovin/sheep-goat and last one to the others
      1. The outgroup could be the human/macmu
   3. 

Yes, it makes more sense, since the more related species are in closer branches

1. –
   1. –
   2. I’d set Zrow as the outgroup protein



* 1. Reroot on cani



* 1. Tree (b) with Zrow as the outgroup better reflects the species phylogeny since it groups the sequences into more plausible evolutionary relationships having Zrow as the most different lineage