

4. Phylogenetic trees - solutions

AUTHOR
EPFL - SV - BIO-463

PUBLISHED
March 12, 2024

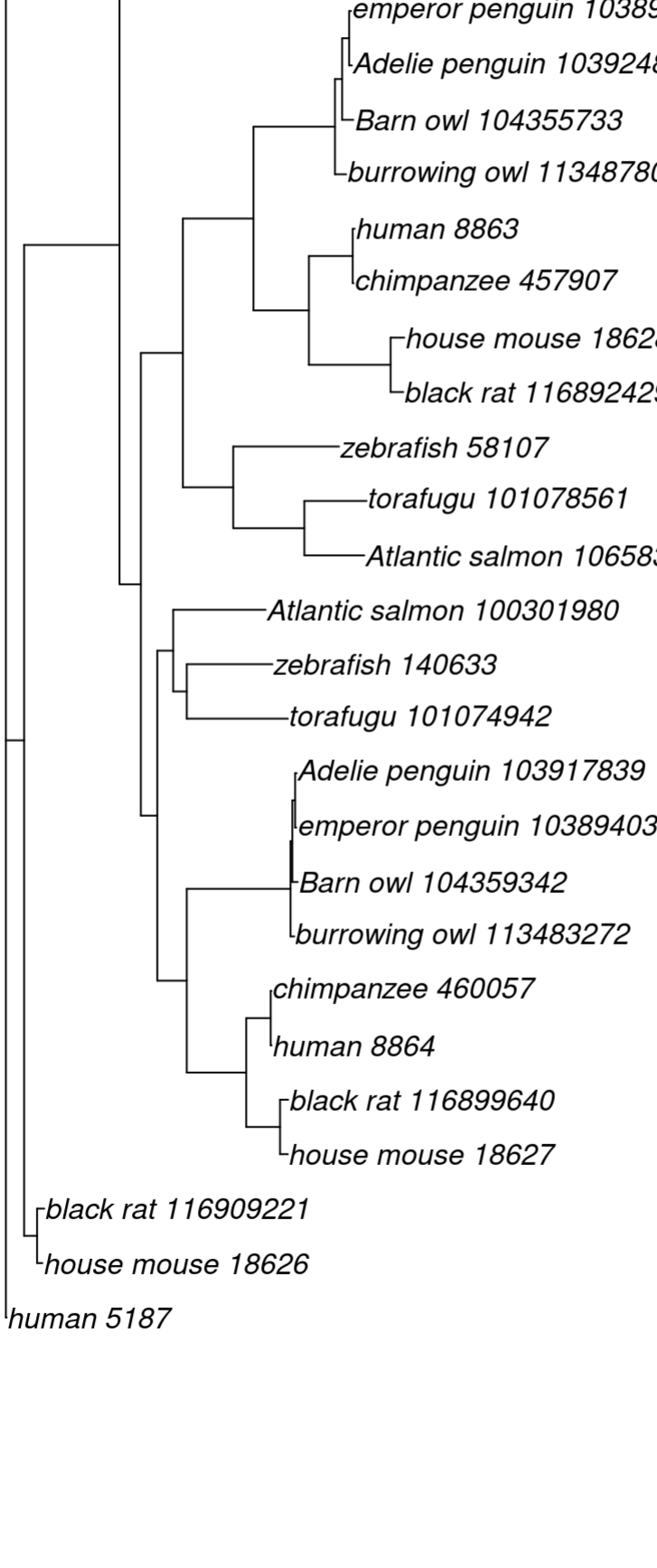
Exercise 2: Plot and analyze the gene tree

Use the R packages *ape* to plot and annotate your tree.

1. Load the file *RAXML_bipartitions.Homologs* using [read.tree](#)
2. Plot the tree using [plot.phylo](#)
3. Re-root the tree at the base of all bird genes (find all tree tips corresponding to *penguin* using [grep](#), then [getMRCA](#) and [root](#)
4. Display the bootstrap values > 80 with [node.labels](#)
5. Annotate the tree by displaying the *Barn_owl*, *human* and *Atlantic_salmon* genes in 3 different colors

```
library(ape)

# read the output of RAXML - the reconstructed phylogeny with bootstrap values,
tree = read.tree("RAXML_bipartitions.Homologs")
plot.phylo(tree)
```

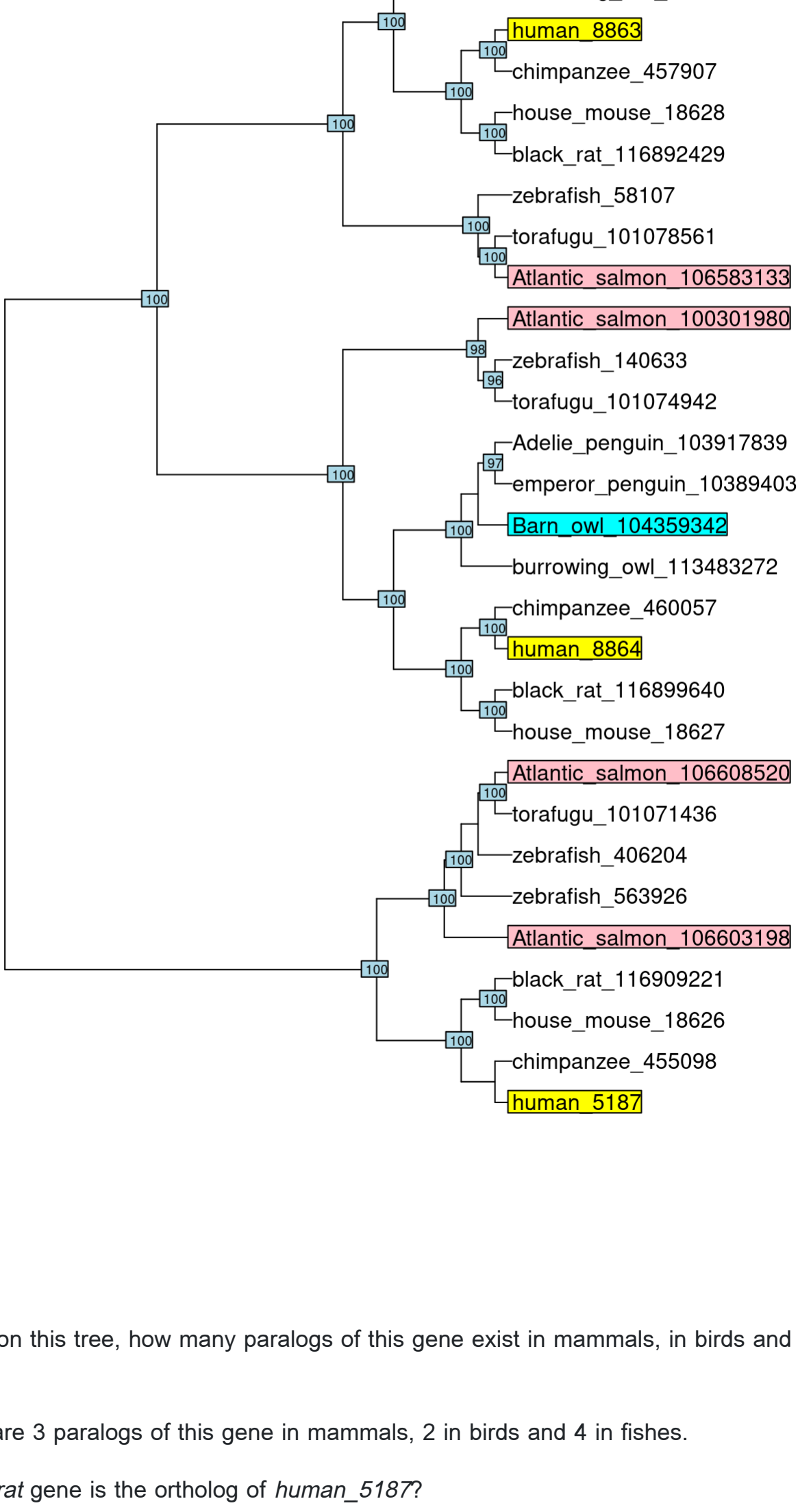


```
# the tree is "unrooted", we force the root at the base of all birds:
# select all tips with the name "penguin" (grep), then find their ancestor (get
root = getMRCA(tree, grep("penguin", tree$tip.label))
tree = root(tree, node=root, resolve.root=T)
```

```
# select bootstrap values > 80 (the node label is the bootstrap values in text,
ns = which(as.numeric(tree$node.label) > 80)
nb.tip = length(tree$tip.label)
```

```
# select: 1. "human" genes, 2. "salmon" genes, 3. "owl" genes, 4. all the rest
# using grep to identify labels containing a given substring
# "all the rest" is 1,...,N from which we remove the numbers already selected
ihs = grep("human", tree$tip.label)
ipt = grep("salmo", tree$tip.label)
icn = grep("Barn_owl", tree$tip.label)
irest = (1:length(tree$tip.label))[-c(ihs,ipt,icn)]
```

```
# re-plot the tree with only bootstrap values>80, and selected genes in color (
par(oma=c(0,0,0,8), xpd=NA)
plot.phylo(tree, show.tip.label=F, use.edge.length=F)
node.labels(tree$node.label[ns], node=ns+nb.tip, cex=0.6)
tiplabels(tree$tip.label[ihs], ihs, adj=c(0,.5))
tiplabels(tree$tip.label[ipt], ipt, bg='pink', adj=c(0,.5))
tiplabels(tree$tip.label[icn], icn, bg='cyan', adj=c(0,.5))
tiplabels(tree$tip.label[irest], irest, bg=0, frame='none', adj=c(0,.5))
```



6. Based on this tree, how many paralogs of this gene exist in mammals, in birds and in fishes?

- There are 3 paralogs of this gene in mammals, 2 in birds and 4 in fishes.

7. Which *rat* gene is the ortholog of *human_5187*?

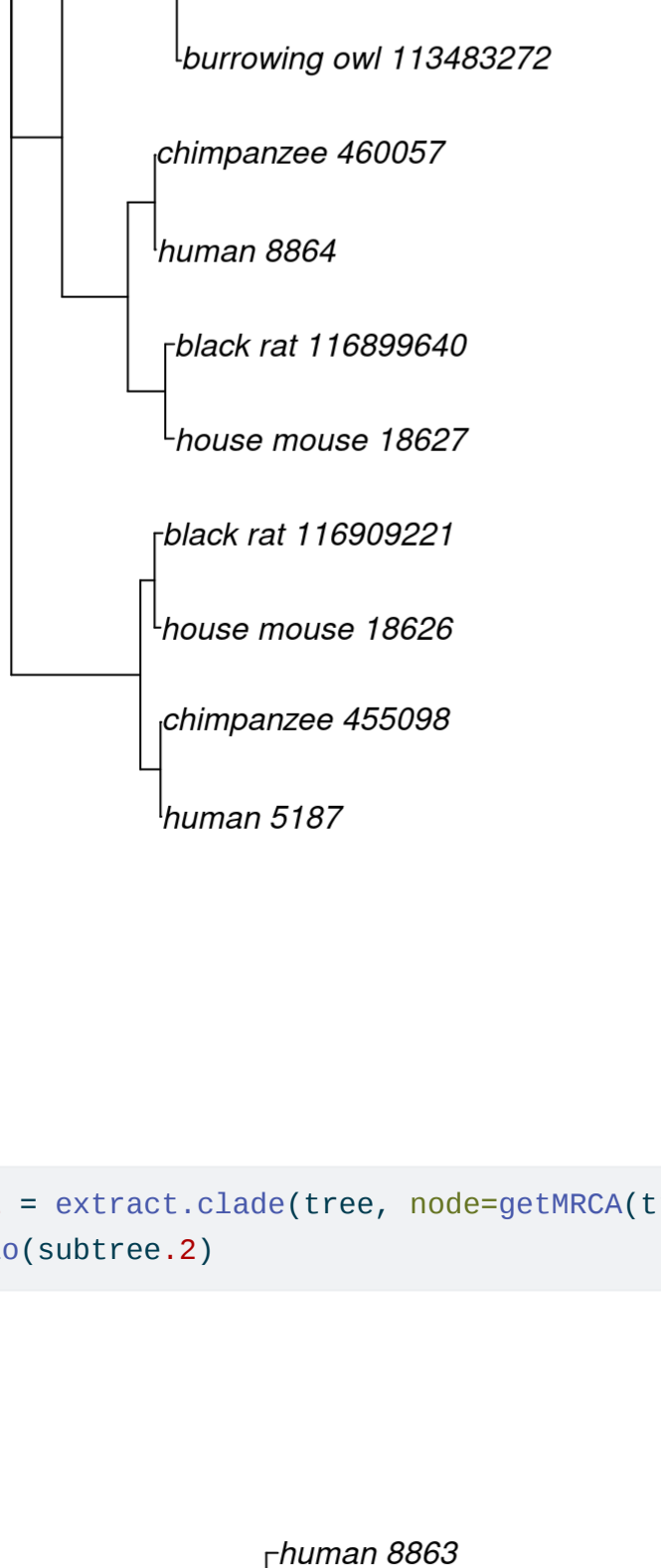
- *black_rat_116909221* is the ortholog of *human_5187*.

Exercise 3: Manipulate the gene tree

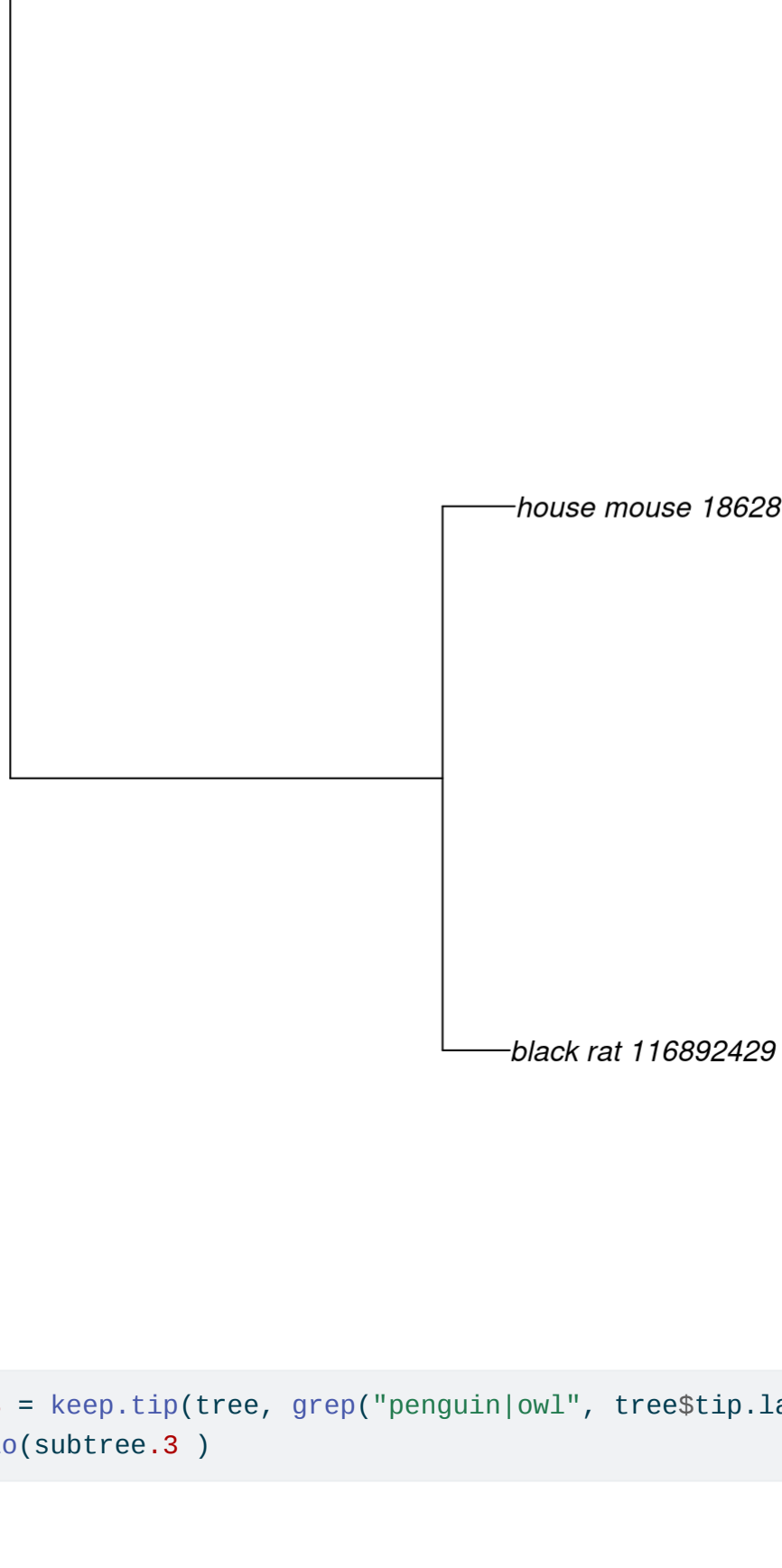
Use the methods [drop.tip](#), [keep.tip](#) and [extract.clade](#) to plot the following subtrees:

1. All non-fish species (remove all *salmon*, *zebrafish* and *torafugu*)
2. The clade containing *human_8863* and *house_mouse_18628* (a clade contains all descendants of the last common ancestor of these 2 leaves)
3. Only the birds (*penguins* and *owls*)

```
par(oma=c(0,0,0,5), xpd=NA)
subtree.1 = drop.tip(tree, grep("fugu|salmo|zebra", tree$tip.label))
plot.phylo(subtree.1)
```



```
subtree.2 = extract.clade(tree, node=getMRCA(tree, grep("_8863|_18628", tree$tip
plot.phylo(subtree.2)
```



```
subtree.3 = keep.tip(tree, grep("penguin|owl", tree$tip.label))
plot.phylo(subtree.3 )
```

