4. Phylogenetic trees - solutions **AUTHOR PUBLISHED** EPFL - SV - BIO-463 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 *nodelabels*
- 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)

ıchimpanzee 455098

Atlantic salmon 106603198 _iemperor penguin 103894682 ^lAdelie penguin 103924883 -Barn owl 104355733 Lburrowing owl 113487809 _Ihuman 8863 chimpanzee 457907 ⊢house mouse 18628 black rat 116892429 zebrafish 58107 torafugu 101078561--Atlantic salmon 106583133 Atlantic salmon 100301980 zebrafish 140633torafugu 101074942 _iAdelie penguin 103917839 lemperor penguin 103894033 Barn owl 104359342 burrowing owl 113483272

ıchimpanzee 460057

rblack rat 116899640

Lhouse mouse 18627

the tree is "unrooted", we force the root at the base of all birds:

root = getMRCA(tree, grep("penguin", tree\$tip.label))

using grep to identify labels containing a given substring

irest = (1:length(tree\$tip.label))[-c(ihs,ipt,icn)]

plot.phylo(tree, show.tip.label=F, use.edge.length=F)

tiplabels(tree\$tip.label[ihs], ihs, adj=c(0,.5))

nodelabels(tree\$node.label[ns], node=ns+nb.tip, cex=0.6)

tree = root(tree, node=root, resolve.root=T)

ns = which(as.numeric(tree\$node.label) > 80)

nb.tip = length(tree\$tip.label)

ihs = grep("human", tree\$tip.label) ipt = grep("salmo", tree\$tip.label) icn = grep("Barn_owl", tree\$tip.label)

par(oma=c(0,0,0,8), xpd=NA)

select all tips with the name "penguin" (grep), then find their ancestor (get

select bootstrap values > 80 (the node label is the bootstrap values in text,

select: 1. "human" genes, 2. "salmon" genes, 3. "owl" genes, 4. all the rest

"all the rest" is 1,..., N from which we remove the numbers already selected

re-plot the tree with only bootstrap values>80, and selected genes in color (

^lhuman 8864

rblack rat 116909221

^Lhouse mouse 18626

^lhuman 5187

-Atlantic salmon 106608520

torafugu 101071436

zebrafish 406204

-zebrafish 563926

tiplabels(tree\$tip.label[ipt], ipt, bg='pink', adj=c(0,.5)) tiplabels(treetip.label[icn], icn, bg='cyan', adj=c(0,.5)) tiplabels(tree\$tip.label[irest], irest, bg=0, frame='none', adj=c(0,.5)) emperor_penguin_10389468 delie_penguin_103924883 Barn_owl_104355733 burrowing owl 113487809 100 human 8863 chimpanzee 457907 -house_mouse_18628 black rat 116892429 -zebrafish_58107 torafugu 101078561 Atlantic salmon 106583133 Atlantic salmon 100301980 -zebrafish_140633 torafugu_101074942 -Adelie_penguin_103917839 emperor_penguin_10389403 Barn owl 104359342 -burrowing_owl_113483272 chimpanzee_460057 human 8864 -black_rat_116899640

100

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

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

Use the methods <u>drop.tip</u>, <u>keep.tip</u> and <u>extract.clade</u> to plot the following subtrees:

subtree.1 = drop.tip(tree, grep("fugu|salmo|zebra", tree\$tip.label))

remperor penguin 103894682

^lAdelie penguin 103924883

-Barn owl 104355733

^lchimpanzee 457907

₁human 8863

Lburrowing owl 113487809

2. The clade containing human_8863 and house_mouse_18628 (a clade contains all

1. All non-fish species (remove all salmon, zebrafish and torafugu)

descendents of the last common ancestor of these 2 leaves)

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

Exercise 3: Manipulate the gene tree

3. Only the birds (penguins and owls)

par(oma=c(0,0,0,5), xpd=NA)

plot.phylo(subtree.1)

black_rat_116909221 is the ortholog of human_5187.

fishes?

-house_mouse_18627

-torafugu_101071436

zebrafish_406204

-zebrafish_563926

-black_rat_116909221

house_mouse_18626

chimpanzee_455098

human 5187

Atlantic_salmon_106608520

Atlantic salmon 106603198

house mouse 18628 └*black rat 116892429*

Adelie penguin 103917839

emperor penguin 103894033

-Barn owl 104359342

_ichimpanzee 460057

^Lhouse mouse 18626

ıchimpanzee 455098

-human 8863

^Lchimpanzee 457907

^lhuman 5187

plot.phylo(subtree.2)

¹human 8864

Lburrowing owl 113483272

rblack rat 116899640 ^Lhouse mouse 18627 black rat 116909221

subtree.2 = extract.clade(tree, node=getMRCA(tree, grep("_8863|_18628", tree\$ti

-house mouse 18628

black rat 116892429

Adelie penguin 103924883

remperor penguin 103894682

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

plot.phylo(subtree.3)

LBarn owl 104355733 Lburrowing owl 113487809 ₁Adelie penguin 103917839 emperor penguin 103894033

-Barn owl 104359342

Lburrowing owl 113483272