CVX Trees

This is a CVX tree and data manipulation markdown file

```r
library(ape)
library("knitr")
cvx.tre<-ape::read.tree("cacti.njtree.nwk")

# Won't use tree tip names to re-import NCBI data
# cvx.seq <- read.GenBank(cvx.tre$tip.label, species.names=TRUE)
#
# Warning message:
# In ape::read.GenBank(cvx.acc, species.names = TRUE) :
#   cannot get the following sequences:
# CMU62963, CCU23414, PMPRNA, CYMRNA, FXMCGA, BBSCG,
# AM1BAMVGR, PVXCGA, PVXX3, PVXCOPR, PVXGEN, PVXCG,
# SYVCPE, MNCCGAA

# Importing file sent by Alexa (note: could replace tip names with accessions and go back to NCBI
cvx.data<-read.csv('seq.metadata.csv')
```

```r
# Output tree info: note has polytomy somewhere
cvx.tre
```

```
##
## Phylogenetic tree with 106 tips and 104 internal nodes.
##
## Tip labels:
##   Hylocereus_polyrhizus_1195597_SRR11190797_potexvirus_-_NC_002815_consensus_sequence, Hylocereus_un
##
## Unrooted; includes branch lengths.
```

```r
# How many unique organisms do we have? (also, coarse test Organism = db_xref)
unique(cvx.data$Organism)
```

```
##  [1] "Cactus virus X"              "Schlumbergera virus X"
##  [3] "Zygocactus virus X"          "Opuntia virus X"
##  [5] "Nandina mosaic virus"        "Alternanthera mosaic virus"
##  [7] "Clover yellow mosaic virus"  NA
##  [9] "Pitaya virus X"              "Papaya mosaic virus"
## [11] ""                           "Plantago asiatica mosaic virus"
## [13] "Senna mosaic virus"         "Hydrangea ringspot virus"
## [15] "Babaco mosaic virus"
```

```r
unique(cvx.data$db_xref)
```

```
## [1] "taxon:112227" "taxon:253700" "taxon:253701" "taxon:253702"
## [5] "taxon:315713" "taxon:85454"  ""             "taxon:1176736"
## [9] "taxon:12181"  "taxon:28354"  "taxon:1881013" "taxon:112228"
## [13] "taxon:2060511"
```

```r
# How many are "* virus X" or "* mosaic virus"
length(grep("virus X",unique(cvx.data$Organism)))
```

```
## [1] 5
```

```r
length(grep("mosaic",unique(cvx.data$Organism)))
```

```
## [1] 7
```

```r
# P.S. They are mutually exclusive!
intersect(grep("virus X",unique(cvx.data$Organism)),grep("mosaic",unique(cvx.data$Organism)))
```

```
## integer(0)
```

```r
# And there there are these one-offs
length(grep("ringspot",unique(cvx.data$Organism)))
```

```
## [1] 1
```

```r
length(grep("exvir",unique(cvx.data$Organism)))
```

```
## [1] 0
```

```r
length(grep("edge",unique(cvx.data$Organism)))
```

```
## [1] 0
```

```r
length(grep("virus 3",unique(cvx.data$Organism)))
```

```
## [1] 0
```

```r
choose.node <- function(col,tree) {
    o <- identify(tree)
    nodelabels(node=o$nodes, pch = 19, col = col)
    print(o)
}

plot(cvx.tre, type="f",cex=0.5)
```

```
# Not run
# Selecting node subtending potexviruses
#choose.node("red",cvx.tre)
#Click close to a node of the tree...
#$nodes
#[1] 379

# Selecting node subtending potexviruses
#choose.node("green",cvx.tre)
#Click close to a node of the tree...
#$nodes
#[1] 382

# Drop the rest of the tree; note name
cvx.tre2<-ape::extract.clade(cvx.tre,382)
```

```
## Error in prop.part(phy)[[node - n]]: subscript out of bounds
```

```
plot(cvx.tre2, type="f",cex=0.75)
```

```
## Error in plot(cvx.tre2, type = "f", cex = 0.75): object 'cvx.tre2' not found
```

```
tip.rename.org = function(tree,dataframe){
   for(i in 1:length(tree$tip.label)){
```

```
        tree$tip.label[i]<-dataframe$Organism[grep(tree$tip.label[i],dataframe$Name)]
    }
tree
}

tip.rename.host = function(tree,dataframe){
    for(i in 1:length(tree$tip.label)){
        tree$tip.label[i]<-dataframe$bi_host_binomial [grep(tree$tip.label[i],dataframe$Name)]
    }
tree
}

#tip.rename.host(

#Remove some Plantago asiatica mosaic virus samples
cvx.tre3<-drop.tip(cvx.tre2,cvx.tre2$tip.label[grep(c("KU159"),cvx.tre2$tip.label)])
```

Error in drop.tip(cvx.tre2, cvx.tre2$tip.label[grep(c("KU159"), cvx.tre2$tip.label)]): object 'cvx.t:

```
cvx.tre4<-drop.tip(cvx.tre3,cvx.tre3$tip.label[grep(c("MH360"),cvx.tre3$tip.label)])
```

Error in drop.tip(cvx.tre3, cvx.tre3$tip.label[grep(c("MH360"), cvx.tre3$tip.label)]): object 'cvx.t:

```
# Adjust outer margin to avoid overplotting, leaking off the edge
par(oma=c(3,3,3,3))

plot(ladderize(tip.rename.host(cvx.tre4,cvx.data)), type="f",cex=0.65, align.tip.label = TRUE)
```

Error in tip.rename.host(cvx.tre4, cvx.data): object 'cvx.tre4' not found

```
plot(ladderize(tip.rename.org(cvx.tre4,cvx.data)), type="f",cex=0.65, align.tip.label = TRUE)
```

Error in tip.rename.org(cvx.tre4, cvx.data): object 'cvx.tre4' not found

```
plot(ladderize(cvx.tre4), type="f",cex=0.65, align.tip.label = TRUE)
```

Error in reorder(phy): object 'cvx.tre4' not found

```
plot(ladderize(tip.rename.org(cvx.tre4,cvx.data)), type="p",cex=0.65, label.offset=0.001)
```

Error in tip.rename.org(cvx.tre4, cvx.data): object 'cvx.tre4' not found

```
plot(ladderize(tip.rename.host(cvx.tre4,cvx.data)), type="p",cex=0.65, label.offset=0.001)
```

Error in tip.rename.host(cvx.tre4, cvx.data): object 'cvx.tre4' not found

```
plot(ladderize(cvx.tre4), type="p",cex=0.65, label.offset=0.001)
```

Error in reorder(phy): object 'cvx.tre4' not found

```
pdf("/trees-phylograms.pdf",
    width=10,
    height=10,
    pointsize=12,
    bg='transparent',
    )
par(oma=c(3,3,3,3))
plot(ladderize(tip.rename.host(cvx.tre4,cvx.data)), type="f",cex=0.65, align.tip.label = TRUE)
plot(ladderize(tip.rename.org(cvx.tre4,cvx.data)), type="f",cex=0.65, align.tip.label = TRUE)
plot(ladderize(cvx.tre4), type="f",cex=0.65, align.tip.label = TRUE)
dev.off()

pdf("/trees-phylograms.pdf",
    width=8.5,
    height=11,
    pointsize=12,
    bg='transparent',
    )
#par(oma=c(3,3,3,3))
plot(ladderize(tip.rename.org(cvx.tre4,cvx.data)), type="p",cex=0.65, align.tip.label = TRUE)
plot(ladderize(tip.rename.host(cvx.tre4,cvx.data)), type="p",cex=0.65, align.tip.label = TRUE)
plot(ladderize(cvx.tre4), type="p",cex=0.65, align.tip.label = TRUE)
dev.off()
```