MERGE DESJARDINS AND ASHTON TREES

Claudia Zirión-Martínez

Setup

```r
library(RRphylo)
library(manipulate)
library(ape)
library(phytools)
library(ggtree)
library(tidyverse)
library(RColorBrewer)
library(ggnewscale)
```

Metadata

Make separate dataframes for each metadata field.

```r
metadata <- read.delim(  "../../data/derived/metadata_fixed.csv", header=TRUE, sep=",")

metadata$vni_subdivision[metadata$vni_subdivision == ""] <- NA
metadata$vni_subdivision <- factor(metadata$vni_subdivision,
                            levels = c("VNIa-4", "VNIa-5", "VNIa-32",
                            "VNIa-93", "VNIa-X", "VNIa-Y", "VNIb",
                            "VNIc", "VNIa-outlier", "VNII"))

sublineage <- metadata %>%
                filter(lineage == "VNI")%>%
                select(strain, vni_subdivision)%>%
                column_to_rownames("strain")%>%
                droplevels()
lineage <- metadata %>%
            select(strain, lineage)%>%
            column_to_rownames("strain")
dataset <- metadata %>%
            select(strain, dataset)%>%
            column_to_rownames("strain")
source <- metadata %>%
            select(strain, source)%>%
            column_to_rownames("strain")
```

Make color vectors for all plots

```r
lineage_colors <- brewer.pal(8, "Dark2")[c(1, 2, 3, 4)]
names(lineage_colors) <- levels(as.factor(metadata$lineage))

sublineage_colors <- c(brewer.pal(12, "Set3")[c(1:10)], "white")
names(sublineage_colors) <- levels(sublineage$vni_subdivision)
```

Desjardins tree

Import the raw Desjardins tree

```
desj_tree_path <- "../../data/raw/CryptoDiversity_Desjardins_Tree.tre"
desj_tree <- read.tree(desj_tree_path)
```

Reroot the tree at the middle of the branch leading to VNII

```
VNII_root <- getMRCA(desj_tree, c("C2","C12"))
edge_length <- subset(desj_tree$edge.length, desj_tree$edge[,2] == VNII_root)
desj_tree <- reroot(desj_tree, VNII_root, edge_length/2)
write.tree(desj_tree, file = "../../data/processed/desj_tree.newick")
```

Plot of Desjardins tree
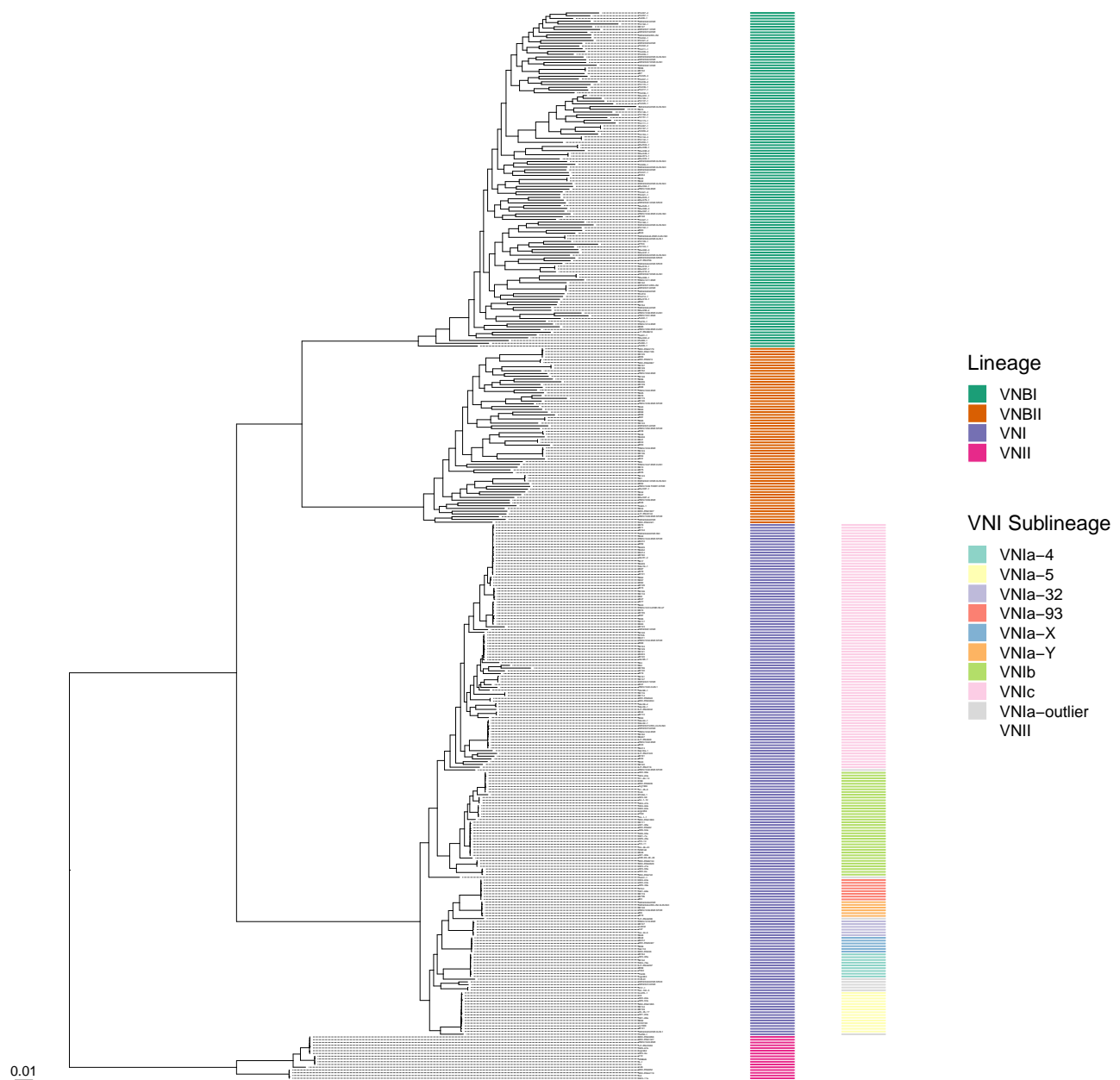
Ashton tree

Import the raw Ashton tree

```
ashton_tree_path <- "../../data/raw/2017.06.09.all_ours_and_desj.snp_sites.mod.fa.cln.tree"
ashton_tree_unrooted <- read.tree(ashton_tree_path)
```

Rename tips to use strain names

```
ashton_tree_unrooted$tip.label <- sapply(ashton_tree_unrooted$tip.label, function(x) {
    if (x %in% metadata$run) {
        metadata$strain[metadata$run == x]
    } else {
        x
    }
})
ashton_tree_unrooted$tip.label <- sapply(ashton_tree_unrooted$tip.label, function(x) {
    if (x %in% metadata$name) {
        metadata$strain[metadata$name == x]
    } else {
        x
    }
})
```
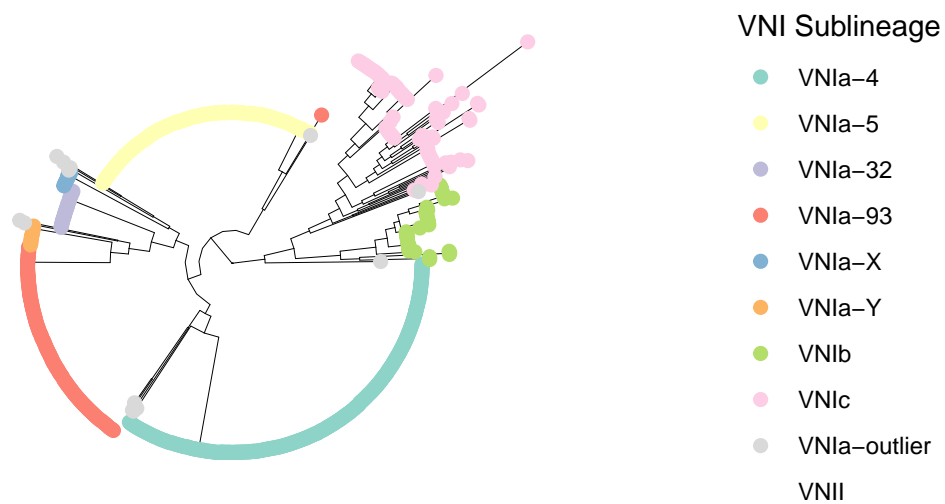
Plot unrooted Ashton tree
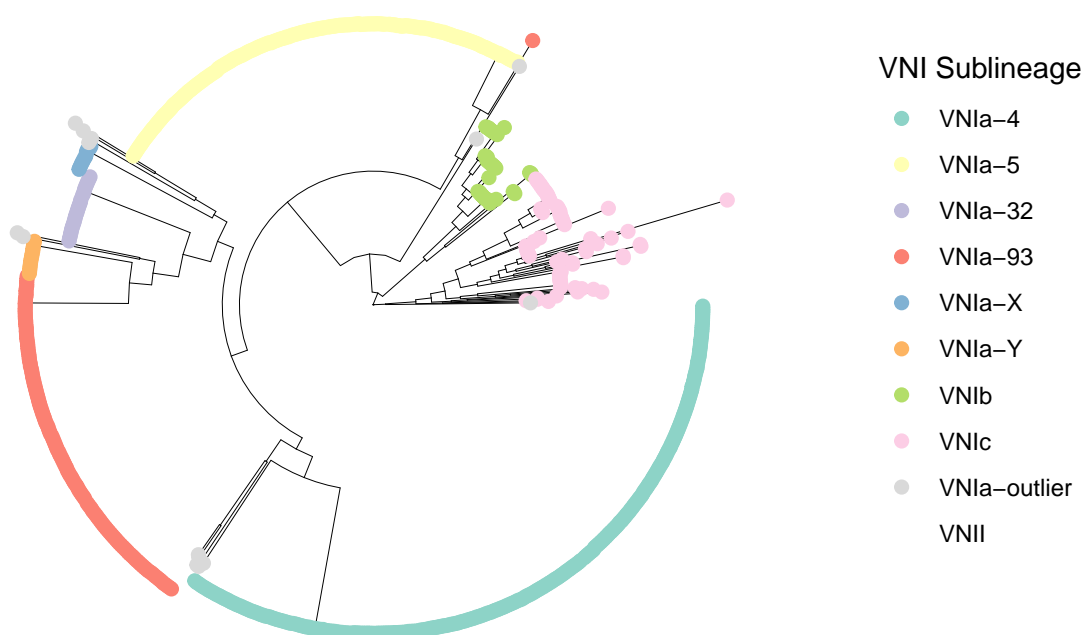
Root Ashton tree at the middle of the branch leading to VNIa

```
VNIa_root <- getMRCA(ashton_tree_unrooted, c("AD3-95a","Tu259-1"))
edge_length <- subset(ashton_tree_unrooted$edge.length,
    ashton_tree_unrooted$edge[,2] == VNIa_root)
ashton_tree <- reroot(ashton_tree_unrooted, VNIa_root, edge_length/2)
write.tree(ashton_tree, file = "../../data/processed/ashton_tree.newick")
```

Plot the rooted Ashton tree

Plot a rectangular version of the Ashton tree

VNI Sublineage

- VNIa−4
- VNIa−5
- VNIa−32
- VNIa−93
- VNIa−X
- VNIa−Y
- VNIb
- VNIc
- VNIa−outlier
- VNII

Merge Desjardins and Ashton trees

Specify clades in Desjardins tree

```r
VNI <- c("Bt92", "Bt79")
VNI_node <- getMRCA(desj_tree, VNI)
VNII <- c("C2","C12")
VNII_node <- getMRCA(desj_tree, VNII)
VNB <- c("Bt7", "Bt34")
VNB_node <- getMRCA(desj_tree, VNB)
```

Get the ages of the nodes from the original Desjardins tree

```r
edge_lengths <- node.depth.edgelength(desj_tree)
node_labels <- c(desj_tree$tip.label, desj_tree$node.label)
edge_length_mapping <- data.frame(
    node = node_labels,
    edge_length = edge_lengths,
    max_length = max(edge_lengths))
edge_length_mapping <- edge_length_mapping %>%
                       mutate(age = max_length - edge_length) %>%
                       rownames_to_column("node_id")
clade_ages <- edge_length_mapping %>%
              filter(node_id %in% c(VNI_node, VNII_node, VNB_node))
nodeages <- c("Bt92-Bt79" = clade_ages$age[clade_ages$node_id == VNI_node],
              "C2-C12" = clade_ages$age[clade_ages$node_id == VNII_node],
              "Bt7-Bt34" = clade_ages$age[clade_ages$node_id == VNB_node])
tip_ages <- edge_length_mapping %>%
            filter(node %in% metadata$strain)
tipages <- tip_ages$age
names(tipages) <- tip_ages$node
```

Remove VNI clade from Desjardins tree to use it as backtree

```r
VNI_tips <- tips(desj_tree, VNI_node)
backtree <- drop.tip(desj_tree, VNI_tips)
```
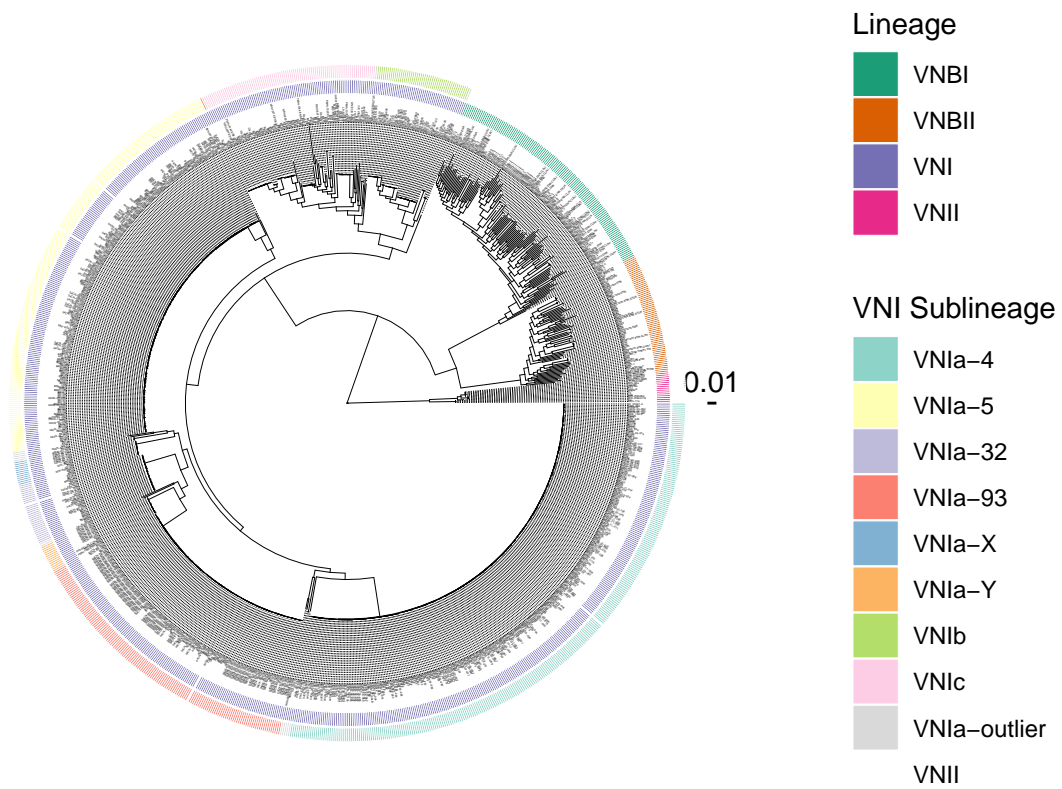
Create the reference tables

```r
reference <- data.frame(bind=c("CNS_289-BK8"),
                        reference=c("Bt7-Bt34"), # "H99"
                        poly=c(FALSE))
```

Merge

```r
merged <- tree.merger(backbone = backtree,
                      data=reference,
                      source.tree = ashton_tree,
                      plot=FALSE,
                      node.ages = nodeages,
                      tip.ages = tipages)
```
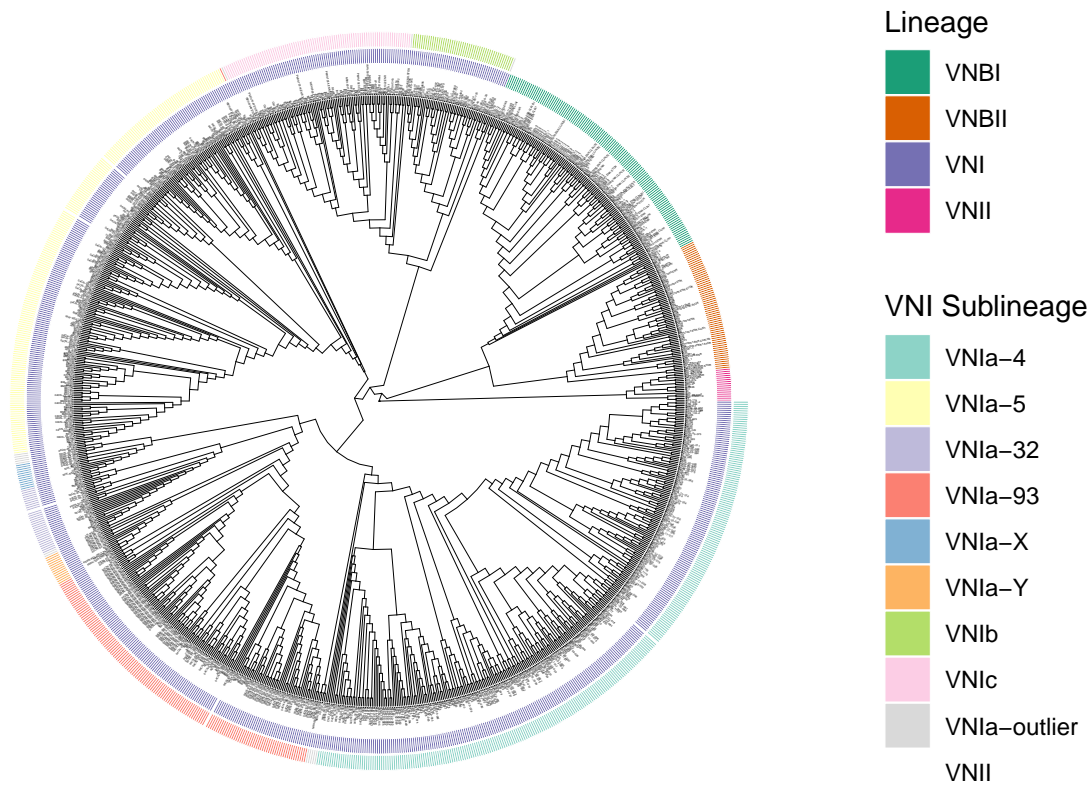
Plot merged tree with branchlengths (not real)

Lineage

- VNBI
- VNBII
- VNI
- VNII

VNI Sublineage

- VNIa−4
- VNIa−5
- VNIa−32
- VNIa−93
- VNIa−X
- VNIa−Y
- VNIb
- VNIc
- VNIa−outlier
- VNII

0.01

Plot cladogram of merged tree

Plot minimal version of the tree

Get one sample of each lineage, and VNI sublineage, and all VNIa-outlier

```
VNI <- metadata %>%
    filter(lineage == "VNI", vni_subdivision != "VNIa-outlier") %>%
    group_by(vni_subdivision) %>%
    slice(1) %>%
    ungroup()
VNIa_outlier <- metadata %>%
    filter(vni_subdivision == "VNIa-outlier")
VNII <- metadata %>%
    filter(lineage == "VNII") %>%
    slice(1) %>%
    ungroup()
VNBI <- metadata %>%
    filter(lineage == "VNBI") %>%
    slice(1) %>%
    ungroup()
VNBII <- metadata %>%
    filter(lineage == "VNBII") %>%
    slice(1) %>%
    ungroup()
tips <- rbind(VNI, VNIa_outlier, VNII, VNBI, VNBII)%>%
    select(strain)
```

Make a small version of the merged tree only with the tips in `tips`

```
small_tree <- drop.tip(merged, setdiff(merged$tip.label, tips$strain))
```

Plot the small tree