DETECT CHROMOSOMAL DUPLICATIONS

Claudia Zirión-Martínez

Setup

```
library(tidyverse)
library(ggtree)
library(ggtreeExtra)
library(ape)
library(ggnewscale)
library(RColorBrewer)
setwd("/FastData/czirion/Crypto_Diversity_Pipeline/analyses/tree_duplications/scripts")
```

Metadata

Load the necessary data

```
metadata <- read.csv(
    "../../data/processed/metadata_ashton_desj_all_fungalpop_H99.csv",
    header = TRUE)
```

Get one dataframe for each variable to be plotted as a separate metadata column in the tree

```
lineage <- metadata %>%
    select(strain, lineage)%>%
    column_to_rownames("strain")

source <- metadata %>%
    select(strain, source)%>%
    column_to_rownames("strain")

sublineage <- metadata %>%
    select(strain, vni_subdivision)%>%
    column_to_rownames("strain")

dataset <- metadata %>%
    select(strain, dataset)%>%
    column_to_rownames("strain")
```

Duplications

```
duplications <- read.delim(
    "../results/tables/duplications_polished.tsv",
    sep = "\t", header = TRUE, stringsAsFactors = TRUE)
```

```r
duplications_full <- duplications %>%
    select(strain, chromosome) %>%
    distinct()
```

Make matrix of duplicated chromosomes

```r
dup_chroms <- duplications_full %>%
    select(strain, chromosome)%>%
    mutate(duplicated_full = 1)%>%
    arrange(chromosome)%>%
    pivot_wider(names_from = chromosome, values_from = duplicated_full, values_fill = 0)%>%
    column_to_rownames("strain")%>%
    mutate(across(everything(), ~ ifelse(. == 1, cur_column(),"Euploid")))

euploid_strain <- metadata %>%
    filter(!strain %in% duplications_full$strain)%>%
    select(strain)

for (chrom in colnames(dup_chroms)){
    euploid_strain[chrom] <- "Euploid"
}

dup_chroms <- euploid_strain %>%
    column_to_rownames("strain") %>%
    bind_rows(dup_chroms)
```

Tree

```r
merged_tree_path <- "/FastData/czirion/Crypto_Diversity_Pipeline/analyses/data/processed/merged_tree.n
tree <- read.tree(merged_tree_path)
```

Remove tips that are not in metadata$strain

```r
tree <- drop.tip(tree, setdiff(tree$tip.label, metadata$strain))
```

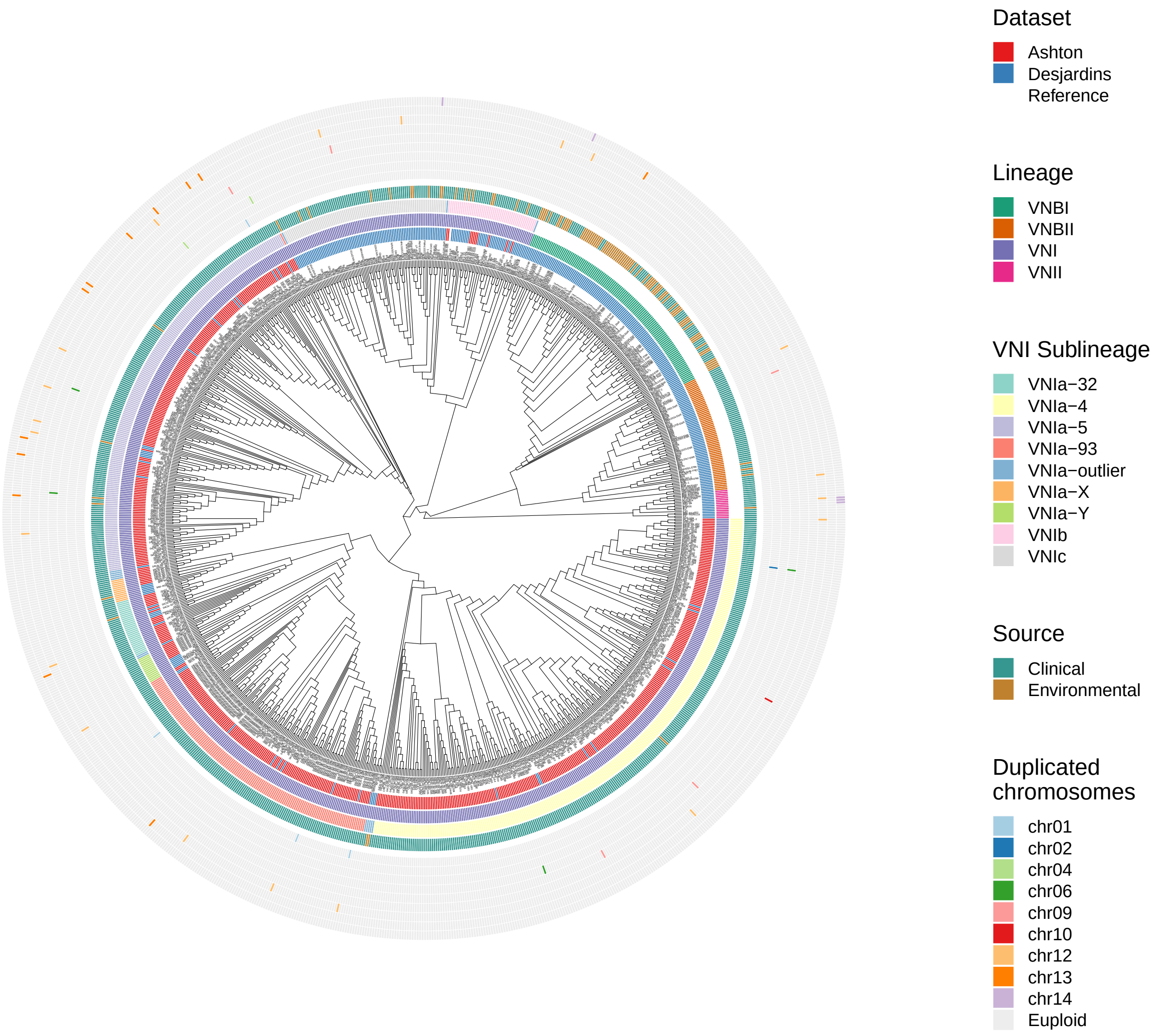
Plots

Create vectors of colors for each metadata variable with the name of the levels as names of the colors

```r
dataset_colors <- c(brewer.pal(9, "Set1")[c(1, 2)], "white")
names(dataset_colors) <- levels(as.factor(dataset$dataset))

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

sublineage_colors <- c(brewer.pal(12, "Set3")[c(1:9)])
names(sublineage_colors) <- levels(sublineage$vni_subdivision)

source_colors <- brewer.pal(11, "BrBG")[c(9, 3)] # 9, 3 are the colors for the two sources
```

```r
names(source_colors) <- levels(as.factor(source$source))

chrom_colors <- c(brewer.pal(nlevels(duplications$chromosome), "Paired"), "grey93")
names(chrom_colors) <- c(levels(duplications$chromosome), "Euploid")
```

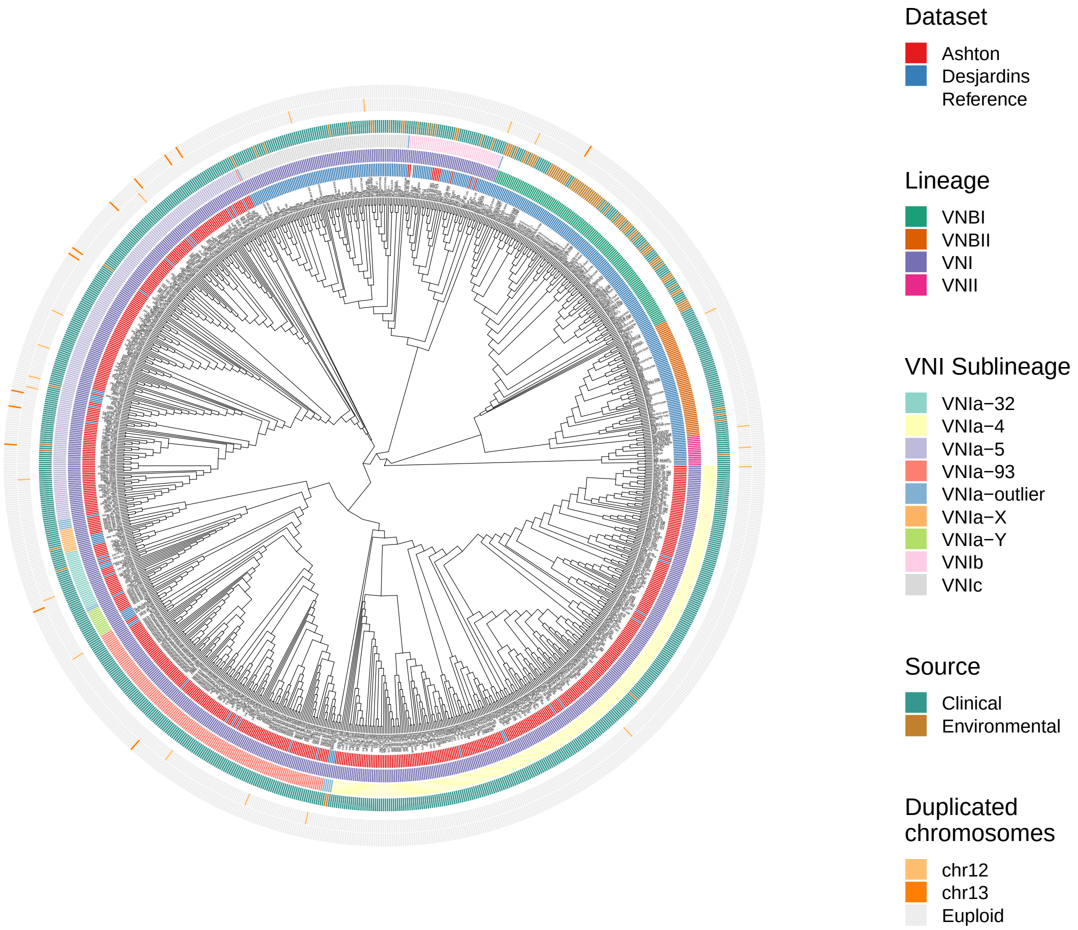
Tree of all samples with duplications of all chromosomes



Tree of all samples with duplications of chromosomes 12 and 13

Subset the duplications_full data frame to only include strains with duplications of chromosomes 12 and 13

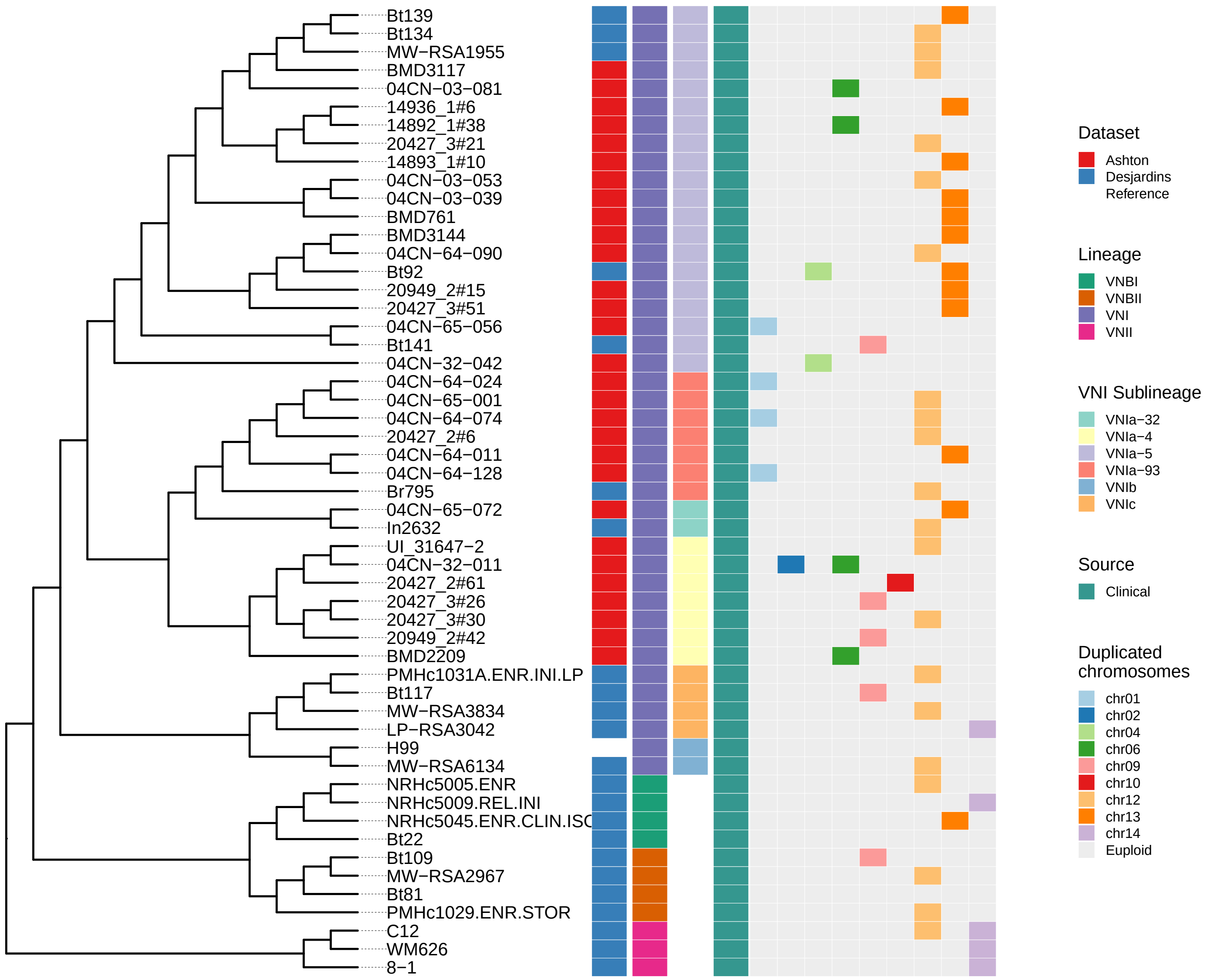
```r
dup_chroms_12_13 <- dup_chroms %>%
    select(chr12, chr13)
```



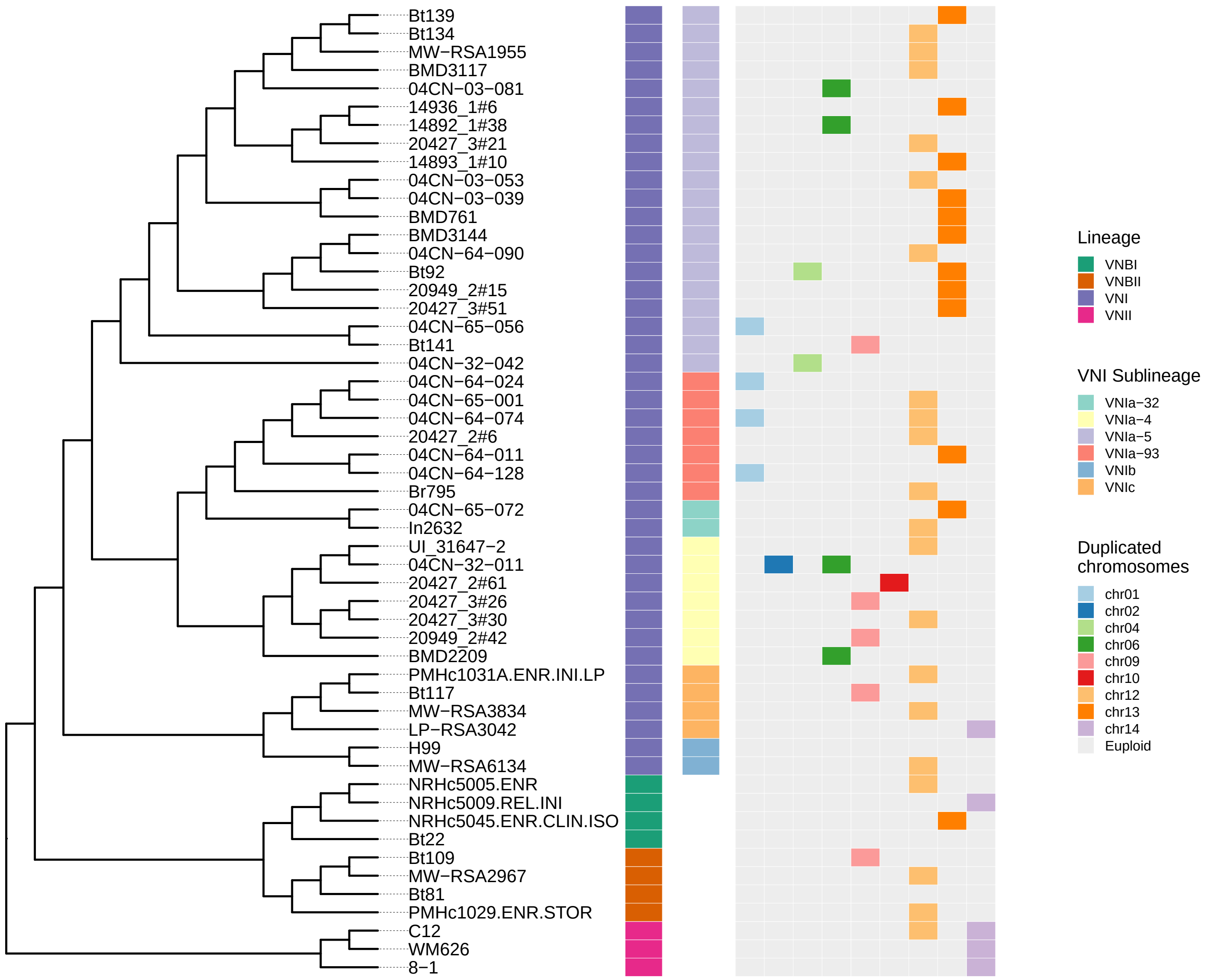
Tree with only the samples that have duplications and the references

```r
keep_strains <- c(levels(duplications_full$strain), "H99", "Bt22", "Bt81")
tree_dups <- drop.tip(tree, setdiff(tree$tip.label, keep_strains))
```

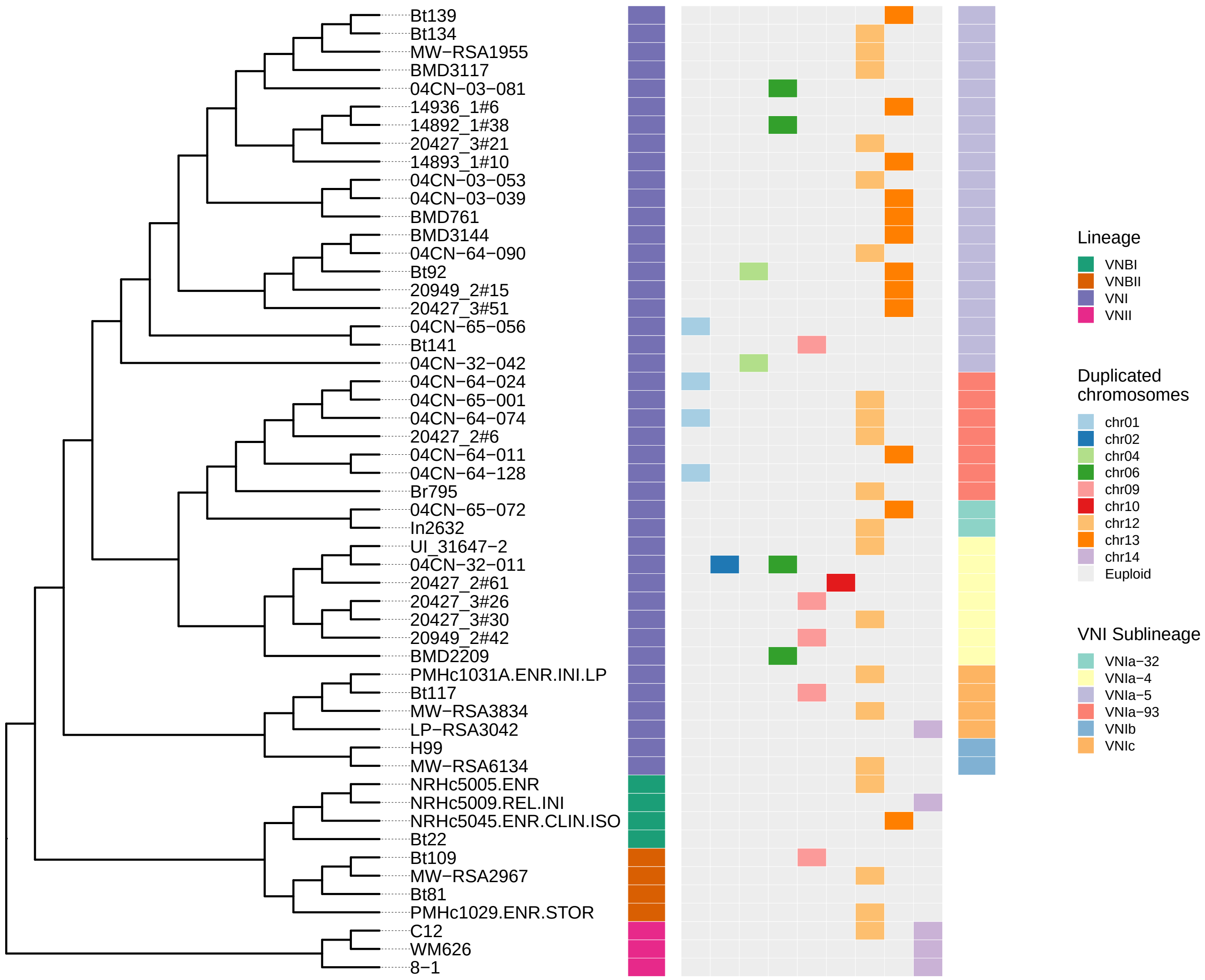
Dataset, lineage, sublineage, source, duplications



Lineage, sublineage, duplications



Lineage, duplications, sublineage


Bt139
Bt134
MW−RSA1955
BMD3117
04CN−03−081
14936_1#6
14892_1#38
20427_3#21
14893_1#10
04CN−03−053
04CN−03−039
BMD761
BMD3144
04CN−64−090
Bt92
20949_2#15
20427_3#51
04CN−65−056
Bt141
04CN−32−042
04CN−64−024
04CN−65−001
04CN−64−074
20427_2#6
04CN−64−011
04CN−64−128
Br795
04CN−65−072
In2632
UI_31647−2
04CN−32−011
20427_2#61
20427_3#26
20427_3#30
20949_2#42
BMD2209
PMHc1031A.ENR.INI.LP
Bt117
MW−RSA3834
LP−RSA3042
H99
MW−RSA6134
NRHc5005.ENR
NRHc5009.REL.INI
NRHc5045.ENR.CLIN.ISO
Bt22
Bt109
MW−RSA2967
Bt81
PMHc1029.ENR.STOR
C12
WM626
8−1
Lineage
VNBI
VNBII
VNI
VNII
Duplicated chromosomes
chr01
chr02
chr04
chr06
chr09
chr10
chr12
chr13
chr14
Euploid
VNI Sublineage
VNIa−32
VNIa−4
VNIa−5
VNIa−93
VNIb
VNIc