Guillaume

```
library("RIdeogram")
```

```
## Warning: le package 'RIdeogram' a été compilé avec la version R 4.1.3
```

```
library("dplyr")
```

```
## Warning: le package 'dplyr' a été compilé avec la version R 4.1.3
```

```
##
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

```
## Warning: le package 'tidyr' a été compilé avec la version R 4.1.3
```

```
library(plyr)
```

```
## Warning: le package 'plyr' a été compilé avec la version R 4.1.3
```

```
## ------------------------------------------------------------------------------
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
```

```
## ------------------------------------------------------------------------------
```

```
##
## Attachement du package : 'plyr'
```

```
## Les objets suivants sont masqués depuis 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
library(stringr)
```

```
## Warning: le package 'stringr' a été compilé avec la version R 4.1.3
```

Reading the data

```
gene_mapping <- read.csv('gene_mapping.tsv', sep='\t', header=TRUE)
dongola <- read.csv("DONGOLA_genes.tsv", sep='\t', header=TRUE)
zanu <- read.csv("ZANU_genes.tsv", sep='\t', header=TRUE)
```

```
head(gene_mapping)
```

```
##   contig middle.position strand ord      name ref.genes
## 1      2           31135     -1   0 gene_3542         1
## 2      2           38868     -1   1 gene_3543         1
## 3      2           42746      1   2   gene_80         1
## 4      2           46243     -1   3 gene_3544         1
## 5      2           53442     -1   4 gene_3545         1
## 6      2           60574      1   5   gene_81         1
##                                                    DONG
## 1  NC_053517.1,111908344,1,6540,DONG_gene-LOC120894913
## 2  NC_053517.1,111899667,1,6539,DONG_gene-LOC120904110
## 3 NC_053517.1,111895084,-1,6538,DONG_gene-LOC120904105
## 4  NC_053517.1,111891588,1,6537,DONG_gene-LOC120904096
## 5  NC_053517.1,111884408,1,6536,DONG_gene-LOC120895288
## 6 NC_053517.1,111877309,-1,6535,DONG_gene-LOC120895290
```

Making the karyotype table

```
karyotype <- setNames(data.frame(matrix(ncol=7, nrow=0)), c("Chr", "Start", "End", "fill", "species", "
karyotype <- rbind(karyotype, data.frame(Chr=c('X','2','3'), Start=c(1, 1, 1), End=c(27238055, 11478317
karyotype <- rbind(karyotype, data.frame(Chr=c('X','2','3'), Start=c(1, 1, 1), End=c(26910000, 11199000
```

Preparing DONG column

```
gene_mapping <- separate(data=gene_mapping, col=DONG, into=c("seq_id_dong", "mid_dong", 'strand_dong',
```

Choose in contig column only 2, 3, X chromosomes

```
gene_mapping <- gene_mapping[gene_mapping$contig %in% c('2', '3', 'X'),]
```

```
head(gene_mapping)
```

```
##   contig middle.position strand ord      name ref.genes seq_id_dong  mid_dong
## 1      2           31135     -1   0 gene_3542         1 NC_053517.1 111908344
## 2      2           38868     -1   1 gene_3543         1 NC_053517.1 111899667
## 3      2           42746      1   2   gene_80         1 NC_053517.1 111895084
## 4      2           46243     -1   3 gene_3544         1 NC_053517.1 111891588
## 5      2           53442     -1   4 gene_3545         1 NC_053517.1 111884408
## 6      2           60574      1   5   gene_81         1 NC_053517.1 111877309
##   strand_dong len_dong             name_dong
## 1           1     6540 DONG_gene-LOC120894913
## 2           1     6539 DONG_gene-LOC120904110
## 3          -1     6538 DONG_gene-LOC120904105
## 4           1     6537 DONG_gene-LOC120904096
## 5           1     6536 DONG_gene-LOC120895288
## 6          -1     6535 DONG_gene-LOC120895290
```

###Perform mapping between chromosomes names and sequences IDs ###From NCBI: ###Chr 2 : NC_053517.1 ###Chr 3 : NC_053518.1
###Chr X : NC_053519.1

```
gene_mapping$seq_id_dong[gene_mapping$seq_id_dong == 'NC_053517.1'] <- '2'
gene_mapping$seq_id_dong[gene_mapping$seq_id_dong == 'NC_053518.1'] <- '3'
gene_mapping$seq_id_dong[gene_mapping$seq_id_dong == 'NC_053519.1'] <- 'X'
head(gene_mapping)
```

```
##   contig middle.position strand ord      name ref.genes seq_id_dong  mid_dong
## 1      2           31135     -1   0 gene_3542         1           2 111908344
## 2      2           38868     -1   1 gene_3543         1           2 111899667
## 3      2           42746      1   2   gene_80         1           2 111895084
## 4      2           46243     -1   3 gene_3544         1           2 111891588
## 5      2           53442     -1   4 gene_3545         1           2 111884408
## 6      2           60574      1   5   gene_81         1           2 111877309
##   strand_dong len_dong             name_dong
## 1           1     6540 DONG_gene-LOC120894913
## 2           1     6539 DONG_gene-LOC120904110
## 3          -1     6538 DONG_gene-LOC120904105
## 4           1     6537 DONG_gene-LOC120904096
## 5           1     6536 DONG_gene-LOC120895288
## 6          -1     6535 DONG_gene-LOC120895290
```

Choosing only 2, 3, X chromosomes in DONGOLA

```
gene_mapping <- gene_mapping[gene_mapping$seq_id %in% c('2', '3', 'X'),]
```

Removing DONG from gene names

```
gene_mapping$name_dong <- gsub("^DONG_(\\w+)", "\\1", gene_mapping$name_dong)
head(gene_mapping)
```

```
##   contig middle.position strand ord      name ref.genes seq_id_dong  mid_dong
## 1      2           31135     -1   0 gene_3542         1           2 111908344
## 2      2           38868     -1   1 gene_3543         1           2 111899667
## 3      2           42746      1   2   gene_80         1           2 111895084
## 4      2           46243     -1   3 gene_3544         1           2 111891588
## 5      2           53442     -1   4 gene_3545         1           2 111884408
## 6      2           60574      1   5   gene_81         1           2 111877309
##   strand_dong len_dong         name_dong
## 1           1     6540 gene-LOC120894913
## 2           1     6539 gene-LOC120904110
## 3          -1     6538 gene-LOC120904105
## 4           1     6537 gene-LOC120904096
## 5           1     6536 gene-LOC120895288
## 6          -1     6535 gene-LOC120895290
```

Remove duplicated genes

```
gene_mapping <- gene_mapping[!duplicated(gene_mapping$name),]
```

Synteny table

```
colnames(zanu) <- c('ID_1', 'Start_1', 'End_1', 'Strand_1')
colnames(dongola) <- c('ID_2', 'Start_2', 'End_2', 'Strand_2')
synteny_table <- merge(gene_mapping, zanu, by.x='name', by.y='ID_1')
synteny_table <- merge(synteny_table, dongola, by.x='name_dong', by.y='ID_2')
names(synteny_table)[names(synteny_table) == 'contig'] <- 'Species_1'
names(synteny_table)[names(synteny_table) == 'seq_id_dong'] <- 'Species_2'
synteny_table$Species_1 <-mapvalues(synteny_table$Species_1,
                                    from=c('X', '2', '3'),
                                    to=c(1, 2, 3))
synteny_table$Species_2 <-mapvalues(synteny_table$Species_2,
                                    from=c('X', '2', '3'),
                                    to=c(1, 2, 3))
synteny_table$Species_1 <- as.integer(synteny_table$Species_1)
synteny_table$Species_2 <- as.integer(synteny_table$Species_2)
head(synteny_table)
```

```
##           name_dong      name Species_1 middle.position strand  ord ref.genes
## 1 gene-LOC120893177 gene_5019         2        48531603     -1 2862         1
## 2 gene-LOC120893178 gene_6182         2        86040949     -1 5204         1
## 3 gene-LOC120893179 gene_2643         2        86040395      1 5203         1
## 4 gene-LOC120893180 gene_5313         2        58398932     -1 3461         1
## 5 gene-LOC120893183 gene_2537         2        82790246      1 4995         1
## 6 gene-LOC120893185 gene_6082         2        82797727     -1 4998         1
##   Species_2 mid_dong strand_dong len_dong  Start_1    End_1 Strand_1  Start_2
```

```
## 1         2 65514822           1     3925 48528403 48534803       -1 65511152
## 2         2 28681053           1     1788 86040710 86041188       -1 28680597
## 3         2 28681607          -1     1789 86040192 86040598        1 28681316
## 4         2 55921684           1     3534 58381587 58416277       -1 55853085
## 5         2 31941591          -1     1998 82789431 82791062        1 31940683
## 6         2 31934112           1     1995 82796508 82798947       -1 31932898
##      End_2 Strand_2
## 1 65519724        1
## 2 28681368        1
## 3 28681908       -1
## 4 55941166        1
## 5 31942410       -1
## 6 31935462        1
```

```
red <- 'FF0000'
blue <- '5891bf'
dong_max_2 <- 111990000
dong_max_3 <- 95710000
color <- function(strand1, strand2, red, blue){
  if (strand1 == strand2)
    return(red)
  else
    return(blue)
}
synteny_table$fill <- mapply(color,
                             synteny_table$Strand_1,
                             synteny_table$Strand_2,
                             red,
                             blue)
# inverse forc hr 2 ad chr3
two_to_three_color <- function(chr1, strand1, strand2, prev_fill, red, blue){
  if (chr1 == 2 || chr1 == 3){
    if (strand1 == strand2)
      return(red)
    else
      return(blue)
  }
  return(prev_fill)
}
synteny_table$fill <- mapply(two_to_three_color,
                             synteny_table$Species_1,
                             synteny_table$Strand_1,
                             synteny_table$Strand_2,
                             synteny_table$fill,
                             red,
                             blue)
two_to_three <- function(chr1, pos2, dong_max_2, dong_max_3){
  if (chr1 == 2 || chr1 == 3){
    if (chr1 == 2)
      return(dong_max_2 - pos2 + 1)
    else
      return(dong_max_3 - pos2 + 1)
  }
  return(pos2)
```

```
}
synteny_table$Start_2 <- mapply(two_to_three,
                                synteny_table$Species_1,
                                synteny_table$Start_2,
                                dong_max_2,
                                dong_max_3)
synteny_table$End_2 <- mapply(two_to_three,
                              synteny_table$Species_1,
                              synteny_table$End_2,
                              dong_max_2,
                              dong_max_3)
synteny_table <- synteny_table[c('Species_1', 'Start_1', 'End_1', 'Species_2', 'Start_2', 'End_2', 'fil
synteny_table <- synteny_table[synteny_table$Species_1==synteny_table$Species_2, ]
head(synteny_table)
```

```
##   Species_1  Start_1    End_1 Species_2  Start_2    End_2   fill
## 1         2 48528403 48534803         2 46478849 46470277 5891bf
## 2         2 86040710 86041188         2 83309404 83308633 5891bf
## 3         2 86040192 86040598         2 83308685 83308093 5891bf
## 4         2 58381587 58416277         2 56136916 56048835 5891bf
## 5         2 82789431 82791062         2 80049318 80047591 5891bf
## 6         2 82796508 82798947         2 80057103 80054539 5891bf
```

Generating the .svg and converting it in.png

```
ideogram(karyotype=karyotype, synteny=synteny_table)
convertSVG("chromosome.svg", device="png")
```