HW3_Antonets

Antonets Svetlana

07 06 2022

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

```
## Warning: пакет 'RIdeogram' был собран под R версии 4.1.2
```

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

```
##
## Присоединяю пакет: 'dplyr'
```

```
## Следующие объекты скрыты от 'package:stats':
##
##     filter, lag
```

```
## Следующие объекты скрыты от 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

#data obtaining

```
gene_mapping <-read.csv("gene_mapping.tsv", sep='\t')
DONGOLA <- read.csv("DONGOLA_genes.tsv", sep='\t')
ZANU <- read.csv("ZANU_genes.tsv", sep='\t')
```

#info separation

```
gene_mapping <- gene_mapping %>% separate(DONG, into = c("sequence_id", "middle_co
ordinate", "strand_Dong", "length_gene", "gene_name_Dong"), sep = ",")
```

#Removal of not 2,3 and X chromosomes for ZANU and rename DONGOLA's chromosomes

```
gene_mapping <- gene_mapping %>% filter(contig == "2"|contig == "3"|contig == "X")
gene_mapping$sequence_id[gene_mapping$sequence_id=="NC_053517.1"]<-"2"
gene_mapping$sequence_id[gene_mapping$sequence_id=="NC_053518.1"]<-"3"
gene_mapping$sequence_id[gene_mapping$sequence_id=="NC_053519.1"]<-"X"
gene_mapping <- gene_mapping %>% filter(sequence_id == "2"|sequence_id == "3"|sequ
ence_id == "X")
gene_mapping$gene_name_Dong <- gsub("DONG_","",as.character(gene_mapping$gene_name
_Dong))
```

#table construction

```
karyotype_dual_comparison <- data.frame(matrix(ncol = 7, nrow = 0))
karyotype_dual_comparison <- setNames(karyotype_dual_comparison,
                                      c("Chr", "Start", "End", "fill",
                                        "species", "size", "color"))
karyotype_dual_comparison <- rbind(karyotype_dual_comparison,
                                   data.frame(Chr=c('X','2','3'), Start=c(1, 1, 1),
                                              End=c(27238055, 114783175, 97973315),
                                          fill='969696',
                                              species='ZANU', size=12, color='25252
5'))
karyotype_dual_comparison <- rbind(karyotype_dual_comparison,
                                   data.frame(Chr=c('X','2','3'), Start=c(1, 1, 1),
                                              End=c(26910000, 111990000, 95710000),
                                          fill='969696',
                                              species='DONGOLA', size=12, color='25
2525'))
```

#Counting of distance and sorting. Then removal of duplicates will remain the first variant for each repeated gene ZANU - with smaller distance between mapped genes

```
gene_mapping$middle_coordinate <- as.numeric(gene_mapping$middle_coordinate)
gene_mapping$distance <- abs(gene_mapping$middle.position - gene_mapping$middle_co
ordinate)
gene_mapping <- gene_mapping[order(gene_mapping$distance),]
gene_mapping <- distinct(gene_mapping, name, .keep_all= TRUE)
```

#function for table construction

```
synteny_table_construction <- function(gene_mapping, ZANU, DONGOLA) {
  synteny_dual_comparison <- setNames(data.frame(matrix(ncol = 7, nrow = 0)),
                                      c("Species_1", "Start_1", "End_1", "Species_2",
                                        "Start_2", "End_2", "fill"))
  for(i in 1:nrow(gene_mapping)) {
    mapped_gene <- gene_mapping[i, ]
    ZANU_gene = mapped_gene$name[1]
    DONGOLA_gene = mapped_gene$gene_name_Dong[1]
    chr_ZANU = mapped_gene$contig[1]
    chr_DONGOLA = mapped_gene$sequence_id[1]
    ZANU_mapped <- ZANU[ZANU$ID == ZANU_gene,]
    DONGOLA_mapped <- DONGOLA[DONGOLA$ID == DONGOLA_gene,]
    if (chr_ZANU == chr_DONGOLA) {
      if (ZANU_mapped$strand[1] == DONGOLA_mapped$strand[1]) {
        color_to_fill <- 'ffb7c5'
      }
      else {
        color_to_fill <- '4682bf'
      }
      if (chr_ZANU == "X") {
        synteny_dual_comparison <- rbind(synteny_dual_comparison,
                                   data.frame(Species_1 = 'X', Start_1=ZANU_mapped$st
art,
                                              End_1=ZANU_mapped$end,
                Species_2 = 'X', Start_2=DONGOLA_mapped$start, End_2=DONGOLA_mapped
$end,
```

```
                fill=color_to_fill))
    }
        else if (chr_ZANU == "2"){
          synteny_dual_comparison <- rbind(synteny_dual_comparison,
                                data.frame(Species_1 = '2', Start_1=ZANU_mapped$st
art,
                                           End_1=ZANU_mapped$end,
                Species_2 = '2', Start_2=111990000-DONGOLA_mapped$start, End_2=1119
90000-DONGOLA_mapped$end,
                fill=color_to_fill))
      }
        else if (chr_ZANU == "3"){
          synteny_dual_comparison <- rbind(synteny_dual_comparison,
                                data.frame(Species_1 = '3', Start_1=ZANU_mapped$st
art,
                                           End_1=ZANU_mapped$end,
                Species_2 ='3', Start_2=95710000-DONGOLA_mapped$start, End_2=957100
00-DONGOLA_mapped$end,
                fill=color_to_fill))
        }
    }
    else {
      next
    }
  }
  return (synteny_dual_comparison)
}
```

#table construction

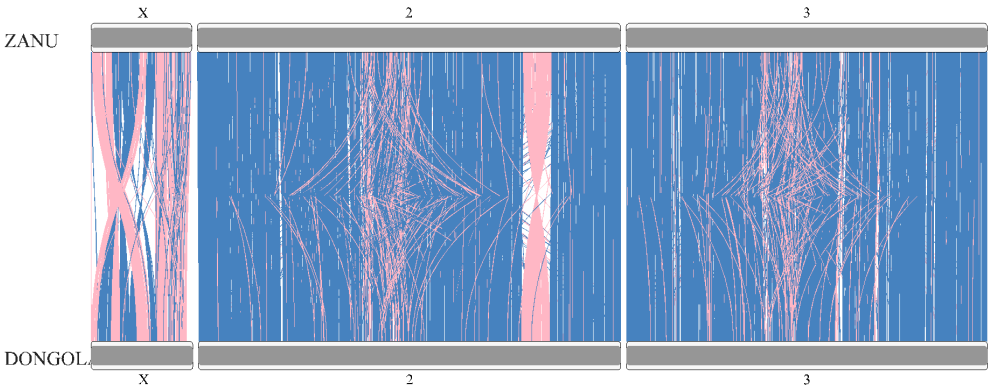
```
synteny_dual_comparison <- synteny_table_construction(gene_mapping, ZANU, DONGOLA)
```

#table preparation

```
synteny_dual_comparison$Species_1 <- gsub("X","1",as.character(synteny_dual_compar
ison$Species_1))
synteny_dual_comparison$Species_2 <- gsub("X","1",as.character(synteny_dual_compar
ison$Species_2))
synteny_dual_comparison$Species_1 <- as.numeric(synteny_dual_comparison$Species_1)
synteny_dual_comparison$Species_2 <- as.numeric(synteny_dual_comparison$Species_2)
```

#Building of graph

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


X
2
3
ZANU
DONGOL
X
2
3