HW3

Alevtina Bogoliubova

```
#install.packages("rsvg")
#install.packages("RIdeogram")
```

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("tidyr")
```

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

head(dongola)
```

```
##                   ID start   end strand
## 1 gene-LOC120906950 59885 60345     -1
## 2 gene-LOC120906947 61728 64249      1
## 3 gene-LOC120906949 88010 88555     -1
## 4 gene-LOC120906948 90190 90789     -1
## 5 gene-LOC120906980   657  1316     -1
## 6 gene-LOC120906964 23986 24588      1
```

```
head(zanu)
```

```
##           ID  start    end strand
## 1 gene_13164   5022  23194     -1
## 2 gene_13165  40014  45938     -1
## 3 gene_13166  92876  97357     -1
## 4 gene_12497  99657 102434      1
## 5 gene_13167 106482 122413     -1
## 6 gene_13168 129453 131721     -1
```

```
head(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
```

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

mapping <- separate(data=mapping, col=DONG, into=c("dong_id", "dong_mid", 'dong_strand', 'dong_len', 'do

gene_id_map = data.frame(id=c('2',"3","X"), val=c('NC_053517.1', 'NC_053518.1', 'NC_053519.1'))
mapping$dong_id <- with(gene_id_map, id[match(mapping$dong_id, val)])
head(mapping)
```

```
##   contig middle.position strand ord      name ref.genes dong_id  dong_mid
## 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
##   dong_strand dong_len             dong_name
## 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
```

```
mapping <- mapping[mapping$dong_id %in% chr,]

mapping$dong_name <- as.character(lapply(mapping$dong_name, gsub, pattern = '^DONG_', replacement =""))
mapping$distance <- abs(mapping$middle.pos - as.numeric(mapping$dong_mid))
```

```
karyotype_zanu <- data.frame(
  c('X', 2, 3),
  c(1, 1, 1),
  c(27238055, 114783175, 97973315),
  c(229926, 969696, 969696),
  c('ZANU', 'ZANU', 'ZANU'),
  c(12, 12, 12),
  c(252525, 252525, 252525))
colnames(karyotype_zanu) <- c('Chr','Start','End','fill','species','size','color')

karyotype_dongola <- data.frame(
  c('X', 2, 3),
  c(1, 1, 1),
  c(26913133, 111988354, 95710210),
```

```
  c(229926, 969696, 969696),
  c('DONGOLA', 'DONGOLA', 'DONGOLA'),
  c(12, 12, 12),
  c(252525, 252525, 252525))
colnames(karyotype_dongola) <- c('Chr','Start','End','fill','species','size','color')

karyotype_table <- rbind(karyotype_zanu, karyotype_dongola)
head(karyotype_table)
```

```
##   Chr Start       End   fill species size  color
## 1   X     1  27238055 229926    ZANU   12 252525
## 2   2     1 114783175 969696    ZANU   12 252525
## 3   3     1  97973315 969696    ZANU   12 252525
## 4   X     1  26913133 229926 DONGOLA   12 252525
## 5   2     1 111988354 969696 DONGOLA   12 252525
## 6   3     1  95710210 969696 DONGOLA   12 252525
```

```
dong_2_end = 111988354
dong_3_end = 95710210
dong_X_end = 26913133
```

```
mapping$contig[mapping$contig == "X"] <- 1
mapping$dong_id[mapping$dong_id == "X"] <- 1

blue = "0540ca"
red = "d40000"

start_z <- c()
end_z <- c()
fill <- c()
for (i in (1:nrow(mapping))){
    name <- mapping[i, "name"]
    fill <- if (mapping[i, "strand"] == mapping[i, "dong_strand"]) append(fill, red)
    else append(fill, blue)
    start_z <- append(start_z, zanu[zanu$ID == name, "start"])
    end_z <- append(end_z, zanu[zanu$ID == name, "end"])
}
```

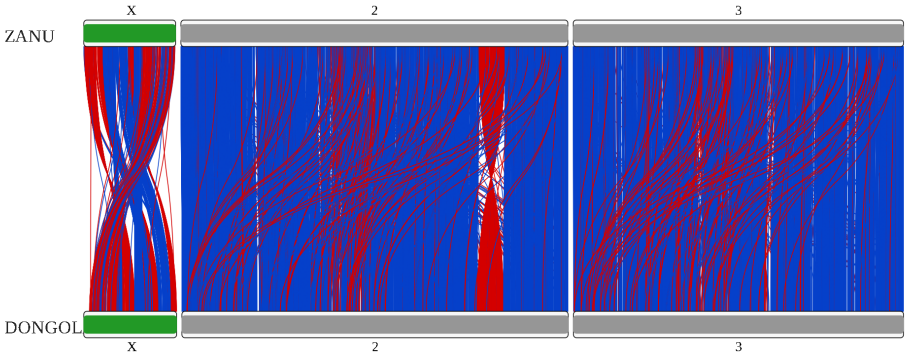
```
start_d <- c()
end_d <- c()
for (i in (1:nrow(mapping))){
    name <- mapping[i, "dong_name"]
    if (mapping[i, "contig"] == 1){
    start <- dong_X_end - dongola[dongola$ID == name, "start"]
    end <- dong_X_end - dongola[dongola$ID == name, "end"]
    } else if ((mapping[i, "contig"] == 2)){
      start <- dong_2_end - dongola[dongola$ID == name, "start"]
      end <- dong_2_end - dongola[dongola$ID == name, "end"]
    } else {
      start <- dong_3_end - dongola[dongola$ID == name, "start"]
      end <- dong_3_end - dongola[dongola$ID == name, "end"]
    }
  start_d <- append(start_d, start)
  end_d <- append(end_d, end)
}
```

```
synteny_table <- data.frame(Species_1 = as.numeric(mapping$contig),
                            Start_1 = start_z,
                            End_1 = end_z,
                            Species_2 = as.numeric(mapping$dong_id),
                            Start_2 = start_d,
                            End_2 = end_d,
                            fill = fill)

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   29035 33235         2   84130  78163 0540ca
## 2         2   37467 40269         2   90351  85137 0540ca
## 3         2   41638 43855         2   94470  91861 0540ca
## 4         2   44541 47945         2   99138  94685 0540ca
## 5         2   50702 56183         2  108404  99834 0540ca
## 6         2   58892 62256         2  113299 108976 0540ca
```

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


X
2
3
ZANU
DONGOL
X
2
3