Hw3

Anna Shchetsova

```
require(data.table)
```

```
## Loading required package: data.table
```

```
require(RIdeogram)
```

```
## Loading required package: RIdeogram
```

```
library(stringr)
```

Load data:

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

Prepare data:

```
colnames(gene_mapping) <-
  c('z_contig', 'z_middle_pos', 'z_strand', 'z_ord', 'z_id', 'ref.genes', 'DONG')
gene_mapping[c('d_contig', 'd_middle_pos', 'd_strand', 'd_ord', 'd_id')] <-
  str_split_fixed(gene_mapping$DONG, ',', 5)

gene_mapping[gene_mapping == 'NC_053517.1'] <- 2
gene_mapping[gene_mapping == 'NC_053518.1'] <- 3
gene_mapping[gene_mapping == 'NC_053519.1'] <- 1

gene_mapping$'z_contig'[gene_mapping$'z_contig' == 'X'] <- 1

gene_mapping <- gene_mapping[gene_mapping$z_contig == gene_mapping$d_contig,]

zanu <- zanu[c('ID', 'start', 'end')]
colnames(zanu) <- c('z_id', 'z_start', 'z_end')
gongola <- gongola[c('ID', 'start', 'end')]
colnames(gongola) <- c('d_id', 'd_start', 'd_end')

tmp <- gene_mapping[ c('z_contig', 'z_strand', 'z_id', 'd_contig', 'd_strand', 'd_id')]
tmp$d_id <- gsub("DONG_(*)", "\\1", tmp$d_id)
tmp$fill <- ifelse(tmp$z_strand == tmp$d_strand, '0000ff', 'ff0000')

tmp <- merge(x = tmp, y = zanu, by = "z_id", all.x = TRUE)
tmp <- merge(x = tmp, y = gongola, by = "d_id", all.x = TRUE)
```

Create synteny and karyotype dataFrame:

```r
synteny <-
  tmp[c('z_contig', 'z_start', 'z_end', 'd_contig', 'd_start', 'd_end', 'fill')]
colnames(synteny) <-
  c('Species_1', 'Start_1', 'End_1', 'Species_2', 'Start_2', 'End_2', 'fill')
synteny$Species_1 <- as.numeric(synteny$Species_1)
synteny$Species_2 <- as.numeric(synteny$Species_2)

karyotype <- data.frame(
  Chr = c('X', '2', '3', 'X', '2', '3'),
  Start = c(1, 1, 1, 1, 1, 1),
  End = c(27238055, 114783175, 97973315, 26913133, 111988354, 95710210),
  fill = c('969696', '969696', '969696', '969696', '969696', '969696'),
  species = c("ZANU", "ZANU", "ZANU", "DONGOLA", "DONGOLA", "DONGOLA"),
  size = c(12, 12, 12, 12, 12, 12),
  color = c('252525', '252525', '252525', '252525', '252525', '252525')
)
```

Visualization result:

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