HW 3

Code

Denis Fedorov

Hide

```
library(RIdeogram)
library(stringr)
```

Hide

```
dg <- read.csv("DONGOLA_genes.tsv", sep='\t')
zg <- read.csv("ZANU_genes.tsv", sep='\t')
gm <- read.csv("gene_mapping.tsv", sep='\t')

#procesing DONG string and removing excessive info
gm <- cbind(gm,str_split_fixed(gm$DONG, "[,]", 5))
colnames(gm) <- seq(1:12)
gm <- gm[,-which(names(gm) %in% c("4","5","6","7","12"))]

colnames(gm) <- c('Zchr','Zmid','Zstrand','Dchr','Dmid','Dstrand','Dlen')
head(gm)
```

	Zchr <chr>	Zmid <int>	Zstrand <int>	Dchr <chr>	Dmid <chr>	Dstrand <chr>	Dlen <chr>
1	2	31135	-1	NC_053517.1	111908344	1	6540
2	2	38868	-1	NC_053517.1	111899667	1	6539
3	2	42746	1	NC_053517.1	111895084	-1	6538
4	2	46243	-1	NC_053517.1	111891588	1	6537
5	2	53442	-1	NC_053517.1	111884408	1	6536
6	2	60574	1	NC_053517.1	111877309	-1	6535

6 rows

Hide

```
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','colo
r')

karyotype_dual_comparison <- rbind(karyotype_ZANU, karyotype_DONGOLA)
head(karyotype_dual_comparison)
```

	Chr <chr>	Start <dbl>	End <dbl>	fill <dbl>	species <chr>	size <dbl>	color <dbl>
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

6 rows

Hide

```
NA
```

Hide

```
gm[gm$Dchr == 'NC_053517.1', ]['Dchr'] = '2'
gm[gm$Dchr == 'NC_053518.1', ]['Dchr'] = '3'
gm[gm$Dchr == 'NC_053519.1', ]['Dchr'] = 'X'

#choosing only X, 2, 3
gm <- subset(gm, gm$Dchr %in% c('2', '3', 'X'))
gm <- subset(gm, gm$Zchr %in% c('2', '3', 'X'))
gm[gm$Dchr == 'X', ]['Dchr'] = '1'
gm[gm$Zchr == 'X', ]['Zchr'] = '1'
gm <- gm[gm$Zchr == gm$Dchr,]
gm$Dlen <- as.integer(gm$Dlen)
gm$Dmid <- as.integer(gm$Dmid)
gm$Dstrand <- as.integer(gm$Dstrand)
head(gm)
```

	Zchr <chr>	Zmid <int>	Zstrand <int>	Dchr <chr>	Dmid <int>	Dstrand <int>	Dlen <int>
1	2	31135	-1	2	111908344	1	6540
2	2	38868	-1	2	111899667	1	6539
3	2	42746	1	2	111895084	-1	6538

4	2	46243	-1	2	111891588	1	6537
5	2	53442	-1	2	111884408	1	6536
6	2	60574	1	2	111877309	-1	6535

6 rows

Hide

```
NA
```

Hide

```
#reversing 2 and 3
for(i in 1:nrow(gm)){
  row <- gm[i, ]
  if (row$Zchr == 2){
    gm[i, 'ZmidReverse'] <- 114783175 - row$Zmid
  }
  else if (row$Zchr == 3){
    gm[i, 'ZmidReverse'] <- 97973315 -row$Zmid
  }
  else {
    gm[i, 'ZmidReverse'] <- row$Zmid
  }
}

gm$chr1 <- as.numeric(gm$Zchr)
gm$str1 <- as.integer(gm$ZmidReverse - gm$Dlen/2)
gm$end1 <- as.integer(gm$ZmidReverse + gm$Dlen/2)
gm$chr2 <- as.numeric(gm$Dchr)
gm$str2 <- as.integer(gm$Dmid - gm$Dlen/2)
gm$end2 <- as.integer(gm$Dmid + gm$Dlen/2)

#coloring
for(i in 1:nrow(gm)){
  row <- gm[i, ]
  gm[i, 'fill'] <- ifelse(row$Zstrand == row$Dstrand, 'D88974', '74D87A')
}
gm <- gm[,-which(names(gm) %in% c('Zchr','Zmid','Zstrand','Dchr','Dmid','Dstrand',
'Dlen','ZmidReverse'))]
head(gm)
```

	chr1 <dbl>	str1 <int>	end1 <int>	chr2 <dbl>	str2 <int>	end2 <int>	fill <chr>
1	2	114748770	114755310	2	111905074	111911614	74D87A
2	2	114741037	114747576	2	111896397	111902936	74D87A
3	2	114737160	114743698	2	111891815	111898353	74D87A
4	2	114733663	114740200	2	111888319	111894856	74D87A
5	2	114726465	114733001	2	111881140	111887676	74D87A

6	2	114719333	114725868	2	111874041	111880576	74D87A

6 rows

Hide

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