

Step 7: Test for Similarity

```
# ::: Set the main directory in "treedist" for step 7
mainDir <- "../treedist"
# Change the directory to let the working directory "treedist."
# For example, mainDir <- "../project-team-12/treedist"
# You can type getwd() to find your current working directory.

allDir <- "allblocks"
cnsDir <- "consecutiveblocks"
```

a. Are the observed tree distances closer to 0 than expected if the 2 trees were chosen at random uniformly? We would think so, if each plant was from a distinct population and if populations did not mix.

```
# Set the working directory for a
setwd(file.path(mainDir, allDir))

# Load data
chr2.all.tre <- read.table("chr2-all.tre")
chrC.all.tre <- read.table("chrC-all.tre")
chrM.all.tre <- read.table("chrM-all.tre")

#chr2.all.rfdist
chr2.all.rfdist <- read.table("chr2-all.tre.rfdist", skip=1, row.names = 1)

#chrC.all.rfdist
chrC.all.rfdist <- read.table("chrC-all.tre.rfdist", skip=1, row.names = 1)

#chrM.all.rfdist
chrM.all.rfdist <- read.table("chrM-all.tre.rfdist", skip=1, row.names = 1)

# 1) find D, distances between all pairs of trees.
chr2.all.D <- chr2.all.rfdist[lower.tri(chr2.all.rfdist)]
chrC.all.D <- chrC.all.rfdist[lower.tri(chrC.all.rfdist)]
chrM.all.D <- chrM.all.rfdist[lower.tri(chrM.all.rfdist)]
```

b. Do trees from 2 consecutive blocks tend to be more similar to each other (at smaller distance) than trees from 2 randomly chosen blocks from the same chromosome? We would expect so if blocks were small, due to less “recombination” between neighboring blocks than between blocks at opposing ends of the chromosome.

```
# Set the working directory for a
setwd(file.path(mainDir, cnsDir))

# Load data
chr2.cns.tre <- read.table("chr2-cns.tre")
chrC.cns.tre <- read.table("chrC-cns.tre")
chrM.cns.tre <- read.table("chrM-cns.tre")

#chr2.cns.rfdist
chr2.cns.rfdist <- read.table("chr2-cns.tre.rfdist", skip=1)
```

```

#chrC.cnsc.rfdist
chrC.cnsc.rfdist <- read.table("chrC-cnsc.tre.rfdist", skip=1)

#chrM.cnsc.rfdist
chrM.cnsc.rfdist <- read.table("chrM-cnsc.tre.rfdist", skip=1)

# 1) find D, distances between all pairs of trees.
chr2.cnsc.D <- as.numeric(chr2.cnsc.rfdist[-length(chr2.cnsc.rfdist)])
chrC.cnsc.D <- as.numeric(chrC.cnsc.rfdist[-length(chrC.cnsc.rfdist)])
chrM.cnsc.D <- as.numeric(chrM.cnsc.rfdist[-length(chrM.cnsc.rfdist)])

par(mfrow=c(2,3))
# ::: a :::

# :: for chromosome 2
hist(chr2.all.D, main="Density of Chr. 2 with all blocks", xlab="Chromosome 2", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chr2.all.D)/2), to=round(216-3-min(chr2.all.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# :: for chromosome C
hist(chrC.all.D, main="Density of Chr. C with all blocks", xlab="Chromosome C", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrC.all.D)/2), to=round(216-3-min(chrC.all.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

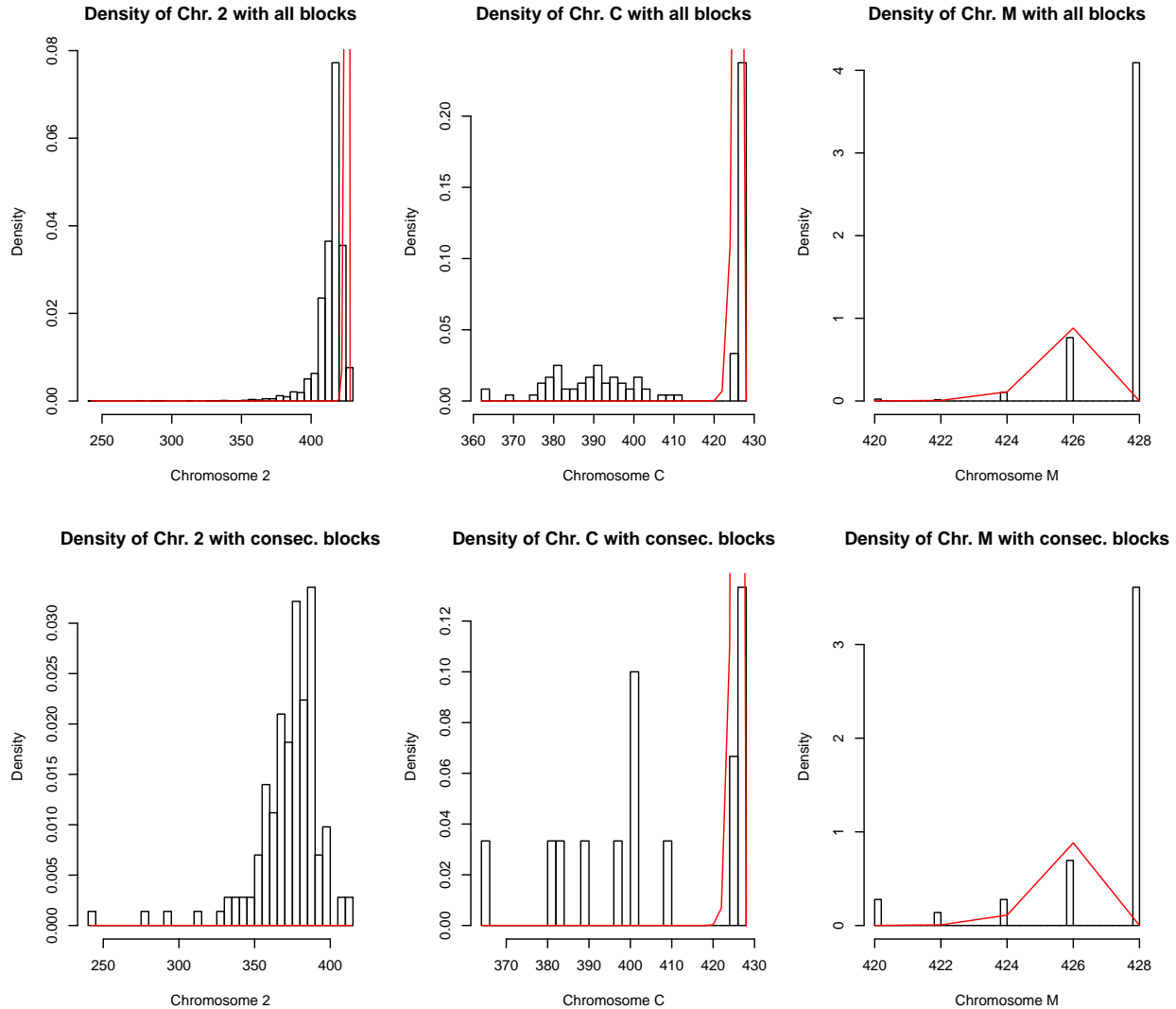
# :: for chromosome M
hist(chrM.all.D, main="Density of Chr. M with all blocks", xlab="Chromosome M", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrM.all.D)/2), to=round(216-3-min(chrM.all.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# ::: b :::
# :: for chromosome 2
hist(chr2.cnsc.D, main="Density of Chr. 2 with consec. blocks", xlab="Chromosome 2", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chr2.cnsc.D)/2), to=round(216-3-min(chr2.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# :: for chromosome C
hist(chrC.cnsc.D, main="Density of Chr. C with consec. blocks", xlab="Chromosome C", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrC.cnsc.D)/2), to=round(216-3-min(chrC.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# :: for chromosome M
hist(chrM.cnsc.D, main="Density of Chr. M with consec. blocks", xlab="Chromosome M", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrM.cnsc.D)/2), to=round(216-3-min(chrM.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

```



```
mean.dist <- rbind(cbind(mean(chr2.all.D), mean(chr2.cnsc.D)),
                  cbind(mean(chrC.all.D), mean(chrC.cnsc.D)),
                  cbind(mean(chrM.all.D), mean(chrM.cnsc.D)))
colnames(mean.dist) <- c("all blocks", "consec. blocks")
rownames(mean.dist) <- c("chr. 2", "chr. C", "chr. M")
mean.dist
```

```
##      all blocks consec. blocks
## chr. 2  414.5892      372.8671
## chr. C  410.3667      406.5333
## chr. M  427.5556      426.8889
```

```
par(mfrow=c(2,3))
```

```
# ::: a :::
```

```
# :: for chromosome 2
```

```
hist((216-3-chr2.all.D)/2, main="Density of Chr. 2 with all blocks", xlab="Chromosome 2", prob=TRUE, br
```

```
x = seq(from=round(216-3-max(chr2.all.D)/2), to=round(216-3-min(chr2.all.D)/2), by=1)
```

```
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")
```

```

# :: for chromosome C
hist(chrC.all.D, main="Density of Chr. C with all blocks", xlab="Chromosome C", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrC.all.D)/2), to=round(216-3-min(chrC.all.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

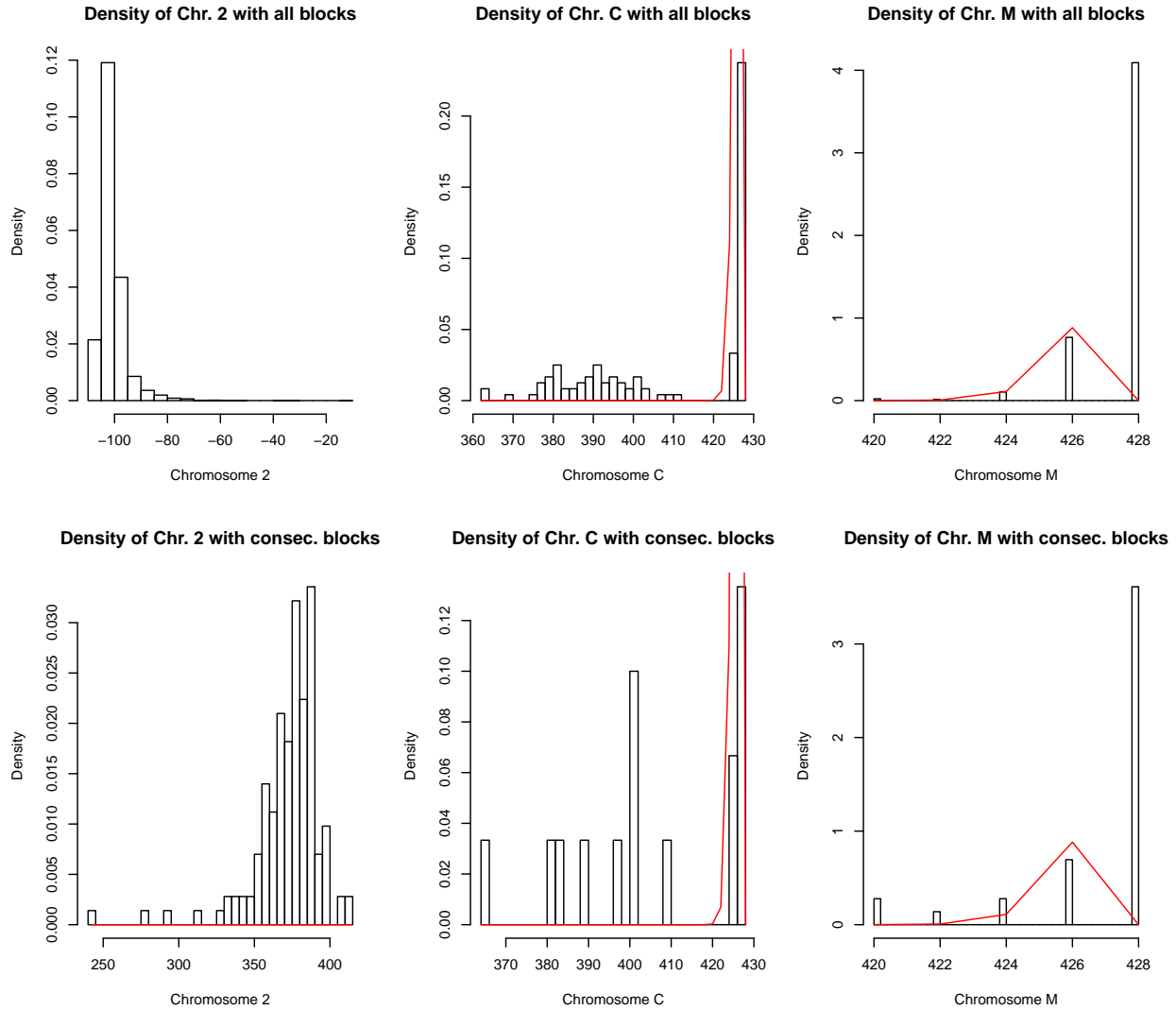
# :: for chromosome M
hist(chrM.all.D, main="Density of Chr. M with all blocks", xlab="Chromosome M", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrM.all.D)/2), to=round(216-3-min(chrM.all.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# ::: b :::
# :: for chromosome 2
hist(chr2.cnsc.D, main="Density of Chr. 2 with consec. blocks", xlab="Chromosome 2", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chr2.cnsc.D)/2), to=round(216-3-min(chr2.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# :: for chromosome C
hist(chrC.cnsc.D, main="Density of Chr. C with consec. blocks", xlab="Chromosome C", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrC.cnsc.D)/2), to=round(216-3-min(chrC.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

# :: for chromosome M
hist(chrM.cnsc.D, main="Density of Chr. M with consec. blocks", xlab="Chromosome M", prob=TRUE, breaks=30)
x = seq(from=round(216-3-max(chrM.cnsc.D)/2), to=round(216-3-min(chrM.cnsc.D)/2), by=1)
lines(x=I(2*(216-3-x)), y=dpois(x, lambda=1/8), col="red")

```



```
mean.dist <- rbind(cbind(mean(chr2.all.D), mean(chr2.cnsc.D)),
                  cbind(mean(chrC.all.D), mean(chrC.cnsc.D)),
                  cbind(mean(chrM.all.D), mean(chrM.cnsc.D)))
colnames(mean.dist) <- c("all blocks", "consec. blocks")
rownames(mean.dist) <- c("chr. 2", "chr. C", "chr. M")
mean.dist
```

```
##      all blocks consec. blocks
## chr. 2  414.5892      372.8671
## chr. C  410.3667      406.5333
## chr. M  427.5556      426.8889
```

(a) We think that the observed tree distances are closer to 0 than expected if the 2 trees were chosen at random uniformly based the top three plots.

(b) We think that trees from 2 consecutive blocks tend to be more similar to each other based on the bottom three plots and mean comparison between with all blocks and consecutive blocks.