ToxApresence\_absence Figure

Filtering REPET GFF files and making TE.DNAseg objects for plotting Note: If you want Rstudio to inherit the correct PATH (i.e. /anaconda/bin) you need to start RStudio from command line “open -a RStudio”

## Important chromosomes  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/  
  
##Make Geneious friendly gff files  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/CS10/CS10\_TA\_a0\_GFF3chr  
cat CS10\_Chromosome\_08.gff3 |grep "CS10\_TA\_a0\_REPET\_TEs" | grep "match\_part" | sed 's/ID=\(.\*\);Parent=\(.\*\);Target/ID=\2;Parent=\1;Target/' > CS10\_Chromosome\_08\_edit.gff3  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/SN15/SN15\_STA\_a0\_GFF3chr  
cat SN15\_SOL\_Chromosome\_05.gff3 |grep "SN15\_STA\_a0\_REPET\_TEs" | grep "match\_part" | sed 's/ID=\(.\*\);Parent=\(.\*\);Target/ID=\2;Parent=\1;Target/' > SN15\_SOL\_Chromosome\_05\_edit.gff3   
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/PTR/PTR\_TA\_a0\_GFF3chr  
cat supercont1.4.gff3 |grep "PTR\_TA\_a0\_REPET\_TEs" | grep "match\_part" | sed 's/ID=\(.\*\);Parent=\(.\*\);Target/ID=\2;Parent=\1;Target/' > supercont1.4\_edit.gff3  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/CS27/CS27\_TA\_a0\_GFF3chr  
cat CS27\_Chromosome\_08.gff3 | grep "CS27\_TA\_a0\_REPET\_TEs" |grep "match\_part" | sed 's/ID=\(.\*\);Parent=\(.\*\);Target/ID=\2;Parent=\1;Target/' > CS27\_Chromosome\_08\_edit.gff3  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/SN79\_SOL/SN79\_STA\_a0\_GFF3chr  
cat SN79\_SOL\_Chromosome\_05.gff3 | grep "SN79\_STA\_a0\_REPET\_TEs" | grep "match\_part" | sed 's/ID=\(.\*\);Parent=\(.\*\);Target/ID=\2;Parent=\1;Target/'> SN79\_SOL\_Chromosome\_05\_edit.gff3  
  
  
##Make TE.dnaseg files for plotting transposons  
## will attach header with following columns to match dna\_segs object c(name,start,end,strand,col,gene\_type)  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/CS10/CS10\_TA\_a0\_GFF3chr  
cat /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/TE\_Header.txt > /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/CS10\_TEsegs.dnasegs  
  
cat CS10\_Chromosome\_08\_edit.gff3 | awk -F "\t" '{print $1"\t"$4"\t"$5"\t"$7"\t""blue""\t""blue""\t""blocks"}' >> /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/CS10\_TEsegs.dnasegs  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/SN15/SN15\_STA\_a0\_GFF3chr  
cat /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/TE\_Header.txt > /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/SN15\_SOL\_TEsegs.dnasegs  
  
cat SN15\_SOL\_Chromosome\_05\_edit.gff3 | awk -F "\t" '{print $1"\t" $4"\t"$5"\t"$7"\t""blue""\t""blue""\t""blocks"}' >> /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/SN15\_SOL\_TEsegs.dnasegs  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/PTR/PTR\_TA\_a0\_GFF3chr  
cat /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/TE\_Header.txt > /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/PTRs\_TEsegs.dnasegs  
  
cat supercont1.4\_edit.gff3 | awk -F "\t" '{print $1"\t" $4"\t"$5"\t"$7"\t""blue""\t""blue""\t""blocks"}' >> /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/PTRs\_TEsegs.dnasegs  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/CS27/CS27\_TA\_a0\_GFF3chr  
cat /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/TE\_Header.txt > /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/CS27\_TEsegs.dnasegs  
  
cat CS27\_Chromosome\_08\_edit.gff3 | awk -F "\t" '{print $1"\t" $4"\t"$5"\t"$7"\t""blue""\t""blue""\t""blocks"}' >> /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/CS27\_TEsegs.dnasegs  
  
cd /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Ben\_REPET/REPET2/TEanno/SN79\_SOL/SN79\_STA\_a0\_GFF3chr  
cat /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/TE\_Header.txt > /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/SN79\_TEsegs.dnasegs  
  
cat SN79\_SOL\_Chromosome\_05\_edit.gff3 | awk -F "\t" '{print $1"\t" $4"\t"$5"\t"$7"\t""blue""\t""blue""\t""blocks"}' >> /Users/meganm/Dropbox\ \(Solomon\ lab\)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence/SN79\_TEsegs.dnasegs

Generate Lastz+TRF filtered outputs (mimeo-map) into DNAcomp files for plotting Note:The orientation that you align the chromosomes will determine if “(+ + =1) or (+ - =1)” This is important to get the color the same for all alignments. Easiest thing is to make RC a chromosome if needed and re-do the alignments.

Exampe mimeo-map run: mimeo-map –afasta SN15\_SOL\_Chromosome\_05.fasta –bfasta SN79\_SOL\_Chromosome\_05.fasta –outfile SN15vsSN79\_map –minIdt 60 –writeTRF

Formated Mimeo output for import as a dnaseg object should have the following tab delim columns “start1 end1 start2 end2 strand identity”

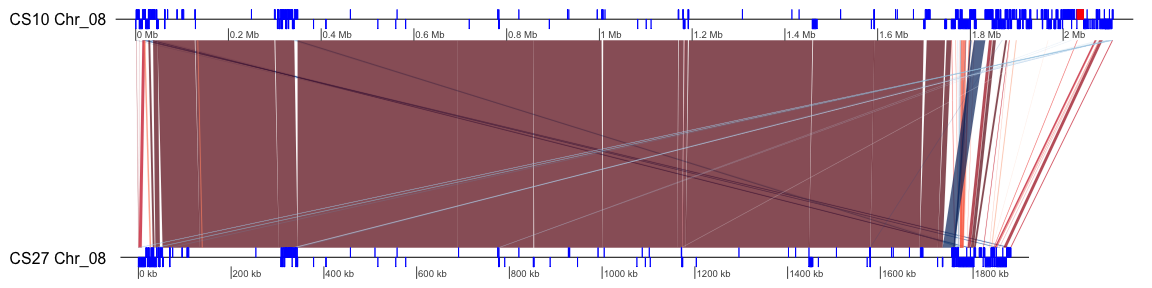
## Import data for Plotting Chromosomes and their Synteny  
library(genoPlotR)

## Loading required package: ade4

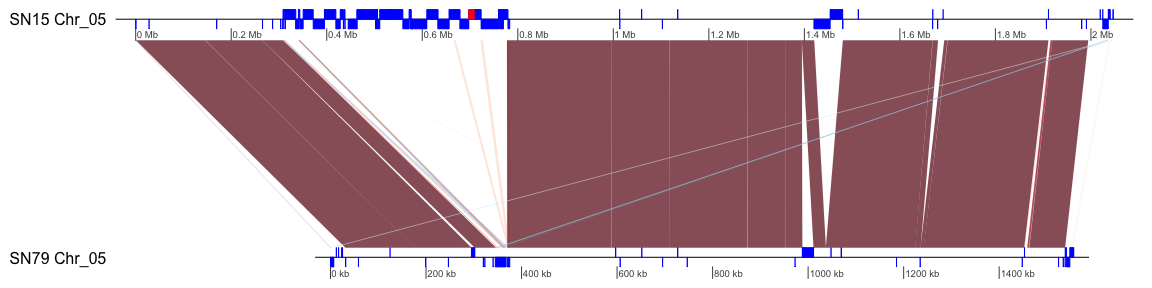
## Loading required package: grid

setwd("~/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence")  
  
  
## ToxA locations Keep these for Later and add to DNAseqs objects  
PtrToxA<-data.frame(name="ToxA",start=1442202,end=1453529,strand="+",col="red", fill="red",gene\_type="blocks")  
  
CS10ToxA<-data.frame(name="ToxA", start=2029860, end=2043926,strand="+", col="red", fill="red",gene\_type="blocks")  
  
SN15ToxA<-data.frame(name="ToxA",start=697480,end=708560,strand="+",col="red", fill="red",gene\_type="blocks")

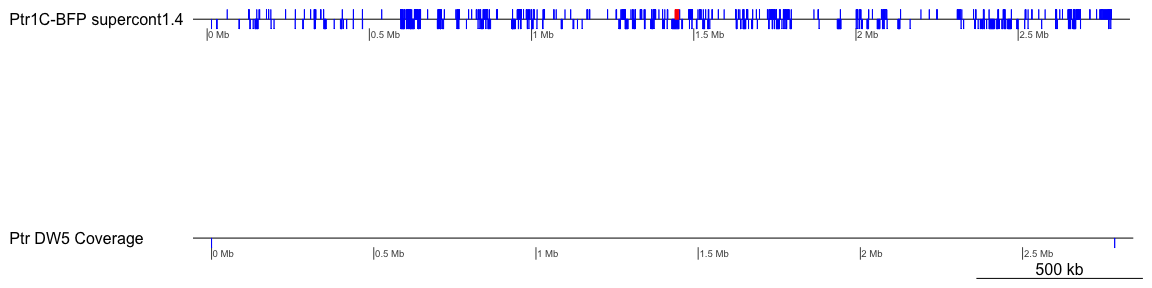
### Now import Transposons DNAseg object to plot  
  
setwd("~/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ToxAPresenceAbsence")  
## Combine ToxA seg and TE annotations  
  
### Bipolaris only plot  
CS10.TE.annot<-data.frame(read.delim("CS10\_TEsegs.dnasegs", sep = "\t"))  
CS10.TE.annot<-rbind(CS10.TE.annot,CS10ToxA)  
CS10.TE.annot<-dna\_seg(CS10.TE.annot)  
  
CS27.TE.annot<-data.frame(read.delim("CS27\_TEsegs.dnasegs", sep = "\t"))  
CS27.TE.annot<-dna\_seg(CS27.TE.annot)  
teonly<-list(CS10.TE.annot ,CS27.TE.annot)  
names(teonly)<- c("CS10 Chr\_08", "CS27 Chr\_08")  
  
Bipolar\_comp<-read\_comparison\_from\_tab("CS10vsCS27\_chr8\_map.dnacomp", header=T)  
Bipolar\_comp$direction<-Bipolar\_comp$strand  
  
plot\_gene\_map(dna\_segs = teonly , comparisons = list(Bipolar\_comp), gene\_type = "side\_blocks", dna\_seg\_scale= T, scale=F, global\_color\_scheme = c("identity", "auto","red\_blue",0.7))



SN15.TE.annot<-data.frame(read.delim("SN15\_SOL\_TEsegs.dnasegs", sep = "\t"))  
SN15.TE.annot<-rbind(SN15.TE.annot,SN15ToxA)  
SN15.TE.annot<-dna\_seg(SN15.TE.annot)  
  
SN79.TE.annot<-try(read\_dna\_seg\_from\_tab("SN79\_TEsegs.dnasegs", header=T))  
SNtes<-list(SN15.TE.annot, SN79.TE.annot)  
names(SNtes)<- c("SN15 Chr\_05", "SN79 Chr\_05")  
  
Pnod\_comp<-read\_comparison\_from\_tab("SN15vsSN79\_map.dnacomp", header=T)  
Pnod\_comp$direction<-Pnod\_comp$strand  
  
plot\_gene\_map(dna\_segs = SNtes , comparisons = list(Pnod\_comp), gene\_type = "side\_blocks", dna\_seg\_scale= T, scale=F, global\_color\_scheme = c("identity", "auto","red\_blue",0.7))



PTRs.TE.annot<-data.frame(read.delim("PTRs\_TEsegs.dnasegs", sep="\t"))  
PTRs.TE.annot<-rbind(PTRs.TE.annot, PtrToxA)  
PTRs.TE.annot<-dna\_seg(PTRs.TE.annot)  
dummy.annot <-try(read\_dna\_seg\_from\_tab("Ptr\_dummy.seg", header = F))  
ptrte<-list(PTRs.TE.annot,dummy.annot)  
names(ptrte)<-c("Ptr1C-BFP supercont1.4", "Ptr DW5 Coverage")  
Ptrxlims<-list(c(0,2787600))  
  
#plot\_gene\_map(dna\_segs = list(Ptr.TE.annot) ,gene\_type = "side\_blocks", dna\_seg\_scale = T)  
plot\_gene\_map(dna\_segs = ptrte ,gene\_type = "side\_blocks", dna\_seg\_scale = T)



### Plot all three graphs together and print to a PDF  
  
library(grid)  
  
cairo\_pdf(paste0("Figure\_X.pdf"), width = 12, height = 9)  
pushViewport(viewport(layout=grid.layout(3,1, heights=unit(c(1,1,1), rep("null",3))), name="Synteny between ToxA +/- Isolates"))  
## Panel A  
pushViewport(viewport(layout.pos.row = 1, name="panelA"))  
plot\_gene\_map(dna\_segs = teonly , comparisons = list(Bipolar\_comp), gene\_type = "side\_blocks", global\_color\_scheme = c("identity", "auto","red\_blue",0.7),dna\_seg\_scale = T, scale=F, main = "A",main\_pos = "left",plot\_new = F)  
  
upViewport()  
## Panel B  
pushViewport(viewport(layout.pos.row = 2,name="panelB"))  
  
plot\_gene\_map(dna\_segs = SNtes , comparisons = list(Pnod\_comp), gene\_type = "side\_blocks", global\_color\_scheme = c("identity","auto","red\_blue", 0.7), dna\_seg\_scale= T, scale=F, override\_color\_schemes = T,main="B", main\_pos="left", plot\_new=F)  
  
##Panel C  
upViewport()  
pushViewport(viewport(layout.pos.row = 3,name="panelC"))  
plot\_gene\_map(dna\_segs = ptrte ,gene\_type = "side\_blocks", dna\_seg\_scale = T, scale=F,main = "C",main\_pos = "left", plot\_new = F)  
dev.off()

## quartz\_off\_screen   
## 2

### Import feature counts accross 10kb windows  
###Note need to create new column which counts the average "coverage" accross these windows assumming  
### 150 bp reads  
library(ggplot2)  
RawCounts <- read.table("/Users/meganm/Dropbox (Solomon lab)/Movement\_of\_ToxA/Ptr\_Mapped\_Reads/DW5\_coveragecount\_10kb\_1.4.txt", sep="\t", header=F)  
  
#Calculate coverage for each window  
RawCounts<-cbind(RawCounts,(RawCounts$V4\*150)/10000)  
  
colnames(RawCounts)<- c("Chrom", "Start","Stop","Reads","Non-zero-bases", "length", "fractionnonzero","Coverage")  
  
cairo\_pdf(paste0("FigureXC\_readcov.pdf"), width = 11.5, height = 2)  
ggplot(RawCounts, aes(x=Start,y=Coverage))+geom\_line(aes(col=fractionnonzero))+theme\_classic()+coord\_cartesian(xlim=c(0,2787645), expand = c(0,0))+scale\_color\_gradient(low="pink",high=" dark green")+theme(axis.line.x=element\_blank(), axis.title.x =element\_blank())+scale\_x\_continuous(breaks=seq(0,2787645,500000))

## Warning in if (expand) expand\_default(scale) else c(0, 0): the condition  
## has length > 1 and only the first element will be used  
  
## Warning in if (expand) expand\_default(scale) else c(0, 0): the condition  
## has length > 1 and only the first element will be used

dev.off()

## quartz\_off\_screen   
## 2