Figure2 Extent of HGT

Megan McDonald

14/11/2018

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Import repeat annotations from REPET pipeline Filter datsets to only the ToxA containing chromosomes

## Reading in Data and formatting it for Filtering ####  
library(ggplot2)  
### Had a lot of issues getting this to work correctly, set the "root.dir" in the R setup chunk above  
#setwd("~/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ExtentofHGT/")  
getwd()

## [1] "/Users/meganm/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ExtentofHGT"

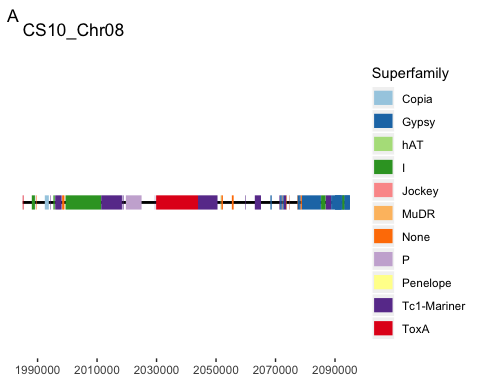
## Set up chromosome segments  
CS10\_Chr08<-data.frame(c(1),c(2108098))  
names(CS10\_Chr08)<-c("start","stop")  
  
WAI2406\_Chr01<-data.frame(c(1),c(4014556))  
names(WAI2406\_Chr01)<-c("start","stop")  
  
WAI2411\_tig17<-data.frame(c(1),c(776846))  
names(WAI2411\_tig17)<-c("start","stop")  
### NOTE This Document was prepared with an older name of SN15, Chr05 is used in this document but the re-named version is Chr08, to match the names of Richards et. al. 2018  
SN15\_Chr05<-data.frame(c(1),c(2046703))  
names(SN15\_Chr05)<-c("start","stop")  
  
PTR\_tig1.4<-data.frame(c(1),c(2787645))  
names(PTR\_tig1.4)<-c("start","stop")  
  
## Import Tabdelim file with all ToxA chromosomes REPET annoations  
AllToxAtransposons<-read.delim("ToxA\_Chromosome\_Transposons.txt", header=T, stringsAsFactors=F)  
  
### Separate into Five different genome objects to plot separately ###  
  
## First get levels of Chromosomes ##  
unique(AllToxAtransposons$Chromosome)

## [1] "CS10\_Chromosome\_08" "WAI2406\_try2\_racon5\_Chromosome\_01"  
## [3] "WAI2411\_contig\_17" "SN15\_SOL\_Chromosome\_05"   
## [5] "supercont1.4"

###[1] "CS10\_Chromosome\_08" "WAI2406\_try2\_racon5\_Chromosome\_01" "WAI2411\_contig\_17"   
###[4] "SN15\_SOL\_Chromosome\_05" "supercont1.4"   
CS10\_transposons<-AllToxAtransposons[AllToxAtransposons$Chromosome=="CS10\_Chromosome\_08",]  
CS10\_transposons$Superfamily[is.na(CS10\_transposons$Superfamily)]<-"None"  
  
WAI2406\_transposons<-AllToxAtransposons[AllToxAtransposons$Chromosome=="WAI2406\_try2\_racon5\_Chromosome\_01",]  
WAI2406\_transposons$Superfamily[is.na(WAI2406\_transposons$Superfamily)]<-"None"  
  
WAI2411\_transposons<-AllToxAtransposons[AllToxAtransposons$Chromosome=="WAI2411\_contig\_17",]  
WAI2411\_transposons$Superfamily[is.na(WAI2411\_transposons$Superfamily)]<-"None"  
  
SN15\_transposons<-AllToxAtransposons[AllToxAtransposons$Chromosome=="SN15\_SOL\_Chromosome\_05",]  
SN15\_transposons$Superfamily[is.na(SN15\_transposons$Superfamily)]<-"None"  
  
PTR\_transposons<-AllToxAtransposons[AllToxAtransposons$Chromosome=="supercont1.4",]  
PTR\_transposons$Superfamily[is.na(PTR\_transposons$Superfamily)]<-"None"

# Make objects for the xlim coord\_cartesian

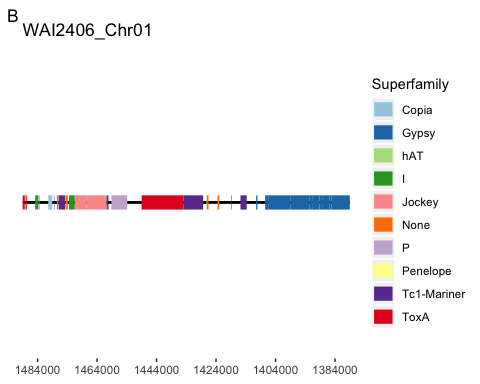
CS10limits<-c(1990000,2090000)  
WAI2411limits<-c(685000,776846)  
WAI2406limits<-c(1384000,1484000)  
SN15limits<-c(665000,765000)  
PTRlimits<-c(1390000,1490000)  
  
CS10breaks<-seq(1990000,2090000,20000)  
WAI2411breaks<-seq(685000,776846,20000)  
WAI2406breaks<-seq(1384000,1484000,20000)  
SN15breaks<-seq(665000,765000,20000)  
PTRbreaks<-seq(1390000,1490000,20000)  
  
CS10\_ChromPlot<-ggplot(CS10\_Chr08, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=CS10limits,ylim = c(-0.02,0.02))+scale\_x\_continuous(breaks = CS10breaks)+  
 geom\_segment(data=CS10\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#fb9a99','#fdbf6f','#ff7f00','#cab2d6','#ffff99','#6a3d9a','#e31a1c'))+  
 labs(title = "CS10\_Chr08", tag="A")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
CS10\_ChromPlot



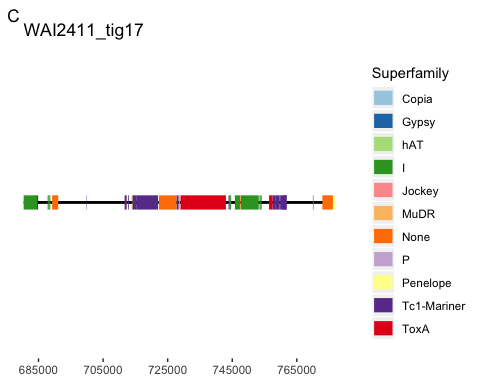
ggsave("CS10test.pdf",width = 12,height = 2, dpi=300)

Plot Other Bipolaris

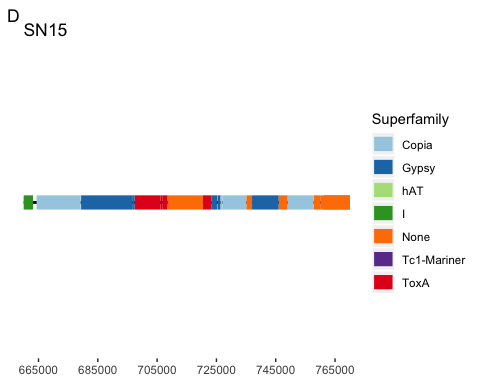
WAI2406\_ChromPlot<-ggplot(WAI2406\_Chr01, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=WAI2406limits,ylim = c(-0.02,0.02))+scale\_x\_reverse(breaks = WAI2406breaks)+  
 geom\_segment(data=WAI2406\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#fb9a99','#ff7f00','#cab2d6','#ffff99','#6a3d9a','#e31a1c'))+  
 labs(title = "WAI2406\_Chr01", tag="B")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
WAI2406\_ChromPlot



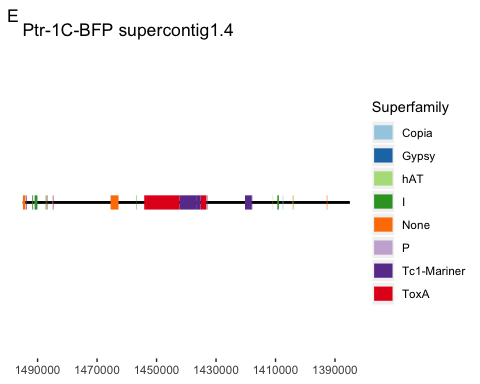
WAI2411\_ChromPlot<-ggplot(WAI2411\_tig17, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=WAI2411limits,ylim = c(-0.02,0.02))+scale\_x\_continuous(breaks = WAI2411breaks)+  
 geom\_segment(data=WAI2411\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#fb9a99','#fdbf6f','#ff7f00','#cab2d6','#ffff99','#6a3d9a','#e31a1c'))+  
 labs(title = "WAI2411\_tig17", tag="C")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
WAI2411\_ChromPlot



SN15\_ChromPlot<-ggplot(SN15\_Chr05, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=SN15limits,ylim = c(-0.02,0.02))+scale\_x\_continuous(breaks = SN15breaks)+  
 geom\_segment(data=SN15\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#ff7f00','#6a3d9a','#e31a1c'))+  
 labs(title = "SN15", tag="D")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
SN15\_ChromPlot



PTR\_ChromPlot<-ggplot(PTR\_tig1.4, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=PTRlimits,ylim = c(-0.02,0.02))+scale\_x\_reverse(breaks=PTRbreaks)+  
 geom\_segment(data=PTR\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#ff7f00','#cab2d6','#6a3d9a','#e31a1c'))+  
 labs(title = "Ptr-1C-BFP supercontig1.4", tag="E")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
PTR\_ChromPlot

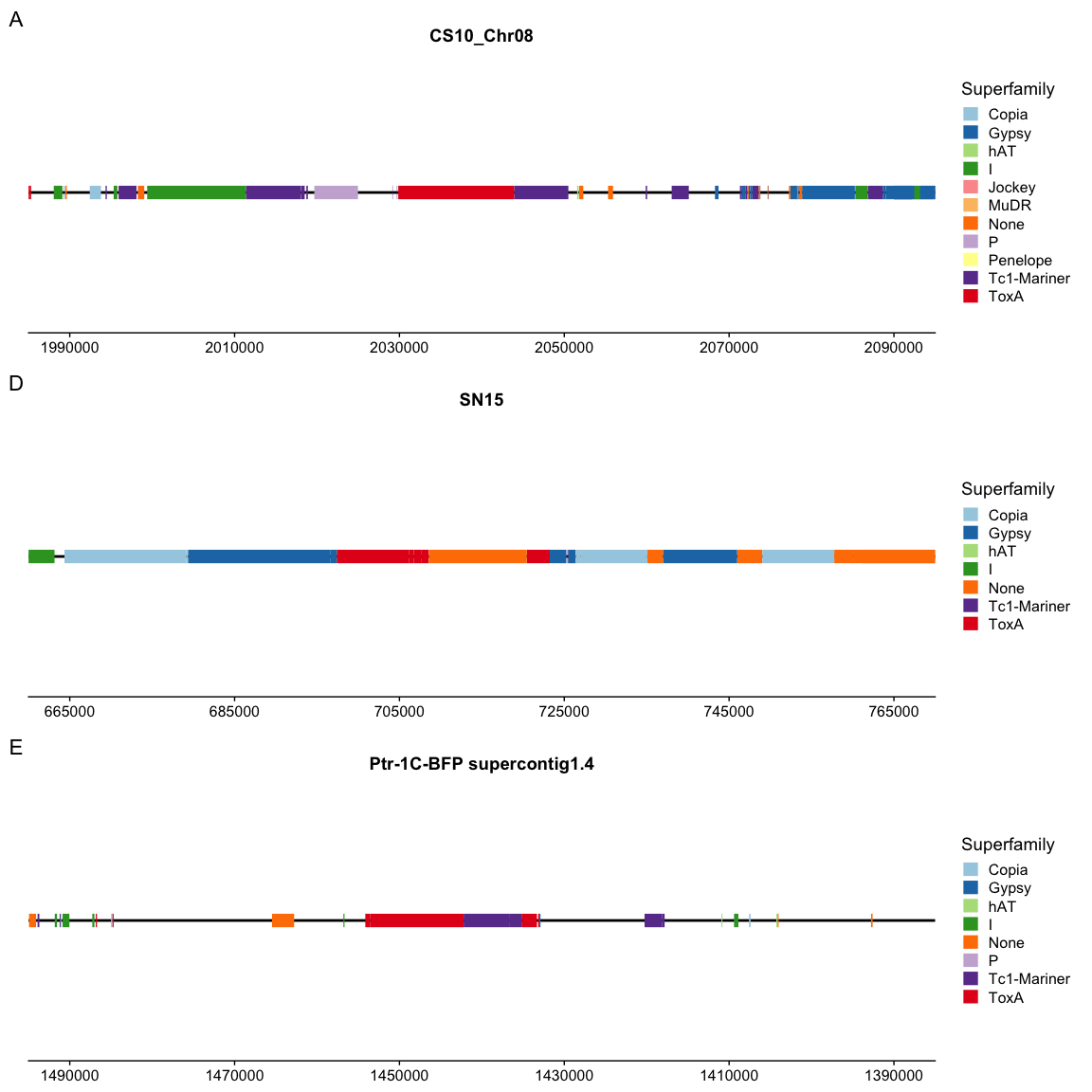


library(cowplot)

##   
## Attaching package: 'cowplot'

## The following object is masked from 'package:ggplot2':  
##   
## ggsave

Fig2B<-plot\_grid(CS10\_ChromPlot,SN15\_ChromPlot,PTR\_ChromPlot, align="v", nrow=3)  
Fig2B



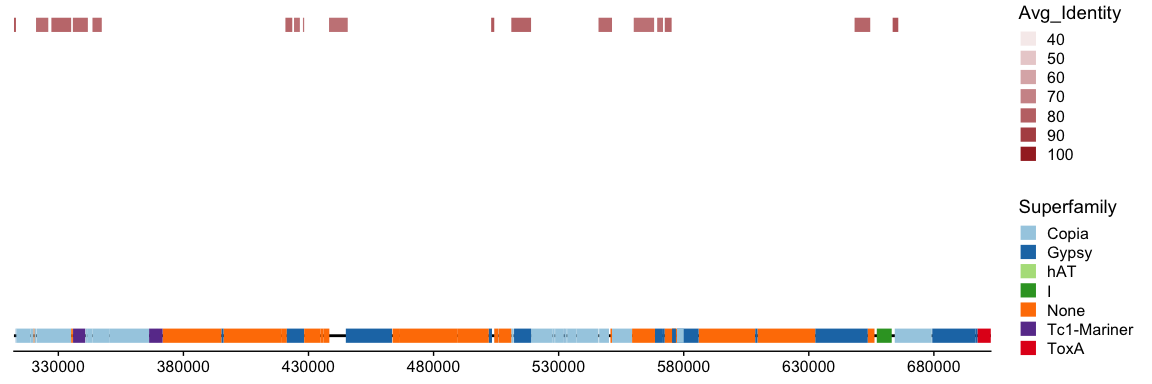
ggsave("RepeatIslands\_3speciesonly.pdf", Fig2B, width = 32, height =24, units=c("cm"),dpi=600 )

## Plot SN15 repeat island just to see what it looks like

PTR\_HGTs\_hits<-read.delim("~/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ExtentofHGT/HGT\_PTR\_SN15\_blastSummary.txt", header=T)  
head(PTR\_HGTs\_hits)

## Chromosome Manual\_begin Manual\_end strand Order Height  
## 1 SN15\_SOL\_Chromosome\_05 312126 313050 + HGT\_Pnod 0.02  
## 2 SN15\_SOL\_Chromosome\_05 321080 325974 + HGT\_Pnod 0.02  
## 3 SN15\_SOL\_Chromosome\_05 327214 335121 + HGT\_Pnod 0.02  
## 4 SN15\_SOL\_Chromosome\_05 335808 341797 + HGT\_Pnod 0.02  
## 5 SN15\_SOL\_Chromosome\_05 343642 347363 + HGT\_Pnod 0.02  
## 6 SN15\_SOL\_Chromosome\_05 420767 423514 + HGT\_Pnod 0.03  
## UniqueID Avg\_Identity  
## 1 HGT\_1 82  
## 2 HGT\_2 76  
## 3 HGT\_3 78  
## 4 HGT\_4 77  
## 5 HGT\_5 77  
## 6 HGT\_6 78

SN15island\_limit<-c(330000,685000)  
SN15island\_break<-seq(330000,685000, 50000)  
  
  
SN15\_Island<-ggplot(SN15\_Chr05, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=SN15island\_limit)+scale\_x\_continuous(breaks = SN15island\_break)+  
 geom\_segment(data=SN15\_transposons, aes(x=begin,xend=end, y=0,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#ff7f00','#6a3d9a','#e31a1c'))+  
 geom\_segment(data=PTR\_HGTs\_hits, aes(x=Manual\_begin, xend=Manual\_end, y=0.01,yend=0.01, alpha=Avg\_Identity),color="Brown",size=5)+scale\_alpha\_continuous(limits=c(40,100))+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
SN15\_Island



## Plot Ptr HGT regions that are spread out in Pnod

PTR\_HGTs\_orig<-read.delim("~/Dropbox (Solomon lab)/Movement\_of\_ToxA/Figure\_ExtentofHGT/HGT\_PTRorig\_locations.txt", header=T)  
PTR\_Islandlimits<-c(1450000,1550000)  
PTR\_Islandbreaks<-seq(1450000,1550000,25000)  
  
PTR\_HGT\_Island<-ggplot(PTR\_tig1.4, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=PTR\_Islandlimits,ylim = c(-0.04,0.04))+scale\_x\_reverse(breaks=PTR\_Islandbreaks)+  
 geom\_segment(data=PTR\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Superfamily), size=5)+  
 scale\_color\_manual(values=c('#a6cee3','#1f78b4','#b2df8a','#33a02c',  
 '#ff7f00','#cab2d6','#6a3d9a','#e31a1c'))+  
 labs(title = "Ptr-1C-BFP supercontig1.4")+  
 geom\_segment(data=PTR\_HGTs\_orig, aes(x=Manual\_begin, xend=Manual\_end, y=Height, yend=Height, alpha=Avg\_Identity),color="Brown",size=5)+scale\_alpha\_continuous(limits=c(40,100))+  
 theme(axis.line.y=element\_blank(),  
 # axis.line.x=element\_blank(),  
 # axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 # axis.ticks.x=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 legend.position = "none",  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
PTR\_HGT\_Island

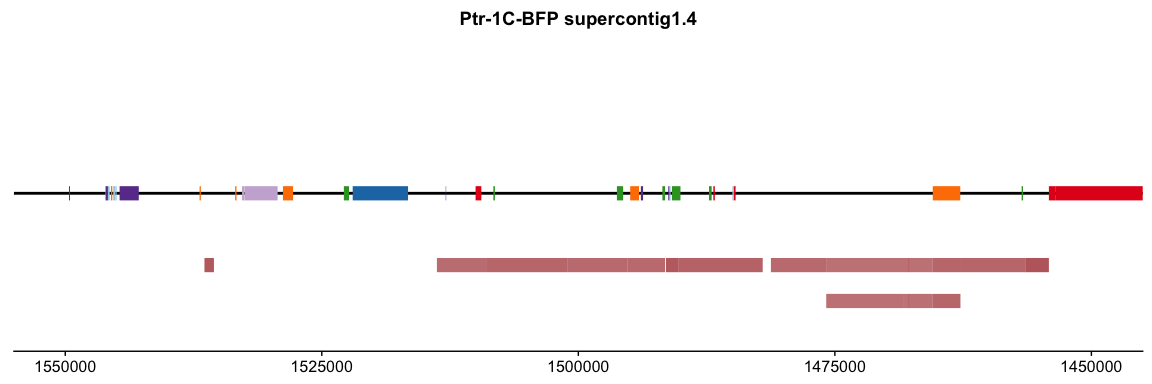


Fig2C<-ggdraw()+draw\_plot(SN15\_Island + theme(legend.justification = "bottom"))+draw\_plot\_label(c("SN15\_Chr08"))  
Fig2C

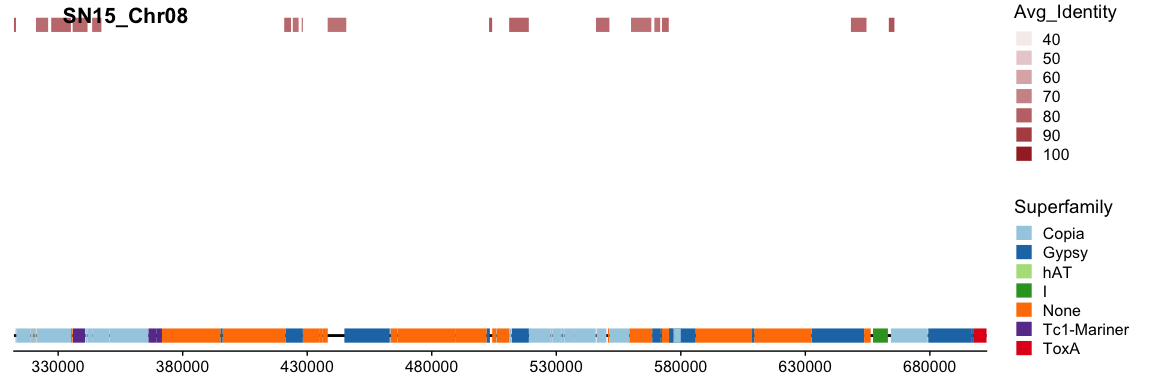
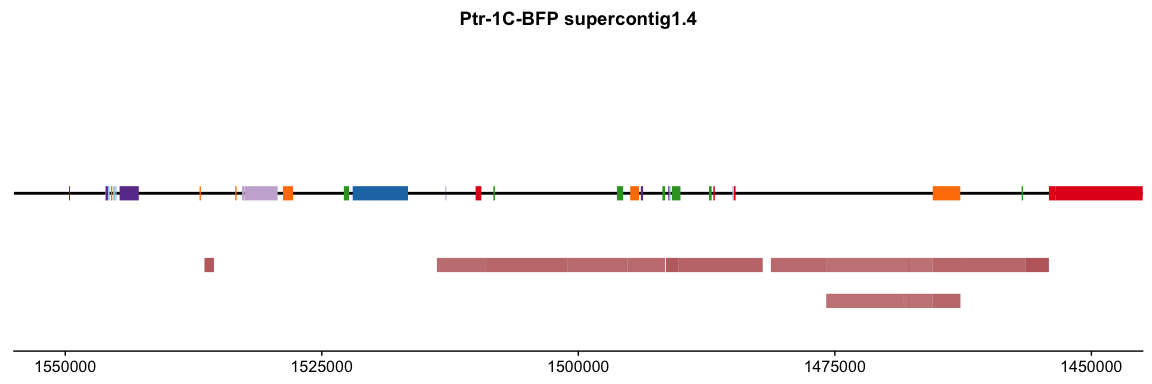
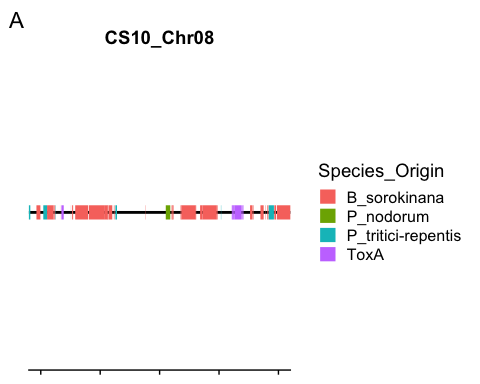


Fig2D<-ggdraw()+draw\_plot(PTR\_HGT\_Island + theme(legend.justification = "bottom"))  
Fig2D



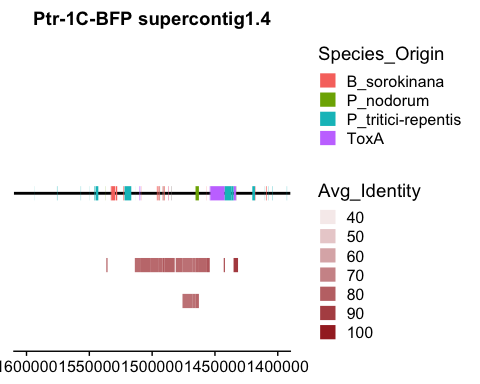
ggsave("ExtentofHGT\_Pnod.pdf", Fig2C, width = 32, height =16, units=c("cm"),dpi=600 )  
  
ggsave("ExtentofHGT\_PTR.pdf", Fig2D, width = 24, height =8, units=c("cm"),dpi=600 )

## Draw transposons according to species of origin  
  
CS10\_species<-ggplot(CS10\_Chr08, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=c(1900000,2000000),ylim = c(-0.02,0.02))+  
 geom\_segment(data=CS10\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Species\_Origin), size=5)+  
 labs(title = "CS10\_Chr08", tag="A")+  
 theme(axis.line.y=element\_blank(),  
 axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
CS10\_species



ggsave("CS10test.pdf",width = 12,height = 2, dpi=300)

PTR\_HGT\_species<-ggplot(PTR\_tig1.4, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=c(1400000,1600000),ylim = c(-0.04,0.04))+scale\_x\_reverse()+  
 geom\_segment(data=PTR\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Species\_Origin), size=5)+  
 labs(title = "Ptr-1C-BFP supercontig1.4")+  
 geom\_segment(data=PTR\_HGTs\_orig, aes(x=Manual\_begin, xend=Manual\_end, y=Height, yend=Height, alpha=Avg\_Identity),color="Brown",size=5)+scale\_alpha\_continuous(limits=c(40,100))+  
 theme(axis.line.y=element\_blank(),  
 # axis.line.x=element\_blank(),  
 # axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 # axis.ticks.x=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
  
PTR\_HGT\_species



SN15\_species<-ggplot(SN15\_Chr05, aes(x=start,y=0,xend=stop,yend=0))+geom\_segment(size=1)+   
 coord\_cartesian(xlim=c(600000,800000),ylim = c(-0.02,0.02))+  
 geom\_segment(data=SN15\_transposons, aes(x=begin,y=0,xend=end,yend=0, color=Species\_Origin), size=5)+  
 geom\_segment(data=PTR\_HGTs\_hits, aes(x=Manual\_begin, xend=Manual\_end, y=0.01,yend=0.01, alpha=Avg\_Identity),color="Brown",size=5)+scale\_alpha\_continuous(limits=c(40,100))+  
 labs(title = "SN15", tag="D")+  
 theme(axis.line.y=element\_blank(),  
 #axis.text.x=element\_blank(),  
 axis.text.y=element\_blank(),  
 axis.ticks.y=element\_blank(),  
 axis.title.y=element\_blank(),  
 axis.title.x=element\_blank(),  
 panel.background=element\_blank(),  
 plot.background=element\_blank())  
SN15\_species

