

het

TMJ

April 18, 2016

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:

read the 012 file and combine genotype with individual

```
setwd("~/Google Drive/final_analysis_LRS/pilot_22Apr2016_final/workdir")
genos <- read.delim("out.012", header=FALSE)
indiv <- read.table("out.012.indv", quote="\"", comment.char="")
genos$V1 <- NULL
gi <- cbind(indiv, genos)
row.names(gi) <- gi$V1
gi=gi[,-1]
tgi <- as.data.frame(t(gi))
```

divide into separate files and calculate each hybrid tool combination separately

```
samBC <- tgi[,1:3]
samBM <- tgi[,4:6]
gatBC <- tgi[,7:9]
gatBM <- tgi[,10:12]
```

```
script used to build loop: samBCt <- samBC[samBC[,1]==0 & samBC[,2]==2 & samBC[,3]!=-1,]
dim(samBCt) totLoci <- sum(samBCt[,1]==0) totLoci refAllele <- sum(samBCt[,3]==0) refAllele altAllele
<- sum(samBCt[,3]==2) altAllele hetAllele <- sum(samBCt[,3]==1) hetAllele
```

```
hetaccuracy <- hetAllele/totLoci hetaccuracy*100
```

function for finding hybrid accuracy rate for each file. I.e. if parents are 0 and 2, are hets 1?

```
hfx<-function(file){
filet <- file[file[,1]==0 & file[,2]==2 & file[,3]!=-1,]
dim(filet)
refAllele <- sum(filet[,3]==0)
refAllele
altAllele <- sum(filet[,3]==2)
altAllele
hetAllele <- sum(filet[,3]==1)
hetAllele
totLoci <- sum(filet[,1]==0)
totLoci
hetaccuracy <- hetAllele/totLoci
hetaccuracy*100
}
```

run the function across the 4 subfiles

```

hetfiles <- list(samBC, samBM, gatBC, gatBM)
resultcorrecthybrid <- lapply(hetfiles, hfx)
#hetfilenames <- list("samBC", "samBM", "gatBC", "gatBM")
#correcthybrid <- as.data.frame(cbind(hetfilenames, resultcorrecthybrid))
#correcthybrid$resultcorrecthybrid <- as.numeric(correcthybrid$resultcorrecthybrid)

```

count number of hets in hybrid

script used to build loop: `length(which(samBC$Xpress_0183.R_2015_08_24_16_56_10_user_PG1-118-NCSU-2.fastq == 1))` `gatBC2 <- gatBC[gatBC[,1]!=-1 & gatBC[,2]!=-1 & gatBC[,3]==1,]` `dim(gatBC2)` `totLoci <- sum(gatBC2[,3]==1)` `totLoci` `correctcalls <- sum(gatBC2[,1]==0 & gatBC2[,2]==2)` `correctcalls` `accuracy <- correctcalls/totLoci` `accuracy`

make a file if parents are correct. I.e. if hybrid is a 1 are parents 0 and 2?

```

correctparentfx <- function(file){
file2 <- file[file[,1]!=-1 & file[,2]!=-1 & file[,3]==1,]
dim(file2)
totLoci <- sum(file2[,3]==1)
totLoci
correctcalls <- sum(file2[,1]==0 & file2[,2]==2)
correctcalls
accuracy <- correctcalls/totLoci
accuracy*100
}

```

make a table with het accuracy rates for each hybrid pair

```

resultcorrectparents <- lapply(hetfiles, correctparentfx)
hetfilenames <- list("samBC", "samBM", "gatBC", "gatBM")
hetaccuracy <- as.data.frame(cbind(hetfilenames, resultcorrecthybrid, resultcorrectparents))
hetaccuracy$resultcorrecthybrid <- as.numeric(hetaccuracy$resultcorrecthybrid)
hetaccuracy$resultcorrectparents <- as.numeric(hetaccuracy$resultcorrectparents)
hetaccuracy$hetfilenames <- as.character(hetaccuracy$hetfilenames)
hetaccuracy

```

```

##   hetfilenames resultcorrecthybrid resultcorrectparents
## 1      samBC           82.70440           91.00346
## 2      samBM           78.68020           88.06818
## 3      gatBC           97.66355           91.26638
## 4      gatBM           95.08197           93.54839

```

```
write.csv(hetaccuracy, file="hetaccuracies.csv")
```

for table with averages

```

samavghyb <- (hetaccuracy[1,2]+hetaccuracy[2,2])/2
gatavghyb <- (hetaccuracy[3,2]+hetaccuracy[4,2])/2
samavgpar <- (hetaccuracy[1,3]+hetaccuracy[2,3])/2
gatavgpar <- (hetaccuracy[3,3]+hetaccuracy[4,3])/2

```

output table and file with averages

```
tool <- list("samavg", "gatavg")
correcthyb <- list(samavghyb, gatavghyb)
correctpar <- list(samavgpar, gatavgpar)
averages <- as.data.frame(cbind(tool, correcthyb, correctpar))
averages$correcthyb <- as.numeric(averages$correcthyb)
averages$correctpar <- as.numeric(averages$correctpar)
averages$tool <- as.character(averages$tool)
averages
```

```
##      tool correcthyb correctpar
## 1 samavg   80.69230   89.53582
## 2 gatavg   96.37276   92.40738
```

```
write.csv(averages, file="hetaccuraciesaverages.csv")
```