# Pragmatic R for Biologists

10/22/10

# R

- An environment for statistical computing
  - Statistics
  - Visualization

# Strengths and Weaknesses

- Great for
  - Statistics
  - Graphics
  - Tabular data
  - Reproducible research
- Not great for
  - Parsing, string manipulation, web
- Also...
  - Functional programming is weird
  - Monolithic memory
  - Confusing gotchas

# Online Resources

- http://stackoverflow.com/questions/tagged/r
- http://stackoverflow.com/questions/tagged/ggplot2
- http://stats.stackexchange.com/
- http://onertipaday.blogspot.com/

# Getting R

- http://www.r-project.org/
- Essential cheat sheet
  - http://cran.r-project.org/doc/contrib/Short-refcard.pdf

# Getting help with R

- ?melt
- ls("package:plyr")
- vignette()

# Using R Effectively

- Interactive conversation

```
savehistory(file = "myWork.Rhistory")
```

- q() to quit

# R Basic Operations

```
> 1+2+3

[1] 6

> 1+2*3

[1] 7

> (1+2)*3

[1] 9

> c(0,1,1,2,3,5,8)

[1] 0 1 1 2 3 5 8

> str(c(0,1,1,2,3,5,8))

 num [1:7] 0 1 1 2 3 5 8

> 1:10

 [1]  1  2  3  4  5  6  7  8  9 10

> str(1:10)

 int [1:10] 1 2 3 4 5 6 7 8 9 10

> c(1,2,3,4)+1

[1] 2 3 4 5
```

```
> c(1,2,3,4)*2

[1] 2 4 6 8

> theCount<-c("one","two","three")

> str(theCount)

 chr [1:3] "one" "two" "three"

> 3==4

[1] FALSE

> b<-(1:12)

> b

 [1]  1  2  3  4  5  6  7  8  9 10 11 12

> b[7]

[1] 7

> b[-1]

 [1]  2  3  4  5  6  7  8  9 10 11 12

> b[-2]

 [1]  1  3  4  5  6  7  8  9 10 11 12
```

```
> b[-7]

 [1]  1  2  3  4  5  6  8  9 10 11 12

> b[ b %% 2 == 0]

[1]  2  4  6  8 10 12

> b %% 2

 [1] 1 0 1 0 1 0 1 0 1 0 1 0

> b %% 2 == 0

 [1] FALSE  TRUE FALSE  TRUE FALSE

 TRUE FALSE  TRUE FALSE

 TRUE FALSE  TRUE
```

# R in a Nutshell Chapter 3

# Getting stuff into R and out of R

read.table()

```
read.table(file, header = FALSE, sep = "", quote = "\"'",
           dec = ".", row.names, col.names,
           as.is = !stringsAsFactors,
           na.strings = "NA", colClasses = NA, nrows = -1,
           skip = 0, check.names = TRUE, fill = !blank.lines.skip,
           strip.white = FALSE, blank.lines.skip = TRUE,
           comment.char = "#",
           allowEscapes = FALSE, flush = FALSE,
           stringsAsFactors = default.stringsAsFactors(),
           fileEncoding = "", encoding = "unknown")
```

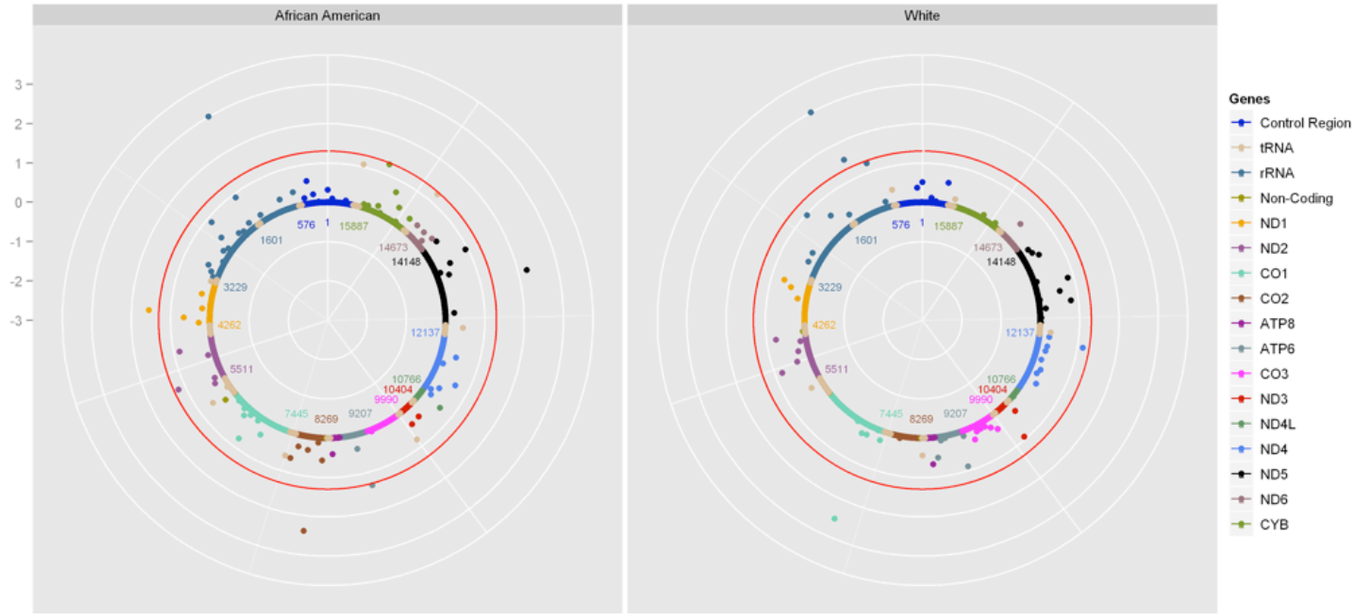
load()

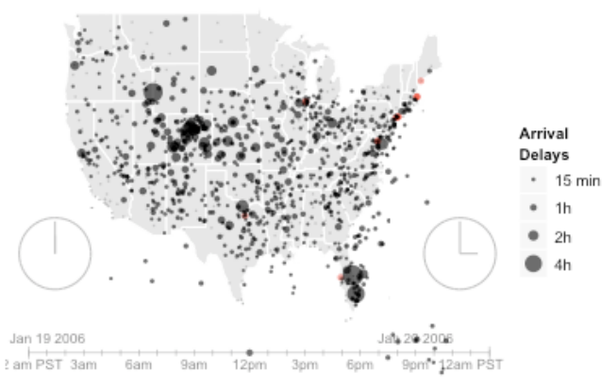
save

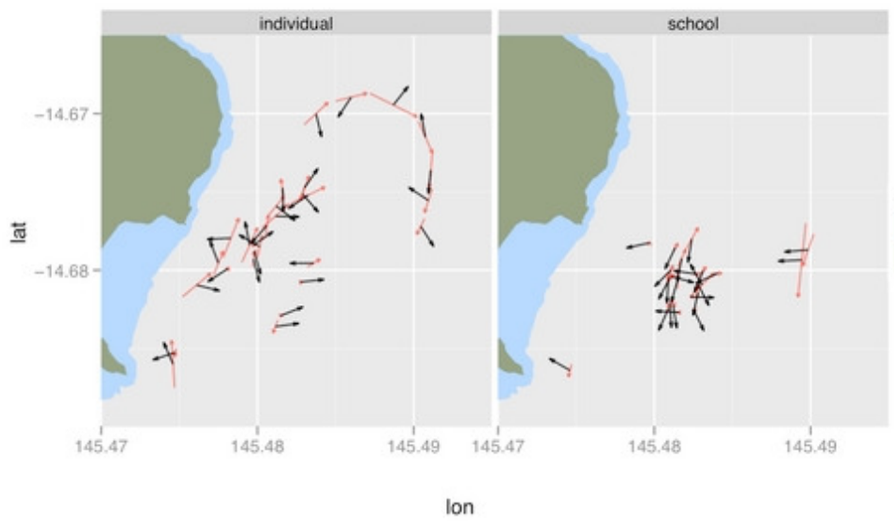
```
save(..., list = character(0L),
     file = stop("'file' must be specified"),
     ascii = FALSE, version = NULL, envir = parent.frame(),
     compress = !ascii, compression_level,
     eval.promises = TRUE, precheck = TRUE)
```

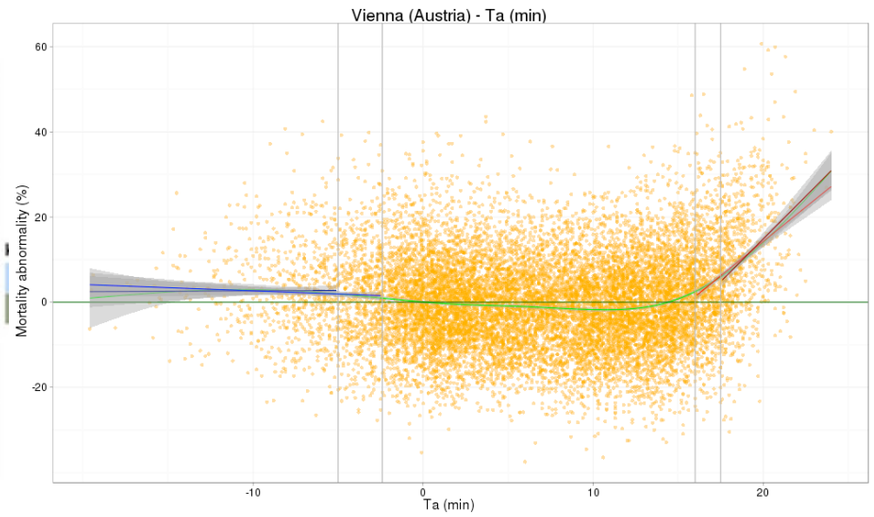
# ggplot2

- Layered
- Sensible defaults



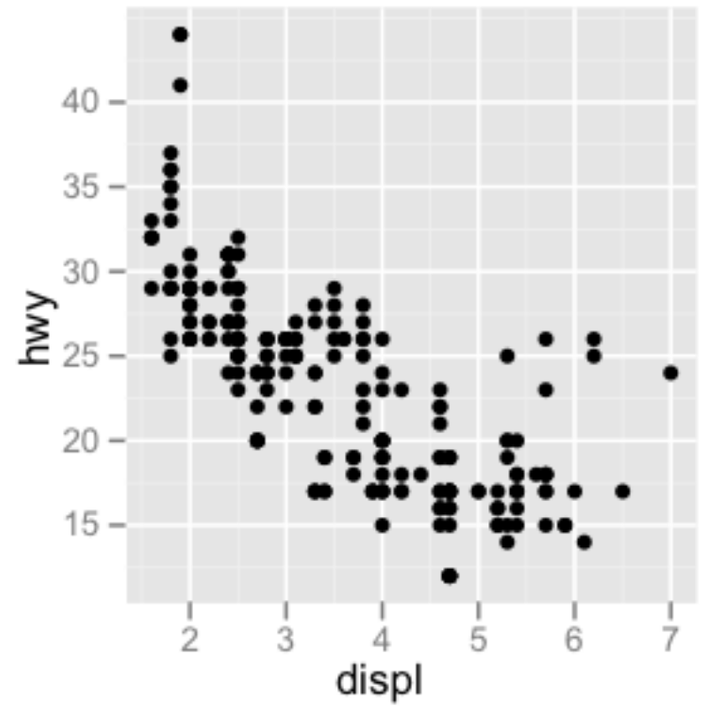

http://github.com/hadley/ggplot2/wiki

# Getting ggplot2

```
> install.packages("ggplot2")
> library("ggplot2")
> data()
> ?mpg
> head(mpg)
> str(mpg)
```

*this is like Data Dumper*

```
> qplot(displ,hwy,data=mpg)
```




http://www.slideshare.net/hadley/01-intro-5041287

# Aesthetics and facetting

- qplot(displ,hwy,data=mpg,color=class)
- qplot(displ,hwy,data=mpg)+facet_grid(. ~ cyl)

http://www.slideshare.net/hadley/01-intro-5041287

# Reorder

- qplot(class,hwy,data=mpg)
- qplot(reorder(class,hwy),hwy,data=mpg)
- qplot(reorder(class,hwy),hwy,data=mpg,geom="boxplot")

# Fun with Diamonds

- # With only one variable, qplot guesses that # you want a bar chart or histogram qplot(cut, data = diamonds)
- qplot(carat, data = diamonds)
- qplot(carat, data = diamonds, binwidth = 1)
- qplot(carat, data = diamonds, binwidth = 0.1)
- qplot(carat, data = diamonds, binwidth = 0.01)
- last_plot() + xlim(0, 3)
- qplot(depth,data=diamonds,binwidth=0.2) + xlim(55, 70) + facet_wrap(~ cut)

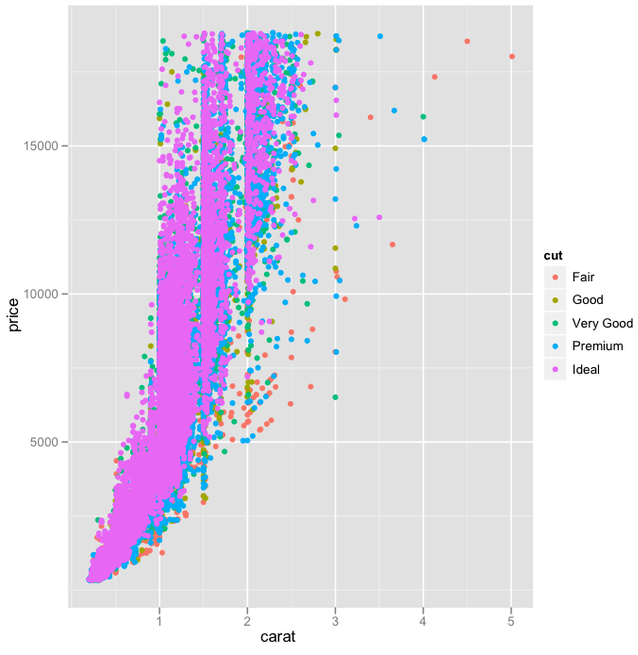
http://had.co.nz/stat405/lectures/02-large.pdf

# ggplot2 grammar

```
qplot(carat,price, data = diamonds, color=cut)
```

## is equivalent to

```
p<-ggplot(diamonds,aes(carat,price,color=cut))+layer(geom="point")
```




Wickham's ggplot pg42

# Sweave

R + LaTeX = .pdf

- A pretentious mix of R and LaTeX designed to intimidate reviewers
- in R: Sweave(“myReport.Rnw”,output=”myReport.tex”)
- pdflatex myReport.tex

From shell:

```
echo "Sweave(\"myReport.Rnw\",output=\"myReport.tex\");" | R --no-save --quiet
```

# 2009 RNAseq Standard Report

```
\documentclass{article}
\usepackage{longtable}
\usepackage{rotating}
\begin{document}

\title{RNAseq Standard Report}
\author{Benjamin Blackman\\Aurelie Bonin\\Suzanne Joneson\
\Eduardo Reis\\ \\TA: Jeremy Leipzig}

\maketitle

\tableofcontents
\pagebreak
\section{Introduction}
The objective of this project was to align and analyze a short
read dataset obtained by Illumina cDNA sequencing. The study
species is \emph{Mimulus guttatus} (monkey flower).


<<basicStuff,echo=FALSE>>=
library(xtable)
library(ggplot2)
@
```

RNAseq Standard Report

Benjamin Blackman
Aurelie Bonin
Suzanne Joneson
Eduardo Reis

TA: Jeremy Leipzig

October 27, 2009

## Contents



## 1 Introduction

The objective of this project was to align and analyze a short read dataset obtained by Illumina cDNA sequencing. The study species is *Mimulus guttatus* (monkey flower).

# Report_1.txt



| rdName | totRds | bpReads | refName | refBpLength | mappedRds | bpmappedRds | predName | numPredGenes | bpPredGenes | rdsMapPredGe | bpmapPredGer | knownName | numKnownGen | bpKnownGenes | rdsMapKnown( | bpmapKnownG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| mimulus_scaff1 | 30246800 | 2268510000 | scaffold_1 | 4921665 | 113123 | 8484225 | scaffold_1.pred | 330 | 453541 | 11390 | 854250 | scaffold_1.annc | 701 | 1133679 | 122646 | 9198450 |
| mimulus_scaff1 | 30246800 | 2268510000 | scaffold_1 | 4921665 | 79273 | 5945475 | scaffold_1.pred | 330 | 453541 | 8131 | 609825 | scaffold_1.annc | 701 | 1133679 | 86286 | 6471450 |

```
\section{Short read dataset and reference sequence - Basic
information}
Here is some basic information on the short read dataset and the
reference sequence used to perform the alignment:\\
\\
<<table1,echo=FALSE,fig=FALSE>>=
table1<-read.table("Report_1.txt")
totRds <- table1[2,2]
bpReads <-table1[2,3]
RefbpLength <-table1[2,5]
NumPredicted <-table1[2,9]
LengthPredicted <-table1[2,10]
NumKnown <-table1[2,14]
LengthKnown <-table1[2,15]
@
\textbf{Short read dataset}\\
Number of reads: \Sexpr{totRds}\\
Total length in bp: \Sexpr{bpReads}\\
\\
\textbf{Reference sequence}\\
Total length in bp: \Sexpr{RefbpLength}\\
Number of predicted genes: \Sexpr{NumPredicted}\\
Total length of predicted gene sequence in bp: \Sexpr
{LengthPredicted}\\
Number of annotated genes: \Sexpr{NumKnown}\\
Total length of annotated gene sequence in bp: \Sexpr
{LengthKnown}\\

\section{Comparison of alignments obtained with different
programs}
We tested two programs (Bowtie and BWA) to align the short reads
to the reference sequence. BWA allows for indels whereas Bowtie
doesn't. Bowtie was used for the rest of the analyses.\\


<<table1,echo=FALSE,fig=FALSE>>=
tableReport <- subset (table1, select = c(rdName, mappedRds,
bpmappedRds))
colnames (tableReport)<-c('Program','Number of aligned
reads','Total length in bp ')
tableReport [,1] <-c('BWA', 'Bowtie')
@

<<tableSetup2,results=tex,echo=FALSE>>=
myXtable<-xtable(tableReport,type=tex,caption="BWA and Bowtie -
Performance comparison")
print(myXtable,include.rownames=FALSE)
@
```

## 2 Short read dataset and reference sequence - Basic information

Here is some basic information on the short read dataset and the reference sequence used to perform the alignment:

**Short read dataset**
Number of reads: 30246800
Total length in bp: 2268510000

**Reference sequence**
Total length in bp: 4921665
Number of predicted genes: 330
Total length of predicted gene sequence in bp: 453541
Number of annotated genes: 701
Total length of annotated gene sequence in bp: 1133679

## 3 Comparison of alignments obtained with diferent programs

We tested two programs (Bowtie and BWA) to align the short reads to the reference sequence. BWA allows for indels whereas Bowtie doesn't. Bowtie was used for the rest of the analyses.

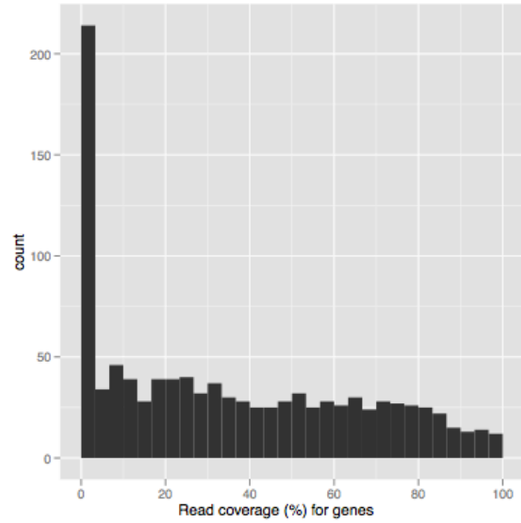
| Program | Number of aligned reads | Total length in bp |
|---|---|---|
| BWA | 113123 | 8484225 |
| Bowtie | 79273 | 5945475 |

Table 1: BWA and Bowtie - Performance comparison

```
\pagebreak
\section{Read coverage - Genes}
The following histogram presents the distribution of read coverages for genes,
as calculated by the percentage of a gene's coding sequences with at least one
read aligned.

<<genereads,echo=FALSE,fig=TRUE>>=
table3<-read.table("Jerm_table3.txt",header=TRUE)
p<-qplot(X.CDS_COVERAGE,data=table3,geom="histogram",xlab="Read coverage (%)
for genes", ylab= "Gene count")
print(p)
@

\pagebreak
\section{Read coverage - Intergenic regions}
The following histogram presents the distribution of read coverages for
intergenic regions, as calculated by the percentage of the region with at least
one read aligned.

<<intergenicreads,echo=FALSE,fig=TRUE>>=
table4<-read.table("table4intergenicredo.txt",header=TRUE)
r<-qplot((INTERGENIC_COVERAGE)*100,data=table4,geom="histogram",xlab="Read
coverage (%) for intergenic regions", ylab= "Intergenic region count")
print(r)
@
\pagebreak
\section{Length of intergenic regions}
The following histogram presents the distribution of intergenic lengths.

<<intergenicregion,echo=FALSE,fig=TRUE>>=
table4<-read.table("table4intergenicredo.txt",header=TRUE)
r<-qplot(INTERGENIC_LENGTH,data=table4,geom="histogram",xlab="Intergenic length
in bp", ylab= "Intergenic region count")
print(r)
@
```

## 4 Read coverage - Genes

The following histogram presents the distribution of read coverages for genes, as calculated by the percentage of a gene's coding sequences with at least one read aligned.

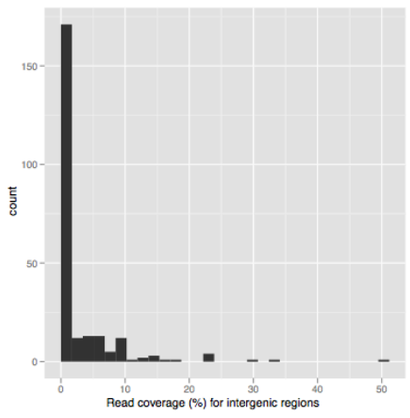



| READ_NAME | GENE_NAME | GENE_LENGTI | CDS_LENGTH | %CDS_COVEI |
|---|---|---|---|---|
| Mimulus_reads | mgf021408m | 607 | 576 | 36.805555556 |
| Mimulus_reads | mgf024713m | 623 | 265 | 0 |
| Mimulus_reads | mgf003544m | 244 | 243 | 67.489711934 |
| Mimulus_reads | mg021480m | 366 | 366 | 0 |
| Mimulus_reads | mgf023543m | 410 | 349 | 0 |
| Mimulus_reads | mg012117m | 298 | 297 | 0 |
| Mimulus_reads | mgf019623m | 440 | 409 | 0 |
| Mimulus_reads | mgf022738m | 393 | 393 | 0 |
| Mimulus_reads | mgf019826m | 504 | 473 | 0 |
| Mimulus_reads | mgf022195m | 604 | 321 | 43.613707165 |
| Mimulus_reads | mgf012634m | 744 | 742 | 94.339622642 |
| Mimulus_reads | mg014291m | 849 | 262 | 100 |
| Mimulus_reads | mg015640m | 677 | 677 | 11.225997046 |

Jerm_table3.txt

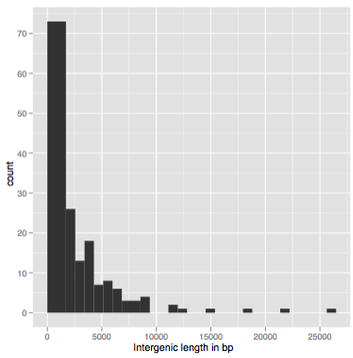
## 5 Read coverage - Intergenic regions

The following histogram presents the distribution of read coverages for intergenic regions, as calculated by the percentage of the region with at least one read aligned.




## 6 Length of intergenic regions

The following histogram presents the distribution of intergenic lengths.

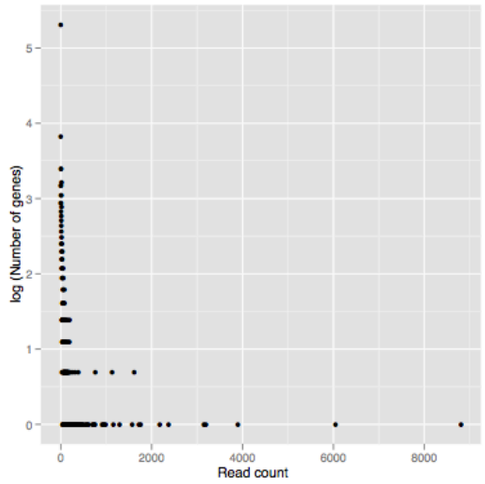


```
\pagebreak
\section{Number of genes per read count}
The following histogram presents the number of genes having a given read count.

<<genenumber,echo=FALSE,fig=TRUE>>=
table5<-read.table("Table5.txt", header=TRUE)
q<-qplot(Read_count,log (Genes_number), data=table5,xlab="Read count", ylab="log
(Number of genes)")
print(q)
@
```

| Read_set_name | Read_count | Genes_number |
|---|---|---|
| Mimulus_reads | 0 | 202 |
| Mimulus_reads | 1 | 19 |
| Mimulus_reads | 2 | 46 |
| Mimulus_reads | 3 | 24 |
| Mimulus_reads | 4 | 30 |
| Mimulus_reads | 5 | 24 |
| Mimulus_reads | 6 | 17 |
| Mimulus_reads | 7 | 30 |
| Mimulus_reads | 8 | 19 |
| Mimulus_reads | 9 | 21 |
| Mimulus_reads | 10 | 21 |
| Mimulus_reads | 11 | 16 |
| Mimulus_reads | 12 | 14 |
| Mimulus_reads | 13 | 15 |
| Mimulus_reads | 14 | 11 |
| Mimulus_reads | 15 | 13 |
| Mimulus_reads | 16 | 10 |

## 7 Number of genes per read count

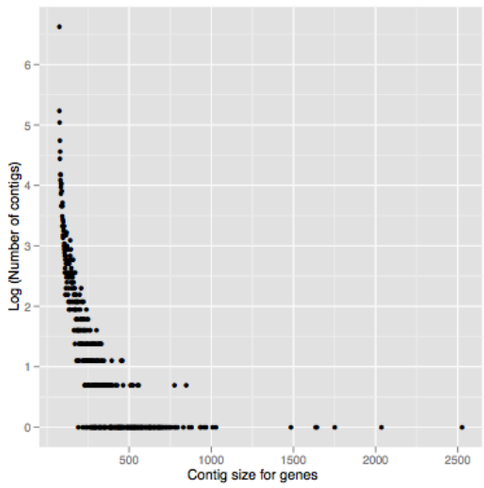
The following histogram presents the number of genes having a given read count.




| Contig_Size | Number_Contig |
|---|---|
| 76 | 755 |
| 77 | 188 |
| 78 | 155 |
| 79 | 115 |
| 80 | 85 |
| 81 | 96 |
| 82 | 60 |
| 83 | 66 |
| 84 | 65 |

## 8 Distribution of contig sizes - Genes

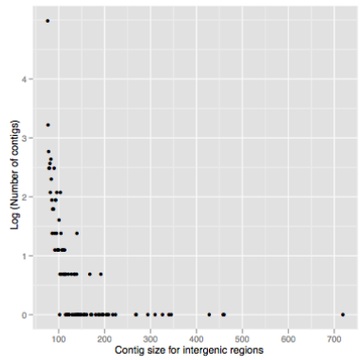
The following histogram presents the distribution of contig sizes for genes.




```
\pagebreak
\section{Distribution of contig sizes - Genes}
The following histogram presents the distribution of contig sizes for genes.

<<genecontigs,echo=FALSE,fig=TRUE>>=
ContigRNA<-read.table("table4contigsize_output_bowtie_mRNAs.txt",header=TRUE)
s<-qplot(Contig_Size,log(Number_Contigs),data=ContigRNA,xlab="Contig size for genes",
ylab= "Log (Number of contigs)")
print(s)
@
\pagebreak


\section{Distribution of contig sizes - Intergenic regions}
The following histogram presents the distribution of contig sizes for intergenic
regions.

<<intercontigs,echo=FALSE,fig=TRUE>>=
ContigInter<-read.table("table4contigsize_output_bowtie_intergenic.txt",header=TRUE)
t<-qplot(Contig_Size,log(Number_Contigs),data=ContigInter,xlab="Contig size for
intergenic regions", ylab= "Log (Number of contigs)")
print(t)
@
\pagebreak
\end{document}
```

## 9 Distribution of contig sizes - Intergenic regions

The following histogram presents the distribution of contig sizes for intergenic regions.




| Contig_Size | Number_Contig |
|---|---|
| 76 | 147 |
| 77 | 25 |
| 78 | 16 |
| 79 | 12 |
| 80 | 12 |
| 81 | 13 |
| 82 | 8 |
| 83 | 14 |
| 84 | 10 |
| 85 | 7 |

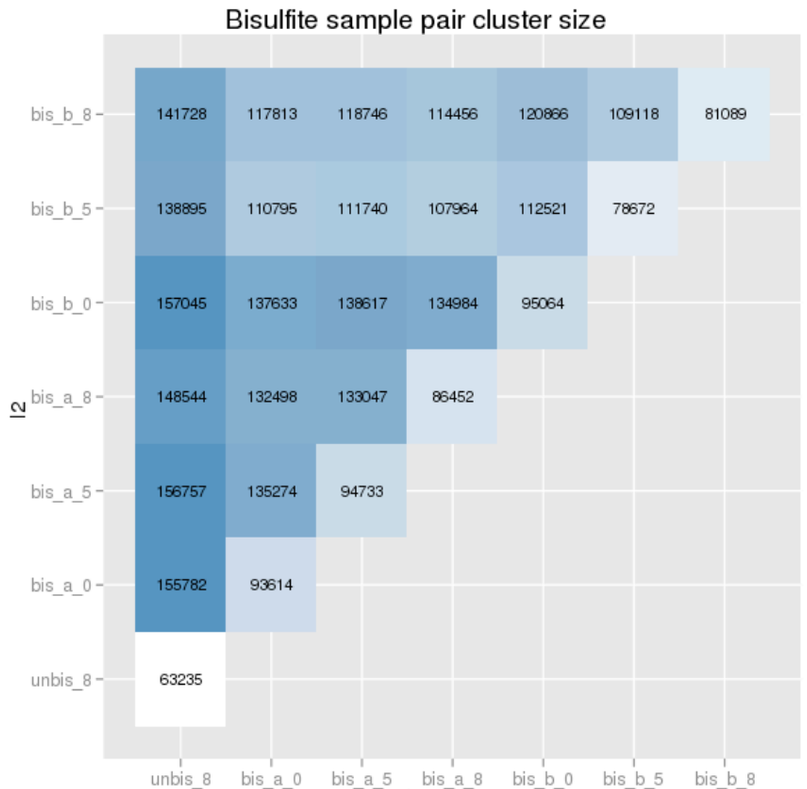
# Bisulfite Analysis

## matrix.txt

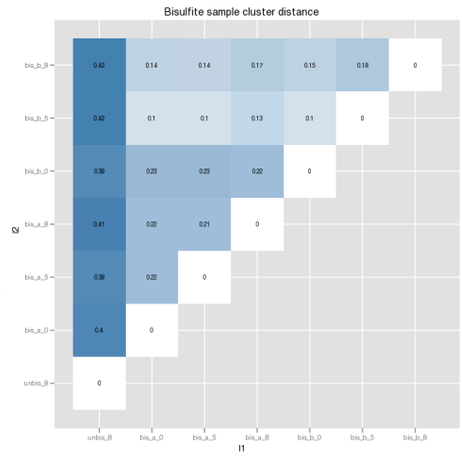
| | | |
|---|---|---|
| 1 | 1 | 63235 |
| 1 | 2 | 155782 |
| 1 | 3 | 156757 |
| 1 | 5 | 148544 |
| 1 | 6 | 157045 |
| 1 | 7 | 138895 |
| 1 | 8 | 141728 |
| 2 | 2 | 93614 |
| 2 | 3 | 135274 |
| 2 | 5 | 132498 |
| 2 | 6 | 137633 |
| 2 | 7 | 110795 |
| 2 | 8 | 117813 |
| 3 | 3 | 94733 |

## lanes.txt

| | | |
|---|---|---|
| 1 | unbis | 8 |
| 2 | bis_a | 0 |
| 3 | bis_a | 5 |
| 5 | bis_a | 8 |
| 6 | bis_b | 0 |
| 7 | bis_b | 5 |
| 8 | bis_b | 8 |




```
library(ggplot2)
bisulfite<-read.table("matrix.txt",col.names=c("l1","l2","clusters"))
bisulfite$l1<-as.factor(bisulfite$l1)
bisulfite$l2<-as.factor(bisulfite$l2)
samples<-read.table("lanes.txt",col.names=c("lane","bis","phu"))
scale_labels<-paste(samples$bis,samples$phu,sep="_")

#you'll need this on mac
X11(type ="cairo")

#let's build this slowly
p<-ggplot(bisulfite,aes(l1,l2))+geom_tile(aes(fill=clusters))
p<-p+scale_fill_gradient(limits = c(min(bisulfite$clusters), max(bisulfite$clusters)), low="white",high="steelblue")
p<-p+scale_x_discrete(breaks=levels(bisulfite$l1),labels=scale_labels)
p<-p+scale_y_discrete(breaks=levels(bisulfite$l2),labels=scale_labels)
p<-p+geom_text(aes(size=3,label=clusters))+opts(title = "Bisulfite sample pair cluster size")+opts(legend.position = "none")

print(p)

savePlot(filename="plot.png",type="png")

selfLane<-function(x){bisulfite$clusters[bisulfite$l1 == x & bisulfite$l2 == x]}

bisulfite$allCor<-mapply(max,laply(bisulfite$l1,selfLane),laply(bisulfite$l2,selfLane))
bisulfite$noCor<-laply(bisulfite$l1,selfLane)+laply(bisulfite$l2,selfLane)
bisulfite$dist<-(bisulfite$clusters-bisulfite$allCor)/bisulfite$noCor

p<-ggplot(bisulfite,aes(l1,l2))+geom_tile(aes(fill=dist))
print(p+scale_fill_gradient(limits = c(min(bisulfite$dist), max(bisulfite$dist)), low="white",high="steelblue")+scale_x_discrete(breaks=levels(bisulfite
$l1),labels=scale_labels)+scale_y_discrete(breaks=levels(bisulfite$l2),labels=scale_labels)+geom_text(aes(size=3,label=round(dist,2)))+opts
(legend.position = "none")+opts(title = "Bisulfite sample cluster distance"))
```



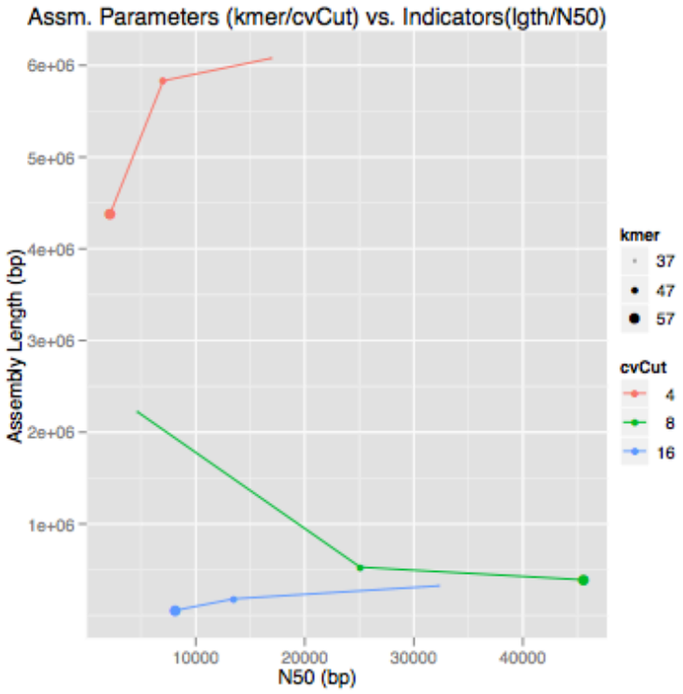
# SVAR

## 3 Effect of kmer and cvCut on Assembly Profile

### 3.1 Assembly coherency

The following scatterplot illustrates the effect of these variables on N50 and assembly length. N50 refers to the length of the shortest contig in an assembly such that the sum of contigs of equal length or longer is at least 50% of the total length of all contigs.

We can plot this to show the effect of kmer value (as progressively larger points) and coverage cutoff (strata connected by lines) on N50 and assembly length:




```
> load("BD574-2.frame.RData")
> ls()
[1] "myDataFrame"
> myDataFrame
    totBP   reads   tiles goodContigs kmer cvCut expCov                groupDir ctgs meanLgth medianLgth
1  323399 1263928  170121          57   37    16     32 out_BD574-2_37_16_32_dir   80 4057.925      143.5
2 6058426 1263928 1092568         998   37     4      8   out_BD574-2_37_4_8_dir 1390 4372.006      370.0
3 2214471 1263928  547583        1176   37     8     16  out_BD574-2_37_8_16_dir 1296 1713.209      606.0
4  182324 1263928  111783          33   47    16     32 out_BD574-2_47_16_32_dir   43 4256.372      596.0
5 5820563 1263928  985530        1584   47     4      8   out_BD574-2_47_4_8_dir 1772 3291.328     1773.5
6  525714 1263928  208144         212   47     8     16  out_BD574-2_47_8_16_dir  252 2097.060      325.5
7   56002 1263928   42878          19   57    16     32 out_BD574-2_57_16_32_dir   22 2556.727      613.0
8 4379805 1263928  725921        3020   57     4      8   out_BD574-2_57_4_8_dir 3094 1417.416      907.0
9  388091 1263928  161205          54   57     8     16  out_BD574-2_57_8_16_dir   76 5130.447      187.0
  totalCoverage maxLgth over1k   N50
1        324634   61354     24 32297
2       6077088   75922    585 16918
3       2220319  101303    488  4677
4        183024   60374     19 13476
5       5832234   46530   1139  7018
6        528459  101732     51 25073
7         56248   13848      9  8116
8       4385486   63866   1438  2166
9        389914  101691     22 45544
```

```
<<plotCov,fig=TRUE,echo=FALSE>>=
kmerDotSizes<-seq(1:length(levels(as.factor(myDataFrame$kmer))))*1
p<-ggplot(myDataFrame[order(myDataFrame$cvCut),],aes(N50,totalCoverage))+
geom_point(aes(size=as.factor(kmer))) +
aes(colour=as.factor(cvCut)) +
scale_size_manual(name="kmer",value = kmerDotSizes) +
geom_line() +
scale_colour_discrete("cvCut") +
xlab("N50 (bp)") +
ylab("Assembly Length (bp)") +
opts(title="Assm. Parameters (kmer/cvCut) vs. Indicators(lgth/N50)")
print(p)
@
```

# Five Prime Enrichment

## leftReads

| pco | pos | tLength | depth |
|---|---|---|---|
| pco500123 | 124 | 686 | 1 |
| pco500123 | 125 | 686 | 1 |
| pco500123 | 126 | 686 | 1 |
| pco500123 | 127 | 686 | 1 |
| pco500123 | 128 | 686 | 1 |
| pco500123 | 129 | 686 | 1 |
| pco500123 | 130 | 686 | 1 |
| pco500123 | 131 | 686 | 1 |

```
library(xtable)
geneDefs<-read.table("geneDefs.txt",sep="\t",header=TRUE,quote='"')
geneDefs$desc<-gsub("Top Blast","",geneDefs$desc)
top20left<-sort(xtabs(depth ~ pco,data=leftReads),decreasing=TRUE)[1:20]
top20leftpcos<-attr(top20left,"dimnames")
displayTableLeft<-merge(as.data.frame(top20leftpcos),geneDefs)
displayTableLeft$desc<-substr(displayTableLeft$desc,1,50)
```

## geneDefs

| pco | desc |
|---|---|
| pco500123 | top blast hypothetical protein |
| pco500134 | zinc ion binding |
| pco644893 | alcohol dehydrogenase |

```
<<top20leftcounts,echo=FALSE,results=tex>>=
xtable(as.data.frame(top20left))
@
```

### 2.1.1 Left reads

The counts here represent total base pairs aligned to that pco. To obtain the reads aligned, divide by average read length.

| | top20left |
|---|---|
| pco598314 | 12610568.00 |
| pco644849 | 10710320.00 |
| pco640072 | 7495096.00 |
| pco651827 | 7476770.00 |
| pco647057 | 7234518.00 |
| pco642907 | 6980828.00 |
| pco664216 | 6651398.00 |
| pco591868 | 5847596.00 |

With gene definitions:

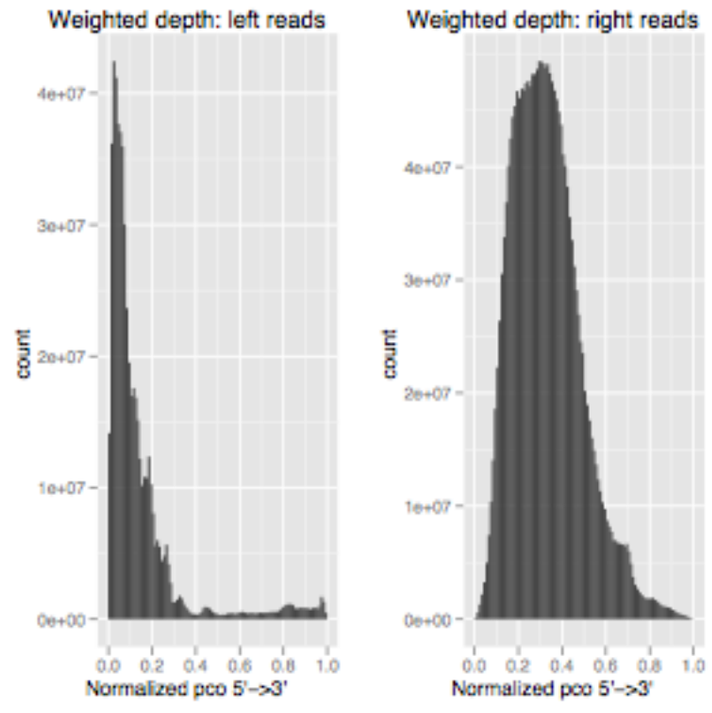
```
<<top20leftdisp,echo=FALSE,results=tex>>=
xtable(displayTableLeft)
@
```

With gene definitions:

| | pco | desc |
|---|---|---|
| 1 | pco522245 | Q8W1C9 Ribosomal protein L35A [Zea mays] E=3e-73 |
| 2 | pco591868 | P45633 60S ribosomal protein L10 [Zea mays] E=1e- |
| 3 | pco595782 | Q6ZJC7 Early nodulin 75-like protein [Oryza sativ |
| 4 | pco598225 | Q41852 40S ribosomal protein S21 [Zea mays] E=6e- |

# Five Prime Enrichment

```
<<prepHisto,echo=FALSE>>=
library("ggplot2")
vplayout <- function(x, y){viewport(layout.pos.row = x, layout.pos.col = y)}

w<-function(myReads,mySide){
qplot(nPos,data=myReads,binwidth=.01,weight=depth,xlim=c(0,1),xlab=("Normalized pco 5'->3'"),main=(paste("Weighted depth:",mySide,"reads")))}

@

<<sidesWeighted,fig=TRUE,echo=FALSE>>=
grid.newpage()
pushViewport(viewport(layout=grid.layout(1,2)))

leftPlot<-w(leftReads,"left")
print(leftPlot,vp=vplayout(1,1))

rightPlot<-w(rightReads,"right")
print(rightPlot,vp=vplayout(1,2))
@
```



# kmer analysis

| | kmer | count | sample | conc |
|---|---|---|---|---|
| 1 | 1 | 2 | b73_mcrbc | 200 |
| 2 | 2 | 10 | b73_mcrbc | 200 |
| 3 | 3 | 32 | b73_mcrbc | 200 |
| 4 | 4 | 136 | b73_mcrbc | 200 |
| 5 | 5 | 512 | b73_mcrbc | 200 |
| 6 | 6 | 2080 | b73_mcrbc | 200 |
| 7 | 7 | 8192 | b73_mcrbc | 200 |
| 8 | 8 | 32896 | b73_mcrbc | 200 |
| 9 | 9 | 131072 | b73_mcrbc | 200 |
| 10 | 10 | 524800 | b73_mcrbc | 200 |
| 11 | 11 | 2093412 | b73_mcrbc | 200 |
| 12 | 12 | 7639104 | b73_mcrbc | 200 |
| 13 | 13 | 18675851 | b73_mcrbc | 200 |
| 14 | 14 | 28596584 | b73_mcrbc | 200 |
| 15 | 15 | 33438709 | b73_mcrbc | 200 |
| 16 | 16 | 35081294 | b73_mcrbc | 200 |
| 17 | 17 | 35452389 | b73_mcrbc | 200 |
| 18 | 18 | 35377437 | b73_mcrbc | 200 |
| 19 | 19 | 35136629 | b73_mcrbc | 200 |

```
load("sampkmer.results.RData")

sampkmer$conc<-factor(sampkmer$conc,levels=kmerLevels)
q<-qplot(kmer,count,color=sample,data=sampkmer,facets = conc ~ .)
print(q)
```