

Lab 13: UNIX for Bioinformatics

Nicholas Do (PID: A15053002)

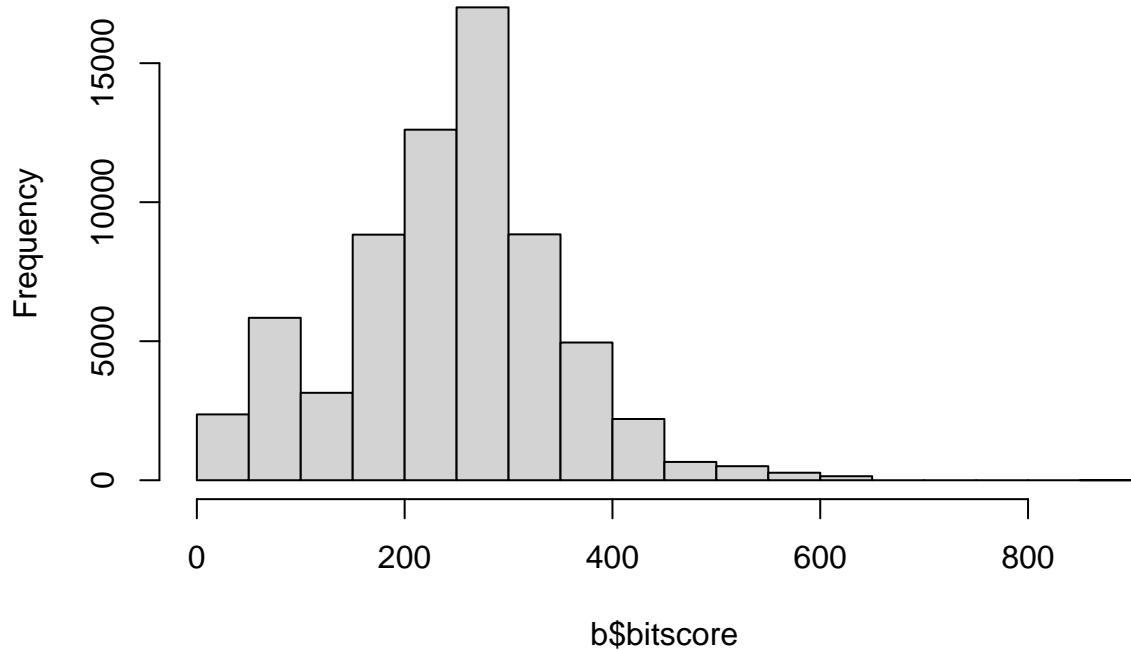
11/9/2021

```
b <- read.delim("mm-second.x.zebrafish.tsv", col.names = c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "sstart", "send", "evalue", "bitscore"))
head(b)

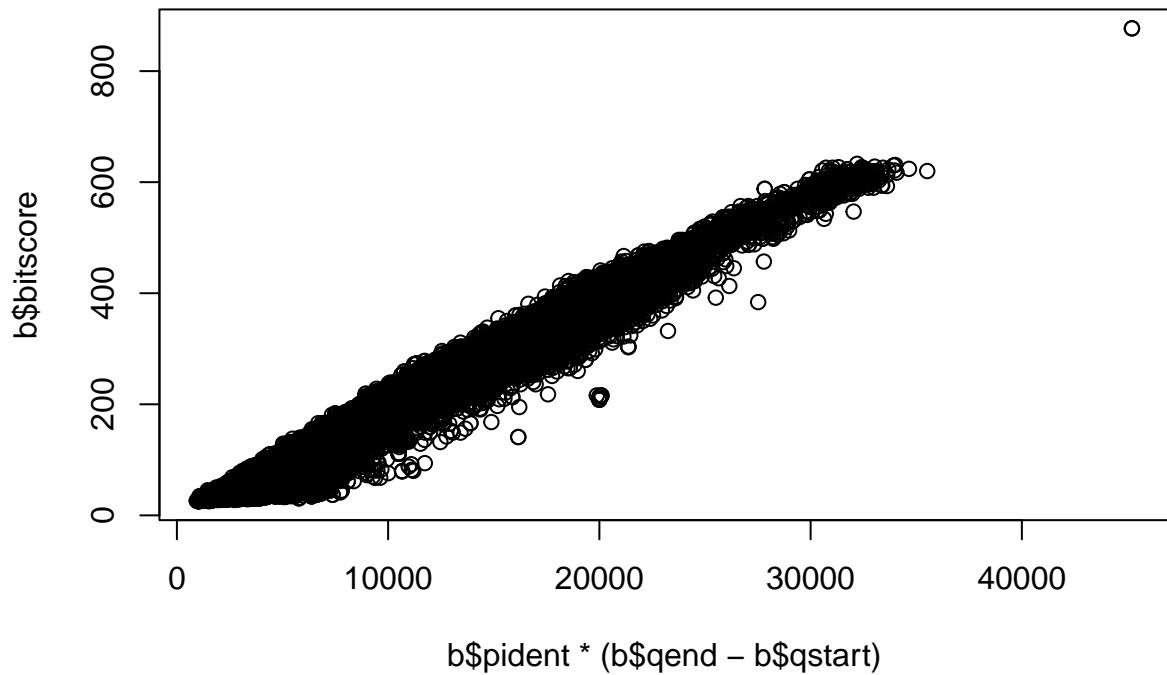
##      qseqid      sseqid pident length mismatch gapopen qstart qend sstart
## 1 YP_220551.1 NP_059332.1 44.509     346      188       3      1 344      1
## 2 YP_220551.1 NP_059341.1 24.540     163      112       3     112 263     231
## 3 YP_220551.1 NP_059340.1 26.804      97       65       2      98 188     200
## 4 YP_220552.1 NP_059333.1 88.132     514       61       0      1 514      1
## 5 YP_220552.1 XP_021326074.1 31.818      66       32       2     427 482      16
## 6 YP_220552.1 NP_001373511.1 31.818      66       32       2     427 482      48
##      send      evalue bitscore
## 1    344 8.62e-92    279.0
## 2    393 5.14e-06    49.7
## 3    296 1.00e-01    35.8
## 4    514 0.00e+00   877.0
## 5     78 6.70e+00    29.3
## 6   110 7.50e+00    29.6

hist(b$bitscore, breaks = 30)
```

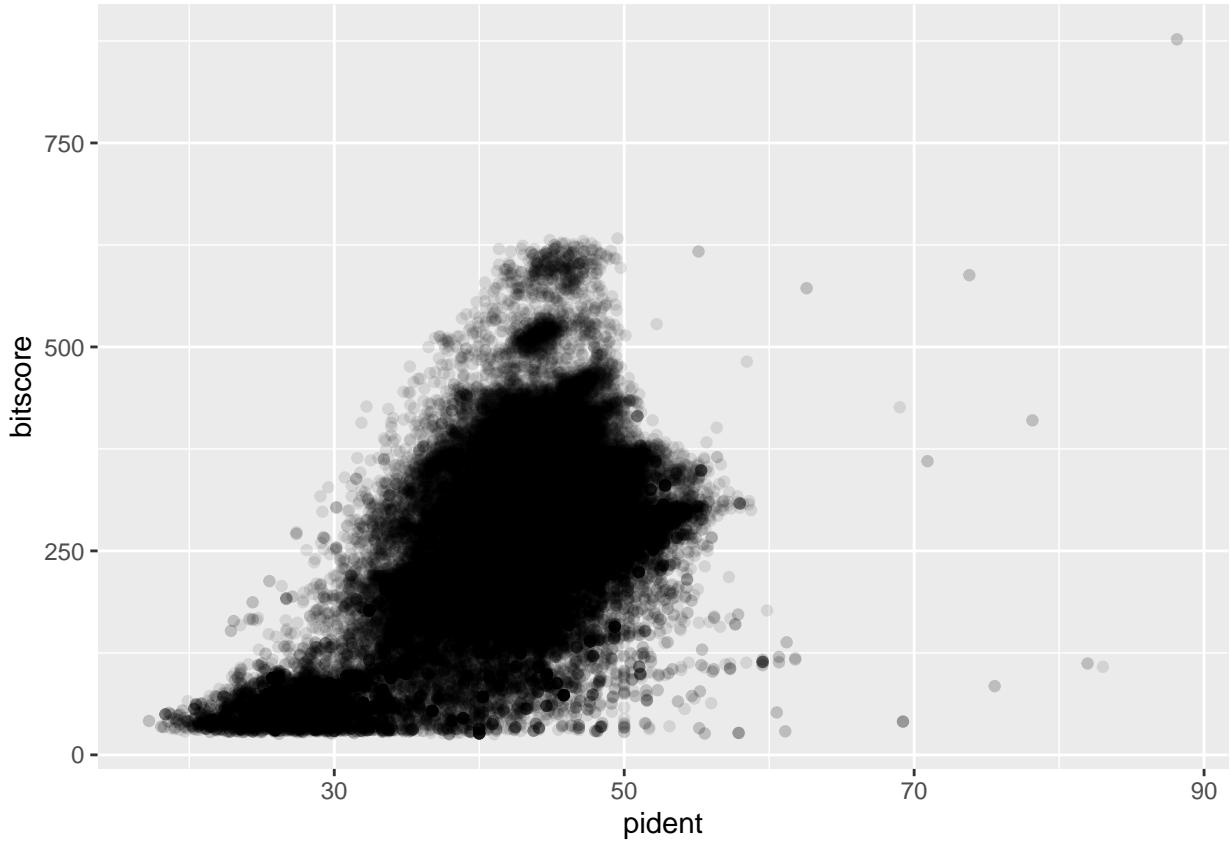
Histogram of b\$bitscore



```
## Assuming your blast results are stored in an object called 'b'  
plot(b$pident * (b$qend - b$qstart), b$bitscore)
```



```
library(ggplot2)
ggplot(b, aes(pident, bitscore)) + geom_point(alpha=0.1)
```



```

ggplot(b, aes((b$pident * (b$qend - b$qstart)), bitscore)) + geom_point(alpha=0.1) + geom_smooth()

## Warning: Use of 'b$pident' is discouraged. Use 'pident' instead.

## Warning: Use of 'b$qend' is discouraged. Use 'qend' instead.

## Warning: Use of 'b$qstart' is discouraged. Use 'qstart' instead.

## Warning: Use of 'b$pident' is discouraged. Use 'pident' instead.

## Warning: Use of 'b$qend' is discouraged. Use 'qend' instead.

## Warning: Use of 'b$qstart' is discouraged. Use 'qstart' instead.

## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

```

