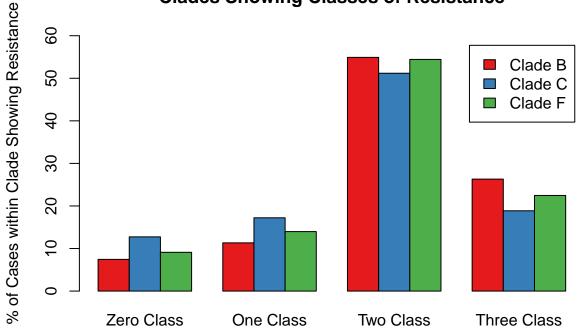# Clade v Classes

James Hunter

September 2, 2015

This uses the "int resist miami.xlsx" dataset. This data is summarized without formulas and without specific references to how it was summarized onto this dataset. The spreadsheet was prepared in 2009.

Read the data into R

```r
suppressPackageStartupMessages(library(RColorBrewer))
cladeclass <- t(read.csv("clade_class_excel.csv", header = FALSE))
```

```
## Warning in read.table(file = file, header = header, sep = sep,
## quote = quote, : incomplete final line found by readTableHeader on
## 'clade_class_excel.csv'
```

```r
# transform to put clades as rows
# Remove NT, Total rows
cladeclass <- cladeclass[-(4:5),]
cladeclass <- 100 * cladeclass # Turn proportions into percentages
colnames(cladeclass) <- c("Zero Class", "One Class", "Two Class", "Three Class")
barplot(cladeclass, beside = TRUE,
                col = brewer.pal(3, "Set1"),
                ylim = c(0, 60),
                xlab = "Number of Classes with Resistance",
                ylab = "% of Cases within Clade Showing Resistance",
                main = "Clades Showing Classes of Resistance",
                legend = c("Clade B", "Clade C", "Clade F"))
```

Clades Showing Classes of Resistance


```r
rownames(cladeclass) <- c("Clade B", "Clade C", "Clade F")
cladeclass
```

```
##          Zero Class One Class Two Class Three Class
## Clade B        7.45     11.33     54.91        26.31
## Clade C       12.74     17.22     51.18        18.87
## Clade F        9.11     13.98     54.43        22.47
```

This is indeed the original graph.

Clades by retroviral classes comparison

With Ester Sabino's spreadsheet, test differences among 3 class and 2 class by clade

```r
clades = factor(c("Clade B", "Clade C", "Clade F"))
# 3 class resistance-- read in 3 class file
cl3bcf <- read.csv("cl3byclade.csv", header = FALSE, stringsAsFactors = FALSE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'cl3byclade.csv'
```

```r
years <- c(as.character(2003:2009), "total")
clades <- c("B", "C", "F")
colnames(cl3bcf) <- years
rownames(cl3bcf) <- clades
cl3bcf
```

```
##      2003   2004   2005   2006   2007   2008   2009  total
## B 0.2684 0.2634 0.2652 0.2711 0.2451 0.2864 0.2543 0.2648
## C 0.2222 0.1186 0.1642 0.2176 0.2500 0.1821 0.1278 0.1718
## F 0.1833 0.1840 0.1982 0.2710 0.2609 0.2551 0.1950 0.2237
```

```r
cladeb3 <- cl3bcf[1,1:7]
cladec3 <- cl3bcf[2,1:7]
t.test(cladeb3, cladec3)
```

```
##
##  Welch Two Sample t-test
##
## data:  cladeb3 and cladec3
## t = 4.2098, df = 6.8244, p-value = 0.00422
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03553837 0.12771877
## sample estimates:
## mean of x mean of y
## 0.2648429 0.1832143
```

```r
# 2 class resistance
cl2bcf <- read.csv("cl2byclade.csv", header = FALSE, stringsAsFactors = FALSE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'cl2byclade.csv'
```

```r
colnames(cl2bcf) <- years
rownames(cl2bcf) <- clades
cl2bcf
```

```
##        2003      2004      2005      2006      2007      2008      2009
## B 0.5243488 0.5714286 0.5554725 0.5487724 0.5328368 0.5166889 0.4896430
## C 0.4444444 0.6101695 0.4925373 0.5233161 0.4239130 0.5257732 0.5027778
## F 0.7000000 0.5600000 0.5622120 0.5140187 0.5217391 0.5102041 0.4905660
##       total
## B 0.5373688
## C 0.5107376
## F 0.5384615
```

```r
#Difference between B and F
cladeb2 <- cl2bcf[1,1:7]
cladef2 <- cl2bcf[3,1:7]
t.test(cladeb2, cladef2)
```

```
##
##  Welch Two Sample t-test
##
## data:  cladeb2 and cladef2
## t = -0.59684, df = 7.7415, p-value = 0.5677
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.08344929  0.04929247
## sample estimates:
## mean of x mean of y
## 0.5341702 0.5512486
```

```r
# 1 class resistance
cl1bcf <- read.csv("cl1byclade.csv", header = FALSE, stringsAsFactors = FALSE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'cl1byclade.csv'
```

```r
colnames(cl1bcf) <- years
rownames(cl1bcf) <- clades
cl1bcf
```

```
##       2003    2004    2005    2006    2007    2008    2009   total
## B 0.11099 0.11502 0.09974 0.10982 0.13067 0.10748 0.11327 0.11129
## C 0.22222 0.15254 0.20896 0.12953 0.21739 0.14433 0.16667 0.16060
## F 0.06667 0.12800 0.14286 0.14953 0.12174 0.13265 0.16981 0.13866
```

```r
cladec1 <- cl1bcf[2,1:7]
cladef1 <- cl1bcf[3,1:7]
t.test(cladec1, cladef1)
```

```
##
##  Welch Two Sample t-test
##
## data:  cladec1 and cladef1
## t = 2.5021, df = 11.673, p-value = 0.02829
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   0.005970721 0.088423565
## sample estimates:
## mean of x mean of y
## 0.1773771 0.1301800
```

```r
# 0 class resistance
cl0bcf <- read.csv("cl0byclade.csv", header = FALSE, stringsAsFactors = FALSE)
```

```
## Warning in read.table(file = file, header = header, sep = sep, quote =
## quote, : incomplete final line found by readTableHeader on 'cl0byclade.csv'
```

```r
colnames(cl0bcf) <- years
rownames(cl0bcf) <- clades
cl0bcf
```

```
##       2003    2004    2005    2006    2007    2008    2009   total
## B 0.09626 0.05021 0.07957 0.07034 0.09140 0.08945 0.14279 0.08659
## C 0.11111 0.11864 0.13433 0.12953 0.10870 0.14777 0.20278 0.15686
## F 0.05000 0.12800 0.09677 0.06542 0.09565 0.10204 0.14465 0.09919
```

```
#Difference between C and F
cladec0 <- cl0bcf[2,1:7]
cladef0 <- cl0bcf[3,1:7]
t.test(cladec0, cladef0)
```

```
##
##  Welch Two Sample t-test
##
## data:  cladec0 and cladef0
## t = 2.2146, df = 11.998, p-value = 0.0469
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.0006229072 0.0766142357
## sample estimates:
##  mean of x  mean of y
## 0.13612286 0.09750429
```