Antibiotic resistance genes thta occur in Biosynthetic gene clusters

## DATA

CARD database contains antibiotic resistance (AR) genes.  
MIBiG database is composed by Biosynthetic gene clusters.  
antiSMASHDB contains predicted BGCs.

## Methodology

### CARD gene families

Genes in CARD were classified into gene families using FastORTHO with default parameters.

### Homology searches between CARD and MIBiG

Homology search was conducted using blast (e-vaule ) in an all vs all comparison with query antibiotic resistance genes from The Comprehensive Antibiotic Resistance Database (CARD) against genes in BGCs from MIBiG database.

In the following sections we will refer to the following questions: How many BGCs contain an AR gene? ¿How many families of AR genes are present in BGCs? Which families are over represented? How many BGCs per AR gene/Family?

# Protein families in CARD

The following figure shows the most populated families in CARD database.

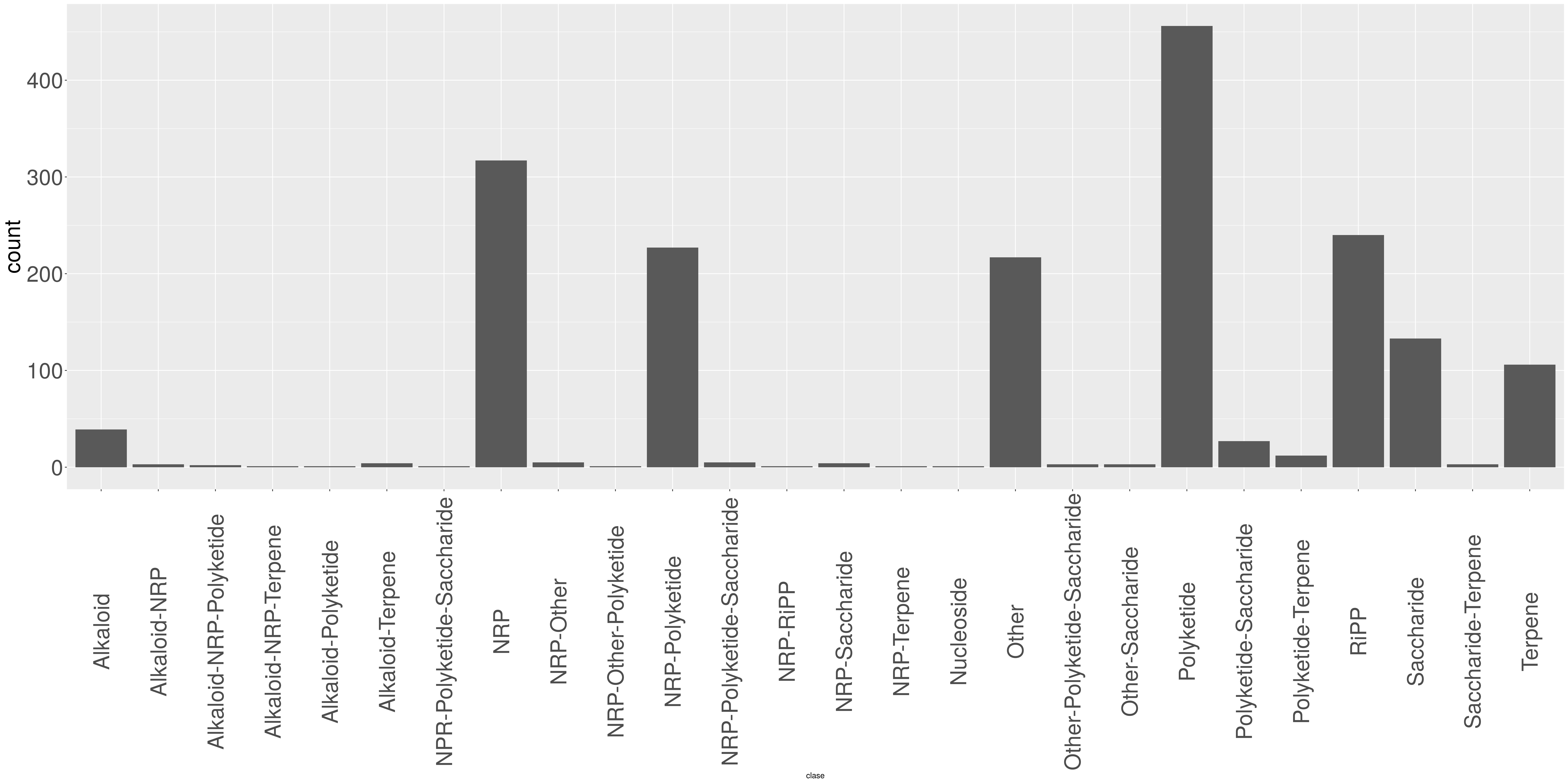
The total number of antibiotic gene resistance families is 76 counting singletons. Without singletons there are 0 families.

These first five families, in a preliminar search corresponds to:

|  |  |  |
| --- | --- | --- |
| Family | Number of genes | Annotation |
| Family 0 | 647 | -lactamase 2 |
| Family 1 | 302 | -lactamase |
| Family 2 | 283 | Transpeptidase |
| Family 3 | 103 | class B extended-spectrum -lactamase |
| Family 4 | 55 | MFS\_1 |

# BGC clases in MiBIG

The following figure shows the most populated classes in MiBIG database.



MIBiG classes

There are 1811 different BGCs reported at MIBiG distributed in several classes as indicated in the following table.

|  |  |
| --- | --- |
| Elements | MIBiG class |
| 39 | Alkaloid |
| 3 | Alkaloid-NRP |
| 2 | Alkaloid-NRP-Polyketide |
| 1 | Alkaloid-NRP-Terpene |
| 1 | Alkaloid-Polyketide |
| 4 | Alkaloid-Terpene |
| 1 | NPR-Polyketide-Saccharide |
| 317 | NRP |
| 5 | NRP-Other |
| 1 | NRP-Other-Polyketide |
| 227 | NRP-Polyketide |
| 5 | NRP-Polyketide-Saccharide |
| 1 | NRP-RiPP |
| 4 | NRP-Saccharide |
| 1 | NRP-Terpene |
| 1 | Nucleoside |
| 217 | Other |
| 3 | Other-Polyketide-Saccharide |
| 3 | Other-Saccharide |
| 456 | Polyketide |
| 27 | Polyketide-Saccharide |
| 12 | Polyketide-Terpene |
| 240 | RiPP |
| 133 | Saccharide |
| 3 | Saccharide-Terpene |
| 106 | Terpene |

# CARD families and their interaction with BGCs in MIBiG

MIBiG is classified in 24 classes, each one contains a variable number of BGCs

.  
.  
.

CARD is divided in 645 families each one with a variable number of genes

.  
.  
.

Average of blast hits of the family over the class where normalized by the size of these sets, and represent the number of elements in family and class respectively.

The following analysis finds the most represented CARD families by MIBiG classes. An average about how many hits normalized by the family size and the number of BGCs elements of the class.

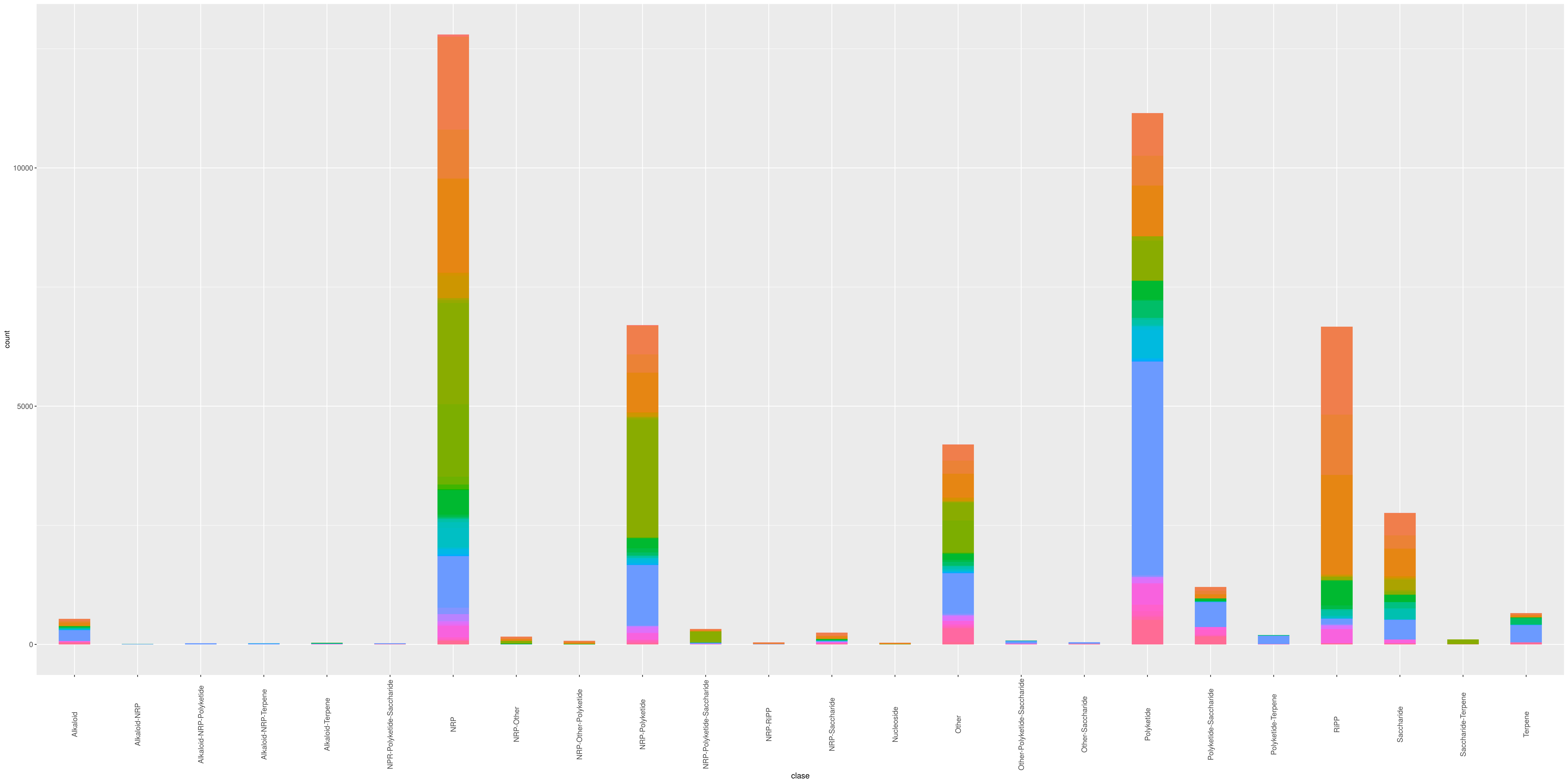
The following figure shows MIBiG Clases, and in each class stacks a CARD family average

CARD families in MIBiG classes  
The following figure shows the average of CARD families by MIBiG classes

CARD\_BIG\_family\_averageHorizontal<-ggplot(AverageFamily, aes(x = FO\_Family , y = average))+ geom\_col(aes(fill = clase), width = 0.7)+ theme(legend.position = "right", plot.title=element\_text(size=20,   
 face="bold",   
 family="American Typewriter",  
 color="tomato",  
 lineheight=1.2), # title  
 axis.title.y=element\_text(size=1), # Y axis title  
 axis.text.x=element\_text(size=1,   
 angle = 90,  
 vjust=.5), # X axis text  
 axis.text.y=element\_text(size=1)) # Y axis text  
#ggsave(file="CARD\_BIG\_family\_averageHorizontal.svg", plot=CARD\_BIG\_family\_averageHorizontal, width=30, height=15)  
ggsave(file="CARD\_BIG\_family\_averageHorizontal\_with\_leyend.svg", plot=CARD\_BIG\_family\_averageHorizontal, width=30, height=15)  
  
#bb\_AverageFamily <- match\_df(AverageFamily, "AAA50325.1|ARO:3003036|oleB", on="id")  
## Now I want to substet the families with the highets average #3   
# I look into inkscape to determine the 10 highest picks  
  
list<-c("AAA50325.1|ARO:3003036|oleB","AAV85982.1|ARO:3000535|macB","APB03219.1|ARO:3003986|TaeA","CAD70268.1|ARO:3002704|fexA","NP\_388442.1|ARO:3004476|vmlR","ORTHOMCL108 (3 genes,2 taxa):","ORTHOMCL150 (2 genes,1 taxa):","ORTHOMCL165 (2 genes,1 taxa):","ORTHOMCL300 (2 genes,1 taxa):","ORTHOMCL65 (4 genes,2 taxa):","ORTHOMCL72 (4 genes,2 taxa):")  
#APB03216.1|ARO:3003982|LlmA  
### Now We substet this list  
#AverageFamily$FO\_Family[which(  
# AverageFamily$FO\_Family == "AAA50325.1|ARO:3003036|oleB"|AverageFamily$FO\_Family == "AAV85981.1|ARO:3000533|macA" | )]  
  
bb\_AverageFamily <- AverageFamily[AverageFamily$FO\_Family %in% list, ]  
CARD\_BIG\_family\_averageHorizontalSubset<- ggplot(bb\_AverageFamily, aes(x = FO\_Family , y = average))+ geom\_col(aes(fill = clase), width = 0.7)+facet\_grid(bb\_AverageFamily$clase ~ .)+ theme(legend.position = "right", plot.title=element\_text(size=50,   
 face="bold",   
 family="American Typewriter",  
 color="tomato",  
 lineheight=1.2), # title  
 axis.title.y=element\_text(size=20), # Y axis title  
 axis.text.x=element\_text(size=20,   
 angle = 90,  
 vjust=.5), # X axis text  
 axis.text.y=element\_text(size=20)) # Y axis text  
ggsave(file="CARD\_BIG\_family\_averageHorizontal\_with\_leyend\_Subset.svg", plot=CARD\_BIG\_family\_averageHorizontalSubset, width=30, height=15)

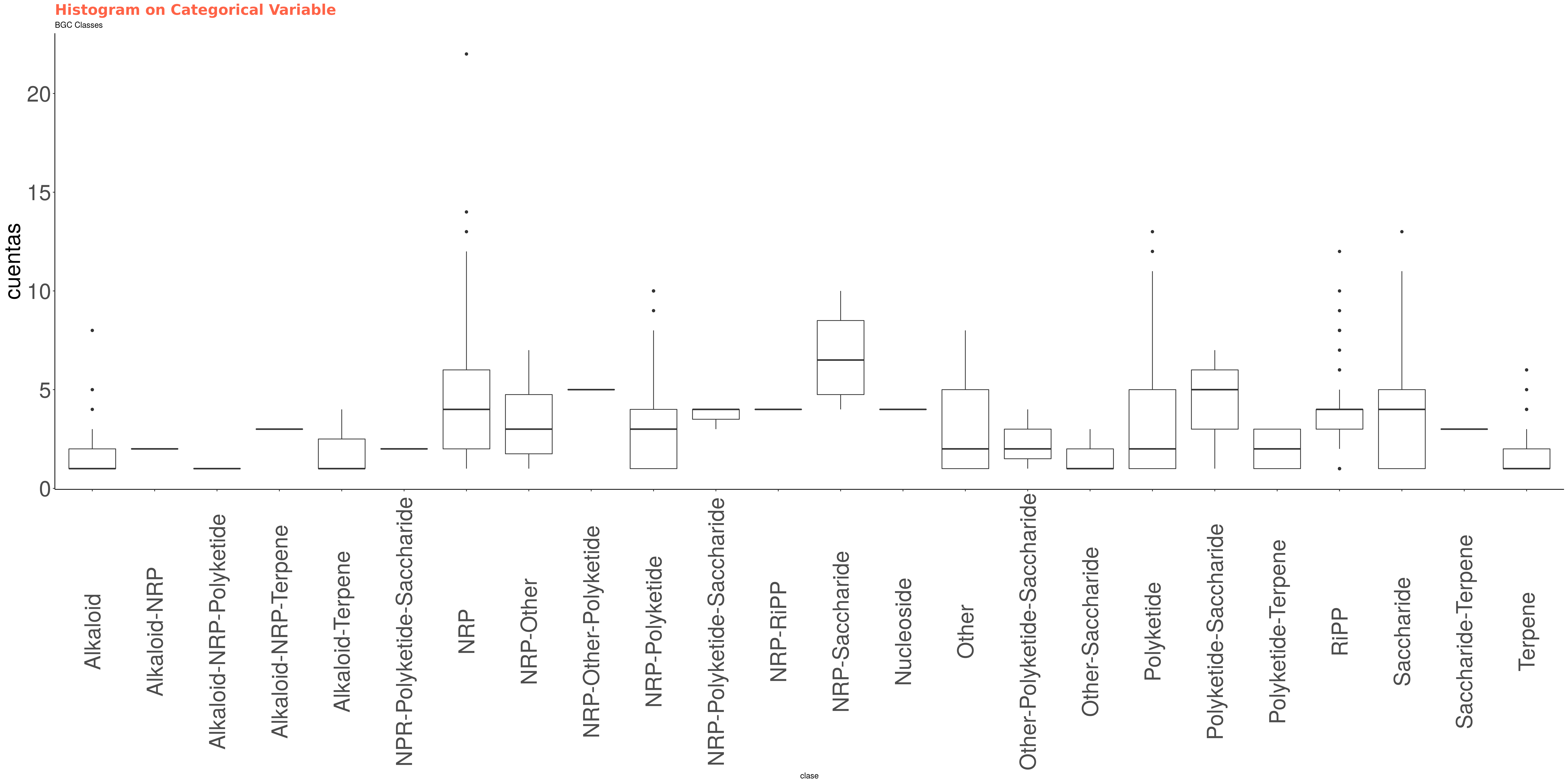
From 1811 different BGCs at MIBiG 1240 contains a CARD hit

The following figure shows which CARD families are the most populated classes in MiBIG database. and now an histogram by class with the number of BGCs with diferente frequecnies or CARd by color



CARD families in MIBiG classes

which MiBiG classes are over represented as CARD hits !–> ### OLD CODE

  
aaaaaqui voy