Antibiotics resistance genes present in Biosynthetic gene clusters

DATA

CARD database contains antibiotic resistance 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 BGC? 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.

```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_character()
## )
```

```
## # A tibble: 6 x 3
##   Gen_ID_CARD secondary_ID_CARD   FO_Family
##   <chr>       <chr>               <chr>
## 1 BAM62794.1  ARO:3002231|IMP-40  BAM62794.1|ARO:3002231|IMP-40
## 2 AIT76101.1  ARO:3002151|DHA-20  ORTHOMCL18 (17 genes,2 taxa):
## 3 AQK48217.1  ARO:3004513|MCR-1.6 ORTHOMCL17 (18 genes,1 taxa):
## 4 AIT76107.1  ARO:3002145|DHA-14  ORTHOMCL18 (17 genes,2 taxa):
## 5 AHA80101.1  ARO:3002112|CMY-100 ORTHOMCL2 (159 genes,2 taxa):
## 6 ADI46626.1  ARO:3002739|QnrB24  ORTHOMCL213 (2 genes,2 taxa):
```

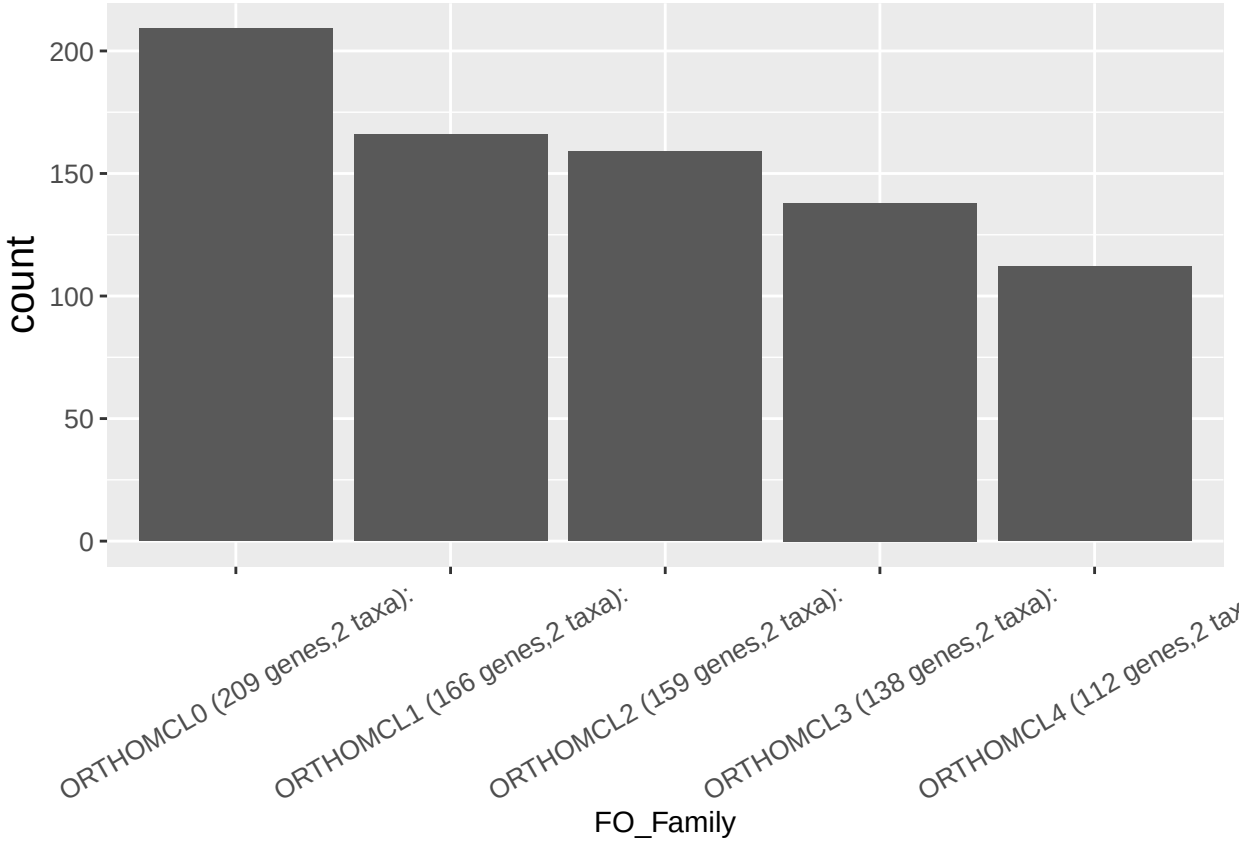
The total number of antibiotic gene resistance families is 652 counting singletons. Without singletons there are 311 families.

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

Family	Number of genes	Annotation
Family 0	209	class A broad spectrum β-lactamase
Family 1	166	class A extended-spectrum β-lactamase
Family 2	159	cephalosporin-hydrolyzing class C β -lactamase
Family 3	138	class A extended-spectrum β-lactamase
Family 4	112	class D β-lactamase

```
## # A tibble: 6 x 4
##   Gen_ID_CARD secondary_ID_CARD   FO_Family                          n
##   <chr>       <chr>               <chr>                          <int>
## 1 ABQ52429.1  ARO:3001636|OXA-104 ORTHOMCL4 (112 genes,2 taxa):   112
## 2 AGU69250.1  ARO:3001683|OXA-312 ORTHOMCL4 (112 genes,2 taxa):   112
## 3 AHN07454.1  ARO:3001526|OXA-338 ORTHOMCL4 (112 genes,2 taxa):   112
## 4 ADX07748.1  ARO:3001664|OXA-203 ORTHOMCL4 (112 genes,2 taxa):   112
## 5 AGU69251.1  ARO:3001684|OXA-313 ORTHOMCL4 (112 genes,2 taxa):   112
## 6 ADI58618.1  ARO:3001658|OXA-174 ORTHOMCL4 (112 genes,2 taxa):   112
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



BGC clases in MiBIG

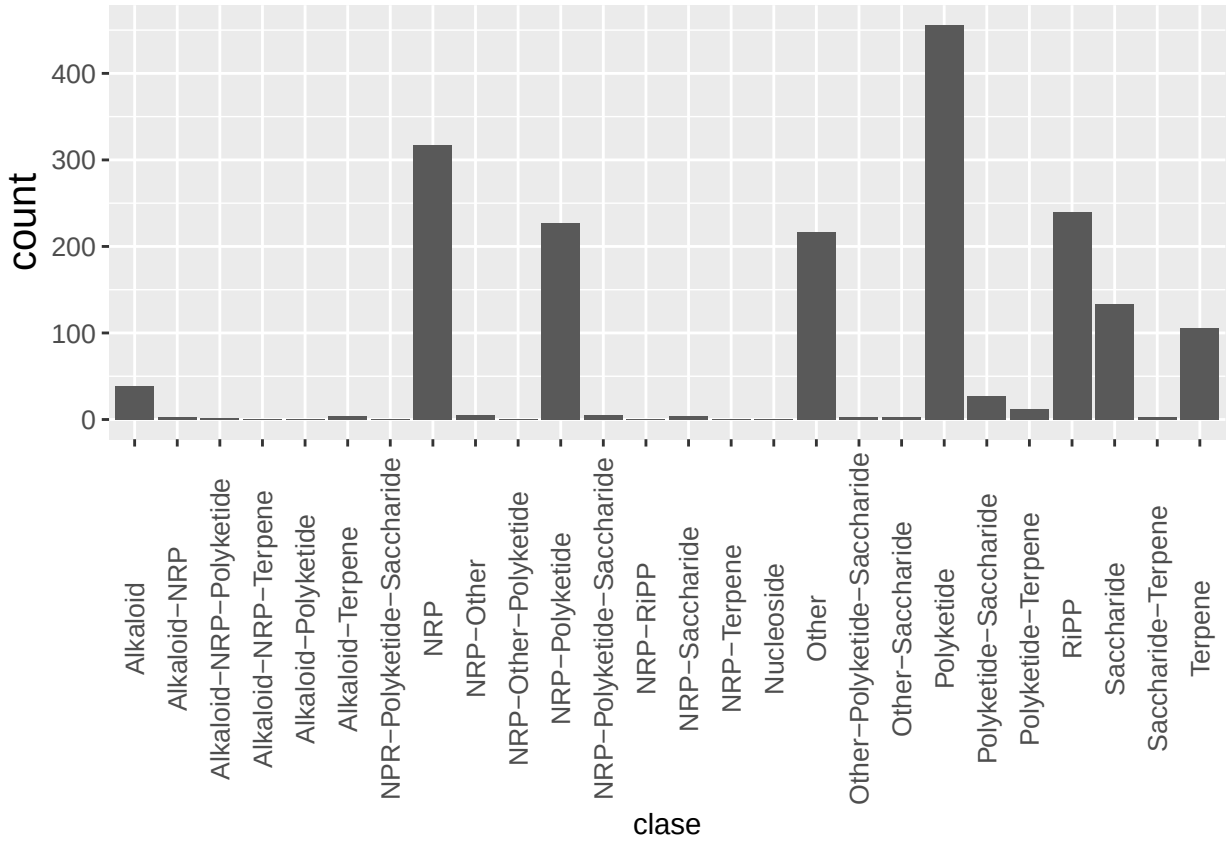
The following figure shows the most populated classes in MiBIG database.

```
## Parsed with column specification:
## cols(
```

```
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_character(),
##   X4 = col_character(),
##   X5 = col_character(),
##   X6 = col_character(),
##   X7 = col_character()
## )
```

```
## # A tibble: 6 x 7
##   BGC_ID   compound   clase    source          phylum   kingdom completness
##   <chr>    <chr>      <chr>    <chr>           <chr>    <chr>   <chr>
## 1 BGC0000~ caerulomy~ NRP-Pol~ Actinoalloteic~ Actinob~ Bacter~ Partial
## 2 BGC0000~ nocathiac~ RiPP     Actinokineospo~ Actinob~ Bacter~ Partial
## 3 BGC0000~ pradimicin Polyket~ Actinomadura h~ Actinob~ Bacter~ Partial
## 4 BGC0000~ kijanimic~ Polyket~ Actinomadura k~ Actinob~ Bacter~ Partial
## 5 BGC0001~ maduropep~ Polyket~ Actinomadura m~ Actinob~ Bacter~ Full
## 6 BGC0000~ AT2433     Alkaloid Actinomadura m~ Actinob~ Bacter~ Partial
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Elements	MIBiG class
39	Alkaloid
3	Alkaloid-NRP
2	Alkaloid-NRP-Polyketide
1	Alkaloid-NRP-Terpene
1	Alkaloid-Polyketide
4	Alkaloid-Terpene

Elements	MIBiG class
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

Interaction CARD - MiBIG

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

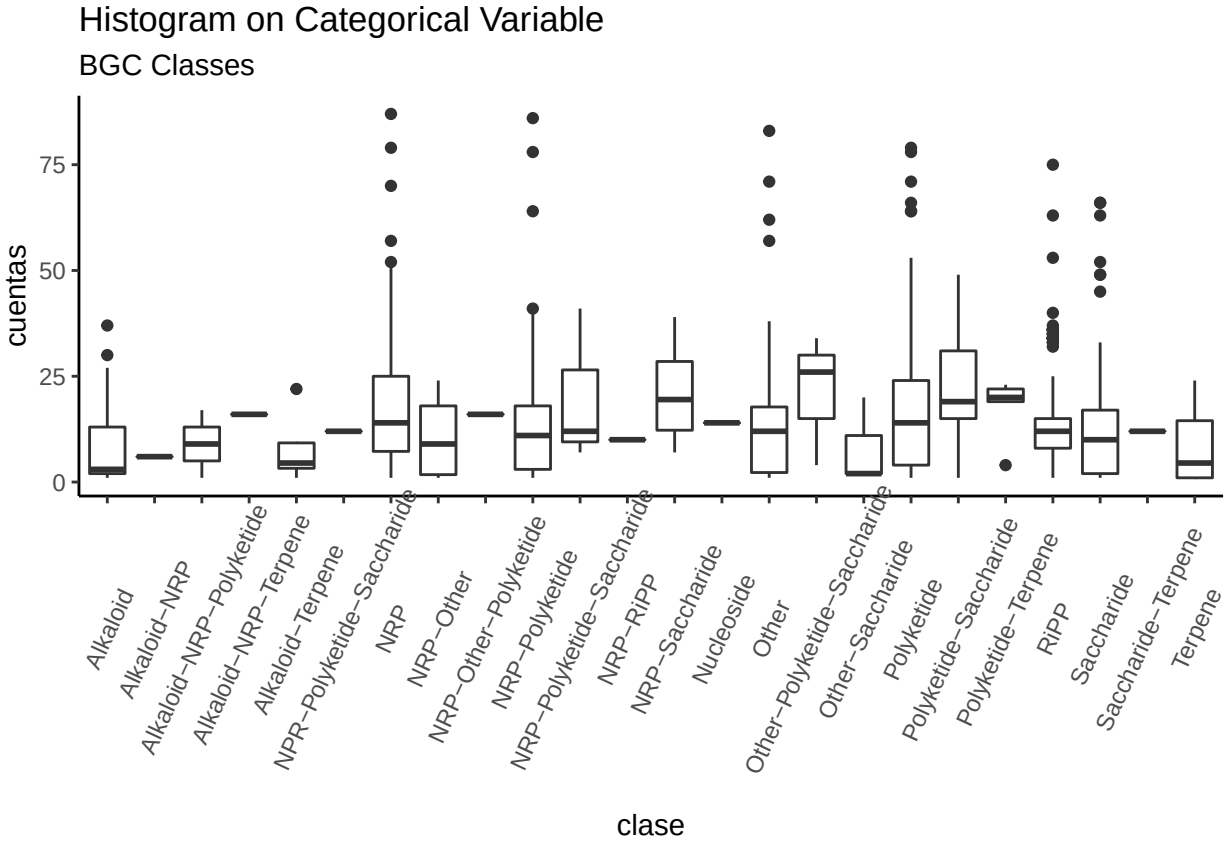
```
## Parsed with column specification:
## cols(
##   X1 = col_character(),
##   X2 = col_character(),
##   X3 = col_character(),
##   X4 = col_character(),
##   X5 = col_character()
## )
```

```
## # A tibble: 6 x 5
##   Gen_ID_CARD CARD_SOURCE BGC_ID    MIBIG_annotation           MIBIG_gen_Id
##   <chr>       <chr>       <chr>     <chr>                      <chr>
## 1 ACT97415.1  CblA-1      BGC00016~ penicillin_binding_transp~ AQX14496.1
## 2 ACT97415.1  CblA-1      BGC00017~ SulD                       AOZ21311.1
## 3 ACT97415.1  CblA-1      BGC00008~ <NA>                       BAA89386.1
## 4 AEJ08681.1  SHV-52      BGC00008~ <NA>                       BAA89386.1
## 5 AEJ08681.1  SHV-52      BGC00003~ beta-lactamase             EDY47102.1
## 6 AEJ08681.1  SHV-52      BGC00017~ SulD                       AOZ21311.1
```

```
## # A tibble: 6 x 2
## # Groups:   BGC_ID [6]
##   BGC_ID     cuentas
##   <chr>        <int>
## 1 BGC0000001      28
## 2 BGC0000002      14
## 3 BGC0000004      22
## 4 BGC0000005      22
```

```
## 5 BGC0000006      20
## 6 BGC0000007      21
```

```
## # A tibble: 100 x 4
## # Groups:   clase, BGC_ID [100]
##    clase    BGC_ID     cuentas     n
##    <chr>    <chr>        <int> <int>
##  1 Alkaloid BGC0000188      22    44
##  2 Alkaloid BGC0000809      30    55
##  3 Alkaloid BGC0000810       1     3
##  4 Alkaloid BGC0000813       5     8
##  5 Alkaloid BGC0000814      37   161
##  6 Alkaloid BGC0000822      16    31
##  7 Alkaloid BGC0000823       2     5
##  8 Alkaloid BGC0000824      24    41
##  9 Alkaloid BGC0000825       2     5
## 10 Alkaloid BGC0000826       2     5
## # ... with 90 more rows
```



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 which CAARD families are over represented in MiBiG which MiBiG classes are over represented as CARD hits