**For exacted smNRPS clusters:**

|  |  |  |  |
| --- | --- | --- | --- |
| bgcName | GroupAttribute | totalCounts | hitRatio |
| smNRPS-CB | C | 4/71 | 5.63% |
| smNRPS-CC | C | 4/71 | 5.63% |
| smNRPS-BP | CD | 2/21 | 9.52% |
| smNRPS-CB | CD | 18/21 | 85.7% |
| smNRPS-CC | CD | 17/21 | 81.0% |
| smNRPS-CB | UC | 15/127 | 11.8% |
| smNRPS-CC | UC | 8/127 | 6.30% |

**smNRPS-CB:** V1.UC61.0\_\_NODE\_930\_length\_26002\_cov\_3.77986\_\_26002\_\_nrps\_\_ANTISMASH\_\_0\_26002

**smNRPS-CC:** V1.CD18.3\_\_NODE\_78\_length\_114577\_cov\_12.9704\_\_43190\_\_nrps\_\_ANTISMASH\_\_49316\_92506

**For not-exacted smNRPS clusters:**

|  |  |  |  |
| --- | --- | --- | --- |
| bgcName | GroupAttribute | totalCounts | hitRatio |
| smNRPS-CB | C | 5/71 | 7.04 |
| smNRPS-CC | C | 3/71 | 4.23 |
| smNRPS-CB | CD | 13/21 | 61.9 |
| smNRPS-CC | CD | 17/21 | 81.0 |
| smNRPS-CB | UC | 12/127 | 9.45 |
| smNRPS-CC | UC | 5/127 | 3.94 |

---------------------------------------------------------------------------------------------------------------------

**For exacted with unique reads smNRPS clusters:**

|  |  |  |  |
| --- | --- | --- | --- |
| bgcName | GroupAttribute | totalCounts | hitRatio |
| smNRPS-CB | C | 3/71 | 4.23 |
| smNRPS-CC | C | 1/71 | 1.41 |
| smNRPS-CB | CD | 13/21 | 61.9 |
| smNRPS-CC | CD | 17/21 | 81.0 |
| smNRPS-CB | UC | 13/127 | 10.2 |
| smNRPS-CC | UC | 4/127 | 3.15 |

**Comparing enriched BGCs**

Enriched CD\_C vs Enriched UC\_C:

Unique CD BGCs = 41

Unique UC BGCs = 5

Common between CD & UC = 2

Enriched C\_CD vs Enriched C\_UC:

Unique C-CD BGCs = 64

Unique C-UC BGCs = 24

Common C BGCS = 13

Enriched CD\_C vs Enriched CD\_UC:

Unique CD-C BGCs = 13

Unique CD-UC BGCs = 16

Common CD BGCs = 30

Enriched UC\_C vs Enriched UC\_CD:

Unique UC-C BGCs = 7

Unique UC-CD BGCs = 60

Common UC BGCs = 0

*E.coli* derived BGCs enriched in CD\_C Spanish MetaHIT\_GUT (data from lefse\_CD\_C\_supp\_table\_1.csv):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BGC | log(HighestMean\_AllClasses) | class\_discriminitive | LDA\_score | LEfSe\_pvalue | manual\_pvalue | NCBI refrence sequence acession number | Taxa Name | Phylum |  |
| V1.UC61.0\_\_NODE\_228\_length\_65988\_cov\_4.67854\_\_26289\_\_nrps\_\_ANTISMASH\_\_39699\_65988 | 3.667408494 | CD | 3.345887623 | 2.24E-05 | 2.24E-05 | NZ\_KE136816 | Escherichia coli KTE38 | Proteobacteria | |
| V1.UC61.0\_\_NODE\_2273\_length\_12464\_cov\_2.76678\_\_12464\_\_t1pks\_\_ANTISMASH\_\_0\_12464 | 3.634168149 | CD | 3.30281596 | 6.55E-05 | 6.55E-05 | NZ\_KE702322 | Escherichia coli UMEA 3304-1 | Proteobacteria | |
| V1.CD41.0\_\_NODE\_43\_length\_139509\_cov\_97.9643\_\_53525\_\_nrps\_\_ANTISMASH\_\_80787\_134312 | 3.575467141 | CD | 3.249382541 | 0.000178234 | 0.000178234 | NZ\_AYJZ01000014 | Escherichia coli FAH1 | Proteobacteria | |
| V1.UC5.0\_\_NODE\_2999\_length\_8272\_cov\_2.30449\_\_8272\_\_bacteriocin\_\_ANTISMASH\_\_0\_8272 | 2.804796624 | CD | 2.668453158 | 0.001282594 | 0.001282594 | NZ\_KE698982 | Escherichia coli HVH 19 (4-7154984) | Proteobacteria | |
| V1.CD35.1\_\_NODE\_3185\_length\_6846\_cov\_4.30231\_\_6846\_\_siderophore\_\_ANTISMASH\_\_0\_6846 | 3.232613991 | CD | 2.858452545 | 0.003434952 | 0.003434952 | NZ\_LSUP01000001 | Escherichia coli | Proteobacteria | |
| V1.CD10.0\_\_NODE\_453\_length\_17115\_cov\_2.32239\_\_17115\_\_siderophore\_\_ANTISMASH\_\_0\_17115 | 3.103324049 | CD | 2.815832757 | 1.31E-07 | 1.31E-07 | NZ\_PSNQ01000016 | Escherichia coli | Proteobacteria | |
| V1.CD18.3\_\_NODE\_40\_length\_166683\_cov\_10.5209\_\_59165\_\_nrps-t1pks\_\_ANTISMASH\_\_61574\_120739 | 3.61749996 | CD | 3.296371957 | 8.70E-06 | 8.70E-06 | NZ\_NOHD01000013 | Escherichia coli | Proteobacteria | |

For enriched CD against HC in MetaHIT with taxonomy info— including class

