**Comparing enriched BGCs in HMP2-IBD to MetaHIT enriched BGCs**

Comparing the enriched **Controls** in LEfSe **(C vs CD)** BGCs in HMP2 and MetaHIT\_Gut cohorts (95%pident, 95%coverage):

1. O2.UC23.2\_\_NODE\_123\_length\_53602\_cov\_5.17215\_\_42532\_\_other\_\_ANTISMASH\_\_8083\_50615 (MetaHIT)

C2075\_\_NODE\_793\_length\_16630\_cov\_377.864\_\_16630\_\_other\_\_ANTISMASH\_\_0\_16630 (HMP2-IBD)

Comparing the enriched **Crohn’s** LEfSe **(C vs CD)** BGCs in HMP2 and MetaHIT\_Gut cohorts (95%pident, 95%coverage):

1. smNRPS-CC:

V1.CD18.3\_\_NODE\_78\_length\_114577\_cov\_12.9704\_\_43190\_\_nrps\_\_ANTISMASH\_\_49316\_92506 (MetaHIT)

C4004\_\_NODE\_591\_length\_10305\_cov\_3.58927\_\_10305\_\_nrps\_\_ANTISMASH\_\_0\_10305 (HMP2-IBD)

The following clusters mapped, but with ~72% coverage at 99%, these are our smNRPS-CB not exacted:

V1.UC61.0\_\_NODE\_930\_length\_26002\_cov\_3.77986\_\_26002\_\_nrps\_\_ANTISMASH\_\_0\_26002

C3008\_\_NODE\_58\_length\_80320\_cov\_9.89467\_\_38512\_\_nrps\_\_ANTISMASH\_\_41808\_80320

The following clusters mapped, but with ~58% coverage at 99%, these are our smNRPS-CB not exacted:

C3017\_\_NODE\_742\_length\_38583\_cov\_142.623\_\_25546\_\_nrps\_\_ANTISMASH\_\_13037\_38583

V1.UC61.0\_\_NODE\_930\_length\_26002\_cov\_3.77986\_\_26002\_\_nrps\_\_ANTISMASH\_\_0\_26002

There were no common enriched **Ulcerative colitis** LEfSe **(C vs UC)** BGCs in HMP2 and MetaHIT\_Gut cohorts (95%pident, 95%coverage).

Comparing the enriched **Controls** LEfSe **(C vs UC)** BGCs in HMP2 and MetaHIT\_Gut cohorts (95%pident, 95%coverage):

1. O2.UC50.0\_\_NODE\_43\_length\_109678\_cov\_2.66805\_\_26723\_\_terpene\_\_ANTISMASH\_\_82955\_109678 (MetaHIT)

C4006\_\_NODE\_15\_length\_222149\_cov\_19.3876\_\_20894\_\_terpene\_\_ANTISMASH\_\_77199\_98093 (HMP2-IBD)

The following clusters mapped, but with 52-67% coverage at >95%, these are not exacted. Wondered if they were exacted they may map to each other.

V1.UC62.0\_\_NODE\_31\_length\_180575\_cov\_19.4374\_\_62233\_\_nrps-ladderane\_\_ANTISMASH\_\_16961\_79194 C2060\_\_NODE\_21\_length\_104722\_cov\_19.3019\_\_47332\_\_nrps\_\_ANTISMASH\_\_8019\_55351 62233 47332 **52** 97.638 0.00e+00 42623 19718

V1.UC62.0\_\_NODE\_31\_length\_180575\_cov\_19.4374\_\_62233\_\_nrps-ladderane\_\_ANTISMASH\_\_16961\_79194 C2060\_\_NODE\_21\_length\_104722\_cov\_19.3019\_\_42199\_\_ladderane\_\_ANTISMASH\_\_53743\_95942 62233 42199 **67** 96.691

**Comparing smNRPS-CC and smNRPS-CB reads that mapped to these clusters in MetaHIT**

1. Unique reads for smNRPS

|  |  |  |  |
| --- | --- | --- | --- |
| bgcName | GroupAttribute | n | hitRatio (%) |
| smNRPS-CB | C | 3 | 4.23 |
| smNRPS-CC | C | 1 | 1.41 |
| smNRPS-CB | CD | 13 | 61.9 |
| smNRPS-CC | CD | 17 | 81.0 |
| smNRPS-CB | UC | 13 | 10.2 |
| smNRPS-CC | UC | 4 | 3.15 |
|  |  |  |  |

1. Aggregated smNRPS:

|  |  |  |  |
| --- | --- | --- | --- |
| bgcName | GroupAttribute | n | hitRatio (%) |
| smNRPS-1/2 | C | 7 | 9.86 |
| smNRPS-1/2 | CD | 18 | 85.7 |
| smNRPS-1/2 | UC | 16 | 12.6 |