Figure 2: **Differences in Biosynthetic Gene Clusters profiles between IBD and Healthy Control subjects in the Spanish cohort of MetaHIT**

1. A bar graph showing the distribution of BGC types detected in MetaHIT samples per disease status.
2. A bar graph showing the total number of BGCs discovered per Phylum-level of MetaHIT samples. See Supplemental Data File X for the full metadata of predicted BGCs.
3. Identified differentially abundant BGCs by LEfSe in Crohn’s Disease and Healthy Control individuals. Proteobacteria-derived and smNRPS1 & 2 BGCs are shown to be in the microbiome of Crohn’s Disease individuals at a higher degree than Healthy Controls individuals.
4. We evaluated the accuracy of disease classification using Random Forest in MetaHIT with ROC and Precision-Recall curves representing the results. Classification between Crohn’s Disease and Healthy controls performed well with ROC AUC= 0.95 and PR AUC = 0.85.
5. Analysis using the machine learning algorithm Random Forest. Top 20 BGCs that distinguish individuals with Crohn’s Disease from Healthy controls were identified. The units on the x axis indicate mean decrease gini impurity.

Figure 4: **Clostridia-derived** **smNRPS 1 & 2 clusters enriched in Crohn’s Disease individuals**

1. LEfSe analysis of BGC profiles of Crohn’s Disease and Healthy controls in iHMP samples. As displayed, the smNRPS 1 & 2 clusters are shown to be independently significantly enriched in Crohn’s Disease individuals in iHMP. Significance levels for LEfSe were p < 0.01 and Linear Discriminant Analysis (LDA) Score > 2.
2. Quantification of smNRPS clusters in MetaHIT and iHMP cohorts. See STAR Methods for a more detailed description of how smNRPS were quantified in samples.
3. Domain annotation for each smNRPS enriched in Crohn’s Disease individuals.

Supplement Fig. 1: **Comparative of IBD biomarkers in MetaHIT**

1. Heatmapof differentially enriched BGCs from MetaHIT (Crohn’s Disease vs. Healthy Control) LEfSe analysis (see Fig 2C), with binary distance metric and Ward’s clustering method.
2. BGCs identified as differentially abundant between Ulcerative colitis and Healthy Control individuals as analyzed by LEfSe.

Supplement Fig. 2: **Differences in Biosynthetic Gene Clusters profiles between IBD and Healthy Control subjects in the iHMP**

1. A bar graph showing the distribution of BGC types detected in iHMP samples per disease status.
2. A bar graph showing the total number of BGCs discovered per Phylum-level of iHMP samples. See Supplemental Data File X for the full metadata of predicted BGCs.