

# The Impact of *NOD2* Genetic Variants on the Gut Mycobiota in Crohn's Disease Patients in Remission and Individuals Without Gastrointestinal Inflammation

Andrew Nelson<sup>1,†</sup>, Christopher J. Stewart<sup>2,†</sup>, Nicholas A. Kennedy<sup>3,4</sup>,  
John K. Lodge<sup>1</sup>, Mark Tremelling<sup>5</sup>, UK IBD Genetics Consortium,  
Chris S. Probert<sup>6,7</sup>, Miles Parkes<sup>8</sup>, John C. Mansfield<sup>2,9</sup>,  
Darren L. Smith<sup>1</sup>, Georgina L. Hold<sup>10,11</sup>, Charlie W. Lees<sup>12,13</sup>,  
Simon H. Bridge<sup>1,2,‡,\*</sup>, Christopher A. Lamb<sup>2,9,‡,\*</sup>

1. Faculty of Health & Life Sciences, Northumbria University, Newcastle upon Tyne, UK
2. Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK
3. IBD Pharmacogenetics Group, University of Exeter, UK
4. Department of Gastroenterology, Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
5. Department of Gastroenterology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK
6. Institute of Translational Medicine, University of Liverpool, Liverpool, UK
7. Department of Gastroenterology, Liverpool University Hospitals NHS Foundation Trust, Liverpool, UK
8. Department of Gastroenterology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK
9. Department of Gastroenterology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne. UK
10. Gastrointestinal Research Group, University of Aberdeen, Aberdeen, UK

© The Author(s) 2020. Published by Oxford University Press on behalf of European Crohn's and Colitis Organisation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

11. Microbiome Research Centre, St George & Sutherland Clinical School, University of New South Wales, Sydney, NSW 2217, Australia
12. Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
13. Western General Hospital, Edinburgh, UK

† Joint first authors

‡ Joint senior authors

\* Authors for correspondence;

Dr Simon H. Bridge; [simon.bridge@northumbria.ac.uk](mailto:simon.bridge@northumbria.ac.uk)

Dr Christopher A. Lamb; [christopher.lamb@newcastle.ac.uk](mailto:christopher.lamb@newcastle.ac.uk)

#### ORCID IDs

Andrew Nelson: <https://orcid.org/0000-0001-6665-4929>

Christopher J Stewart: <https://orcid.org/0000-0002-6033-338X>

Nicholas A Kennedy: <https://orcid.org/0000-0003-4368-1961>

John K Lodge: <https://orcid.org/0000-0001-6014-4952>

Miles Parkes: <https://orcid.org/0000-0002-6467-0631>

John C Mansfield: <https://orcid.org/0000-0003-2490-7750>

Chris S Probert: <http://orcid.org/0000-0003-0477-6714>

Darren L Smith: <https://orcid.org/0000-0003-4925-467X>

Georgina L Hold: <https://orcid.org/0000-0001-7573-3397>

Charlie W Lees: <http://orcid.org/0000-0002-0732-8215>

Simon H Bridge: <https://orcid.org/0000-0002-6212-3255>

Christopher A Lamb: <https://orcid.org/0000-0002-7271-4956>

## ABSTRACT

### Background and Aims

Historical and emerging data implicate fungi in Crohn's disease (CD) pathogenesis. However, a causal link between mycobiota, dysregulated immunity and any impact of *NOD2* variants remains elusive. This study aims to evaluate associations between *NOD2* variants and faecal mycobiota in CD patients and non-CD subjects.

### Methods

Faecal samples were obtained from 34 CD patients (18 *NOD2* mutant, 16 *NOD2* wild-type) identified from the UK IBD Genetics Consortium. To avoid confounding influence of mucosal inflammation CD patients were in clinical remission and had a faecal calprotectin <250 µg/g. 47 non-CD subjects were included as comparator groups, including 22 matched household (4 *NOD2* mutant) and 25 non-household subjects with known *NOD2* genotype (14 *NOD2* mutant) identified by the NIHR BioResource Cambridge. Faecal mycobiota composition was determined using ITS1 sequencing, and compared to 16S rRNA gene sequences and volatile organic compounds.

### Results

CD was associated with higher numbers of observed fungal OTUs ( $P=0.033$ ). Principal coordinates analysis using Jaccard index ( $P=0.018$ ) and weighted Bray-Curtis dissimilarities ( $P=0.01$ ) showed *Candida* spp. clustered closer to CD patients whereas *Cryptococcus* spp. clustered closer to non-CD. In CD, we found higher relative abundance of Ascomycota ( $P=0.001$ ) and lower relative abundance Basidiomycota ( $P=0.019$ ) phyla. An inverse relationship was found between bacterial

and fungal Shannon diversity in *NOD2* wild-type that was independent of CD ( $r=0.349$ ;  $P=0.029$ ).

### **Conclusion**

This study confirms compositional changes in the gut mycobiota in CD and provides evidence that fungi may play a role in CD pathogenesis. No *NOD2* genotype-specific differences were observed in the faecal mycobiota.

Keywords: Crohn's Disease, *NOD2* genotype, gut mycobiota

Accepted Manuscript

## 1. INTRODUCTION

Crohn's disease (CD) is a chronic relapsing inflammatory disease of the gastrointestinal (GI) tract, which affects the quality of life of over 1.5 million individuals in North America and Europe.<sup>1,2</sup> While inflammation may occur at any point along the GI tract, CD commonly affects the terminal ileum and colon,<sup>3</sup> and can lead to significant tissue damage, often necessitating the introduction of biological therapy, and in many patients surgical intervention.<sup>4-6</sup> While the aetiology and pathogenesis of CD remain unresolved, the likely mechanisms involve complex interactions between predisposing genes and environmental exposures, which lead to aberrant immune responses against the gut microbiota, resulting in an imbalanced microbial community.<sup>7-10</sup>

CD patients have altered gut bacterial communities that include imbalances in Bacteroidetes, Firmicutes, and Proteobacteria phyla.<sup>11-13</sup> The gut microbiota of the healthy adult GI tract includes the fungal mycobiota<sup>14-16</sup> and a study by Sokol and colleagues, which examined the faecal mycobiota in adult CD patients reported fungal dysbiosis.<sup>17</sup> Furthermore, a number of studies evaluating the faecal mycobiota in CD report increased prevalence of *Candida* spp.,<sup>17-19</sup> although the link between increased abundance and disease pathogenesis remains obscure.

The strongest genetic association for CD susceptibility is in the gene encoding the nucleotide-binding oligomerisation domain-containing protein 2 (*NOD2*; also known as CARD15).<sup>20,21</sup> The 3 most common CD-associated *NOD2* mutations; R702W and G908R result in amino acid substitutions and L1007fsinsC results in a premature stop codon and dysfunctional *NOD2*. In Caucasians, up to 50% of CD patients carry

at least one *NOD2* mutation and individuals that carry two mutated *NOD2* alleles have a 20-40-fold increased risk of developing CD.<sup>20,21</sup>

*NOD2* is a cytosolic pattern recognition receptor that is highly expressed in dendritic cells and macrophages,<sup>22</sup> Paneth cells,<sup>23,24</sup> and variably expressed in intestinal epithelial cells.<sup>25</sup> *NOD2* is activated by muramyl dipeptide (MDP), a bacterial cell wall component<sup>26,27</sup> and upregulates expression of cytokines, chemokines, defensins, and triggers adaptive immune responses.<sup>28</sup> The 3 major CD-associated mutations are located within the ligand recognition domain of the *NOD2* protein and are therefore defective in their ability to sense MDP and trigger autophagy, which impairs gut bacterial handling.<sup>29,30</sup>

A number of studies have investigated the impact of *NOD2*-bacterial interactions in CD,<sup>31-35</sup> including a retrospective study by Frank *et al.*, which found *NOD2*-specific compositional shifts in the intestinal-associated bacterial community.<sup>34</sup> Recently and in contrast, Kennedy & colleagues evaluated the impact of *NOD2* variants on the faecal bacterial community in well-phenotyped CD patients compared to matched controls and found no *NOD2*-specific alterations in the bacterial communities.<sup>35</sup>

An intriguing study by Wagener and co-workers demonstrated that *NOD2* is also activated by the fungal wall component chitin and induces interleukin 10 secretion, so *NOD2*-dependent recognition of chitin-particles dampens inflammatory responses.<sup>36</sup> Currently, it is unknown what impact *NOD2* genetic variants may have on the gut mycobiota in healthy individuals or patients with CD.

There were three aims of this study. First, to investigate the faecal mycobiota of CD patients in remission and non-CD individuals to determine whether there are unique fungal signatures that distinguish between the two groups, and if *NOD2* genotype

impacts upon the composition of the faecal mycobiota. Second, to compare the relative abundance of fungal and bacterial communities to identify potential correlations that may be implicated in CD and/or *NOD2* genotype. Third, to determine whether there are specific correlations between volatile organic compounds (VOC) and bacteria or fungi detectable in the faecal contents.

## 2. MATERIALS & METHODS

### 2.1 Study participants

This study used clinical data, stool samples, 16S rRNA gene sequences (accessible from the European Nucleotide Archive, accession number PRJEB21593) and faecal volatile organic compounds (VOC) collected previously for a study to investigate the impact of *NOD2* genotype on faecal bacterial community profiles in CD and non-CD individuals.<sup>35</sup> *NOD2* mutant CD patients were selected if they carried 2 copies (homozygous or compound heterozygotes) of the CD-associated *NOD2* mutations (R702W [rs2066844], G908R [rs2066845] or L1000fs [rs2066847]). *NOD2* mutant CD patients were matched for age, gender and geographical location to a wild-type *NOD2* CD patient. To avoid the confounding influence of inflammation, CD patients were confirmed to be in clinical remission as defined by physician assessment and a faecal calprotectin <250 µg/g.<sup>6</sup> All non-CD subjects had a faecal calprotectin <100 µg/g.<sup>6</sup> Healthy controls were stratified by the same *NOD2* genotypes i.e., *NOD2* wild-type and *NOD2* mutant (compound heterozygotes, other homozygotes and single heterozygotes). Two non-CD comparator groups were recruited for this study.<sup>35</sup> Firstly an environmental control group, where possible a household member (usually an unrelated spouse) of CD participants was recruited for stool sample collection and saliva for *NOD2* genotyping. Secondly, a genetic control group of volunteers of

known *NOD2* genotype were recruited from the NIHR Cambridge BioResource for stool sample collection. All stool samples were frozen within 24 hours of collection. Ethical approval was obtained from the North of Scotland Research Ethics Committee (reference 12/NS/0050). All study participants provided written consent.

## 2.2 ITS1 sequencing

DNA was extracted from 0.2 g stool samples with the DNeasy PowerLyzer PowerSoil kit (QIAGEN, Manchester, UK) and processed according to the manufacturers protocol. A DNA extraction kit negative control was processed alongside each batch of 23 samples and sequenced. ITS1 sequencing was carried out by NU-OMICs (Northumbria University, Newcastle upon Tyne, UK). Briefly, PCR amplification of the internal transcribed spacer 1 (ITS1) region of the eukaryotic ribosomal cluster was amplified using primers ITS1F (CTTGGTCATTTAGAGGAAGTAA) and ITS2 (GCTGCGTTCTTCATCGATGC) using a previously described method<sup>14</sup> and adapted as below. PCR reactions were made up in a total volume of 20  $\mu$ L using 1x AccuPrime™ Pfx Reaction Mix, 0.5  $\mu$ M each primer, 500 ng bovine serum albumin, 2.5 U AccuPrime™ Pfx DNA Polymerase and 14.25  $\mu$ L of template DNA. PCR cycling conditions: initial denaturation at 95 °C for 2 min, 35 amplification cycles of 95 °C for 30 s, 52 °C for 30 s, 68 °C for 30 s, followed by a final extension step of 68 °C for 7 min. PCR products were cleaned and normalised using the SequalPrep™ Normalisation Plate Kit (Invitrogen, Paisley, UK) and pooled. Pools were quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen) and diluted to 2 nM. Pools were denatured using sodium hydroxide and diluted to 5 pM, and sequenced using the MiSeq v3 600-cycle (2 x 300 bp) reagent kit on the Illumina MiSeq platform.



### 2.3 Bioinformatic analysis

Sequences were processed using Mothur (v1.39.5),<sup>37</sup> paired end reads were merged using make.contigs with trimoverlap set to true. Any sequence with an ambiguous base was removed from the dataset. Reads were assigned taxonomy using the UNITE database (v8)<sup>38</sup> and only fungal reads were retained for further analysis. A rarefied observed taxonomic units (OTU) table was generated for downstream analyses of taxa relative abundance, alpha and beta diversity. The total number of raw ITS1 reads was 2,497,751 with a median number of reads per sample of 11,039. After rarefaction, the number of samples employed in subsequent analyses were reduced to 81, with each sample normalised to 1,082 ITS1 sequence reads. A total of 90.5% of ITS1 sequences were identified by comparison to the UNITE ITS sequence database. The raw sequence data are available from the European Nucleotide Archive under the study accession number PRJNA607176. Analyses of fungal communities were performed in R (v3.3.1). To evaluate alpha and beta diversity of fungal and bacterial communities, the number of observed OTUs, Shannon diversity, Jaccard indices and Bray-Curtis dissimilarity values were calculated using the “Vegan” package (v 2.5-6).<sup>39</sup>

### 2.4 Volatile organic compounds

The VOC data was generated from a previous study<sup>35</sup> using a previously described method.<sup>40</sup>

## 2.5 Statistical analysis

Statistical analyses were performed using Prism v.8 (GraphPad, San Diego, US), Minitab 19 (Minitab Ltd, Coventry, UK) or R (v3.3.1). The distribution of continuous variables (age, BMI and faecal calprotectin) were assessed by Anderson-Darling normality tests. Proportions of categorical variables were compared using Fisher's exact test. Significance of nonparametric variables was determined using the Mann-Whitney test for two category comparisons or the Kruskal-Wallis test when comparing two or more categories or the Wilcoxon signed rank test for matched pair comparisons. Significance of parametric variables was determined using the two-sample t test for two category comparisons or paired t tests for matched pair comparisons. Correlation between continuous variables were assessed using nonparametric Spearman rank correlation tests or parametric variables were assessed using Pearson's correlation test. Basidiomycota/Ascomycota ratios were calculated for each sample by dividing the percentage relative abundance of Basidiomycota with the percentage relative abundance of Ascomycota. Principle Coordinate analysis (PCoA) using unweighted Jaccard or weighted Bray-Curtis distance was performed to investigate relationships in  $\beta$ -diversity and between variable distances were analysed by PERMANOVA. Where applicable P values were corrected using the false discovery rate algorithm (FDR).<sup>41</sup> To determine the associations between the relative abundance of either fungal genera or bacterial genera and the intensity of stool VOC, we performed sparse partial least squares regression analysis in canonical mode, using MixOmics.<sup>42</sup>

### 3. RESULTS

#### 3.1 Characteristics of the study population

The initial cohort comprised 113 study participants of Caucasian ethnicity (Supplementary **Table 1**), 32 samples were excluded from the study and the exclusion criteria are summarised in **Figure 1A**. Eighty-one participants remained and the characteristics of this cohort are summarised in **Table 1A**. Three household control subjects could not be *NOD2* genotyped leaving 78 participants for genotype-stratified analyses (**Figure 1B**). There were 34 CD patients (53% *NOD2* mutant) in remission and 47 non-CD individuals (38% *NOD2* mutant) without GI disease (22 household matched volunteers and 25 NIHR Cambridge BioResource volunteers). CD patients had lower BMIs and higher faecal calprotectin concentrations. Fifteen percent of CD patients were current smokers. A smaller subgroup of 15 CD patients (60% *NOD2* mutant) were matched with their household control (13% *NOD2* mutant), the characteristics of these groups are summarised in **Table 1B**. CD patients had lower BMIs and higher concentrations of faecal calprotectin.

#### 3.2 Fungal diversity of patients with CD and non-CD individuals

Using the UNITE ITS database we identified 523 fungal OTUs, which were detectable in the stool samples of the cohort and this identified 184 genera.

Comparing the within sample diversity (alpha-diversity) between CD patients and non-CD subjects, we found a significant difference in observed OTUs (**Figure 2A**;  $P = 0.033$ ) but there was no difference in Shannon diversity (**Figure 2A**;  $P = 0.19$ ). We examined beta-diversity which considers between sample variation of the fungal community using Jaccard indices (considers presence or absence) and Bray-Curtis

dissimilarity (considers relative abundance). A significant difference was observed in the PCoA using Jaccard indices (**Figure 2B**;  $R^2 = 0.0184$ ,  $P = 0.018$ ) and Bray-Curtis dissimilarities (**Figure 2B**;  $R^2 = 0.0246$ ,  $P = 0.01$ ; PERMANOVA). The genus *Candida* and *Cryptococcus* appeared most discriminatory where *Candida* was most associated with CD, and *Cryptococcus* was most associated with non-CD. The dominant phyla in the stool mycobiota were Ascomycota and Basidiomycota (Supplementary **Table S2**). Ascomycota were significantly more abundant in CD (**Figure 2C**; FDR-Adj  $P = 0.001$ ), whereas Basidiomycota were significantly less abundant in CD compared to non-CD (**Figure 2C**; FDR-Adj  $P = 0.019$ ). We sought to determine whether there were significant differences in the relative abundance of fungal genera and found no significant difference in genus abundance (**Figure 2D**). The Basidiomycota/Ascomycota abundance ratio was found to be significantly lower in CD patients (**Figure 2E**;  $P = 0.0051$ ). We investigated whether there were correlations in alpha-diversity between the bacterial and fungal communities, but no significant correlations in the number of observed OTUs (**Figure 2F**) or Shannon diversity (**Figure 2G**) by disease status were found.

### 3.3 CD-specific changes in the mycobiota compared to household matched individuals

We sought to compare the fungal mycobiota composition between CD and non-CD shared household contacts. Analysis of the alpha-diversity revealed CD patients had a significantly higher number of OTUs (**Figure 3A**;  $P = 0.0045$ , paired t test) and Shannon diversity (**Figure 3A**;  $P = 0.048$ , paired t test). PCoA using Jaccard and Bray-Curtis indices showed no clustering in the fungal community composition

(**Figure 3B**). There was no significant difference in the relative abundance of any fungal phyla or genera between CD and household contacts (**Figure 3C and 3D**). The Basidiomycota/Ascomycota abundance ratio was also not significantly different between groups (**Figure 3E**;  $P = 0.083$ , Wilcoxon matched pair test). We investigated whether there were correlations in alpha-diversity metrics between bacterial and fungal communities in CD and household matched subjects, and found no significant correlations in either the number of observed OTUs (**Figure 3F**) or Shannon diversity (**Figure 3G**).

### 3.4 Faecal mycobiota-*NOD2* genotype association

To examine the impact of *NOD2* genotype on the stool mycobiota we compared the relative abundance of fungal phyla and genera between 42 *NOD2* wild-type vs. 36 *NOD2* mutant subjects (19 compound heterozygotes, 3 frameshift homozygotes, 10 other homozygotes and 4 single heterozygotes). These analyses were independent of Crohn's disease. Comparing alpha-diversity, we found no significant differences in observed OTUs (**Figure 4A**; FDR Adj.  $P = 0.43$ ) or Shannon diversity (**Figure 4A**; FDR Adj.  $P = 0.57$ ). The fungal profiles were comparable between *NOD2* wild-type and *NOD2* mutant subjects using Jaccard indices (**Figure 4B**;  $R^2 = 0.0131$ ,  $P = 0.494$ ) and Bray-Curtis (**Figure 4B**;  $R^2 = 0.0141$ ,  $P = 0.321$ ). There were no significant differences in the relative abundance of Ascomycota (**Figure 4C**; FDR Adj.  $P = 0.28$ ) or Basidiomycota (**Figure 4C**; FDR Adj.  $P = 0.23$ ) or genera (**Figure 4D**). The Basidiomycota/Ascomycota abundance ratio was not found to be significantly different (**Figure 4E**;  $P = 0.182$ , Mann-Whitney test). We investigated whether there were correlations in alpha-diversity between bacterial and fungal communities and found no significant correlations in either observed OTUs (**Figure**

**4F**) in either wild-type *NOD2* ( $r = -0.072$ ;  $P = 0.662$ ) or mutant *NOD2* ( $r = -0.060$ ;  $P = 0.729$ ). However, there was a significant correlation between the bacterial and fungal Shannon diversity (**Figure 4G**) in wild-type *NOD2* individuals ( $r = -0.349$ ;  $P = 0.029$ ) but this was not significant in *NOD2* mutant subjects ( $r = -0.238$ ;  $P = 0.162$ ).

### 3.5 The impact of *NOD2* mutations on the alpha-diversity of the mycobiota in Crohn's disease

We sought to determine whether *NOD2* was associated with the stool mycobiota in 34 CD patients (16 *NOD2* wild-type vs. 18 *NOD2* mutant) and 44 non-CD individuals (26 *NOD2* wild-type vs. 18 *NOD2* mutant). Three samples could not be conclusively genotyped so were excluded from this analysis (supplementary **Figure S1A**). There were no significant differences between CD patients with *NOD2* wild-type and *NOD2* mutant in either observed OTUs (FDR Adj.  $P = 0.685$ ) or Shannon diversity (FDR Adj.  $P = 0.85$ ). In non-CD individuals, we found no significant differences between *NOD2* wild-type and *NOD2* mutant subjects in either observed OTUs (FDR Adj.  $P = 0.435$ ) or Shannon diversity (FDR Adj.  $P = 0.63$ ). Four non-CD subjects carried a heterozygous *NOD2* mutation. Excluding these samples and repeating the analysis, we found no significant differences between 26 *NOD2* wild-type and 14 mutant subjects in either observed OTUs (FDR Adj.  $P = 0.42$ ) or Shannon diversity (FDR Adj.  $P = 0.80$ ).

Finally, we sought to determine whether *NOD2* would exert the strongest effect on the alpha-diversity of the stool mycobiota in CD patients and non-CD subjects matched for age, gender and geography. We compared 9 CD *NOD2* wild-type subjects vs. 9 CD *NOD2* mutant subjects and found no significant differences in observed OTUs (supplementary **Figure S1B**;  $P = 0.436$ , paired t test) or Shannon

diversity (supplementary **Figure S1B**;  $P > 0.999$ , Wilcoxon test). In non-CD individuals we compared 10 *NOD2* wild-type vs. 10 *NOD2* mutant subjects and found no significant differences in observed OTUs (supplementary **Figure S1C**;  $P = 0.677$ , paired t test) or Shannon diversity (supplementary **Figure S1C**;  $P = 0.544$ , paired t test).

### 3.6 Fungal and bacterial interactions

To determine whether there were correlations in the relative abundance of fungal and bacterial genera, we performed sparse partial least squared (sPLS) regression. We identified pairwise positive correlations in the faecal mycobiota in samples between genera *Wallemia* and *Anaerostipes*, *Debaryomyces* and *Pseudobutyrvibrio*, *Cryptococcus* and *Ruminococcus*, *Saccharomyces* and *Escherichia/Shigella*, and *Sordariomyces* and *Clostridium-sensu-stricto-1* (**Figure 5A**). Previous work on these samples profiled the VOC in the faecal contents of the CD patients and non-CD subjects and found the concentration of pentanoic acid, 2-butanone, acetone and 2-hexanone, 5-methyl were significantly lower in CD and the concentration of 2-piperidinone and butanoic acid, 3-methyl-, ethyl ester were significantly higher in CD.<sup>35</sup> We sought to look at the relationship of these CD-discriminating VOCs with the relative abundance of fungal genera employing a canonical correlation analysis approach (**Figure 5B**) and identified a strong positive correlation between CD-associated *Candida* spp. and the CD-associated VOC 2-piperidinone, and *Debaryomyces* with the CD-associated VOC 2-hexanone, 5-methyl-.

We also sought to determine the relationship of the VOCs with the bacterial genera (**Figure 5C**) and identified strong inverse relationships between *Prevotella* with 2-piperidinone, the bacterial genera *Dorea*, *Blautia* and *Alistipes* showed a negative

correlation with acetone and 2-butanone. The genus *Faecalibacterium* was negatively associated with pentanoic acid and 2-hexanone, 5-methyl- and *Pseudobutyrvibrio* inversely correlated with the concentration of pentanoic acid.

#### 4. DISCUSSION

This study reports that CD patients in clinical remission have higher fungal diversity compared to a non-inflammatory, non-CD comparator group. We found significantly higher observed OTUs in CD and the fungal profiles of CD patients clustered distinctly from that of non-CD subjects, with *Candida* spp. and *Cryptococcus* spp. found to cluster with CD and non-CD subjects, respectively. We identified shifts in the fungal phyla composition in CD patients, notably lower Basidiomycota that include *Cryptococcus* spp. and higher Ascomycota that include *Candida* spp. However, following correction for multiple comparisons, there were no significant differences found in the relative abundance of fungal genera in CD patients and non-CD individuals. We also report significantly higher alpha-diversity in CD patients compared to their respective household matched control. A strength of this study was that it included well-characterised CD and non-CD participants of known *NOD2* genotype. We compared the faecal mycobiota of subjects stratified by *NOD2* genotype and in the presence or absence of CD and found no differences were evident.

We found CD-specific shifts in the abundance of Ascomycota and Basidiomycota. The Basidiomycota/Ascomycota ratio in non-CD subjects was approximately 1:2, and this increased to 1:8 in CD patients. Ascomycota dominated the fungal community in CD patients and this finding was independent of intestinal inflammation. Our findings confirm previously identified shifts in fungal phyla in the



stool mycobiota of CD patients compared to non-CD controls.<sup>17-19,43</sup> The study by Sokol and colleagues reported that during intestinal inflammatory flare the abundance of Basidiomycota increased markedly, whilst the abundance of Ascomycota decreased, and that the Basidiomycota/Ascomycota ratio increased with increasing inflammation.<sup>17</sup> In contrast, our study excluded any participants with evidence of disease activity and found that Ascomycota was the dominant phyla in CD patients and the Basidiomycota/Ascomycota ratio were decreased.

A number of faecal mycobiota studies in CD patients have found an association with *Candida albicans*, although the functional importance of this association remains unknown.<sup>17,19,43</sup> Our data found that *C. albicans* was the most abundant species in CD patients and also in individuals with CD-associated *NOD2* mutations. A study by Hoarau and colleagues found that *Candida Tropicalis* was significantly higher in CD patients and correlated positively with serum ASCA.<sup>18</sup> In our study, *C. tropicalis* was the 8<sup>th</sup> most abundant species identified in CD patients, however, there were no significant differences found between CD and non-CD subjects. It was also interesting to note that one CD patient was detectable for nearly 80% of all *C. Tropicalis* ITS1 reads and this patient had CD-associated *NOD2* mutations.

Recently Limon and co-authors found that *Malassezia restricta* was more abundant in CD patients with the single-nucleotide polymorphism CARD9<sup>S12N</sup>, which normally expresses a key signalling adaptor that plays an important role in anti-fungal immunity.<sup>44</sup> Curiously, we did not detect *M. restricta* in our cohort. This discrepancy is potentially due to sampling differences. We used stool samples, whereas Limon and colleagues used intestinal wash samples, which have different microbial communities than stool.<sup>45</sup> The observation may also relate to demographic

differences of the respective cohorts; all our study participants were recruited from 4 locations in the UK, whereas the subjects in the Limon study were recruited from California, US.

Our study found that fungal diversity is higher in CD and this was independent of inflammation. This contrasts with the well-established lower bacterial diversity seen in CD,<sup>11-13,35</sup> and may indicate fungi fill this niche in CD. The inclusion of matched VOC data from a previous study<sup>35</sup> enabled us to look at associations between metabolites and relative bacterial and fungal abundance. One of the primary routes with which the gut microbiota may interact and cross-communicate with each other, and the host is through the production of volatile metabolites. Previous studies report that the faecal concentration of the ketone 2-piperidinone is higher in CD during active disease and remission.<sup>35,46</sup> The present study sought to evaluate correlations between fungal and bacterial communities with faecal VOC, and identified a positive correlation between *Candida* spp. and 2-piperidinone, and a negative correlation between 2-piperidinone and *Prevotella* spp., which is intriguing. Future studies to assess if this compound has anti-bacterial properties that confer a competitive advantage to yeast and enables *Candida* spp. to outcompete commensal bacteria such as *Prevotella* would be helpful to interrogate the functional relationship of these microbes and VOC.

A study by Wagener and colleagues found that there were chitin-dependent pathways of *NOD2* activation that lead to interleukin-10 secretion and promote intestinal homeostasis<sup>36</sup> leading to the hypothesis that the mycobiota of CD patients with *NOD2* mutations would differ to wild-type individuals. Our findings suggest that there is no difference in the faecal mycobiota between *NOD2* wild-type or mutant

individuals with or without CD. However, CD patients with *NOD2* mutations often have ileal involvement and a more aggressive fistulising and fibrostenotic disease phenotype, and increased risk of postsurgical disease recurrence post-ileal resection.<sup>47-49</sup> Evidence from animal models<sup>50</sup> and human studies<sup>45</sup> show there is a modest correlation between the microbial contents of the stool and ileal wash samples, which suggests future studies are warranted that should include paired samples to determine the impact of *NOD2* genotype on the ileal-associated microbiome.

The present study has a number of strengths. The study included a cohort of well-characterised CD patients and non-CD subjects with known *NOD2* genotypes. Importantly, the study excluded participants with raised faecal calprotectin levels, thereby reducing the confounding influence of inflammation, which is a key stressor to the gut mycobiota.<sup>17</sup> The study design enabled comparisons between CD vs. non-CD controls, CD *NOD2* wild type vs. CD *NOD2* mutant, non-CD *NOD2* wild type vs. non-CD *NOD2* mutant, and CD vs. matched household. The latter grouping was important as unaffected individuals sharing the same household environment with a CD patient may also show signs of bacterial dysbiosis.<sup>51</sup> Our findings show that the fungal community of CD patients are richer and more diverse than the mycobiota of their respective household controls.

Some study limitations should be noted. Reflective of the complexity of undertaking a genotype-stratified study, despite recruitment from the UK IBD Genetics cohort via several IBD centres, the absolute number of participants and the number of participants in each group (CD patients, non-CD subjects, matched household subjects, wild-type *NOD2* and mutant *NOD2*), and samples in this multi-omic

analysis is relatively small. Furthermore, this study evaluated the impact of the most common *NOD2* genetic variants on the mycobiota and stratified participants into two groups (wild-type *NOD2* and mutant *NOD2*) as we recognised it was underpowered to examine the effects of individual genotypes including the effect of rarer *NOD2* variants.<sup>52</sup> To remove the confounding influence of inflammation, our study recruited CD patients in remission. However, it is possible that the greatest impact of *NOD2* genotype on the gut mycobiota will manifest during active disease. We also did not undertake functional analysis regarding *NOD2* activity in the respective CD-associated *NOD2* mutations, and thus did not determine whether there was aberrant fungal sensing or handling in *NOD2* mutant subjects compared to *NOD2* wild-type. Finally, this study characterised fungal taxonomy by employing an ITS1 amplicon DNA sequencing approach and recognise that there may be variation in the taxonomic resolution of the fungal communities by targeting alternative genetic markers such as ITS2, 18S and 28S rRNA.<sup>53</sup>

In conclusion, this study shows that fungal diversity is higher and bacterial diversity is lower in CD. We show CD-specific changes in fungal diversity, which importantly is independent of inflammatory flare. Using stool samples, we could not identify specific associations of *NOD2* genotype with the mycobiota in CD patients or non-CD subjects. *Candida* spp. clustered more closely with CD and highlights the need for further studies to characterise whether *Candida* is a gut commensal or a pathogen in CD patients. Future investigations should also explore the impact of *NOD2* genotype on the ileal-associated mycobiota in larger cohorts stratified by *NOD2* genotype to resolve these complex mucosal interactions.

## **Funding**

This work was supported by CORE, the Digestive Diseases Foundation (now Guts UK), and the Wellcome Trust (grant number 097943 to NAK and 093885 to CAL) for stool collection, bacterial sequencing and VOC profiling. Further financial support for fungal sequencing was provided from Northumbria University and NU-OMICS.

## **Conflict of interest**

All authors declare no conflict of interest.

## **Author Contributions**

SHB, CAL, NAK, JKL, MP, JCM, DLS, GLH & CWL conceived the study. CAL, NAK, MT, MP, JCM & CWL recruited patients. NAK, GLH & AN participated in sample preparation. NAK & CSP measured VOC profiles. AN completed fungal DNA extractions, ITS1 sequencing and bioinformatics. AN, CJS, CAL & SHB analysed data. SHB & CAL drafted and wrote the manuscript. All authors reviewed, revised and approved of the final version of the manuscript.

## **Acknowledgements**

We are grateful to the patients who provided samples for this study. We are grateful for support from the NIHR Newcastle Biomedical Research Centre and the NIHR Biomedical Research Centre award to Addenbrooke's Hospital/University of Cambridge School of Clinical Medicine.

## REFERENCES

1. Jones GR, Lyons M, Plevris N, *et al.* IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut* 2019;**68**:1953-60.
2. Ng SC, Shi HY, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. *Lancet* 2018;**390**:2769-78.
3. Cleynen I, Boucher G, Jostins L, *et al.* Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: A genetic association study. *Lancet* 2016;**387**:156-67.
4. Peyrin-Biroulet L, Loftus EV, Jr., Colombel JF, Sandborn WJ. The natural history of adult Crohn's disease in population-based cohorts. *Am J Gastroenterol* 2010;**105**:289-97.
5. Cosnes J, Bourrier A, Nion-Larmurier I, *et al.* Factors affecting outcomes in Crohn's disease over 15 years. *Gut* 2012;**61**:1140-5.
6. Lamb CA, Kennedy NA, Raine T, *et al.* British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* 2019;**68**:s1-s106.
7. Jostins L, Ripke S, Weersma RK, *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;**491**:119-24.

8. Gevers D, Kugathasan S, Denson LA, *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;**15**:382-92.
9. Pascal V, Pozuelo M, Borruel N, *et al.* A microbial signature for Crohn's disease. *Gut* 2017;**66**:813-22.
10. Chu H, Khosravi A, Kusumawardhani IP, *et al.* Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science* 2016;**352**:1116-20.
11. Frank DN, St Amand AL, Feldman RA, *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* 2007;**104**:13780-5.
12. Sokol H, Pigneur B, Watterlot L, *et al.* Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* 2008;**105**:16731-6.
13. Darfeuille-Michaud A, Boudeau J, Bulois P, *et al.* High prevalence of adherent-invasive escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology* 2004;**127**:412-21.
14. Nash AK, Auchtung TA, Wong MC, *et al.* The gut mycobiome of the human microbiome project healthy cohort. *Microbiome* 2017;**5**:153.
15. Auchtung TA, Fofanova TY, Stewart CJ, *et al.* Investigating colonization of the healthy adult gastrointestinal tract by fungi. *mSphere* 2018;**3**.
16. Suhr MJ, Hallen-Adams HE. The human gut mycobiome: Pitfalls and potentials--a mycologist's perspective. *Mycologia* 2015;**107**:1057-73.
17. Sokol H, Leducq V, Aschard H, *et al.* Fungal microbiota dysbiosis in IBD. *Gut* 2017;**66**:1039-48.

18. Hoarau G, Mukherjee PK, Gower-Rousseau C, *et al.* Bacteriome and mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *mBio* 2016;**7**.
19. Lewis JD, Chen EZ, Baldassano RN, *et al.* Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. *Cell Host Microbe* 2015;**18**:489-500.
20. Hugot JP, Chamaillard M, Zouali H, *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599-603.
21. Ogura Y, Bonen DK, Inohara N, *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603-6.
22. Tada H, Aiba S, Shibata K, Ohteki T, Takada H. Synergistic effect of Nod1 and Nod2 agonists with toll-like receptor agonists on human dendritic cells to generate interleukin-12 and T helper type 1 cells. *Infect Immun* 2005;**73**:7967-76.
23. Lala S, Ogura Y, Osborne C, *et al.* Crohn's disease and the NOD2 gene: A role for paneth cells. *Gastroenterology* 2003;**125**:47-57.
24. Ogura Y, Lala S, Xin W, *et al.* Expression of NOD2 in paneth cells: A possible link to Crohn's ileitis. *Gut* 2003;**52**:1591-7.
25. Ramanan D, Tang MS, Bowcutt R, Loke P, Cadwell K. Bacterial sensor nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal *Bacteroides vulgatus*. *Immunity* 2014;**41**:311-24.



26. Girardin SE, Boneca IG, Viala J, *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* 2003;**278**:8869-72.
27. Maekawa S, Ohto U, Shibata T, Miyake K, Shimizu T. Crystal structure of NOD2 and its implications in human disease. *Nat Commun* 2016;**7**:11813.
28. Philpott DJ, Sorbara MT, Robertson SJ, Croitoru K, Girardin SE. NOD proteins: Regulators of inflammation in health and disease. *Nat Rev Immunol* 2014;**14**:9-23.
29. Cooney R, Baker J, Brain O, *et al.* NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nat Med* 2010;**16**:90-7.
30. Homer CR, Richmond AL, Rebert NA, Achkar JP, McDonald C. ATG16L1 and NOD2 interact in an autophagy-dependent antibacterial pathway implicated in crohn's disease pathogenesis. *Gastroenterology* 2010;**139**:1630-41, 41 e1-2.
31. Knights D, Silverberg MS, Weersma RK, *et al.* Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med* 2014;**6**:107.
32. Li E, Zhang Y, Tian X, *et al.* Influence of Crohn's disease related polymorphisms in innate immune function on ileal microbiome. *PLoS One* 2019;**14**:e0213108.
33. Imhann F, Vich Vila A, Bonder MJ, *et al.* Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut* 2018;**67**:108-19.

34. Frank DN, Robertson CE, Hamm CM, *et al.* Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 2011;**17**:179-84.
35. Kennedy NA, Lamb CA, Berry SH, *et al.* The impact of NOD2 variants on fecal microbiota in Crohn's disease and controls without gastrointestinal disease. *Inflamm Bowel Dis* 2018;**24**:583-92.
36. Wagener J, Malireddi RK, Lenardon MD, *et al.* Fungal chitin dampens inflammation through IL-10 induction mediated by NOD2 and TLR9 activation. *PLoS Pathog* 2014;**10**:e1004050.
37. Schloss PD, Westcott SL, Ryabin T, *et al.* Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537-41.
38. Nilsson RH, Larsson KH, Taylor AFS, *et al.* The unite database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res* 2019;**47**:D259-D64.
39. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens M. The vegan package. *Community ecology package* 2007;**10**:719.
40. Aggio RB, Mayor A, Reade S, Probert CS, Ruggiero K. Identifying and quantifying metabolites by scoring peaks of GC-MS data. *BMC Bioinformatics* 2014;**15**:374.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;**57**:289-300.
42. Le Cao KA, Gonzalez I, Dejean S. Integromics: An R package to unravel relationships between two omics datasets. *Bioinformatics* 2009;**25**:2855-6.

43. Chehoud C, Albenberg LG, Judge C, *et al.* Fungal signature in the gut microbiota of pediatric patients with inflammatory bowel disease. *Inflamm Bowel Dis* 2015;**21**:1948-56.
44. Limon JJ, Tang J, Li D, *et al.* Malassezia is associated with Crohn's disease and exacerbates colitis in mouse models. *Cell Host Microbe* 2019;**25**:377-88 e6.
45. Zmora N, Zilberman-Schapira G, Suez J, *et al.* Personalized gut mucosal colonization resistance to empiric probiotics is associated with unique host and microbiome features. *Cell* 2018;**174**:1388-405 e21.
46. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. *Aliment Pharmacol Ther* 2016;**43**:596-611.
47. Abreu MT, Taylor KD, Lin YC, *et al.* Mutations in NOD2 are associated with fibrostenosing disease in patients with Crohn's disease. *Gastroenterology* 2002;**123**:679-88.
48. Alvarez-Lobos M, Arostegui JI, Sans M, *et al.* Crohn's disease patients carrying nod2/CARD15 gene variants have an increased and early need for first surgery due to stricturing disease and higher rate of surgical recurrence. *Ann Surg* 2005;**242**:693-700.
49. Helio T, Halme L, Lappalainen M, *et al.* CARD15/NOD2 gene variants are associated with familiarly occurring and complicated forms of Crohn's disease. *Gut* 2003;**52**:558-62.

50. Yasuda K, Oh K, Ren B, *et al.* Biogeography of the intestinal mucosal and luminal microbiome in the rhesus macaque. *Cell Host Microbe* 2015;**17**:385-91.
51. Joossens M, Huys G, Cnockaert M, *et al.* Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 2011;**60**:631-7.
52. Rivas MA, Beaudoin M, Gardet A, *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet* 2011;**43**:1066-73.
53. Frau A, Kenny JG, Lenzi L, *et al.* DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies. *Sci Rep* 2019;**9**:9328.

Accepted Manuscript

**TABLE 1. Overview of the clinical characteristics of the study cohort**

<b>A. Crohn's disease patients and non-Crohn's disease individuals (n = 81)</b>			
<b>Characteristic</b>	<b>Crohn's disease (n = 34)</b>	<b>Non-Crohn's disease (n = 47)</b>	<b>P</b>
Gender (Fn,%)	21 (62%)	22 (47%)	0.259
Age (yrs), median (Q1-Q3)	53 (44 - 64)	57 (45 - 64)	0.77
BMI (kg/m <sup>2</sup> ), median (Q1-Q3)	23.4 (21.4 - 27.0)	25.7 (23.0 - 29.1)	<b>0.007</b>
Ethnicity			ND
White - British	32	45	
White - Irish	1	0	
White - Other	1	2	
Faecal calprotectin (µg/g), median (Q1-Q3)	50 (19 - 115)	19 (19 - 30)	<b>&lt;0.001</b>
Current smoker, n (%)	5 (15%)	0 (0%)	<b>0.011</b>
<i>NOD2</i> genotype; n (%)			ND
Wild-type	16 (47%)	26 (55%)	
Compound heterozygote	11 (32%)	8 (17%)	
Frameshift homozygote	3 (9%)	0 (0%)	
Other homozygote	4 (12%)	6 (13%)	
Heterozygote	0 (0%)	4 (9%)	
Not determined	0 (0%)	3 (6%)	
Antibiotics in last 12 months; n (%)	16 (47%)	12 (26%)	0.059
Medication; n (%)			ND
NSAID/aspirin	6 (18%)	9 (19%)	
5ASA	10 (29%)	0 (0%)	
Systemic steroids	1 (3%)	0 (0%)	
Immunomodulator	12 (35%)	0 (0%)	
Anti-TNF	2 (6%)	0 (0%)	
Bile acid sequestrants	9 (26%)	0 (0%)	
Proton pump inhibitors	9 (26%)	3 (6%)	
Probiotics; n (%)	3 (9%)	3 (6%)	0.692
<b>B. Crohn's disease subgroup with household matched individuals (n = 30)</b>			
<b>Characteristic</b>	<b>Crohn's disease (n = 15)</b>	<b>Household matched individuals (n = 15)</b>	
Gender (Fn,%)	9 (60%)	6 (40%)	0.466
Age (yrs), median (Q1-Q3)	44 (38 - 64)	45 (40 - 65)	0.793
BMI (kg/m <sup>2</sup> ), median (Q1-Q3)	21.9 (21.1 - 22.9)	26.2 (22.8 - 28.5)	<b>0.005</b>
Faecal calprotectin (µg/g), median (Q1-Q3)	50 (19 - 170)	19 (19 - 30)	<b>0.024</b>

Current smoker, n (%)	2 (13%)	0 (0%)	0.483
Antibiotics in last 12 months, n (%)	7 (47%)	5 (33%)	0.71
Probiotics, n (%)	2 (13%)	0 (0%)	0.483
<i>NOD2</i> genotype; n (%)			
Wild-type	6 (40%)	10 (67%)	ND
Compound heterozygote	3 (20%)	0 (0%)	
Frameshift homozygote	2 (13%)	0 (0%)	
Other homozygote	4 (27%)	0 (0%)	
Heterozygote	0 (0%)	2 (13%)	
Not determined	0 (0%)	3 (20%)	

Non-CD, Non-Crohn's disease; BMI, Body Mass Index; *NOD2*, Nucleotide-binding oligomerisation domain-containing protein 2; ND, Not Determined. 5-ASA, 5-amino salicylates; TNF, tumour necrosis factor.

Medications: 5-amino salicylates: mesalazine/sulfasalazine, Systemic steroids: prednisolone, Immunomodulators:

Azathioprine/methotrexate, Anti-TNF: Adalimumab, Proton-pump inhibitor:

Esomeprazole/Omeprazole/Lansoprazole/Rabiprazole, Bile acid sequestrants: Colestyramine/Colesevelam

All data variable were non-parametric and were summarised as the median value (Q1-Q3). Proportional categorical variables were compared using a 2 x 2 contingency table Fisher's exact test. The medians of CD and non-CD subjects were compared using the Kruskal-Wallis test.  $P < 0.05$  were considered significant.

Accepted Manuscript

## FIGURE LEGENDS

**FIGURE 1. Flow diagrams summarising study participants inclusion criteria and *NOD2* analyses.** A.) Data from 34 CD patients and 47 non-CD individuals were used in the primary analysis B.) Flow of *NOD2* analyses to evaluate the impact of *NOD2* mutations on the mycobiota independently of disease and to look at disease-*NOD2* specific alterations.

**FIGURE 2. Altered mycobiota diversity in CD patients during remission compared to non-CD individuals.** A.) Observed number of OTU and Shannon diversity of the mycobiota. The centre line denotes the median, the boxes cover the interquartile range (Q1-Q3), and the whiskers extend to the most extreme data point, which is no more than 1.5 times the length of the box away from the box. Points outside the whiskers represent outlier samples. (The ranges and descriptive statistics are the same for panel C and D) B.) Beta diversity metrics (Jaccard Index and Bray-Curtis dissimilarity C.) Percentage relative abundance of Ascomycota and Basidiomycota. D.) Percentage relative abundance of the top 10 most abundant fungal genera E.) Basidiomycota/Ascomycota ratio. The middle lines are the median ratio, the boxes cover the interquartile ranges (Q1-Q3). The whiskers show the range from the minimum ratio to the maximum ratio F.) Association between bacterial and fungal observed OTUs in CD and non-CD (Spearman's rank correlation test applied to both groups) and G.) Association between bacterial and fungal Shannon diversity metrics in CD (Pearson's correlation) and non-CD (Spearman's rank correlation).

Where relevant P values were adjusted for multiple comparisons using FDR and considered significant if  $P < 0.05$ .

**FIGURE 3. Altered mycobiota diversity in CD patients during remission compared to household matched controls.** A.) Alpha-diversity metrics (observed OTUs and Shannon). The centre line denotes the median, the boxes cover the interquartile range (Q1-Q3), and the whiskers extend to the most extreme data point, which is no more than 1.5 times the length of the box away from the box. Points outside the whiskers represent outlier samples. (The ranges and descriptive statistics are the same for panel C and D) B.) Beta-diversity metrics (Jaccard Index (presence/absence of fungal genera) and Bray-Curtis distance (relative abundance of fungal genera)) C.) Relative abundance of Ascomycota and Basidiomycota. D.) Percentage relative abundance of the top 10 most dominant fungal genera E.) Basidiomycota/Ascomycota ratio. The middle lines are the median ratio, the boxes cover the interquartile ranges (Q1-Q3). The whiskers show the range from the minimum ratio to the maximum ratio F.) Association between bacterial and fungal observed OTUs in CD and non-CD (Spearman's rank correlation test applied to both groups) and G.) Association between bacterial and fungal Shannon diversity metrics in CD (Pearson's correlation) and non-CD (Spearman's rank correlation). Where relevant P values were adjusted for multiple comparisons using FDR and considered significant if  $P < 0.05$ .



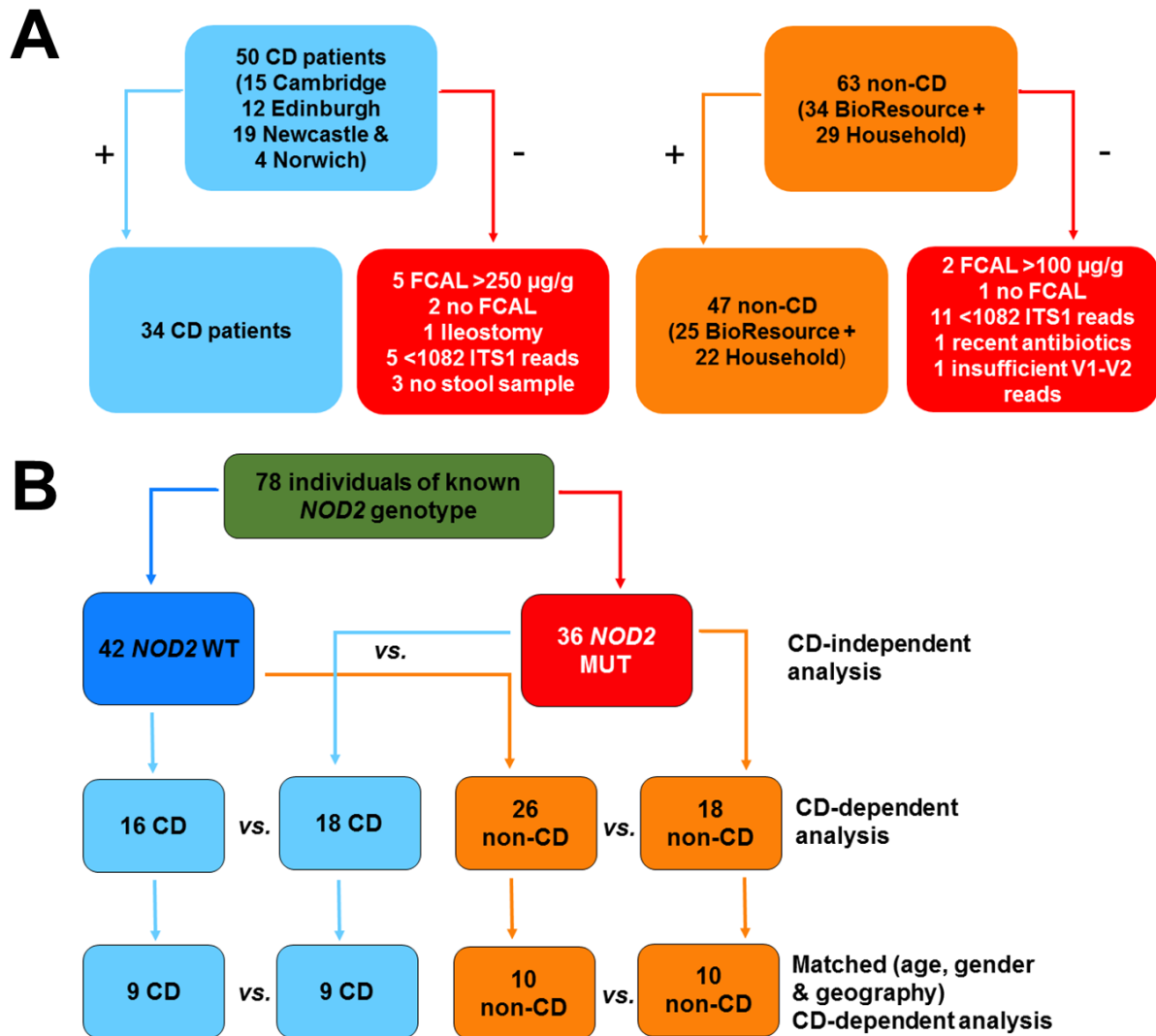
**FIGURE 4. Evaluating the impact of *NOD2* genotype independently of CD.** A.) Alpha-diversity metrics (observed OTUs and Shannon) B.) Beta-diversity metrics (Jaccard Index (presence/absence of fungal genera) and Bray-Curtis distance (relative abundance of fungal genera)) C.) Relative abundance of Ascomycota and Basidiomycota. The centre line denotes the median, the boxes cover the interquartile range (Q1-Q3), and the whiskers extend to the most extreme data point, which is no more than 1.5 times the length of the box away from the box. Points outside the whiskers represent outlier samples. (The ranges and descriptive statistics are the same for panel C and D) D.) Percentage relative abundance of the top 10 most prevalent fungal genera E.) Basidiomycota/Ascomycota ratio. The middle lines are the median ratio, the boxes cover the interquartile ranges (Q1-Q3). The whiskers show the range from the minimum ratio to the maximum ratio F.) Association between bacterial and fungal observed OTUs in *NOD2* wild-type and *NOD2* mutant subjects (Spearman's rank correlation test applied to both groups) and G.) Association between bacterial and fungal Shannon diversity metrics in *NOD2* wild-type subjects (Pearson's correlation) and *NOD2* mutant subjects (Spearman's rank correlation). Where relevant P values were adjusted for multiple comparisons using FDR and considered significant if  $P < 0.05$ .

**FIGURE 5. Canonical correlation analyses showing fungal and bacterial interactions and that certain volatile organic compounds (VOCs) associate strongly with the relative abundance of fungal and bacterial communities.**

These analyses are not stratified by disease or *NOD2* genotype and include n = 81 study participants. The red and blue shading indicate the strength of the positive and negative associations, and yellow shading indicates weak to no association. A.) Shows the association between the 5 most dominant fungal genera with 5 most dominant bacterial genera. B.) Shows the association between 8 fungal genera and the 6 most discriminative VOCs (CD vs. non-CD). C.) Shows the association between 10 bacterial genera and the 6 most discriminative VOCs (CD vs. non-CD).

Accepted Manuscript

Figure 1



ACCEPTED

Figure 2

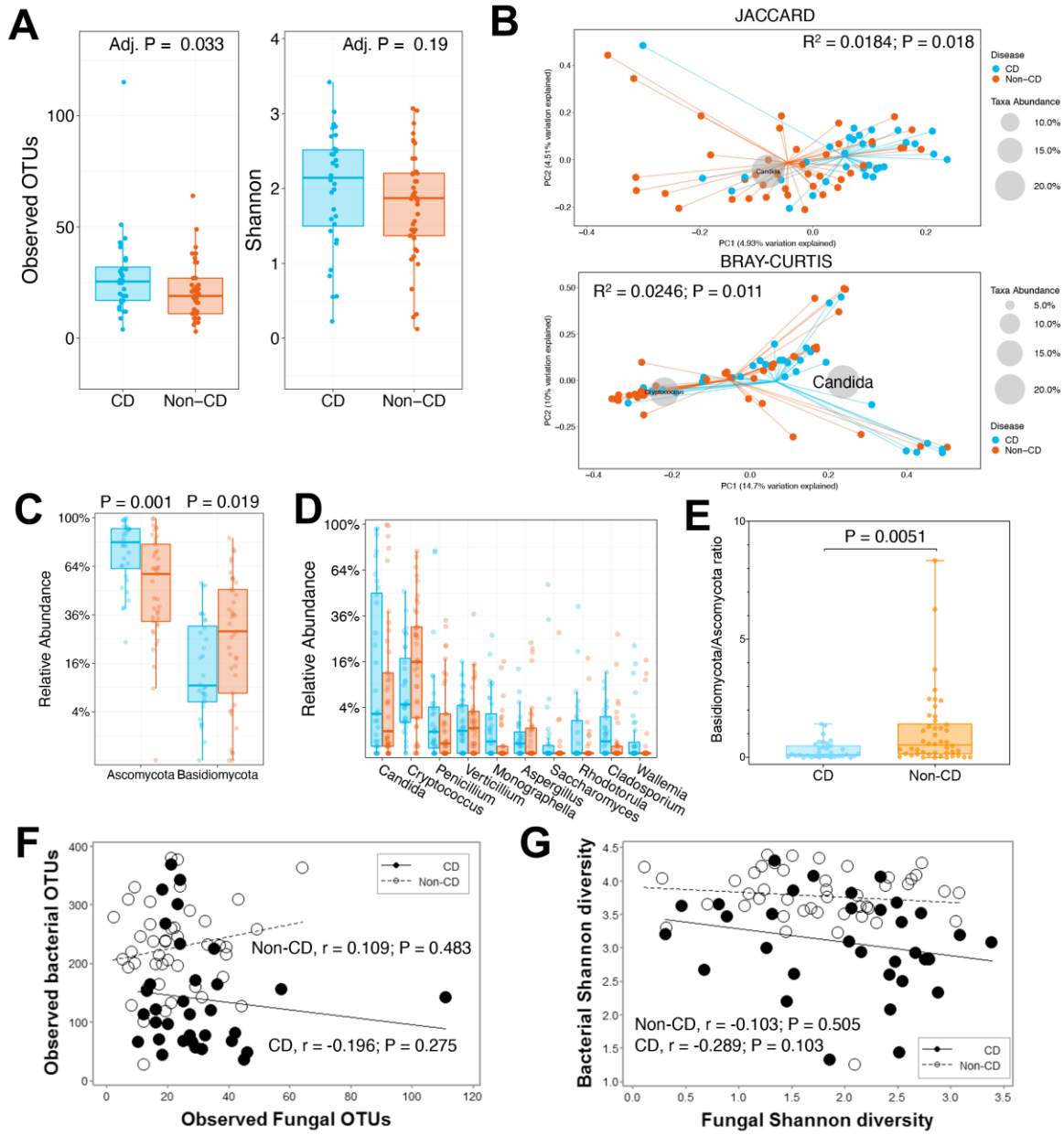


Figure 3

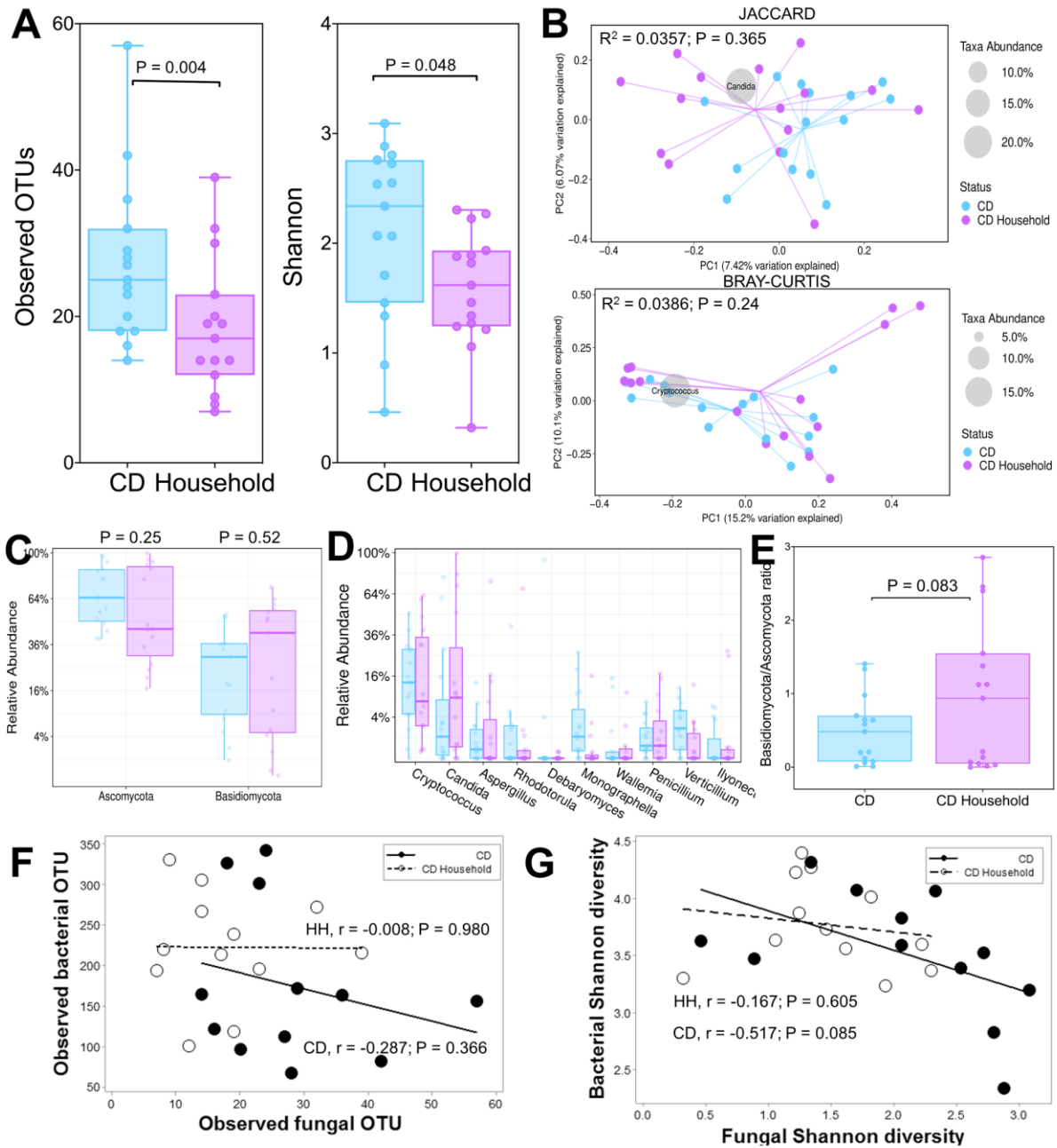


Figure 4

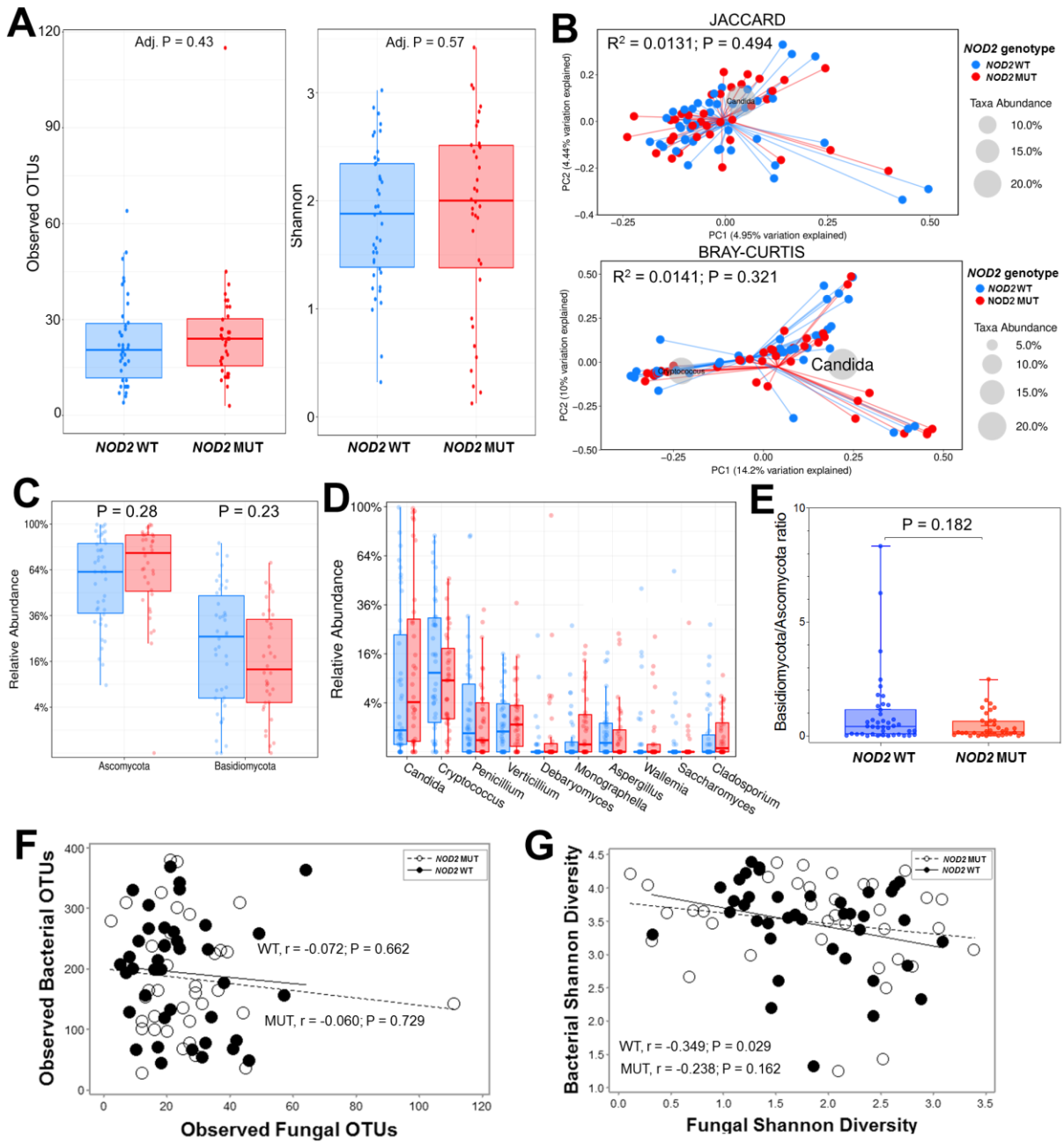


Figure 5

