

# Multi-meta-omics reveal distinct microbial genomic profiles and metabolic dysregulation in non-celiac gluten sensitivity

Kunal Dixit,<sup>1,2</sup> Susheel Bhanu Busi,<sup>3,4</sup> Anam Ahmed,<sup>5,6</sup> Avinash Kshirsagar,<sup>7</sup> Christian Jäger,<sup>3</sup> Alka Singh,<sup>5</sup> Varun Shah,<sup>7,8</sup> Sunil D. Saroj,<sup>1</sup> Vineet Ahuja,<sup>5</sup> Paul Wilmes,<sup>3</sup> Yogesh Shouche,<sup>7,9</sup> Govind Makharia,<sup>5</sup> Dhiraj Dhotre<sup>7</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 13.

**ABSTRACT** Non-celiac gluten sensitivity (NCGS) is an emerging diagnosis, and its symptoms overlap with irritable bowel syndrome (IBS). The gut microbiome is likely to play a role in the pathogenesis of NCGS. We analyzed the gut microbiome in patients with NCGS and in patients with IBS, using shotgun metagenomics and metabolomics of fecal samples. Analyses of taxonomic and functional microbial diversity revealed a higher abundance of methanogenic archaea, such as *Methanobrevibacter filiformis*, *Methanobrevibacter boviskoreani*, *Methanosphaera stadtmanae*, and a higher fold change in urea, uridine 5-monophosphate, and adenosine monophosphate in patients with NCGS compared to patients with IBS, who showed higher fold changes in metabolites gamma-aminobutyric acid and lactic acid. Furthermore, pangenome and metabolome analyses revealed disease-specific gene clusters, as well as genomic and metabolic features differentiating NCGS from IBS. While patients with NCGS did not show lower potential for gluten degradation, a lower synthetic potential for fructan beta-fructosidase was found in them. The present study provides an extensive analysis of taxonomic, genomic, and metabolic features that may play a role in the pathogenesis and symptom development in patients with NCGS.

**IMPORTANCE** Non-celiac gluten sensitivity (NCGS) is an emerging diagnosis with symptoms that overlap with irritable bowel syndrome (IBS). Using shotgun metagenomics and metabolomics, we report deeper insights into the microbiome profile, including viral and archaeal diversity, lower fructan degradation potential, the differential abundance of metabolites, and genomic features of gut bacteria in patients with NCGS. Understanding the microbiome associated with this disorder may shed light on the possible role of the microbiome in the pathophysiology of NCGS.

**KEYWORDS** non-celiac gluten sensitivity, metagenomics, metabolomics, gut microbiome, irritable bowel syndrome, gluten-free diet

Non-celiac gluten sensitivity (NCGS) is an emerging disorder caused due to ingestion of wheat or its protein called gluten (1, 2). NCGS is manifested by both gastrointestinal and extraintestinal symptoms, such as abdominal pain, altered stool frequency, nausea, headache, brain fog, tingling and numbness, fatigue, and musculoskeletal pain (3, 4). A systematic review has reported a global prevalence of 10% with marked regional variation, being more common in developed countries (5). The wide range of prevalence reported for this gluten spectrum disorder is due to the different definitions applied for NCGS categorization across studies. In contrast to celiac disease, there is no diagnostic biomarker, and the diagnosis of NCGS is currently established by the presence of clinical response to a gluten-free diet (GFD) and recurrence of symptoms on the resumption of gluten in the diet (6). Furthermore, the clinical features of NCGS overlap with those of

**Editor** Sonny T. M. Lee, Kansas State University, Manhattan, Kansas, USA

Address correspondence to Govind Makharia, govindmakharia@gmail.com, or Dhiraj Dhotre, dheerajdhotre@gmail.com.

The authors declare no conflict of interest.

**Received** 5 December 2025

**Accepted** 8 February 2026

**Published** 30 March 2026

Copyright © 2026 Dixit et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

patients with irritable bowel syndrome (IBS), and many such patients are misdiagnosed as IBS patients.

While symptoms of NCGS are dependent on ingestion of wheat, it is proposed that the disease may occur either because of high fermentable oligo-, di-, monosaccharides, and polyols (FODMAP) contents of wheat or due to gluten in it. The proposed mechanism for occurrence of symptoms in patients with NCGS is an innate immune response to gliadin peptides, hypersensitivity of the intestine, and intestinal motility disorder. A number of reports suggest gut dysbiosis in patients with IBS (7–10). It is plausible that gut microbiota handle gluten in such a way as to create peptides that induce innate immune response and intestinal dysmotility in patients with NCGS.

There is limited knowledge about the gut microbiota and metabolites in patients with NCGS; thus, it is imperative to study this gluten spectrum disorder with a multi-omic approach to delineate the differences and also to understand the commensal or pathogenic role of the microbiota in them.

To address the current gaps in knowledge, we recruited patients with IBS and NCGS based on standard criteria to map differences in the microbiome structure, functions, and metabolome profiles between the two disorders. We also estimated the effect of GFD on patients with NCGS by comparing NCGS patient samples pre- and post-GFD. We further examined the load of microbially coded genes responsible for gluten degradation between these disorders, as lower gluten degradation has already been found in other gluten spectrum disorders. We identified differences in microbiome-based taxonomic as well as functional profiles for NCGS and IBS patients.

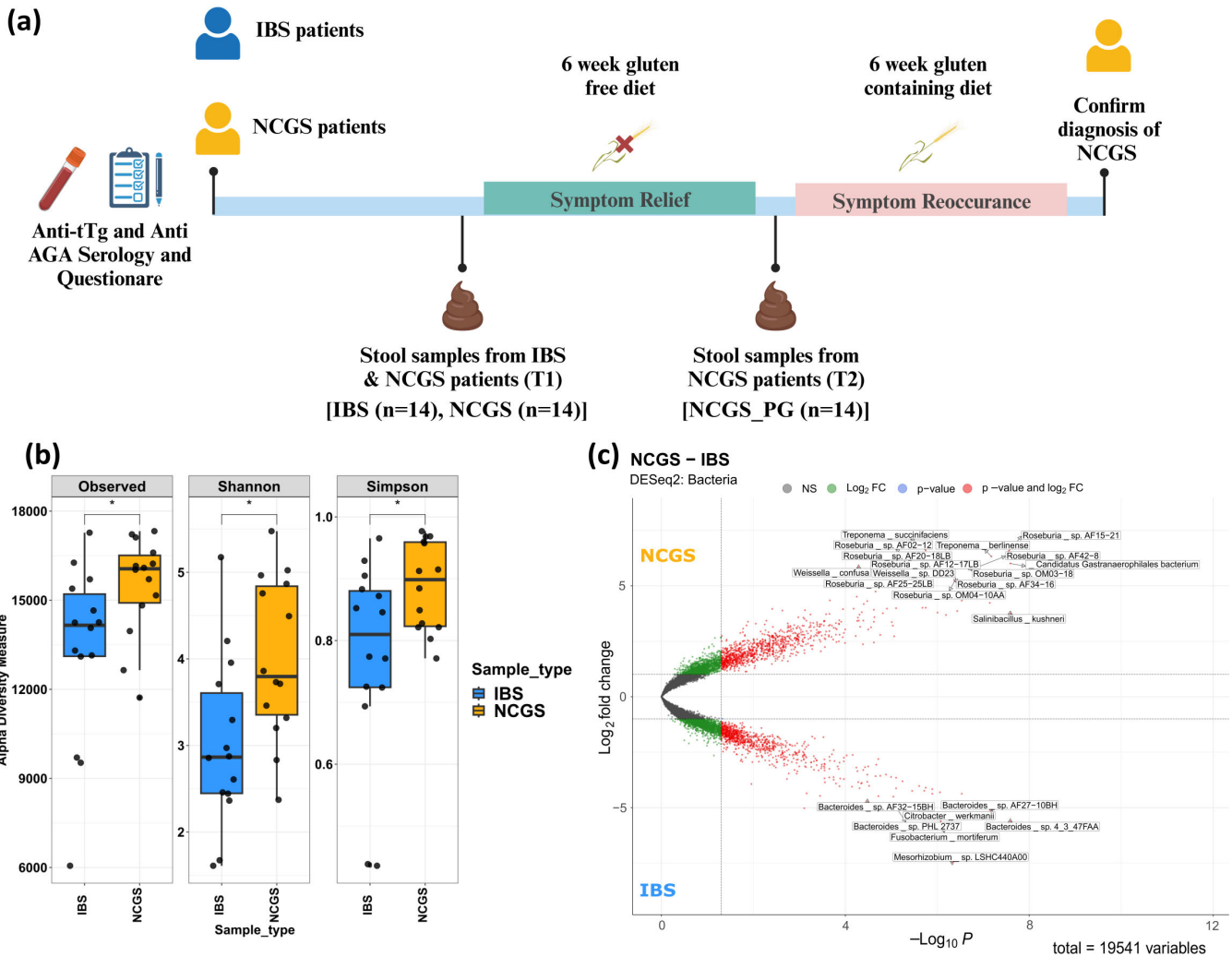
## RESULTS

### Recruitment of patients and study overview

The details of patient recruitment criteria, symptom score, and details for patient response to GFD intervention have been published elsewhere (11). To map the taxonomic diversity, functional potential, and metabolite profiles of the microbiome, we performed shotgun metagenomics and metabolomics on a total of 42 stool samples collected from patients with IBS ( $n = 14$ ), patients with NCGS before GFD ( $n = 14$ ; NCGS), and after GFD ( $n = 14$ ; NCGS patients post-GFD [NCGS\_PG]; Fig. 1a). Whole-genome metagenome sequencing yielded a total of  $7.0 \times 10^7 (\pm 1.0 \times 10^7 \text{ SD})$  reads per sample. Adapter trimming discarded 2.7% of the total reads, and on average, 19.10% of the reads were found to be of human origin and therefore discarded. The sequencing characteristics, pre- and post-quality control, and after assembly are highlighted in Data S1. Additionally, for the IBS, NCGS, and NCGS\_PG groups, a total of 406, 515, and 526 metagenomically assembled genomes (MAGs) were obtained after assembly and binning, out of which 170, 227, and 225 MAGs, respectively, had completeness >70% and contamination <10%. Quality data for MAGs are provided in the Data S2. A total of 171 metabolite features were profiled with gas chromatography-mass spectrometry (GC-MS), of which 82 were identified. Additionally, we could profile and identify a total of 87 metabolites using liquid chromatography-mass spectrometry (LC-MS) in stool samples of IBS patients and NCGS patients. The list of all metabolites detected in the stool samples by GC-MS and LC-MS is provided as Data S3.

### Metagenomics reveals diversity differences across groups

Microbial diversity was captured based on the metagenomic read mapping against the comprehensive Kraken database (see Materials and Methods). Comparative microbiome analyses were performed for IBS patients and NCGS patients (before GFD) as well as NCGS samples before and after GFD (NCGS\_PG). The diversity spanning three broad taxonomic clades (bacteria, archaea, and viruses) was assessed using alpha diversity indices (including richness, Shannon, and Simpson), alongside differential taxa analysis. All three alpha diversity metrics showed higher bacterial diversity in the NCGS group compared with the IBS group. Alpha diversity measures suggested higher bacterial



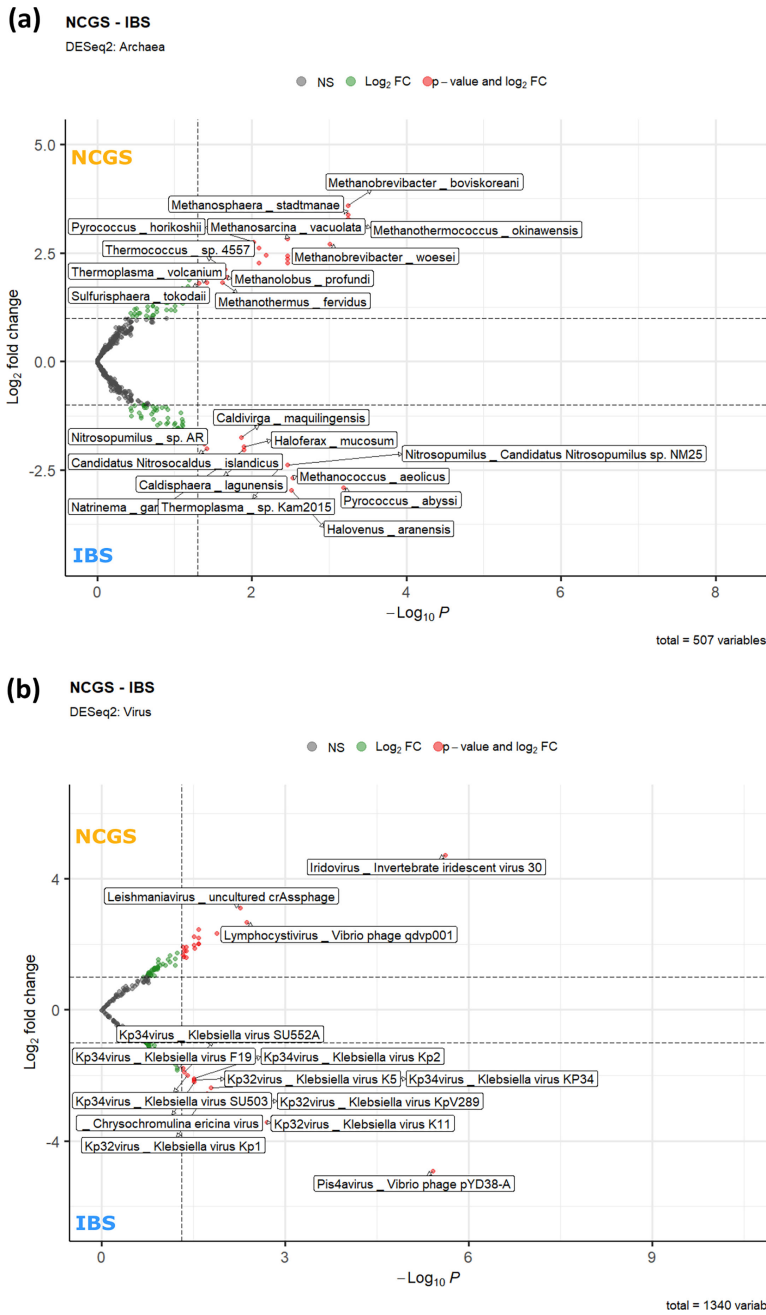
**FIG 1** Bacterial diversity varies between NCGS and IBS patients. (a) Study design scheme depicting patient recruitment and sample collection. (b) Box plot illustrating the bacterial alpha diversity measures of NCGS and IBS patient stool samples. (c) Volcano plot presenting the differentially abundant bacterial species in NCGS and IBS patients. Comparisons shown are between IBS and pre-GFD NCGS only.

diversity in the NCGS group compared to IBS patients (Fig. 1b). DESeq2 analyses (Wald test with Benjamini-Hochberg correction) revealed significant differences in taxa from all three taxonomic groups. The genera *Roseburia*, *Treponema*, *Weissella*, alongside *Salinibacillus kushneri*, were found in greater numbers in patients with NCGS. On the other hand, genus *Bacteroides*, along with *Fusobacterium mortiferum* and *Citrobacter werkmanii*, were higher in IBS patients (Fig. 1c; Data S2). Moreover, the differentially higher profile of NCGS archaea was mainly characterized by methane-producing species *Methanospaera*, *Methanobrevibacter*, *Methanosarcina*, *Methanothermococcus*, *Methanothermus*, etc. (Fig. 2a). Subsequently, we assessed the virome profiles of NCGS and IBS patients and found that they largely overlapped (Fig. S1). However, differentially higher amounts of *Iridovirus*, *Leishmaniovirus*, and *Lymphocystivirus* were found in NCGS patients; whereas, multiple phages targeting *Klebsiella* were found differentially abundant in IBS patients (Fig. 2b). The complete list of differentially abundant taxa can be found as part of Data S2.

**Functional diversity differences between NCGS and IBS patients**

To complement the taxonomic analyses, we assessed the functional potential of the metagenomes, i.e., of the assembled contigs, by mapping them to the Kyoto

Downloaded from https://journals.asm.org/journal/msphere on 24 June 2026 by 192.171.199.129.



**FIG 2** NCGS and IBS have differential archaeal and viral community structure. (a) Volcano plot depicting differentially abundant archaeal diversity between study groups based on DeSeq2 analysis. (b) Volcano plot depicting differentially abundant viral diversity between study groups based on DeSeq2 analysis. Comparisons shown are between IBS and pre-GFD NCGS only.

encyclopedia of genes and genomes (KEGG), carbohydrate-active enzyme (CAZyme), and Pfam databases, respectively. A total of 14,014 KEGG Orthologs (KOs), 338 CAZymes, and 8,988 protein families were identified in total, of which 305 KOs, 16 CAZymes, and 178 protein families were differential between NCGS and IBS (DESeq2 analysis; Wald test; Fig. S2; Data S4). A higher number of KO and protein families was observed in NCGS patients compared to IBS patients; however, the differences were not statistically significant. Interestingly, NCGS patients possessed a significantly higher abundance of K02410, K00575, and K02053 (*Padj.* 0.015, 0.030, and 0.036, respectively), associated with microbial communication, such as bacterial chemotaxis, the two-component system,

Downloaded from https://journals.asm.org/journal/msphere on 24 June 2026 by 192.171.199.129.

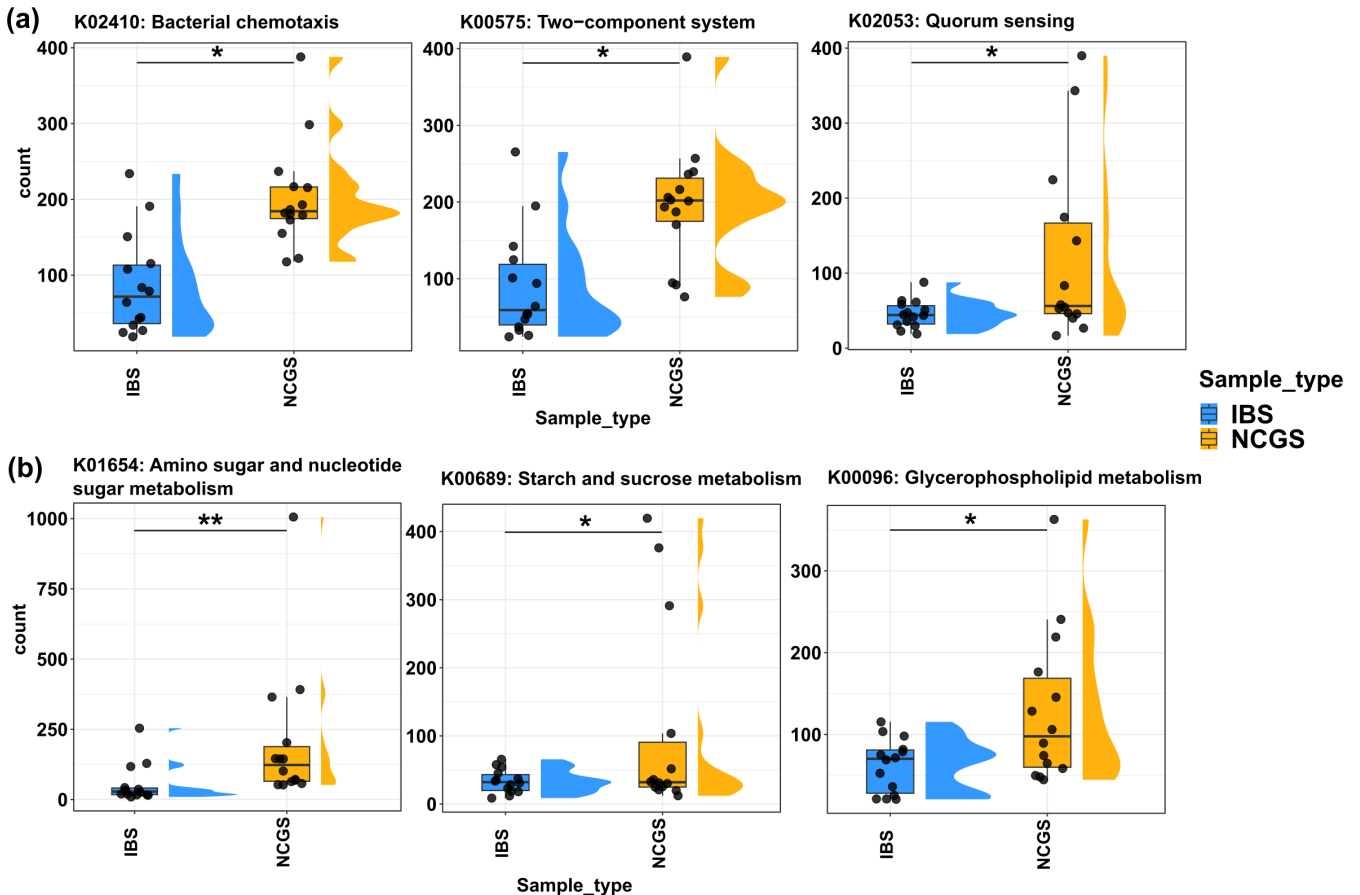
and quorum sensing. Moreover, significantly higher levels of carbohydrate-active enzymes in NCGS patients were reflected by higher levels of K01654, K00689, and K00096 (*P*<sub>adj.</sub> 0.004, 0.015, and 0.040, respectively), which are responsible for sugar metabolism (Fig. 3a and b).

### Effect of gluten-free diet on patients with NCGS

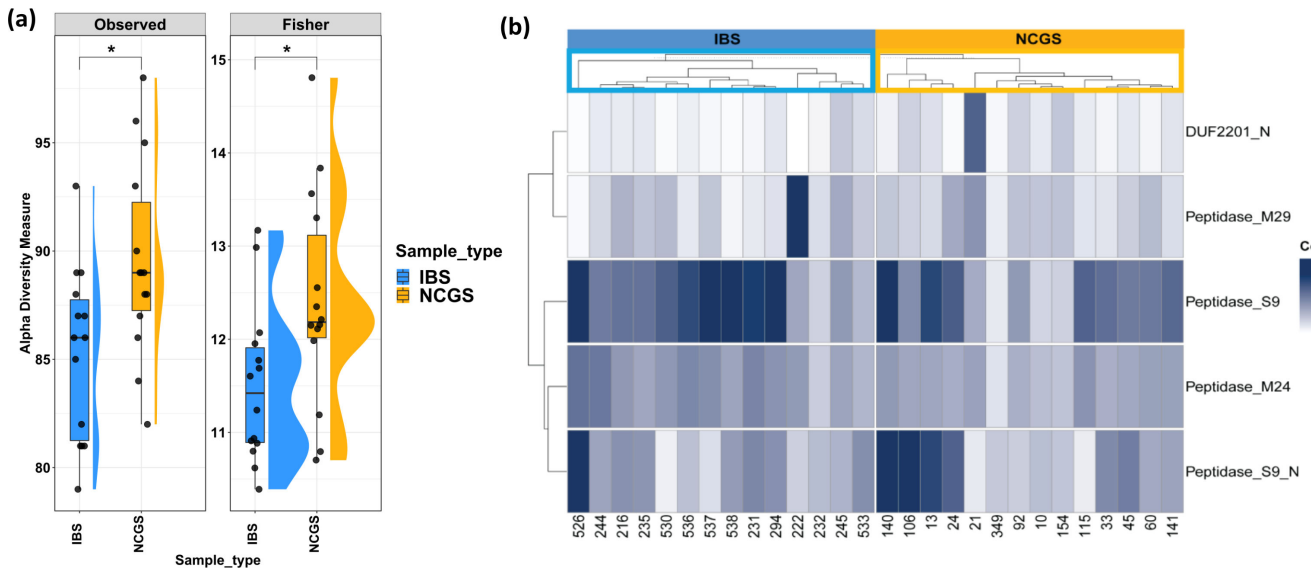
NCGS patients were kept on GFD for 6 weeks to map microbiome changes as a response to the intervention. The results for NCGS samples before and after GFD are mainly presented as part of the supplemental data. Kraken-based taxonomic annotation identified a total of 252 bacterial species and 19 viruses that were differentially abundant between NCGS and NCGS\_PG groups. No archaeal species were found to be significantly different between these groups. The complete list of differentially abundant taxa is present as Data S2. In case of differences in the functional features in NCGS patients, no significant differences were reported in the overall diversity measures (observed features and Fisher index; *P* > 0.05) as an effect of GFD (Fig. S3a to c); however, 54 KOs, 16 CAZymes, and 26 protein families were found significantly different between NCGS and NCGS\_PG groups (Fig. S3d to f and Data S4).

### Gluten degradation potential in NCGS as compared to IBS patients

Gluten-degrading proteolytic enzymes are speculated to correlate with symptom development in gluten spectrum disorders (12, 13). Interestingly, NCGS microbiomes encoded significantly higher proteases compared to IBS (Fig. 4a; *P*<sub>adj.</sub> < 0.05, Wilcoxon's ranked sum test). Enzymes capable of degrading gluten were screened for their



**FIG 3** Microbial functional potential of NCGS and IBS patients is significantly different. KOs responsible for (a) bacterial communication and (b) sugar metabolism were found significantly higher in patients with NCGS (*P*<sub>adj.</sub> < 0.05). Comparisons shown are between IBS and pre-GFD NCGS only.

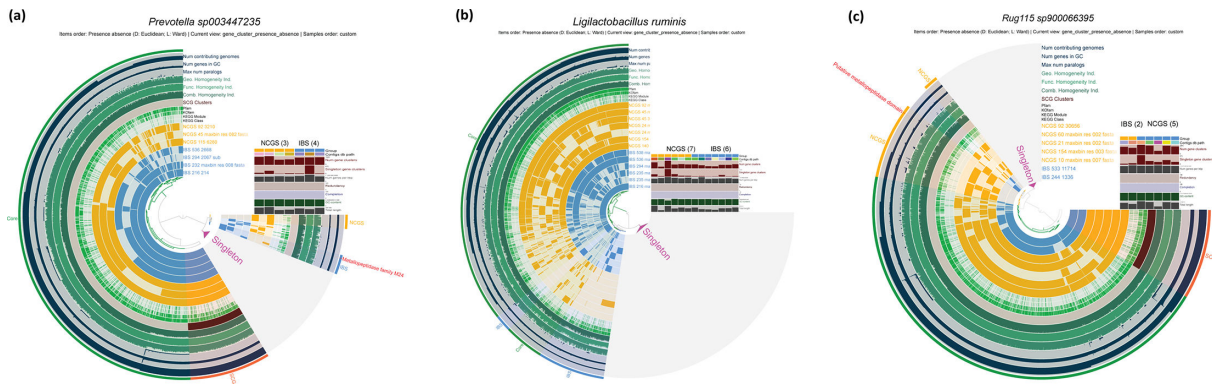


**FIG 4** Protease profiles of NCGS and IBS patients. (a) Plot presenting higher alpha diversity of proteases in patients with NCGS. (b) Heatmap showing similar gluten degradation potential in both NCGS and IBS patients. Comparisons shown are between IBS and pre-GFD NCGS only.

total numbers between the two perturbations and further mapped in the various pangenomes to test the involvement of differentially abundant taxa in gluten degradation. Microbial gluten degradation potential was assessed based on levels of previously reported gluten-degrading enzymes. These included putative metallopeptidase domain (DUF2201\_N), thermophilic metalloprotease (peptidase\_M29), prolyl oligopeptidase (peptidase\_S9), prolyl oligopeptidase, N-terminal beta-propeller domain (peptidase\_S9\_N), and metallopeptidase family M24 (peptidase\_M24). NCGS and IBS patients were found to have similar microbial gluten degradation potential based on the type and distribution of the enzymes. Importantly, a hierarchical clustering approach revealed distinct profiles separating the NCGS and IBS groups based on the abundance of these gluten-degrading enzymes (Fig. 4b).

### Pangenome analysis revealed NCGS-specific gene clusters

To further understand the taxa contributing to the observed functional differences between the NCGS and IBS groups, we assessed the differential abundance of the medium-to-high-quality MAGs generated from the metagenomes (14). These analyses revealed 40 differentially abundant MAGs between NCGS and IBS. These included but were not limited to *Prevotella sp003447235*, *Ligilactobacillus ruminis*, *RUG115 sp900066395* (Data S5), which were further used for pangenome analyses to map the species-specific gluten degradation potential and explore differences in the group-specific gene clusters (Fig. 5a to c). In addition to earlier reported protein families with gluten degradation potential, we investigated disorder-specific gene clusters responsible for gluten degradation using pangenome analysis for these three species. We found that these three species had the potential for gluten degradation in both the NCGS and IBS groups, although no significant differences in their overall abundances were observed. Pangenome analysis revealed gene clusters on the genome of *Ligilactobacillus ruminis* that were unique to IBS. In contrast, the genomes of *RUG115 sp900066395* from NCGS patients had gene clusters unique to this disorder and encoded two additional putative metallopeptidase domains. Meanwhile, *Prevotella sp003447235* was found to possess unique gene clusters for both NCGS and IBS groups, with two metallopeptidase family “M24” gene cluster hits mapped specifically to IBS. A list of group-specific gene clusters in differential species pangenomes is provided in Data S5.



**FIG 5** Pangenome analysis of metagenomically assembled genomes. (a) Pangenome of *Prevotella* sp003447235 shows IBS and NCGS-specific gene clusters. (b) Pangenome of *Ligilactobacillus ruminis* shows IBS-specific gene clusters. (c) Pangenome of *RUG115* sp900066395 shows NCGS-specific gene clusters. Comparisons shown are between IBS and pre-GFD NCGS only.

### Fructan degradation in NCGS

In addition to the gluten degradation capacity, microbial fructan degradation was tested in NCGS patients because fructan is increasingly suspected of being a probable cause of symptoms in NCGS patients (15). To understand if the potential for fructan degradation varied between NCGS and IBS, we extracted the relevant gene abundance information from the individual metagenomes across all samples. We found that the genes mapping to the KO K03332, encoding fructan beta-fructosidase involved in fructan hydrolysis, were observed at significantly lower levels in patients with NCGS (Fig. 6;  $P_{adj} < 0.05$ , Wilcoxon rank-sum test) than in IBS. The median transcripts per million (TPM) abundance and IQR for the NCGS and IBS groups for this gene were 233.58 (174.37–336.66) and 385.43 (255.69–435.59), respectively.

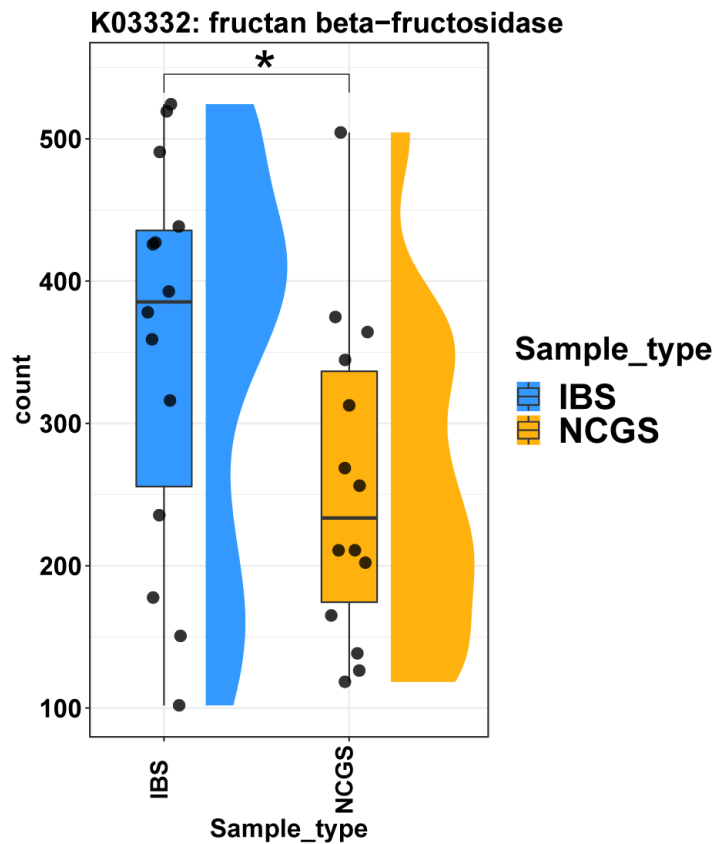
### Metabolome profiles in patients

Glucose, carbonic acid, and glutamic acid were the most highly abundant metabolites found with GC-MS analysis in both NCGS and IBS patient stool samples and were similar between the NCGS and IBS patients. Diethylglycol, 5-aminovaleric acid, and lactic acid were found to be more abundant in IBS patients compared to NCGS patients. The principal coordinate analysis (PCoA) revealed that the stool metabolite profiles between NCGS and IBS patients, as well as NCGS patients before (NCGS) and after GFD (NCGS\_PG), were in fact overlapping (Fig. S4). MetaboAnalyst 6.0 was used to find the fold changes in the metabolite profiles between NCGS and IBS patients. A total of 15 metabolites were found with log2 fold change greater than 2, which include uridine 5-monophosphate, adenosine monophosphate, and urea at higher fold changes in NCGS patients and metabolites such as gamma-aminobutyric acid, lactic acid, indolelactic acid, fucose, etc. at higher fold changes in IBS patients, although the differences were not statistically significant (Fig. 7).

### DISCUSSION

The prevalence of NCGS worldwide and the increasing evidence of its associations with the gut microbiome call for a comprehensive exploration of the microbiome associated with this gluten spectrum disorder (1, 2). Additionally, the highly overlapping manifestations and the lack of biomarkers call for a multi-omic analysis of the gut microbiome (16). Several reports suggest that the gut microbiome has a critical influence on NCGS, in addition to nutritional and environmental confounders (17–19). The present study provides clear findings, as the sampling strategy was aimed at excluding false-positive NCGS patients. We used shotgun metagenomics and metabolomics to compare the microbiome of 42 stool samples from IBS, NCGS, and NCGS post-GFD patients, revealing

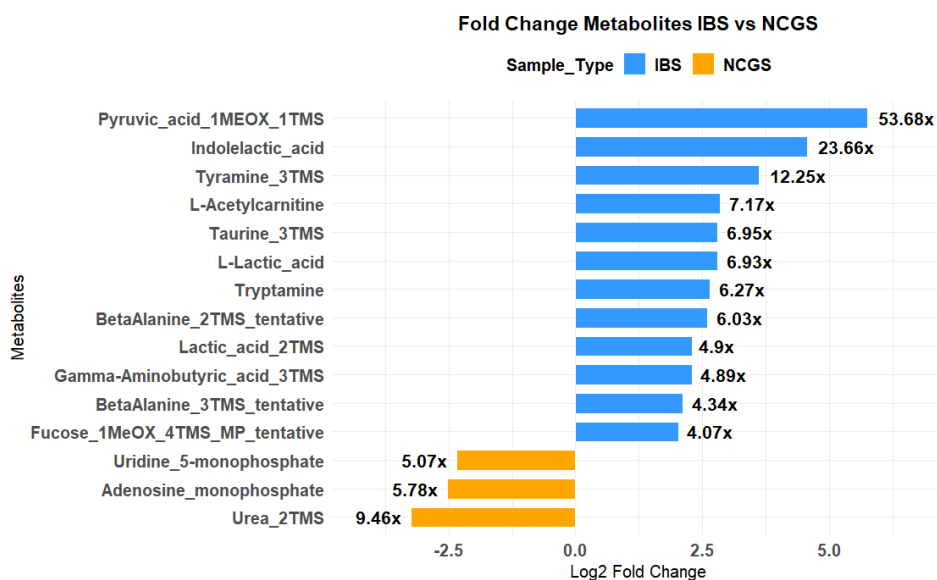
Downloaded from https://journals.asm.org/journal/msphere on 24 June 2026 by 192.171.199.129.



**FIG 6** Lower abundance of fructan beta-fructosidase in patients with non-celiac gluten sensitivity. Comparisons shown are between IBS and pre-GFD NCGS only.

differences between taxonomic profiles, genomic features, and metabolites. The current study is the first extensive report in the context of NCGS that includes differential diversity across archaea and viruses besides bacteria as well as genomic potential and metabolome of stool samples. A recent report by Mohammadzadeh et al. (20) has highlighted that archaea, a key component of the human gut microbiome, remain understudied in terms of their pathogenicity (20). In our study, several methanogenic archaea, such as *Methanobrevibacter filiformis*, *Methanosarcina vacuolata*, *Methanobrevibacter boviskoreani*, and *Methanosphaera stadtmanae*, were found significantly more frequently in NCGS patients. *Methanobrevibacter* species are the main representatives of methanogens in the human system and are known to modulate host nutrient degradation. Several studies have demonstrated the presence of *Methanobrevibacter smithii* in patients with IBS, infectious endocarditis, and Parkinson's disease, and it has been associated with high methane emission (20). Interestingly, we found differentially higher abundances of *Methanobrevibacter filiformis*, *Methanobrevibacter boviskoreani*, and *Methanobrevibacter woesei* in NCGS patients. *Methanosphaera stadtmanae* has also been previously detected in humans and is known to utilize acetate for methane production. It has been reported that higher methane synthesis affects intestinal transit time in IBS patients (21). It is plausible that a similar phenomenon may be responsible for the constipation in NCGS patients.

Furthermore, proteases are known to influence intestinal barrier functions, alter bowel habits, change visceral sensitivity, and immune response (22). In this context, the higher total protease levels in NCGS patients are also of interest. Higher concentrations of proteases in stool and biopsy specimens have previously been reported in patients with IBS and have been suggested to play a role in pathophysiology (22). There are no reports investigating proteases in NCGS patients. Here, we found significantly higher



**FIG 7** Fold change metabolite profile in NCGS and IBS patients. Bar plot showing fold changes for metabolites (log2 fold change > 2 are shown) identified using GC-MS and LC-MS analysis. The y-axis presents metabolites, the x-axis presents log2 fold change, and the actual fold changes are indicated on the respective bars. Comparisons shown are between IBS and pre-GFD NCGS only.

levels of proteases in NCGS patients compared to IBS patients, suggesting that they may affect the intestinal barrier and bowel habits of these patients, leading to the complex symptomatology observed in NCGS. Increased proteolysis could be due to the release of cysteine, serine, and metalloproteases by intestinal bacteria (22). Interestingly, some of these enzymes are known for their gluten-degrading activity and are suspected to have therapeutic potential in celiac disease, as lower levels of these enzymes have been reported in CeD patients (12, 23). We observed no significant difference in the levels of previously reported gluten-degrading enzymes between the study groups, suggesting that both disorders have similar gluten-degrading potential. This is further evidence that symptoms in NCGS patients may be due to a dietary factor other than gluten. A study by Skodje et al. (15) recently reported that fructans, a type of FODMAP, trigger symptoms in self-reported NCGS patients rather than gluten (15). Moreover, apart from GFD, a low-FODMAP diet has been shown to improve physiology in terms of reducing intraepithelial lymphocytes and increasing mucin-producing cells in NCGS patients (24). In parallel, we have observed in patients with NCGS a significantly lower abundance of fructan beta-fructosidase (K03332), which is primarily responsible for the hydrolysis of terminal beta-D-fructofuranose residues in fructans.

Pangenome analysis revealed NCGS- and IBS-specific gene clusters on the genomes of *Prevotella sp003447235* and *Rug115 sp900066395*. The appearance of additional genes encoding the gluten-degrading metallopeptidase S4 family in the IBS-specific cluster and of the putative metallopeptidase domain in the NCGS-specific clusters of these organisms may indicate their specific adaptability to the NCGS and IBS patient groups. In addition, pangenome analyses revealed genes encoding glycosyltransferase family 2, group 1, and WbsX, along with rhamnosyltransferase (EC: 2.4.1), which are present in the NCGS-specific gene cluster of *Rug115 sp900066395*. These proteins were previously found in the genome of *Ruminococcus gnavus* and play a role in the production of complex glucorhamnan polysaccharide, an inflammatory polysaccharide that induces TNF $\alpha$ , an inflammatory cytokine in patients with Crohn's disease (25). Further validation is essential for the presence of this inflammatory molecule in NCGS patients and its link to the symptoms associated with the disorder.

Metabolome profiles of gluten spectrum disorders are limitedly investigated in previous studies to obtain biomarkers. Jacobs et al. (9) have reported that IBS can be characterized by lower gentisate and hydrocinnamate as well as higher tyramine, which was also found to be true in the present study. Also, they found that diarrhea-predominant IBS differs from constipation-predominant IBS in terms of increased bile acids, polyamines, malate, and fumarate (9). While stool metabolome profiles of NCGS and IBS patients appear similar for the most part, we identified a few metabolite targets with high fold change between the disorders. Urea has been found to cause intestinal barrier dysfunction when present in higher than clinically relevant concentrations. Higher amounts of urea have been reported to cause dissipation of transepithelial electrical resistance, causing loss of tight junction proteins (26). On the other hand, a higher concentration of pentoses in NCGS patients could be indicative of impaired metabolism.

We acknowledge that a caveat of our findings is the lack of a control group of healthy individuals without any NCGS or IBS manifestation. Interestingly, an earlier report by Mazcorro et al. (27) highlighted the taxonomic differences between healthy controls and NCGS patients (27). They reported significantly higher amounts of *Ruminococcaceae* in NCGS patients along with *Saccharibacteria*, *Actinobacillus*, and *Fingoldia*. In contrast, lower *Sphingobacterium* counts were found in healthy controls (27). Another study comparing NCGS with the microbiota of healthy individuals reported higher *Peptostreptococcaceae* and lower *Porphyromonadaceae* abundances in patients with NCGS (28). Though these are important findings, as highlighted previously, the diagnosis of NCGS is not straightforward, necessitating the need to develop key taxonomic and functional biomarkers for classifying and differentiating the two disorders. Another limitation of the current study is not taking IBS subtype classification into account for subtle variations within the microbiome (9). However, the current study focused on the overall differences between NCGS and IBS, given the possible misclassification of the two disorders, and a study with a larger sample size is required to track if IBS subtypes vary in microbiome composition as compared to NCGS patients. Our collective findings, including taxonomic, functional capacity, and the metabolomic information, therefore, provide a platform for leveraging key multi-omic modalities. These will not only allow for the classification and identification of NCGS but also will enable the understanding of underlying microbially mediated mechanisms in disease etiology.

## MATERIALS AND METHODS

### Patient recruitment

Patients were recruited from the outpatient clinic of the Department of Gastroenterology and Human Nutrition at the All India Institute of Medical Sciences (New Delhi). Patients were screened for IBS using the Rome IV criteria and the Bristol stool chart. Individuals with regular alcohol consumption (>20 g/week), smokers (>5 cigarettes or bidis/week), patients who had undergone abdominal surgery, intestinal resection, active nonsteroidal anti-inflammatory drugs usage, antibiotics, or prebiotics within 8 weeks, and patients with malabsorption syndromes (tropical sprue, giardiasis, or COVID) were excluded from the study. Routine screening tests were performed in all patients with IBS, such as microscopic examination of the stool (to rule out parasitic infections), fecal occult blood test, and routine hematologic and biochemical tests. Patients were also tested for AGA IgA and IgG titers >30 U/mL (QUANTA Lite Gliadin IgA, IgG). Celiac disease was ruled out in gluten-sensitive patients with negative anti-tTG IgA (QUANTA Lite R h-tTG IgA) and villous atrophy using histology of small intestinal biopsy. Gluten-sensitive patients were treated with a GFD intervention for 6 weeks. Patients who responded to GFD (>30% decrease in symptoms) were maintained on their regular (gluten containing) diet to confirm NCGS diagnosis if symptoms recurred. Data for response to GFD and symptom scores have been published elsewhere (11). Fecal samples were collected from IBS and gluten-sensitive patients. At a second time point, stool samples were collected from gluten-sensitive patients who responded to GFD after a 6-week intervention (Fig. 1a). All

samples were frozen and stored at  $-80^{\circ}\text{C}$  until further processing. More details for the patient recruitment and symptom data are published elsewhere (11).

### DNA extraction and metagenomic sequencing

Forty-two stool samples were collected from IBS patients and NCGS patients (before and after GFD). The QIAamp 96 PowerFecal QIAcube HT kit (QIAGEN, Valencia, CA, USA) was used for fecal DNA extraction using the QIAcube HT liquid handler system. Spectrophotometric methods NanoDrop ND-1000 spectrophotometer (NanoDrop Biotechnologies, USA) and Qubit fluorometer (Thermo Fisher Scientific, USA) were used to quantify the extracted DNA. DNA integrity was checked using agarose gel electrophoresis. The extracted DNA was diluted to  $100\text{ ng}/\mu\text{L}$  and stored at  $-20^{\circ}\text{C}$  until further use.

Sequencing was outsourced to Medgenome Labs, India. The DNA samples were subjected to random shotgun sequencing. Metagenome library preparation was performed using the KAPA DNA HyperPrep kit with nine PCR cycles. Ultrasonication was used to shear the DNA, followed by a sequence of enzymatic treatments. The next step was ligation of indexed adapters followed by purification with reversible immobilization using SPRI beads (Beckman Coulter Lifesciences, USA). Prepared libraries were sequenced on the Illumina HiSeqX platform to generate 10 gigabytes of data per sample using paired-end shotgun sequencing with  $150 \times 2$  chemistry. Raw sequencing output was delivered in FASTQ format, with standard Illumina run-quality metrics indicating  $\geq 75\%$  of bases at Q30 or higher.

### Metagenomic data processing

Pre-processing of the above metagenomic data, genome assembly, as well as taxonomic and functional annotations with various databases, were done using the Integrated Meta-omics Pipeline (IMP) version 3 (29). Pre-processing included adapter trimming and subsequent filtering of *Homo sapiens* DNA (hg38) to obtain only microbiome-based gene sequences and functions. The next step was assembly using MEGAHIT (version 2.0) (30). A consensus approach including MaxBin (31) and MetaBat (32) was used to reconstruct MAGs, followed by selection of a non-redundant collection of MAGs using DASTool with a score threshold of 0.7 (33). CheckM v1.16 was used to assess the quality of MAGs, and taxonomy was assigned using the GTDB toolkit (34). IMP-based annotation of proteins (Pfam) (35), metabolic pathways (KEGG and MetaCyc) (36, 37), and carbohydrate-active enzymes (dbCAN) (38) was performed using contigs. The gene counts for these features were obtained by mapping the pre-processed reads (trimmed, quality-filtered, and human-reads removed) to the contigs and subsequently using featureCounts to obtain gene-specific abundances (39). To address the differences in sequencing depth per sample and variability in contig characteristics, we employed the TPM method for normalization, which accounts for the sequencing depth and the length of each identified gene. Read-based microbial taxonomic identification with Kraken (v2.1.0) (40) was also performed for metagenome data. For the read-based taxonomy, the maxikraken database available at [https://lomanlab.github.io/mockcommunity/mc\\_data-bases.html](https://lomanlab.github.io/mockcommunity/mc_data-bases.html) was used. Specifically, the maxikraken2\_1903\_140GB (March 2019, 140 GB) database was used to capture all domains of life.

Additionally, pangenome analysis was performed for the differentially abundant MAGs between study groups using the workflow of Anvi'o v7.1 (41, 42). Gene cluster abundances were used to generate dendrograms, and pangenome analysis was performed using the `--min-bit 0.5` and `--mcl-inflation 10` parameters. Singletons were not considered for differential analysis and were masked in subsequent figures, as it is hard to establish the functional capacity of singletons in driving disease etiology.

### Metabolite extraction and GC-MS measurements

Stool samples were processed for homogenization based on the previous protocol, and around 50 mg of homogenized stool samples were taken for metabolite analyses

(43). An internal standard mix was added to the supernatant of homogenized samples. Liquid-liquid extraction was carried out from the supernatant with cold chloroform and methanol–water. The extract was split and was vacuum dried to use further for GC-MS and LC-MS analysis. The specific analytical details for sample pre-processing, GC-MS measurements, and information about mass spectrometric data post-processing can be found in a previous publication (43). For quality control, pool samples were introduced in the measurement sequence after every fifth measurement. Compounds were annotated based on retention time and mass spectrum using an in-house mass spectral library, with an overall similarity threshold set at 0.80. Peak areas of two automatically defined quantification ions were normalized by dividing them by the peak area of the internal standard, correcting for procedural variance.

### LC-MS measurements

Metabolite analyses were conducted using a Thermo Vanquish LC coupled to a Thermo Orbitrap Exploris 240 mass spectrometer. Chromatography utilized a SeQuant ZIC-pHILIC 5  $\mu\text{m}$  polymer column (150  $\times$  2.1 mm) connected to the corresponding SeQuant ZIC-pHILIC Guard (20  $\times$  2.1 mm) pre-column, maintained at a column temperature of 45°C. The flow rate was set at 0.2 mL/min, with mobile phases consisting of 20 mmol/L ammonium carbonate in water, pH 9.2 (eluent A), and acetonitrile (eluent B). The gradient profile was as follows: 0 min, 80% B; 3 min, 80% B; 18 min, 20% B; 19 min, 20% B; 20 min, 80% B; 24.5 min, 80%; 25.5 min, 80% B (0.4 mL/min); 29.5 min, 80% B (0.4 mL/min); 30 min, 80% B (0.2 mL/min). An injection volume of 5  $\mu\text{L}$  was used. All MS experiments employed electrospray ionization with polarity switching enabled (+ESI/–ESI). Source parameters were set as follows: sheath gas flow rate, 35; aux gas flow rate, 7; sweep gas flow rate, 0; spray voltage (static), 3 kV ( $\pm$ ); ion transfer tube temperature, 320°C; vaporizer temperature, 275°C. The Orbitrap mass analyzer operated at a resolving power of 60,000 in full-scan mode [scan range:  $m/z$  75–1,000; radio-frequency (RF) lens, 70%, automatic gain control (AGC) target, standard; max injection time, 100 ms]. Data were acquired using Thermo Xcalibur software (Version 4.5.474.0). Data post-processing was performed with TraceFinder (Version 5.1, Build 110).

### Data analyses

All downstream analysis for the metagenome was performed using assembled contigs, and MAGs were only used for pan-genome analysis. Comparative microbiome analysis was performed for IBS and pre-GFD NCGS patient samples as well as pre-GFD (NCGS) and post-GFD (NCGS\_PG) patient samples. As stated above, post-GFD samples (NCGS\_PG) were analyzed solely for within-patient comparisons and were not included in the NCGS–IBS group comparisons. Statistical analyses were performed and figures generated using R v4.1.1 statistical software (<https://www.R-project.org/>) with phyloseq (44), ggplot2 (45), ggpubr (46), ape (47), vegan (48), and reshape2 (42) packages. Nonparametric multidimensional scaling and alpha diversity indices, such as the Shannon index, Simpson index, and observed diversity, were calculated to capture the diversity of microbial taxa, proteins, and cazymes ( $P_{\text{adj.}} < 0.05$ , Wilcoxon signed-rank test). DESeq2 analysis (Wald test) with false discovery rate adjustments (Benjamini-Hochberg method) for multiple testing was used to find proteins, cazymes, KOs, and taxa significantly enriched in NCGS and IBS groups (49). The EnhancedVolcano package was used to create volcano graphs representing statistically significant differences between study groups. For metabolome analysis, the response ratios were calculated as chromatographic peak area normalized with the peak area for internal standards. The analysis refers to the relative abundance of a particular metabolite when comparing IBS and NCGS groups. PCoA was employed to assess the overall metabolite diversity between the study groups. Fold changes and log<sub>2</sub> fold changes for the metabolites were calculated using an open-source pipeline, MetaboAnalyst v6.0 (50), and were plotted using the ggplot2 package.

## ACKNOWLEDGMENTS

The authors are thankful to the Department of Science and Technology (DST), the Government of India (EMR/2016/004735 dated 5 September 2018) for funding the research. Kunal Dixit is grateful to the EMBO Scientific Exchange Grant for funding his visit to the University of Luxembourg.

Authors would like to thank Ms. Gayatri Chandan at the National Center for Cell Sciences, Pune, for her help in extraction of metabolites.

## AUTHOR AFFILIATIONS

<sup>1</sup>Symbiosis School of Biological Sciences, Symbiosis International (Deemed University), Pune, India

<sup>2</sup>Department of Food Science and Technology, University of California, Davis, California, USA

<sup>3</sup>Luxembourg Center for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>4</sup>UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, United Kingdom

<sup>5</sup>Department of Gastroenterology and Human nutrition, All India Institute of Medical Sciences, New Delhi, India

<sup>6</sup>McMaster University, Hamilton, Canada

<sup>7</sup>National Centre for Cell Science, Pune, India

<sup>8</sup>Gujarat Biotechnology University, Gandhinagar, India

<sup>9</sup>SKAN Research Trust, Bengaluru, India

## AUTHOR ORCIDS

Kunal Dixit  <http://orcid.org/0000-0002-1334-6997>

Govind Makharia  <http://orcid.org/0000-0002-2474-2194>

Dhiraj Dhotre  <http://orcid.org/0000-0002-5000-7396>

## AUTHOR CONTRIBUTIONS

Kunal Dixit, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review and editing | Susheel Bhanu Busi, Formal analysis, Investigation, Software, Visualization, Writing – original draft, Writing – review and editing | Anam Ahmed, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Avinash Kshirsagar, Formal analysis, Methodology | Christian Jäger, Formal analysis, Methodology, Writing – original draft | Alka Singh, Data curation, Methodology, Writing – review and editing | Varun Shah, Methodology, Software, Validation, Writing – review and editing | Sunil D. Saroj, Data curation, Investigation, Resources, Supervision, Validation, Writing – review and editing | Vineet Ahuja, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review and editing | Paul Wilmes, Formal analysis, Investigation, Resources, Software, Supervision, Validation, Writing – review and editing | Yogesh Shouche, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review and editing | Govind Makharia, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review and editing | Dhiraj Dhotre, Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review and editing

## DATA AVAILABILITY

The data generated in the study are available at the National Center for Biotechnology Information (NCBI) BioProject Repository with ID [PRJNA981497](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA981497). The STORMS (Strengthening The Organizing and Reporting of Microbiome Studies) checklist is available at <https://github.com/dixit-kunal/NCGS-IBS-Multiomics-Study-Data>.

## ETHICS APPROVAL

Patient recruitment was initiated after approval from the ethics committees of the All India Institute of Medical Sciences (AIIMS), New Delhi (IEC/68/08.01.2016), and the National Center for Cell Science (NCCS), Pune (RP-58/2017). Data acquisition and sample collection were done after obtaining informed and written consent from the participants.

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Data S1 (mSphere00856-25-S0001.xlsx).** Contig characteristics.

**Data S2 (mSphere00856-25-S0002.xlsx).** Metagenome bin summary, differential MAGs, and Kraken-based differential taxa.

**Data S3 (mSphere00856-25-S0003.xlsx).** List of metabolites GCMS and LCMS.

**Data S4 (mSphere00856-25-S0004.xlsx).** Differential features: metagenome analysis.

**Data S5 (mSphere00856-25-S0005.xlsx).** Disorder-specific gene clusters from pangenomes.

**Figure S1 (mSphere00856-25-S0006.tiff).** Viral diversity in NCGS and IBS.

**Figure S2 (mSphere00856-25-S0007.tiff).** Differential metagenome features between NCGS and IBS.

**Figure S3 (mSphere00856-25-S0008.tiff).** Diversity and differential abundance for metagenome features in NCGS and NCGS\_PG.

**Figure S4 (mSphere00856-25-S0009.tiff).** Principal coordinate analysis for metabolite profiles.

**Supplemental legends (mSphere00856-25-S0010.docx).** Legends for Fig. S1–S4 and Data S1–S5.

## REFERENCES

- Molina - Infante J, Santolaria S, Sanders DS, Fernández - Bañares F. 2015. Systematic review: noncoeliac gluten sensitivity. *Aliment Pharmacol Ther* 41:807–820. <https://doi.org/10.1111/apt.13155>
- Cárdenas-Torres FI, Cabrera-Chávez F, Figueroa-Salcido OG, Ontiveros N. 2021. Non-celiac gluten sensitivity: an update. *Medicina (Kaunas)* 57:526. <https://doi.org/10.3390/medicina57060526>
- Catassi C, Bai JC, Bonaz B, Bouma G, Calabrò A, Carroccio A, Castillejo G, Ciacci C, Cristofori F, Dolinsek J, et al. 2013. Non-celiac gluten sensitivity: the new frontier of gluten related disorders. *Nutrients* 5:3839–3853. <https://doi.org/10.3390/nu5103839>
- Sapone A, Bai JC, Ciacci C, Dolinsek J, Green PHR, Hadjivassiliou M, Kaukinen K, Rostami K, Sanders DS, Schumann M, Ullrich R, Villalta D, Volta U, Catassi C, Fasano A. 2012. Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC Med* 10:13. <https://doi.org/10.1186/1741-7015-10-13>
- Shiha MG, Manza F, Figueroa-Salcido OG, Ontiveros N, Caio G, Jansson-Knodell CL, Rubio-Tapia A, Aziz I, Sanders DS. 2025. Global prevalence of self-reported non-coeliac gluten and wheat sensitivity: a systematic review and meta-analysis. *Gut* 75:502–510. <https://doi.org/10.1136/gutjnl-2025-336304>
- Elli L, Branchi F, Tomba C, Villalta D, Norsa L, Ferretti F, Roncoroni L, Bardella MT. 2015. Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. *World J Gastroenterol* 21:7110–7119. <https://doi.org/10.3748/wjg.v21.i23.7110>
- Vervier K, Moss S, Kumar N, Adoum A, Barne M, Browne H, Kaser A, Kiely CJ, Neville BA, Powell N, Raine T, Stares MD, Zhu A, De La Revilla Negro J, Lawley TD, Parkes M. 2022. Two microbiota subtypes identified in irritable bowel syndrome with distinct responses to the low FODMAP diet. *Gut* 71:1821–1830. <https://doi.org/10.1136/gutjnl-2021-325177>
- Sarkar D, Roy P, Saha S. 2026. Meta-analysis of faecal microbiome studies followed by machine learning to identify intestinal disease-specific taxonomic signatures. *Microb Pathog* 211:108221. <https://doi.org/10.1016/j.micpath.2025.108221>
- Jacobs JP, Lagishetty V, Hauer MC, Labus JS, Dong TS, Toma R, Vuyisich M, Naliboff BD, Lackner JM, Gupta A, Tillisch K, Mayer EA. 2023. Multi-omics profiles of the intestinal microbiome in irritable bowel syndrome and its bowel habit subtypes. *Microbiome* 11:5. <https://doi.org/10.1186/s40168-022-01450-5>
- Kim GH, Lee K, Shim JO. 2023. Gut bacterial dysbiosis in irritable bowel syndrome: a case-control study and a cross-cohort analysis using publicly available data sets. *Microbiol Spectr* 11:e02125-22. <https://doi.org/10.1128/spectrum.02125-22>
- Ahmed A, Dixit K, Singh A, Agarwal A, Mehtab W, Prasad S, Rajput MS, Chauhan A, Agarwal A, Mehta S, Ahuja V, Shouche Y, Dhotre D, Makharia GK. 2024. Sieving out non-celiac gluten sensitivity amongst patients with irritable bowel syndrome. *Dig Liver Dis* 56:451–457. <https://doi.org/10.1016/j.dld.2023.10.014>
- Wei G, Helmerhorst EJ, Darwish G, Blumenkranz G, Schuppan D. 2020. Gluten degrading enzymes for treatment of celiac disease. *Nutrients* 12:2095. <https://doi.org/10.3390/nu12072095>
- Bodkhe R, Shetty SA, Dhotre DP, Verma AK, Bhatia K, Mishra A, Kaur G, Pande P, Bangarusamy DK, Santosh BP, Perumal RC, Ahuja V, Shouche YS, Makharia GK. 2019. Comparison of small gut and whole gut microbiota of first-degree relatives with adult celiac disease patients and controls. *Front Microbiol* 10:164. <https://doi.org/10.3389/fmicb.2019.00164>
- Busi SB, Bourquin M, Fodelianakis S, Michoud G, Kohler TJ, Peter H, Pramateftaki P, Styllas M, Tolosano M, De Staercke V, Schön M, de Nies L, Marasco R, Daffonchio D, Ezzat L, Wilmes P, Battin TJ. 2022. Genomic and metabolic adaptations of biofilms to ecological windows of opportunity in glacier-fed streams. *Nat Commun* 13:2168. <https://doi.org/10.1038/s41467-022-29914-0>
- Skodje GI, Sarna VK, Minelle IH, Rolfsen KL, Muir JG, Gibson PR, Veierød MB, Henriksen C, Lundin KEA. 2018. Fructan, rather than gluten, induces symptoms in patients with self-reported non-celiac gluten sensitivity. *Gastroenterology* 154:529–539. <https://doi.org/10.1053/j.gastro.2017.10.040>
- Rej A, Sanders DS. 2019. The overlap of irritable bowel syndrome and noncoeliac gluten sensitivity. *Curr Opin Gastroenterol* 35:199–205. <https://doi.org/10.1097/MOG.0000000000000517>

17. Schuppan D, Junker Y, Barisani D. 2009. Celiac disease: from pathogenesis to novel therapies. *Gastroenterology* 137:1912–1933. <https://doi.org/10.1053/j.gastro.2009.09.008>
18. Chang FY, Lu CL, Chen TS. 2010. The current prevalence of irritable bowel syndrome in Asia. *J Neurogastroenterol Motil* 16:389–400. <https://doi.org/10.5056/jnm.2010.16.4.389>
19. Dixit K, Ahmed A, Singh A, Inamdar M, Chavan S, Bodkhe R, Mehtab W, Chauhan A, Saroj SD, Ahuja V, Shouche Y, Dhotre D, Makharia G. 2024. Site-specific gut microbial signatures in non-celiac gluten sensitivity. *Gut Microbes Reports* 1:2438621. <https://doi.org/10.1080/29933935.2024.2438621>
20. Mohammadzadeh R, Mahnert A, Duller S, Moissl-Eichinger C. 2022. Archaeal key-residents within the human microbiome: characteristics, interactions and involvement in health and disease. *Curr Opin Microbiol* 67:102146. <https://doi.org/10.1016/j.mib.2022.102146>
21. Bin Waqar SH, Rehan A. 2019. Methane and constipation-predominant irritable bowel syndrome: entwining pillars of emerging neurogastroenterology. *Cureus* 11:e4764. <https://doi.org/10.7759/cureus.4764>
22. Steck N, Mueller K, Schemann M, Haller D. 2012. Bacterial proteases in IBD and IBS. *Gut* 61:1610–1618. <https://doi.org/10.1136/gutjnl-2011-300775>
23. Kóiv V, Tenson T. 2021. Gluten-degrading bacteria: availability and applications. *Appl Microbiol Biotechnol* 105:3045–3059. <https://doi.org/10.1007/s00253-021-11263-5>
24. Dieterich W, Schuppan D, Schink M, Schwappacher R, Wirtz S, Agaimy A, Neurath MF, Zopf Y. 2019. Influence of low FODMAP and gluten-free diets on disease activity and intestinal microbiota in patients with non-celiac gluten sensitivity. *Clin Nutr* 38:697–707. <https://doi.org/10.1016/j.clnu.2018.03.017>
25. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. 2019. *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci USA* 116:12672–12677. <https://doi.org/10.1073/pnas.1904099116>
26. Vaziri ND, Yuan J, Norris K. 2013. Role of urea in intestinal barrier dysfunction and disruption of epithelial tight junction in chronic kidney disease. *Am J Nephrol* 37:1–6. <https://doi.org/10.1159/000345969>
27. Garcia-Mazcorro JF, Rivera-Gutierrez X, Cobos-Quevedo ODJ, Grube-Pagola P, Meixueiro-Daza A, Hernandez-Flores K, Cabrera-Jorge FJ, Vivanco-Cid H, Dowd SE, Remes-Troche JM. 2018. First insights into the gut microbiota of Mexican patients with celiac disease and non-celiac gluten sensitivity. *Nutrients* 10:1641. <https://doi.org/10.3390/nu10111641>
28. Catassi C, Elli L, Bonaz B, Bouma G, Carroccio A, Castillejo G, Cellier C, Cristofori F, de Magistris L, Dolinsek J, et al. 2015. Diagnosis of Non-Celiac Gluten Sensitivity (NCGS): the salerno experts' criteria. *Nutrients* 7:4966–4977. <https://doi.org/10.3390/nu7064966>
29. Narayanasamy S, Jarosz Y, Muller EEL, Heintz-Buschart A, Herold M, Kaysen A, Laczny CC, Pintel N, May P, Wilmes P. 2016. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol* 17:260. <https://doi.org/10.1186/s13059-016-1116-8>
30. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
31. Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>
32. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>
33. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
34. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btzb48>
35. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Res* 49:D412–D419. <https://doi.org/10.1093/nar/gkaa913>
36. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>
37. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 44:D471–D480. <https://doi.org/10.1093/nar/gkv1164>
38. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>
39. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>
40. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
41. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2020. Community-led, integrated, reproducible multi-omics with anvio. *Nat Microbiol* 6:3–6. <https://doi.org/10.1038/s41564-020-00834-3>
42. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvio: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>
43. De Saedeleer B, Malabirade A, Ramiro-Garcia J, Habier J, Trezzi J-P, Peters SL, Daujeumont A, Halder R, Jäger C, Busi SB, May P, Oertel W, Mollenhauer B, Laczny CC, Hettich RL, Wilmes P. 2021. Systematic characterization of human gut microbiome-secreted molecules by integrated multi-omics. *ISME Commun* 1:82. <https://doi.org/10.1038/s43705-021-00078-0>
44. McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
45. Wickham H. 2016. *Ggplot2*. Springer International Publishing, Cham.
46. Kassambara A. 2023. *Ggpubr: "ggplot2" based publication ready plots*. Available from: <https://cran.r-project.org/web/packages/ggpubr/index.html>
47. Paradis E, Blomberg S, Bolker B. 2024. *ape: analyses of phylogenetics and evolution*. Available from: <https://cran.r-project.org/web/packages/ape/index.html>
48. Oksanen J, Simpson GL, Blanchet FG, et al. 2025. *Vegan: community ecology package*. Available from: <https://cran.r-project.org/web/packages/vegan/index.html>
49. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
50. Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C, Spigelman AF, MacDonald PE, Wishart DS, Li S, Xia J. 2024. *MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation*. *Nucleic Acids Res* 52:W398–W406. <https://doi.org/10.1093/nar/gkac253>