

Bacterial dysbiosis predicts the diagnosis of Crohn's disease in Saudi children

Mohammad I. El Mouzan, Harland S. Winter¹, Ahmed A. Al Sarkhy, Kirill Korolev², Rajita Menon³, Asaad A. Assiri⁴

Department of Pediatrics, Gastroenterology Division, King Saud University, ⁴Department of Pediatrics, Gastroenterology Division, Supervisor, Prince Abdullah Bin Khalid Celiac Disease Research Chair, King Saud University, Riyadh, Kingdom of Saudi Arabia, ¹Mass General Hospital for Children, Gastroenterology, MA, ²Department of Physics, Bioinformatics Program, Boston University, MA, ³Department of Physics, Boston University, Boston, MA, Boston, USA

Abstract

Background: Studies have reached different conclusions regarding the accuracy of dysbiosis in predicting the diagnosis of Crohn's disease (CD). The aim of this report is to assess the utility of mucosal and fecal microbial dysbiosis as predictors in the diagnosis of this condition in Saudi children.

Methods: Tissue and fecal samples were collected prospectively from children with final diagnosis of CD and from controls. Bacterial DNA was extracted and sequenced using Illumina MiSeq chemistry. The abundance and diversity of bacteria in tissue and fecal samples were determined in relation to controls. Sparse logistic regression was calculated to predict the diagnosis of CD based on subject's microbiota profile.

Results: There were 17 children with CD and 18 controls. All children were Saudis. The median age was 13.9 and 16.3 years for children with CD and controls respectively. Sex distribution showed that 11/17 (65%) of the CD and 12/18 (67%) of the control subjects were boys. The mean area under the curve (AUC) was significantly higher in stool (AUC = 0.97 ± 0.029) than in tissue samples (AUC = 0.83 ± 0.055) ($P < 0.001$).

Conclusions: We found high AUC in mucosal and fecal samples. The higher AUC for fecal samples suggests higher accuracy in predicting the diagnosis of CD.

Keywords: Crohn's disease, microbiota, Saudi children

Address for correspondence: Prof. Mohammad I. El Mouzan, Department of Pediatrics, King Saud University, P O Box 2925, Riyadh 11461, Kingdom of Saudi Arabia.

E-mail: melmouzan@ksu.edu.SA

Submitted: 19-Jul-2020 **Revised:** 05-Oct-2020 **Accepted:** 23-Oct-2020 **Published:** 09-Feb-2021

INTRODUCTION

Inflammatory bowel diseases (IBD) are chronic conditions mainly represented by Crohn's disease (CD) and ulcerative colitis (UC). The incidence of the disease is highest in Western populations, with an upward trend observed in recent years.^[1-4] Although IBD is less common elsewhere, its incidence is increasing in developing countries.^[4,5] The

etiology of IBD is not known but believed to be multifactorial. Dietary components and microbial imbalance may trigger inflammation.^[6-9] Experimental and clinical data provide strong evidence of the role of bacterial microbiota in IBD and pro-inflammatory and anti-inflammatory taxa have been described.^[10] Bacterial dysbiosis in mucosa and stool was reported as predictor in the diagnosis of IBD including

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

How to cite this article: El Mouzan MI, Winter HS, Al Sarkhy AA, Korolev K, Menon R, Assiri AA. Bacterial dysbiosis predicts the diagnosis of Crohn's disease in Saudi children. Saudi J Gastroenterol 2021;27:144-8.

Access this article online	
Quick Response Code:	Website: www.saudijgastro.com
	DOI: 10.4103/sjg.SJG_409_20

CD but with conflicting results on the relative accuracy of different samples. These reports are both limited in number and are based on Western populations.^[11-13] Differences in culture and dietary lifestyle may affect microbiota composition and structure in non-Western populations, where IBD may be considered a new condition. Therefore, it is important to include data from non-Western populations in microbiome studies. The objective of this report, therefore, is to assess the accuracy of bacterial dysbiosis in mucosa and stools in predicting the diagnosis of CD in newly diagnosed Saudi children, a non-Western population.

METHODS

Study population, data collection and processing

In this prospective study, children were recruited from January 2012 to October 2014, from King Saud University, King Khalid University Hospital, providing primary, secondary and tertiary level free of charge services and one private gastroenterology clinic providing diagnostic and therapeutic services for a fee. Both institutions are in Riyadh, Kingdom of Saudi Arabia. The children presenting to the gastroenterology clinics with complaints compatible with the diagnosis of CD were considered but were recruited after the diagnosis of CD was confirmed.

Inclusion criteria for children with CD included age below 18 years, Saudi nationality, no history of antibiotic intake at least 6 months before sample collection, and confirmation of the diagnosis of CD. *Inclusion criteria for controls* included age below 18 years, Saudi nationality, no history of antibiotic intake at least 6 months before sample collection, and lack of clinical, laboratory or endoscopic features of CD, UC, IBD-unclassified, infections, or other inflammatory diseases. *Exclusion criteria for children with CD and controls* included age above 18 years, non-Saudi nationals, children with the diagnosis of UC, IBD-unclassified, inflammation, infection, co-morbidities, or antibiotic intake at least 6 months before sample collection. *Diagnostic procedures* included a history of symptoms in the form of, but not limited to, abdominal pain, diarrhea with or without blood, weight loss, perianal disease; laboratory, imaging, endoscopic and histopathologic investigations.

Confirmation of the diagnosis of CD is made according to the guidelines published by the European Society for Pediatric Gastroenterology Hepatology and Nutrition.^[14] The main final diagnoses in controls were recurrent abdominal pain and polyps. There were 44 mucosal tissue samples obtained from the children with CD (8 from the Ileum and 6 from each part of the colon); whereas, 14 mucosal tissue samples were collected from controls including

6 from the Ileum, 3 from the transverse colon, 2 from the sigmoid colon, and 3 from the rectum. For various practical reasons, mucosal samples were not obtained from all colonic segments of all subjects. Similarly, stool samples were not available from all participants and only 20 samples were given (10 from children with CD and 10 from controls). To minimize the wash out effect of bowel preparation required for colonoscopy, stools were collected in the clinic or the hospital ward before bowel preparation in the majority of children (about 75%), but when this was not possible (about 25%), a sample was taken from the first bowel motion after starting bowel cleaning. Antibiotic intake was not documented in any subject in this report. Tissue and fecal samples were put in cryotubes, not containing fixatives or stabilizers, and transported in ice to the research center, generally reaching within 5 to 20 minutes after dispatch, where they were stored at -80°C. At the end of the study, all samples were dispatched in dry ice container to the USA for analysis.

Bacterial DNA extraction and sequencing

Bacterial DNA extraction was performed using the Mobio Powersoil Kit. Amplicon pyrosequencing (bTEFAP®) was used and the primers 515F GTGCCAGCMGCCGCGGTAA and 806R GGACTACHVGGGTWTCTAAT were utilized to assess the microbiome using the Illumina MiSeq with a methodology based upon the bTEFAP® following manufacturer's protocol, producing sequences deep enough for comparison purposes.^[15,16] Subsequently, paired sequences were merged and depleted of barcodes and primers. Short sequences <150 base pairs (bp), those with ambiguous base calls, those with homopolymer runs exceeding 6bp, and chimeras were removed. Operational taxonomic units (OTUs) were defined after removal of singleton sequences, clustering at 3% divergence (97% similarity).^[17,18] OTUs were then taxonomically classified using the Nucleotide Basic Local Alignment Search Tool against a 16S National Center for Biotechnology Information (NCBI)-derived database (www.ncbi.nlm.nih.gov, <http://rdp.cme.msu.edu>).^[18,19]

Statistics

Python and R packages were used.^[20-22] All statistical analyses were performed on log-transformed data after adding pseudo counts of 1 read for each taxonomic group. Therefore, the reported means are exponentials of the mean log abundances (they are geometric rather than arithmetic means with pseudo counts included). Exact Fisher's test was used to assess statistical significance, followed by a correction for multiple hypothesis testing using the

Benjamini-Hochberg procedure.^[23] The associations were considered statistically significant when FDR-corrected *P* values (q value) were less than 0.05.

The bacterial community alpha diversity was assessed by the Shannon index and beta diversity was measured by Bray-Curtis distance. We used the Fisher's exact test to determine *P* values for alpha diversity and an exact permutation test to determine *P* value for beta diversity separations. The significance level of variation between mucosa and stool was defined as a q value <0.05.

The bacterial dysbiosis score, the metric used by the logistic regression classifier for discriminating the samples of controls from those of children with CD, was calculated. First, we performed a logistic regression to infer sparse linear coefficients for bacterial taxa for each subject. The dysbiosis score was defined by the weighted sum of the taxa abundances, with the weights learned from the logistic regression.

The accuracy of the dysbiosis classifier was tested by computing the receiver operating characteristic (ROC) curve with 5-fold stratified cross validation under 100 permutations of the training data partition. ROC curves show the trade-off between true positive rate, also known as sensitivity, vs. false positive rate which is 1-specificity for every possible cut-off for the logistic regression classification test. Sparse logistic regression was performed to predict the diagnosis of Crohn's disease status based on participant's bacterial microbiota. The accuracy was tested by calculating the area under the curve (AUC) of the receiver operating characteristic.

Ethics

The protocol of this study was evaluated and approved by the Institutional Board Review of the College of Medicine, King Saud University in Riyadh, Kingdom of Saudi Arabia [No: 10/2647/IRB, 26/6/2010]. Guardians and/or children signed informed consent and/or assent before enrollment in the study.

RESULTS

Demographic and clinical characteristics

Comparison of the overall characteristics of the children with CD and controls are detailed in Table 1. Briefly, there were 17 children with Crohn's disease and 18 non-IBD controls. The median age was 15 and 16.3 years for children with CD and controls respectively. Sex distribution showed that 11/17 (65%) of the children with CD and 12/18 (67%) of the non-IBD-controls were males.

Bacterial microbiota structure in Saudi children with CD in relation to controls has been reported recently defining significantly- abundant and -depleted bacteria at all levels as well as reduced diversity.^[24]

Bacterial dysbiosis classification

The AUC for the bacterial dysbiosis classifier is shown in Figure 1, which indicates higher AUC in stools (0.97 ± 0.029) than in mucosa (0.83 ± 0.056) ($P < 0.001$).

DISCUSSION

Microbial dysbiosis can be defined as a change of microbial community in disease, relative to subjects with no signs of disease.^[25] Variation of mucosal microbial structure across the gastrointestinal tract as well as between fecal and mucosal samples have been reported in the literature.^[26,27] Therefore, in this study mucosal tissue samples were taken from the ileum, colon, and stools, and results are reported together.

To our knowledge, this is the first report to address the role of bacterial dysbiosis in the diagnosis of CD from a non-Western population, who have a different culture and dietary lifestyle. All the participants in this study were newly diagnosed and proved to have dysbiosis.^[24] In addition, the fact that they did not receive any treatment before sample collection suggests that dysbiosis may be accurate in predicting the diagnosis of CD.

The most important finding of this study is the high AUC for bacterial dysbiosis in our children with CD with significantly higher AUC in stool ($AUC = 0.97 \pm 0.029$) than mucosal tissue samples ($AUC = 0.83 \pm 0.055$) ($P < 0.001$), a finding consistent with a report of an AUC of 0.91 in stool samples from patients with active IBD.^[12] In contrast, our findings conflict with the results of a larger study reporting lower dysbiosis in fecal (AUC of 0.66) than in mucosal samples.^[11] Although this conclusion was highlighted by Hofer,^[28] a re-analysis of the same dataset of Gevers *et al.*

Table 1: Demographic and clinical characteristics

Items	Children with CD	Controls
Number	17	18
Age (yrs): median (range)	15 (7.3-17.8)	16.3 (3.9-18.6)
Gender: No (% males)	11 (65)	12 (67)
Positive consanguinity	6/15 (40%)	6/17 (35%)
CD location:		
Ileal (L1)	2 (11.8%)	Not applicable
Colonic (L2)	2 (11.8%)	Not applicable
Ileocolonic (L3)	13 (76.4%)	Not applicable
CD behavior:		
Non-constricting, nonpenetrating (B1)	11 (64%)	Not applicable
Constricting (B2)	3 (18%)	Not applicable
Penetrating (B3)	3 (18%)	Not applicable

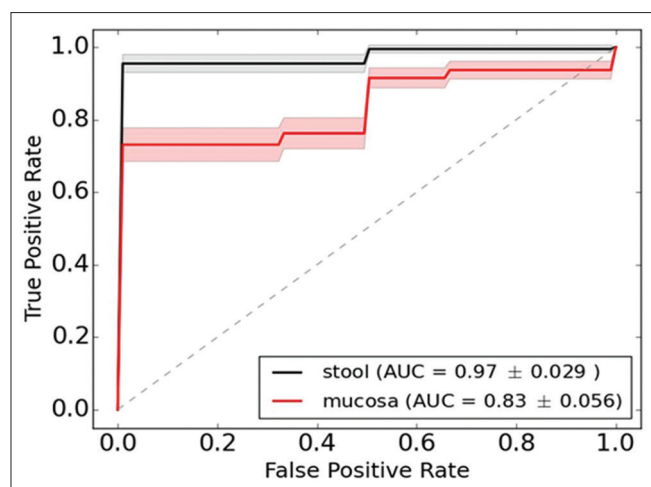


Figure 1: Receiver Operating Characteristic (ROC) curve for the logistic regression dysbiosis classifier. The mean ROC curve for dysbiosis in stool (solid black line) and in mucosa samples (solid red line). The standard deviation is shown in gray and light red shading. The area under the curve for bacterial dysbiosis is significantly higher in stool than in mucosa, with a P value of < 0.001

using another statistical approach (log transformation), similar to statistical methods used in our study, revealed higher mean AUC for stool dysbiosis approaching clinically useful values (AUC of 0.72 (95% CI: 0.663–0.770)).^[13] This higher value of AUC (0.72 vs 0.66) underscores the importance of analytical methodology. The discrepancy between prior research findings and our results of showing higher AUC in stool than in mucosal samples, could be related to our small sample size, possibly augmented by the role of lifestyle difference in this population affecting microbiota profile, differences in mucosal location, and to log transformation statistical methodology used in the present study. Since this is the first report from a non-Western population and the third report worldwide, further studies from different populations are needed to clarify the role of bacterial dysbiosis in predicting the diagnosis of CD.

The clinical implication of the results of this study, as well as previous dysbiosis studies,^[11,12] is limited at present in view of the high cost of microbiome analysis. However, these studies are important from a scientific point of view to stimulate further research in this direction, hopefully leading to the discovery of additional dysbiosis non-invasive screening tests for CD.

The main limitation of this study is the small number of participants. However, Crohn's disease is a new condition in developing countries and the fact that this is the first prediction study from Saudi Arabia, a non-Western population, makes it of interest for the understanding of the importance of microbiota in Crohn's disease.

CONCLUSIONS

We found high AUC in mucosal and fecal samples. The higher AUC for fecal samples suggests higher accuracy in predicting the diagnosis of CD. Although, the clinical application of this finding is limited at present by the high cost of microbiota analysis, these results are expected to stimulate future research to clarify the role of microbiota in CD, leading to the development of a non-invasive dysbiosis test.

Acknowledgement

The authors thank the Deanship of Scientific Research at King Saud University in Riyadh, Kingdom of Saudi Arabia for funding this work through Research Group No (RGP-1441-007).

Financial support and sponsorship

Funded by the Deanship of Scientific Research; Grant no. RG-1441-007.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Shivashankar R, Tremaine WJ, Harmsen WS, Loftus EV Jr. Incidence and prevalence of Crohn's disease and ulcerative colitis in Olmsted County, Minnesota from 1970 through 2010. *Clin Gastroenterol Hepatol* 2017;15:857-63.
2. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012;142:46-54.
3. Benchimol EI, Fortinsky KJ, Gozdyra P, den Heuvel M, Limbergen JV, Griffiths AM. Epidemiology of pediatric inflammatory bowel disease: A systematic review of international trends. *Inflamm Bowel Dis* 2011;17:423-39.
4. Duricova D, Fumery M, Annese V, Lakatos PL, Peyrin-Biroulet L, Gower-Rousseau C. The natural history of Crohn's disease in children: A review of population-based studies. *Eur J Gastroenterol Hepatol* 2017;29:125-34.
5. El Mouzan MI, Saadah O, Al-Saleem K, Al Edreesi M, Hasosah M, Alanazi A, et al. Incidence of pediatric inflammatory bowel disease in Saudi Arabia: A multicenter national study. *Inflamm Bowel Dis* 2014;20:1085-90.
6. Chapman-Kiddell CA, Davies PSW, Gillen L, Radford-Smith GL. Role of diet in the development of inflammatory bowel disease. *Inflamm Bowel Dis* 2010;16:137-51.
7. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334:105-8.
8. Leone V, Chang EB, Devkota S. Diet, microbes, and host genetics: The perfect storm in inflammatory bowel diseases. *J Gastroenterol* 2013;48:315-21.
9. Robinson AM, Gondalia SV, Karpe AV, Eri R, Beale DJ, Morrison PD, et al. Fecal microbiota and metabolome in a mouse model of spontaneous chronic colitis: Relevance to human inflammatory bowel disease. *Inflamm Bowel Dis* 2016;22:2767-87.
10. Dasgupta S, Kasper DL. Relevance of commensal microbiota in the

- treatment and prevention of inflammatory bowel disease. *Inflamm Bowel Dis* 2013;19:2478-89.
11. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, *et al.* The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 2014;15:382-92.
 12. Papa E, Docktor M, Smillie C, Weber S, Preheim SP, Gevers D, *et al.* Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS One* 2012;7:e39242.
 13. Wang F, Kaplan JL, Gold BD, Bhasin MK, Ward NL, Kellermayer R, *et al.* Detecting microbial dysbiosis associated with pediatric Crohn disease despite the high variability of the gut microbiota. *Cell Rep* 2016;14:945-55.
 14. Levine A, Koletzko S, Turner D, Escher JC, Cucchiara S, de Ridder L, *et al.*; European Society of Pediatric Gastroenterology, Hepatology, and Nutrition. ESPGHAN revised Porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr* 2014;58:795-806.
 15. Dowd SE, Callaway TR, Wolcott RD, Sun Y, McKeehan T, Hagevoort RG, *et al.* Evaluation of the bacterial diversity in the feces of cattle using 16S rDNA bacterial tag-encoded FLX amplicon pyrosequencing [bTEFAP]. *BMC Microbiol* 2008;8:125.
 16. Dowd SE, Sun Y, Wolcott RD, Domingo A, Carroll JA. Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) for microbiome studies: Bacterial diversity in the ileum of newly weaned Salmonella-infected pigs. *Foodborne Pathog Dis* 2008;5:459-72.
 17. Eren AM, Zozaya M, Taylor CM, Dowd SE, Martin DH, Ferris MJ. Exploring the diversity of *Gardnerella vaginalis* in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. *PLoS One* 2011;6:e26732.
 18. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460-1.
 19. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, *et al.* Greengenes, a chimera checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72:5069-72.
 20. R Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2015.
 21. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
 22. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
 23. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodological)* 1995;57:289-300.
 24. El Mouzan MI, Winter HS, Assiri AA, Korolev KS, Al Sarkhy AA, Dowd SE, *et al.* Microbiota profile in new-onset pediatric Crohn's disease: Data from a non-Western population. *Gut Pathog* 2018;10:49.
 25. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cell Microbiol* 2014;16:1024-33.
 26. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 2016;14:20-32.
 27. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, de Vos WM. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol* 2002;68:3401-7.
 28. Hofer U. Bacterial imbalance in Crohn's disease. *Nat Rev Microbiol* 2014;12:312.