90755_Auto_Edited.docx

Observational Study

The Gut Microbiota Predicts the Diagnosis of Ulcerative Colitis in Saudi Children

Mohammad El Mouzan, Ahmed Al Sarkhy, Asaad Assiri

Abstract

BACKGROUND

Ulcerative colitis (UC) is an immune-mediated chronic inflammatory condition with a worldwide distribution. Although the etiology of this disease is still unknown, the understanding of the role of the microbiota is becoming increasingly strong.

AIM

To investigate the predictive power of the gut microbiota for the diagnosis of UC in a cohort of newly diagnosed treatment-naïve Saudi children with UC.

METHODS

The study population included 20 children with a confirmed diagnosis of UC and 20 healthy controls. Microbial DNA was extracted and sequenced, and shotgun metagenomic analysis was performed for bacteria and bacteriophages. Biostatistics and bioinformatics demonstrated significant dysbiosis in the form of reduced alpha diversity, beta diversity, and significant difference of abundance of taxa between children with UC and control groups. The receiver operating characteristic (ROC) curve, a probability curve, was used to determine the difference between the UC and control groups. The area under the curve (AUC) represents the degree of separability between the UC group and the control group. The AUC was calculated for all identified

bacterial species and for bacterial species identified by the random forest classification algorithm as important potential biomarkers of UC. A similar method of AUC calculation for all bacteriophages and important species was used for bacteriophages.

RESULTS

The median age and range were 14 (0.5-21) and 12.9 (6.8-16.3) years for children with UC and controls, respectively, and 40% and 35% were male for children with UC and controls, respectively. The AUC for all identified bacterial species was 89.5%. However, when using the bacterial species identified as important by random forest classification algorithm analysis, the accuracy increased to 97.6%. Similarly, the AUC for all the identified bacteriophages was 87.4%, but this value increased to 94.5% when the important bacteriophage biomarkers were used.

CONCLUSION

The very high to excellent AUCs of fecal bacterial and viral species suggest the use of a potential noninvasive microbiota-based test for the diagnosis of unusual cases of UC in children. In addition, the identification of important bacteria and bacteriophages whose abundance is reduced in children with UC suggests the potential of preventive and adjuvant microbial therapy for UC.

INTRODUCTION

Ulcerative colitis (UC) is an immune-mediated inflammatory bowel disease (IBD). Although the incidence of this disease is highest in Western populations, it is increasing globally [1-3]. The etiology of UC is unknown; however, multifactorial factors involving interactions between genetics, host immunity, the mucosal barrier, and the gut microbiome are highly suspected [4-6]. The role of the microbiota has been extensively reported mainly in Western populations, with strong evidence of an association with UC.

In Saudi Arabia, a developing country in transition, the incidence and clinical patterns of UC have been reported [7-10]. In addition, the microbiota profile of Saudi children with Crohn's disease has been reported to be significantly associated with not only the presence of bacteria but also the high AUC for bacteria in fecal samples, suggesting high accuracy in predicting the diagnosis of CD [11,12]. However, there are no reports on the predictive power of the microbiota for the diagnosis of UC. The objective of this study was to evaluate the role of the microbiota in predicting the diagnosis of UC in Saudi children.

MATERIALS AND METHODS

Study population: Children with a confirmed diagnosis of UC were enrolled in the study. The children were recruited from multiple hospitals in Riyadh, Kingdom of Saudi Arabia (KSA). The inclusion criteria included new-onset and untreated disease, as well as no antibiotic exposure for at least 6 months before stool collection. Fecal samples from the children with UC were collected before bowl preparation. Healthy school children were randomly selected as controls. Stool samples from children with UC and controls were collected in cryovials without fixatives or stabilizers and immediately stored at −80 °C until analysis.

DNA extraction and sequencing: Bacterial and viral DNA from fecal samples was isolated using the QIAGEN DNeasy PowerSoil Pro Kit according to the manufacturer's protocol. DNA libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina) and IDT Unique Dual Indexes with a total DNA input of 1ng. Library were subsequently sequenced on an Illumina NovaSeq S4 platform .

Statistical and bioinformatics analysis: Shannon alpha diversity metrics were calculated in R using the R package "vegan". Wilcoxon rank-sum tests were performed between groups using the R package ggsignif [13,14]. Bray−Curtis dissimilarity was calculated in R using the vegan package with the function vegdist, and PCoA tables were generated using the ape function pcoa. PERMANOVA tests for each distance matrix were generated using the vegan's6 function adonis2, and beta dispersion was calculated and

compared using the ANOVA method for the betadispering function from vegan [15]. DESeq2 was used to estimate differential abundance between cohorts based on count data [16]. The random forest classification algorithm was applied to the relative abundance data to predict bacterial and viral species biomarkers that might improve prediction [17].

The receiver operating characteristic (ROC) curve was used to determine the difference between the UC and control groups. The area under the curve (AUC) represents the degree of separability between the UC group and the control group. The AUC was calculated for all identified bacterial and bacteriophage species in this study and for bacterial and bacteriophage species identified by the random forest classification algorithm as important potential biomarkers of UC [18].

Ethical aspects: The study was 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

The median age and range were 14 (0.5-21) and 12.9 (6.8-16.3) years for children with UC and controls, respectively, and 40% and 35% were male for children with UC and controls, respectively. A high number of significant bacterial and bacteriophage dysbiosis events were found (unpublished data). Among these, 11 bacterial species biomarkers were identified. These included the *Bifidobacterium angulatum*, *Alistipes putredinis*, *Bacteroides caccae, and Bifidobacterium adolescentis*. (Table 1). Similarly, among the high number of bacteriophages, four were identified as biomarkers. These included the *Salmonella phage SEN4*, *Streptococcus phage YMC-2011*, and *uncultured crAssphage* (Table 2).

The AUC for all identified bacterial species was 89.5% (79.1%−100.0%) , but when based on the biomarkers, the accuracy increased to 97.6% (94.2%-100.0%) indicating very good to excellent predictive power (Figure 1 A, B). Similarly, the AUC for all the identified

bacteriophages was 87.4% (75.9%−98.8%), but the AUC increased to 94.5 % (87.8%-100%), when the identified important species were used, indicating very good to excellent predictive power (Figure 2 A, B).

DISCUSSION

Shotgun metagenomic analysis of bacterial and viral bacteriophage species in fecal samples of children with new-onset untreated UC revealed significant differential abundances between the UC group and the control group, indicating significant dysbiosis (unpublished data). The AUC of the ROC curve represents the degree of separability between the UC group and the control group, indicating the predictive power of the ROC curve for UC diagnosis.

In this study, we calculated the AUC based not only on the entire bacterial species and bacteriophages but also on important species identified by the random forest classification algorithm. The calculated AUC based on the abundance of all the bacterial species increased from 89.5% to 97.6% when only 11 bacterial species biomarkers were considered, indicating increased predictive power of the important bacterial species biomarkers. Similarly, the AUC calculated based on the bacteriophages increased from 87.4% to 94.5% when only four biomarkers were considered, indicating that the use of these bacteriophage biomarkers has greater predictive power for distinguishing UC patients from controls. The excellent predictive power of these biomarkers indicates the potential for the development of microbiota-based diagnostic tests. Among the bacteria and bacteriophages, *Bifidobacterium angulatum* and *uncultured crAssphage* had the highest median importance scores. *Bifidobacterium angulatum* is a species that belongs to the Bifidobacterium genus that is known to modulate the immune system and may be considered protective against UC [19,20] . *Uncultured crAssphage* is the most abundant human-associated virus and is found in the gut virome in approximately 50% of humans. This virus infects species of Bacteroides with mostly beneficial effects on health. Accordingly, *Bifidobacterium angulatum* and uncultured *crAssphage* could constitute the basis of prophylactic or therapeutic options [21-24].

The excellent predictive diagnostic power for UC in this report is slightly greater but consistent with the 93% accuracy for UC diagnosis reported within a multiclass disease in an adult study in Hong Kong [25] and the 91% accuracy in a group of children with UC in which shotgun metagenomic bacterial species-level abundance was used [26]. Finally, the 84.4% to 95% predictive power of the bacteriophage species identified in this study has not been reported thus far and deserves further study.

Study limitations: This study had a relatively small sample size, but it may be acceptable for this is the first study to use metagenomic analysis in a non-Western childhood population to determine the accuracy of the microbiota in predicting the diagnosis of UC.

CONCLUSION

The very high to excellent AUCs of fecal bacterial and viral species indicate the potential for the development of noninvasive microbiota-based tests for the diagnosis of UC and for preventive and adjuvant microbial therapy for UC. In addition, the identification of important bacteria and bacteriophages whose abundance is reduced in children with UC suggests the potential of preventive and adjuvant microbial therapy.

90755_Auto_Edited.docx

ORIGINALITY REPORT

3%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	15 words — 1%
2	gutpathogens.biomedcentral.com Internet	13 words — 1%
3	www.ecco-ibd.eu Internet	13 words — 1%
4	Feng Zhou, Guo Dong Zhang, Yang Tan, Shi An Hu, Qun Tang, Gang Pei. "NOD-like receptors mediate homeostatic intestinal epithelial barrier function: promising therapeutic targets for inflammatory bowel disease", Therapeutic Advances in Gastroenterology, 2023 Crossref	12 words — 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS