TITLE: A Co-abundant Gene Analysis for Identification of Gut Microbial Strains Associated with Immune Responses to Oral Cholera Vaccine

AUTHORS: Fred J. Heller[a], Hasan Al Banna[b], M. Hasanul Kaisar[b], Denise Chac[c], Fahima Chowdhury[b], Taufiqur R. Bhuiyan[b], Afroza Akter[b], Ashraful I. Khan[b], Polash Chandra Karmakar[b], Pinki Dash[b], Regina C. LaRocque[d,e], Edward T. Ryan[d,e,f], Samuel S. Minot[g], Jason B. Harris*[h,i], Firdausi Qadri*[b], and Ana A. Weil*[c]

*shared sr. authorship

AUTHOR AFFILIATIONS:

[a]Duke University, School of Medicine, Durham, NC, USA

[b]Infectious Diseases Division, International Centre for Diarrheoal Disease Research, Bangladesh, Dhaka, Bangladesh

[c]Department of Medicine, University of Washington, Seattle, WA, USA

[d]Department of Medicine, Harvard Medical School, Boston, MA, USA

[e] Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA

[f]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, MA, USA

[g]Data Core, Fred Hutchison Cancer Research Center, Seattle, WA, USA

[h]Department of Pediatrics, Harvard Medical School, Boston, MA, USA

[i]Division of Global Health, Massachusetts General Hospital Children, Boston, MA, USA

Author emails:

Fred Heller: fjh2@duke.edu

Hasan Banna: hb.bd@outlook.com

Hasanul Kaisar: mhasanulkaisar@gmail.com

Denise Chac: dchac@uw.edu

Fahima Chowdhury: fchowdhury@icddrb.org

Taufiq Bhuiyan: taufiqur@icddrb.org

Afroza Akter: afrozaakter@icddrb.org

Ashraf Khan: ashrafk@icddrb.org

Polash Chandra Karmakar: polash.chandra@icddrb.org

Pinki Dash: pinki.dash@icddrb.org

Regina LaRocque: rclarocque@partners.org

Edward Ryan: etryan@mgh.harvard.edu

Sam Minot: sminot@fredhutch.org

Jason Harris: jbharris@partners.org

Firdausi Qadri: fqadri@icddrb.org

Ana Weil: anaweil@uw.edu

Correspondence: Ana Weil anaweil@uw.edu

ABSTRACT (287/350):

Background: Oral cholera vaccines (OCVs) are an important tool for reduction of the worldwide cholera burden, but some individuals who receive an OCV do not develop protective immune responses. The gut microbiota is a potential explanation for these differences. Components of the gut microbiota associated with differences in OCV response have not been identified.

Results: We used metagenomic sequencing to identify predicted protein-coding genes in the gut microbiota at the time of OCV administration, and then measured immune responses to vaccination. Vaccine recipients were classified as OCV "responders" if they developed a post-vaccination increase in memory B cell populations that produce IgA or IgG specific for cholera toxin and the *V. cholerae* O-specific polysaccharide. We next analyzed microbial genes seen at similar abundances across individual samples and classified these into co-abundant gene groupings (CAGs), and correlated CAGs with OCV responses. Next, to identify specific bacterial strains associated with OCV responses, we mapped CAGs to bacterial genomes and generated a "priority score" for each strain detected in the study population. This score reflects both the number of CAGs aligning to a specific bacterial genome and the strength of the association between the CAGs and the vaccine response. This strain-level analysis revealed relationships between the gut microbiota and immune response to OCV that were not detected at the genus or species level. Bacterial strains which produce short-chain fatty acids and those with sphingolipid-containing cell membranes were correlated with more robust immune responses to vaccination.

Conclusion: Our study demonstrates a method for translating metagenomic sequencing data into strain-specific results associated with a biological outcome. Using this approach, we identified strains for the study of bacterial-derived molecules or metabolites associated with immune responses; such agents might have potential utility as vaccine adjuvants.

KEYWORDS: cholera vaccine*, Vibrio cholerae,* cholera, gut microbiome, microbiota

Text: 5251

INTRODUCTION

Cholera is a severe diarrheal illness caused by the bacterium *Vibrio cholerae*, and millions of cases are reported annually [1]. In 2017, the World Health Organization set the goal of reducing cases of cholera by 90% by 2030 [2]. In addition to cholera surveillance and improved sanitation, use of oral cholera vaccines (OCV) are a critical tool in achieving this aim.

The immunologic mechanisms of long-term protection from cholera are not fully understood. The vibriocidal titer is a commonly used measure of humoral response to *V. cholerae,* and although higher titers are moderately associated with protection from disease, titers wane prior to loss of protective immunity [3, 4]. Another response that is correlated with protective immunity is the production of *V. cholerae*-specific O-specific polysaccharide (OSP) cells (MBCs), that rise following infection and remain elevated for at least one year [5]. In a person exposed to *V. cholerae*, antigen-specific MBCs may differentiate rapidly into antibody-secreting plasma cells to produce antibodies that may act at the mucosal surface to protect from infection. In studies of household contacts of cholera patients*,* those with elevated OSP-specific MBCs at the time of exposure were protected from subsequent infection [6, 7].

Antibody responses to natural infection are largely directed toward cholera toxin (CT) and the OSP of the *V. cholerae* lipopolysaccharide (LPS) [8]. CT consists of a B subunit (CtxB), a pentamer that contacts the mucosal epithelium, and an A subunit, which is endocytosed and mediates fluid secretion into the intestinal lumen, causing the severe watery diarrhea characteristic of cholera [9]. The *V. cholerae* O1 serogroup is responsible for nearly all endemic and epidemic cholera and serogroup specificity is defined by the composition of the OSP. The

O1 serogroup can be divided into two biotypes, Classical and El Tor, and the current seventh pandemic of cholera is caused by the El Tor biotype [10]. Serogroup O1 *V. cholerae* are further defined by the serotypes Inaba and Ogawa, which differ from one another by one methyl group on the terminal saccharide of *V. cholerae* O1 OSP [11].

Several whole-cell (WC) killed OCVs are prequalified by WHO, including Dukoral (Valneva, Saint-Herblain, France) that is composed of inactivated O1 strains Inaba El Tor biotype, and Inaba classical biotype, and two strains of Ogawa classical biotype. Dukoral also contains recombinant CtxB (rCTB), which is thought to provide some short-term protection against *V. cholerae* in clinical trials [12, 13]. Studies in patients who have recovered from cholera demonstrate that adaptive immunity lasts at least 3 years [14, 15]. WC vaccine efficacy in adults lasts for several years, but is more limited in children, especially in those under 5 years of age [16-18], and the overall protective efficacy of several WC OCVs is 30% in children under 5 years of age, compared to 64% in persons over 5 years of age [19].

On a population level, vaccination with a killed WC OCV stimulates an OSP-specific MBC response, but some vaccine recipients have no detectable OSP-specific MBC response [5, 20, 21]. In particular, young children have a reduced responder frequency of MBC response targeting LPS, and this same group is known to have a shorter duration of protection after WC compared with older participants [12, 18, 21]. A proposed explanation for these differences in immunity among adult individuals and between age groups is that the gut microbiota may impact the host response to OCVs [22, 23]. This may be supported by the observation that immune responses to a live OCV were reduced in vaccine recipients with bacterial overgrowth, an effect that the authors proposed could be due to immunomodulation from metabolites of gut microbiota [24]. We have previously explored this question using community-level microbiome sequencing to detect associations between the gut microbiota at the time of vaccination and

OSP-specific MBC responses to a killed WC vaccine [20]. To gain increased resolution on this relationship and to identify specific strains that may mediate this association, we here report use of a novel metagenomic sequencing of stool from children and adults receiving a WC OCV to examine this relationship using a gene-directed analytic method.

METHODS

Study design and subject enrollment

Adult and child participants over 2 years of age and <60 years of age living in urban Dhaka, Bangladesh were enrolled through the International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b) if they had not previously been vaccinated with an OCV and had no major comorbid conditions [25]. After obtaining written informed consent from patients or guardians, study participants received two doses of WC-rCTB doses (Dukoral), per the package insert, on study Day 0 and 14. Blood specimens were collected from the participants on Day -14 (two weeks prior to the first dose of vaccination), Day 2, Day 7, and Day 44. All timepoints are considered in reference to Day 0, the day of administration of the first dose. Vibriocidal titer was assessed at each of these timepoints; OSP-specific and CT-specific MBC were measured at Day -14 and Day 44. Demographic data were collected at enrollment and stool samples were collected at baseline prior to the first dose of WC-rCTB.

Vibriocidal antibody titer

We measured vibriocidal antibody levels in plasma as previously described [26] using guinea pig complement (icddr,b Animal Resources Branch, Dhaka, Bangladesh) and the homologous serotype of *V. cholerae* O1 Ogawa (strain X-25049) and *V. cholerae* O1 Inaba (T-19479) as the target organism. We defined the vibriocidal titer as the reciprocal of the highest dilution resulting in >50% reduction of the optical density compared to that of control wells without plasma [26].

Isolation of peripheral blood mononuclear cells (PBMCs) and MBC enzyme-linked immune absorbent spot assay

PBMCs were recovered by differential centrifugation using Ficoll-Isopaque (Pharmacia, Piscataway, USA). PBMCs were washed and suspended at a concentration of 1 x 10^7 cells/ml in Roswell Park Memorial Institute Medium (RPMI) 1640 complete medium (Gibco, Carlsbad,

USA) with 10% heat-inactivated fetal bovine serum (FBS; HyClone, Logan, USA). MBC assays were performed using fresh PBMCs on day -14 and 44 as previously described [7, 13, 27, 28]. We placed 5 x 10^5 PBMCs/well in cell culture plates (BD Biosciences, San Jose, CA) containing RPMI 1640 and 10% FBS. To stimulate antigen-independent proliferation and differentiation of MBCs into antibody secreting cells (ASCs), a mixture of three B-cell mitogens containing 6 µg/ml CpG oligonucleotide (Operon, Huntsville, USA), a 1/100,000 dilution of crude pokeweed mitogen extract, and a 1/10,000 dilution of fixed *Staphylococcus aureus* Cowan (Sigma, St. Louis, USA) was added to all wells except negative controls, to which only media was added. Plates were incubated at 37°C in 5% CO_2 for 5–6 days, after which the cells were harvested and washed.

For the memory B cell enzyme-linked immune absorbent spot (ELISPOT) assay, we coated nitrocellulose-bottom plates (#MSHAN-4550; Millipore, Bedford, USA) with OSP: Bovine Serum Albumin (BSA) (10 µg/ml), LPS (25 µg/ml), or GM1 ganglioside (3 nmol/ml) followed by recombinant CtxB (2.5 µg/ml), 5 µg/ml affinity-purified goat anti-human immunoglobulin (Jackson Immunology Research, USA) as a positive control, or 2.5 µg/ml keyhole limpet hemocyanin (KLH) (Pierce Biotechnology, USA) as a negative control. After the plates were blocked with RPMI 1640 containing 10% FBS, 20% of the cells from each culture plate well were added for detection of total ASCs and 80% were used for detection of ASCs to the antigens CT and OSP. Plates were incubated for 5 hours at 37°C in 5% CO_2. Next, plates were washed, and alkaline phosphatase-conjugated goat anti-human IgG or horseradish peroxidase-conjugated goat anti-human IgA (Southern Biotech, Birmingham, USA) was added at a dilution of 1:500. Following an overnight incubation at 4°C, plates were developed with 3-amino-9-ethyl carbazole for IgA, or with (5-bromo-4-chloro-3-indolylphosphate)-nitroblue tetrazolium for IgG.

We expressed ELISPOT counts as the percentage of antigen-specific MBCs out of the total IgG, IgA, and IgM MBCs. We used wells coated with KLH and un-stimulated samples as negative controls. We counted the number of MBC ASC per well independently using two reviewers in a sample-blinded fashion with a stereomicroscope and averaged the counts. We applied previously determined criteria for inclusion in the analysis, excluding wells that failed stimulation, exhibited nonspecific stimulation, or were stimulated in response to the negative control [21]. Sufficient stimulation of total antibody isotype-specific MBCs was defined as a 3-fold increase for stimulated versus non-stimulated wells, and we also required fewer than three cells specific to either antigen (CT or OSP) in the corresponding sample's unstimulated wells, and fewer than three antigen-specific cells in the corresponding sample's KLH MBC assay.

Participants were classified as vaccine responders or non-responders according to four vaccine response measures (VRMs). The four VRMs correspond to IgA or IgG OSP-specific MBC responses and IgA or IgG CT-specific MBC responses. A VRM responder was defined as a vaccine recipient that developed an increase in the specific MBC response between pre- and post-vaccination timepoints, and non-responders were defined as having no change in MBC measures between these timepoints. Where participants lacked MBC ELISPOT results for inclusion in our analysis for either timepoint due to insufficient blood volume acquisition, participants were categorized as having "insufficient data" for that VRM.

Stool processing and sequencing

Stool samples were selected for sequencing if the participant also had MBC ELISPOT results for both pre-and post-vaccination timepoints for at least two out of four VRMs. Application of this criteria resulted in sequencing the stool from 90% of study participants. Stool was collected in cryovials and immediately frozen at -80°C and shipped to US investigators. At the University of Washington, stool microbial DNA was extracted using Powersoil DNA isolation kit (Qiagen, Germany) with a modified protocol as previously described (29). Briefly, stool was thawed on ice

and approximately 100 mg of stool was added to bead beating tubes. The stool was treated with C1 solution, heated at 65°C for 10 minutes, 95°C for 3 minutes, and vortexed for 10 minutes. Samples were treated using C2–C5 solutions provided by the Powersoil kit. Resulting DNA was eluted in DNase and RNase free water, and quality and quantity were measured using a NanoDrop ND-1000 (Thermo Scientific, USA).

Sequencing, read processing, and co-abundant gene (CAG) grouping

Whole genome sequencing (WGS) was performed on the NovaSeq 6000 PE150 platform using standard Illumina adapters targeting a minimum of 10M reads per sample. Fastq files were processed using *geneshot*, a pipeline developed to streamline the analysis of gene-level metagenomic data and its association with experimental metadata [29]. Using default settings for this pipeline (v0.9), raw WGS data from each individual specimen was quality-filtered and *de novo* assembled using Megahit [30]; human reads were excluded using human reference genome GRCh37 hg19, and protein-coding sequences were predicted using Prodigal [31] and deduplicated across all specimens using linclust (at 90% amino acid similarity); coding sequences were annotated with taxonomic labels using DIAMOND (by alignment against NCBI RefSeq 01/22/2020) and with functional labels using eggNOG-mapper [32]; WGS reads from each specimen were aligned against the gene catalog with DIAMOND; coding sequences were grouped by co-abundance into CAGs [33]. The relationship between CAGs containing ≥2 genes were correlated with participant VRMs using the corncob algorithm implementation of beta-binomial regression and accounting for multiple comparisons using the Benjamini-Hochberg false-discovery rate correction procedure [34, 35]. CAGs containing only one gene were excluded to reduce the total number of tested hypotheses, considering the relative lack of biological interpretability for singleton coding sequences in the context of gene-level metagenomic analysis. “Top” CAGs were defined prior to the analysis as those with a q-value ≤0.1.

Assessment of gut microbial communities

Reads from all participant samples were used to estimate the relative abundance of taxonomic groups using metaphlan2 [36], which uses the alignment of raw reads against a set of previously identified marker genes that are conserved across taxonomic groupings of genomes. Output of metaphlan2 was merged across all samples and non-bacterial taxa were removed. Alpha diversity measures including Shannon Index, and beta diversity measures including Bray Curtis Dissimilarity Index were calculated using divnet (v3.6) in R (v4.0.2)[37]. Principal coordinate of analysis (PCoA) plots to visualize these results were generated using the R package vegan.

CAG alignment to bacterial genomes

Top CAGs found to be significantly associated with VRMs were aligned to bacterial genomes to identify the strains containing the most top CAGs. A total of 61,918 bacterial reference genomes were selected from the in the National Center for Biotechnology Information prokaryote genome database using DIAMOND v0.9.10 via the AMGMA (Annotation of Microbial Genomes By Microbiome Association) workflow (https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/ and https://github.com/fredhutch/amgma) [33, 38, 39]. We used all bacterial sequences in the database apart from bacterial sequences with low assembly quality and completeness (chromosomes and scaffolds). We included all complete genomes and 85% of contig-level assemblies and excluded 15% of contig-level assemblies with the lowest levels of completeness (>200 contigs). Using the alignment of coding sequences from the metagenomic analysis against this set of partial and full genomes, a relative abundance measure was generated for each genome in each sample by summing the proportion of reads from a given sample that aligned to those genes found in each genome. Using that strain-level relative abundance

measure, we applied the same statistical approach (beta-binomial modeling using the corncob algorithm) to estimate the coefficient of association with VRMs on a per-genome basis.

Development of CAG-based strain identification method

Once CAGs were aligned to bacterial genomes [29, 33], we next quantified the association between these strains and VRMs. For this purpose, we created a priority score designed to 1) account for the strength of the association between top CAGs aligning to a genome and the VRM of interest, 2) preserve of the directionality of the association between the VRM and top CAGs, 3) be independent of CAG size, in order to avoid bias toward large CAGs containing more genes that could potentially align, and 4) discriminate between strains within the same species.

For a given strain with *n* number of significant CAGs aligning to strain genomes, this priority score was calculated as follows:

$$PS_{strain\ X} = \frac{\sum_{CAG\ 1}^{CAG\ n}\left(Num_Genes_Align_{CAG\ 1} \times \pm\sqrt{{est_coefficient_{CAG\ 1}}^2 + {wald_{CAG\ 1}}^2}\right)}{\sum_{CAG\ 1}^{CAG\ n} Num_Genes_Align_{CAG\ 1}} \times \sqrt{{mean_est_coefficient_{Strain\ X}}^2 + {mean_wald_{Strain\ X}}^2}$$

The estimated (est) coefficient reflects the abundance of a CAG in responders compared to non-responders for each VRM, and the Wald score integrates standard error which reflects the uniformity of the association [33] . The "±" in the first term corresponds to the sign of the CAG's estimated coefficient. The number of genes aligning from a CAG to the strain genome is represented in the estimated coefficient and Wald score, creating a weighted measure of CAG-strain association which is summed in the first term's numerator for all top CAGs with genes aligning to the strain. By normalizing this weighted sum in the numerator of the first term by the

unweighted number of genes aligning from top CAGs, we minimize bias based on CAG size. In the 2nd term, the score is further weighted by integrating the number of genes aligned to reference strains without regard for CAG grouping, and thereby discriminates between strains that may have otherwise identical first terms due to the alignment of the same number of genes from the same top CAG(s). A "+" priority score indicates a positive association between the VRM and strain, and a "-" indicates an inverse relationship. Code for computing priority scores can be found in the Additional Materials (see Availability of data and material).

Statistical Analysis

Visualizations, figures, and statistical testing were generated using R (v4.0.2) and Graphpad Prism (v9.0.2) unless otherwise specified. Strain trees were generated using GToTree (v1.5.51) [40] and visualized using the Interactive Tree of Life (v6.1.2) [41]. Comparison of vibriocidal titer and MBC responses between baseline and follow-up timepoints within each age group was completed using two-tailed Mann Whitney U testing with a predetermined significance level of $p \leq 0.05$. Shannon indices were compared between sexes using a two-tailed Mann Whitney U test. A Kruskal-Wallis test with Dunn's adjustment for multiple comparisons was used to assess differences in Shannon index between age groups as well as between responders and non-responders for each vaccine response measures.

RESULTS

Study population and immune responses to vaccination

Ninety-nine participants were enrolled between February 2017 and July 2018 in Dhaka, Bangladesh, and ten were excluded due to an insufficient volume of blood draw that precluded testing all MBC responses. Eighty-nine persons with at least 3 of 4 MBC results (baseline IgG and IgA measures, and follow up IgG and IgA measures) were included in the analysis and their stool underwent metagenomic sequencing. Twenty-two participants were aged 2-5 years (11/22, 50% female), 22 participants >5-17 years (12/22, 54% female), and 45 participants ≥18-46 years (32/45, 71% female). We measured vibriocidal responses to *V. cholerae* O1 Ogawa and Inaba 14 days prior to vaccination and at day 7 and 44 after the first dose (**Suppl Fig 1***).* After the first vaccine dose, the population mean vibriocidal titers significantly increased by day 7 and remained elevated at day 44 (Suppl Fig 1). As previously observed in cholera-endemic areas, the pre-vaccination baseline vibriocidal titer increased with age (comparing the 2-5 yrs age group to each of the latter groups, tested using one way ANOVA, Inaba p=0.03 Ogawa p=0.02, Suppl Fig 1)[42]. Two doses of WC-rCTB also resulted in an increase in MBC responses in some study participants (shown in **Fig 1A**), as seen in prior studies [13, 21]. At a population level, mean increases in IgG MBCs specific for CT and OSP before and after vaccination were statistically significantly different, and CT or OSP-specific MBCs producing IgA were not, as previously observed [13, 21] (shown in **Fig 1B**). Immune responses were as expected based on prior studies of WC vaccine responses from populations living in cholera-endemic areas [5, 18, 20, 43].

Gut microbial diversity and MBC response

Across all participants, the baseline microbiota was dominated by the orders *Lactobacillales*, *Clostridiales*, *Bifidobacteriales*, and *Enterobacteriales* (**Suppl Fig 2**). We first compared taxonomic diversity measures between age and sex groups and found no significant differences

by age (Shannon Index or by Bray Curtis community structure, **Suppl Fig 3A-B**). Shannon diversity (a measure of both species diversity and abundance) showed a difference in gut microbial diversity by sex (**Suppl Fig 3C-D**). These differences were driven by fewer *Lactobacillales, Veillonellaceae, and Eubacteriaceae,* and more *Erysipelotrichales, Ruminococcaceae, and Lachnospiraceae* in women (**Suppl Fig 4**). Although gut microbiota differences between sexes are not well studied in low and middle income countries, the gut microbiota between men and women have demonstrated differences across populations in numerous studies in high-income countries [44]. No significant differences were found by Shannon Index or by Bray Curtis community structure between vaccine responders and non-responders for in the four VRMs (**Suppl Fig 5**).

Identification of co-abundant gene groups (CAGs) associated with vaccine responses

To address the high dimensionality of the metagenomic data we used gene groups as the fundamental unit of analysis. Co-abundant genes (CAGs) are likely to be found in proximity to one another or on a common mobile element and grouping genes together has been successfully used to infer the relationship between genes and clinical outcomes [29, 33, 45]. We identified CAGs associated with VRMs and predetermined a significance threshold of a False-Discovery Rate (FDR)-adjusted q-value of ≤0.1, based on prior studies associating metagenomic data with biological responses [46] (**Fig 2**). We calculated significant CAGs for each VRM separately (IgA OSP, IgA CT, IgG OSP, and IgG CT), and identified CAGs associated with an MBC response or lack of MBC response in each group. A total of 4147 CAGs were identified, and 323 CAGs containing ≥2 genes met our significance threshold (**Suppl Table 1**). Among the 323 significant CAGs, 299 were unique and 24 CAGs were associated with multiple VRMs (**Fig 3**). We found that significant CAGs were not specific to sex, but several CAGs identified did differ significantly between the 2-5 year and >18 age groups (**Suppl Fig 6**)**.** Overall, 219/323 (68%) of all significant CAGs that correlated with VRMs did not

align to any bacterial genomes, likely due to either gaps in reference libraries and unannotated bacterial diversity, or non-bacterial genes that correlate with VRM responses (Suppl Table 1). Because all CAGs are composed of non-human genes, this represents an enormous number of underexplored and unidentified organism diversity of potential clinical significance that may be bacterial, fungal, or viral in origin. The greatest number of significant CAGs that mapped to bacterial genomes were associated with IgA OSP-specific MBC responses, and most of these CAGs (98.2%) were positively associated with a higher immune response (Suppl Table 1).

Estimates of gene function based on significant CAGs

CAG mapping to whole bacterial genomes demonstrated that many CAGs align to multiple reference strains (**Fig 4**, **Suppl Fig 7**). Most significant CAGs contained a small number of genes, and most genes composing significant CAGs contained genes of unknown function (based on the results of the eggnog functional annotation database shown in **Suppl Table 2**). Due to the uncertain quality of gene annotations, we interpret single gene annotations as possibilities, and instead use patterns in this data as possible gut microbial functions that could impact MBC responses to oral vaccination. Among annotated genes in significant CAGs, most encode for nonspecific cellular functions, relate to amino acid metabolism, or were phage associated. Several genes in significant CAGs were involved in handling of heavy metals such as nickel and iron (Suppl Table 2). Among the 21 CAGs that were associated with multiple VRMs, genes in these CAGs were either not annotated, minimally annotated, or nonspecific (annotations listed in **Suppl Table 3**).

Strains associated with Vaccine Response Measures

Every strain found in the stool of our study population that mapped to a significant CAG received a priority score for each VRMs (full results, priority scores, and strain rank, **Suppl Table 4-7**). Top strains associated with each VRM were from diverse taxonomic groups

(summarized by species, **Fig 5, Suppl Fig 8**). We found that typically only a small fraction of strains within a given taxonomic group had a high priority score, indicating that the association between VRMs and gut microbes are strain specific (**Fig 6**). Among 412 total genera detected in the stool of the study population, only 34 genera contained five or more strains that mapped to a significant CAG (Fig 6). To demonstrate the relationship between high priority and unscored strains within one species, we used phylogenetic trees to compare high priority scored strains to related, unscored strains present in the study population (representative tree, **Fig 7**). Our results suggest that even closely related strains may differ in gene content enough to be associated with different biological effects on the gut mucosa. There was one exception to this finding in which nearly all strains of one species strongly correlated with OSP IgA responses. Ninety-nine percent *Clostridioides difficile* (792/800) strains detected in our study population were highly associated with this VRM (Fig 5 and 6).

Overall, strains with high priority scores for OSP IgG and OSP IgA MBC responses were nearly all positively associated (the strain correlated with an MBC response), with few significant CAGs identified that were negatively associated with OSP responses (Table 1). Among the 1,080 strains positively correlated with OSP IgA MBC results, strict anaerobes including strains of *Bacteroides* species (n=32) and *Prevotella copri* (n=53), both bacterial groups with sphingolipid-containing cell walls, were prominent among the highest priority scores (Fig 5, counts in Suppl Table 4). OSP IgG positively associated strains were taxonomically varied and included members of *Bifidobacterium* and *Ruminococcus/Blautia* genera and several Enterobacteriaceae (strains from *Escherichia coli, Enterobacter,* and *Klebsiella,* Suppl Table 5*).* Strains correlating with CT MBC responses included a wide range of positive and negatively associated strains (Suppl Fig 8). Strains from one species were sometimes both negatively and positively associated with a CT VRM (both IgG and IgA), including strains of *Eubacterium rectale*, *Faecalibacterium prausnitzii*, and *Ruminococcus gnavus* (Suppl Fig 8, full data in Supplemental

Table 4-7). Overall, our results indicate that strain-specific analyses may be needed to select appropriate strains for validation experiments of a correlative host-microbe interaction. These findings support the concept that taxonomic groupings alone are not reliable markers of specific genes or associations with biological effects.

Discussion:

Oral cholera vaccines are an essential tool for reducing the global impact of cholera; however, the protection they provide differs between individuals for unclear reasons. Here we investigate components of the host gut microbiome at the time of vaccination and the association with *V. cholerae* antigen-specific memory B cell responses to a whole-cell killed cholera vaccine. We also describe a quantitative method of applying gene-based metagenomic sequencing data to the identification and ranking of bacterial strains associated with a biological outcome.

We focused primarily on *V. cholerae* OSP-specific MBC results because the presence of detectable OSP-specific MBCs is associated with long-term protection from infection, and a potential mechanism for OSP-specific antibody protective immunity has recently been established [6, 7, 47]. In contrast with other toxin-mediated infections, CT-specific antibody responses are associated with only short-term protection [6, 17, 48]. In studies of human cholera, immune responses to T-cell dependent antigens such as CT are longer lasting than responses to the T-cell independent antigen OSP, which suggests that T-cell help may be needed for long-lasting MBC responses [49]. Differences in the immune stimulation pathway between OSP and CT is one possible explanation for differences we observed in the gut microbes associated with CT compared to OSP MBC responses.

Previous studies have identified a relationship between the gut microbiota and immune responses to oral vaccinations to prevent typhoid and rotavirus infections [50, 51], and we have previously studied this relationship in OCV recipients [20]. However, the sequencing methods used in these studies for taxonomic profiling do not permit the gene content resolution needed to distinguishing strain-level associations. To advance our understanding of the relationship between the gut microbiota and immune responses to vaccination, experimental validation and mechanistic testing of bacterial isolates is needed, and specific strains of bacteria must be

chosen to design these downstream studies [52-54]. To inform this process, we examined strain-specific correlations with immune responses to OCVs. We observed no differences in alpha or beta diversity between responders and non-responders for any of the vaccine responses measures we studied, and we found that strain-specific results were not generalizable by genera or species.

Several of the bacterial strains that we found to have a strong correlation with protective immune responses are from taxa known to have potential immunomodulatory effects in the gut that have not yet been studied in the context of oral vaccination. For example, strains from the sphingolipid-cell wall containing genera *Bacteroides* and *Prevotella* were highly associated with positive OSP-specific MBC responses to vaccination. Sphingolipids from gut microbial species can be taken up by the gut epithelium in outer membrane vesicles and act as signaling molecules impacting host inflammation and cell proliferation, and in animal models have shown that these molecules may alter cytokine responses independent of TLR signaling to modulate mucosal B cell responses [55, 56]. Another group of strains highly correlated with responses to vaccination in our study, including strains from the *Ruminococcaceae* family (*Faecalibacterium prausnitzii*, *Ruminococcus gnavus)* and genera from the *Lachnospiraceae* family (including *Blautia, Eubacterium, Coprococcus,* and *Dorea*) are known to produce short chain fatty acids (SCFA) that are carboxylic acids produced as a byproduct of anaerobic fermentation. Common SCFAs made by the gut microbiota are acetate, propionate, and butyrate, molecules known to act as both histone-deacetylase (HDAC) inhibitors and as ligands for G-protein coupled receptors (GPCRs) that detect signals outside the cell and activate cellular responses [57]. For example, activation of GPCR109A expressed on intestinal epithelial cells downregulates the LPS-induced Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-κB) pathway activation in a mouse model in the presence of butyrate [58]. NFκB innate immune signaling is essential for B cell activation and isotype switching, and modulation of this pathway by SCFAs

may influence creation of MBCs in response to OCVs and may also mediate MBC differentiation and development through additional GPCR mechanisms [59-61]. SCFA are also known to impact intestinal permeability, and this could alter the accessibility of OCV antigens to lamina propria macrophages and other innate immune effector cells [62, 63].

Several co-abundant gene groups (CAGs) that were highly associated with immune responses to OCVs in our study mapped to pathogens, such as strains in the *Shigella* genus and strains of *Salmonella enterica*. Persons with diarrhea were excluded from enrolling in this study, and these pathogens may have been detected in the stool due to subclinical or recent infection, or could be a marker of exposure to contaminated food and water, as has been described in other study participants living in areas lacking modern sanitation [64-68]. Although these pathogens are known to interact with host microbes during their respective symptomatic infections in humans, studies of how these microbes impact the gut mucosa as “bystander” members is not known. Notably, we found that many strains of *Clostridiales difficile* were highly associated with IgA-specific OSP responses, but not other immune responses*. C. difficile* is an opportunistic pathogen that can also be a part of the “healthy” gut microbiota and has been previously observed in this population [69]. Two *C. difficile* peptides (the fragment of the receptor-binding domain of toxin A TxA[C314] and a fragment of the 36 kDa surface-layer protein [SLP-36kDa] from strain C253) are known to have immunoadjuvant properties; investigators previously noted an increase in mucosal IgA and systemic IgG targeting antigen when these peptides were co-administered to mucosal surfaces in mice [70]. This mechanism is a potential explanation for the strong correlation observed in our study between *C. difficile* and OSP IgA responses, and the specific pathways activated in response to OCV antigen will require further study.

Two-thirds of the CAGs that we found to be significantly associated with more robust immune responses to vaccination did not align to any bacterial reference genomes. This supports our

methodology for application of this method not only for the identification of strains for use in *in vitro* mechanistic testing, but also for the discovery of novel genes of interest in bacteria, viruses, fungi, and other organisms. Furthermore, many of the genes in significant CAGs that did align to bacterial genomes are not well characterized.

Our study was limited by unequal enrollment by sex among adults. As expected, we found differences in the gut microbiota between males and females in this study, but this was independent of overall immune responses to vaccination, and CAG distribution did not differ between sexes. We did not analyze blood group data for participants, which has been reported in some studies to impact vaccine responses [71, 72]. We also did not collect data on antibiotic use or helminth infection, which are both widespread in this population, and could theoretically impact immune responses [73, 74]. We were not able to control for the presence or absence of helminths in study participant microbiota using sequencing data due to the limitations of metagenomic analysis including a relatively low abundance of eukaryotic reads compared to bacterial reads, and reference libraries for eukaryotic genes that are frequently contaminated with bacterial reads [75, 76]. Our analysis was focused on bacteria, which excludes other organisms found in stool. Also, we measured genomic DNA in stool to evaluate gene content, and we did not assess cell viability. Our analysis also does not account for potential gut microbe intra-community interactions and how this may impact immune responses to OCV. Assessments of mRNA transcripts and the stool proteome would further shed light on the relationship between gene content, specific bacterial strains, and VRMs.

We have demonstrated a method for quantifying the relationship between gene-level genomic data and a biological outcome for the purpose of identifying candidate strains for *in vitro* study. This approach will be more powerful as the size and accuracy of genome reference libraries and predicted functional annotations increases. Using this method, we have identified bacterial

strains associated with MBC responses linked to protective immunity. Human microbiota gene-based analyses that determine specific strains of interest for validation experiments and mechanistic testing are a requirement for both increasing our understanding of immune response to oral vaccines, and harnessing the gut microbiota for development of probiotic, prebiotic, or bacterial metabolite vaccine adjuvants.

DECLARATIONS

Ethics approval and consent for publication

This study was approved by the Research Review and Ethics Review Committee of the icddr,b, in Dhaka, Bangladesh. Approvals for this work were also granted by the Institutional Review Board of the Massachusetts General Hospital and the University of Washington. Written informed consent was obtained for all participants in this study. For children below 17 years of age, permission of at least one parent/guardian was required for participation. For those aged 11-17 years of age, participant written assent was also obtained. The authors declare that they have no competing interests.

Availability of data and material

The raw data can be accessed through NCBI, accession #PRJNA782606. Additional dataset(s) supporting the conclusions of this article are included within the article (and its additional files). In **Additional file 1**, participant metadata is presented including sex (female=0, male=1), age (yrs), ELISPOT-measured MBC response at baseline and follow-up timepoints for each VRM, and participant designation as responder (R) or non-responder (NR) for each VRM (R=1, NR=0; blanks indicate a participant had data not meeting quality control criteria for that VRM). In **Additional file 2**, a record of scripts used for priority score generation is recorded (also available in GitHub: https://github.com/letsgetthisfred/Dukoral_gene_level_analysis/blob/main/Generic%20Code%20for%20priority%20score%20and%20geneshot%20output.R).

Funding

This work was supported by the NIH [NIAID R01 AI103055 to J. B. H. and R. C. L.; R01 A1099243 to J. B. H. and F. Q, K08 AI123494 to A. A. W.; R01 106878 to E.T.R, and NIGMS

R35 GM133420 (support of SSM, PI: Amy D. Willis)], Fogarty International Center [D43 TW005572 to T.R.B, and K43 TW010362 to T.R.B], the Government of the People's Republic of Bangladesh (to the International Centre for Diarrheal Disease Research [icddr,b]), Global Affairs Canada (to the icddr,b), the Swedish International Development Cooperation Agency (to the icddr,b), and the UK Department for International Development (to the icddr,b).

Acknowledgments

We would like to thank Meti Debela, Dr. Tsoni Peled, the study participants, and the icddr,b laboratory, field, and data management staff. The icddr,b (International Centre for Diarrheal Disease Research, Bangladesh) gratefully acknowledges the government of the People's Republic of Bangladesh, Global Affairs Canada, the Swedish International Development Cooperation Agency, and the Department for International Development. The funding bodies had no role in the design of the study, data collection, analysis, interpretation of data or writing of the manuscript. We would also like to thank Pavol Kováč and Peng Xu of NIDDK for assistance with OSP-based reagents.

Competing interests

All authors report no conflicts.

Author's Contributions

Fred Heller: Immunologic and microbiome data analysis, creation of priority score, figure generation, manuscript writing.

Hasan Banna: Generation of immunologic data, interpretation of data, and manuscript writing.

Hasanul Kaisar: Generation of immunologic data, interpretation of data, and manuscript writing.

Denise Chac: Immunologic and microbiome data analysis, figure generation, manuscript writing.

Fahima Chowdhury: Recruitment of study participants, management of clinical study, manuscript editing.

Taufiq Bhuiyan: Recruitment of study participants, management of clinical study, interpretation of results, manuscript editing.

Afroza Akter: Generation of immunologic data, manuscript writing.

Ashraf Khan: Recruitment of study participants, management of clinical study, manuscript editing.

Polash Chandra Karmakar: Management of clinical study, generation of immunologic data

Pinki Dash: Management of clinical study, generation of immunologic data

Regina LaRocque: Study design and manuscript editing.

Edward Ryan: Study design, reagent generation, interpretation of results, manuscript writing and editing.

Sam Minot: Microbiome and immunologic data analysis, priority score creation, manuscript writing.

Jason Harris: Study design and manuscript editing.

Firdausi Qadri: Study design, recruitment of study participants, and manuscript editing.

Ana Weil: Study design, immunologic and microbiome data analysis, interpretation of results, creation of priority score, figure generation, and manuscript writing.

FIGURES AND TABLES

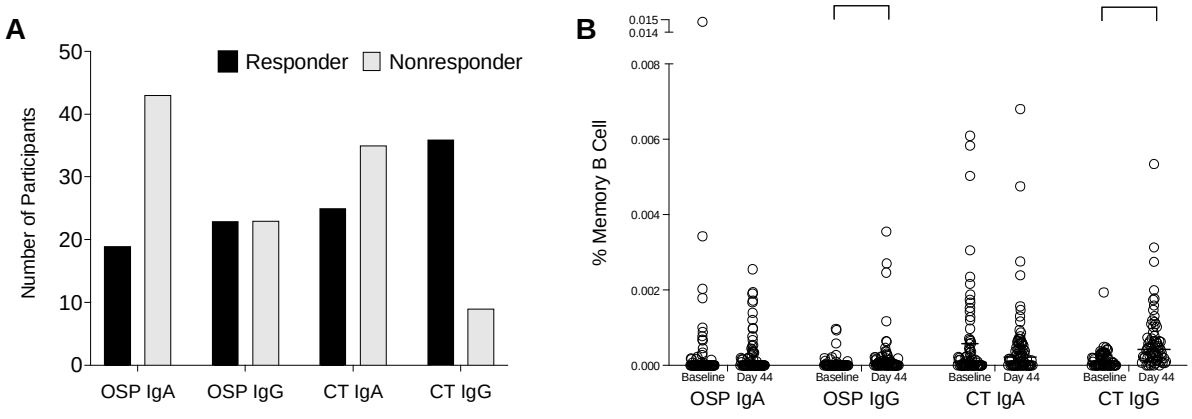

Figure 1: CT and OSP-specific Memory B cell responses after OCV. (**A**) Counts of responders (R) and non-responder (NR) participants. Responders were defined as an OCV recipient who developed an increase in MBC response between pre- and post-vaccination timepoints, and non-responders were defined as having no change in MBC measures between timepoints. (**B**) Memory B cell responses pre- and post-vaccination. Bars represent the median value and interquartile range. Multiple Mann-Whitney U test was used for analysis, and an asterisk denotes statistically significant differences ($p<0.05$) between pre- and post-vaccination timepoints.

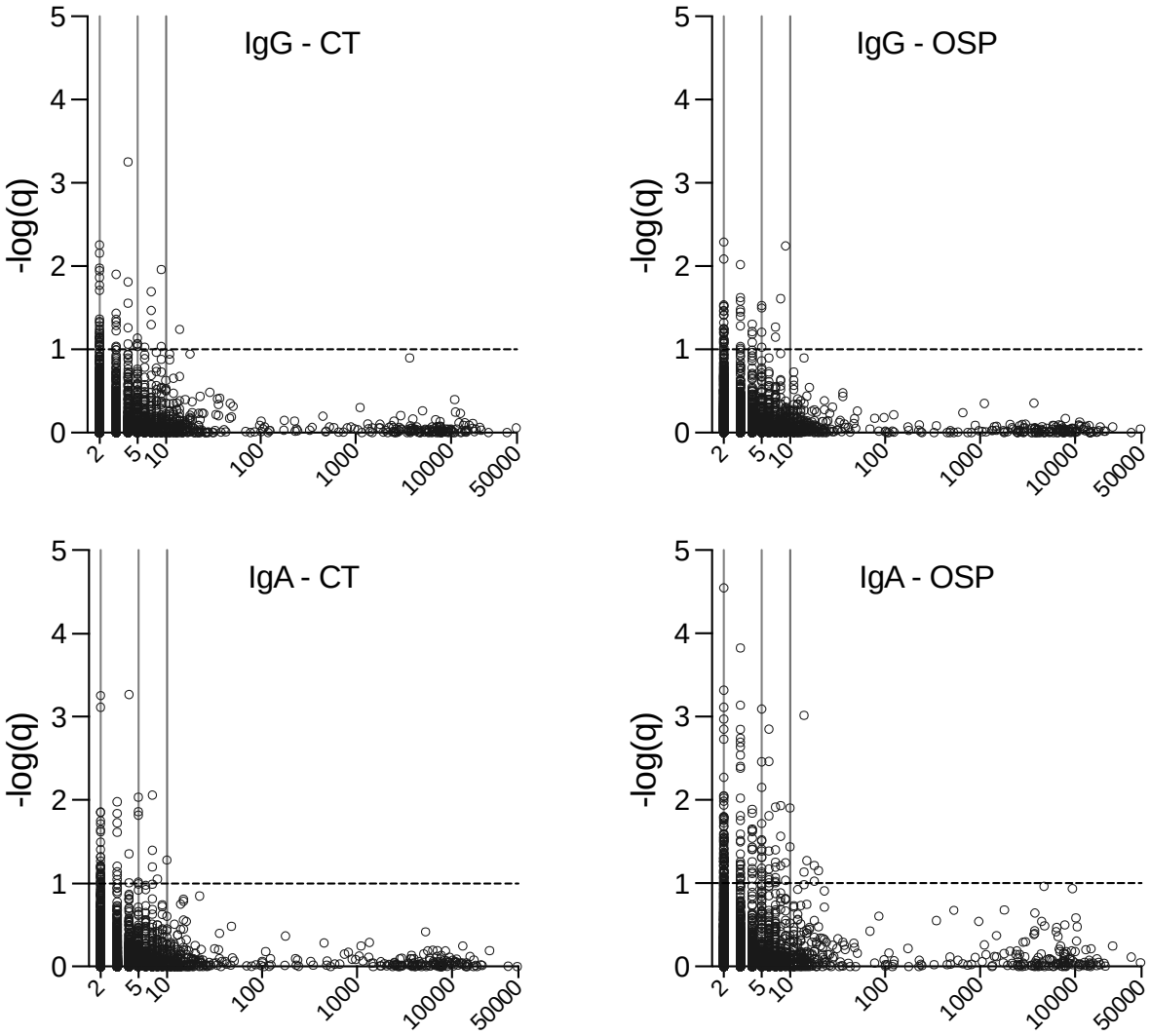

Figure 2: Co-abundant gene groups (CAGs) shown by size distribution and association with VRMs. CAGs are plotted as circles by the number of genes contained in a CAG on the x axis and the -log of their FDR adjusted q-value on the Y axis. The dotted line corresponds to q=0.1 and represents our threshold for CAG significance, demonstrating that most of the significant CAGs identified contained 2-15 genes.

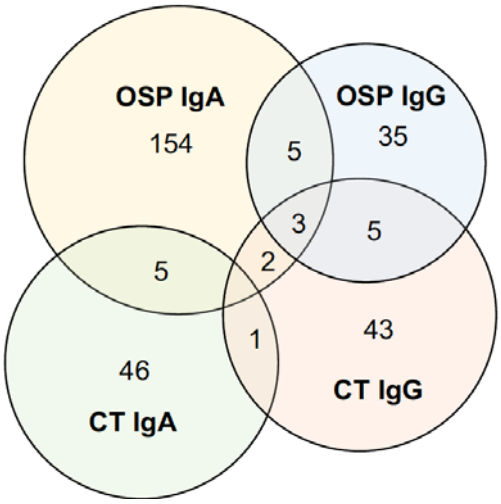

Figure 3: Significant CAG distribution within and between each VRM. The number of CAGs meeting significance criteria for each VRM are indicated within each oval and counts of CAGs found to correlate with multiple VRMs shown in overlapping circles. If VRM circles are not overlapping, there were no significant CAGs found that corresponded to those VRMs. Circle size corresponds with the number of significant CAGs associated with each VRM.

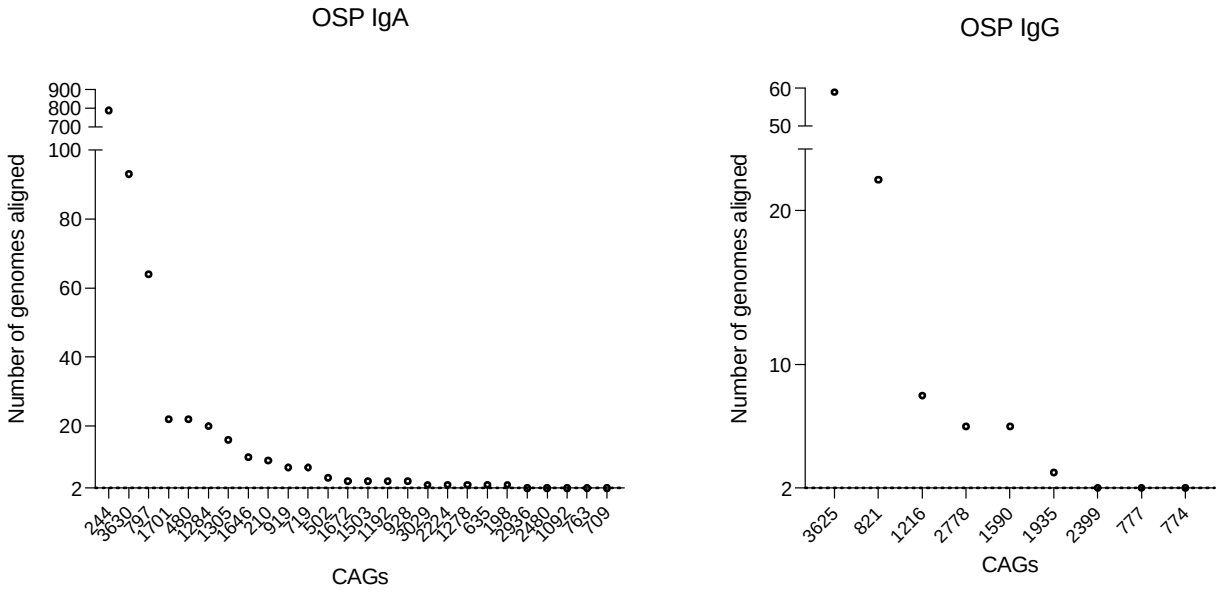

Figure 4: Number of genomes aligning to significant CAGs found to correlate with OSP-specific IgA and IgG producing MBCs after OCV. Each circle represents one top CAG containing ≥2 genes with q≤0.1 association with the listed VRM. The Y axis represents the number of reference genomes to which genes in those CAGs align with >90% sequence identity. Significant CAGs aligning to only one bacterial genome are not shown: 20 CAGs for OSP IgA, 3 CAGs for OSP IgG.

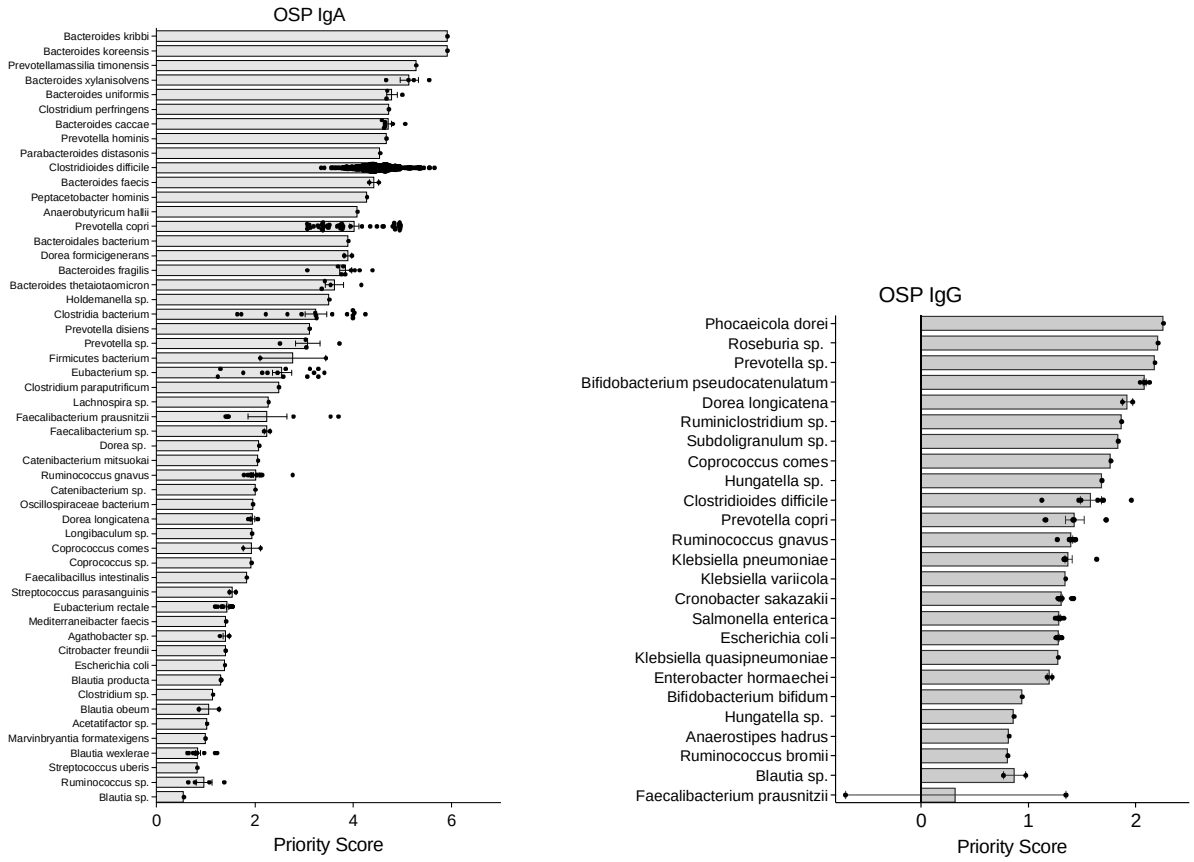

Figure 5: Top strains associated with OSP-specific MBC responses after OCV. Species containing strains with a priority score greater than 0.5 are shown for (**A**) OSP IgA and (**B**) OSP IgG MBC responses. Each dot indicates the priority score of a specific strain within the species listed on the left in each row. For each species, bars represent mean with SEM of priority scored strains. Full strain data and priority scores for each VRM are shown in Suppl Tables 4-7.

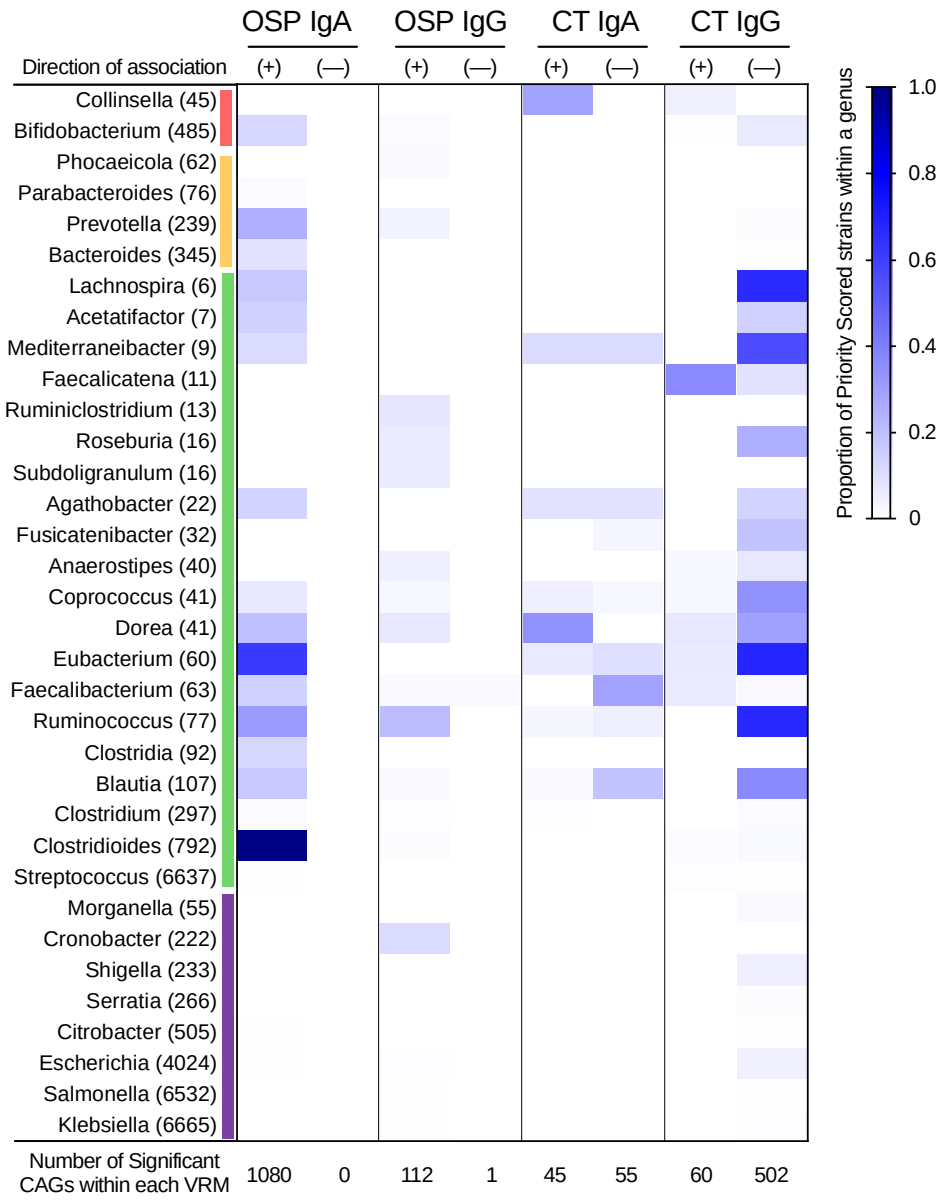

Figure 6: Proportion of strains within a genus that are associated with a specific VRM. Rows shown in this heatmap are genera found in the gut microbiota of the study population which contained 5 or more strains containing significant CAGs. Columns represent VRMs and indicate a positive or negative association. Parentheses indicate the total number of distinct strains in each genus found in our study population, and blue shading represents the % of the total strains that contained significant CAGs. Phylum identifications are shown in a color bar vertically to the right of genus names (red-Bacteroidetes; orange-Actinobacteria; green-

Firmicutes; purple-Proteobacteria). For example, approximately 15% the 485 *Bifidobacterium* strains detected in the study population were assigned a positive priority score associated with OSP IgA responses.

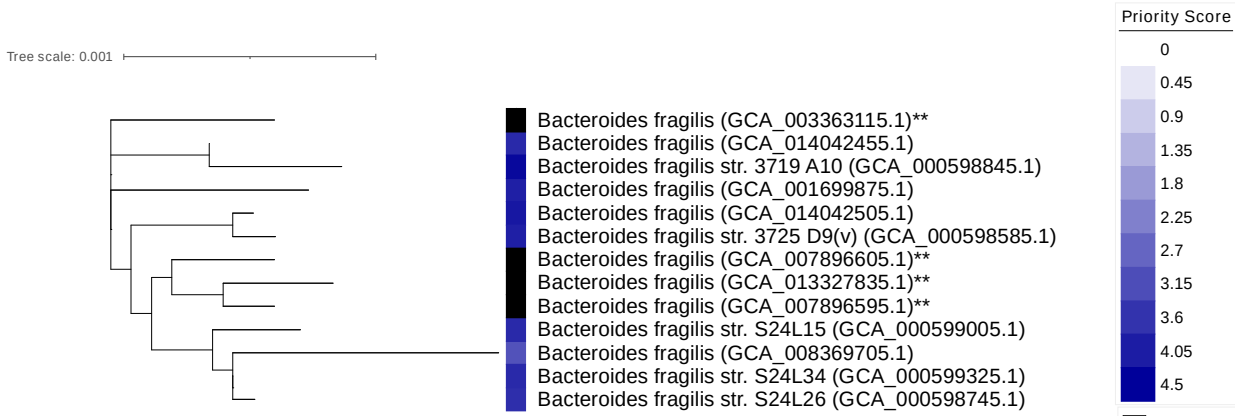

Figure 7: Closely related bacterial strains may differ in their associations with VRMs. A phylogenetic tree of *Bacteroides fragilis* strains and their associations with the OSP IgA VRM is shown. All *B. fragilis* strains listed were detected in the study population, and the color gradient of priority score quantifies the strength of the association found between the OSP IgA MBC response and the strain listed. Strains shown with black squares and asterisks after their names were identified in the stool of the study participants but did not contain significant CAGs (i.e., were not associated with OSP IgA responses).

	Number of positively associated strains	Number of negatively associated strains	Mean positive priority score (SD)	Mean negative priority score (SD)
OSP IgA	1107	0	3.90 (1.37)	NA
OSP IgG	112	1	1.39 (0.33)	-0.704 (NA)
CT IgA	45	55	1.81 (0.95)	-2.30 (0.70)
CT IgG	59	503	1.20 (1.17)	-2.78 (0.95)

Table 1: Summary of strains associated with specific vaccine response measures. Counts of strains positively and negatively associated with each VRM based on alignment of the 323 significant CAGs. Mean priority scores for positively and negatively associated strains, along with standard deviations (SD), are also shown by VRM. Please see Supplemental Tables 4-7 for complete strain listings.

REFERENCES

1. Ali M, Nelson AR, Lopez AL, Sack DA: **Updated global burden of cholera in endemic countries**. *PLoS neglected tropical diseases* 2015, **9**(6):e0003832.
2. **Ending Cholera - A Global Roadmap to 2030. World Health Organization.** . http://wwwwhoint/cholera/publications/global-roadmap/en/ 2017(*Last accessed Oct 20, 2021*).
3. Ritter AS, Chowdhury F, Franke MF, Becker RL, Bhuiyan TR, Khan AI, Saha NC, Ryan ET, Calderwood SB, LaRocque RC *et al*: **Vibriocidal Titer and Protection From Cholera in Children**. *Open Forum Infect Dis* 2019, **6**(4):ofz057.
4. Saha D, LaRocque RC, Khan AI, Harris JB, Begum YA, Akramuzzaman SM, Faruque AS, Ryan ET, Qadri F, Calderwood SB: **Incomplete correlation of serum vibriocidal antibody titer with protection from Vibrio cholerae infection in urban Bangladesh**. *The Journal of infectious diseases* 2004, **189**(12):2318-2322.
5. Falkard B, Charles RC, Matias WR, Mayo-Smith LM, Jerome JG, Offord ES, Xu P, Kovac P, Ryan ET, Qadri F *et al*: **Bivalent oral cholera vaccination induces a memory B cell response to the V. cholerae O1-polysaccharide antigen in Haitian adults**. *PLoS neglected tropical diseases* 2019, **13**(1):e0007057.
6. Aktar A, Rahman MA, Afrin S, Akter A, Uddin T, Yasmin T, Sami MIN, Dash P, Jahan SR, Chowdhury F *et al*: **Plasma and memory B cell responses targeting O-specific polysaccharide (OSP) are associated with protection against Vibrio cholerae O1 infection among household contacts of cholera patients in Bangladesh**. *PLoS Negl Trop Dis* 2018, **12**(4):e0006399.
7. Patel SM, Rahman MA, Mohasin M, Riyadh MA, Leung DT, Alam MM, Chowdhury F, Khan AI, Weil AA, Aktar A *et al*: **Memory B cell responses to Vibrio cholerae O1 lipopolysaccharide are associated with protection against infection from household contacts of patients with cholera in Bangladesh**. *Clinical and vaccine immunology : CVI* 2012, **19**(6):842-848.
8. Kauffman RC, Bhuiyan TR, Nakajima R, Mayo-Smith LM, Rashu R, Hoq MR, Chowdhury F, Khan AI, Rahman A, Bhaumik SK *et al*: **Single-Cell Analysis of the Plasmablast Response to Vibrio cholerae Demonstrates Expansion of Cross-Reactive Memory B Cells**. *mBio* 2016, **7**(6).
9. Weil AA, Becker RL, Harris JB: **Vibrio cholerae at the Intersection of Immunity and the Microbiome**. *mSphere* 2019, **4**(6).

10. Son MS, Megli CJ, Kovacikova G, Qadri F, Taylor RK: **Characterization of Vibrio cholerae O1 El Tor biotype variant clinical isolates from Bangladesh and Haiti, including a molecular genetic analysis of virulence genes**. *Journal of clinical microbiology* 2011, **49**(11):3739-3749.

11. Azman AS, Lessler J, Luquero FJ, Bhuiyan TR, Khan AI, Chowdhury F, Kabir A, Gurwith M, Weil AA, Harris JB *et al*: **Estimating cholera incidence with cross-sectional serology**. *Sci Transl Med* 2019, **11**(480).

12. Clemens JD, Sack DA, Chakraborty J, Rao MR, Ahmed F, Harris JR, van Loon F, Khan MR, Yunis M, Huda S *et al*: **Field trial of oral cholera vaccines in Bangladesh: evaluation of anti-bacterial and anti-toxic breast-milk immunity in response to ingestion of the vaccines**. *Vaccine* 1990, **8**(5):469-472.

13. Alam MM, Riyadh MA, Fatema K, Rahman MA, Akhtar N, Ahmed T, Chowdhury MI, Chowdhury F, Calderwood SB, Harris JB *et al*: **Antigen-specific memory B-cell responses in Bangladeshi adults after one- or two-dose oral killed cholera vaccination and comparison with responses in patients with naturally acquired cholera**. *Clinical and vaccine immunology : CVI* 2011, **18**(5):844-850.

14. Levine MM, Black RE, Clements ML, Cisneros L, Nalin DR, Young CR: **Duration of infection-derived immunity to cholera**. *The Journal of infectious diseases* 1981, **143**(6):818-820.

15. Glass RI, Becker S, Huq MI, Stoll BJ, Khan MU, Merson MH, Lee JV, Black RE: **Endemic cholera in rural Bangladesh, 1966-1980**. *Am J Epidemiol* 1982, **116**(6):959-970.

16. Arifuzzaman M, Rashu R, Leung DT, Hosen MI, Bhuiyan TR, Bhuiyan MS, Rahman MA, Khanam F, Saha A, Charles RC *et al*: **Antigen-specific memory T cell responses after vaccination with an oral killed cholera vaccine in Bangladeshi children and comparison to responses in patients with naturally acquired cholera**. *Clinical and vaccine immunology : CVI* 2012, **19**(8):1304-1311.

17. Clemens JD, Sack DA, Harris JR, Van Loon F, Chakraborty J, Ahmed F, Rao MR, Khan MR, Yunus M, Huda N *et al*: **Field trial of oral cholera vaccines in Bangladesh: results from three-year follow-up**. *Lancet* 1990, **335**(8684):270-273.

18. Sur D, Kanungo S, Sah B, Manna B, Ali M, Paisley AM, Niyogi SK, Park JK, Sarkar B, Puri MK *et al*: **Efficacy of a low-cost, inactivated whole-cell oral cholera vaccine: results from 3 years of follow-up of a randomized, controlled trial**. *PLoS Negl Trop Dis* 2011, **5**(10):e1289.

19. Bi Q, Ferreras E, Pezzoli L, Legros D, Ivers LC, Date K, Qadri F, Digilio L, Sack DA, Ali M *et al*: **Protection against cholera from killed whole-cell oral cholera vaccines: a systematic review and meta-analysis**. *Lancet Infect Dis* 2017, **17**(10):1080-1088.
20. Chac D, Bhuiyan TR, Saha A, Alam MM, Salma U, Jahan N, Chowdhury F, Khan AI, Ryan ET, LaRocque R *et al*: **The gut microbiota and development of Vibrio cholerae-specific long-term memory B cells in adults after whole-cell killed oral cholera vaccine**. *Infection and immunity* 2021:Iai0021721.
21. Leung DT, Rahman MA, Mohasin M, Patel SM, Aktar A, Khanam F, Uddin T, Riyadh MA, Saha A, Alam MM *et al*: **Memory B cell and other immune responses in children receiving two doses of an oral killed cholera vaccine compared to responses following natural cholera infection in Bangladesh**. *Clinical and vaccine immunology : CVI* 2012, **19**(5):690-698.
22. Praharaj I, John SM, Bandyopadhyay R, Kang G: **Probiotics, antibiotics and the immune responses to vaccines**. *Philos Trans R Soc Lond B Biol Sci* 2015, **370**(1671).
23. Qadri F, Bhuiyan TR, Sack DA, Svennerholm AM: **Immune responses and protection in children in developing countries induced by oral vaccines**. *Vaccine* 2013, **31**(3):452-460.
24. Lagos R, Fasano A, Wasserman SS, Prado V, San Martin O, Abrego P, Losonsky GA, Alegria S, Levine MM: **Effect of small bowel bacterial overgrowth on the immunogenicity of single-dose live oral cholera vaccine CVD 103-HgR**. *The Journal of infectious diseases* 1999, **180**(5):1709-1712.
25. Saha A, Khan A, Salma U, Jahan N, Bhuiyan TR, Chowdhury F, Khan AI, Khanam F, Muruganandham S, Reddy Kandukuri S *et al*: **The oral cholera vaccine Shanchol when stored at elevated temperatures maintains the safety and immunogenicity profile in Bangladeshi participants**. *Vaccine* 2016, **34**(13):1551-1558.
26. Qadri F, Wenneras C, Albert MJ, Hossain J, Mannoor K, Begum YA, Mohi G, Salam MA, Sack RB, Svennerholm AM: **Comparison of immune responses in patients infected with Vibrio cholerae O139 and O1**. *Infection and immunity* 1997, **65**(9):3571-3576.
27. Crotty S, Aubert RD, Glidewell J, Ahmed R: **Tracking human antigen-specific memory B cells: a sensitive and generalized ELISPOT system**. *J Immunol Methods* 2004, **286**(1-2):111-122.
28. Aktar A, Rahman MA, Afrin S, Faruk MO, Uddin T, Akter A, Sami MIN, Yasmin T, Chowdhury F, Khan AI *et al*: **O-Specific Polysaccharide-Specific Memory B Cell**

Responses in Young Children, Older Children, and Adults Infected with Vibrio cholerae O1 Ogawa in Bangladesh. *Clin Vaccine Immunol* 2016, **23**(5):427-435.

29. Minot SS, Barry KC, Kasman C, Golob JL, Willis AD: **geneshot: gene-level metagenomics identifies genome islands associated with immunotherapy response**. *Genome Biol* 2021, **22**(1):135.

30. Li D, Liu CM, Luo R, Sadakane K, Lam TW: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph**. *Bioinformatics* 2015, **31**(10):1674-1676.

31. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification**. *BMC bioinformatics* 2010, **11**:119.

32. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P: **Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper**. *Mol Biol Evol* 2017, **34**(8):2115-2122.

33. Minot SS, Willis AD: **Clustering co-abundant genes identifies components of the gut microbiome that are reproducibly associated with colorectal cancer and inflammatory bowel disease**. *Microbiome* 2019, **7**(1):110.

34. Benjamini Y HY: **Controlling the false discovery rate: a practical and powerful approach to multiple testing on JSTOR**. *J R Stat Soc Ser B* 1995, **57**:289-300.

35. Martin BD, Witten D, Willis AD: **MODELING MICROBIAL ABUNDANCES AND DYSBIOSIS WITH BETA-BINOMIAL REGRESSION**. *Ann Appl Stat* 2020, **14**(1):94-115.

36. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C: **Metagenomic microbial community profiling using unique clade-specific marker genes**. *Nat Methods* 2012, **9**(8):811-814.

37. Willis AD, Martin BD: **Estimating diversity in networked ecological communities**. *Biostatistics* 2020.

38. Buchfink B, Reuter K, Drost HG: **Sensitive protein alignments at tree-of-life scale using DIAMOND**. *Nature methods* 2021, **18**(4):366-368.

39. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND**. *Nature methods* 2015, **12**(1):59-60.

40. Lee MD: **GToTree: a user-friendly workflow for phylogenomics**. *Bioinformatics* 2019, **35**(20):4162-4164.

41. Letunic I, Bork P: **Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation**. *Nucleic Acids Res* 2021.

42. Mosley WH, Ahmad S, Benenson AS, Ahmed A: **The relationship of vibriocidal antibody titre to susceptibility to cholera in family contacts of cholera patients**. *Bull World Health Organ* 1968, **38**(5):777-785.

43. Leung DT, Uddin T, Xu P, Aktar A, Johnson RA, Rahman MA, Alam MM, Bufano MK, Eckhoff G, Wu-Freeman Y *et al*: **Immune responses to the O-specific polysaccharide antigen in children who received a killed oral cholera vaccine compared to responses following natural cholera infection in Bangladesh**. *Clinical and vaccine immunology : CVI* 2013, **20**(6):780-788.

44. Kim YS, Unno T, Kim BY, Park MS: **Sex Differences in Gut Microbiota**. *World J Mens Health* 2020, **38**(1):48-60.

45. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, Lipson KS, Knight R, Caporaso JG, Segata N *et al*: **Species-level functional profiling of metagenomes and metatranscriptomes**. *Nature methods* 2018, **15**(11):962-968.

46. Kishikawa T, Ogawa K, Motooka D, Hosokawa A, Kinoshita M, Suzuki K, Yamamoto K, Masuda T, Matsumoto Y, Nii T *et al*: **A Metagenome-Wide Association Study of Gut Microbiome in Patients With Multiple Sclerosis Revealed Novel Disease Pathology**. *Front Cell Infect Microbiol* 2020, **10**:585973.

47. Charles RC, Kelly M, Tam JM, Akter A, Hossain M, Islam K, Biswas R, Kamruzzaman M, Chowdhury F, Khan AI *et al*: **Humans Surviving Cholera Develop Antibodies against Vibrio cholerae O-Specific Polysaccharide That Inhibit Pathogen Motility**. *mBio* 2020, **11**(6).

48. Harris JB, Larocque RC, Chowdhury F, Khan AI, Logvinenko T, Faruque AS, Ryan ET, Qadri F, Calderwood SB: **Susceptibility to Vibrio cholerae Infection in a Cohort of Household Contacts of Patients with Cholera in Bangladesh**. *PLoS neglected tropical diseases* 2008, **2**(4):e221.

49. Harris AM, Bhuiyan MS, Chowdhury F, Khan AI, Hossain A, Kendall EA, Rahman A, LaRocque RC, Wrammert J, Ryan ET *et al*: **Antigen-specific memory B-cell responses to Vibrio cholerae O1 infection in Bangladesh**. *Infection and immunity* 2009, **77**(9):3850-3856.

50. Eloe-Fadrosh EA, McArthur MA, Seekatz AM, Drabek EF, Rasko DA, Sztein MB, Fraser CM: **Impact of oral typhoid vaccination on the human gut microbiota and correlations with s. Typhi-specific immunological responses**. *PloS one* 2013, **8**(4):e62026.

51. Harris VC, Armah G, Fuentes S, Korpela KE, Parashar U, Victor JC, Tate J, de Weerth C, Giaquinto C, Wiersinga WJ *et al*: **Significant Correlation Between the Infant Gut Microbiome and Rotavirus Vaccine Response in Rural Ghana**. *The Journal of infectious diseases* 2017, **215**(1):34-41.
52. Laudadio I, Fulci V, Palone F, Stronati L, Cucchiara S, Carissimi C: **Quantitative Assessment of Shotgun Metagenomics and 16S rDNA Amplicon Sequencing in the Study of Human Gut Microbiome**. *OMICS* 2018, **22**(4):248-254.
53. Ley RE: **Gut microbiota in 2015: Prevotella in the gut: choose carefully**. *Nat Rev Gastroenterol Hepatol* 2016, **13**(2):69-70.
54. Gupta VK, Chaudhari NM, Iskepalli S, Dutta C: **Divergences in gene repertoire among the reference Prevotella genomes derived from distinct body sites of human**. *BMC Genomics* 2015, **16**:153.
55. Duan RD, Nilsson A: **Metabolism of sphingolipids in the gut and its relation to inflammation and cancer development**. *Prog Lipid Res* 2009, **48**(1):62-72.
56. Sepahi A, Cordero H, Goldfine H, Esteban M, Salinas I: **Symbiont-derived sphingolipids modulate mucosal homeostasis and B cells in teleost fish**. *Scientific reports* 2016, **6**:39054.
57. Zhang Z, Tang H, Chen P, Xie H, Tao Y: **Demystifying the manipulation of host immunity, metabolism, and extraintestinal tumors by the gut microbiome**. *Signal Transduct Target Ther* 2019, **4**:41.
58. Thangaraju M, Cresci GA, Liu K, Ananth S, Gnanaprakasam JP, Browning DD, Mellinger JD, Smith SB, Digby GJ, Lambert NA *et al*: **GPR109A is a G-protein-coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon**. *Cancer Res* 2009, **69**(7):2826-2832.
59. Gerondakis S, Siebenlist U: **Roles of the NF-kappaB pathway in lymphocyte development and function**. *Cold Spring Harb Perspect Biol* 2010, **2**(5):a000182.
60. Yang W, Xiao Y, Huang X, Chen F, Sun M, Bilotta AJ, Xu L, Lu Y, Yao S, Zhao Q *et al*: **Microbiota Metabolite Short-Chain Fatty Acids Facilitate Mucosal Adjuvant Activity of Cholera Toxin through GPR43**. *Journal of immunology (Baltimore, Md : 1950)* 2019, **203**(1):282-292.
61. Kang S, Keener AB, Jones SZ, Benschop RJ, Caro-Maldonado A, Rathmell JC, Clarke SH, Matsushima GK, Whitmire JK, Vilen BJ: **IgG-Immune Complexes Promote B Cell Memory by Inducing BAFF**. *Journal of immunology (Baltimore, Md : 1950)* 2016, **196**(1):196-206.

62. Zheng L, Kelly CJ, Battista KD, Schaefer R, Lanis JM, Alexeev EE, Wang RX, Onyiah JC, Kominsky DJ, Colgan SP: **Microbial-Derived Butyrate Promotes Epithelial Barrier Function through IL-10 Receptor-Dependent Repression of Claudin-2**. *J Immunol* 2017, **199**(8):2976-2984.
63. Lewis K, Lutgendorff F, Phan V, Söderholm JD, Sherman PM, McKay DM: **Enhanced translocation of bacteria across metabolically stressed epithelia is reduced by butyrate**. *Inflammatory bowel diseases* 2010, **16**(7):1138-1148.
64. Lindsay B, Ochieng JB, Ikumapayi UN, Toure A, Ahmed D, Li S, Panchalingam S, Levine MM, Kotloff K, Rasko DA *et al*: **Quantitative PCR for detection of Shigella improves ascertainment of Shigella burden in children with moderate-to-severe diarrhea in low-income countries**. *Journal of clinical microbiology* 2013, **51**(6):1740-1746.
65. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin J, Ahmed S, Alonso PL *et al*: **Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study**. *Lancet (London, England)* 2016, **388**(10051):1291-1301.
66. Pavlinac PB, Platts-Mills JA, Tickell KD, Liu J, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin MJ *et al*: **The clinical presentation of culture-positive and culture-negative, qPCR-attributable shigellosis in the Global Enteric Multicenter Study and derivation of a Shigella severity score: implications for pediatric Shigella vaccine trials**. *Clin Infect Dis* 2020.
67. Levade I, Saber MM, Midani FS, Chowdhury F, Khan AI, Begum YA, Ryan ET, David LA, Calderwood SB, Harris JB *et al*: **Predicting Vibrio cholerae Infection and Disease Severity Using Metagenomics in a Prospective Cohort Study**. *J Infect Dis* 2021, **223**(2):342-351.
68. Midani FS, Weil AA, Chowdhury F, Begum YA, Khan AI, Debela MD, Durand HK, Reese AT, Nimmagadda SN, Silverman JD *et al*: **Human Gut Microbiota Predicts Susceptibility to Vibrio cholerae Infection**. *J Infect Dis* 2018, **218**(4):645-653.
69. Midani FS, Weil AA, Chowdhury F, Begum YA, Khan AI, Debela MD, Durand HK, Reese AT, Nimmagadda SN, Silverman JD *et al*: **Human Gut Microbiota Predicts Susceptibility to Vibrio cholerae Infection**. *The Journal of infectious diseases* 2018.
70. Brun P, Scarpa M, Grillo A, Palu G, Mengoli C, Zecconi A, Spigaglia P, Mastrantonio P, Castagliuolo I: **Clostridium difficile TxAC314 and SLP-36kDa enhance the immune response toward a co-administered antigen**. *J Med Microbiol* 2008, **57**(Pt 6):725-731.

71. Ramamurthy T, Wagener D, Chowdhury G, Majumder PP: **A large study on immunological response to a whole-cell killed oral cholera vaccine reveals that there are significant geographical differences in response and that O blood group individuals do not elicit a higher response**. *Clinical and vaccine immunology : CVI* 2010, **17**(8):1232-1237.

72. Clemens JD, Sack DA, Harris JR, Chakraborty J, Khan MR, Huda S, Ahmed F, Gomes J, Rao MR, Svennerholm AM *et al*: **ABO blood groups and cholera: new observations on specificity of risk and modification of vaccine efficacy**. *The Journal of infectious diseases* 1989, **159**(4):770-773.

73. Alexandrova L, Haque F, Rodriguez P, Marrazzo AC, Grembi JA, Ramachandran V, Hryckowian AJ, Adams CM, Siddique MSA, Khan AI *et al*: **Identification of Widespread Antibiotic Exposure in Patients With Cholera Correlates With Clinically Relevant Microbiota Changes**. *J Infect Dis* 2019, **220**(10):1655-1666.

74. Harris JB, Podolsky MJ, Bhuiyan TR, Chowdhury F, Khan AI, Larocque RC, Logvinenko T, Kendall J, Faruque AS, Nagler CR *et al*: **Immunologic Responses to Vibrio cholerae in Patients Co-Infected with Intestinal Parasites in Bangladesh**. *PLoS neglected tropical diseases* 2009, **3**(3):e403.

75. Lind AL, Pollard KS: **Accurate and sensitive detection of microbial eukaryotes from whole metagenome shotgun sequencing**. *Microbiome* 2021, **9**(1):58.

76. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T *et al*: **A human gut microbial gene catalogue established by metagenomic sequencing**. *Nature* 2010, **464**(7285):59-65.