

SCIENTIFIC REPORTS

OPEN

A cross-sectional comparative study of gut bacterial community of Indian and Finnish children

Shreyas V. Kumbhare^{1,2}, Himanshu Kumar³, Somak P. Chowdhury^{2,4}, Dhiraj P. Dhotre², Akihito Endo⁵, Jaana Mättö⁶, Arthur C. Ouwehand⁷, Samuli Rautava^{3,8}, Ruchi Joshi⁹, Nitinkumar P. Patil¹⁰, Ravindra H. Patil¹, Erika Isolauri⁸, Ashish R. Bavdekar⁹, Seppo Salminen³ & Yogesh S. Shouche²

The human gut microbiome plays a crucial role in the compositional development of gut microbiota. Though well documented in western pediatrics population, little is known about how various host conditions affect populations in different geographic locations such as the Indian subcontinent. Given the impact of distinct environmental conditions, our study assess the gut bacterial diversity of a small cohort of Indian and Finnish children and investigated the influence of *FUT2* secretor status and birth mode on the gut microbiome of these populations. Using multiple profiling techniques, we show that the gut bacterial community structure in 13–14-year-old Indian (n = 47) and Finnish (n = 52) children differs significantly. Specifically, Finnish children possessed higher *Blautia* and *Bifidobacterium*, while genera *Prevotella* and *Megasphaera* were predominant in Indian children. Our study also demonstrates a strong influence of *FUT2* and birth mode variants on specific gut bacterial taxa, influence of which was noticed to differ between the two populations under study.

Microbial communities associated with the human gut respond and interact with host immune and digestive functions, thus contributing greatly to the overall well-being of the human body^{1–3}. Previous studies have shown that such microbe-human interactions begin as early as during the gestation period and evolve in a major way at birth, followed by multiple environmental challenges. Host genetics, dietary habits and lifestyle play key roles in compositional development of the microbiome^{2,4–15}. Conditions such as malnutrition in the early stages of life are known to delay the maturation of gut microbiome and thus to hamper the development of metabolic functions in the gut^{1,16,17}. Additionally, the precocious alterations in the gut microbiome are known to be the major contributors for the underlying mechanism of early age obesity^{18–20}. These alterations can lead to long-term consequences with increased risk of diseases in the later life. In summary, understanding the implications of development trajectory of gut microbiota in health and disease is necessary to determine the influence of such elements.

There is growing evidence to suggest the association of host genotype with the microbial community structure. A well-characterized association of this type in western population is *FUT2* gene polymorphism. This gene encodes for an enzyme fucosyltransferase-2 (α -1,2-fucosyl transferase) which is involved in synthesis of the ABO antigen^{21,22}. Single Nucleotide Polymorphism (SNP) in the exonic region of this gene is also known to be associated with certain gastrointestinal diseases such as Inflammatory Bowel Disease (IBD) and Crohn's disease^{23–25}. It is evident from few studies that this molecule is also secreted from mucosal surfaces such as the intestinal lining and may serve as the source of glycans and adhesion site to gut microbes like lactic acid bacteria^{26–28}. Very few

¹Department of Microbiology, R.C. Patel Arts, Science, and Commerce College, Shirpur, Dist. Dhule, Maharashtra, 425405, India. ²National Centre for Cell Science, Savitribai Phule University of Pune campus, Ganeshkhind, Pune, Maharashtra, 411007, India. ³Functional Foods Forum, Faculty of medicine, University of Turku, Turku, 20520, Finland. ⁴Max Planck Institute for Biogeochemistry, Jena, 07747, Germany. ⁵Department of Food and Cosmetic Science, Tokyo University of Agriculture, Hokkaido, Japan. ⁶Finnish Red Cross Blood Service, Kivihaantie 7, Helsinki, 00310, Finland. ⁷Active Nutrition, DuPont Nutrition & Health, Kantvik, 02460, Finland. ⁸Department of Paediatrics, University of Turku, Turku, 20520, Finland. ⁹King Edward Memorial hospital research centre, Pune, Maharashtra, 411011, India. ¹⁰Department of Microbiology, Smt. Chandibai Himathmal Mansukhani College, Ulhasnagar, Thane, Maharashtra, 421003, India. Shreyas V. Kumbhare and Himanshu Kumar contributed equally to this work. Correspondence and requests for materials should be addressed to S.S. (email: sepsal@utu.fi) or Y.S.S. (email: yogesh@nccs.res.in)

Description	Finnish	Indian
1. Sex (%)		
Male	27 (51.92%)	22 (46.80%)
Female	25 (48.07%)	25 (53.19%)
Total	52	47
2. Age (Mean \pm SD)		
	13.73 \pm 0.23	13.37 \pm 0.46
3. Mode of delivery (%)		
Vaginal	44 (84.6%)	33 (70.2%)
Cesarean	8 (15.3%)	14 (29.8%)
4. Secretor description (%)		
Secretor	45 (86.53%)	40 (85.10%)
Non secretor	7 (13.46%)	7 (14.89%)
5. Genotype (%)		
AA	7 (13.46%)	7 (14.89%)
GA	23 (44.23%)	18 (38.29%)
GG	22 (42.30%)	21 (44.68%)

Table 1. Demographic and genotypic characters of the subjects. The table describes demographic characteristics of the Finnish and Indian cohort along with the *FUT2* secretor status as determined by genotyping and information about the birth mode of subjects.

attempts have been made to assess the association of this polymorphism with specific gut bacterial taxa^{22, 26, 28, 29} and to our knowledge, no efforts have been undertaken to compare this effect in different geographical settings.

Furthermore, the early vaginal contact during normal delivery of an individual is a major source of microbial diversity followed by the microbiome acquired from the breast milk^{30–32}. Vaginal secretions ensure inoculation of the fetus at birth with specific microbial groups which carry out important functions necessary for normal development of the neonate^{11, 33, 34}. Previous studies indicate that the neonatal gut microbiome is strongly associated with the mode of delivery and impacts the immune system development of the neonate^{9, 35, 36}. The impetuous microbial disturbances introduced due to an altered mode of delivery are associated with increased risk of diseases^{37, 38}. Overall, the association of practices in the early life of infants, with the gut microbial alterations is now clear from the extensive studies carried out in the western cohorts; however, there is scarce information available about such influences in the Indian population.

Investigations worldwide have suggested differences in the gut microbial composition between distinct geographical locations or populations^{39, 40}. Also, international cohort studies suggest that inter-country variation in gut microbiota composition is significantly greater than inter-personal variations^{41–43}. Collectively, though well documented, what remains unknown is whether the effect of these host elements and environmental challenges that cause microbial composition shifts in the western population also hold true in a heterogeneous and ethnically distinct non-western population. The present work aims at comparing the gut microbiota of 13–14-year-old Indian and Finnish children in a pilot study cohort, and also to assess the influence of *FUT2* polymorphism and mode of delivery on the gut microbial composition of these populations.

Results

***FUT2* secretor status of Indian and Finnish children.** Genotyping of *FUT2* SNP rs601338 (W143X) revealed the secretor status of 99 subjects (52 Finnish and 47 Indian children). The study cohort included 45 and 40 secretors from Finnish and Indian population respectively, and 7 non-secretors each, from both the population. Secretor state GG/GA and non-secretor state AA were observed in both cohorts (Table 1).

Targeted quantification of specific microorganisms in the human gut. qPCR-based quantifications of 12 bacterial species (Supplementary Table S1) were carried out to determine their absolute abundance in the gut of Indian and Finnish children. Closer investigation of the heat map analysis demonstrated an overall higher abundance of bacterial species from *Clostridium* group in Indian children along with *Bifidobacterium bifidum* and *Bifidobacterium longum* group (Fig. 1a). Extended bar plot analysis revealed significant differences in 7 out of 12 bacterial species (Fig. 1b), with *Bifidobacterium longum* subsp. *infantis* and *Bifidobacterium animalis* subsp. *lactis* being significantly higher in Finnish children (Welch's *t*-test, $q < 0.001$) along with *Staphylococcus aureus* (Welch's *t*-test, $q < 0.001$) and *Akkermansia muciniphila* (Welch's *t*-test, $q < 0.001$). While, *Bifidobacterium bifidum*, *Clostridium coccoides* group, and *Clostridium leptum* were significantly higher in Indian children (Welch's *t*-test, $q < 0.001$). Distinct clades of Indian and Finnish samples in the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based dendrogram analysis also indicates differences in gut microbiota between these populations (See Fig. 1a). Principal component analysis showed distinct clustering of Indian and Finnish samples, while it also revealed that most of these bacterial taxa contributed to explaining the uniqueness of Finnish cohort, while only species from *Clostridium* group were found to be driving component for distinct clustering of the Indian cohort (Supplementary Fig. S1).

Differences in gut microbial community structure between Indian and Finnish children. PCR-DGGE was carried out using the total DNA extracted from fecal samples resulting in obtaining

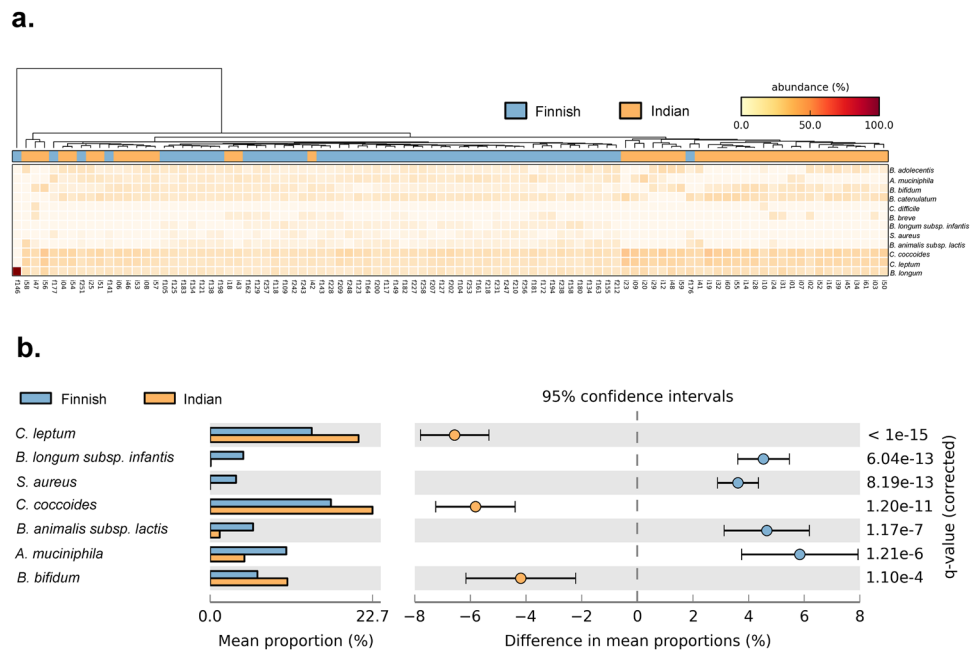


Figure 1. Real time quantification of bacterial species. **(a)** Heat Map is illustrating the extent of differences in the abundance of targeted bacterial species across samples from a Finnish and Indian population. **(b)** Extended bar plot is representing the differential abundance of significantly different bacterial species based on real time quantification data.

99 individual bacterial community fingerprints (standard strains used are listed in Supplementary Table S2). Further, this fingerprint-based dendrogram analysis demonstrated distinct clustering of two populations under study (Fig. 2a). Additionally, computer-assisted data matrix (distance matrix) was generated based on microbial community signature, to estimate various distances between samples and construct ordination plots using Principal Coordinate Analysis (PCoA). Ordination analysis performed on the DGGE data also revealed distinct clustering of Indian and Finnish subjects based on the prevalence of bands (Fig. 2b). Further, microbial diversity analysis performed using High Throughput Sequencing (HTS) suggested the presence of Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria and Verrucomicrobia as the most dominant phyla in both Finnish and Indian children (Fig. 3a). Bacterial composition at family level was dominated by *Bifidobacteriaceae* (34%) and *Lachnospiraceae* (32%) in Finnish subjects. While *Veillonellaceae* (29%) group was found to be the most dominant bacterial family in Indian children followed by *Bifidobacteriaceae* (23%) and *Prevotellaceae* (21%) (See Fig. 3b). Genera such as *Bifidobacterium* and *Blautia* were dominant in Finnish children, while the gut bacterial composition of Indian children was dominated by *Prevotella* and *Megasphaera*.

Variation in the gut microbial composition was observed in between Indian and Finnish children by both PCR-DGGE and 16S rRNA gene-based amplicon sequencing. Alpha diversity analysis showed that Finnish children harbored a marginally more diverse bacterial community than Indian children as indicated by the diversity measures such as observed species and Shannon index (Supplementary Table S3). Furthermore, phyla Actinobacteria (Welch's *t*-test, $q = <0.001$), Bacteroidetes (Welch's *t*-test, $q = <0.001$) and TM7 (Welch's *t*-test, $q = 0.026$) were differentially abundant in between these two populations (Supplementary Fig. S2a). Apart from few bacterial families (Supplementary Fig. S2b), genera such as *Bifidobacterium*, *Blautia*, *Megasphaera*, *Prevotella*, *Ruminococcus*, *Dorea*, and *Coprococcus* were found to be differentially abundant in these two populations (Supplementary Fig. S2c). Of note, bacteria belonging to genera *Bifidobacterium* (Mann-Whitney U-test, $p = <0.0013$) and *Blautia* (Mann-Whitney U-test, $p = <0.001$) were significantly higher in Finnish children, while Indian children displayed significantly higher abundance of *Prevotella* (Mann-Whitney U-test, $p = <0.001$) and *Megasphaera* (Mann-Whitney U-test, $p = <0.001$) (See Fig. 4a). Beta diversity analysis based on the Non-metric Multidimensional Scaling (NMDS) method demonstrated distinct grouping of gut bacterial communities of Indian and Finnish children (Fig. 4b). Additionally, a total of 213 genera (63.6%) were found to be common amongst Indian and Finnish children, while 65 genera (19.4%) and 57 genera (17%) were unique in Finnish and Indian children respectively (Fig. 4c).

Our investigation on the prevalence (presence/absence) of different bacterial groups gave deeper insight into the gut bacterial community structure of Indian and Finnish children (Supplementary Table S4). It was observed that phyla such as Chloroflexi and Thermi were prevalent in almost all Indian children as compared to the Finnish population (Welch's *t*-test, $p = <0.001$) (Supplementary Table S4a). Likewise, a comparison between Indian and Finnish gut bacterial community showed that genera such as *Acidaminococcus*, *Catenibacterium*, *Fusobacterium*, *Megasphaera*, *Mitsuokella*, and *Slackia* were significantly prevalent in Indian children (Welch's *t*-test, $p = <0.001$). While, genera such as *Adlercreutzia*, *Anaerofustis*, and *Coprobacillus* were highly prevalent in Finnish (Welch's

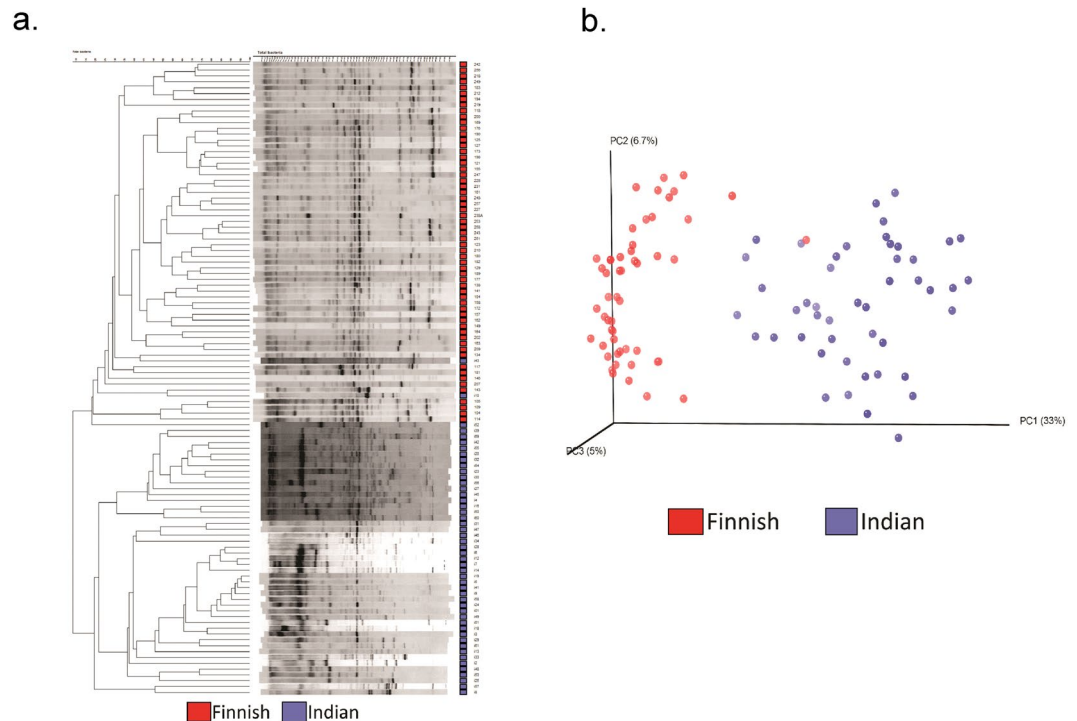


Figure 2. DGGE based diversity analysis. **(a)** UPGMA Cluster analysis based on PCR-DGGE analysis fingerprints from all samples. Finnish samples (red) and Indian samples (purple). **(b)** PCoA plot using Euclidean distances based on the prevalence of bands, representing differences in diversity between two populations: Finnish samples (red) and Indian samples (purple).

t -test, $p = <0.001$) as compared to Indian children (Supplementary Table S4c). Together, these results suggested that there are substantial differences in gut microbial community structure of Indian and Finnish children.

Influence of *FUT2* secretor status on gut microbial composition. Our observations revealed that the gut microbiota composition is strongly associated with *FUT2* secretor status of an individual. Principal component analysis revealed that the influence pattern of *FUT2* variants differs substantially within the two populations (Fig. 5a), suggesting that different set of gut bacterial taxa are affected in a particular population as compared to the other. Phyla such as Actinobacteria and Verrucomicrobia were significantly higher in both Indian and Finnish secretor children (Welch's t -test, $q = <0.001$), while Firmicutes and Proteobacteria were abundant in the non-secretor children in both the population (Welch's t -test, $q = <0.001$) (Supplementary Fig. S3a and c). However, Bacteroidetes were found higher in Indian secretors (Welch's t -test, $q = <0.001$), whereas they were observed to be higher in Finnish non-secretors (Welch's t -test, $q = <0.001$).

The abundance of bacterial families such as *Bifidobacteriaceae*, *Ruminococcaceae*, *Verrucomicrobiaceae* and *Enterobacteriaceae* (Welch's t -test, $q = <0.001$) exhibited a significant association with the secretor status of children in both populations (Supplementary Fig. S4a and c). *Bifidobacteriaceae* and *Verrucomicrobiaceae* were significantly abundant (Welch's t -test, $q = <0.001$) in secretors as compared to non-secretors, while *Veillonellaceae* and *Enterobacteriaceae* were found at significantly higher levels (Welch's t -test, $q = <0.001$) in non-secretors. Additionally, bacterial families like *Coriobacteriaceae*, *Prevotellaceae* and *Clostridiaceae* were highly abundant in secretors when compared to non-secretors of the Indian population (Welch's t -test, $q = <0.001$), while non-secretors harbored higher *Planococcaceae* as compared to secretors (Welch's t -test, $q = <0.001$) within Indian cohort. *Lachnospiraceae* group was highly significant in secretor children (Welch's t -test, $q = <0.001$) within Finnish cohort (Supplementary Fig. S4a and c).

Further investigations to understand the effect of *FUT2* secretor status on different bacterial genera demonstrated that within the Finnish population, *Coprococcus*, *Succinilasticum*, *Streptococcus*, and *Blautia* were highly abundant in secretor children (Welch's t -test, $q = <0.001$), while bacteria belonging to genus *Lactococcus* were highly abundant in non-secretors (Welch's t -test, $q = <0.001$). Similarly, genera like *Megasphaera*, *Veillonella*, *Rummeliibacillus*, and *Megasphaera* were abundant in Indian children with non-secretor status (Welch's t -test, $q = <0.001$); However the genera *Prevotella*, *Roseburia* and *Clostridium* were significantly higher in secretors (Welch's t -test, $q = <0.001$). Bacterial genera such as *Bifidobacterium*, *Akkermansia*, and *Ruminococcus* were observed to have higher abundance in secretors, whereas non-secretors of both Indian and Finnish cohort showed higher abundance of *Escherichia*. (Supplementary Fig. S4b and d). Collectively, this analysis indicated that *FUT2* secretor status did influence the gut microbial composition in both the cohorts; the pattern of which differed between these populations.

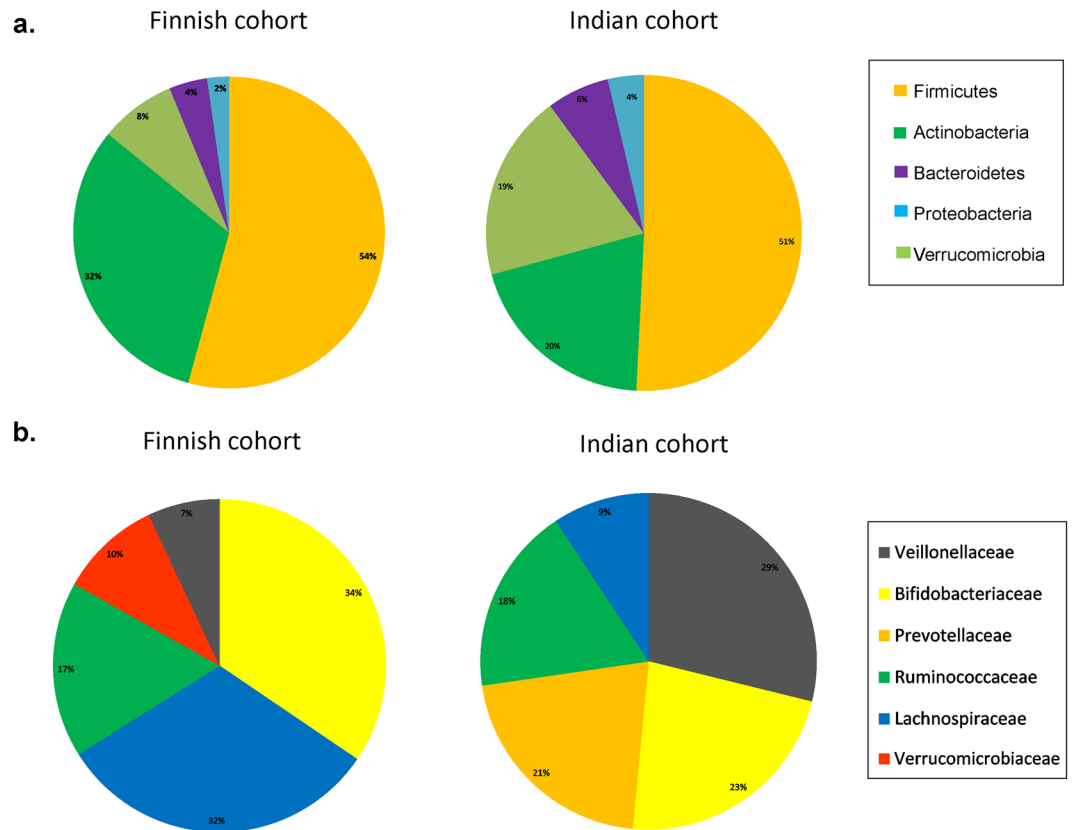


Figure 3. Gut bacterial community composition. (a) Phylum level distribution of dominant gut bacterial groups within Indian and Finnish children. (b) Family level distribution of dominant gut bacterial groups within Indian and Finnish children.

Influence of Mode of delivery on gut microbial composition. Gut bacterial communities were also observed to have a significant association with the mode of delivery of an individual. We also performed principal component analysis to assess whether the influence of birth mode on the gut microbial composition of Indian children was similar to that of Finnish children. Our analysis showed substantial differences in the pattern, suggesting that different set of taxa are affected due to birth mode variants (cesarean VS. vaginal) between these populations (see Fig. 5b). Amongst the significantly affected bacterial phyla, Bacteroidetes and Firmicutes were higher (Welch's t -test, $q = <0.001$) in vaginally born children in both Indian and Finnish population, while Proteobacteria and Verrucomicrobia were seen to be more abundant (Welch's t -test, $q = <0.001$) in cesarean born children. It was noted that the phylum Actinobacteria was significantly abundant (Welch's t -test, $q = <0.001$) in cesarean born as compared to vaginally born Finnish children (Supplementary Fig. S3b and d).

Furthermore, mode of delivery had a significant impact on bacterial families such as *Enterobacteriaceae*, *Verrucomicrobiaceae* (abundant in Cesarean born, Welch's t -test, $q = <0.001$), *Bacteroidaceae*, *Veillonellaceae* and *Streptococcaceae* (abundant in vaginally born, Welch's t -test, $q = <0.001$) in Finnish and Indian children as well (Supplementary Fig. S5a and c). Genera such as *Escherichia*, *Akkermansia*, *Bacteroides*, and *Streptococcus* had a similar impact of birth mode within both the cohorts under study. Former two being higher in cesarean born, while the others abundant in vaginally born children. Conversely, distinct influence on abundance of genera *Bifidobacterium* and *Faecalibacterium* were noted (Supplementary Fig. S5b and d) between these populations. Also, genera like *Dialister*, *Succiniclacticum*, *Ruminococcus*, and *Blautia* were abundant in vaginally born children as compared to cesarean within Finnish cohort (Welch's t -test, $q = <0.001$). It was also noted that vaginally delivered Indian children harbored higher abundance of *Megamonas*, *Veillonella*, *Rummeliibacillus*, *Clostridium*, *Megasphaera*, and *Collinsella*, while cesarean born Indian children harbored higher *Lactobacillus* as compared to their respective counterparts (Welch's t -test, $q = <0.001$). Together, these observations showed that birth mode had an influence on the gut microbial composition. However, the pattern differed when compared between Indian and Finnish population.

Differences in the predicted functions of gut bacterial community. We further analyzed the microbial community data to predict their metabolic functions. This analysis revealed significant differences between the functional profile of Indian and Finnish gut bacterial community as illustrated in Fig. 6. The PCoA plot (see Fig. 6a) represents distinct clustering of Finnish and Indian samples based on the predicted metabolic functions of the respective bacterial community. Also, the microbiota of Indian children was predicted to have significantly higher genes involved in glycan biosynthesis and metabolism, cell motility, lipopolysaccharide (LPS) biosynthesis

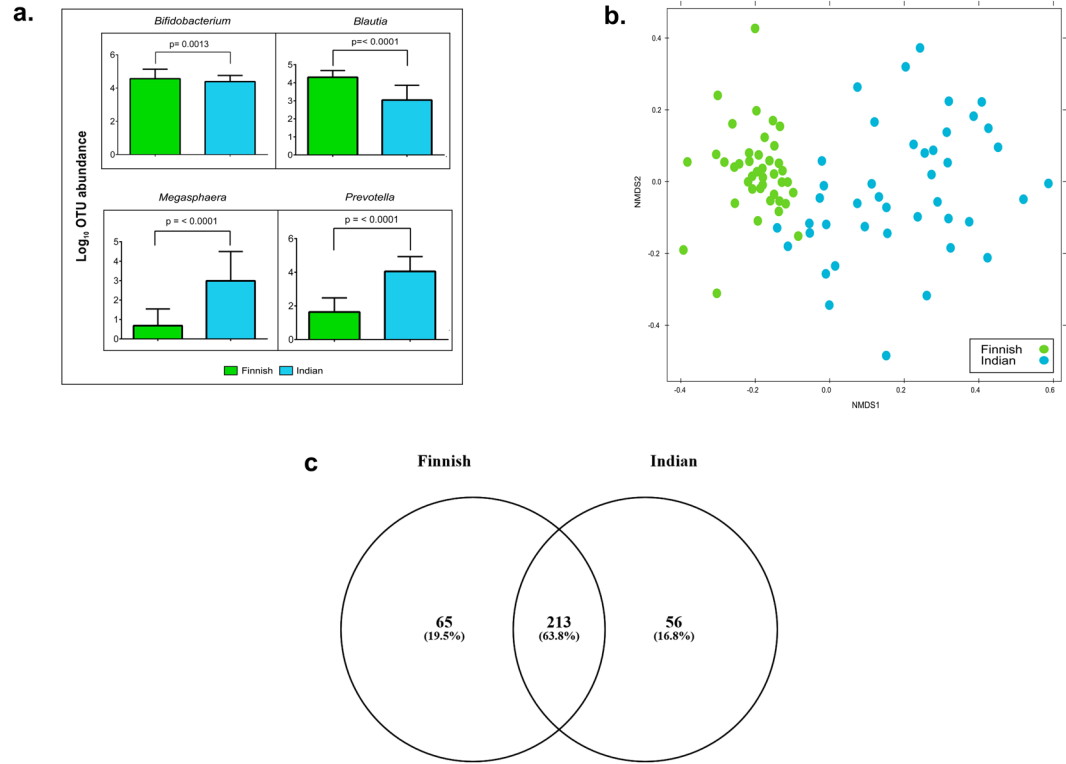


Figure 4. Beta diversity analysis. **(a)** Box plot illustrating differences in relative abundances of key contributors (most dominant) of bacterial community between Indian (blue) and Finnish children (green). **(b)** Beta diversity analysis using NMDS plot based on OTU level differences between Finnish (green) and Indian (blue) children **(c)** Venn diagram representing shared and unique bacterial genera between the gut bacterial community of Finnish and Indian children.

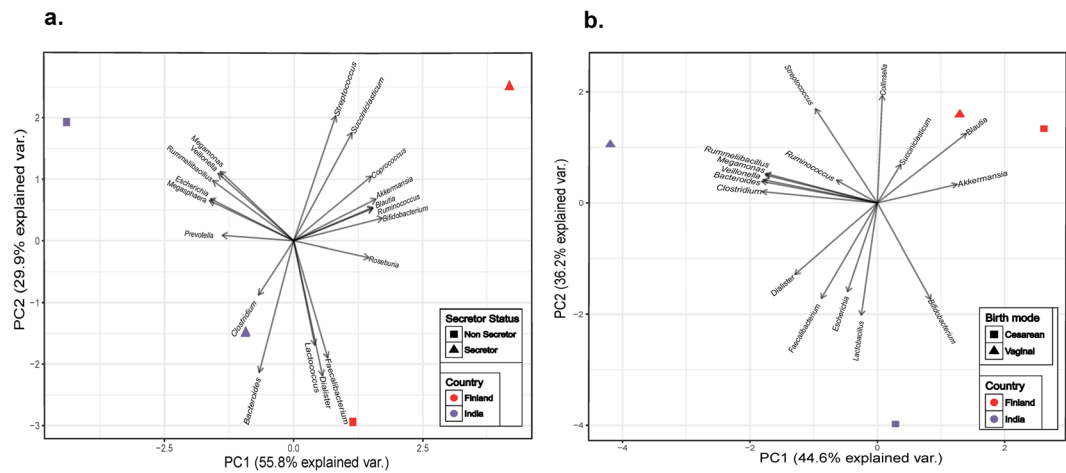


Figure 5. Principal Component Analysis. PCA was performed using the bacterial genera data; relative abundances were observed to be differing with *FUT2* secretor state and birth mode in both the populations (Supplementary Figs S4 and S5). Results were plotted according to the PC1 and PC2 scores with the percent variation explained by respective axis. Specific colors were used to indicate the region and different shapes to indicate either the secretor state **(a)** or birth mode **(b)**. Arrows indicate bacterial genera with their names.

proteins (Kruskal-Wallis test, Linear Discriminant analysis (LDA) score effect size filter = 2). Likewise, Finnish gut microbiota had significantly higher genes contributing to processes such as Carbohydrate metabolism, Methane metabolism, and transport systems (Kruskal-Wallis test, LDA score effect size filter = 2) as represented in Fig. 6b. Additional information of differentially abundant genes between Indian and Finnish population is illustrated in Supplementary Fig. S6.

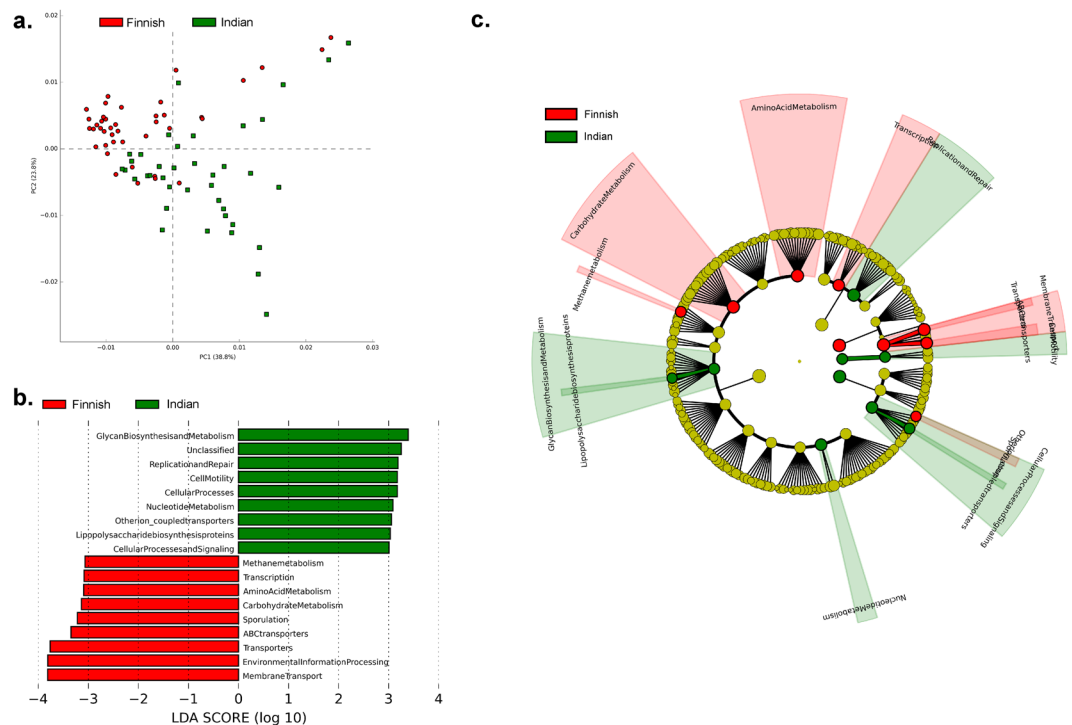


Figure 6. Imputed metagenomic functions of the gut bacterial community. **(a)** PCA plot is illustrating differences in metabolic functions of the gut bacterial community of Finnish (red) and Indian (green) children. **(b)** Bar chart is representing specific metabolic function differential abundant in Finnish and Indian children (Effect size filter = >2.0). **(c)** Cladogram representing differences in abundance of microbial genes between the gut bacterial community of Finnish and Indian children contributing in specific metabolic pathways.

Furthermore, cladogram analysis based on KEGG orthologs using Linear discriminative analysis Effect Size (LEfSe) also supported the previous observation of distinct metabolic profiles of Indian and Finnish children (see Fig. 6c). We directed our investigation to decipher association of *FUT2* secretor status and birth mode with functional profiles of the bacterial community. This approach revealed significant association of secretor status and abundance of genes involved in starch and sucrose metabolism (higher in secretors, Welch's *t*-test, $q < 0.001$), lipopolysaccharide biosynthesis proteins and porphyrin and chlorophyll metabolism (more abundant in non-secretors, Welch's *t*-test, $q < 0.001$) amongst Indian and Finnish children (see Fig. 7a and b). Moreover, genes contributing to methane metabolism and synthesis of Branched Chain Amino Acids (BCAA): valine, leucine, and isoleucine were found to be higher in secretors as compared to the non-secretor status of Finnish cohort (Welch's *t*-test, $q < 0.001$). Beta-Alanine metabolism and glyoxylate and dicarboxylate metabolism genes were found to be higher in non-secretors as compared to the secretor children of Indian cohort (Welch's *t*-test, $q < 0.001$).

The impact of mode of delivery on the abundance of carbohydrate metabolism (fructose, mannose, sucrose and starch metabolism, etc.) genes differed between the two populations. Interestingly, these genes which were higher in vaginally born Finnish children, were observed to be abundant in cesarean born Indian children (see Fig. 7c and d). However, mode of delivery had a similar effect on the functional features of biosynthesis of unsaturated fatty acids (observed highly abundant in cesarean born children, Welch's *t*-test, $q < 0.001$) in both the cohorts under study. Also, these factors had a significant influence on other metabolic features of bacterial community as illustrated in Fig. 7a–d. Altogether, these results indicate that the metabolic profiles of gut microbiota were substantially different amongst two populations and also between *FUT2* and birth mode variants within the population.

Discussion

In this cross-sectional comparative study, we found that the gut microbial composition of Finnish and Indian children differed significantly, with a distinct set of bacterial taxa being predominant in each population. Our investigation further reveals the association of *FUT2* and birth mode variants with gut microbial composition at 13–14 years of age, patterns of which were discrete in the two geographically and ethnically distinct populations considered in this study.

We quantified a select group of bacterial species which represent the early colonizers of the gut using qPCR, and were previously reported to be associated with factors such as nutrition, allergy, and immune-modulation in the western population^{44, 45}. We evaluated their prevalence in the gut microbiome at 13–14 years of age in both populations, which provided us primary leads about differences in the gut microbial community structure between the two populations. Interestingly, it was observed that few *Bifidobacterium* species were present in

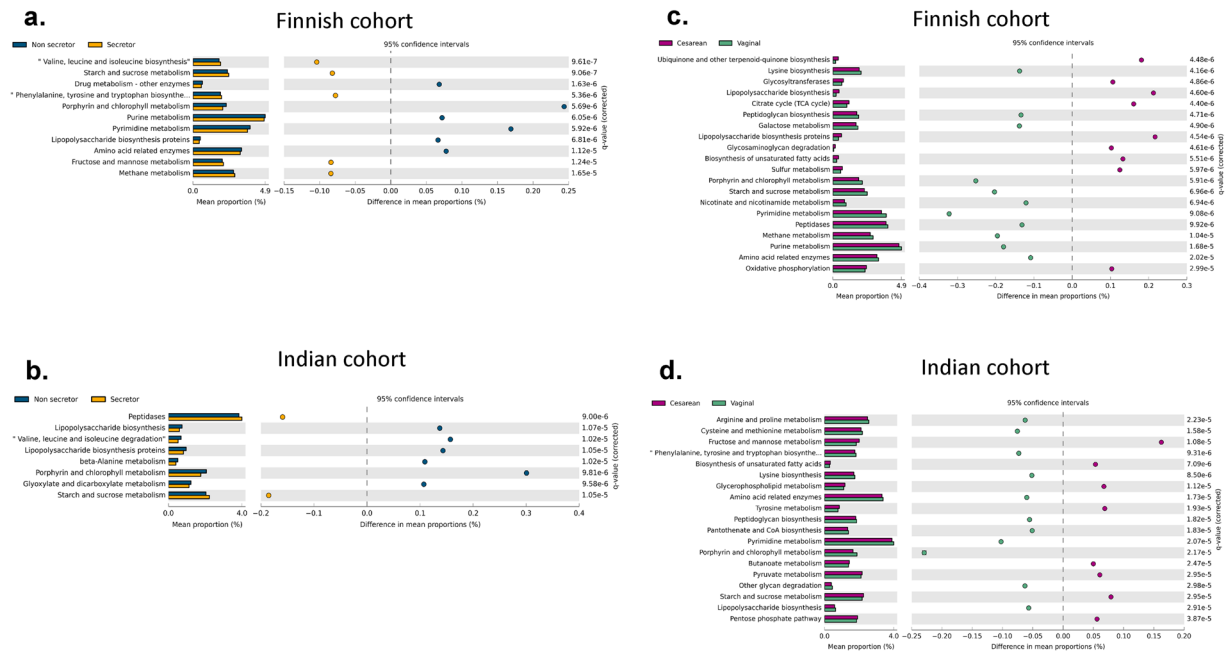


Figure 7. Effect of *FUT2* secretor status and birth mode on the pattern of microbial metabolic functions. (a and b) Extended bar plots are representing significantly different KEGG orthologs associated with microbial metabolic functions between secretor and non-secretor children of the Finnish and Indian population respectively. (c and d) Extended bar plots are representing significantly different KEGG orthologs associated with microbial metabolic functions between vaginally born and cesarean born children of the Finnish and Indian population respectively. Appropriate effect size filter was employed to represent highly affected features.

higher number in one population, while lower in another. These observations could be related to the differences in their genomes encoding for varied, complex polysaccharide degrading enzymes (CAZymes)^{46–48}. Further, the abundance of *Akkermansia*, a known mucin degrader in the intestine, was also found to be substantially higher in Finnish children. This bacterium is shown to possess a strong negative correlation with high-fat diet^{49–51}, indicating probable differences in the fat content consumed by these populations. Thus, these findings are indicative of the fact that such common intestinal colonizers are responsive to the distinct dietary habits across the globe, playing a key role in shaping the gut microbiome^{7, 39}. We also performed principal component analysis, which indicated that out of the targeted bacteria in the qPCR experiment, only a few i.e. *Clostridium* species could explain the variation and distinct clustering of the Indian population. We further tried to correlate the qPCR data with that obtained from the 16S rRNA gene-based microbial profiling. Although not significant (except for *Bifidobacterium*), similar differential trend in the abundances of these genera was observed in the HTS data. This unsubstantial discrepancy could be attributed to the technical differences, such as the use of species-specific primers in qPCR, as against universal bacterial primers in the HTS. Collectively, this integrated analysis of qPCR and HTS data suggested an urge for exploring such untapped and functionally important taxa in the Indian population.

Our analysis based on the PCR-DGGE and HTS clearly demonstrated the extent of distinctness and diversity in the gut microbial composition of children of similar age but from distinct populations. The outcomes of our multi-pronged approach suggests that the gut microbial communities harbored by Finnish and Indian children are not only distinct but are driven by specific bacterial groups such as *Prevotella*, *Megasphaera* (Indian subjects), *Blautia* and *Bifidobacterium* (Finnish subjects). It also provided important information about the differential abundances of genera such as *Blautia* and *Ruminococcus*, increased numbers of which are reported to be induced by whole grain consumption⁵². Further, these differences were also reflected in the predicted functions of gut microbial community when compared between populations. Thus such variations in the microbial composition observed through qPCR and HTS, as well as in their imputed metabolic functions could be collectively attributed to the response of dietary and genetic variances between these two geographically distinct countries.

The key caveat that needs to be considered in this study is the sample size (especially for *FUT2* non-secretors) which doesn't allow generalization about the entire population. However, this study gives us clues about differences within the same age groups and especially in growing children. This pilot study has generated a preliminary report which already shows promise regarding the impact that these factors have on the intestinal microbiota. Given the total populations of the two countries being considered in this “Pilot study” and the proportion of non-secretors globally (~30% with minor allele frequency $A = 0.3217$ globally; www.ncbi.nlm.nih.gov/variation/tools/1000genomes), extensive sampling would be needed for generalizing these findings at population level.

The sample size conundrum led us to further employ a statistically robust approach to determine whether similar influence patterns of host elements exist in two distinct populations. Our analysis revealed a strong association of gut microbial composition with secretor status of the subject amongst both the populations. According

to previous studies, most of the taxa that were found higher in secretor subjects, are known to harbor enzymatic machinery like α -L-fucosidases and β -galactosidases which enable these bacterial groups to degrade and utilize the glycan chains of the blood group antigens expressed in the intestine^{53,54}. Besides, the exact reason for the higher abundance of genera such as *Streptococcus* (Finnish children) and *Clostridium* (Indian children) in secretor subjects is unclear but could be concomitant to their ability to use the human mucin glycans and compete for intestinal colonization^{48,54}. A key result of this study is the association of organisms other than LAB (as reported before) with the *FUT2* genotype. Interestingly, it was noted that the genes encoding for lipopolysaccharide (LPS) biosynthesis were significantly more abundant in non-secretors in both the populations. Previous studies suggest that high level of LPS production triggers the immune responses leading to systematic inflammation and dysbiosis in the gut⁵⁵. These observations collectively indicate that *FUT2* polymorphism (non-secretor status) contributes to the selection of bacterial community and in turn modulates the innate immune responses in the gut, which makes an individual more prone to inflammation associated diseases like IBD, obesity, T2D^{5,23–25}. However, from our investigations, it was noted that the pattern of influence of *FUT2* variants on gut microbiota differed amongst the Indian and Finnish children. These differences in the influence could be related to the distinct dietary habits and environmental conditions which are known to play a key role in shaping the gut microbial composition, as also described by Kashyap *et al.*⁵⁶. Despite the fact that mode of delivery is associated with specific gut bacterial groups in the neonates, very few studies have been directed towards understanding this association in the later life^{30–32,57} and comparing those effects amongst differing geographical locations. Our efforts towards understanding this association in 13–14 years children (age of adolescence) revealed that there is a long-term effect of mode of delivery on the gut microbiome of an individual. Of particular note are higher abundances of Actinobacteria and Proteobacteria in cesarean born children. Similar patterns are reported in the patients suffering from gastrointestinal diseases such as IBD⁵⁸, suggesting that cesarean born children may have increased risk of such diseases due to microbial alterations. As also for *FUT2* secretor state, differences were also reflected in the predicted functional profiles about the birth mode variants in both the populations. However, the pattern of influence differed substantially between the two populations as clearly evident from the principal component analysis. Our observations suggest that such variances in pattern indicate the role of gut microbial alterations, caused due to discrete environmental challenges faced by the mother and child during birth (delivery) in these two populations. Further, the nature of microbiota of the Indian subcontinent suggests more extensive studies being needed before concluding if factors such as secretor status and birth mode can partition the population in a meaningful way.

Sample size is an important caveat of this study which could have masked the differences occurring within low abundance taxa between groups. We, therefore, suggest that future comparisons of this type will require power analysis (for estimated sample size) taking into consideration factors investigated here. Despite these shortfalls, our preliminary study using such factors does reveal for the first time differences in gut microbial composition in these distinct populations. Factors such as age, socio-economic status, and ethnicity which are a substantial part of environmental variables remain largely consistent within each cohort thus reducing confounding effects on observed differences.

Altogether, we hypothesize that there exists a possibility that well-documented knowledge obtained about how exactly *FUT2* secretor state and birth mode influence the gut microbial community, may not be valid across different populations. Our findings emphasize the need for careful considerations of these differences during comparison and also in the view of designing any future studies, specifically those involving international intervention techniques aiming at remission of microbial alterations caused due to such host and environmental elements. It thus highlights the need for thorough understanding of the influence of such associations in a particular population with extensive sampling, especially in a heterogeneous community such as India.

In summary, we demonstrated that there are differences in the gut microbial community of 13–14-year-old children of the Indian and Finnish population. We also showed that there are differences in the mechanism by which host genotype and other elemental host factors such as birth mode influence the gut microbiome in distinct populations. Also, our findings underscore the importance of such comparisons in relation with these key determinants (*FUT2* and birth mode variants) involved in shaping the gut microbiome.

Methods

Sample collection. A total of 99 children (Indian, $n = 47$ and Finnish, $n = 52$), of age 13–14 years old, with no indications of infections during the sampling time and with a medication history of no antibiotic consumption for three months preceding sample collection, were recruited for the study. The study was approved by the ethics committees of National Centre for Cell Science, King Edward's Memorial Hospital research Centre and Turku University Hospital. Written informed consent from the parents/guardians of the participants was obtained using the methods approved according to the guidelines of respective centers. All methods and experiments were performed by following relevant guidelines and regulations. A detailed questionnaire was designed and was addressed by the parents/guardians of the participants. The demographic and general characteristics of the cohorts under study are described in Table 1. The questionnaire included information about the medical and medication history etc. Fecal and EDTA anticoagulated whole blood samples were collected and stored at -80°C until further processing.

DNA extraction. Fecal and blood samples collected from the subjects were processed for total DNA extraction using QIAmp Mini stool DNA extraction and Blood & Tissue DNA extraction kit (Qiagen, USA) following the manufacturer's instructions. Extracted DNA samples were quantified using NanoDrop spectrophotometer ND1000 (Thermo Scientific, USA).

SNP analysis. Single Nucleotide Polymorphism (SNP) analysis to determine the *FUT2* SNP rs601338 was carried out by using the genotyping method. Briefly, the exon region (exon 2) of *FUT2* gene was amplified using

specific primers as described previously²⁸. Sequencing of the amplicon was carried out on ABI 3730 XL sequencer (Applied Biosystems, USA). Mutation analysis was carried out using SeqScape 2.6.0 (Applied Biosystems, USA) and heterozygous bases were detected (see details in Supplementary Table S5).

DGGE based Bacterial community fingerprinting. DGGE of the DNA extracted from fecal samples was carried out as described in previous studies to obtain a bacterial community fingerprint¹⁵. PCR primers F357: 5'-GCCGCCCGCCG

GCCCCGCGCCCGGCCCGCCGCCGCCCGCCCTACGGGAGGCAGCAG-3' and R518: 5'-ATTACCGCGGCTGCTGG-3' were used as described previously^{15,59} to target the total bacterial community with a gradient range of 35–70%. The standard bacterial strains used are listed in Supplementary Table S2. The gel images were processed using Bionumerics tool version 6.6 (Applied Maths, Belgium) and were analyzed as described by Kumar *et al.*, 2013²⁶. Furthermore, a dendrogram was constructed for comparison of the total bacterial community between Indian and Finnish subjects as described by Endo *et al.*¹⁵. Euclidean distances based on the data obtained from Bionumerics were calculated to construct PCoA plots.

Quantification of specific bacterial species. qPCR-based quantification of 12 bacterial groups from the fecal samples was carried out as described previously by Collado *et al.*⁶⁰. List of bacterial species along with the primers used is presented in Supplementary Table S1.

Library preparation and high throughput sequencing. DNA extracted from the fecal samples was used to amplify the V4 region of 16S rRNA gene to determine the gut bacterial community structure. Primer set 515 F/806 R was employed to target the V4 region as described in earlier studies^{61,62}. The amplified products were further subjected to library preparation and sequencing on the Illumina HiSeq. 2000 platform as per the manufacturer's instructions (Illumina technologies, USA). Sequencing was performed using V4 paired-end (2×250 bp) chemistry. The paired end reads obtained, were quality checked using the FastQC tool [Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>] and assembled using FLASH assembler tool⁶³. Details of the sequencing are added in Supplementary Table S5.

Community structure analysis. The assembled reads were used for community structure analysis using QIIME pipeline⁶⁴. OTU picking method was carried out using UCLUST closed reference method⁶⁵, and the representative OTUs were assigned taxonomy using RDP classifier method⁶⁶ with GreenGenes (13.8 release) as reference dataset⁶⁷. Alpha and beta diversity analysis were performed, and further statistical analysis was carried out using R and STAMP⁶⁸. Normalization of the data was carried out using rarefaction step included in QIIME package. A detailed description of the normalization carried out for various comparisons is given in the statistical methods and analyses section.

Prediction of metabolic functions of the bacterial community. Metagenomic prediction tool 'Phylogenetic Investigation of Communities by Reconstruction of Unobserved States' (PICRUSt)⁶⁹ was utilized to determine the metabolic functions of the bacterial community. Nearest Sequence Taxonomic Identity (NSTI) values were obtained for each sample and samples with NSTI values more than 0.10 were excluded from further analysis to avoid false predictions as described by Langille *et al.* 2013 (which suggest the NSTI values for human samples to be lesser than 0.10 with higher prediction accuracy)⁶⁹. Metagenomic predictions were obtained based on KEGG Orthology (KO) and were further analyzed using LEfSe⁷⁰ to obtain differentially abundant genes, especially those involved in various metabolic processes of the bacterial communities. This approach was used to carry out comparisons of imputed metabolic functions of the bacterial community, between the Finnish and Indian population and also to decipher effect of *FUT2* secretor status and birth mode on metabolic functions within these populations.

Statistical methods and analyses. UPGMA based dendrogram analysis using Pearson similarity coefficient was performed on the PCR-DGGE fingerprinting data, while UPGMA based dendrogram analysis was carried out using STAMP for qPCR data. Euclidean distance based matrices were used to construct PCoA plots for DGGE data using QIIME. Statistical packages provided in the STAMP tool were utilized to determine the differences in the bacterial community as well as the imputed metabolic functions within groups under consideration. To overcome the sample size bias during the comparisons carried out between secretor vs. non-secretor and vaginally born vs. cesarean born children for both the populations, a randomization approach was used. This approach included consolidating the OTU data of all the samples from a particular group (e.g. secretor or non-secretor) thus considering the total obtained set as one microbial community representing a single group. Furthermore, normalization was carried out against the other respective group using three iterations of randomization (triplicate), and data for all the replicates obtained was further used for comparison. Welch t-test (two-sided) with Benjamini-Hochberg FDR correction (q -value filter = 0.05 and appropriate effect size filter) was applied for multiple comparisons considering the unequal variances of the study groups. Bar graph analysis was carried out using the non-parametric Mann-Whitney test in GraphPad prism for evaluating significant differences in abundances of specific genera between two groups. Furthermore, NMDS plots based on Bray-Curtis distance matrix were constructed to carry out beta diversity analysis. Factorial Kruskal-Wallis test ($\alpha = 0.05$) was employed for inter-class comparisons with appropriate logarithmic LDA score filter using LEfSe.

Data availability. The sequence data is made available at NCBI SRA submission with accession number SRP077882 (Bioproject ID PRJNA327770).

References

- Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome, and immune system: envisioning the future. *Nature* **474**, 327–336 (2012).
- Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313–323 (2009).
- Thaiss, C. A., Zmora, N., Levy, M. & Elinav, E. The microbiome and innate immunity. *Nature* **535**, 65–74 (2016).
- Bonder, M. J. *et al.* The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412 (2016).
- Brown, K., DeCoffe, D., Molcan, E. & Gibson, D. L. Diet-induced dysbiosis of the intestinal microbiota and the effects on immunity and disease. *Nutrients* **4**, 1095–1119 (2012).
- Chong, C. W. *et al.* Effect of ethnicity and socioeconomic variation to the gut microbiota composition among pre-adolescent in Malaysia. *Sci. Rep.* **5**, 13338 (2015).
- Conlon, M. A. & Bird, A. R. The Impact of Diet and Lifestyle on Gut Microbiota and Human Health. *Nutrients* **2015**, 17–44 (2015).
- Collado, M. C., Rautava, S., Aakko, J., Isolauri, E. & Salminen, S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci. Rep.* **6**, 23129 (2016).
- Romano-Keeler, J. & Weitkamp, J.-H. Maternal influences on fetal microbial colonization and immune development. *Pediatr. Res.* **77**, 189–195 (2015).
- Diaz Heijtz, R. Fetal, neonatal, and infant microbiome: Perturbations and subsequent effects on brain development and behavior. *Semin. Fetal Neonatal Med.* **46**, 4–11 (2016).
- Neu, J. The microbiome during pregnancy and early postnatal life. *Semin. Fetal Neonatal Med.* **21**, 373–379 (2016).
- Nuriel-Ohayon, M., Neuman, H. & Koren, O. Microbial changes during pregnancy, birth, and infancy. *Front. Microbiol.* **7**, 1–13 (2016).
- Chu, D. M. *et al.* The early infant gut microbiome varies in association with a maternal high-fat diet. *Genome Med.* **8**, 77 (2016).
- Rodríguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* **26**, 26050 (2015).
- Endo, A., Prtty, A., Kalliomäki, M., Isolauri, E. & Salminen, S. Long-term monitoring of the human intestinal microbiota from the 2nd week to 13 years of age. *Anaerobe* **28**, 149–156 (2014).
- Kane, A. V., Dinh, D. M. & Ward, H. D. Childhood malnutrition and the intestinal microbiome. *Pediatr. Res.* **77**, 256–62 (2015).
- Voreades, N., Kozil, A. & Weir, T. L. Diet and the development of the human intestinal microbiome. *Front. Microbiol.* **5**, 1–9 (2014).
- Gomez-Arango, L. F. *et al.* Connections between the gut microbiome and metabolic hormones in early pregnancy in overweight and obese women. *Diabetes* **65**, 2214–2223 (2016).
- Turta, O. & Rautava, S. Antibiotics, obesity and the link to microbes - what are we doing to our children? *BMC Med.* **14**, 1–6 (2016).
- Cox, L. M. & Blaser, M. J. Antibiotics in early life and obesity. *Nat Rev Endocrinol* **73**, 389–400 (2015).
- Kelly, R. J., Rouquier, S., Giorgi, D., Lennon, G. G. & Lowe, J. B. Sequence and expression of a candidate for the human Secretor blood group alpha-(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. *Journal of Biological Chemistry* **270**, 4640–4649 (1995).
- Wacklin, P. *et al.* Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. *PLoS One* **9** (2014).
- Parmar, A. S. *et al.* Association study of FUT2 (rs601338) with celiac disease and inflammatory bowel disease in the Finnish population. *Tissue Antigens* **80**, 488–493 (2012).
- McGovern, D. P. B. *et al.* Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum. Mol. Genet.* **19**, 3468–3476 (2010).
- Hu, D. *et al.* Association of Ulcerative Colitis with FUT2 and FUT3 Polymorphisms in Patients from Southeast China. *PLoS One* **11**, e0146557 (2016).
- Kumar, H. *et al.* Secretor status is strongly associated with microbial alterations observed during pregnancy. *PLoS One* **10**, 1–10 (2015).
- Uchida, H. *et al.* Lactic acid bacteria (LAB) bind to human B- or H-antigens expressed on intestinal mucosa. *Biosci. Biotechnol. Biochem.* **70**, 3073–6 (2006).
- Wacklin, P. *et al.* Secretor genotype (FUT2 gene) is strongly associated with the composition of bifidobacteria in the human intestine. *PLoS One* **6**, e20113 (2011).
- Davenport, E. R. *et al.* ABO antigen and secretor statuses are not associated with gut microbiota composition in 1,500 twins. *BMC Genomics* **17**, 1–14 (2016).
- Madan, J. C. *et al.* Association of Cesarean Delivery and Formula Supplementation With the Intestinal Microbiome of 6-Week-Old Infants. *JAMA Pediatr.* **170**, 1–8 (2016).
- Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. USA* **107**, 11971–5 (2010).
- Milani, C. *et al.* Exploring Vertical Transmission of Bifidobacteria from Mother to Child. *Appl. Environ. Microbiol.* **81**, 7078–87 (2015).
- Ma, B., Forney, L. J. & Ravel, J. Vaginal microbiome: rethinking health and disease. *Annu. Rev. Microbiol.* **66**, 371–89 (2012).
- Romero, R. *et al.* The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. *Microbiome* **2**, 4 (2014).
- Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578–1593 (2012).
- Russo, E., Taddei, A., Ringressi, M. N., Ricci, F. & Amedei, A. The interplay between the microbiome and the adaptive immune response in cancer development. *Therap. Adv. Gastroenterol.* **9**, 594–605 (2016).
- Almqvist, C., Cnattingius, S., Lichtenstein, P. & Lundholm, C. The impact of birth mode of delivery on childhood asthma and allergic diseases—a sibling study. *Clin. Exp. Allergy* **42**, 1369–1376 (2012).
- Feng, X. L., Xu, L., Guo, Y. & Ronsmans, C. Factors influencing rising caesarean section rates in China between 1988 and 2008. *Bull. World Health Organ.* **90**, 30–39A (2012).
- Bhute, S. *et al.* Molecular characterization and meta-analysis of gut microbial communities illustrate enrichment of Prevotella and Megasphaera in Indian Subjects. *Front. Microbiol.* **7**, 1–14 (2016).
- Nam, Y. D., Jung, M. J., Roh, S. W., Kim, M. S. & Bae, J. W. Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLoS One* **6**, e22109 (2011).
- The Integrative HMP (iHMP) Research Network Consortium. Perspective The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease. *Cell Host Microbe* **16**, 276–289 (2014).
- Balzola, F., Bernstein, C., Ho, G. T. & Lees, C. A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. *Nature* **464**, 59–65 (2010).
- Nakayama, J. *et al.* Diversity in gut bacterial community of school-age children in Asia. *Sci. Rep.* **5**, 8397 (2015).
- Rautava, S., Kalliomäki, M. & Isolauri, E. New therapeutic strategy for combating the increasing burden of allergic disease: Probiotics - A Nutrition, Allergy, Mucosal Immunology and Intestinal Microbiota (NAMI) Research Group report. *J. Allergy Clin. Immunol.* **116**, 31–37 (2005).

45. Isolauri, E. & Salminen, S. Probiotics: Use in Allergic Disorders. *J. Clin. Gastroenterol.* **42**, S91–S96 (2008).
46. Lee, J., Sullivan, D. J. O., Lee, J. & Sullivan, D. J. O. Genomic Insights into Bifidobacteria. *Microbiol. Mol. Biol. Rev.* **74**, 378–416 (2010).
47. Bottacini, F. *et al.* Comparative genomics of the Bifidobacterium breve taxon. *BMC Genomics* **15**, 170 (2014).
48. Taillford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **6**, 81 (2015).
49. Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int. J. Syst. Evol. Microbiol.* **54**, 1469–1476 (2004).
50. Collado, M. C., Derrien, M., Isolauri, E., De Vos, W. M. & Salminen, S. Intestinal integrity and Akkermansia muciniphila, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl. Environ. Microbiol.* **73**, 7767–7770 (2007).
51. Everard, A. *et al.* Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci. USA* **110**, 9066–9071 (2013).
52. Weber, C. G. *et al.* Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J.* **7**, 269–280 (2013).
53. Anderson, K. M. *et al.* A Clostridial Endo-β-galactosidase That Cleaves Both Blood Group A and B Glycotopes pable of liberating both the A trisaccharide A-Tri. *J. Biol. Chem.* **280**, 7720–7728 (2005).
54. Paixão, L. *et al.* Host glycan sugar-specific pathways in streptococcus pneumonia: Galactose as a key sugar in colonisation and infection. *PLoS One* **10**, 1–33 (2015).
55. Fei, N. & Zhao, L. An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice. *ISME J.* **7**, 880–4 (2013).
56. Kashyap, P. C. *et al.* Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota. *Proc. Natl. Acad. Sci. USA* **110**, 17059–64 (2013).
57. Salminen, S., Gibson, G. R., McCartney, A. L. & Isolauri, E. Influence of mode of delivery on gut microbiota composition in seven year old children Intestinal. *Gut* **53**, 1386–1390 (2004).
58. Sartor, R. B. & Mazmanian, S. K. Intestinal Microbes in Inflammatory Bowel Diseases. *Am. J. Gastroenterol. Suppl.* **1**, 15–21 (2012).
59. Muyzer, G., De Waal, E. C. & Uitterlinden, A. G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **59**, 695–700 (1993).
60. Collado, M. C., Isolauri, E., Laitinen, K. & Salminen, S. Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. *Am. J. Clin. Nutr.* **88**, 894–899 (2008).
61. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **108**, 4516–4522 (2011).
62. Tremblay, J. *et al.* Primer and platform effects on 16S rRNA tag sequencing. *Front. Microbiol.* **6**, 771 (2015).
63. Magoc, T. & Salzberg, S. L. FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies Tanja Mago. *Bioinformatics* **27**, 2957–2963 (2011).
64. Caporaso, J. G. *et al.* correspondence QIIME allows analysis of high-throughput community sequencing data Intensity normalization improves color calling in SOLiD sequencing. *Nat. Methods* **7**, 335–336 (2010).
65. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
66. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73**, 5261–5267 (2007).
67. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–8 (2012).
68. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124 (2014).
69. Langille, M. *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* **31**, 814–21 (2013).
70. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).

Acknowledgements

This work was supported by funding from the Department of Biotechnology, Government of India and Academy of Finland through a bilateral project entitled “The origins and process of microbiota development in different geographic areas: creating new nutritional tools for microbiota modulation” (DBT Grant Number: YS/BT/IN/FINLAND/10/YSS/2012-15 and Academy of Finland, Grant number: 255736). The authors would like to acknowledge the Director, KEMHRC, Pune and Director, NCCS, Pune for their support. Author SVK would like to thank Dr. Kiran Mahale and Diptaraj Chaudhari for their suggestions in the statistical analysis.

Author Contributions

S.S. and Y.S.S. conceived the study. S.V.K., S.P.C., H.K., S.S., A.C.O., J.M., E.I., A.R.B. and Y.S.S. designed the study. R.J., A.R.B., and E.I. provided the samples. S.V.K., H.K., S.P.C., and A.E. performed the experiments. S.S., Y.S.S., S.R., H.K., N.P.N., and R.H.P. provided guidance with the experiments. S.V.K. and H.K. performed the bioinformatics and statistical analyses. D.P.D. provided guidance with bioinformatics and statistical analysis. S.V.K. drafted the manuscript with input from S.S., Y.S.S., and H.K. All authors contributed to manuscript revisions, have read and approved the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-11215-y](https://doi.org/10.1038/s41598-017-11215-y)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017