



OPEN

Transcription shifts in gut bacteria shared between mothers and their infants

Tommi Vatanen^{1,2}✉, O. Sakwinska³, B. Wilson¹, S. Combremont³, W. S. Cutfield^{1,4}, S. Y. Chan^{5,6}, K. M. Godfrey^{7,8}, NiPPeR Study Group* & Justin M. O'Sullivan^{1,8,9,10}✉

The infant gut microbiome contains a portion of bacteria that originate from the maternal gut. In the infant gut these bacteria encounter a new metabolic environment that differs from the adult gut, consequently requiring adjustments in their activities. We used pilot community RNA sequencing data (metatranscriptomes) from ten mother-infant dyads participating in the NiPPeR Study to characterize bacterial gene expression shifts following mother-to-infant transmission. Maternally-derived bacterial strains exhibited large scale gene expression shifts following the transmission to the infant gut, with 12,564 activated and 14,844 deactivated gene families. The implicated genes were most numerous and the magnitude shifts greatest in *Bacteroides* spp. This pilot study demonstrates environment-dependent, strain-specific shifts in gut bacteria function and underscores the importance of metatranscriptomic analysis in microbiome studies.

Large-scale microbial colonization of the infant gut begins at birth^{1–3} and these early microbial residents contribute to host development and health. For example, many of these initial inhabitants interacting with immune cells^{4,5}, metabolise complex sugars (human milk oligosaccharides, HMOs) in breast milk^{6–8} and reduce the intestinal pH to restrict the growth of opportunistic pathogens⁹. The absence of bacteria with such core functions may therefore have lasting health consequences for the human host.

A portion of the pioneering infant gut bacteria are vertically transmitted from the maternal gut at birth^{10,11}. The metabolic environment of the infant gut is markedly different to that of the adult gut, due to the sole reliance on breast milk and/or infant formula. Maternally derived bacteria are therefore likely to adjust their gene expression to survive in this new environment. Indeed, an initial investigation highlighted high-level functional shifts, including decreased bacterial fermentation, following mother-to-infant transmission¹². Previous studies have not, however, investigated changes in expression of individual genes nor strain-specific transcription adjustments in maternally-derived bacterial strains.

Here, we used gut metagenomes and metatranscriptomes from 10 mother-infant dyads from the United Kingdom and New Zealand participating in the NiPPeR Study¹³ to investigate variations in the gene expression of gut bacterial strains shared between mother and infant. We specifically asked to what degree did the shared strains differ in terms of their gene expression in the infant vs. maternal gut and whether any bacterial species were more proficient in adjusting to their new environment in the infant gut.

Results

Gut bacteria differed taxonomically in NiPPeR mothers and their infants. The NiPPeR double-blinded, randomized controlled trial recruited 1,729 women prior to conception and followed 585 mother-infant dyads from delivery to examine the impact of a nutritional supplement [containing additional micronutrients,

¹Liggins Institute, The University of Auckland, Private Bag 102904, Auckland, New Zealand. ²The Broad Institute of MIT and Harvard, Cambridge, MA, USA. ³Nestlé Institute of Health Sciences, Nestlé Research, Société des Produits Nestlé S.A., 1000 Lausanne, Switzerland. ⁴A Better Start – National Science Challenge, Auckland, New Zealand. ⁵Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁶Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore, Singapore. ⁷NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust and University of Southampton, Southampton, UK. ⁸MRC Lifecourse Epidemiology Centre, University of Southampton, Southampton, UK. ⁹The Maurice Wilkins Centre, The University of Auckland, Auckland, New Zealand. ¹⁰Brain Research New Zealand, The University of Auckland, Auckland, New Zealand. *A list of authors and their affiliations appears at the end of the paper. ✉email: t.vatanen@auckland.ac.nz; justin.osullivan@auckland.ac.nz

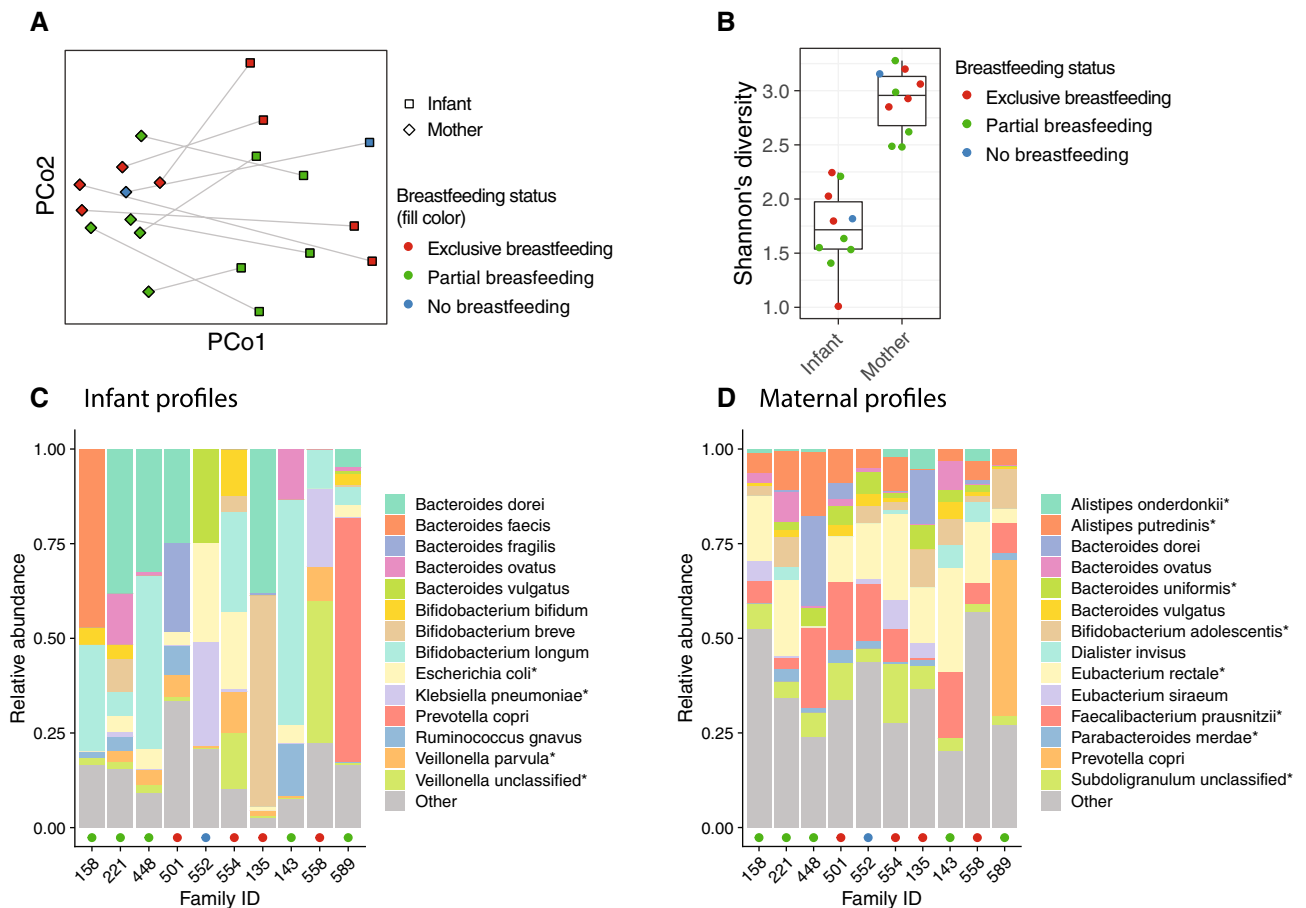


Figure 1. Distinctive microbial profiles separated infants from their mothers. **(A)** Principal Coordinate Analysis ordination of bacterial taxonomic profiles. Samples from the same family are connected by line. **(B)** Shannon's diversity of microbial species profiles. The box of the boxplot shows the interquartile range, whiskers show the range of data and horizontal line in the box shows the median. **(C–D)** Relative abundances of the 14 most abundant species (on average) in the infant **(C)** and maternal **(D)** stool samples. Coloured points on the x-axis indicate breastfeeding status as on panel B.

myo-inositol and probiotics—*Lactobacillus rhamnosus* NCC 4007 (CGMCC 1.3724; LPR) and *Bifidobacterium animalis* sp. *lactis* NCC 2818 (CNCM I-3446; Bl818)] taken before and during pregnancy on the health of mothers and their infants¹³. In this pilot study, we analysed 20 stool samples, collected six months after birth, from 10 mother–infant dyads within the NiPPeR cohort (five from Auckland, New Zealand; five from Southampton, United Kingdom) with appropriate positive and negative control samples to characterize vertical mother-to-infant microbial transmission and subsequent shifts in bacterial transcription (Table S1).

We first characterised the fecal microbiomes using metagenomic DNA sequencing and investigated the microbial populations taxonomically at the species level. As expected, mothers and infants harbored distinctive microbial communities separated on Principal Coordinate Analysis ordination (Fig. 1A, PERMANOVA test $R^2 = 0.314$, $p < 0.001$), mirrored by a contrasting difference in microbial diversity (Fig. 1B, Student's t-test $p = 1.1 \times 10^{-5}$). The most abundant species in infants included several species from the genera *Bifidobacterium* and *Bacteroides* as well as opportunistic pathogens such as *Escherichia coli* and *Klebsiella pneumoniae* (Fig. 1C), consistent with other infant cohorts. Mothers harbored more diverse microbial communities with many species characteristic of more complex diets such as *Faecalibacterium prausnitzii* and *Prevotella copri* (Fig. 1D)^{14,15}. Analysis of breastfeeding status was underpowered and did not identify any apparent, consistent effects on the taxonomic profiles (Fig. 1A–C).

Microbial metabolic functions differed between mothers and infants. We next asked how these apparent taxonomic differences were reflected in bacterial gene abundances related to metabolic pathways in the MetaCyc database. To this end, we first compared the overall reference database coverage between adult and infant metagenomes and found that a higher proportion of infant metagenomes mapped to our reference protein database (UniRef protein clusters) compared to the adult metagenomes ($p = 0.0003$, Fig. 2A). Paradoxically, on average, a smaller proportion of the mapped reads in infants tended to have confident functional annotations in the MetaCyc pathway database, although this difference was not statistically significant ($p = 0.16$, Fig. 2B). This indicates that the bacterial genome databases cover the infant gut bacteria better than a more diverse set of bacteria harbored by adult humans but, counterintuitively, genes on the infant gut reference genomes were more

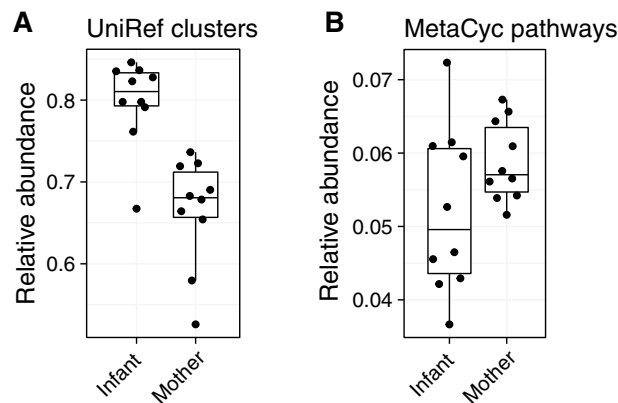


Figure 2. Mapping metagenomic DNA to protein databases and MetaCyc pathways. **(A)** Larger proportion of DNA reads from the infant metagenomes were mapped against UniRef protein clusters compared to adult metagenomes (Wilcoxon test, $p=0.0003$). **(B)** Proportion of mapped DNA reads in **(A)** that also contributed to quantifications of MetaCyc pathways. The difference between mothers and infants was not statistically significant (Wilcoxon test, $p=0.16$).

poorly annotated with metabolic functions compared to adult gut genomes. Metatranscriptomic data provides additional information to prioritize such unannotated genes for in vitro functional characterization.

We then tested 354 MetaCyc pathways observed in at least two samples for differences in DNA abundance between mothers and infants and found that 152 pathways (42.9% of tested pathways) were differentially abundant between mothers and infants (linear mixed effects model, q -value < 0.05 , Table S2). Of these, 104 were more abundant in infants and 48 were more abundant in mothers. We went on to confirm whether these differences persisted in the metatranscriptomic data (RNA sequencing) and quantified MetaCyc pathways by transcript (RNA) abundance. Of the slightly higher number of metabolic pathways detected by their transcripts in at least two samples (385 MetaCyc pathways), only 40 (10.4%) were differentially abundant between mothers and infants (q -value < 0.05 , Table S2). For example, pyrimidine deoxyribonucleotide biosynthesis (MetaCyc identifier PWY-6545) and purine ribonucleosides degradation (PWY0-1296) were more abundant in mothers by DNA abundance ($q = 9.8 \times 10^{-4}$ and $q = 7.4 \times 10^{-4}$, respectively; Table S2) but equally abundant in mothers and infants by RNA abundance ($q = 0.63$ and $q = 0.87$, respectively). The number of differentially abundant pathways being markedly lower in the metatranscriptomic ($n = 40$ pathways) data compared to the metagenomic data ($n = 152$ pathways) could indicate that the bacterial communities were compensating for variation in DNA gene abundance by collectively adjusting gene expression to maintain diverse and stable community activity profiles.

Gut bacterial strains shared between mothers and infants. Maternal gut strains vertically transmitted at birth contribute to the seeding of infant gut microbial communities^{10,11}. We sought to identify such vertically transmitted, or otherwise shared strains between mother and infant using a single nucleotide polymorphism (SNP)-based haplotyping approach. This approach assigned a SNP haplotype profile for the dominant strain of each species in all samples. We then compared these SNP haplotypes within species and median-normalized these nucleotide identities to account for differences in population diversity between species. Family-specific mother-infant strain pairs were, on average, more similar compared to random strain pairs (Fig. 3A). We then assigned a conservative threshold of 0.2 median-normalized DNA distance to identify shared strains (Fig. 3A). Using this threshold, we found 51 bacterial strains that were shared between mothers and their infants, representing 30 different bacterial species and seven bacterial genera (Table 1). Most shared strains (36/51, 70.6%) were species in the genus *Bacteroides* and six (11.8%) were bifidobacteria. Family-specificity of the shared strains was evident, as exemplified on the strain phylogeny of *Bacteroides ovatus* with the highest number of strains ($n = 4$) shared between a mother and her infant (Fig. 3B). Given the maternal gut microbiome is a known source of strains colonizing the infant gut, these shared strains most likely represent vertical mother-to-infant transmission events but could also be independently acquired from other sources by both the mother and the infant, or even transmitted from the infant to the mother.

Strains shared within a family adjust by changing gene expression. Gut bacteria respond to environmental changes through, for example, environmental and quorum sensing mechanisms. Since the environments in maternal and infant gut differ, we next investigated transcriptome (RNA) changes in the strains that were shared within a family. We detected a transcriptome signal in both maternal and infant samples for 21 of 51 (40%) shared strains with the total of 68,850 gene families, ranging between 959–6,393 unique gene families per strain per sample. We then compared gene family-specific RNA abundances, measured in copies-per-million, to corresponding DNA abundances (Fig. 4A). Since both measurements are relative (not absolute) quantifications, the average amounts of RNA and DNA detected from any given species may vary significantly. For example, if one species in a community is transcriptionally highly active (i.e. there are more RNA reads from that species compared to DNA reads), the amounts of RNA reads from all other species/strains will appear depleted com-

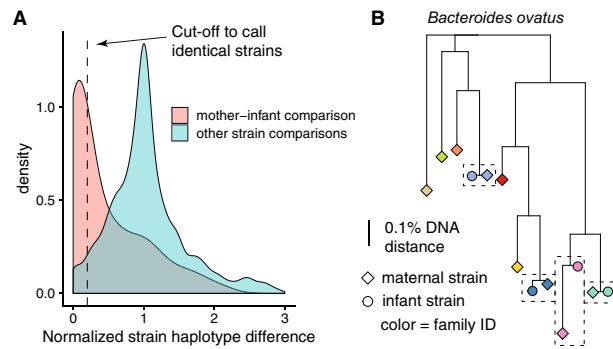


Figure 3. Bacterial strains shared between a mother and her infant. **(A)** Density plot of within-species strain comparisons between all detected strains per species. X-axis shows median normalized SNP-haplotype DNA distance (Jukes-Cantor model) divided by the median DNA distance per species (0 = equal strain haplotypes, 1 = median SNP haplotype dissimilarity). Red area shows comparisons within mother-infant pairs (given that a strain was detected from both infant and mother), teal area shows all other strain comparisons representing strain comparisons between random gut microbiomes. **(B)** Phylogenetic tree of *Bacteroides ovatus* strains detected in this cohort. Colors represent different mother-infant pairs (families) and scale shows the branch length corresponding to 0.1% DNA dissimilarity. Four family-specific strains are highlighted with dashed boxes.

Bacterial species	Number of transmission events	Events by breastfeeding status at 6 months		
		Exclusive	Partial	None
<i>Bacteroides ovatus</i>	4	0	4	0
<i>Bacteroides dorei</i>	3	1	2	0
<i>Bacteroides</i> sp 3 1 33FAA	3	1	1	1
<i>Bacteroides</i> sp 3 1 40A	3	1	1	1
<i>Bacteroides</i> sp 4 3 47FAA	3	1	1	1
<i>Bacteroides</i> sp D1	3	0	3	0
<i>Bacteroides caccae</i>	2	0	2	0
<i>Bacteroides</i> sp 2 1 16	2	2	0	0
<i>Bacteroides</i> sp 2 1 7	2	1	1	0
<i>Bacteroides</i> sp 9 1 42FAA	2	1	1	0
<i>Bifidobacterium adolescentis</i>	2	0	2	0
<i>Bifidobacterium longum</i>	2	0	2	0
<i>Parabacteroides distasonis</i>	2	1	1	0
<i>Parabacteroides</i> sp 20 3	2	1	1	0
<i>Bacteroides cellulosilyticus</i>	1	1	0	0
<i>Bacteroides fragilis</i>	1	1	0	0
<i>Bacteroides</i> sp 1 1 30	1	0	1	0
<i>Bacteroides</i> sp 2 1 22	1	0	1	0
<i>Bacteroides</i> sp 2 2 4	1	0	1	0
<i>Bacteroides</i> sp 4 1 36	1	0	1	0
<i>Bacteroides thetaiotaomicron</i>	1	1	0	0
<i>Bacteroides uniformis</i>	1	0	1	0
<i>Bacteroides vulgatus</i>	1	0	0	1
<i>Bifidobacterium bifidum</i>	1	0	1	0
<i>Bifidobacterium</i> sp 12 1 47BFAA	1	0	1	0
<i>Clostridium</i> sp SS2 1	1	1	0	0
<i>Collinsella aerofaciens</i>	1	1	0	0
<i>Lachnospiraceae bacterium</i> 5 1 63FAA	1	1	0	0
<i>Parabacteroides merdae</i>	1	0	1	0
<i>Roseburia inulinivorans</i>	1	1	0	0

Table 1. Number of mother-infant pairs that shared a strain per bacterial species.

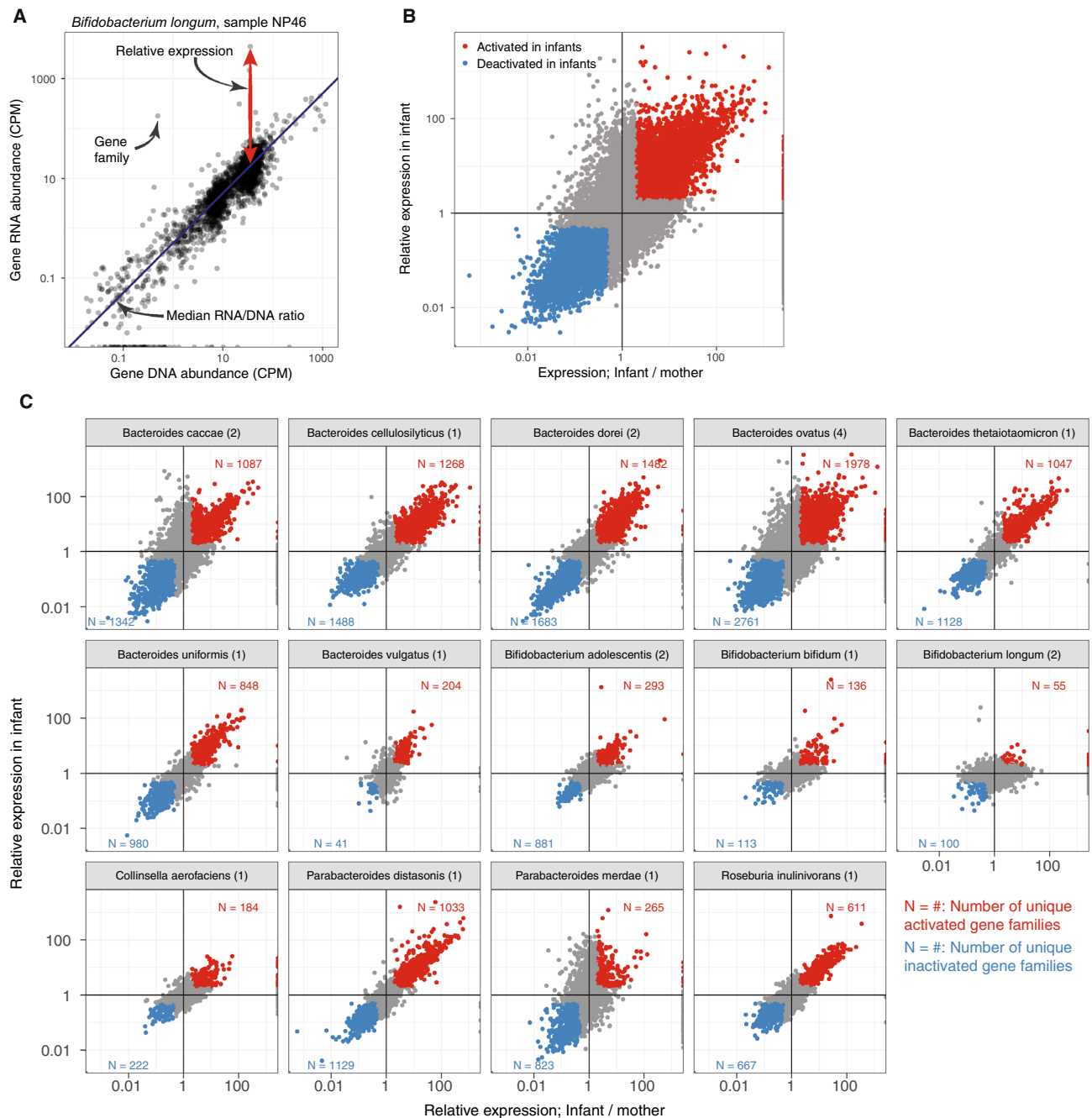


Figure 4. Gene expression changes in the strain shared within a family. **(A)** DNA and RNA abundances in copies-per-million (CPM) in *Bifidobacterium longum* in an infant sample (NP46). Gene family relative expression is calculated by normalizing the RNA/DNA ratio with the species and sample specific median RNA/DNA ratio. **(B)** A scatterplot of gene family relative expression change in infants vs. mothers (x-axis) and infant relative expression (y-axis). The gene families that were activated or inactivated after mother-to-infant transmission are highlighted in red and blue, respectively; activation denotes to infant relative expression > 2 (y-axis) and Infant/mother relative expression > 2 (x-axis); deactivation denotes to infant relative expression $< \frac{1}{2}$ and infant/mother relative expression $< \frac{1}{2}$. **(C)** Panel **(B)** stratified by species. Numbers in parentheses following the species name shows the number of families sharing a strain in the given species and panel-specific gene family frequencies (N = ...) denote the number of unique gene families that were activated or deactivated in red and blue, respectively (if the same gene family was activated in multiple infants it was counted only once).

pared to DNA. Within the shared strains, median RNA/DNA ratio per species per sample varied between 0.38 and 1.08. We therefore median-normalized the RNA/DNA ratio of each species sample combination and refer to this normalized RNA/DNA ratio as *relative expression*. Figure 4A illustrates the normalization for *B. longum* in one infant sample (NP46).

Since our ultimate goal was to investigate transcriptome changes between maternal and infant gut bacteria, we also measured the change in the relative expression between mothers and infants within the family (Fig. 4B, x-axis). This corresponded to asking whether a gene family on a shared strain was activated in the infant gut in comparison to how it was expressed in the maternal gut. As the analyses lacked sufficient statistical power to assess strain-specific gene expression changes, we applied thresholds in expression quantifications (relative expression in infants, ratio of relative expressions in infants and mother) to identify tentative gene activation and inactivation events. We identified 12,574 gene family activation events (18.3% of all gene families harbored on the shared strains) where the gene families were overexpressed in infants (relative expression, RNA/DNA > 2) and had at least twofold higher relative expression in infants compared to their mothers (Fig. 4B, red points; Table S3). Slightly more gene families, $n = 14,844$ (21.6%), were deactivated in infants (relative expression < 0.5) and had at least twofold higher relative expression in mothers (Fig. 3B, blue points; Table S3). Stratifying these findings by bacterial species revealed marked differences in the frequency of (de)activated gene families and the changes in relative expression (Fig. 4C). Many species in the genus *Bacteroides* activated and deactivated more than one thousand genes following tentative mother-to-infant transmission, with many gene families showing an over 100-fold increase in relative expression between the infant and the mother. While we also detected (de)activated gene families in bifidobacteria, the frequency and magnitude of the changes in bifidobacteria were less pronounced (Fig. 4C). For example, *Bifidobacterium longum*, known for its versatile breast milk metabolism, only activated 55 unique gene families in two tentative mother-to-infant transmission events. Stratifying these observations by breastfeeding status was underpowered and did not show any consistent trends (Table S3). Overall, of 68,850 gene families harbored by the shared bacterial strains, 37,156 (54.0%) did not have any homology-based function annotation (i.e. were unannotated), whereas 5,764 of 12,564 (45.9%, strain-by-strain paired Wilcoxon test, $p = 9.5 \times 10^{-7}$) activated genes and 9,935 of 14,844 (66.9%, $p = 9.5 \times 10^{-6}$) deactivated genes were unannotated.

Finally, we asked whether the gene families activated in infants were consistent between mother-infant pairs using data from three *Bacteroides* and two *Bifidobacterium* species which had shared strains in more than one mother-infant pair. This “replication rate” was highest in *Bacteroides dorei* and *Bacteroides ovatus*, where 46.2% and 45.7% of the infant-activated gene families were consistent in two or more infants, respectively. In *Bifidobacterium adolescentis* and *Bifidobacterium longum* only 3.4% and 1.8% of activated gene families were consistent between multiple infants.

Discussion

We demonstrated differences in functional activity and strain-specific gene expression in the gut microbiomes of ten mother-infant dyads. Several strains from the genera *Bacteroides* and *Bifidobacterium* that were shared between mother-infant pairs showed marked adjustments to the infant gut environment by thousands of gene (de)activation events. The magnitudes of these shifts were largest in *Bacteroides* spp. which also contribute to the metabolism of dietary carbohydrates, including oligosaccharides in breast milk^{16–18}. Larger, longitudinal studies are needed to assess reproducibility and temporal stability of such bacterial adjustments in the infant gut.

The relative abundance of metagenomic gene families that could be mapped to metabolic pathways in the MetaCyc database was lower in infants compared to their mothers. Mirroring this, roughly half of the genes tentatively contributing to bacterial adjustments were unannotated providing another priority research question; what are these genes doing? Metatranscriptomic data will be useful for prioritizing bacterial genes and their protein products for in vitro functional characterization. Priority could be given, for example, to genes that are highly expressed in infants, are widely prevalent (present in most infants) or infant-specific, are associated with important infant phenotypes or are harbored by bacteria of interest (e.g. bifidobacteria); or any combination of such criteria. New computational annotation methodologies and gut metabolite profiles could also provide insights into the functions and metabolic targets/outputs of these genes in a less targeted manner.

This pilot study included data from ten mother-infant pairs, with one stool sample per participant, and is therefore limited in statistical power. Additionally, our study design did not allow us to evaluate the null distribution of bacterial gene expression changes in an individual (a mother or an infant) over time. In addition to including more participants, future studies should preferably incorporate multiple stool samples from both mothers and infants to assess and quantify day-to-day fluctuations in bacterial gene expression. This would enable identification of gene expression shifts in maternally-derived strains that are statistically different from day-to-day variations. Optimally, longitudinal sampling in infants should also cover the transition from exclusive breastfeeding to introduction of solid foods and weaning to investigate how this common early feeding pattern affects microbial transcription patterns. The gene expression shifts in *Bifidobacterium* spp. and *Bacteroides* spp., for example, could be more pronounced during exclusive breastfeeding and then stabilize towards adult profiles after weaning. Finally, larger studies could more robustly assess species-specific replication rates of gene expression changes in specific gene families and whether the same genes are consistently up- or down-regulated across individuals.

We focused on bacterial species tentatively transmitted from the maternal to infant gut yet a large proportion of infant gut bacteria are obtained from other sources¹⁰. Future studies may therefore fruitfully explore gene expression and metabolic adjustments of bacteria obtained from other, ecologically more distant environments, since such adjustments may require even more radical activity alterations. Researchers should also investigate whether and to what degree bacterial transcription and its adjustments play a role in their engraftment capacity and early gut microbiome assembly.

Conclusions

We observed large scale gene expression shifts in gut bacterial strains shared between mothers and their infants. These gene expression shifts were most prominent in *Bacteroides* spp. potentially reflecting bacterial adjustments to the new environment. Taken together, metatranscriptomic data provides an important 'omics data layer adding information on environment-dependent community metabolic activities and adjustments to microbiome investigations.

Methods

Trial registration and ethics approval. NiPPeR is an academic-led study by the EpiGen Global Research Consortium (ClinicalTrials.gov NCT02509988, Universal Trial Number U1111-1171-8056; 16/7/2015, <https://www.clinicaltrials.gov/ct2/show/NCT02509988>). After independent, full, external peer review the study protocol and subsequent amendments have been approved by the Research Ethics Committees at each of the three study sites (Southampton: Health Research Authority NRES Committee South Central Research Ethics Committee (REC), reference 15/SC/0142), the Health and Disability Ethics Committee (HDEC) (New Zealand) reference 15/NTA/21 and the National Healthcare Group Domain Specific Review Board (NHG DSRB) (Singapore) reference 2015/00205). This study is being conducted in compliance with the NiPPeR study protocol, Good Clinical Practice and the applicable regulatory requirements. All data analysed in this study was stored and analysed in accordance with the UK Data Protection Act, and applicable regulations and guidance of each country and institution.

Sample collection and preservation. Stools were collected from mothers and infants in Southampton and New Zealand according to a standardised S.O.P to ensure uniformity (6SOP7, Stool [collection and processing], effective date 13/08/2016, Revision V1.0). Briefly, on the sixth month post-delivery clinical visit, mothers were given a bedpan liner (Vernacare) and asked to (1) pass urine into the toilet prior to placing the tray on the toilet seat; (2) pass the stools; (3) place a small scoop into the collection tube (Sarstedt Faeces container with screw cap, polypropylene, sterile, Size: 101 mm [L] × 16.5 mm [Dia.]) containing RNAlater Stabilisation Solution (5 ml). Participants were instructed to screw down the cap tightly and shake vigorously at least 10 times to ensure that the sample has been well-mixed into the solution. Infant samples were taken directly from the nappy and placed in collection tubes containing RNAlater, as per the maternal samples. The time and date of the collection were immediately recorded on the tube and specimen containers placed on ice and taken to the laboratory. If collected at home, samples were held at -20°C until delivery to the clinic.

Faecal samples in the RNAlater tubes were processed [placed in a Falcon tube, centrifuged (4600 rpm, 4°C , 5 min)] within a maximum of 60 min from receipt of samples from subject (start and end times were recorded). RNAlater was discarded and 0.5 g of sample was transferred into each of 6×3 ml micronics tubes. Sample weights of each aliquot were recorded and samples stored (-80°C) until analysis.

DNA extraction controls. Negative and positive DNA extraction controls were extracted and included in the analysis. The negative control was 200 μl RNase free water. Two microbial community standards were used as positive controls; (1) 20 Strain Even Mix Whole Cell Material (ATCC MSA-2002, Virginia, USA) embedded in a stool matrix (#NATROTA-GP, ZeptoMetrix Corporation, New York, USA), and (2) ZymoBIOMICS Microbial Community Standard (#D6300, Zymo Research, California, USA).

Nucleic acid extraction and library preparation. DNA and RNA extractions were performed at the Liggins Institute (University of Auckland, New Zealand) using a modified protocol of the AllPrep DNA/RNA Mini Kit (#80204, Qiagen, Germany) as described previously¹⁹. Briefly, stool aliquots were thawed and incubated in 100 μl of lysis buffer (30 mM Tris-HCl, 1 mM EDTA, 15 mg/ml lysozyme) and 10 μl proteinase K for 10 min at room temperature with regular agitation. Samples were then mixed with 1.2 ml RLT plus buffer (Qiagen, Germany), 12 μl beta-mercaptoethanol, and 1 ml of acid washed glass beads ($\leq 106 \mu\text{m}$, -140 U.S. sieve; #G4649-100G, Sigma Aldrich, USA) and shaken vigorously at 30 Hz for 10 min on a TissueLyser II (Qiagen, Germany). The homogenate was then passed through a QIAshredder spin column (#79654, Qiagen, Germany), before continuing with the standard AllPrep DNA/RNA Mini kit protocol (#80204, Qiagen, Germany). DNA was eluted in 75 μl of EB buffer and RNA eluted in 50 μl RNase free water. DNA was quantified by Qubit dsDNA High Sensitivity Assay Kit (#Q32851, Invitrogen, California, USA). RNA was quantified by Qubit RNA High Sensitivity Assay Kit (#Q32852, Invitrogen, California, USA) and quality assessed using Agilent RNA 6000 Nano Kit and Agilent 2100 Bioanalyzer instrument (Agilent technologies, California, USA). Sample data (including RNA integrity Number, RIN) is available in Table S1.

DNA shotgun sequencing. DNA sequencing was performed following a previously described procedure²⁰. Genomic DNA was fragmented (random shearing) into ~ 350 bp fragments. Sequencing libraries were prepared from the fragmented DNA using the NEBNext Ultra Library Prep Kit for Illumina (New England Biolabs). Fragment size distribution within the prepared sequencing libraries was evaluated (Qubit 2.0 fluorometer quantitation and Agilent 2100 Bioanalyzer). Prior to sequencing, DNA concentrations of the sequencing libraries were measured by quantitative real-time PCR (qPCR). Sequencing libraries were sequenced on an Illumina platform (2×150 bp paired-end sequencing).

RNA shotgun sequencing. The Ribo-Zero kit was used to deplete samples of rRNA before random fragmentation of the mRNA. cDNA synthesis was performed according to²¹. cDNA was synthesized from mRNA using random hexamers primer, after which a custom second-strand synthesis buffer, dNTPs, RNase H and DNA polymerase I were added to initiate the second-strand synthesis. After terminal repair, A ligation, and sequencing adaptor ligation, the double-stranded cDNA library was completed through size selection and PCR enrichment. The cDNA library was sequenced using paired-end 150 bp sequencing.

Bioinformatics. Both DNA and RNA sequencing data were subject to standard quality control, including read trimming based on base call qualities and host contamination removal. We obtained, on average, [mean (min–max)] 22.6 M (16.7–27.7 M) and 45.6 M (32.4–55.5 M) quality controlled, paired sequencing reads for DNA and RNA sequencing data, respectively. Bacterial species abundances were quantified using the metagenomes and MetaPhlAn2 v2.7.7²². Alpha-diversities of the taxonomic profiles were measured by Shannon's diversity index and beta-diversities were measured by Bray–Curtis dissimilarity. Principal Coordinate ordination was generated using Nonmetric Multidimensional Scaling (metaMDS function in vegan R package v 2.5-6, <https://cran.r-project.org/web/packages/vegan/>). SNP-haplotypes of the dominant bacterial strains of a given species per sample were determined by StrainPhlAn²³ by requiring a minimum coverage of 5 bases for SNP calling ('-min_read_depth 5' command line parameter for sample2markers.py Python script). The phylogenetic tree (Fig. 3B) was generated using the phangorn R package (v2.5.5, <https://cran.r-project.org/web/packages/phangorn/>). Species-stratified gene family abundances from both metagenomes and metatranscriptomes were quantified by HUMAnN2 v0.11.2²⁴ which involves mapping sequencing reads against custom species pangenomes. Gene family abundances were further stratified to quantify MetaCyc pathways by methodology implemented in HUMAnN2. Using HUMAnN2 for both metagenomic and metatranscriptomic analyses enabled co-analysis of DNA and RNA abundances of any given gene detected by both sequencing methods.

Statistical analysis. Homology-based functional annotations of (de)activated genes were compared to annotations of all genes harbored by the shared strains as follows. For each shared strain, percentage of annotated genes among all genes and (de)activated genes were computed. These percentages were compared using paired Wilcoxon rank sum test across all transmitted strains. For all statistical testing scenarios with multiple testing, false discovery rate corrected p-values (q-values), are reported.

Ethics approval and consent to participate. The study protocol and subsequent amendments have been approved by the Research Ethics Committees at each of the three study sites (Southampton: Health Research Authority NRES Committee South Central Research Ethics Committee (REC), reference 15/SC/0142), the Health and Disability Ethics Committee (HDEC) (New Zealand) reference 15/NTA/21 and the National Healthcare Group Domain Specific Review Board (NHG DSRB) (Singapore) reference 2015/00205). All participating mothers gave written, informed consent at the first preconception study visit.

Data availability

Sequencing data will be made publicly available in NCBI Sequence Read Archive upon publication of this manuscript under BioProject accession PRJNA707065. Analysis scripts and intermediate data products to reproduce all analyses are available at <https://github.com/tvatanen/nipper-metatranscriptomics>.

Received: 16 July 2021; Accepted: 10 December 2021

Published online: 24 January 2022

References

- Mitchell, C. M. *et al.* Delivery mode affects stability of early infant gut microbiota. *Cell Rep. Med.* **1**(9), 100156 (2020).
- Shao, Y. *et al.* Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**(7776), 117–121 (2019).
- de Goffau, M. C. *et al.* Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**(7769), 329–334 (2019).
- Dominguez-Bello, M. G., Godoy-Vitorino, F., Knight, R. & Blaser, M. J. Role of the microbiome in human development. *Gut* **68**(6), 1108–1114 (2019).
- Olin, A. *et al.* Stereotypic immune system development in newborn children. *Cell* **174**(5), 1277–1292.e14 (2018).
- Kirmiz, N., Robinson, R. C., Shah, I. M., Barile, D. & Mills, D. A. Milk glycans and their interaction with the infant-gut microbiota. *Annu. Rev. Food Sci. Technol.* **25**(9), 429–450 (2018).
- Kujawska, M. *et al.* Succession of bifidobacterium longum strains in response to a changing early life nutritional environment reveals dietary substrate adaptations. *iScience* **23**(8), 101368 (2020).
- Vatanen, T. *et al.* Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life. *Nat. Microbiol.* **4**(3), 470–479 (2019).
- Henrick, B. M. *et al.* Elevated fecal pH indicates a profound change in the breastfed infant gut microbiome due to reduction of *Bifidobacterium* over the past century. *mSphere* **3**(2), e00041-18 (2018).
- Ferretti, P. *et al.* Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**(1), 133–145.e5 (2018).
- Yassour, M. *et al.* Strain-level analysis of mother-to-child bacterial transmission during the first few months of life. *Cell Host Microbe* **24**(1), 146–154.e4 (2018).
- Asnicar, F. *et al.* Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. *mSystems* <https://doi.org/10.1128/mSystems.00164-16> (2017).
- Godfrey KM, Cutfield W, Chan S-Y, Baker PN, Chong Y-S, NiPPeR Study Group. Nutritional intervention preconception and during pregnancy to maintain healthy glucose metabolism and offspring health ("NiPPeR"): Study protocol for a randomised controlled trial. *Trials* **18**(1), 131 (2017).

14. Asnicar, F. *et al.* Microbiome connections with host metabolism and habitual diet from 1098 deeply phenotyped individuals. *Nat. Med.* **27**(2), 321–332 (2021).
15. Wang, D. D. *et al.* The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nat. Med.* **27**(2), 333–343 (2021).
16. Marcobal, A. *et al.* Consumption of human milk oligosaccharides by gut-related microbes. *J. Agric. Food Chem.* **58**(9), 5334–5340 (2010).
17. Borewicz, K. *et al.* The association between breastmilk oligosaccharides and faecal microbiota in healthy breastfed infants at two, six, and twelve weeks of age. *Sci. Rep.* **10**(1), 4270 (2020).
18. Vatanen, T. *et al.* Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* **165**(4), 842–853 (2016).
19. Wilson, B. C. *et al.* Strain engraftment competition and functional augmentation in a multi-donor fecal microbiota transplantation trial for obesity. *Microbiome* **9**(1), 107 (2021).
20. Mokkalá, K. *et al.* GlycA, a novel marker for low grade inflammation, reflects gut microbiome diversity and is more accurate than high sensitive CRP in reflecting metabolomic profile. *Metabolomics* **16**(7), 76 (2020).
21. Dudhate, A., Shinde, H., Tsugama, D., Liu, S. & Takano, T. Transcriptomic analysis reveals the differentially expressed genes and pathways involved in drought tolerance in pearl millet [*Pennisetum glaucum* (L.) R. Br]. *PLoS ONE* **13**(4), e0195908 (2018).
22. Truong, D. T. *et al.* MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**(10), 902–903 (2015).
23. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* **27**(4), 626–638 (2017).
24. Franzosa, E. A. *et al.* Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**(11), 962–968 (2018).

Acknowledgements

We would like to thank the participants for their enthusiastic involvement in the study. A list of consortium members and their affiliations appears at the Supplement; consortium representative Justin O’Sullivan (justin.osullivan@auckland.ac.nz).

Author contributions

K.M.G., W.S.C., and S.Y.C. are Principal Investigators on the NiPPeR study Trial Management Group, which manages study delivery at the three sites, data analysis and interpretation and manuscript preparation. J.O.S. and O.S. prepared the microbiome S.O.P. for the NiPPeR trial. J.O.S. and T.V. conceived of the microbiome analysis. B.W. and S.C. extracted the microbiome samples. O.S. coordinated the sequencing. T.V. performed the analyses and wrote the manuscript. J.O.S. co-wrote the manuscript and takes responsibility for its content. All authors have commented on the manuscript.

Funding

Public good funding for this investigator-led study was through the UK Medical Research Council (as part of an MRC award to the MRC Lifecourse Epidemiology Unit), the Singapore National Research Foundation under its Translational and Clinical Research (TCR) Flagship Programme, the Singapore Government administered through the Agency of Science, Technology and Research (A*STAR; Singapore Institute for Clinical Sciences) and the New Zealand Government (as part of the Gravida, Centre of Research Excellence: Growth and Development). Funding for provision of the intervention and control drinks and to cover aspects of the fieldwork for the NiPPeR study was provided by Société des Produits Nestlé SA under a Research Agreement with the University of Southampton, Auckland UniServices Ltd, Singapore Institute for Clinical Sciences, National University Hospital Singapore PTE Ltd., National University of Singapore. Funding for the microbiome sequencing was provided by Société des Produits Nestlé SA under a Research Agreement with Auckland UniServices Ltd.

Competing interests

OS and SC are employees of Société des Produits Nestlé (SPN). KMG has received reimbursement for speaking at conferences sponsored by companies selling nutritional products, and is part of an academic consortium that has received research funding from Abbott Nutrition, Nestec, BenevolentAI Bio Ltd. and Danone. TV, BW, SYC, WSC and JOS declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-04848-1>.

Correspondence and requests for materials should be addressed to T.V. or J.M.O.

Reprints and permissions information is available at www.nature.com/reprints.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

NiPPeR Study Group

O. Sakwinska³, W. S. Cutfield^{1,4}, S. Y. Chan^{5,6}, K. M. Godfrey^{7,8}, J. M. O'Sullivan^{1,8,9,10}, Sheila J. Barton⁸, Mary Cavanagh¹, Yap Seng Chong^{5,6,11}, Paula Costello⁸, Vanessa Cox⁸, Sarah El-Heis^{8,12}, Mrunalini Jagtap⁶, Karen Lillycrop¹³, Heidi Nield⁸, Gernalia Satianegara⁶, Irma Silva-Zolezzi¹⁴, Shu E. Soh⁶, Gladys Woon⁶, Tim Kenealy¹, Mark Vickers¹ & Jonathan Swann¹²

¹¹National University Hospital, Singapore, Singapore. ¹²Faculty of Medicine, University of Southampton, Southampton, UK. ¹³Faculty of Biomedical Sciences, University of Southampton, Southampton, UK. ¹⁴Nestlé Research, Société des Produits Nestlé S.A., Singapore, Singapore.