









Indigenous infants in remote Australia retain an ancestral gut microbiome despite encroaching Westernization

Received: 25 February 2025

Accepted: 20 October 2025

Published online: 03 December 2025

 Check for updates

Leonard C. Harrison ^{1,2,25} ✉, Theo R. Allnutt ^{1,3,25}, Sarah Hanieh⁴, Alexandra J. Roth-Schulze¹, Katrina M. Ngui¹, Natalie L. Stone¹, Esther Bandala-Sanchez¹, Lilly Backshell¹, George Gurruwiwi⁴, Veronica Gondarra⁴, Jennifer J. Couper^{5,6}, Maria E. Craig ^{7,8,9}, Elizabeth A. Davis ^{10,11,12}, Tony Huynh^{13,14,15}, Georgia Soldatos^{16,17}, John M. Wentworth ^{1,18}, Peter Vuillermin ^{19,20}, Megan A. S. Penno ⁶, Beverley-Ann Biggs ^{4,21} & ENDIA Study Group Consortium*

Studies of traditional Indigenous compared to ‘Western’ gut microbiomes are underrepresented, and lacking in young children, limiting knowledge of early-life microbiomes in different cultural contexts. Here we analyze the gut metagenomes of 50 Indigenous Australian infants (median age <one year) living remotely with variable access to Western foods, compared to age- and sex-matched non-Indigenous infants living in urban Australia. Indigenous infants had greater alpha diversity and significant differences in bacterial beta diversity, with 114 species and 38 genera differing in abundance. Indigenous infants almost exclusively had higher carriage of *Megaspaera*, *Streptococcus*, *Caecibacter*, *Parolsenella* and *Prevotella* species, and markedly higher numbers of gut viruses and fungi. *Bifidobacteria* ssp. were dominant in Indigenous infants. Despite encroaching Westernisation, the gut microbiome of Indigenous infants retains key features of traditional societies worldwide, attesting to the dominant influence of remote environment and traditional lifestyle in maintaining microbiome diversity.

Indigenous Australians are distinguished as the oldest, continuous living culture, successful as hunter-gatherers and subsistence farmers for over 60,000 years prior to European settlement¹. Along with the degradation of their traditional culture, Indigenous Australians have experienced a marked increase in cardiometabolic non-communicable diseases (NCDs), including obesity, diabetes, cardiovascular diseases, chronic kidney disease, and cancers^{2,3}. NCDs are associated with chronic low-grade, systemic inflammation^{4,5} and a shift to a less diverse gut microbiome (dysbiosis)^{6,7}. These pathophysiological features are considered to be a maladaptive response to a “Western” lifestyle, especially to the “Western” diet rich in saturated fat, simple carbohydrates, additives, and processed components, and low in fiber^{7,8}, and to other less well understood factors such as the cumulative burden of spiritual, cultural, and environmental stressors⁹.

Following colonization at birth, the gut harbors the largest and most diverse microbiome, critical for the development and ongoing health of the gut, immune, and other body systems. Distinct taxonomic and functional differences have been identified between the gut microbiomes of traditional hunter-gatherer or subsistence societies compared to urban Western societies^{10,11}, providing insights into the evolution of the gut microbiome under the influence of ethnicity, environment, and lifestyle. The gut microbiome of Western compared to traditional societies is less diverse, with a lower abundance of fiber-degrading bacteria that produce anti-inflammatory short-chain fatty acids and a higher abundance of mucus-degrading bacteria, which may impair gut epithelium integrity, leading to leakiness of bacterial products and systemic inflammation¹². The Western gut microbiome contains a higher abundance of *Bacteroides*, *Enterobacteria*, and

A full list of affiliations appears at the end of the paper. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: harrison@wehi.edu.au

Akkermansia and a lower abundance of beneficial *Bifidobacteria* and *Lactobacilli*; in the traditional gut microbiome, *Prevotella*, *Treponema*, *Proteobacteria*, *Clostridiales*, and *Ruminobacter* predominate, some having disappeared from the Western gut microbiome^{10,11}.

Counted among the hundreds of clans of Indigenous Australians are the Yolngu people, living in a remote area of northern Australia, whose traditional customs endure despite an increasingly pervasive Western lifestyle. Typically, they live in a community of modest, single-storey concrete dwellings, multiple generations sharing a household. The interiors are sparse, with minimal furniture. Some also live in temporary demountable structures installed following cyclone damage. The community comprises distinct groups, each with its own networks of family and kinship ties. Daily life takes place mostly outdoors, particularly on verandahs, which serve as important spaces for gathering and social interactions. The Yolngu diet comprises mainly Western-style food and beverages with a variable mix of traditional foods (see “Methods”). Yolngu children are increasingly exposed to sugar-sweetened beverages and ultra-processed and takeaway foods, and experience periods of food scarcity¹³. Indigenous Australians appear to have a genetic propensity for strong inflammatory responses¹⁴. Indeed the Yolngu children display elevated concentrations of circulating inflammatory cytokines (Hasthi Dissanayake, personal communication), consistent with gut microbiome dysbiosis and a predisposition to NCDs. Analysis of the oral¹⁵ and upper respiratory¹⁶ microbiomes of Indigenous Australian children by 16S rRNA gene amplicon sequencing has revealed the carriage of bacteria potentially associated with an increased prevalence of NCDs, but knowledge of the gut microbiome in young Indigenous children is lacking. Whether the gut microbiome of these Indigenous infants is ancestral and to what extent and at what age it might exhibit features of Westernization is unknown. The infant gut microbiome is shaped initially by vertical transmission from the mother, but also by genetic background and multiple environmental exposures encompassing living conditions, diet, water, soil, animals, toxins, parasitic and other infections, antibiotics, and psycho-social stress, which vary between populations and may distinguish Indigenous children living remotely from non-Indigenous children living in urban settings.

In order to characterize the remote, Indigenous gut microbiome in early childhood, and as a basis on which to understand the impact of the Western lifestyle, we used shotgun metagenomic sequencing to compare the gut microbiomes of 50 randomly-selected Indigenous Australian infants living in a remote Yolngu community to those of 50 age- and sex-matched non-Indigenous infants living in different urban areas of Australia. We show that Indigenous infants have greater alpha diversity and significant differences in bacterial beta diversity, with 114 species and 38 genera differing in abundance. Some taxa were unique to Indigenous infants, who had higher carriage of *Bifidobacteria* at younger ages and *Prevotella* at older ages; in contrast, non-Indigenous infants had a high abundance of *Phocaeicola* (*Bacteroides*) across ages. Indigenous infants also had markedly higher numbers of gut viruses and fungi. Thus, despite encroaching Westernization, these Indigenous infants begin life with a gut microbiome that retains key features of traditional societies worldwide. The Western gut microbiome has not been transmitted intergenerationally and has not yet emerged, attesting to the dominant influence of a remote environment and traditional lifestyle in maintaining gut microbiome diversity.

Results

Study populations

The study groups comprised 24 Indigenous and 24 non-Indigenous females, median (IQR) ages 294 (153, 428) and 293 (114, 433) days, respectively, and 26 Indigenous and 26 non-Indigenous males, median (IQR) ages 360 (222, 476) and 377 (226, 467) days, respectively. Matched pairs and individual data for mode of delivery, breastfeeding, and furred pets in the household are detailed in Supplementary Data 1. The characteristics of the groups are summarized in Table 1.

Table 1 | Characteristics of Indigenous and non-Indigenous subjects

| | Indigenous | Non-Indigenous | P value ^a |
|---|-------------------|-------------------|----------------------|
| Sex: female, male | 24, 26 | 24, 26 | |
| Age (days) | | | |
| Median (IQR) | 331 (166, 462) | 332 (175, 458) | 0.958 |
| Mean (SD) | 317 (170) | 317 (171) | |
| Gestational age at delivery (weeks) | | | |
| Median (IQR) | 38.0 (37.0, 39.0) | 38.0 (37.3, 39.4) | 0.200 |
| Mean (SD) | 37.6 (2.82) | 38.2 (1.63) | |
| Mode of delivery | | | |
| Vaginal | 32/50 (64%) | 26/50 (52%) | 0.311 |
| Cesarean section (elective) | 11/50 (22%) | 13/50 (26%) | 0.815 |
| Cesarean section (emergency) | 7/50 (14%) | 11/50 (22%) | 0.436 |
| Intrapartum antibiotics | | | |
| Vaginal | Unknown | 15/26 (58%) | |
| Cesarean section (elective) | “ | 12/13 (92%) | |
| Cesarean section (emergency) | “ | 11/11 (100%) | |
| Birth weight (Kg) | | | |
| Median (IQR) | 3.15 (2.77, 3.37) | 3.53 (3.15, 3.86) | 0.0004 |
| Mean (SD) | 3.09 (0.614) | 3.52 (0.552) | |
| Breastfeeding | | | |
| Current | 47/50 (94%) | 24/50 (48%) | <0.0001 |
| Exclusive for the first 6 months | 42/50 (84%) | 3/50 (6%) | <0.0001 |
| Complementary food started (months) | | | |
| Median (IQR) | 5.00 (4.50, 6.00) | 4.60 (4.37, 5.52) | 0.006 |
| Mean (SD) | 5.46 (1.38) | 4.80 (0.924) | |
| Maternal body mass index (BMI) ^b | | | |
| Median (IQR) | 22.6 (18.6, 25.8) | 25.2 (21.9, 30.7) | <0.0001 |
| Mean (SD) | 22.7 (4.9) | 26.3 (4.9) | |
| Consumption of traditional foods by the mother, at least twice weekly | 46/50 (92%) | Not applicable | |
| Number of people living in the home | | | |
| Median (IQR) | 4.00 (3.00, 7.25) | 4.00 (3.00, 4.75) | 0.017 |
| Mean (SD) | 6.70 (8.14) | 3.87 (1.13) | |
| At least one furred animal in the household | 25/50 | 30/50 | 0.422 |
| Type 1 diabetes proband | | | |
| Father | 0 | 12/50 (24%) | |
| Mother | 0 | 27/50 (54%) | |
| Sibling | 0 | 10/50 (20%) | |
| Fecal calprotectin (normal 100 mg/g) | | | |
| Median (IQR) | 1318 (515, 1809) | 39.9 (4.40, 102) | <0.0001 |
| Mean (SD) | 1274 (734) | 71.5 (103) | |
| Serum intestinal fatty acid binding protein (normal <1664 pg/ml) | | | |
| Median (IQR) | 637 (419–1091) | 689 (517–1572) | 0.538 |
| Mean (SD) | 954 (962) | 1062 (777) | |
| Fecal parasites by microscopy | 15/50 (30%) | Not determined | |
| Antibiotic (amoxicillin in the majority) within 1 month of fecal collection | 5 (10%) | 8 (16%) | 0.554 |
| Any antibiotic ever | Unknown | 11/50 (22%) | |

^aUnpaired t-test (two-tailed) or Fisher’s Exact test for categorical variables.

^bIndigenous was post-pregnancy when the infant was assessed; non-Indigenous was pre-pregnancy.

Bar charts of observed taxa (bacteria) proportions (25 most abundant) classified by MetaPhlAn

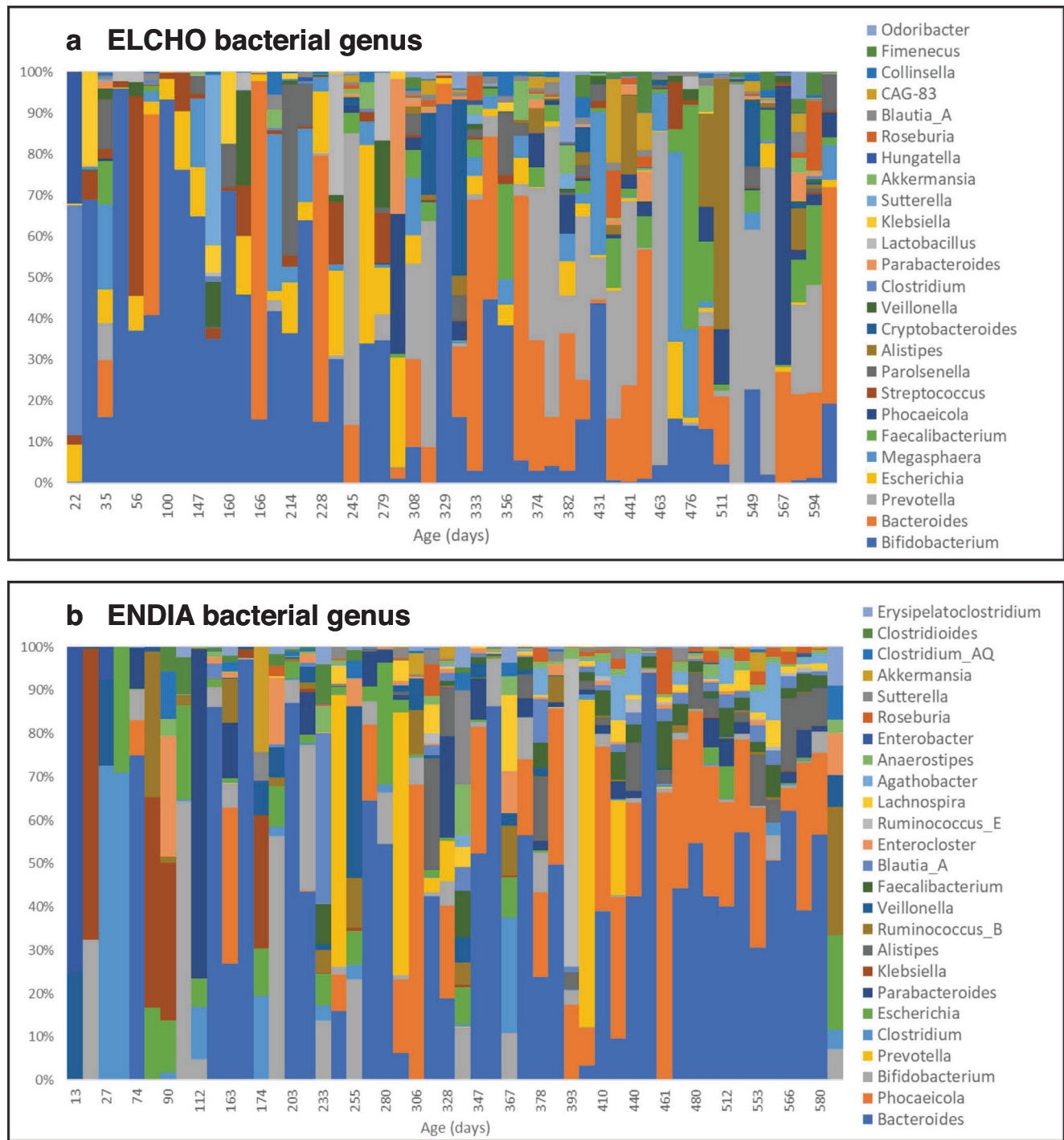


Fig. 1 | Bar charts of observed taxa (bacteria) proportions (25 most abundant) classified by MetaPhlAn. a Elcho (Indigenous) infants. **b** ENDIA (Non-indigenous) infants. For each population, samples were left to right by increasing age. Relative proportions of taxa are ranked highest from bottom to top.

Indigenous and non-Indigenous infants did not differ significantly by gestational age at delivery, mode of delivery, number of furred pets in the household, recent antibiotic exposure, or serum intestinal fatty acid binding protein (iFABP). They were distinguished, however, by lower birth weight, a markedly higher frequency of current and exclusive breastfeeding, later introduction of complementary feeding,

and markedly higher fecal calprotectin. BMI was significantly lower in Indigenous women but could only be obtained with their consent after pregnancy, when their infant was seen; in non-Indigenous women, BMI was documented at the time of pregnancy diagnosis. Significantly more Indigenous women were underweight and significantly more non-Indigenous women were obese (Supplementary Data 1).

Bar charts of observed taxa (viruses) proportions (25 most abundant) classified by Kraken2

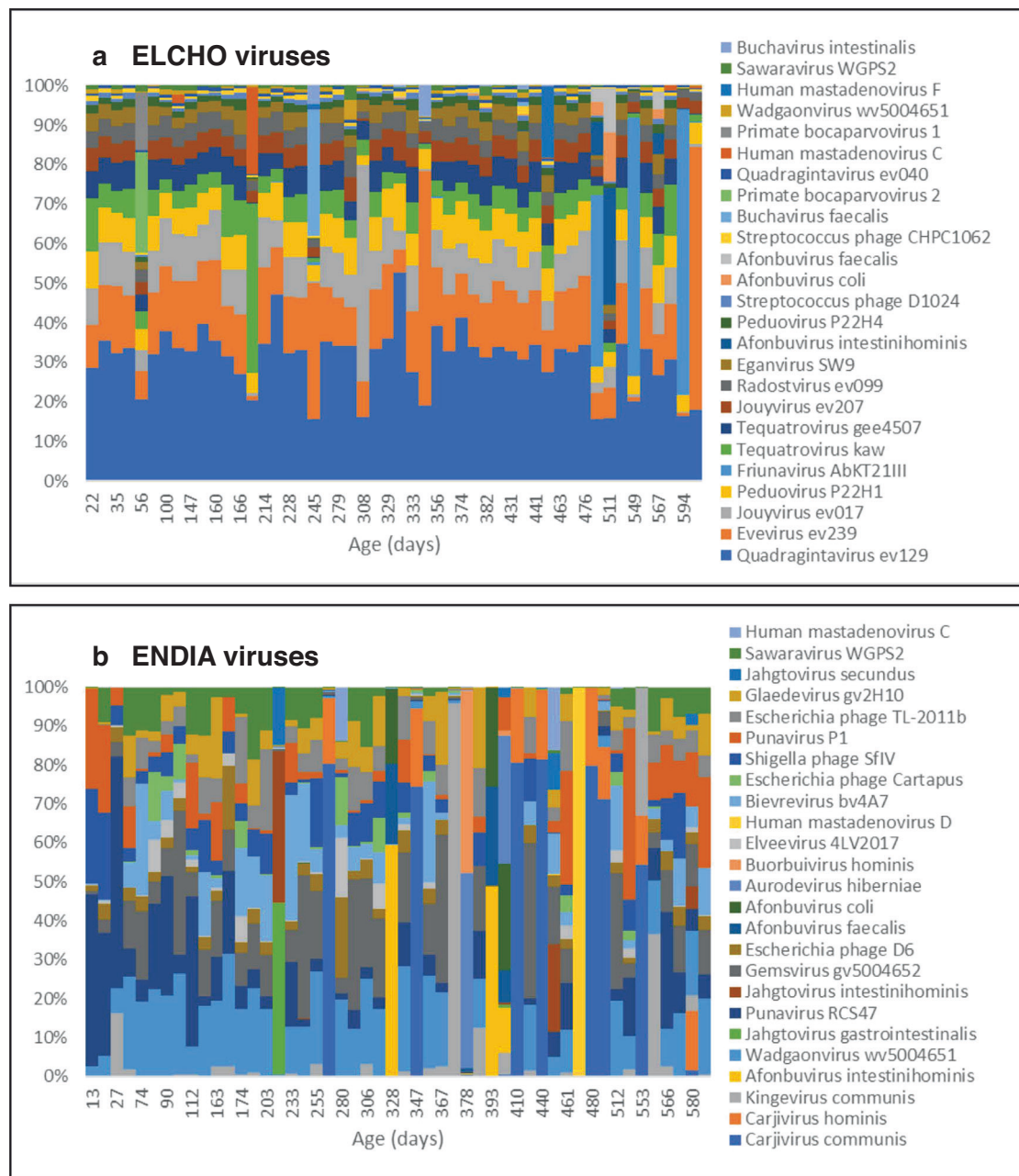


Fig. 2 | Bar charts of observed taxa (viruses) proportions (25 most abundant) classified by Kraken2. a Elcho (Indigenous) infants. **b** ENDIA (Non-Indigenous) infants. For each population, samples were ranked left to right by increasing age. Relative proportions of taxa are ranked highest from bottom to top.

Indigenous infants have distinct taxonomic profiles

MetaPhlan4 classified 1679 bacterial species (Indigenous = 1581; non-Indigenous = 1310) in the 100 samples. Kraken2-classified 371 viral (Indigenous = 367; non-Indigenous = 353), 31 fungal (Indigenous = 31; non-Indigenous = 31), and 15 other eukaryote species following their taxonomic confirmation with BLAST (Indigenous = 12; non-Indigenous = 10). Complete bacterial abundance data are provided in Supplementary Data 2; all Kraken2 virus and eukaryote abundance data are provided in Supplementary Data 3. Bacterial taxa proportions at the genus level (Fig. 1) revealed a high prevalence of *Bifidobacteria* in Indigenous infants, especially at younger ages, and of *Prevotella*, especially at older ages, in contrast

to non-Indigenous infants who had a high prevalence of *Phocaeicola* (*Bacteroides*) across the ages.

Viruses (Fig. 2) were significantly more abundant in Indigenous than non-Indigenous infants (17.2 vs. 2.9 million counts, respectively) and were different, and their abundance profile was more consistent in Indigenous infants, possibly due to their geographic homogeneity. In Indigenous infants, the dominant viruses were *Enterobacteria* (*Escherichia*) phages (as classified by the International Committee on Taxonomy of Viruses, <https://ictv.global/>), viz., *Quadragintavirus ev129*, *Tequatrovirus*, *Evevirus ev239*, and *Jouyvirus ev017*. These viruses were virtually non-existent in non-Indigenous infants. The most common viruses in non-Indigenous infants were the CrAssphages,

Bar charts of observed taxa (fungi) proportions (25 most abundant) classified by Kraken2

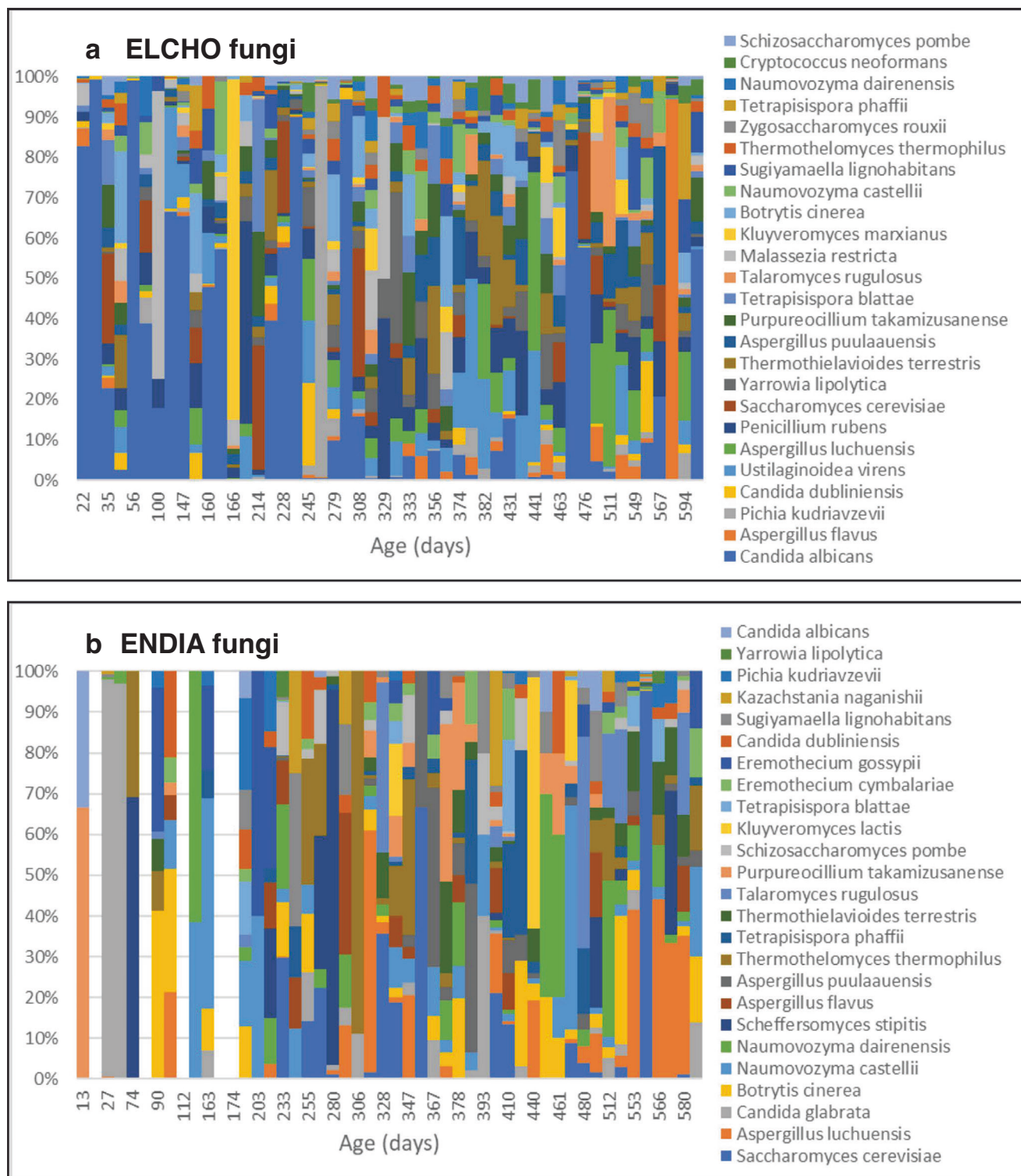


Fig. 3 | Bar charts of observed taxa (fungi) proportions (25 most abundant) classified by Kraken2. a Elcho (Indigenous) infants. **b** ENDIA (Non-indigenous) infants. For each population, samples were ranked left to right by increasing age.

Relative proportions of taxa are ranked highest from bottom to top. White columns are samples in which no classified fungi were detected.

Heatmap of non-fungal eukaryote counts classified by Kraken2

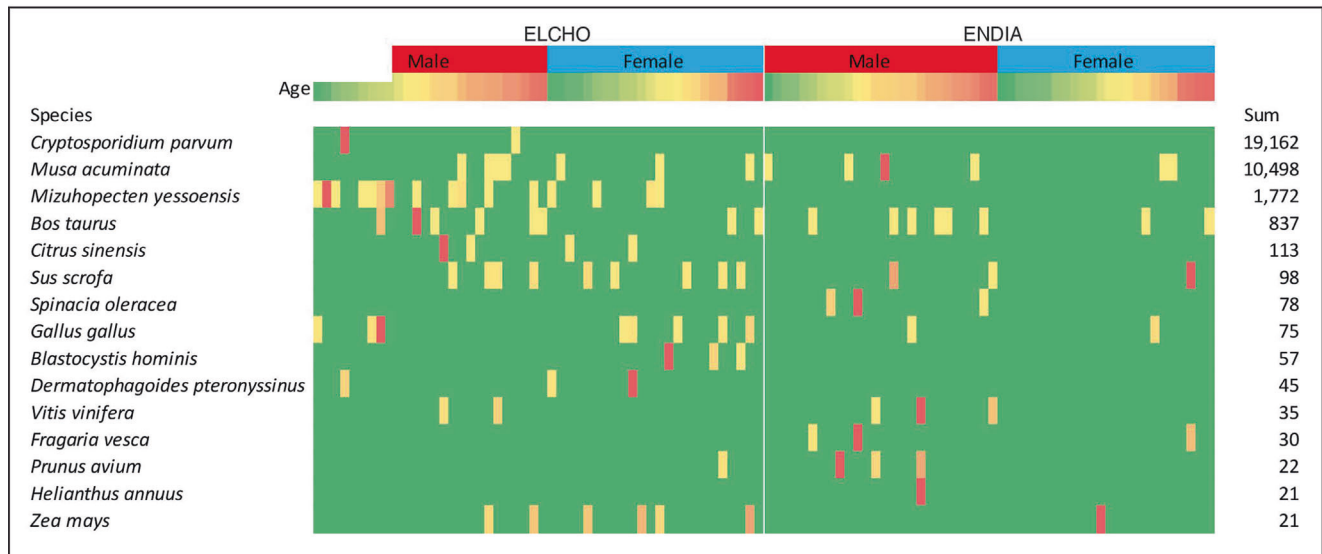


Fig. 4 | Heatmap of non-fungal eukaryote counts classified by Kraken2. Highest to lowest abundance is indicated by a red to green gradient.

Carjivirus communis, *Carjivirus hominis*, and *Kingevirus communis*, which infect *Bacteroides*¹⁷. The potential pathogen, human *Mastadenovirus*, appeared in the top 25 viruses in both Indigenous and non-Indigenous infants, being strain F and strains C and D, respectively. However, its occurrence was sporadic in both populations in only a few infants, but at high counts. The only other potentially pathogenic virus identified was *Primate bocaparvovirus 2* (*Human bocavirus 2c*), in Indigenous infants, which is known to cause respiratory tract infections.

Fungi (Fig. 3) were also significantly more abundant in Indigenous than non-Indigenous infants (mean counts/sample 4835 vs 71, respectively). *Candida albicans* (total counts 226,104) was by far the most abundant in Indigenous infants, especially at a younger age, and *Saccharomyces cerevisiae* (total counts 1122) was the most abundant in non-Indigenous infants, although this was predominantly due to a single individual (888 counts). After *S. cerevisiae*, *Aspergillus luchuensis* was most abundant in non-Indigenous infants (total counts 248), with counts spread more evenly over several samples.

Due to the limited number of non-fungal eukaryotic genomes classified in the Kraken2 database, species-level classification was not reliable without a BLAST check step. The abundances of the 15 identified and BLAST-confirmed non-fungal eukaryote taxa are shown as a heatmap (Fig. 4). Total counts of these classified eukaryotes were higher in Indigenous than non-Indigenous infants (22,190 and 10,674, respectively). Several animal food taxa were present in both Indigenous and non-Indigenous infants, namely *Bos taurus* (beef), *Sus scrofa* (pork), and *Gallus gallus* (chicken), as well as plant foods, e.g., *Zea mays* (maize), *Musa acuminata* (banana), and *Vitis vinifera* (grape). The mollusc, *Mizuhopecten yessoensis* (scallop) and *Citrus sinensis* (sweet orange) were present only in Indigenous infants, and *Spinacia oleracea* (spinach) and *Fragaria vesca* (strawberry) only in non-Indigenous infants. Non-food eukaryotes, two protozoan parasites, *Cryptosporidium parvum* and *Blastocystis hominis*, and the house dust mite, *Dermatophagoides pteronyssinus*, were present in two, three, and three Indigenous samples (seven infants), respectively. Parasites were detected microscopically in 15 (30%) Indigenous infants: protozoans (either *Cryptosporidium* spp or *Giardia intestinalis*) in 10 and helminths (*Trichiuris trichiuria* or *Ascaris lumbricoides*) in five (see also ref. 18).

Indigenous infants have greater bacterial alpha diversity

Indigenous infants had significantly greater alpha diversity than non-Indigenous infants, when observed at the family taxonomic level and above (e.g., Shannon [family] REML, $P=0.012$; see Table 2). Females from both populations had significantly higher diversity than males at all levels above species (e.g., Shannon [genus] $P=0.030$). As expected for the developing gut microbiome, alpha diversity increased with age at the species level in both populations (Shannon $P=0.042$), although this effect was not evident at higher taxonomic levels, except for the richness metric, indicating that while some degree of diversification occurs, it is more a case of species succession (see beta diversity below). Mode of delivery and presence of furred pets had no significant effect on alpha diversity. The exclusive breastfeeding for the first 6 months category contrasted with the currently breastfeeding category in having a significant positive effect on alpha diversity at most taxonomic levels above genus. REML tests were also carried out with corrections for both categories (shown in Supplementary Data 4). These corrections only slightly altered the observed significant comparisons in Table 2, with P values becoming non-significant in one currently breastfeeding and eight exclusively breastfeeding cases. Examples of significant Shannon indices ($P < 0.05$) for the two groups by population (family), sex (genus), and age category (species) are shown in Fig. 5. Alpha diversity results are presented in full in Supplementary Data 4.

Bacterial beta diversity distinguishes Indigenous and non-Indigenous infant populations

Beta diversity was significantly different between populations and between age categories at all taxonomic levels (see Table 3). This is evident in the distribution of samples in principal coordinate analysis (PCO) plots of beta diversity (Fig. 6). Breastfeeding was also significant in ADONIS tests, as expected because of its close linkage to population. A 3-D PCO plot, online at https://html-preview.github.io/?url=https://github.com/theo-allnut-bioinformatics/Indigenous_gut_microbiome_2023/blob/main/pco3d.html, was used to explore further potential clusters in the data, but none were clearly discerned. As for alpha diversity, currently breastfed and exclusively breastfed categories corrected for in ADONIS tests had no effect on the significant

Table 2 | Summary of P values (REML tests) for alpha diversity metrics (Richness, Shannon index, Simpson index) among groups of independent variables at each taxonomic level (phylum, class, order, family, genus, and species)

| P values for alpha diversity metrics | | | |
|--------------------------------------|---------------|---------------|---------------|
| Taxonomic level | | | |
| Group category | Richness | Shannon | Simpson |
| Phylum | | | |
| Population | 0.0049 | 0.0109 | 0.0219 |
| Age category | 0.0001 | 0.5152 | 0.8113 |
| Sex | 0.8870 | 0.0173 | 0.0087 |
| Mode of delivery | 0.3993 | 0.4292 | 0.3940 |
| Currently breastfeeding | 0.4828 | 0.0544 | 0.0881 |
| Exclusive breastfeeding | 0.0099 | 0.0008 | 0.0017 |
| Furred pets | 0.6724 | 0.7256 | 0.9178 |
| Class | | | |
| Population | 0.0092 | 0.0014 | 0.0060 |
| Age category | 0.0000 | 0.3535 | 0.6116 |
| Sex | 0.8122 | 0.0105 | 0.0057 |
| Mode of delivery | 0.3349 | 0.2763 | 0.2798 |
| Currently breastfeeding | 0.4095 | 0.0214 | 0.0474 |
| Exclusive breastfeeding | 0.0171 | 0.0001 | 0.0003 |
| Furred pets | 0.8059 | 0.5539 | 0.8864 |
| Order | | | |
| Population | 0.0705 | 0.0179 | 0.0172 |
| Age category | 0.0000 | 0.4704 | 0.8008 |
| Sex | 0.4600 | 0.0088 | 0.0048 |
| Mode of delivery | 0.7088 | 0.4918 | 0.4244 |
| Currently breastfeeding | 0.1048 | 0.1041 | 0.0610 |
| Exclusive breastfeeding | 0.1395 | 0.0017 | 0.0011 |
| Furred pets | 0.4965 | 0.8722 | 0.8437 |
| Family | | | |
| Population | 0.0061 | 0.0119 | 0.0232 |
| Age category | 0.0000 | 0.0394 | 0.3979 |
| Sex | 0.3975 | 0.0029 | 0.0024 |
| Mode of delivery | 0.7568 | 0.8261 | 0.9313 |
| Currently breastfeeding | 0.1148 | 0.6957 | 0.4565 |
| Exclusive breastfeeding | 0.0184 | 0.0019 | 0.0020 |
| Furred pets | 0.6440 | 0.7665 | 0.5839 |
| Genus | | | |
| Population | 0.3584 | 0.8368 | 0.7984 |
| Age category | 0.0002 | 0.0003 | 0.0252 |
| Sex | 0.1361 | 0.0295 | 0.0360 |
| Mode of delivery | 0.5407 | 0.6071 | 0.5852 |
| Currently breastfeeding | 0.0130 | 0.2224 | 0.3382 |
| Exclusive breastfeeding | 0.7728 | 0.3527 | 0.1857 |
| Furred pets | 0.7593 | 0.8887 | 0.9781 |
| Species | | | |
| Population | 0.3725 | 0.5741 | 0.6689 |
| Age category | 0.0000 | 0.0000 | 0.0002 |
| Sex | 0.2235 | 0.2012 | 0.2468 |
| Mode of delivery | 0.6456 | 0.7836 | 0.9151 |
| Currently breastfeeding | 0.0224 | 0.0773 | 0.2460 |
| Exclusive breastfeeding | 0.5002 | 0.4378 | 0.2619 |
| Furred pets | 0.6929 | 0.9846 | 0.8658 |

Significant P values (<0.05) are in bold.

uncorrected comparisons (see Supplementary Data 4, which also details the statistics).

Mode of delivery did not differ by frequency between the Elcho and ENDIA populations, and, as shown (Tables 2 and 3), alpha and beta diversity were not different by delivery mode across both populations. However, this would not exclude a population-specific effect. Therefore, we performed REML testing on alpha diversity and Adonis testing on beta diversity within each population for cesarean vs. vaginal delivery. All comparisons were non-significant (Supplementary Data 5), which demonstrates that the populations were not differentially affected by mode of delivery.

Bacterial abundance differentiates infant populations

Differential abundance (DA) data are summarized in Table 4. Detailed tables of classified taxa and abundances are presented in Supplementary Data 6 for population, and in Supplementary Data 7 for breastfeeding. For the population overall, DA was significant for 114 species, 38 genera, 12 families, 8 orders, 3 classes, and 2 phyla. The 25 most differentially abundant species prevalent in both populations (most positive or negative ALDEx2 effect) are shown in Fig. 7. Species virtually exclusive to Indigenous infants were *Megasphaera* spp., *Streptococcus lactarius*, *Caecibacter* spp., *Parolsenella* spp., and *Prevotella* spp; those almost exclusive to non-Indigenous infants were *Muricomes oroticus* and *Fecalibacillus* spp. *Bifidobacteria* spp. were dominant in Indigenous infants, and *Intestinibacter bartlettii*, *Clostridium AQ innocuum*, and *Ruminococcus* spp. were dominant in non-Indigenous infants.

At the genus level, *Megasphaera*, *Caecibacter*, *Parolsenella*, *Prevotella*, *Allisonella*, *Dialister*, *Thermophilibacter*, *Paratractidigestiibacter*, *Acidominococcus*, *UBA7748*, *Olsenella*, and *Coriobacterium* were virtually unique to Indigenous infants, whereas *Intestinibacter*, *Sellimonas*, *Ruminococcus*, *Muricomes*, *Clostridium*, *Ventrimonas*, *UBA9414*, *Hespelia*, and *Erysipelatoclostridium* were the dominant genera in non-Indigenous infants.

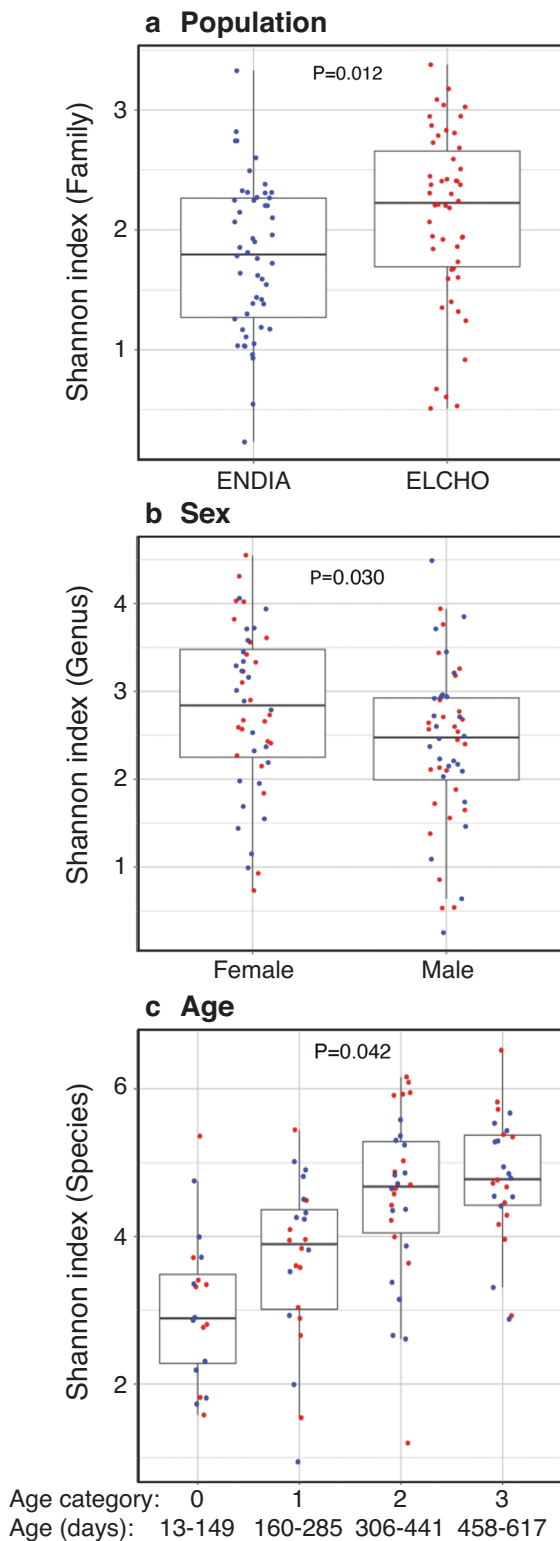
For currently breastfeeding, seven of the top 10 most differentially abundant taxa in Indigenous infants (*Megasphaera* spp., *Streptococcus lactarius*, *Megasphaera stantonii*, *Megasphaera elsdenii*, *Megasphaera cerevisiae*, *Caecibacter* spp., *Caecibacter hominis*) were shared with the population, as well as several *Bifidobacteria* spp, which is consistent with the much higher frequency of breastfeeding in this population. The effect of exclusively breast fed on DA was not examined separately because only three mothers were exclusively breastfeeding and not also currently breastfeeding, a category too small to correct for any effect of exclusive breastfeeding alone, and likely to be contained in the observed DA for currently breast fed.

DA was also analyzed separately on each age category (see Supplementary Data 8), aiming to reveal more details about age-dependent differences between the populations. Because of the smaller sample size in each category, fewer differentially abundant taxa were present in each, although the ages at which some taxa were also differentially abundant in the total populations were revealed. Thus, Indigenous infants had higher carriage of *Streptococcus* at younger ages and *Prevotella* and *Megasphaera* at older ages. *Clostridia* spp. were more abundant in non-Indigenous infants at older ages.

Functional profiling distinguishes infant populations

Metagenomic functional annotation using SUPER-FOCUS with the SEED database¹⁹ and analysis with Aldex2 revealed many differentially expressed functions, e.g., at SEED level 2, 18 functions were increased in Indigenous compared to non-Indigenous infants, and 48 were increased in non-Indigenous compared to Indigenous infants ($P < 0.05$) (see Supplementary Data 9). In general, metabolism-related functions (involving RNA, nucleotides, proteins, DNA, lipids) were

Dotplots of alpha diversity (Shannon index)



significantly increased in Indigenous infants, in some cases reflecting the relative abundance of certain taxa we had identified (e.g., Streptococci, viruses). Other functions (e.g., tetrapyrroles, monosaccharides) were increased in non-Indigenous infants. It is not possible to assign these functions to specific bacterial taxa, because reads are mapped to the SEED database in SUPER-FOCUS

Fig. 5 | Dotplots of alpha diversity (Shannon index) for taxonomic levels found to be significantly different between groups (REML, $P < 0.05$). **a** Population (family); **b** sex (genus); and **c** age (species). Elcho (Indigenous) = red dots; ENDIA (non-Indigenous) = blue dots. Vertical lines = 5th–95th percentile; boxes = 25th–75th percentile; horizontal lines = means. The selection of age ranges is described in “Methods.” In (c), pair-wise comparisons between age groups were significant (each $P = 0.006$), except for 13–149 vs. 160–285 days and 306–441 vs. 458–617 days ($P = 1.000$). This information is also in Supplementary Data 4.

independently of those mapped for taxonomic classification. Also, without transcriptomic data, any inferences from changes in bacterial function between Indigenous and non-Indigenous infants would be questionable. Overall, the differences in function reflect those in phenotypic diversity between the populations, and further comment on the possible significance of individual functions would be speculative.

Markers of gut pathology in infant populations

Derived from neutrophils, calprotectin in feces is employed as a marker of gut inflammation²⁰. The concentration of fecal calprotectin is higher in very young children^{21,22}, but the reference range is not well defined. In young Finnish children, the upper limit is regarded as 100 mg/g²¹. At this value, fecal calprotectin was increased in 46/50 (92%) Indigenous infants and 12/50 (24%) non-Indigenous infants (median [IQR]: 1318 [515, 1809] vs 39.9 [4.40,102], respectively; $P = 0.0001$) (Table 1). Males (1287 mg/g) and females (1123 mg/g) did not differ.

Serum iFABP is a marker of gut epithelial integrity²³. A reference range is not available for young children. Therefore, based on the median value of a control group in a study of childhood inflammatory bowel disease²⁴, we used the $1.5 \times$ IQR outlier rule to define the upper limit as 1664 pg/ml. Serum iFABP was increased in 6 (12%) of Indigenous infants and 7 (14%) of non-Indigenous infants (median [IQR]: 637 [419–1091] vs 689 [517–1572], respectively; $P = 0.126$) (Table 1).

Discussion

Metagenomic sequencing revealed major differences in the gut microbiomes between Australian Indigenous infants living remotely and age- and sex-matched Australian non-Indigenous infants living in urban environments. This is the first such comparative study of the gut microbiome in infants of which we are aware. The gut microbiome of the Indigenous infants shared features with gut microbiomes of other pre-industrialized societies. It comprised significantly greater numbers of bacteria, viruses, and fungi than the non-Indigenous gut microbiome, and displayed greater bacterial diversity, 114 bacterial species being differentially abundant. Some taxa present in Indigenous infants, e.g., species from the families *Prevotellaceae*, *Spirochaetaceae*, *Succinivibrionaceae*, and *Veillonellaceae*, were absent in non-Indigenous infants. This is reminiscent of the “VANISH” (volatile and/or associated negatively with industrialized societies of humans) bacterial species, reported as missing in other studies of modern, urban societies compared to traditional hunter-gatherer or agricultural societies^{11,25}. Indigenous infants had a higher abundance of *Bifidobacteria*, commensurate with their high frequency of breastfeeding, provisioning milk oligosaccharides that promote the growth of *Bifidobacteria*²⁶. *Prevotella*, a marker of non-urban, pre-industrial microbiomes¹⁰, was prevalent in the older Indigenous infants, but not in the non-Indigenous infants, whereas *Phocaeicola* (*Bacteroides*), previously noted to be more abundant in urban societies^{10,11}, characterized non-Indigenous infants across the ages. Differences in observed taxa could not be attributed to recent antibiotic (mainly amoxicillin) usage, which applied to only a minority in each population.

Our findings show that the gut microbiome of Yolngu Indigenous infants retains key features of traditional gut microbiomes and appears not to have been substantially modified at this stage by encroaching

Table 3 | Summary of beta diversity tests

| Beta diversity category | Phylum | Class | Order | Family | Genus | Species |
|-------------------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Population | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| Age category | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| Sex | 0.3730 | 0.3760 | 0.4080 | 0.3520 | 0.4340 | 0.6990 |
| Mode of delivery | 0.8960 | 0.8060 | 0.8310 | 0.7430 | 0.7390 | 0.6950 |
| Breastfeeding | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| Exclusive breastfeeding | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 | 0.0010 |
| Furred pets | 0.7450 | 0.7310 | 0.7350 | 0.6780 | 0.8920 | 0.8040 |

ADONIS *P* values among categories are shown for each taxonomic level (phylum, class, order, family, genus, and species). Significant *P* values (<0.05) are in bold.

Westernization. The infant gut microbiome is shaped initially from birth by vertical transmission of microbiota primarily from the mother's gut, and to a lesser extent from the vagina and skin, and breast milk²⁷. Mode of birth for each population and time of introduction of complementary food did not differ between Indigenous and non-Indigenous infants. Therefore, it seems reasonable to suggest that the unique gut microbiome of Indigenous infants reflects a persisting inter-generational influence of the remote environment, traditional diet, and lifestyle. Host-microbiota relationships evolve in response to diverse environmental modifiers, including exposure to family members and other humans, animals, plants, air, soil, and water, as well as man-made products (e.g., chemicals, antibiotics, drugs, food additives, etc.) and poorly-defined psycho-social "stressors." These exposures obviously differ between the Indigenous and non-Indigenous infants. We did not set out to identify them or the relative roles of environment and host genetics, as this would require a separate study of Yolngu infants away from their traditional environment, which is unlikely to be feasible. Nevertheless, the greater diversity of the gut microbiome in Indigenous infants is likely to be due to their distinctive exposures, and those of their mothers and other family members, to a more diverse natural environment compared to the built, urban, and more "hygienic" environment of the non-Indigenous infants.

Two markers of gut pathology were measured in this study. Neutrophil-derived fecal calprotectin was significantly higher in Indigenous infants. While this would be expected to reflect gut inflammation, this is not necessarily the case. Increased fecal calprotectin has been associated with gut bacteria that promote inflammation²⁰, but these were not present in the Indigenous infants. Parasites modify the gut microbiome²⁸, but they were detected in only a small minority of Indigenous infants, and fecal calprotectin was similar in infants with parasites compared to the population as a whole (median: 1133 vs 1318 mg/g). Fecal calprotectin was shown to be a marker of Environmental Enteric Dysfunction (EED), a low-grade inflammatory disorder associated with stunting in children living in poor communities²⁹, but the Indigenous infants did not display criteria for EED. Furthermore, iFABP, a marker of intestinal epithelial damage and permeability^{23,24}, was not increased in Indigenous compared to non-Indigenous infants. Importantly, fecal calprotectin in infancy may also be elevated for non-pathological reasons, a major one being exclusive breastfeeding^{30,31}. This is a plausible explanation for the difference between the Indigenous and non-Indigenous infants, in whom the proportions exclusively breastfed were 84% and 22%, respectively, closely matching those for raised fecal calprotectin. Follow-up studies as the children age and cease breastfeeding may resolve the fecal calprotectin question.

A caveat of this study, which limits our conclusions, was the unavailability of microbiome data on the mothers, as permission could not be obtained to collect samples from them. A further caveat is that while infants were closely matched for age and sex, the non-Indigenous infants had a first-degree relative with type 1 diabetes and were therefore at increased genetic risk for type 1 diabetes. Although none had detectable pancreatic islet autoantibodies, the earliest known marker of sub-clinical disease, we can't exclude the

possibility that their gut microbiome differs from that of children without genetic susceptibility to type 1 diabetes. Nevertheless, we show that the Indigenous infants begin life with a distinctive ancestral gut microbiome. This suggests that if Westernization of the gut microbiome occurs in them, it will be acquired and not be transmitted intergenerationally. Our findings may only be a snapshot in the ongoing inter-generational extinction of the ancestral microbiome³², but this remains to be seen. They extend knowledge of the infant gut microbiome and are a foundation on which to explore environmental and lifestyle factors that shape the development of the gut microbiome and its relationship to future health in Indigenous children.

Methods

Study populations

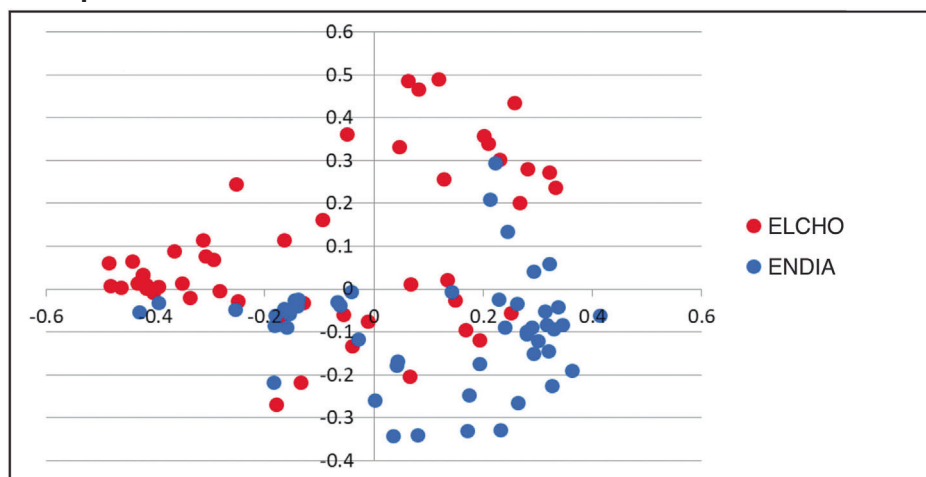
Fifty Indigenous infants, 24 females and 26 males, aged 22–617 days, numbering all available infants under 2 years of age who could be accessed, were recruited in October 2017 in a remote community in north-east Arnhem Land, Northern Territory, Australia, in the Early Life Child Health Observation (Elcho) study. As described previously¹⁸, prior to commencing the study and after community engagement, the local research team participated in a week-long codesign and training program which involved telling the research story in the local language, with discussion of the study protocol, means of recruitment, data collection, and consent. In this program, the concept of the microbiome was discussed using metaphors and microscopy to develop an understanding of the role of microscopic organisms in human health. Prior to enrollment, parents or guardians gave written informed consent on behalf of infant participants. After explanation and discussion in both English and the local language, trained research staff collected maternal socio-demographic, nutritional, environmental, breastfeeding, and dietary data using a structured questionnaire. The majority of infants were being breastfed. On up to 3 days a week, their mothers consumed traditional foods, viz., seafoods such as turtle, shellfish, fish, oysters and crabs, mangrove worms, game such as kangaroo, bush fruits, plant roots, and tuber-like yams. The study protocol was approved by the Human Research Ethics Committee of the Northern Territory Department of Health and Menzies School of Health Research (Ref. 2017-2814), Melbourne Health Human Research Ethics Committee (Ref. 2017.064), Miwatj Health Indigenous Corporation Board, and the Local Shire Authority.

Fecal samples were obtained within 2 h from freshly soiled diapers, transferred to sterile 5 mL screw cap containers, immediately frozen at -20°C , and transported on dry ice to the Peter Doherty Institute, Melbourne, where they were stored at -80°C for 5 months before DNA extraction and metagenomic sequencing were performed at the Walter and Eliza Hall Institute, Melbourne. All samples were coded and de-identified before being analyzed in a blinded manner.

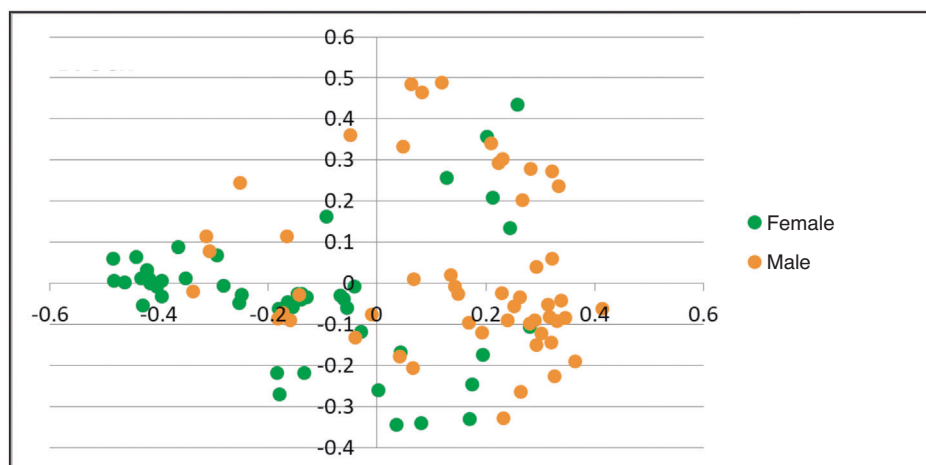
Indigenous Australian infants in this Elcho study were matched (see Supplementary Data 1) for sex, and as closely as possible for age, with non-Indigenous infants participating in the Australia-wide Environmental Determinants of Islet Autoimmunity (ENDIA) pregnancy-birth cohort study (Australia New Zealand Clinical Trials Registry

Principal coordinate plots of beta diversity

a Population



b Sex



c Age

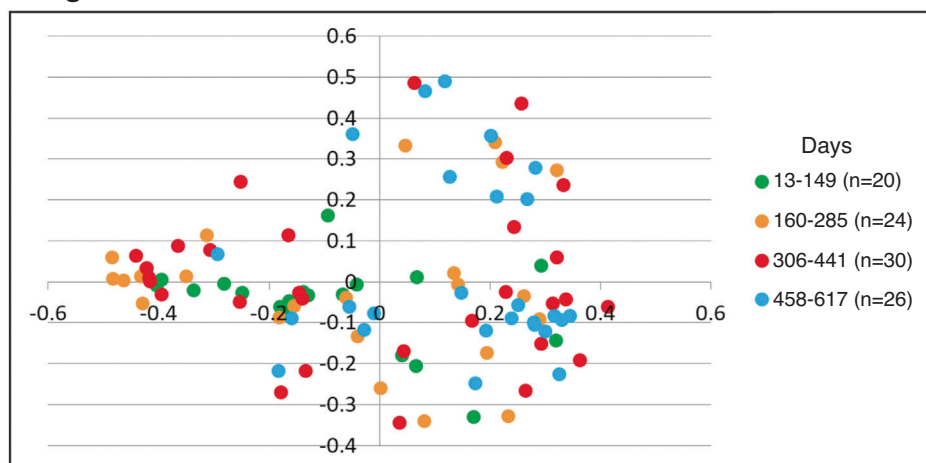


Fig. 6 | Principal coordinate analysis (PCO) plots of beta diversity at the species level. Each plot is the same PCO with the points colored to indicate different categories or groups. **a** Population; **b** sex; **c** age; PCO1 (x-axis) = 13.3% and PCO2 (y-axis) = 6.7% of total variation. Selection of age ranges is described in “Methods”.

ACTRN12613000794707), in which the child has a first-degree relative with type 1 diabetes³³. Participants selected for this study were from urban areas of NSW, Victoria, South Australia, Western Australia, and Queensland. Parents or guardians gave written informed consent for their children to participate in ENDIA research, including collaborative

studies. Less than 50% of ENDIA infants in this study were being breastfed at the time of collection of stool samples, which were processed similarly to those of the Indigenous infants. Human Research Ethics Committee (HREC) approval was obtained at each clinical site, with the Women’s and Children’s Hospital, Adelaide, acting as the lead

Table 4 | Numbers of differentially abundant taxa between populations (Elcho vs. ENDIA) and between currently and not currently breastfeeding categories according to ALDEx2 effect size statistic corrected P value < 0.05

| | | Phylum | Class | Order | Family | Genus | Species |
|-------------------------|-------|--------|-------|-------|--------|-------|---------|
| Population | Up | 0 | 1 | 2 | 5 | 21 | 73 |
| | Down | 2 | 2 | 6 | 7 | 17 | 41 |
| | Total | 2 | 3 | 8 | 12 | 38 | 114 |
| Currently breastfeeding | Up | 2 | 2 | 4 | 4 | 8 | 21 |
| | Down | 1 | 1 | 4 | 5 | 17 | 98 |
| | Total | 3 | 3 | 8 | 9 | 25 | 119 |

“Up” = increased abundance in Elcho vs. ENDIA or currently breastfeeding vs. not currently breastfeeding; “Down” = decreased abundance in the same.

HREC site under the Australian National Mutual Acceptance Scheme (HREC/16/WCHN/066). Fecal samples were collected similarly to those from the Indigenous children. None of the matched non-Indigenous ENDIA infants had autoantibodies to pancreatic islet antigens, a marker of sub-clinical T1D, although five subsequently became seropositive after 4 years of age.

Whole metagenome sequencing and taxonomic analysis

DNA was extracted with the MoBio PowerSoil kit (MoBio Laboratories, Carlsbad, CA) and whole metagenome sequencing (WMS) libraries generated as previously described³⁴. Sequencing by 2 × 150 bp paired-end chemistry was performed on an Illumina NovaSeq 6000 (Illumina, San Diego, California, USA) machine by the Ramaciotti Centre for Genomics (Sydney, Australia).

Illumina reads for each sample were filtered by KneadData (v0.7.7 <https://github.com/biobakery/kneaddata>) using default settings. Data were then further filtered to remove low entropy reads using a script based on the Shannon information index (<https://github.com/theo-allnut-bioinformatics/scripts/blob/master/shannons-filter.py>). Where possible, following filtering, read counts were capped at 10 million per sample.

Diversity analysis

All bioinformatic pipelines, scripts, and program settings are available at https://github.com/theo-allnut-bioinformatics/Indigenous_gut_microbiome_2023. The bacterial taxonomic composition of infant gut metagenomic samples was profiled with MetaPhlan 4.0³⁵. MetaPhlan classifications were converted to GTDB species taxonomy (<https://gtdb.ecogenomic.org/>) and taxa counts normalized to counts per million (cpm). Only taxa with >100 total counts (prior to normalization) and containing at least three non-zero samples were retained for analysis. Alpha diversity (diversity within microbial communities) metrics, viz., Richness, Shannon, and Simpson indices, were calculated for the taxonomic levels Phylum, Class, Order, Family, Genus, and Species using USEARCH v10.0.240³⁶. Viruses, fungi, and higher eukaryotes were classified and quantified using Kraken2³⁷, as previously described³⁸. The identity of Kraken2-classified taxa was checked by BLAST (nt database, 28/8/2022). Taxa with a predominant BLAST match other than the Kraken2 classification were excluded. Counts obtained from Kraken2 classifications were not normalized or adjusted.

Bacterial differences in alpha diversity were assessed using restricted maximum likelihood (REML) models implemented in the R package “lmer.” Each independent variable (“population,” “sex,” “currently breastfeeding,” “exclusively breastfeeding in first six months,” “presence of furry pets,” and “age category”) was initially tested separately in univariate models. Age was treated as a categorical variable to reflect distinct stages of microbiota development during infancy. The four age categories were: age 0 (13–149 days, $n = 20$), age 1 (160–285 days, $n = 24$), age 2 (306–441 days, $n = 30$), and age 3 (458–617 days, $n = 26$). These categories were balanced between Indigenous and non-Indigenous infants to reduce confounding in

comparisons by population. Given the modest sample size ($n = 100$ total), we did not adjust for age in multivariable REML models due to the risk of overfitting and loss of statistical power.

Beta diversity (diversity between microbial communities) was analyzed using the R package “pairwiseAdonis” with Bray-Curtis distances, an implementation of Permanova (<https://github.com/pmartinezarbizu/pairwiseAdonis>), for the same variables as alpha diversity. Models were fitted separately for each covariate of interest, including “population,” “sex,” and feeding-related variables. Age category was not included as a covariate in these models due to concerns about statistical power and model complexity, given the sample size.

Alpha and beta diversity were analyzed separately for the taxonomic levels: phylum, class, order, family, genus, and species.

Differential abundance

DA of raw counts between Indigenous Elcho and non-Indigenous ENDIA populations overall, and within each age category, was tested at each bacterial taxonomic level using ALDEx2 v1.31.0³⁹. The Benjamini–Hochberg-corrected P value of Welch’s t -test was used to determine significance ($P < 0.05$), and significant results were ranked by the ALDEx2 effect-size metric. DA of Kraken2 counts was also tested with ALDEx2, but no lower threshold was applied to the number of non-zero count samples, and only the species level was tested; a total abundance count threshold of 300 was used for viruses and 100 for fungi and other eukaryotes. It should be noted that, due to its limited coverage, the eukaryotic Kraken2 database classification of species³⁵ should be regarded as indicative only and not necessarily quantitative.

Metagenomic functional profiling

Functional annotation of metagenomic data was performed with SUPER-FOCUS¹⁹.

Fecal calprotectin

Fecal calprotectin (mg/g) was measured by quantitative, enzyme-linked immunoassay (CALPRO Oslo, Norway), according to the manufacturer’s instructions.

Serum intestinal fatty acid binding protein (iFABP)

Serum iFABP (pg/ml) was measured by ELISA (Enzyme-Linked Immunosorbent Assay; Hycult Biotech, The Netherlands), according to the manufacturer’s instructions.

Fecal parasites

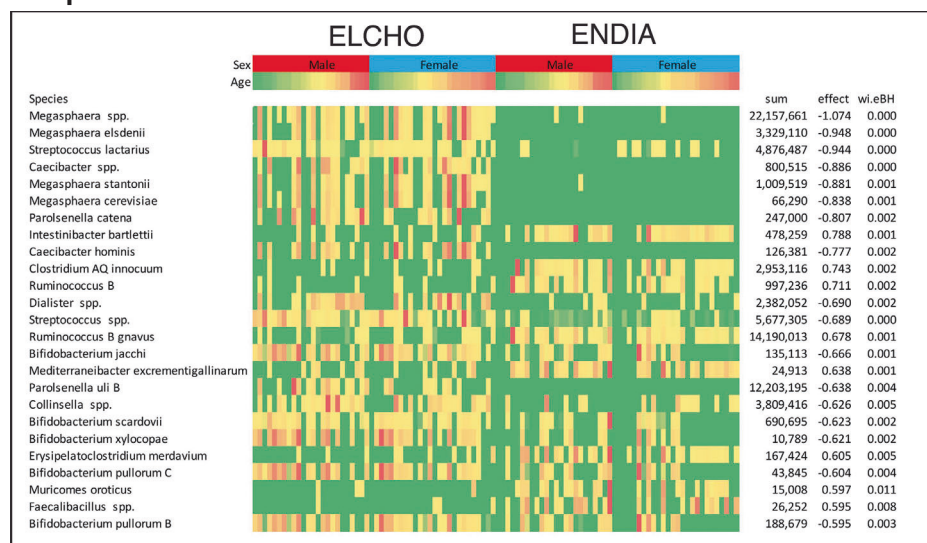
Parasites in fecal samples were analyzed directly by microscopy, both in the field and following fixation and storage in sodium-acetate formalin, as previously described¹⁸.

Statistics

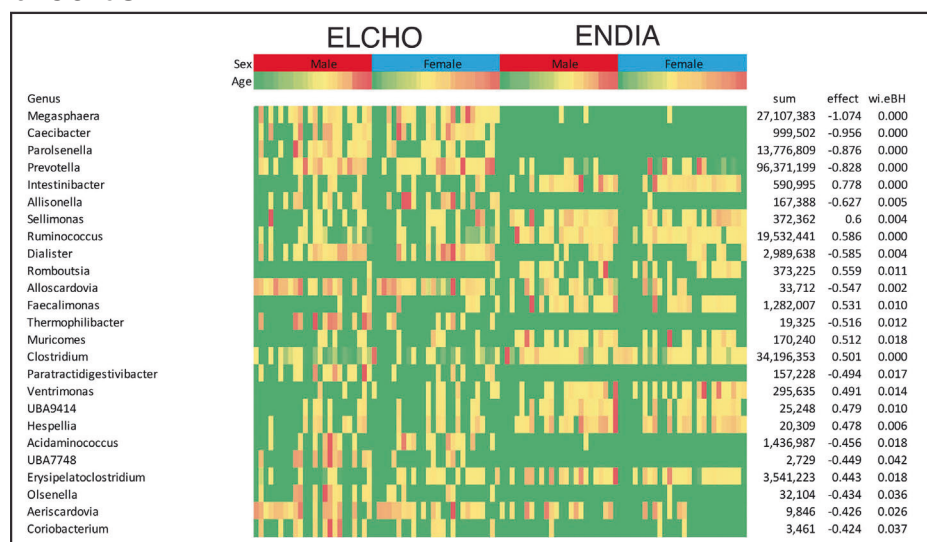
Group differences between calprotectin and iFABP biomarkers were analyzed by a non-parametric, two-tailed Mann–Whitney test, and proportions were compared by Fisher’s Exact test, using GraphPad

Significant differentially abundant taxa between ELCHO and ENDIA infants

a Species



b Genus



c Family

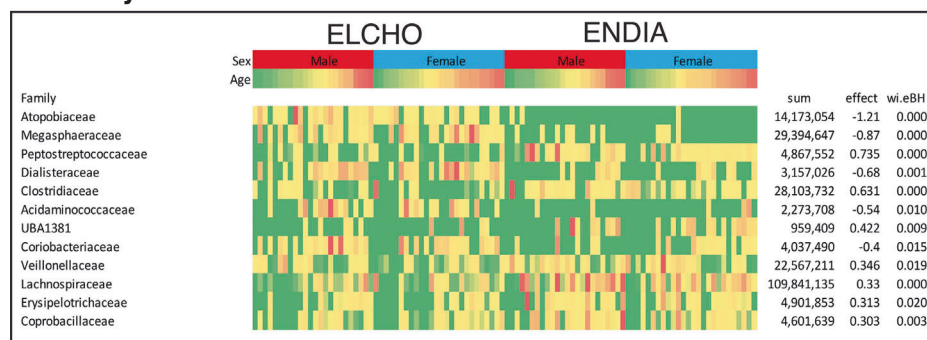


Fig. 7 | Differentially abundant taxa between ELCHO and ENDIA infants. Significantly differentially abundant taxa are shown at the species (a), genus (b) and family (c) levels (ALDEx2 BH < 0.05) (BH = expected Benjamini–Hochberg-corrected *P* value of Welch’s *t*-test). ALDEx2 was performed using CLR

normalization, but the figures depict raw counts. For each taxon, highest to lowest abundance is indicated by red to green gradient. Taxa are sorted by effect size, lowest to highest, top to bottom, i.e., increased abundance in Elcho vs. ENDIA.

Prism version 10.4.1 for macOS, GraphPad Software. www.graphpad.com.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data presented in this study are included in the article and Supplementary Data. In addition, the raw data and their descriptions are deposited in the European Nucleotide Archive (ENA) (<https://www.ebi.ac.uk/ena/browser/submit>) BioProject, under the ID PRJEB87209. Further inquiries can be directed to the corresponding author. Data on individual living humans cannot be publicly available due to their sensitive nature, as regulated by privacy legislation. Individual-level data cannot be made publicly available because disclosure would risk re-identification of participants and would contravene the Australian Privacy Principles Guidelines under Section 28(1) of the Privacy Act 1988 as administered by the Office of the Australian Information Commissioner (OAIC). The ethical framework set by the Australian National Statement on Ethical Conduct in Human Research also requires confidentiality and proportional, risk-minimizing data sharing. De-identified individual-level data can be made available to qualified researchers upon reasonable request, subject to prior approval or waiver by an accredited Human Research Ethics Committee (HREC) or international equivalent, execution of a Material Access Agreement with the relevant data custodian, and commitment to OAIC-aligned safeguards (e.g., secure storage, no re-identification, destruction/return at project end). Requests should be directed to L.C.H. (harrison@wehi.edu.au).

Code availability

All scripts and code written for this study are available in a GitHub repository: https://github.com/theo-allnut-bioinformatics/Indigenous_gut_microbiome_2023.

References

- Rasmussen, M. et al. An Indigenous Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
- Zhao, Y., Connors, C., Wright, J., Guthridge, S. & Bailie, R. Estimating chronic disease prevalence among the remote Aboriginal population of the Northern Territory using multiple data sources. *Aust. N. Z. J. Public Health* **32**, 307–313 (2008).
- Horwood, P. F. et al. Health challenges of the Pacific Region: Insights from history, geography, social determinants, genetics, and the microbiome. *Front. Immunol.* **10**, 2184 (2019).
- Bennett, J. M., Reeves, G., Billman, G. E. & Sturmberg, J. P. Inflammation—nature’s way to efficiently respond to all types of challenges: Implications for understanding and managing “the epidemic” of chronic diseases. *Front. Med.* **5**, 316 (2018).
- Hotamisligil, G. S. Inflammation, metaflammation and immunometabolic disorders. *Nature* **542**, 177–185 (2017).
- Byndloss, M. & Bäumlner, A. The germ-organ theory of non-communicable diseases. *Nat. Rev. Microbiol.* **16**, 103–110 (2018).
- West, C. E. et al. The gut microbiota and inflammatory non-communicable diseases: Associations and potentials for gut microbiota therapies. *J. Allergy Clin. Immunol.* **135**, 3–13 (2015).
- Singh, R. K. et al. Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15**, 73 (2017).
- Mitchell, T. Colonial trauma: complex, continuous, collective, cumulative, and compounding effects on the health of Indigenous Peoples in Canada and beyond. *Int. J. Indig. Health* **14**, 74–94 (2019).
- Gupta, V. K., Paul, S. & Dutta, C. Geography, ethnicity or subsistence specific variations in human microbiome composition and diversity. *Front. Microbiol.* **8**, 1162 (2017).
- Sonnenberg, E. D. & Sonnenberg, J. L. The ancestral and industrialized gut microbiota and implications for human health. *Nat. Rev. Microbiol.* **17**, 383–390 (2019).
- Di Vincenzo, F. et al. Gut microbiota, intestinal permeability, and systemic inflammation: a narrative review. *Intern. Emerg. Med.* **19**, 275–293 (2024).
- Tonkin, E. et al. Dietary intake of Indigenous Australian children aged 6–36 months in a remote community: a cross-sectional study. *Nutr. J.* **19**, 34 (2020).
- Cox, A. J., Moscovis, S. M., Blackwell, C. C. & Scott, R. J. Cytokine gene polymorphism among Indigenous Australians. *Innate Immun.* **20**, 431–439 (2014).
- Handsley-Davis, M. et al. Heritage-specific oral microbiota in Indigenous Australian dental calculus. *Evol. Med. Public Health* **10**, 352–362 (2022).
- Coleman, A. et al. Upper respiratory tract microbiome of Australian Indigenous and Torres Strait Islander children in ear and nose health and disease. *Microbiol. Spectr.* **9**, e0036721 (2021).
- Shkoporov, A. N. et al. ΦCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides intestinalis*. *Nat. Commun.* **9**, 4781 (2018).
- Hanieh, S. et al. Enteric pathogen infection and consequences for child growth in young Indigenous Australian children: a cross-sectional study. *BMC Infect. Dis.* **21**, 9 (2021).
- Silva, G. G. Z., Green, K., Dutilh, B. E. & Edwards, R. A. SUPER-FOCUS: A tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* **32**, 354–3561 (2016).
- Jukic, A. et al. Calprotectin: from biomarker to biological function. *Gut* **70**, 1978–1988 (2021).
- Kolho, K. L. & Alfthan, H. Concentration of fecal calprotectin in 11,255 children aged 0–18 years. *Scand. J. Gastroenterol.* **55**, 1024–1027 (2020).
- Peura, S. et al. Normal values for calprotectin in stool samples of infants from the population-based longitudinal Born Into Life study. *Scand. J. Clin. Lab. Investig.* **78**, 120–124 (2017).
- Huang, X., Zhou, Y., Sun, Y. & Wang, Q. Intestinal fatty acid binding protein: a rising therapeutic target in lipid metabolism. *Prog. Lipid Res.* **87**, 101178 (2022).
- Logan, M. et al. Intestinal fatty acid binding protein is a disease biomarker in paediatric coeliac disease and Crohn’s disease. *BMC Gastroenterol.* **22**, 260 (2022).
- Carter, M. M. et al. Ultra-deep sequencing of Hadza hunter-gatherers recovers vanishing gut microbes. *Cell* **186**, 3111–3124.e13 (2023).
- Marcobal, A. et al. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **10**, 507–514 (2011).
- Ferretti, P. et al. Mother-to-infant microbial transmission from different body sites shapes the developing infant gut microbiome. *Cell Host Microbe* **24**, 133–145 (2018).
- Leung, J. M., Graham, A. L. & Knowles, S. C. L. Parasite-microbiota interactions with the vertebrate gut: synthesis through an ecological lens. *Front. Microbiol.* **9**, 843 (2018).
- Crane, R. J., Jones, K. D. J. & Berkley, J. A. Environmental enteric dysfunction: an overview. *Food Nutr. Bull.* **36**, S76–S87 (2015).
- Dorosko, S. M., Mackenzie, T. & Connor, R. I. Fecal calprotectin concentrations are higher in exclusively breastfed infants compared to those who are mixed-fed. *Breastfeed. Med.* **3**, 117–119 (2008).
- Savino, F. et al. High fecal calprotectin levels in healthy, exclusively breast-fed infants. *Neonatology* **97**, 299–304 (2010).
- Sonnenburg, E. D. et al. Diet-induced extinctions in the gut microbiota compound over generations. *Nature* **529**, 212–215 (2016).

33. Penno, M. A. S. et al. Environmental determinants of islet autoimmunity (ENDIA): a pregnancy to early life cohort study in children at risk of type 1 diabetes. *BMC Pediatr.* **13**, 124 (2013).
34. Roth-Schulze, A. J. et al. Type 1 diabetes in pregnancy is associated with distinct changes in the composition and function of the gut microbiome. *Microbiome* **9**, 167 (2021).
35. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlan 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
36. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
37. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
38. Allnutt, T. R., Roth-Schulze, A. J. & Harrison, L. C. Expanding the taxonomic range in the fecal metagenome. *BMC Bioinform.* **22**, 312 (2021).
39. Fernandes, A. D. et al. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. *PLoS ONE* **8**, e67019 (2013).

Acknowledgements

The authors acknowledge members of the Early Life Child Health Observation Project Team: Elizabeth Bungawara, Lloyd Dhamarandji, Yalurr Dhamarandji, David Djilimara, Janice Djiliri, Jess Gatti, Noella Goveas, Jannie Kraayenhof, Norbert Ryan, and Jenny Shield. The authors thank the participants, their families, and health workers in the community, and, for their support, Miwatj Health Aboriginal Corporation, the Marthakal Homelands Health Service, Families as First Teachers, and Beth Hilton-Thorpe and Christalla Hajisava. The Indigenous studies were supported by a grant from the Hallmark Indigenous Research Initiative at the University of Melbourne and a Royal Melbourne Hospital Home Lottery Grant (GIA-060-2018). The ENDIA studies were supported by JDRF Australia, the recipient of the Commonwealth of Australia grant for Accelerated Research under the Medical Research Future Fund (grant keys 3-SRA-2023-1374-M-N, 3-SRA-2020-966-M-N, 1-SRA-2019-871-M-B, 4-SRA-2015-127-M-B), and with funding from The Leona M. and Harry B. Helmsley Charitable Trust. LCH was the recipient of an Investigator Grant (APP 1173945) from the National Health and Medical Research Council of Australia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

B.-A.B., S.H. and L.C.H. conceived the study. L.C.H., T.R.A., S.H., A.J.R.-S. and B.-A.B. provided methodology. S.H., G.G., V.G., J.J.C., M.E.C., E.A.D., T.H., G.S., J.M.W., P.V., M.A.S.P., L.C.H. and B.-A.B. provided resources. L.C.H., B.-A.B., S.H., G.G., V.G., J.J.C., M.E.C., E.A.D., T.H., G.S., J.M.W.,

P.V. and M.A.S.P. supervised the study. L.C.H., T.R.A., S.H., A.J.R.-S., M.A.S.P. and B.-A.B. curated data. K.M.N., N.L.S., E.B.-S., and L.B. performed assays. T.R.A., A.J.R.-S., L.C.H., S.H., N.L.S., E.B.-S. and M.A.S.P. analyzed data. L.C.H. and T.R.A. wrote the original draft. All authors reviewed and edited the manuscript. B.-A.B., L.C.H., M.A.S.P. and J.J.C. obtained funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-65758-0>.

Correspondence and requests for materials should be addressed to Leonard C. Harrison.

Peer review information *Nature Communications* thanks Brandon Hickman and the other, anonymous, reviewer for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025

¹Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia. ²The University of Melbourne, Department of Medical Biology, Parkville, VIC, Australia. ³Royal Botanic Gardens, Melbourne, VIC, Australia. ⁴Department of Infectious Diseases, The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia. ⁵Robinson Research Institute, Adelaide Medical School, University of Adelaide, Adelaide, SA, Australia. ⁶Women's and Children's Hospital, Adelaide, SA, Australia. ⁷Discipline of Paediatrics and Child Health, School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, Australia. ⁸Institute of Endocrinology and Diabetes, Children's Hospital at Westmead, Sydney, NSW, Australia. ⁹Specialty of Child and Adolescent Health, Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia. ¹⁰Children's Diabetes Centre, Telethon Kids Institute, The University of Western Australia, Perth, WA, Australia. ¹¹Department of Endocrinology and Diabetes, Perth Children's Hospital, Perth, WA, Australia. ¹²Discipline of Paediatrics, Medical School, The University of Western Australia, Perth, WA, Australia. ¹³Department of Endocrinology and Diabetes, Queensland Children's Hospital, Brisbane, QLD, Australia. ¹⁴Children's Health Research Centre, Faculty of Medicine, The University of Queensland, Brisbane, QLD, Australia. ¹⁵Department of Chemical Pathology, Mater Pathology, Brisbane, QLD, Australia. ¹⁶Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia. ¹⁷Diabetes and Vascular Medicine Unit, Monash Health, Melbourne, VIC, Australia. ¹⁸Department of Diabetes and Endocrinology, Royal Melbourne Hospital, Melbourne, VIC, Australia. ¹⁹Faculty of Health, School of Medicine, Deakin University, Geelong, VIC, Australia. ²⁰Child Health Research Unit, Barwon Health, Geelong, VIC, Australia. ²¹Victorian Infectious Diseases Service, Royal Melbourne Hospital, Melbourne, VIC, Australia. ²⁵These authors contributed equally: Leonard C. Harrison, Theo R. Allnutt.

✉ e-mail: harrison@wehi.edu.au

ENDIA Study Group Consortium

Leonard C. Harrison ^{1,2,25} ✉, Jennifer J. Couper^{5,6}, Maria E. Craig ^{7,8,9}, Elizabeth A. Davis ^{10,11,12}, Tony Huynh^{13,14,15}, Georgia Soldatos^{16,17}, John M. Wentworth ^{1,18}, Peter Vuillermin ^{19,20}, Megan A. S. Penno ⁶, Patricia Ashwood⁵, Simon C. Barry^{5,6}, Peter G. Colman¹⁸, Emma E. Hamilton-Williams²², William D. Rawlinson^{7,23}, Aveni Haynes¹⁰, Ki Wook Kim^{7,23}, Kelly J. McGorm⁵, Grant Morahan²⁴, Rebecca L. Thomson⁵ & Jason Tye-Din^{1,2}

²²Frazer Institute, The University of Queensland, Woolloongabba, Brisbane, QLD, Australia. ²³Virology Research Laboratory, Serology and Virology Division (SAViD), NSW Health Pathology, Prince of Wales Hospital, Sydney, NSW, Australia. ²⁴Harry Perkins Institute of Medical Research, The University of Western Australia, Perth, WA, Australia.