Food & Function

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d5fo01640a

Association of lactose intake and lactase persistence genotype with microbial taxa and function in healthy multi-ethnic U.S. adults

Yirui Tang,[a] Andrew Oliver,[b] Zeynep Alkan,[b] Ian Korf,[c] Liping Huang,[a,b] Mary E. Kable[a,b] and Danielle G. Lemay *[a,b]

Lactase persistence is a genetically inherited trait that enables continued lactose digestion into adulthood. Lactase non-persistence (LNP) individuals often experience incomplete lactose digestion, allowing undigested lactose to reach the colon, where it may shape microbial composition and function. We investigated the relationship between the lactase persistence (LP) genotype, lactose consumption, and the taxonomic and functional profiles of the fecal microbiome. Participants from the USDA Nutritional Phenotyping Study, a cross-sectional observational study designed to assess how dietary factors impact human health, whose fecal microbiome profile was measured using shotgun metagenomic sequencing (n = 330) were included in this analysis. Fecal SCFA levels were measured using GC-MS. Fecal microbiome taxonomy and gene abundance were quantified using shotgun metagenomic sequencing. Lactose consumption and yogurt intake were estimated based on Automated Self-Administered 24h Dietary Assessment Tool (ASA24®) dietary recalls or Food Frequency Questionnaire. The LP/LNP genotype was determined by a single nucleotide polymorphism (SNP ID: rs4988235). Several genera of lactic acid bacteria (*Veillonella*, *Lactobacillus*, *Lacticaseibacillus*, and *Lactococcus*) were differentially abundant between recent high-lactose consuming (>10.0 g lactose per day) and low-lactose consuming (<3.3 g lactose per day) individuals. Among the LNP participants who self-identified as Caucasian or Hispanic, high-lactose consumers (>10.0 g per day *via* 24-h recall) had significantly higher relative abundances of lactic acid bacteria and lactate-utilizing bacteria (*Lacticaseibacillus*, *Lactobacillus*, *Megamonas*, and *Veillonella*) than low-lactose consumers (<3.3 g per day). Independent of lactose intake, LNP participants had a higher abundance of fecal microbial β-galactosidase genes than LP participants. Among the LNP participants, those with high recent lactose consumption also showed a significant shift towards more fecal propionate. The abundance of the yogurt-associated microbe, *Streptococcus thermophilus*, was positively associated with yogurt intake independent of the genotype. Alternative milk consumption was significantly negatively associated with fecal SCFAs both in the full cohort and the Caucasian/Hispanic subset, regardless of the genotype. Our results suggest that functional and persistent host lactase enzymes may work to competitively exclude lactic acid bacteria, contributing to a smaller realized niche for lactic acid bacteria in LP individuals compared to LNP individuals. However, regardless of the host genotype, consumption of alternative milk may be associated with reduced production of health-promoting intestinal metabolites, such as SCFAs.

Received 5th April 2025,
Accepted 1st August 2025

DOI: 10.1039/d5fo01640a

rsc.li/food-function

1. Introduction

Lactose, a dominant carbohydrate in dairy products, can be digested in the human small intestine by lactase (E.C. 3.2.1.108). Lactase is a hydrolysis enzyme that can break down lactose into glucose and galactose; these monosaccharides are subsequently absorbed and metabolized. Lactase activity is highly expressed during infancy, and although it decreases after weaning in approximately 33% of the population, around 35% of individuals can produce lactase into adulthood.[1] This phenomenon, known as lactase persistence (LP), is a heritable phenotype, mainly conferred by one single nucleotide polymorphism (SNP ID: rs49882235) in people of European descent. In these populations, lactase activity is

[a]Department of Nutrition, University of California, Davis, Davis, California, USA. E-mail: Danielle.Lemay@usda.gov
[b]United States Department of Agriculture-Agricultural Research Service, Western Human Nutrition Research Center, Davis, California, USA
[c]Genome Center, University of California, Davis, Davis, California, USA

high, and lactose can be fully digested through adulthood. Conversely, populations with lactase non-persistence (LNP) have low lactase activity in the small intestine and may experience lactose intolerance and uncomfortable gastrointestinal symptoms upon consuming lactose-rich dairy products.[2,3]

Colonic microorganisms can also break down lactose when it is not completely digested in the small intestine. Bacteria, such as lactic acid bacteria, *Bifidobacterium* and Proteobacteria, synthesize β-galactosidase that hydrolyzes lactose into glucose and galactose. Microbes in the colon then metabolize these monosaccharides into lactate, short-chain fatty acids (SCFAs) (mainly acetate, propionate, and butyrate) and gases (H_2, CO_2, and CH_4) through fermentation and cross-feeding.[4–6] After consuming lactose, if the intestinal lactase level is low, more lactose may travel to the large intestine and become metabolized by the colonic bacteria.

In a previous study, we conducted a microbiome analysis using 16S amplicon sequencing of fecal samples from healthy U.S. adults in the USDA Nutritional Phenotyping Study and found a higher abundance of families *Lactobacillaceae* and *Lachnospiraceae* in samples from LNP individuals consuming more than 12 g d^{-1} of lactose than LP individuals.[7] The family *Lactobacillaceae* contains lactic acid bacteria that can ferment lactose into lactic acid, and the members of the family *Lachnospiraceae* can subsequently convert lactate into acetate. Therefore, we speculated that the greater production of acetate, especially in LNP individuals, depends in part on cross-feeding networks which may form between bacteria such as *Lactobacillaceae* and *Lachnospiraceae*. However, this previous study was limited by the low resolution of microbial taxa, and the absence of microbial gene and SCFA measurements.

We conducted the current study to address these limitations. First, we sequenced and analyzed 330 shotgun metagenomes from the participants in the cohort. We hypothesized that fecal samples from LNP individuals consuming higher amounts of lactose would contain more DNA from lactose-fermenting species and more microbial β-galactosidase genes compared to LP individuals. Additionally, we conducted an SCFA analysis of participant fecal samples to test the hypothesis that fecal samples from LNP participants consuming higher amounts of lactose contained more acetate.

Although previous clinical studies have investigated how lactose consumption interacts with the host genotype to influence gut microbial composition and function, particularly in controlled settings,[8–10] relatively few studies have done so in large, free-living multiethnic populations with high-resolution metagenomic and fecal SCFA data. Our current study aims to expand on the previous findings, with a specific goal of investigating how lactose consumption and LP genotype influence fecal microbiome composition and function in healthy multiethnic US adults.

Most studies that examined the association between lactose consumption and the fecal microbiome treated lactose as a single, isolated compound.[9–12] However, it is well-known that different dairy products contain different amounts of lactose and also confer different physiological effects due to various dairy matrix effects.[13] For example, the lactose in fluid milk would be digested at a different rate from a solid piece of fermented cheese or semi-solid yogurt.[14] Therefore, we additionally examined whether the intake of different types of dairy products was associated with microbiome taxa and function in a lactase persistence trait-specific manner.

2. Materials and methods

2.1 Participants

Healthy US adults, aged 18 to 65 years, male or female, with a body mass index (BMI) ranging from 18 to 45 kg m^{-2} living near Davis, California, USA, were recruited in the Nutritional Phenotyping Study conducted by the USDA, Agriculture Research Service at the Western Human Nutrition Research Center. This was a cross-sectional observational study in which fecal microbial metabolites, specifically SCFAs acetate, butyrate, and propionate, were assessed in relation to self-reported dietary intake, genotypes, physiological status, and anthropometric measurements.[15] Males and females were recruited to fill nine groups based on a combination of three age ranges (18.00–33.99, 34.00–49.99, and 50.00–65.00 years old) and three BMI categories (normal: 18.50–24.99, overweight: 25.00–29.99, and obese: 30.00–45.00 kg m^{-2}) to balance enrollment in each sex. Participants were excluded if they had high blood pressure (systolic blood pressure > 140 mm Hg or diastolic blood pressure > 90 mm Hg) when measured on-site during their first visit, or if they were currently taking daily medication for diagnosed chronic disease(s) including, but not limited to, diabetes mellitus, cardiovascular disease, cancer, gastrointestinal disorders, kidney disease, liver disease, bleeding disorders, asthma, autoimmune disorders, hypertension, or osteoporosis. Individuals who were pregnant or lactating, had a known allergy to egg, or had recently received antibiotic therapy in the past three months were also excluded from this study. Participants whose dietary records were considerably incomplete or whose extracted fecal DNA was low in concentration (<100 ng μL^{-1}) or in quality ($A_{260/280}$ < 1.78; $A_{260/230}$ < 1.72) were excluded from fecal metagenomic sequencing (Fig. 1). This study is registered on ClinicalTrial.gov (Identifier NCT02367287) and received ethical approval from the University of California Davis Institutional Review Board (IRB). All study procedures involving human participants were conducted in accordance with the ethical standards of the IRB. Written informed consent was obtained from each participant prior to study enrollment in compliance with the IRB guidelines.

2.2 Stool collection and processing

Participants were instructed to collect one stool sample in a Ziploc bag enclosed in a hard, plastic container with a lid, immediately place it in a cooler with cold packs and deliver it

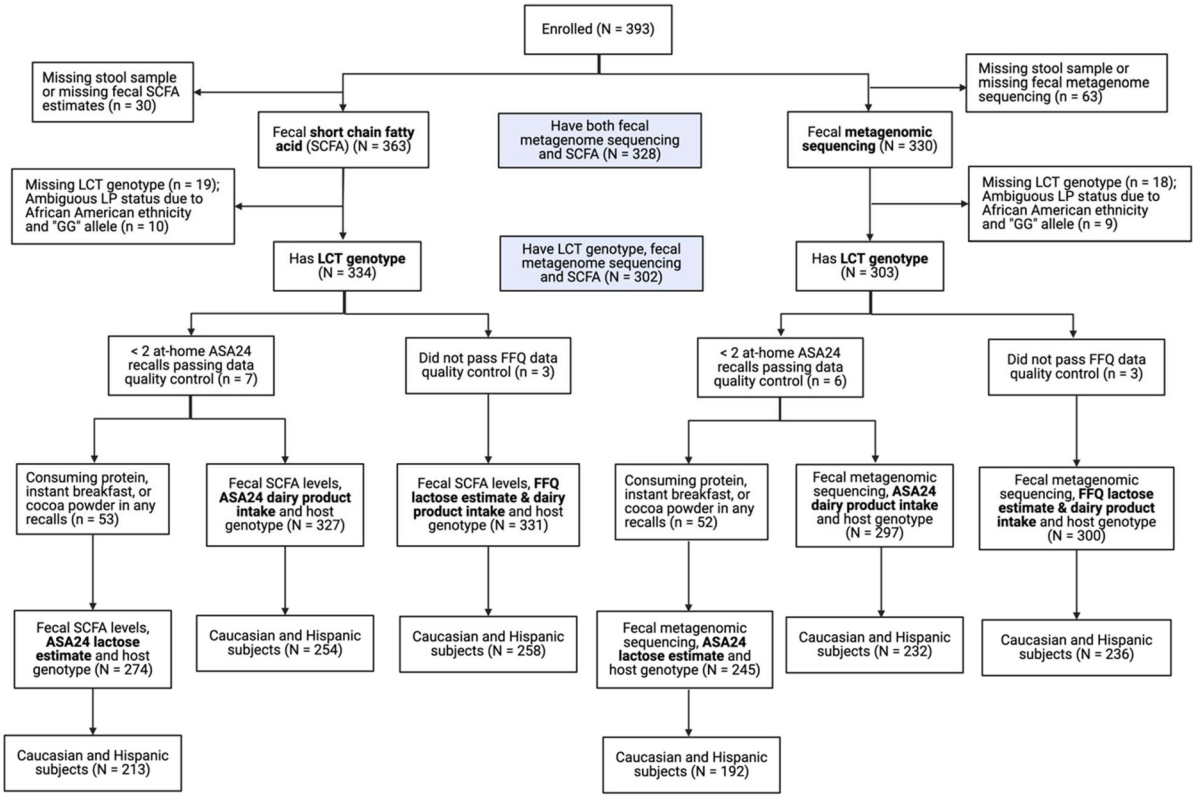

Fig. 1 STROBE diagram indicating inclusion of participant data.

to the USDA research center for same-day processing, as previously described.[16] Stool samples were homogenized in a Stomacher paddle blender for three minutes, flash-frozen on dry ice, broken into aliquots, and stored at −70 °C until fecal DNA isolation.

2.3 Quantification of fecal short-chain fatty acids

Fecal SCFA concentrations were quantified as previously published.[17] Briefly, 150 mg of stool was spiked with deuterated SCFA surrogates, and SCFAs were extracted into 200 μL 1 : 1 methanol : acetonitrile by bead beating with a Geno/Grinder 2010 homogenizer (Cole-Palmer) for 8 minutes at 1200 rpm. Filtered and clarified extracts were mixed with one volume of internal standard and run on a 7890 gas chromatograph interfaced with a 5977B mass selective detector (Agilent). The final concentrations were computed from standard curves and corrected by the percent recoveries of the deuterated surrogates in each sample.

2.4 Fecal DNA extraction, library preparation, and sequencing

ZymoBIOMICS™ DNA Miniprep Kit (Zymo Research) was used to isolate fecal DNA, as previously described.[17,18] DNA Technologies & Expression Analysis Core at the Genome Center at the University of California Davis performed the shotgun metagenome sequencing library preparation, library QC, qualification, and pooling.[17,18]

2.5 Dietary assessment

The Automated Self-Administered 24-hour (ASA24®) Dietary Assessment Tool, versions 2014 and 2016,[19] was used to assess recent dietary intake, as previously published.[20] Recent average intake of dietary components was computed using the mean across all 24-hour recalls that passed quality control.[21] The 2014 Block Food Frequency Questionnaire (FFQ) (NutritionQuest, Berkeley, CA) was used to estimate habitual dietary intake over the past 12 months. A registered dietitian performed a quality check manually.[22] Dairy intake variables (including servings of total yogurt and total cheese, servings of total dairy and total cow's milk, servings of chocolate milk and plant-based alternative milks, *etc.*) were estimated from the standard Block FFQ output, as previously described.[23] While the FFQ output includes a summary variable for habitual lactose intake, it does not distinguish between lactose-containing and lactose-free versions of dairy products. Likewise, ASA24® does not directly estimate lactose intake or differentiate lactose-free products. Therefore, we developed a method to quantify lactose intake based on ASA24® records by automated mapping to the Nutrition Data System for Research (NDSR),[24] which includes lactose content for each food item, and applied that method to the current cohort to estimate recent lactose intake for each individual.[7] Habitual alternative milk intake was calculated by summing soy, rice, and almond milk consumption (the food codes in the raw FFQ records

were 'group_soy_milk_total_grams', 'group_rice_milk_total_grams', and 'group_other_milk_almond_total_grams').

2.6 Blood collection

Fasting blood was collected by certified phlebotomists in the morning after a 12-hour overnight fasting (water was allowed to maintain hydration) following consumption of a standard meal the evening before. All blood specimens were then processed as described elsewhere.[15]

2.7 Genomic DNA purification and genotyping

Genomic DNA was extracted from whole blood.[7,25] Genotyping reactions were performed using a single nucleotide polymorphism (SNP) assay-based polymerase chain reaction (PCR), as described in previous publications.[7,25]

2.8 Metagenomic sequence analysis

The metagenome analysis pipeline is shown in Fig. S1. Human reads were removed using BMTagger 3.101[26] by aligning to the human genome version GRCh38.p13.[27] Then, Trimmomatic version 0.33[28] was used to remove adapters and trim paired-end reads with a sliding window of 4 bp, a minimum average quality of 15, and a minimum length of 99 bp as described in detail previously.[29] Afterward, FastUniq version 1.1[30] with default settings was used to remove duplicate reads. FLASH version 1.2.11[31] was used to merge paired-end reads with an overlapping read length range between 10 bp and 100 bp and a mismatch ratio of 0.1. MetaPhlAn version 4.0.6,[32,33] along with the vOct22 CHOCOPhlAnSGB 202212 database, was used for profiling the microbial composition at the species level. HUMAnN 3.6.1[34] was used to profile microbiome functional pathway abundance and gene abundance using the ChocoPhlAn v201901_v31 database.

MicrobeCensus was used to calculate the average genome size (AGS) of the metagenomes.[35] Custom protein reference databases of microbial β-galactosidase (EC 3.2.1.23) and phospho-β-galactosidase (EC 3.2.1.85) were constructed, respectively, using the most up-to-date experimentally verified genes from Uniprot (SI1 and SI2).[36,37] Microbial β-galactosidase gene abundance was estimated using the high-throughput protein alignment program, DIAMOND, and was normalized by AGS.[38]

2.9 Statistical methods

For lactose-consumption grouping, individuals were divided into three lactose-intake groups based on terciles of recent and habitual lactose intakes, respectively. For example, the low recent-lactose-intake group included individuals within the 33rd percentile or lower of the total recent lactose intake. The medium recent-lactose-intake individuals had recent lactose consumption between the 33rd and the 66th percentiles. Finally, the high recent-lactose-intake group contained individuals with recent lactose intake within the upper tercile, containing the highest recent lactose intake. This strategy was implemented to isolate individuals with low and high lactose intakes. Previous research in our study cohort has shown that microbial abundance was significantly different between the lactose intake tercile groups.[7]

Multivariable associations between the lactase genotype, lactose intake, covariates, and microbial metagenomic features (gene and pathway abundances) were determined using the comprehensive R package, MaAsLin2.[39] Differential abundance of microbiome taxonomy between groups was analyzed using the Wald and likelihood ratio tests in the DESeq2 package version 1.45.0[40] using the R version 4.3.2 programming language.[41] The models used to assess how the interaction effect of genotype and lactose intake influenced the relative abundance of fecal microbiome taxonomy are represented as follows, for recent and habitual intake, respectively:

$$\begin{aligned}\text{Full model} = \text{taxa} \sim\ & \text{lactose intake group} + \text{sex} + \text{age} \\ & + \text{BMI} + \text{fiber intake} + \text{LP status} \\ & + \text{lactose intake group} : \text{LP status}\end{aligned}$$

$$\begin{aligned}\text{Reduced model} = \text{taxa} \sim\ & \text{lactose intake group} + \text{sex} \\ & + \text{age} + \text{BMI} + \text{fiber intake} + \text{LP status}\end{aligned}$$

A Student's *t*-test was used to determine whether the LP status was significantly correlated with fecal SCFA concentration. An analysis of covariance (ANCOVA) test was used to determine the association of each fecal SCFA (acetate, butyrate, propionate, and total SCFAs) with lactase genotype, lactose intake, and multiple covariates (sex, BMI, age, and fiber intake). The Shapiro–Wilk normality test was used to ensure normal distribution of the residuals. An appropriate transformation was applied to the dependent variables, if needed. The bestNormalize R package was used to select the best transformation method.[42,43] A Benjamini–Hochberg adjustment was applied to the *P* values for multiple testing correction.[44]

3. Results

3.1. Participant characteristics

Participants with stool and blood samples selected for fecal metagenomic sequencing analysis and SNP-based PCR, respectively (n = 303), varied by age, sex, BMI and ethnicity, as shown in Table 1. Most participants identified as Caucasian (n = 196), while others identified as Hispanic (n = 41), Asian (n = 33), African American (n = 6, GA or AA allele), multi-ethnic (n = 17), and others (n = 10). African American participants with the GG allele were excluded from analysis due to their potential for having a lactase persistence phenotype *via* other SNPs.[45,46] Both sexes were distributed nearly equally, with 161 males and 173 females. The study cohort represents the multiethnic population in Davis, California, and surrounding areas.

Due to the strong confounding effect between ethnicity and LP genotype in our cohort, where nearly all Asian participants were homozygous for the LNP genotype (GG) and African American participants with the GG allele had been excluded, we conducted secondary analyses restricted to Caucasian and

Table 1 General characteristics of each lactase persistence group for all individuals[a]

Characteristics	Data from the lactase persistence genotype group		
	Lactase persistent (LP)	Lactase non-persistent (LNP)	*P*-Value
Avg age (years [range])	42.7 (19–66)	38.6 (18–65)	<0.05
BMI (kg m^{-2} [range])	27.02 (18.04–43.87)	27.11 (18.56–43.25)	0.80
No. of males	99	50	0.46
No. of females	95	59	0.46
Ethnicity (*n*)			<0.05
African American	6	0[b]	0.15
Asian	1	32	<0.05
Caucasian	164	32	<0.05
Hispanic	12	29	<0.05
Multi-ethnic	9	8	0.49
Other	2	8	<0.05

[a] n = 303. [b] African Americans with GG allele were excluded due to the ambiguous LNP status.

Hispanic participants (n = 236). Recent lactose consumption data from 163 participants and habitual lactose intake data from 203 participants in the Caucasian and Hispanic subcohort were categorized into low vs high intake groups, respectively (Table S1).

3.2 Recent lactose intake, LP genotype, and fecal microbial taxonomy

Since colonic microorganisms have the capacity of metabolizing residual lactose in the large intestine, we investigated whether the interaction between human lactase genotype and lactose consumption was associated with differences in fecal microbial taxonomy abundance. There were 245 participants whose recorded recent lactose intake passed quality control; these participants were divided into low lactose (n = 82) and high (n = 85) intake terciles. The low-intake group included individuals who recently consumed ≤ 3.2 g per day lactose, whereas the high-intake group had recently consumed ≥ 10.0 g per day lactose (Table S1). The relative abundances of several genera containing lactic acid bacteria—*Veillonella*, *Lactobacillus*, *Lacticaseibacillus*, and *Lactococcus*—were significantly higher (p-adj < 0.05) in the recent lactose high-intake group with LNP genotype than the other groups as determined by the likelihood ratio test with sex, BMI, age, and recent fiber intake as covariates (Fig. 2A; Table S2). LNP individuals with higher recent lactose intake also had significantly lower fecal abundance of *Erysipelatoclostridium*, *Anaerotignum*, and *Holdemanella* genera (Fig. 2B) (p-adj < 0.05).

After correcting for multiple comparisons, 90 genera were significantly associated with the interaction between the LP status and lactose consumption in Caucasian and Hispanic participants, when sex, BMI, age, and fiber consumption were included as covariates (SI3). Among Caucasian and Hispanic participants, lactic acid bacteria and lactate-utilizing bacteria (*Lacticaseibacillus*, *Lactobacillus*, *Lactonifactor*, *Megamonas*, *Eggerthella*, and *Veillonella*) were significantly more abundant in LNP with higher recent lactose intake compared to LNP with lower intake (p-adj < 0.05) (Fig. 3A; Table S3). The relative abundances of genera *Paraprevotella*, *Phascolarctobacterium*, *Slackia*, *Erysipelatoclostridium*, and *Coprobacter* were significantly lower with high recent lactose consumption compared

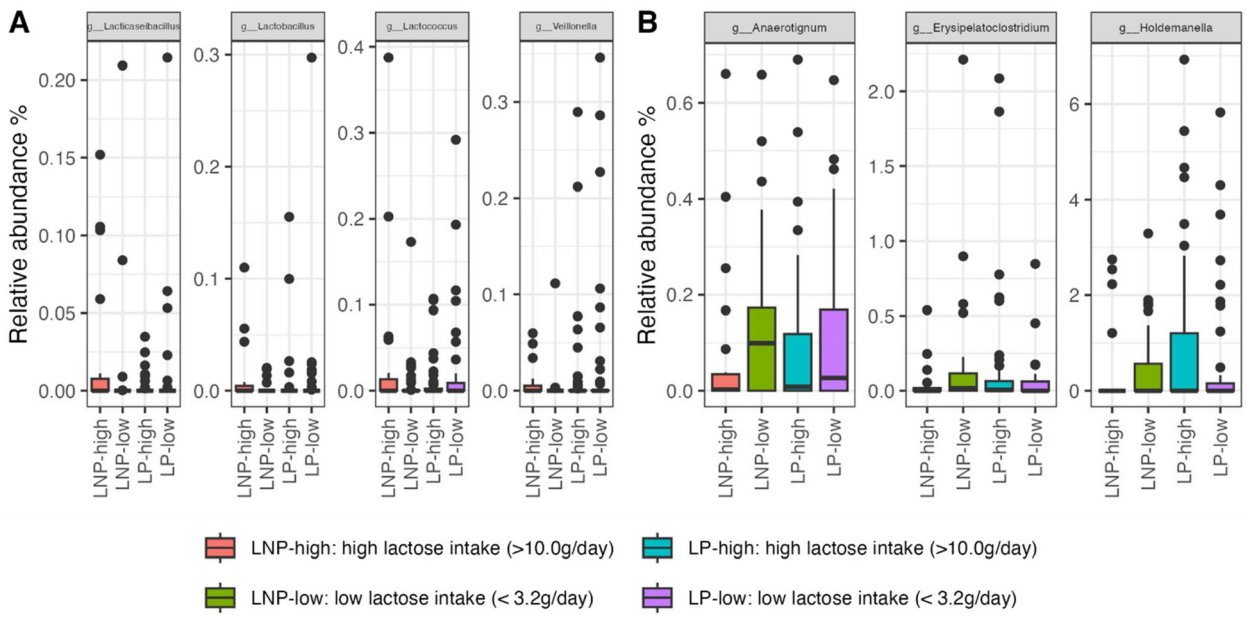

Fig. 2 (A) The relative abundances of genera *Veillonella*, *Lactobacillus*, *Lacticaseibacillus*, and *Lactococcus* were significantly elevated (p-adj < 0.05) in the recent lactose high-consumption group of lactase non-persistent (LNP) individuals compared to the low-consumption group as determined by the likelihood ratio test with sex, BMI, age, and recent fibre intake as covariates. (B) LNP individuals with higher recent lactose intake (as measured by ASA 24-hr recalls) had significantly lower fecal abundance of *Erysipelatoclostridium*, *Anaerotignum*, and *Holdemanella* genera (p-adj < 0.05). LP, lactase persistence; LNP, lactase non-persistent.

to low consumption, but only within the LNP group (p-adj < 0.05) (Fig. 3B).

3.3 Habitual lactose intake, LP genotype, and fecal microbial taxonomy

Habitual lactose consumption data from 302 participants, collected from an FFQ covering the previous 12 months, passed quality control and were categorized into low (n = 91) and high (n = 107) lactose intake tercile groups. The low-intake group included participants with habitual lactose intake ≤ 6.1 g per day, while the high-intake group consumed more than 13.0 g per day of lactose (Table S1). After correcting for multiple comparisons, 94 genera were significantly different by lactose genotype and intake group (SI4). *Lactococcus* and *Turicibacter* were significantly elevated in participants with high habitual lactose intake compared to the low-intake group, irrespective of genotype (p-adj < 0.05). LNP individuals with higher habitual lactose consumption also had decreased abundances of genera *Parasutterella*, *Lachnoclostridium*, *Escherichia*, and *Eggerthella* (p-adj < 0.05) as determined by the likelihood ratio test with sex, BMI, age, and habitual fiber intake as covariates (SI4).

Among Caucasian and Hispanic individuals, lactic acid bacteria and acetate producers (*Lacticaseibacillus*, *Anaerococcus*, and *Victivallis*) were significantly higher in lactase non-persisters with greater habitual lactose intake than those with lower habitual intake (p-adj < 0.05). The relative abundances of *Erysipelatoclostridium*, *Negativibacillus*, *Slackia*, and *Anaerofustis* genera were negatively associated with habitual lactose intake among LNP Caucasians and Hispanics (p-adj < 0.05) (SI5).

3.4 Lactose intake, LP genotype, and microbial genes

To determine the functional capacity of the gut microbiome for breaking down lactose to glucose and galactose, we quantified the abundance of microbial β-galactosidase genes (EC 3.2.1.23) in the shotgun metagenomes. We built a custom protein database of microbial β-galactosidase comprising 55 experimentally verified genes (SI1). Our analysis revealed a significant association between host lactase genotype and microbial β-galactosidase gene abundance (Fig. 4A). Specifically in all participants, LNP individuals had more microbial β-galactosidase genes than the LP individuals, independent of lactose intake, after adjusting for age, sex, BMI, and fiber intake (p-adj = 0.038) (Fig. 4A). However, this result did not extend to our subgroup analysis of Caucasian and Hispanic participants (p-adj > 0.05). The interaction of genotype and lactose consumption was not significant when predicting fecal microbial β-galactosidase gene abundance (p-adj > 0.05) (Fig. 4B).

Some microorganisms cannot hydrolytically cleave lactose directly and instead use a phospho-β-galactosidase to hydrolyze the phosphorylated derivative of lactose into glucose and galactose-6-phosphate (EC 3.2.1.85; MetaCyc ID = LACTOSECAT-PWY). We constructed another database for phospho-β-galactosidase genes and examined its association with lactose intake and host genotype (SI2). Neither lactase genotype nor lactose intake was found to be significantly associated with the abundance of phospho-β-galactosidase genes in fecal samples.

Microbial contributions to LACTOSECAT-PWY abundance showed variable abundance across the 330 shotgun metagen-

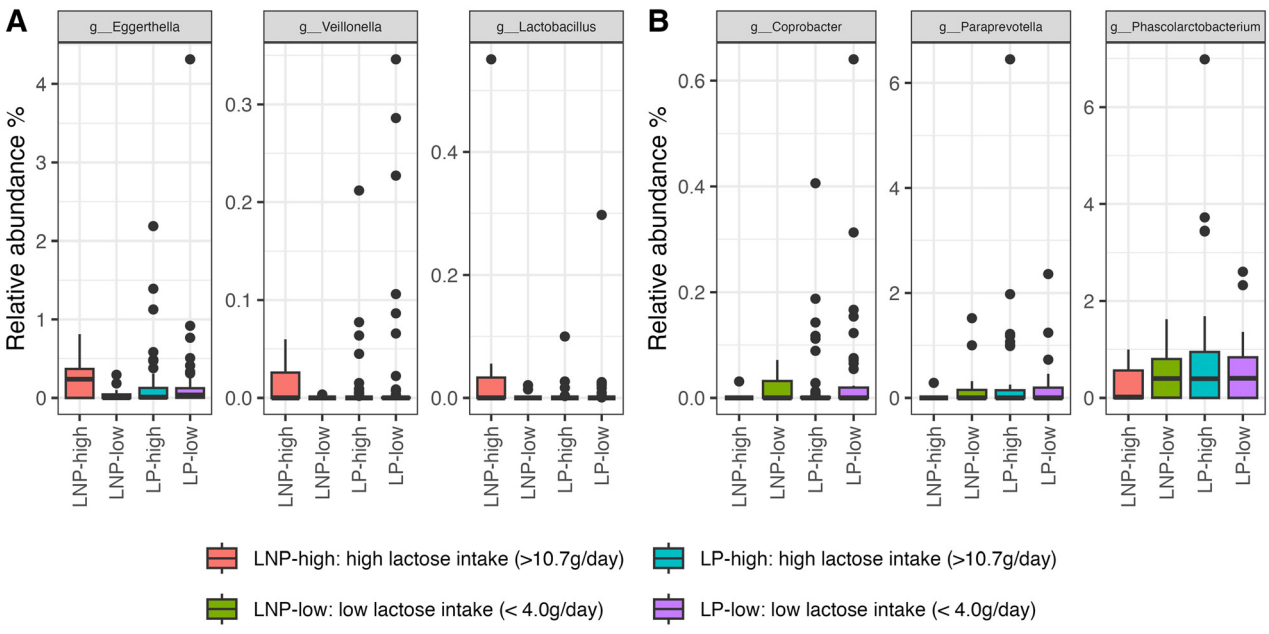

Fig. 3 Among the Caucasian and Hispanic subset of participant fecal microbiomes, (A) *Eggerthella*, *Veillonella*, and *Lactobacillus* genera had significantly higher abundance in lactase non-persistent (LNP) individuals with higher recent lactose intake than in LNP individuals with lower lactose intake (p-adj < 0.05). *Veillonella* uses lactate for growth; *Eggerthella* can ferment lactose. (B) The relative abundances of genera *Coprobacter*, *Paraprevotella*, and *Phascolarctobacterium* were significantly lower in the lactase non-persistent group with high recent lactose consumption compared to LNP with low consumption (p-adj < 0.05). LP, lactase persistent; LNP, lactase non-persistent.

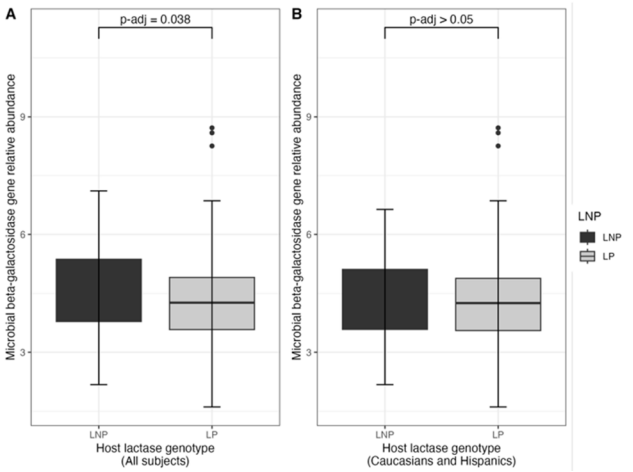

Fig. 4 (A) Lactase non-persisters (LNP) had significantly higher abundance fecal microbial β-galactosidase genes than the persisters (LP), after adjusting for age, sex, BMI, and fiber intake (p-adj = 0.038), n = 330. (B) The pattern was not significant among Caucasian and Hispanic subjects (p-adj > 0.05), n = 252. LP, lactase persistent; LNP, lactase non-persistent.

omes with major contributions from *Collinsella aerofaciens*, *Anaerostipes hadrus*, and species from *Lactococcus*, *Lactobacillus*, and *Streptococcus* (Fig. S2). Multivariate analysis using MaAsLin2 demonstrated no significant association of lactose intake or lactase genotype with the abundance of this pathway, when sex, BMI, age, and fiber intake were included in the multivariate models as covariates.

3.5 Association of the LP genotype with fecal SCFAs

Using data from all participants (n = 327), there was a significant effect of LP genotype on fecal propionate concentrations (ANCOVA, p-adj = 0.033 after Benjamini–Hochberg (BH) correction for multiple testing), when age, BMI, sex, and recent fiber consumption were included as covariates (Fig. 5A). Covariate-adjusted differences in fecal butyrate, acetate, and total SCFAs between genotypes were not significant after multiple testing correction (Fig. S3). Among Caucasian and Hispanic participants (n = 254), fecal butyrate, propionate, and total SCFA levels were significantly different between lactase genotypes, after adjusting for covariates and multiple comparisons (ANCOVA, p-adj = 0.039, 0.023, and 0.039, respectively) (Fig. 5B–D).

The ratio of acetate : propionate : butyrate is approximately 60 : 20 : 20 in the colon and stool of healthy individuals.[47] We calculated the deviation of each SCFA from its expected ratio (*e.g.* the relative to expected ratio of acetate $= \frac{\text{acetate concentration}}{\text{acetate + propionate + butyrate concentrations}} - \frac{60}{100}$. A positive value means a higher acetate ratio than expected) and studied the association between the relative change of fecal SCFAs with the host lactase genotype. Among Caucasians and Hispanics, LNP participants had a lower than expected ratio of fecal acetate and a higher than expected ratio of fecal propionate (ANCOVA, p-adj = 0.024 and 0.036, respectively) (Fig. S4), when adjusted for covariates: sex, BMI, age, and recent fiber intake. While not statistically significant, we noted this trend extended across all participants (ANCOVA, p-adj = 0.058 and 0.058, respectively).

3.6 Recent lactose intake, LP genotype, and fecal SCFAs

We used ANCOVA models to examine the association between fecal SCFA levels (acetate, propionate, butyrate, and total SCFAs), LP genotype, and recent lactose intake adjusted for sex, age, BMI, and recent fiber consumption in each model. LP individuals who did not consume much lactose (LP-low) had

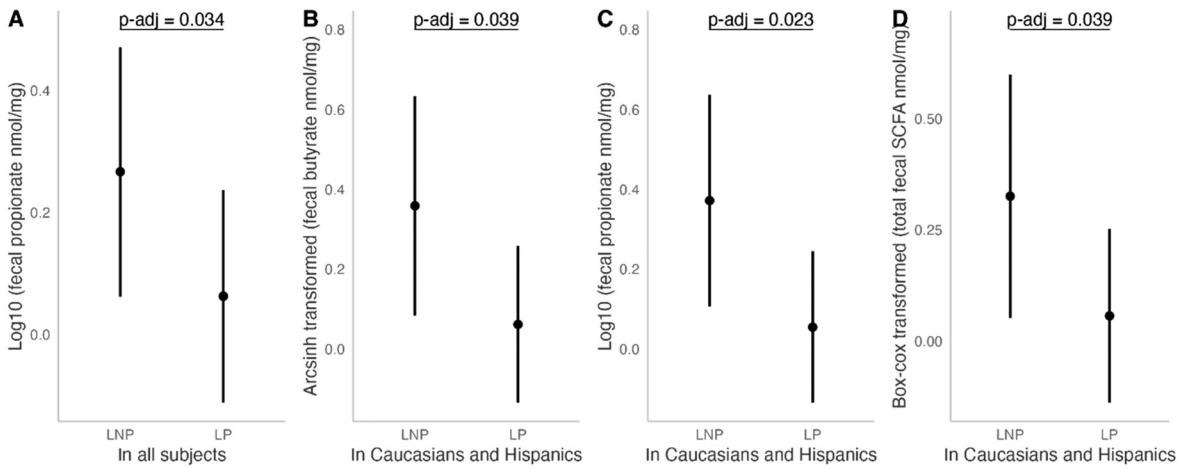

Fig. 5 Fecal short-chain fatty acid levels in lactase persistent (LP) and non-persistent (LNP) study subjects, including sex, BMI, age, and recent fiber consumption as covariates. LNP subjects had significantly elevated (A) fecal propionate (log transformed) than LP subjects (p-adj = 0.034, n = 353). Among Caucasian and Hispanic subjects (n = 274), LNP individuals had significantly elevated (B) fecal butyrate (arcsinh transformed), (C) propionate (log transformed), and (D) total SCFA (Box-Cox transformed) levels than LP individuals (p-adj = 0.039, 0.023, and 0.039, respectively). LP, lactase persistent; LNP, lactase non-persistent.

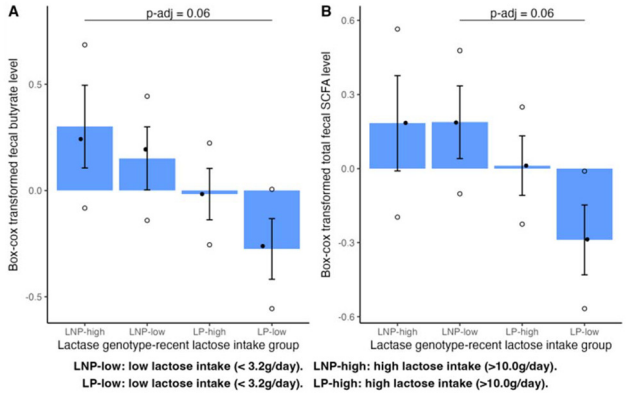

Fig. 6 (A) Decreased fecal butyrate in lactase persistent (LP) individuals who consume less lactose (<3.2 g) compared to lactase non-persistent (LNP) individuals who consume more recent lactose (>10 g). (B) Less total fecal SCFAs in LP individuals who consume less lactose than in LNP individuals who consume less recent lactose. (A and B) Means were adjusted for sex, age, BMI, and recent fiber consumption. LP, lactase persistent; LNP lactase, non-persistent.

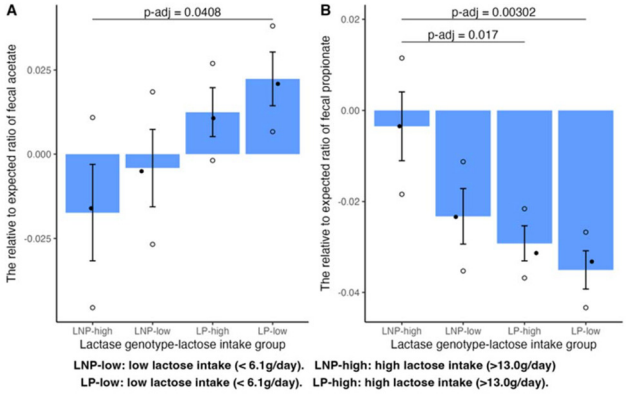

Fig. 7 Among Caucasian and Hispanic participants, LNP participants who habitually consumed more lactose had (A) lower acetate ratio (higher propionate + butyrate) than LP participants who consumed less lactose and (B) higher propionate ratio (lower acetate + butyrate) than LP participants (p-adj = 0.04 and 0.003, respectively). Means were adjusted for sex, age, BMI, and habitual fiber consumption. LP, lactase persistent; LNP lactase non-persistent.

less fecal butyrate than LNP individuals who consumed more lactose (LNP-high) (Fig. 6A) (ANCOVA, p-adj = 0.06); LP-low individuals also had lower total SCFA levels than lactase non-persistent (LNP-low) individuals who did not consume much lactose (Fig. 6B) (ANCOVA, p-adj = 0.06). In the subset of Caucasian and Hispanic participants, there were no significant differences in the fecal SCFA levels between LP genotypes by recent lactose intake (ANCOVA, p-adj > 0.10). Recent lactose intake and lactase genotype were not significantly associated with a shift toward any single fecal SCFA level (ANCOVA, p-adj > 0.10).

3.7 Habitual lactose intake, LP genotype, and fecal SCFAs

There were no differences in the absolute fecal SCFA levels between LP genotypes by habitual lactose intake, either in the full cohort or only among Caucasian and Hispanic participants. The relative to expected ratios of fecal acetate and propionate were significantly different between the habitual lactose intake group and the lactase genotype group among Caucasian and Hispanic participants (ANCOVA, p-adj = 0.045 and 0.012, respectively), when covariates (sex, BMI, age, and habitual fiber intake) were included. *Post-hoc* testing showed that LP individuals who habitually consumed less lactose (LP-low) had relatively more acetate than expected (Fig. 7A) and, consequently, less propionate than expected (Fig. 7B), compared to LNP individuals who habitually consumed more lactose (LNP-high) (ANCOVA, p-adj = 0.04 and 0.003, respectively). Also, in Caucasian and Hispanic participants, LP-high individuals with high habitual lactose intake had significantly less propionate than expected compared to LNP-high individuals (ANCOVA, p-adj = 0.017) (Fig. 7B).

3.8 Dairy product consumption and fecal SCFAs

In addition, we quantified the intake of multiple dairy products from ASA24 dietary recalls and the FFQ, including total dairy, cheese, milk, and yogurt. We then performed ANCOVA with multiple covariates to investigate the influence of the dairy matrix effect and lactase genotype on fecal microbial metabolism.

Independent of genotype, consumption of total dairy, milk, yogurt, or cheese, as measured *via* ASA24 was not significantly correlated with fecal SCFAs in either the full cohort or the Caucasian and Hispanic subset. When examining habitual intake of dairy products, as reported on the FFQ, participants who reported more cheese consumption over the past 12 months had lower than expected ratio of fecal butyrate (and consequently a higher ratio of acetate + propionate) (ANCOVA, p-adj = 0.019) (SI6); however, this relationship was not significant in the Caucasian and Hispanic subset. The other types of habitual dairy product intake (total dairy, milk, and yogurt) were not significantly associated with fecal SCFA either in the full cohort or in only Caucasian and Hispanic participants.

Considering the genotype, the effect of interactions between recent milk consumption and lactase persistent genotype on fecal acetate, butyrate, and total SCFA levels was significant in the full cohort (ANCOVA, p-adj = 0.033, 0.040, and 0.028, respectively) (SI6). However, these interactions were not significant in the Caucasian and Hispanic subsets. Among the Caucasian and Hispanic participants, we observed a significant interaction effect between recent yogurt intake and lactase genotype on the relative to expected ratio of fecal acetate (ANCOVA, p-adj = 0.049) (SI6). Two significant associations were identified in the full cohort: the interaction between habitual dairy intake and lactase genotype and the interaction between habitual cheese intake and lactase genotype were each significantly associated with the relative to expected ratio of butyrate (ANCOVA, p-adj = 0.044 and 0.044, respectively) (SI6). However, these relationships were not significant in the Caucasian and Hispanic subset. In short, there was no association between dairy product intake and absolute

SCFA abundance; this finding was robust across both the full cohort and Caucasian and Hispanic subsets.

3.9 Yogurt consumption, LP genotype, and fecal yogurt-associated microbe abundance

Lactobacillus delbrueckii subsp. *Bulgaricus* and *Streptococcus thermophilus* are considered the signature yogurt-associated microbes when investigating the survival of yogurt bacteria during gastrointestinal transit, so we further investigated the association between yogurt-associated microbes, lactase persistence genotype, and yogurt consumption in this study. Multivariate analysis revealed that in all participants, the relative abundance of *S. thermophilus* was positively associated with recent (Fig. S5A) and habitual (Fig. S5B) yogurt intake, respectively (coefficient = 0.91 and 1.23; p-adj = 1.75×10^{-3} and 1.78×10^{-6}, respectively). The correlation between *S. thermophilus* abundance and habitual yogurt intake was also positively significant in Caucasian and Hispanic participants (coefficient = 1.03; p-adj = 1.19×10^{-3}) (Fig. S5C). The relative abundance of *S. thermophilus* was positively associated with the relative ratio of the fecal propionate level in Caucasian and Hispanic subjects (coefficient = 0.76, p = 0.007), independent of the lactase genotype. Sex, age, BMI and fiber intake were included as covariates in the multivariate model using MaAsLin2. There was no association between *S. thermophilus* abundance and fecal SCFA levels in all subjects. No significant association was observed between yogurt consumption, either recent or habitual, and the LP genotype.

3.10 Association of alternative milk consumption with fecal SCFAs

In the full cohort, 342 and 349 participants provided recent and habitual alternative milk consumption information, respectively, which passed quality control for dietary intake assessment. Out of these participants, 63 participants (18.42%) were recent alternative milk consumers (intake > 0); 89 participants (25.50%) were habitual alternative milk consumers; and 48 of them were both recent and habitual consumers.

In the full cohort, fecal SCFA levels were significantly lower in recent consumers of alternative milk compared to non-consumers of alternative milk, independent of genotype (Fig. 8). This was true for fecal acetate, butyrate, propionate, and total SCFA levels (ANCOVA, p-adj = 0.0045, 0.0045, 0.0045 and 0.0043, respectively), in all participants, with sex, BMI, age, and recent fiber intake included as covariates. The same observation was true in a secondary analysis of only Caucasian and Hispanic participants, where recent consumers of alternative milk had significantly lower fecal acetate, butyrate, propionate, and total SCFA levels compared to non-consumers (ANCOVA, p-adj = 0.032, 0.032, 0.023, and 0.023, respectively). There were no differences in the fecal SCFA levels between habitual consumers of alternative milk and non-consumers in the full cohort, either independent of genotype or accounting for the lactase genotype. There were also no differences in a secondary analysis of only Caucasian and Hispanic participants. In summary, recent, but not habitual, consumers of alternative milk had lower SCFA levels compared to non-consumers in both the full cohort and Caucasian and Hispanic participants.

4. Discussion

This study explored the association between lactose intake, LP genotype, and gut microbial composition and function in a healthy multi-ethnic cohort of US adults. In a previous study, we estimated recent lactose intake from ASA24 dietary recalls,

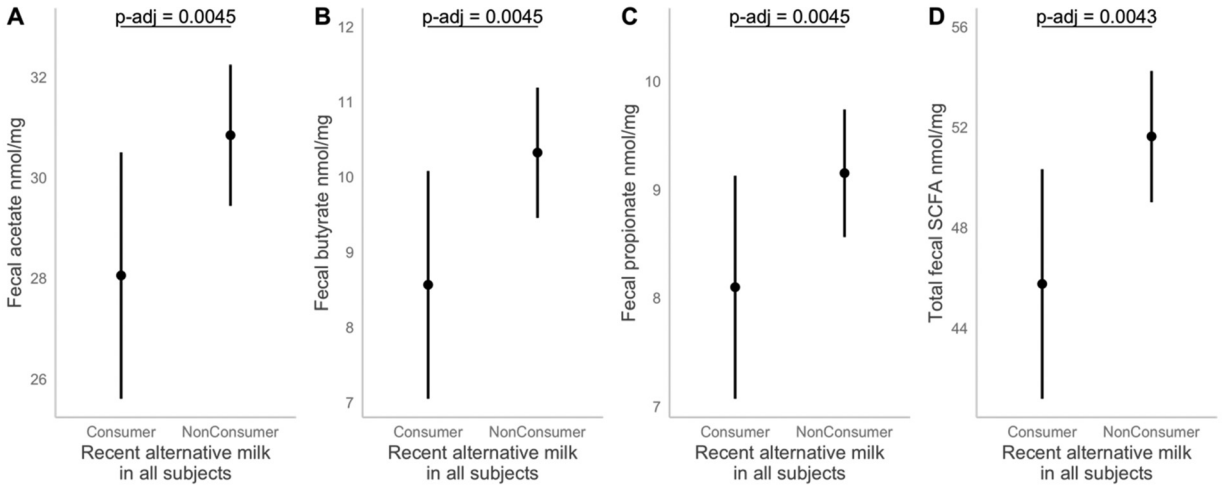

Fig. 8 Abundance of fecal short-chain fatty acid levels associated with recent alternative milk consumers *vs.* non-consumers, holding covariates constant. There were significant effects of recent alternative milk consumption on fecal (A) acetate, (B) butyrate, (C) propionate, and (D) total SCFA (p-adj = 0.0045, 0.0045, 0.0045 and 0.0043, respectively, using Benjamini–Hochberg correction), in all subjects, when sex, BMI, age, and recent fiber intake were included as covariates. These patterns were also statistically significant in the subset of Caucasian/Hispanic participants only (data not shown).

habitual lactose intake from an FFQ, LP genotype based on the SNP rs49882235, and fecal microbial abundance by 16S rRNA sequencing.[7] In the current study, we extended these analyses to include 330 shotgun metagenomes and the measurement of fecal SCFAs to better characterize the interaction of lactose intake and host genotype on microbial taxonomy and function. Several key findings emerged from our analysis.

First, we found that among LNP participants with higher lactose intake, fecal samples contained a significantly increased relative abundance of lactic acid bacteria and lactate-utilizing bacteria compared to those with lower intake. This pattern among all participants was also consistent using the Caucasian and Hispanic subgroup, suggesting a robust association between lactose consumption and specific microbial taxa. The difference in fecal microbe levels by genotype coincided with a trend (p-adj = 0.06) of elevated fecal butyrate and total SCFA levels in LNP individuals with higher recent lactose intake. The elevated abundance of lactic acid bacteria and lactate-utilizing bacteria, such as *Veillonella*, likely result from the fermentation of undigested lactose in the colon. Lactic acid bacteria, such as *Lactobacillus*, *Lactococcus*, and *Lacticaseibacillus*, ferment lactose into lactate and glucose and excrete lactate extracellularly. Cross-feeding bacteria can then utilize lactate, thereby increasing the levels of butyrate and other SCFAs in the feces.[5,11,48]

Previous studies have shown that lactose treatment suppressed *Bacteroidaceae* abundance and upregulated the relative abundances of lactic acid bacteria, *Bifidobacterium*, and *Veillonellaceae* (a lactate-utilizer) in *in vitro* fermentations.[11,49] These taxonomic changes also coincided with elevated fecal SCFA levels, specifically acetate and lactate. In studies looking at infant and toddler gut microbiomes, lactose significantly increased SCFA production and was bifidogenic in *in vitro* fermentations.[50] In a 12-week intervention study among 22 healthy Asian-origin adults with the LNP genotype, a progressively increasing lactose dosage led to significantly elevated *Bifidobacterium* relative abundance and higher fecal β-galactosidase activity, although fecal SCFA levels were not reported.[9] That we did not find *Bifidobacterium* to be elevated illustrates the likelihood that other lactic acid bacteria and lactate-utilizers can serve the same functional niche.

Interestingly, LNP individuals with higher recent lactose intake also showed a decrease in the relative abundance of several opportunistic pathogens, such as *Erysipelatoclostridium* and *Clostridium symbiosum.*[51,52] This finding was consistent among Caucasian and Hispanic participants, suggesting that the decrease in opportunistic pathogens is not merely due to ethnicity. These results suggest a potential protective effect of higher lactose consumption on the gut microbiome in LNP individuals.[55]

By examining the functional capacity of the gut microbiome to break down lactose, we found that individuals with lactase non-persistence exhibited higher abundance of fecal microbial β-galactosidase genes than persisters, regardless of lactose intake. This finding indicates a functionally stable microbial niche in LNP individuals, potentially reflecting a genotype-associated adaptation of the gut microbiota. Notably, this difference in microbial gene abundance between LP and LNP was not statistically significant in the Caucasian and Hispanic subset. It is therefore unclear whether the differences found with all participants are due to ethnicity.

Our study found no significant association between lactase genotype, lactose intake, or their interaction with the abundance of microbial phospho-β-galactosidase genes or of the abundance of the overall lactose degradation pathway *via* phospho-β-galactosidase (LACTOSECAT-PWY). These results are consistent with prior observations that the intracellular β-galactosidase is a key enzyme in microbial lactose degradation.[48]

The observation that LNP individuals exhibited higher fecal β-galactosidase gene abundance, even independent of lactose intake, points to a plausible microbial functional adaptation. In the absence of host lactase activity, undigested lactose chronically reaches the colon, even at small or infrequent amounts, potentially supporting a stable niche for lactose-degrading microbes. Additionally, microbial β-galactosidases may have overlapping substrate specificity with host endogenous glycans, such as mucin-derived galactose-containing oligosaccharides.[53] Over time, this selective pressure could maintain, or enrich, the microorganisms capable of expressing β-galactosidase enzymes regardless of current lactose intake.

Consistent with this hypothesis, we also observed that LNP participants had elevated fecal SCFA concentrations compared to LP participants, independent of lactose or dairy intake. This association was marginally significant in the Caucasian and Hispanic subset analysis. It potentially reflects a lifelong microbial adaptation to the absence of host lactase. Mucins secreted into the gut lumen are a continuous source of glycan substrates, which may be cleaved by host or microbially produced[53] β-galactosidases, potentially supporting sustained microbial fermentation and SCFA production. The fact that there are more microbial β-galactosidases and more SCFAs produced by LNP individuals raises the question of whether people with this genotype are more efficient degraders of mucin. In addition, most people in the U.S. are exposed to at least some dietary lactose, which may be sufficient to promote microbial cross-feeding and SCFA generation in LNP individuals.[5] Together, these findings suggest that LNP individuals may harbor a gut microbiome primed to compensate for the host enzymatic deficiencies.

Interestingly, greater cheese consumption over the past 12 months was associated with a lower-than-expected ratio of fecal butyrate and consequently a higher ratio of acetate + propionate, suggesting that cheese consumption may influence SCFA proportions. A previous study with the same cohort using machine learning and food-level data showed consumption of processed cheese and cheese spread as the strongest predictive food taxon of fecal SCFA abundances.[17] A recent cross-sectional study of mucosal-associated microbes and dairy intake demonstrated that participants who consumed more cheese in the past year had lower relative abundance of *Bacteroides* and *Subdoligranulum.*[54] Although SCFAs were not

analyzed in that study, the shifts in microbial taxa could be expected to alter ratios of SCFA production.

The fact that the association of cheese consumption with altered SCFA profiles was independent of genotype would support the idea that the cheese matrix shuttles lactose past the small intestine, where it would typically be degraded by the host enzymes. In this scenario, lactose embedded within the cheese matrix could reach the large intestine and act as a prebiotic, fostering microbial fermentation and SCFA production. However, the relatively low lactose content in processed cheese (approximately 1–5 g per 100 g) raises questions about whether this amount is sufficient to exert significant prebiotic effects.[55] Alternatively, the cheese matrix may protect and deliver intrinsic SCFAs (approximately 1.41% of total fatty acid could be butyric acid) directly to the large intestine, bypassing host digestion in the small intestine and influencing microbial metabolism.[56,57] This mechanism could be further complicated by the distinct matrix of processed cheese, which differs from natural cheese due to the inclusion of emulsifiers, salts, and high-temperature processing to ensure homogeneity.[58–60] These factors may alter the bioavailability of nutrients and microbial substrates within the cheese matrix.[61–64] Additionally, the presence of sodium propionate or calcium propionate in processed cheese, commonly added as mold inhibitors, could contribute to the observed shifts in fecal SCFA ratios.[65,66] These additives may directly introduce propionate into the gut, thereby influencing the SCFA ratio, or become metabolized by microorganisms.

The interaction between LP genotype and dairy intake revealed that genetic differences in lactose digestion significantly modulate the impact of milk and yogurt on SCFA profiles, with recent milk consumption affecting acetate, butyrate, and total SCFA levels in LP individuals. These genotype-dependent effects were particularly evident in the full cohort, though ethnic-specific differences emerged in the Caucasian and Hispanic subsets, where yogurt intake interacted with lactase genotype to influence acetate ratios. Collectively, these findings emphasize that the relationship between dairy consumption and gut microbial metabolism is not uniform but is instead shaped by the type of dairy product, its processing methods, and the genetic predisposition of the host. These findings advance our understanding of diet–microbiome interactions and highlight the need to consider both dietary composition and genetic factors when evaluating the impact of dairy on gut health. Future research should explore the mechanisms underlying these interactions, particularly the role of dairy components, processing methods, and host genetics in modulating microbial metabolic pathways.

Lactobacillus delbrueckii subsp. *bulgaricus* and *Streptococcus thermophilus* are two commonly used bacteria in yogurt fermentation.[67,68] However, results to date of the survival of yogurt bacteria during gastrointestinal transit remain conflicting.[69–73] Some researchers reported that both yogurt-associated bacteria were not detected in feces from daily yogurt consumption,[69] whereas other clinical studies demonstrated that the relative abundance of *S. thermophilus*, but not *L. delbrueckii*, significantly increased in fecal samples when administered in the form of yogurt intervention.[72] In our study, the relative abundance of *S. thermophilus* was significantly linked to both recent and habitual yogurt intakes, supporting our dietary intake estimates. Previous studies have demonstrated that individuals with higher yogurt consumption have an increased abundance of yogurt-fermenting bacteria.[67–73] One explanation for why *L. delbrueckii* was not significantly correlated with yogurt consumption might be the limited detection of this microbe in our cohort: only 11 participants in the cohort had detectable levels of *L. delbrueckii*. The fact that the yogurt-associated microbe, *S. thermophilus*, was not associated with SCFA levels suggests that recent yogurt consumption may not influence SCFA production.

An unexpected finding was that participants who had consumed alternative milk recently, as reported in the 24-h recalls, had significantly lower fecal SCFA concentrations compared to those who had consumed dairy milk. This result was robust in that it was true for both the full cohort and the Caucasian and Hispanic participant subset. It is possible that alternative milk consumption may influence SCFA production in the gut differently from regular milk intake due to the distinct nutrient differences between these products such as lactose, oligosaccharides, milk fat globule membrane, lactoferrin, minerals, and vitamins.[74–79] While dairy milk does not contain fermentable dietary fiber, except for perhaps the role of lactose as a fermentation substrate in lactase non-persistent individuals, most alternative milk does not contain fiber either (<1 g per serving) as the solids are filtered, removing fibrous parts of plants when producing the liquid. Oat milk contains the most fiber (still low at 2 g fiber per serving); however, this product also contains free sugars (~7 g per serving) that are enzymatically liberated from the oats during processing. At the time of study enrollment (2015–2019), oat milk was not yet a common product in local stores and no participants reported consumption. As there may be multiple reasons to consume plant-based milks,[80] a randomized-controlled trial should be pursued to determine whether there are differences in SCFAs between dairy and non-dairy milk consumption.

That habitual consumption did not show the same statistical significance as recent consumption could be due to the fact that the timing of the fecal sample collection is more closely matched to the assessment of recent diet. However, the discordant results between habitual and recent intake of alternative milk were likely due to the differences in the two dietary intake instruments (ASA24 for recent, FFQ for habitual) as the FFQ was not designed to probe consumption of alternative milk.

While our study utilized ASA24 and FFQ to capture recent and habitual dietary intake, respectively, it is important to acknowledge that the gut microbiome composition and functions are likely more responsive to recent dietary exposures. The FFQ was particularly useful for evaluating long-term trends, such as habitual dairy intake, and for examining intake patterns of foods with less frequent consumption (*e.g.* cheese and yogurt in a US cohort). However, interpretation of micro-

biome associations with the habitual dietary data should be made cautiously, as temporal mismatches between long-term intake and fecal sampling may obscure or attenuate associations.

Our study has several limitations. The cross-sectional design limits causal inference, and the dietary intake was self-reported, which may introduce recall bias. Additionally, the abundance of microbial β-galactosidase genes was used as a proxy for functional potential and may not reflect enzymatic activity. The cohort's ethnic diversity is both a strength and a challenge, as it introduces variability that can complicate data interpretation. Future intervention trials with lactase non-persistent adults in the U.S. are much needed to improve dietary guidance for these individuals.

Our findings provide multiple important translational insights for nutrition guidance in individuals with lactase non-persistence. The observation that LNP participants with more lactose consumption exhibited higher abundances of SCFA-producing microbes and microbial β-galactosidase genes suggested that the gut microbiome may functionally adapt to compensate for low host lactase activity. This supports the idea that LNP individuals may tolerate small or gradual increments in lactose intake, especially through fermented or low-lactose dairy products, such as yogurt and cheeses, which are naturally lower in lactose and contain live cultures that may aid digestion. Additionally, the reduction in potentially pathogenic taxa in high-lactose-consuming LNP participants highlights a possible protective role of lactose fermentation by commensal microbes. These findings reinforce the emerging view that complete dairy could confer microbial and metabolic benefits. This points to the potential for genotype-informed dietary guidance that supports microbial function while minimizing intolerance symptoms. Future dietary recommendations could integrate individual genetic and microbial profiles to promote gut health in a personalized and inclusive manner.

5. Conclusions

An interaction between the host genotype and dietary intake of lactose was associated with changes in the gut microbiome. Among adults who reportedly consumed >10 g of recent lactose per day, those with the LNP genotype had higher amounts of lactic acid bacteria and lactate-utilizing bacteria in their fecal samples compared to those with the LP genotype. Regardless of the host genotype, consumption of alternative milk may be associated with reduced production of health-promoting intestinal metabolites, which should be further investigated given the increasing popularity of these beverages.

Conflicts of interest

This research was supported, in part, by the California Dairy Research Foundation. The authors report no other conflicts of interest.

Data availability

Dietary intake data and fecal SCFA data are deposited at **https://github.com/vicky291/CDRF_metagenome/tree/main/data**. Metagenomic reads for 330 individuals are deposited in the NCBI Sequence Read Archive under two accession numbers: SRP354271 and SRP497208. The code to reproduce our results can be found in the GitHub repository: **https://github.com/vicky291/CDRF_metagenome**.

Supplementary information is available. Supplementary tables with adjusted *p*-values from all pairwise comparisons of genus relative abundance among lactase genotype-lactose intake groups, as well as supplementary figures are available. See DOI: **https://doi.org/10.1039/d5fo01640a**.

Acknowledgements

This study was supported by the California Dairy Research Foundation and the United States Department of Agriculture, Agricultural Research Service, 2032–51530–026–000D, 2032–10700–002–00D, 2032–51000–004–000D and 2032–51000–005–000D. We thank Ellen Bonnel, Eduardo Cervantes, Dustin Burnett, Annie Kan, Yasmine Bouzid, and Joanne Arsenault for their assistance with subject recruitment and dietary data cleaning. We thank Catherine Kirschke and Yining (Elaine) Wang for their assistance with blood DNA purification and single-nucleotide polymorphism genotyping. We thank Charles Stephensen for access to USDA Nutritional Phenotyping Study data and Zeya Xue for preliminary analyses. We thank Kevin Comerford for valuable feedback.

References

1 C. J. E. Ingram, C. A. Mulcare, Y. Itan, M. G. Thomas and D. M. Swallow, Lactose digestion and the evolutionary genetics of lactase persistence, *Hum. Genet.*, 2009, **124**(6), 579–591.
2 D. M. Swallow, Genetics of Lactase Persistence and Lactose Intolerance, *Annu. Rev. Genet.*, 2003, **37**, 197–219.
3 L. Ségurel and C. Bon, On the Evolution of Lactase Persistence in Humans, *Annu. Rev. Genomics Hum. Genet.*, 2017, **18**, 297–319.
4 G. Gottschalk, *Bacterial Metabolism*, Springer-Verlag, New York, 2nd edn, 1986.
5 E. J. Culp and A. L. Goodman, Cross-feeding in the gut microbiome: Ecology and mechanisms, *Cell Host Microbe*, 2023, **31**(4), 485–499, DOI: **10.1016/j.chom.2023.03.016**.
6 D. Ríos-Covián, P. Ruas-Madiedo, A. Margolles, M. Gueimonde, C. G. De los Reyes-Gavilán and N. Salazar, Intestinal short chain fatty acids and their link with diet and human health, *Front. Microbiol.*, 2016, **7**, 1–9.
7 M. E. Kable, E. L. Chin, L. Huang, C. B. Stephensen and D. G. Lemay, Association of Estimated Daily Lactose Consumption, Lactase Persistence Genotype (rs4988235),

and Gut Microbiota in Healthy Adults in the United States, *J. Nutr.*, 2023, **153**(8), 2163–2173.

8 A. Szilagyi, Adaptation to Lactose in Lactase Non Persistent People: Effects on Intolerance and the Relationship between Dairy Food Consumption and Evalution of Diseases, *Nutrients*, 2015, **7**(8), 6751–6779.

9 L. JanssenDuijghuijsen, E. Looijesteijn, M. van den Belt, B. Gerhard, M. Ziegler, R. Ariens, *et al.*, Changes in gut microbiota and lactose intolerance symptoms before and after daily lactose supplementation in individuals with the lactase nonpersistent genotype, *Am. J. Clin. Nutr.*, 2024, **119**(3), 702–710.

10 S. Hertzler and D. Savaiano, Colonic adaptation to daily lactose feeding in lactose maldigesters reduces lactose intolerance, *Am. J. Clin. Nutr.*, 1996, **64**(2), 232–236.

11 J. Firrman, L. S. Liu, K. Mahalak, W. Hu, K. Bittinger, A. Moustafa, *et al.*, An in vitro analysis of how lactose modifies the gut microbiota structure and function of adults in a donor-independent manner, *Front. Nutr.*, 2023, **9**, 1040744.

12 M. Ito and M. Kimura, Influence of lactose on faecal microflora in lactose maldigestors, *Microb. Ecol. Health Dis.*, 1993, **6**(2), 73–76.

13 T. K. Thorning, H. C. Bertram, J. P. Bonjour, L. De Groot, D. Dupont and E. Feeney, *et al.*, *Whole dairy matrix or single nutrients in assessment of health effects: Current evidence and knowledge gaps*, in *American Journal of Clinical Nutrition*, American Society for Nutrition, 2017, pp. 1033–1045.

14 D. A. Savaiano and M. D. Levitt, Milk Intolerance and Microbe-Containing Dairy Foods, *J. Dairy Sci.*, 1987, **70**(2), 397–406.

15 L. M. Baldiviez, N. L. Keim, K. D. Laugero, D. H. Hwang, L. Huang, L. R. Woodhouse, *et al.*, Design and implementation of a cross-sectional nutritional phenotyping study in healthy US adults, *BMC Nutr.*, 2017, **3**(1), 1–13.

16 D. G. Lemay, L. M. Baldiviez, E. L. Chin, S. S. Spearman, E. Cervantes, L. R. Woodhouse, *et al.*, Technician-Scored Stool Consistency Spans the Full Range of the Bristol Scale in a Healthy US Population and Differs by Diet and Chronic Stress Load, *J. Nutr.*, 2021, **151**(6), 1443–1452.

17 A. Oliver, Z. Alkan, C. B. Stephensen, J. W. Newman, M. E. Kable and D. G. Lemay, Diet, Microbiome, and Inflammation Predictors of Fecal and Plasma Short-Chain Fatty Acids in Humans, *J. Nutr.*, 2024, **154**(11), 3298–3311.

18 A. Oliver, Z. Xue, Y. T. Villanueva, B. Durbin-Johnson, Z. Alkan, D. H. Taft, *et al.*, Association of Diet and Antimicrobial Resistance in Healthy U.S. Adults, *mBio*, 2022, **13**(3), 1–18.

19 *Epidemiology and Genomics Research Program, Automated Self-Administered 24-Hour (ASA24®) Dietary Assessment Tool*, NIH.

20 D. G. Lemay, L. M. Baldiviez, E. L. Chin, S. S. Spearman, E. Cervantes, L. R. Woodhouse, *et al.*, Technician-Scored Stool Consistency Spans the Full Range of the Bristol Scale in a Healthy US Population and Differs by Diet and Chronic Stress Load, *J. Nutr.*, 2021, **151**(6), 1443–1452.

21 Y. Y. Bouzid, J. E. Arsenault, E. L. Bonnel, E. Cervantes, A. Kan, N. L. Keim, *et al.*, Effect of Manual Data Cleaning on Nutrient Intakes Using the Automated Self-Administered 24-Hour Dietary Assessment Tool (ASA24), *Curr. Dev. Nutr.*, 2021, **5**(3), 1–10.

22 E. L. Chin, M. Van Loan, S. S. Spearman, E. L. Bonnel, K. D. Laugero, C. B. Stephensen, *et al.*, Machine Learning Identifies Stool pH as a Predictor of Bone Mineral Density in Healthy Multiethnic US Adults, *J. Nutr.*, 2021, **151**(11), 3379–3390.

23 E. L. Chin, L. Huang, Y. Y. Bouzid, C. P. Kirschke, B. Durbin-Johnson, L. M. Baldiviez, *et al.*, Association of lactase persistence genotypes (Rs4988235) and ethnicity with dairy intake in a healthy u.s. population, *Nutrients*, 2019, **11**(8), 1–23.

24 E. L. Chin, G. Simmons, Y. Y. Bouzid, A. Kan, D. J. Burnett, I. Tagkopoulos and D. G. Lemay, Nutrient Estimation from 24-Hour Food Recalls Using Machine Learning and Database Mapping- A Case Study with Lactose, *Nutrients*, 2019, **11**(12), 3045.

25 E. L. Chin, L. Huang, Y. Y. Bouzid, C. P. Kirschke, B. Durbin-Johnson, L. M. Baldiviez, *et al.*, Association of lactase persistence genotypes (Rs4988235) and ethnicity with dairy intake in a healthy u.s. population, *Nutrients*, 2019, **11**(8), 1–23.

26 K. Rotmistrovsky and R. Agarwala, BMTagger: Best Match Tagger for removing human reads from metagenomics datasets. 2011.

27 V. A. Schneider, T. Graves-Lindsay, K. Howe, N. Bouk, H. C. Chen, P. A. Kitts, *et al.*, Evaluation of $GRCh_3$8 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly, *Genome Res.*, 2017, **27**(5), 849–864.

28 A. M. Bolger, M. Lohse and B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data, *Bioinformatics*, 2014, **30**(15), 2114–2120.

29 D. H. Taft, J. Liu, M. X. Maldonado-Gomez, S. Akre, M. N. Huda, S. M. Ahmad, *et al.*, Bifidobacterial Dominance of the Gut in Early Life and Acquisition of Antimicrobial Resistance, *mSphere*, 2018, **3**(5), 1–24.

30 H. Xu, X. Luo, J. Qian, X. Pang, J. Song, G. Qian, *et al.*, FastUniq: A Fast De Novo Duplicates Removal Tool for Paired Short Reads, *PLoS One*, 2012, **7**(12), 1–6.

31 T. Magoč and S. L. Salzberg, FLASH: Fast length adjustment of short reads to improve genome assemblies, *Bioinformatics*, 2011, **27**(21), 2957–2963.

32 A. Blanco-Miguez, F. Beghini, F. Cumbo, L. J. Mciver, K. N. Thompson, M. Zolfo, *et al.*, Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4.

33 D. T. Truong, A. Tett, E. Pasolli, C. Huttenhower and N. Segata, Microbial strain-level population structure & genetic diversity from metagenomes, *Genome Res.*, 2017, **27**(4), 626–638.

34 F. Beghini, L. J. McIver, A. Blanco-Míguez, L. Dubois, F. Asnicar, S. Maharjan, *et al.*, Integrating taxonomic, func-

tional, and strain-level profiling of diverse microbial communities with bioBakery 3, *Elife*, 2021, **10**, e65088, Available from: https://elifesciences.org/articles/65088.

35 S. Nayfach and K. S. Pollard, Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome, *Genome Biol.*, 2015, **16**(1), 51.

36 A. Bateman, M. J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, *et al.*, UniProt: the Universal Protein Knowledgebase in 2023, *Nucleic Acids Res.*, 2023, **51**(D1), D523–D531.

37 M. L. Treiber, D. H. Taft, I. Korf, D. A. Mills and D. G. Lemay, Pre-and post-sequencing recommendations for functional annotation of human fecal metagenomes, *BMC Bioinf.*, 2020, **21**(1), 1–15.

38 B. Buchfink, C. Xie and D. H. Huson, Fast and sensitive protein alignment using DIAMOND, *Nat. Methods*, 2015, **12**(1), 59–60.

39 H. Mallick, A. Rahnavard, L. J. McIver, S. Ma, Y. Zhang, L. H. Nguyen, *et al.*, Multivariable association discovery in population-scale meta-omics studies, *PLoS Comput. Biol.*, 2021, **17**(11), e1009442.

40 M. I. Love, W. Huber and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biol.*, 2014, **15**(12), 550.

41 R Core Team. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. Available from: https://www.R-project.org/.

42 R. A. Peterson and J. E. Cavanaugh, Ordered quantile normalization: a semiparametric transformation built for the cross-validation era, *J. Appl. Stat.*, 2020, **47**(13–15), 2312–2327.

43 R. A. Peterson, Finding Optimal Normalizing Transformations via bestNormalize, *The R Journal*, 2021, **13**(1), 310.

44 Y. Benjamini and Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B: Methodol.*, 1995, **57**(1), 289–300.

45 S. A. Tishkoff, F. A. Reed, A. Ranciaro, B. F. Voight, C. C. Babbitt, J. S. Silverman, *et al.*, Convergent adaptation of human lactase persistence in Africa and Europe, *Nat. Genet.*, 2007, **39**(1), 31–40.

46 M. C. Campbell and A. Ranciaro, Human adaptation, demography and cattle domestication: An overview of the complexity of lactase persistence in Africa, *Hum. Mol. Genet.*, 2021, **30**(R1), R98–109.

47 G. den Besten, K. van Eunen, A. K. Groen, K. Venema, D. J. Reijngoud and B. M. Bakker, The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism, *J. Lipid Res.*, 2013, **54**(9), 2325–2340.

48 C. F. Iskandar, C. Cailliez-Grimal, F. Borges and A. M. Revol-Junelles, Review of lactose and galactose metabolism in Lactic Acid Bacteria dedicated to expert genomic annotation, in *Trends in Food Science and Technology*, Elsevier Ltd, 2019, Vol. 88, pp. 121–132.

49 L. M. A. Jakobsen, U. K. Sundekilde, H. J. Andersen, D. S. Nielsen and H. C. Bertram, Lactose and Bovine Milk Oligosaccharides Synergistically Stimulate B. longum subsp. longum Growth in a Simplified Model of the Infant Gut Microbiome, *J. Proteome Res.*, 2019, **18**(8), 3086–3098.

50 P. Van den Abbeele, N. Sprenger, J. Ghyselinck, B. Marsaux, M. Marzorati and F. Rochat, A comparison of the in vitro effects of 2′fucosyllactose and lactose on the composition and activity of gut microbiota from infants and toddlers, *Nutrients*, 2021, **13**(3), 1–23.

51 M. N. Milosavljevic, M. Kostic, J. Milovanovic, R. Z. Zaric, M. Stojadinovic, S. M. Jankovic, *et al.*, Antimicrobial treatment of Erysipelatoclostridium ramosum invasive infections: a systematic review, *Rev. Inst. Med. Trop. Sao Paulo*, 2021, **63**, e30.

52 Y. M. Ren, Z. Y. Zhuang, Y. H. Xie, P. J. Yang, T. X. Xia, Y. L. Xie, *et al.*, BCAA-producing Clostridium symbiosum promotes colorectal tumorigenesis through the modulation of host cholesterol metabolism, *Cell Host Microbe*, 2024, **32**(9), 1519–1535.e7.

53 M. D. Berkhout, C. M. Plugge and C. Belzer, How microbial glycosyl hydrolase activity in the gut mucosa initiates microbial cross-feeding, *Glycobiology*, 2022, **32**(3), 182–200.

54 E. Chen, N. J. Ajami, D. L. White, Y. Liu, S. Gurwara, K. Hoffman, *et al.*, Dairy Consumption and the Colonic Mucosa-Associated Gut Microbiota in Humans-A Preliminary Investigation, *Nutrients*, 2025, **17**(3), 567.

55 L. A. Shakerdi, L. Wallace, G. Smyth, N. Madden, A. Clark, U. Hendroff, *et al.*, Determination of the lactose and galactose content of common foods: Relevance to galactosemia, *Food Sci. Nutr.*, 2022, **10**(11), 3789–3800.

56 S. M. Rafiq and B. C. Ghosh, Effect of Peanut Addition on the Fatty Acid Profile and Rheological Properties of Processed Cheese, *J. Food Process. Technol.*, 2017, **08**(08), 690.

57 M. G. Schmitt, K. H. Soergel and C. M. Wood, Absorption of short chain fatty acids from the human jejunum, *Gastroenterology*, 1976, **70**(2), 211–215.

58 P. L. H. McSweeney, G. Ottogalli and P. F. Fox, Diversity and Classification of Cheese Varieties: An Overview, in *Cheese*, Elsevier, 2017, pp. 781–808.

59 R. N. Salek, F. Buňka and M. Černíková, The use of different cheese sources in processed cheese, in *Processed Cheese Science and Technology*, Elsevier, 2022, pp. 79–113.

60 B. Riebel, S. Govindasamy-Lucey, J. J. Jaeggi and J. A. Lucey, Functionality of process cheese made from Cheddar cheese with various rennet levels and high-pressure processing treatments, *J. Dairy Sci.*, 2024, **107**(1), 74–90.

61 J. M. Aguilera, The food matrix: implications in processing, nutrition and health, *Crit. Rev. Food Sci. Nutr.*, 2019, **59**(22), 3612–3629.

62 A. I. Mulet-Cabero, M. Torres-Gonzalez, J. Geurts, A. Rosales, B. Farhang, C. Marmonier, *et al.*, The Dairy Matrix: Its Importance, Definition, and Current Application

in the Context of Nutrition and Health, *Nutrients*, 2024, **16**(17), 2908.

63 A. Asensio-Grau, I. Peinado, A. Heredia and A. Andrés, In vitro study of cheese digestion: Effect of type of cheese and intestinal conditions on macronutrients digestibility, *LWT–Food Sci. Technol.*, 2019, **113**, 108278.

64 K. Žolnere, M. Arnold, B. Hull and D. W. Everett, Cheese proteolysis and matrix disintegration during in vitro digestion, *Food Struct.*, 2019, **21**, 100114.

65 L. D. Kagliwal, S. B. Jadhav, R. S. Singhal and P. R. Kulkarni, PRESERVATIVES | Permitted Preservatives – Propionic Acid, in *Encyclopedia of Food Microbiology*, Elsevier, 2014, pp. 99–101.

66 M. Jelena and M. Zorana, Flavors, colors, and preservatives used in processed cheese, in *Processed Cheese Science and Technology*, Elsevier, 2022, pp. 125–147.

67 Y. H. Chang, C. H. Jeong, W. N. Cheng, Y. Choi, D. M. Shin, S. Lee, *et al.*, Quality characteristics of yogurts fermented with short-chain fatty acid-producing probiotics and their effects on mucin production and probiotic adhesion onto human colon epithelial cells, *J. Dairy Sci.*, 2021, **104**(7), 7415–7425, DOI: **10.3168/jds.2020-19820**.

68 S. Nagaoka, Yogurt Production, in *Lactic Acid Bacteria*, ed. M. Kanauchi, Humana Press, New York, NY, 2019. Available from: **https://link.springer.com/protocol/10.1007/978-1-4939-8907-2_5#publish-with-us**.

69 R. Del Campo, D. Bravo, R. Cantón, P. Ruiz-Garbajosa, R. García-Albiach, A. Montesi-Libois, *et al.*, Scarce evidence of yogurt lactic acid bacteria in human feces after daily yogurt consumption by healthy volunteers, *Appl. Environ. Microbiol.*, 2005, **71**(1), 547–549.

70 M. Elli, M. L. Callegari, S. Ferrari, E. Bessi, D. Cattivelli, S. Soldi, *et al.*, Survival of yogurt bacteria in the human gut, *Appl. Environ. Microbiol.*, 2006, **72**(7), 5113–5117.

71 D. D. G. Mater, L. Bretigny, O. Firmesse, M. J. Flores, A. Mogenet, J. L. Bresson, *et al.*, Streptococcus thermophilus and Lactobacillus delbrueckii subsp. bulgaricus survive gastrointestinal transit of healthy volunteers consuming yogurt, *FEMS Microbiol. Lett.*, 2005, **250**(2), 185–187.

72 C. Vizioli, R. Jaime-Lara, S. G. Daniel, A. Franks, A. F. Diallo, K. Bittinger, *et al.*, Administration of Bifidobacterium animalis subsp. lactis strain BB-12® in healthy children: characterization, functional composition, and metabolism of the gut microbiome, *Front. Microbiol.*, 2023, **14**, 1–16.

73 H. Aslam, F. Collier, J. A. Davis, T. P. Quinn, M. O'Hely, J. A. Pasco, *et al.*, Gut Microbiome Diversity and Composition Are Associated with Habitual Dairy Intakes: A Cross-Sectional Study in Men, *J. Nutr.*, 2021, **151**(11), 3400–3412.

74 Y. P. Huang, B. Paviani, N. K. Fukagawa, K. M. Phillips and D. Barile, Comprehensive oligosaccharide profiling of commercial almond milk, soy milk, and soy flour, *Food Chem.*, 2023, **409**, 135267.

75 O. Sadeghi, A. Milajerdi, S. D. Siadat, S. A. Keshavarz, A. R. Sima, H. Vahedi, *et al.*, Effects of soy milk consumption on gut microbiota, inflammatory markers, and disease severity in patients with ulcerative colitis: a study protocol for a randomized clinical trial, *Trials*, 2020, **21**(1), 565.

76 M. Dineva, M. P. Rayman and S. C. Bath, Iodine status of consumers of milk-alternative drinks *v.* cows' milk: data from the UK National Diet and Nutrition Survey, *Br. J. Nutr.*, 2021, **126**(1), 28–36.

77 S. C. Bath, S. Hill, H. G. Infante, S. Elghul, C. J. Nezianya and M. P. Rayman, Iodine concentration of milk-alternative drinks available in the UK in comparison with cows' milk, *Br. J. Nutr.*, 2017, **118**(7), 525–532.

78 E. Kosmerl, D. Rocha-Mendoza, J. Ortega-Anaya, R. Jiménez-Flores and I. García-Cano, Improving Human Health with Milk Fat Globule Membrane, Lactic Acid Bacteria, and Bifidobacteria, *Microorganisms*, 2021, **9**(2), 341.

79 M. Chichlowski, N. Bokulich, C. L. Harris, J. L. Wampler, F. Li, C. L. Berseth, *et al.*, Effect of Bovine Milk Fat Globule Membrane and Lactoferrin in Infant Formula on Gut Microbiome and Metabolome at 4 Months of Age, *Curr. Dev. Nutr.*, 2021, **5**(5), nzab027.

80 A. A. Paul, S. Kumar, V. Kumar and R. Sharma, Milk Analog: Plant based alternatives to conventional milk, production, potential and health concerns, *Crit. Rev. Food Sci. Nutr.*, 2020, **60**(18), 3005–3023.