Reports

BioTechniques

Prediction of high fecal propionate-to-butyrate ratios using 16S rRNA-based detection of bacterial groups with liquid array diagnostics

Pranvera Hiseni*[,1,2], Lars Snipen[2], Robert C Wilson[3], Kari Furu[1], Finn Terje Hegge[1] & Knut Rudi[2,3]

*[1]Genetic Analysis AS, Kabelgata 8, Oslo, 0580, Norway; [2]Department of Chemistry, Biotechnology & Food Sciences, Norwegian University of Life Sciences, PO Box 5003, Aas, 1432, Norway; [3]Department of Biotechnology, Inland Norway University of Applied Sciences, PO Box 400 Vestad, Elverum, 2418, Norway; *Author for correspondence: Tel.: +47 4864 4835; ph@genetic-analysis.com*

BioTechniques 74: 9–21 (January 2023) 10.2144/btn-2022-0045

First draft submitted: 4 April 2022; Accepted for publication: 5 December 2022; Published online: 5 January 2023

ABSTRACT

Butyrate and propionate represent two of three main short-chain fatty acids produced by the intestinal microbiota. In healthy populations, their levels are reportedly equimolar, whereas a deviation in their ratio has been observed in various diseased cohorts. Monitoring such a ratio represents a valuable metric; however, it remains a challenge to adopt short-chain fatty acid detection techniques in clinical settings because of the volatile nature of these acids. Here we aimed to estimate short-chain fatty acid information indirectly through a novel, simple quantitative PCR-compatible assay (liquid array diagnostics) targeting a limited number of microbiome 16S markers. Utilizing 15 liquid array diagnostics probes to target microbiome markers selected by a model that combines partial least squares and linear discriminant analysis, the classes (normal vs high propionate-to-butyrate ratio) separated at a threshold of 2.6 with a prediction accuracy of 96%.

METHOD SUMMARY

We present a quantitative PCR-compatible test based on the liquid array diagnostics method to be used as a tool for detecting/classifying fecal samples with an atypically high propionate-to-butyrate ratio. The liquid array diagnostics-based test presented here targets the 16S rRNA gene of a limited number of bacterial markers to infer their presence and abundance in fecal samples. The classification of samples (normal vs high propionate-to-butyrate ratio) is performed utilizing an algorithm combining partial least squares and linear discriminant analysis.

KEYWORDS:

butyrate ● gut microbiome ● LAD ● propionate ● qPCR ● SCFA

The human gut microbiome affects the health of the host through a variety of mechanisms, including the fermentation of nondigestible carbohydrates that escape small intestinal digestion and absorption [1]. The end products of this fermentation, short-chain fatty acids (SCFAs), serve a variety of functions, including acting as the main energy source for colonocytes [2], enhancing the intestinal epithelial barrier [3], regulating mucus production [4], modulating inflammatory responses [5], inducing apoptosis in colon cancer cells [6], regulating blood pressure [7], mediating gut–brain cross-talk [8], regulating glucose homeostasis [9] and lipid metabolism and adjusting satiety levels [10].

In healthy adult populations, it is estimated that the three major SCFAs (acetic, propionic and butyric acid) accumulate in a 3:1:1 molar ratio [11–13]. A deviation in such proportions, with a significant decrease in butyrate levels, has been observed in people consuming a diet high in protein and low in carbohydrates [14]. Butyrate production is solely dependent on the intake of nondigestible fiber, whereas the major propionate producers, such as Bacteroidetes, metabolize peptides as well, thus leaving propionate levels unaltered [15]. Lower butyrate levels have also been linked to a slower fecal transition time, and both are associated with a higher colonic pH, which in turn promotes the production of propionate [16]. A low pH environment protects against the overgrowth of pathogens [17]; thus, in this context, an increase in the propionate-to-butyrate (P:B) ratio may indicate a vulnerable gastrointestinal state.

A deviant ratio in favor of propionate was proposed to act as a diagnostic marker for irritable bowel syndrome (IBS) [18]. Increased levels of this acid (but not butyrate) were also reported in overweight and obese people [11], individuals with an increased risk of type 2 diabetes [19], patients with Alzheimer's disease [20] and those with nonalcoholic fatty liver disease [21]. Additionally, a reduced butyrate (but not propionate) concentration was observed in people with a high risk of stroke [22]. Although the evidence linking disproportionately low levels of butyrate and/or high levels of propionate with various diseases is expanding, routine diagnostic measurement for SCFA content remains challenging, mainly due to the high volatility of SCFAs and the complex sample clean-up procedure [23,24].

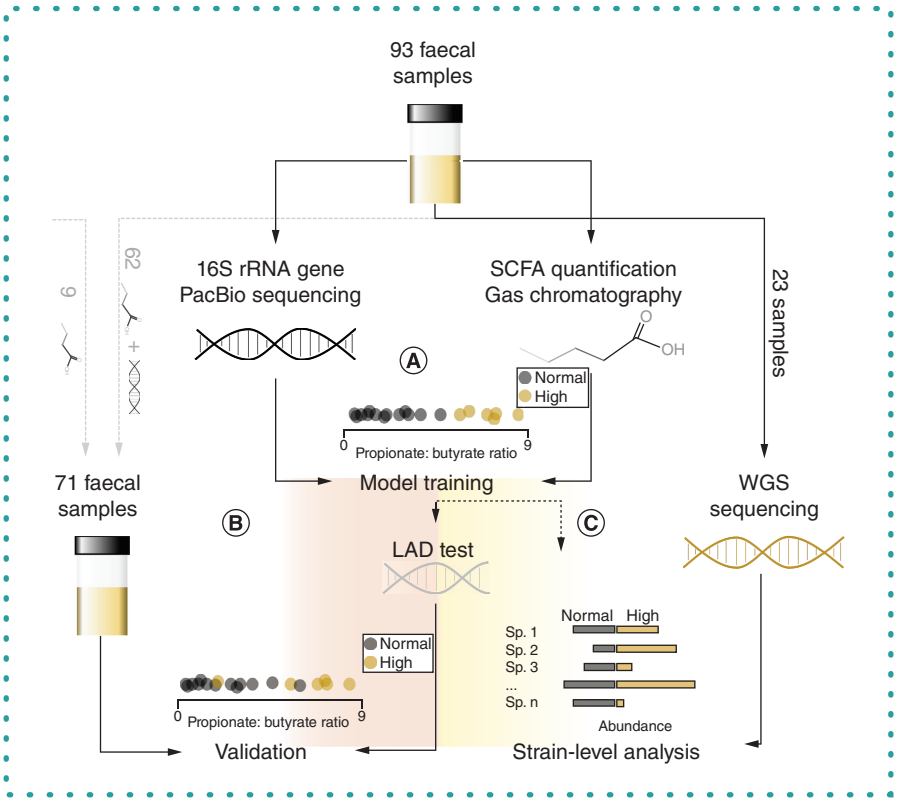

Figure 1. Building and validating a propionate-to-butyrate ratio prediction model. (A) Identification of taxonomic biomarkers for propionate-to-butyrate ratio. In this step, 93 fecal samples were analyzed for both their taxonomic composition (16S rRNA gene sequencing with PacBio single-molecule real-time technology) and SCFA content (GC). A model combining partial least squares and linear discriminant analysis was built, selecting a limited number of operational taxonomic units to act as predictors of normal versus high propionate-to-butyrate ratio. **(B)** Validation of the prediction model using an LAD-based test. In total, 71 fecal samples, nine of which were not PacBio-sequenced, were tested with a set of LAD probes designed to target operational taxonomic units selected by the combined partial least squares and linear discriminant analysis model in the previous step. **(C)** Functional and strain resolution associations with propionate-to-butyrate ratios.
LAD: Liquid array diagnostics; SCFA: Short-chain fatty acid; SP: Species; WGS: Whole-genome shotgun.

Here we aimed to infer SCFA levels by targeting a limited number of key bacteria using a novel quantitative PCR (qPCR)-compatible method, liquid array diagnostics (LAD) [25], circumventing the need to utilize GC-based methods. A LAD test targets variable regions within the 16S rRNA gene and allows the detection of up to 25 bacterial markers in a single tube. We focused on the P:B ratio, a single variable with the potential of providing an indication of functional dysbiosis in clinical samples. The analytical strategy followed in this study is outlined in Figure 1.

Materials & methods

Fecal samples & gDNA extraction

In total, 115 anonymized adult fecal samples, biobanked at Genetic Analysis AS, Oslo, Norway (research biobank no. 4071), were used for this study. Samples were collected and anonymized in accordance with the ruling by the Norwegian Regional Committee for Medical and Health Research Ethics (reference no. 2010/3209).

All fecal samples were stored at -40°C prior to gDNA extraction or GC sample prep. The gDNA of all fecal samples was extracted using a mag midi kit (LGC Biosearch Technologies, Hoddesdon, UK) following the steps suggested by the manufacturer. Genomic extracts were further analyzed with LAD, PacBio single-molecule real-time technology (PacBio, CA, USA) [26], or whole-genome shotgun sequencing (Illumina, CA, USA).

Measurement of SCFA content with GC

The SCFA content of 115 fecal samples was measured with GC (TRACE™ 1310 with autosampler; Thermo Fisher Scientific, MA, USA). Fecal samples were diluted in water (1:10) to a total volume of 1500 µl and then homogenized for 2 × 40 s at 1800 rpm using a Fastprep®-96 (MP Biomedicals, CA, USA). After a gentle spin, 300 µl of supernatant was transferred to a new tube, to which 300 µl of internal standard was added. The internal standard consisted of 0.4% formic acid and 2 mM 2-methylvaleric acid. The samples were centrifuged at 13,000 rpm for 10 min. Subsequently, 300 µl of supernatant was transferred to spin columns (0.2-µm filters) and centrifuged at

10,000 rpm for 5 min. The solution that passed the membrane was transferred to GC vials for SCFA measurement. An internal standard (1 mM 2-methylvaleric acid) was used as a reference for sample-to-sample normalization of results. A total of nine samples did not pass quality control by failing to produce a measurement on acetic acid. Given that this acid is the most volatile, its depletion was taken as an indication that the samples were compromised; therefore, they were excluded from further processing. In addition, a sample erroneously handled during laboratory work was removed. The P:B ratio of the 105 remaining samples was computed and used for further data analyses.

PacBio sequencing of 16S rRNA gene

Ninety-six samples randomly selected from 115 with SCFA content determined by GC were sent for PacBio Sequencing (Full-Length 16S Amplification SMRTbell® Library Preparation and Sequencing) at the Norwegian Sequencing Center (Oslo, Norway) (www.sequencing.uio.no). The first round of amplification was performed using the in-house 16S primers (GA-map® forward primer 5′-TCCTACGGGAGGCAGCAG-3′, GA-map® reverse primer 5′-CGGTTACCTTGTTACGACTT-3′, both protected by patent US20110104692 A1) tailed with universal sequences, as recommended in the PacBio protocol.

On average, there were approximately 28,480 sequencing reads per sample with an average length of 1175 nucleotides. The reads sharing at least 0.97 sequence identity were clustered into operational taxonomic units (OTUs) using the open-source metagenomics tool VSEARCH [27]. Excluding singletons, clustering resulted in 598 total OTUs, and the average number of OTUs per sample was 184. The OTU read counts were centered log ratio-transformed [28] (after the addition of one pseudo-read count) prior to further processing. One of the samples was unfortunately mislabeled when sent for sequencing, and two of the 96 samples did not meet the GC criteria (no measured acetate) and were thus excluded from the downstream analysis.

Identification of bacterial targets through combined partial least squares & linear discriminant analysis modeling

Centered log ratio-transformed OTU read counts from 93 samples were used as input for a combined partial least squares and linear discriminant analysis algorithm, with the aim of selecting variables (OTUs) to act as markers for classifying samples with normal versus high P:B ratios [29]. The aim was to correctly identify and classify the samples with the highest ratios, as they represent the deviation from the norm. The border between the two types of samples (normal vs high ratio) was allowed to go as low as possible without losing model prediction accuracy. The highest accuracy was reached at a P:B border of 2.5, with 37 OTUs acting as predictors, spanning 15 dimensions (leave-one-out cross-validated sensitivity = 90%, specificity = 99%, positive prediction rate = 90% and negative prediction rate = 99%). These OTUs were subsequently considered targets for LAD assay development.

Probe design for LAD

Eight-mer sequences containing a C at their 3′ ends – shared between only 16S *in silico* amplicons of target organisms – were computed using the in-house TNTProbeTool [30]. These sequences were considered the 3′ end segments of potential LAD labeling probes (LPs). Probes had to have a minimum melting temperature (*T*m) of 70°C (computed by the nearest neighbor method) hybridizing to the target group and a maximum *T*m of 30°C hybridizing to a nontarget group. The final LP sequences did not contain a C at their 3′ ends. In this way, the presence of the corresponding bacterial target would ensure that they became extended with a quencher-labeled ddCTP.

A reverse complementary reporter probe (RP) was designed for each of the LPs. The RPs were designed with a fluorophore tag on their 5′ ends, ensuring proximity to the quencher in duplexes harboring a 3′-labeled, RP-complementary LP. Duplexes containing the same fluorophore were designed with varying lengths to produce distinct temperature-dependent signals on the same qPCR channel of detection. The quenching effect of a longer and more stable duplex is observed as a dissociation curve with a higher *T*m. The DNA duplex *T*ms were calculated using the web-based OligoAnalyzer Tool™ 3.1 (Integrated DNA Technologies, Inc., IA, USA) based on the nearest neighbor method.

Basic Local Alignment Search Tool searches (nucleotide collection [nt/nr] database with Blastn) with each OTU sequence as query were performed to infer OTU taxonomy. Initially, 21 probes were designed, covering all 37 OTUs. However, six of the probes, targeting 11 OTUs (*Coprococcus eutactus*, *Alistipes indistinctus*, *Bacteroides eggerthii*, [*Clostridium*] *spiroforme*, *Ruthenibacterium lactatiformans*, [*Clostridium*] *glycyrrhizinilyticum*), failed to produce a signal; therefore, they were excluded from the assay.

Because of sequence similarity between F_3_1 and R_12_1 LPs, which were designed to detect *Dorea longicatena* and *Fusicatenibacter saccharivorans*, respectively, it was impossible to keep them in a single test tube, as this would risk producing double signals when only one target was present. Therefore the test was split into two tubes and the number of probes was divided proportionally between them (eight probes in group 1, seven probes in group 2). We used ROX_12_1 as an RP for both LPs. A list of final probes and their *T*ms and target species are presented in Table 1.

Generation of templates for LAD labeling reaction

Genomic DNA from 71 available samples was PCR-amplified. The SCFA content of these samples had been measured in previous steps; however, nine of the samples were not PacBio-sequenced. Each PCR reaction, with a total volume of 25 µl, consisted of 5 µl bacterial lysate (catalog no. 01-02-00500; Solis BioDyne, Tartu, Estonia), 3.75 U HOT FIREPol® DNA Polymerase (catalog no. 01-02-00500; Solis BioDyne), 1× B1 buffer (catalog no. 01-02-00500; Solis BioDyne), 2.5 mM $MgCl_2$ (catalog no. 01-02-00500; Solis BioDyne), 0.2 mM dNTPs

Vol. 74 | No. 1 | © 2023 The Authors

Table 1. Liquid array diagnostics test probes.

LP name	LP sequence, 5′–3′	RP name	RP sequence, 5′–3′[†]	*Tm* (°C)	Target species	Group
F_2_8	GCTACACACGTGCTACAATGGCGCATA	FAM_2_8	tTATGCGCCATTG	43.7	*Escherichia coli*	1
F_3_1	CGGGACTGCATTTGGAACTGCTGAG	ROX_12_1	tAGCCAGACAGTTTCCAATGCAGTCCCA	52.9	*Dorea longicatena*	1
H_6_13	GGTGGATGCTGGATGTGGGGAC	HEX_6_13	ttGTCCCCACATCC	45.5	*Bifidobacterium adolescentis*	2
R_9_2	CCGGGACTGCTTTGGAAACTATGCAG	ROX_9_2	ttCTGCATACTTTCCAAAGC	43	*Coprococcus comes*	2
R_10_3	GGAGCGTAGAAGGCATTGCAAGC	ROX_10_3	ttGCTTGCAATGCCTTC	52.8	*Blautia* sp. Marseille-P3313	2
R_12_1	TGGGACTGCATTGGAAACTGTCTGGCT	ROX_12_1	tAGCCAGACAGTTTCCAATGCAGTCCCA	69.3	*Fusicatenibacter saccharivorans*	2
C_14_5	CCCGTCACTCCATGAGAGTTGGAGATAC	CY5_14_5	ttGTATCTCCAACTCTC	45.5	Uncultured bacterium clone AP07S.190	1
C_15_4	CCGTACTGGCTCTGGAAACTGTTCAG	CY5_15_4	ttCTGAACAGTTTCCAGAGC	55.4	*Holdemanella biformis*	1
C_17_1	GGCCACACACGTACTACAATGGTGGTTAA	CY5_17_1	tTTAACCACCATTGTAGTACGTGTGTGG	64.6	*Flavonifractor plautii*, *Flintibacter* sp. KGMB00164	1
C.5_18_8	TGGAAGCCGGGAGTACCTGAAG	CY5.5_18_8	ttCTTCAGGTACcCCC	35.9	*Barnesiella* sp. strain mt172, *Barnesiella* sp. strain mt155	1
C.5_19_2	CGCGAGGGGGAGCAAAACTGGAAAA	CY5.5_19_2	tTTTTCCAGTTTTGC	44.8	Uncultured bacterium isolate DGGE gel band RB1-25	1
C.5_20_2	GCGGACTACTGGGCACCAA	CY5.5_20_2	tTTGGTGCCCAGTAGTC	55.2	*Faecalibacterium prausnitzii*; uncultured bacteria clones 2-002-f10, A3-213 and TS3_a01c08	2
C.5_21_1	GGAAGCGACTGGGCAACCAGAAG	CY5.5_21_1	ttCTTCTGGTTGCCCAGTCGCTTC	64.9	Uncultured organism clone ELU0116-T290-S-NI_000152	1
Fecali291	TTGCTCCACCTCGCGGTCTTGCTTCTCTTTGTTTAA	Fecali291 CY5	TTAAACAAAGAGAAGCAAGACCGCGAGGTGGAGCAA	72.2°C	*Faecalibacterium prausnitzii*; [*Eubacterium*] *siraeum* V10Sc8a; Ruminococcaceae bacterium strain MT139; uncultured bacteria clones PB1_aai26e05, C3-2 16S, A3-213, TS3_a01c08, SJTU_A2_03_71 and A5_016	2
Rum1167	CACTCTAGCCTGACAGTT	Rum1167 CY5	AACTGTCAGGCTAG	47.1°C	[*Ruminococcus*] gnavus; uncultured bacteria clones SJTU_G_10_25, Cadhufec15ml and CFT114H1	2

[†] Lowercase t nucleotides represent 5′-end T-tails. These were introduced with the purpose of securing physical distance between the fluorophore and Gs (within either the RP sequence itself or the G's in the 3′ end of the complementary LP sequence). Gs are known to have an intrinsic quenching property.
LP: Labeling probe; RP: Reporter probe; *Tm*: Melting temperature.

(catalog no. 18427088; Thermo Fisher Scientific) and 0.2 μM in-house primers (GA-map® forward primer 5′-TCCTACGGGAGGCAGCAG-3′, GA-map® reverse primer 5′-CGGTTACCTTGTTACGACTT-3′, both protected by patent US20110104692 A1). The amplification was carried out using an Applied Biosystems Veriti™ Thermal Cycler (Thermo Fisher Scientific) with an initiation period of 15 min at 95°C followed by 30 cycles of 30-s denaturation at 95°C, 30-s annealing at 55°C and 80-s elongation at 72°C, ending with a final step of elongation at 72°C for 7 min. PCR products were then treated with 2.7 U Exonuclease I (catalog no. M0293L; New England Biolabs, MA, USA) and 7.36 U recombinant shrimp alkaline phosphatase (catalog no. M0371L; New England Biolabs) and set for incubation at 37°C for 10 min, followed by 15 min at 80°C to inactivate the enzymes.

Single nucleotide extension of LPs & melting curve analysis with LAD

A total of 10 μl of PCR products treated with Exonuclease I–shrimp alkaline phosphatase (14.5–25.6 ng/μl) were used as templates for LP labeling. The labeling reaction also comprised LPs at a final concentration of 0.1 μM (biomers.net GmbH, Ulm, Germany), 1× buffer C (catalog no. 01-06-00500; Solis BioDyne), 1 mM $MgCl_2$ (catalog no. 01-06-00500; Solis BioDyne), 7.5 U HOT TERMIPol® DNA Polymerase (catalog no. 01-06-00500; Solis BioDyne) and 0.96 μM ddCTP-DYQ660 (catalog no. NU-850–660Q; Jena Bioscience, Jena, Germany). The reaction was performed in a PCR instrument with an initiation step at 95°C for 12 min followed by 40 cycles of denaturation (96°C for 20 s) and annealing/elongation (68°C for 40 s).

Following labeling, a mixture of RPs and $MgCl_2$ was added to the reactions to achieve final concentrations of 0.01 μM and 5 mM, respectively. Reagent S, available from Inland Norway University of Applied Sciences (Hamar, Norway), was also added to a final concentration of 1%. The melting curve analysis (31–85°C) was performed using a CFX96 qPCR instrument (Bio-Rad Laboratories, Inc., CA, USA).

The extraction of peaks and determination of positive signals were performed as described by Hiseni *et al.* [25] with a slight modification. Prior to extracting the signals, the fluorescence measurements within each channel were centered with the purpose of minimizing the range of measurements across wells at any given temperature. Next, the baseline within each channel was corrected (flattened) by subtracting the centered values of each sample from the average no template control centered values. As an ultimate step, for group 1 samples only, a further correction of FAM and CY5 baselines was performed by subtracting the values from one another (FAM = FAM - CY5 and CY5 = CY5 - FAM).

Bioinformatics evaluation of probe specificity

OTU sequences (PacBio) were used as subjects to check for sequences complementary to 3′ C-labeled probes. A search for the occurrence of probes, allowing one mismatch anywhere along the sequence was performed (excluding the probe 3′-C). The intention of this step was to prove that probes precisely targeted the intended bacteria. OTU sequences containing sites complementary to probe sequences were considered to act as 'labeling templates'. The read counts of all such sequences were considered *in silico* signals, which were then used to compute the correlation with real LAD signals.

Whole-genome shotgun sequencing

A total of 24 samples were sent for whole-genome shotgun sequencing at the Norwegian Sequencing Center. Libraries were prepared using a Nextera™ DNA Flex Library Preparation Kit (Illumina) following the protocol recommended by the manufacturer. Samples had different SCFA levels that spanned well the P:B values. One of the samples failed the GC quality check (no measured acetate) and was excluded from further analysis.

Processing of whole-genome shotgun sequencing results

DIAMOND software [31] was used to search for genes related to propionate and butyrate. Raw whole-genome shotgun sequencing reads were used as an input. For propionate, we searched for the genes *mmdA*, *lcdA* and *pduP* (markers for the succinate, acrylate and propanediol pathways, respectively [32]). For butyrate, the process involved searching for *but* and *buk* genes. For each read, only the hit with the highest bit score per pathway was kept (e-value $\leq$1e-05). For each sample, the reads that got a hit with one of the genes were counted and then grouped and summed according to the SCFAs to which they were related. After normalizing for the query sequence size and sequencing depth, the total number of hits related to propionate and butyrate was compared with the relative abundance of these acids. Taxonomic assignment of the sequencing reads was performed with a combination of Kraken2 [33], KrakenUniq [34] and Bracken [35] using HumGut_975 as a custom database, as described by Hiseni et al. [36].

Results & discussion

Identification of taxonomic biomarkers for P:B ratio

We examined the microbiome composition (PacBio sequencing of 16S rRNA gene) and SCFA content of 93 adult fecal samples. The aim was to identify potential associations between different members of the microbiome and levels of propionate and butyrate, and use this information to build a simple, predictive test based on LAD technology.

We computed the correlation between centered log ratio-transformed OTU read counts and relative abundance of propionate and butyrate. Only OTUs with $>$0.2 or $<$-0.2 correlations ($p < 0.05$) were considered. A total of 65 OTUs correlated with propionate levels (39

positively, 26 negatively), and 62 correlated with butyrate levels (28 positively, 34 negatively). Of these, 11 correlated with both butyrate and propionate, albeit in opposite directions. A simplified network of SCFA/OTU relationships is presented in Supplementary Figure 1.

We performed a Basic Local Alignment Search Tool search for highly similar sequences (nucleotide collection [nt/nr] database) using OTU sequences as queries. Among the OTUs positively correlated with butyrate, we found some that shared high sequence identity (>98.5%) with typical butyrate producers, including *Faecalibacterium prausnitzii* [37,38] (correlation = 0.21; $p < 0.05$), *Agathobaculum butyriciproducens* [39] (correlation = 0.23; $p < 0.05$) and *Coprococcus catus* [40] (correlation = 0.21; $p < 0.05$). However, we also found a positive relationship between butyrate and the read counts of sequences sharing high identity (>99%) with *Lactobacillus acidophilus* (correlation = 0.22; $p < 0.05$), *Fusicatenibacter saccharivorans* (correlation = 0.33; $p < 0.005$) and *Blautia wexlerae* (correlation = 0.26; $p < 0.05$) – species not known to produce this acid [41–43]. Furthermore, *Dysosmobacter welbionis* (correlation = -0.25; $p < 0.05$) and *Flavonifractor plautii* (correlation = -0.3; $p < 0.005$) – both butyrate producers [44,45] – exhibited a negative correlation with the relative abundance of butyrate. Similarly, propionate levels did not exclusively correlate with well-described propionate producers.

In light of this complex outcome, we decided to build a model based on a binary classification system (i.e., classifying samples as having a high or normal acid level). Aiming for a simple method, we chose to detect and classify samples based on a single variable that inferred information about both acid concentrations: the P:B ratio. Classification of samples based on this ratio makes biological sense, as the molar ratio between propionate and butyrate in healthy adults is nearly 1.0 [11–13]. Given the role of butyrate in maintaining human health [2,6,46–48], our goal was to detect samples with depleted butyrate levels, inferred by a deviant P:B ratio in favor of propionate (i.e., P:B ratio >>1.0).

We computed the P:B ratio from GC data for all samples. We then built a model combining partial least squares and linear discriminant analysis (PLS + LDA) using OTU read counts as predictors and aimed to find the ratio that best separated the two groups (normal vs high ratio) while selecting a reasonably small number of OTUs to act as markers. These marker OTUs did not exclusively represent propionate and butyrate producers. GC measurements for each sample are presented in Supplementary Table 1, and a list of all OTUs correlated with propionate and/or butyrate is presented in Supplementary Table 2.

Building a predictive LAD-based test

We designed 21 LAD probes to cover the 37 OTUs selected by the combined partial least squares and linear discriminant analysis model, with the intention of converting the dry lab results to a routine molecular diagnostic tool for classification. Six of the probes failed to produce a signal, so they were removed from the assay. The remaining LAD probes were used to analyze 71 random samples, nine of which were not PacBio-sequenced. The performance of the LAD probes is presented in Supplementary Figure 2.

When signals from the 15 LAD probes were used as an input, the best separation, yielding the highest model prediction accuracy (leave-one-out cross-validated), was observed at 2.6 (Figure 2A). This value corresponded well to the value derived by applying a formula designed to find outliers in positively skewed data like ours (i.e., median value + 3 × median absolute deviation → 0.92 + 3 × 0.54 = 2.54) [49]. A detailed distribution of P:B ratios among the samples tested is presented in Figure 2B.

To ensure that a high P:B ratio (≥2.6) implied increased levels of propionate relative to butyrate (and not acetate), we computed the average levels of these acids within the different groups. Indeed, the average butyrate concentration for the normal ratio group was 20%, whereas the average butyrate concentration for the high ratio group was 7.2%. Samples with a normal P:B ratio had, on average, a propionate level of 16.6%, whereas samples with a high P:B ratio had a level of 29.8% (Figure 2C). These results support our theory that a disturbed ratio between propionate and butyrate elucidates information about the levels of both acids.

Validation of the prediction model using LAD

Given the limited number of samples, we validated the LAD test by performing leave-one-out cross-validation; that is, classification of each sample was performed by taking the rest of the samples into account, excluding from the training set the one to be classified. Of nine samples with a P:B ratio ≥2.6, the algorithm correctly classified seven and missed two; however, of 62 samples with a P:B ratio <2.6, 61 were classified correctly (Figure 3). The positive predictive value showed that for any sample classified as having a high ratio, the chance of that sample indeed having a ratio >2.6 was 87.5%. The negative predictive value was 97%. All nine samples that were not PacBio-sequenced and therefore not included in the initial model for selecting OTU markers were correctly classified (all normal ratio).

We acknowledge that the number of tested samples not forming the basis for marker selection by the PLS + LDA algorithm is low (i.e., nine of 71). Therefore, testing of more independent samples will be crucial in the next phase of LAD characterization as well as further development and implementation. However, we here present a solid proof of concept to serve as a foundation for future work.

We do not possess clinical details regarding the individuals whose samples were tested, and that may present another limitation of this study. It would be of particular interest to learn whether these people suffer from health conditions for which high propionate or low butyrate has been reported. Nevertheless, we screened the metadata of 130 samples used by Zeng *et al.*, who reported that significantly increased propionate levels were associated with a high risk of stroke [22]. We found that, on average, people with a low risk of stroke had a P:B ratio <2.6, whereas significantly higher P:B ratios were observed in people with a medium and high risk of stroke (average P:B ratios of 2.04, 3.22 and 2.84 for low, medium and high risk, respectively) ($p < 0.05$).

A P:B ratio threshold of approximately 2.6 was determined using two different approaches (PLS + LDA algorithm and outlier formula). It represents a limit separating normal samples from biological outliers in terms of both SCFAs and microbiome composition. It is

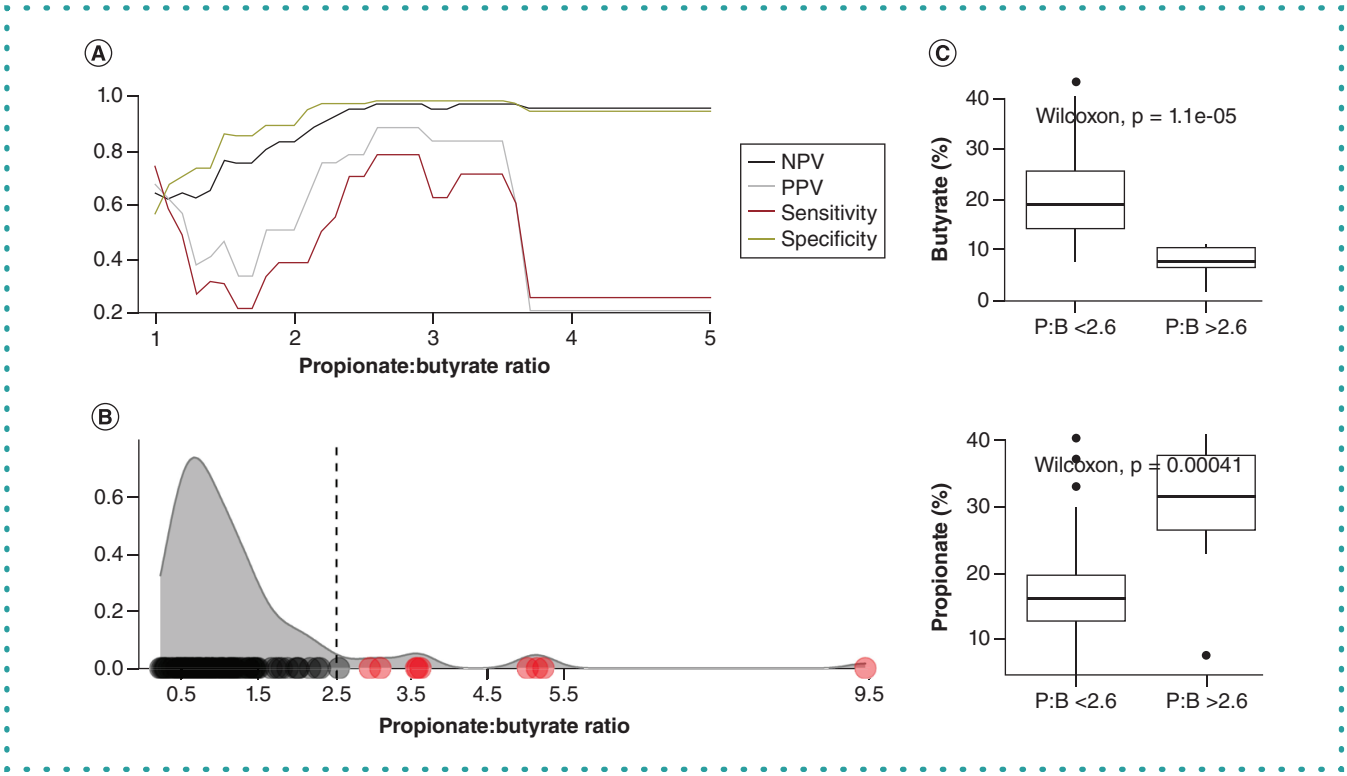

Figure 2. Propionate-to-butyrate ratio model selection. (A) Diagnostic testing accuracy of various P:B ratio thresholds using liquid array diagnostics signals as predictors. Different thresholds were tested to determine the border between high and normal ratio samples for further combined partial least squares and linear discriminant analysis classification. The dashed vertical line depicts the 2.6 ratio, the lowest ratio to yield the highest sensitivity, specificity, NPV and PPV. **(B)** Ratio density among 71 tested samples. Most of the samples (three quartiles) had a ratio <1.5, whereas the median ratio was 0.9. The dashed vertical line at 2.54 separates the outliers from the data (median + 3 × median absolute deviation). Dots along the x-axis show measured ratios for each sample, colored based on classes they belong to according to the combined partial least squares and linear discriminant analysis model (black = normal P:B ratio, red = high P:B ratio). **(C)** Box plots showing difference in distribution of butyrate (upper panel) and propionate (lower panel) levels across two different groups of samples (normal = P:B ratio <2.6, high = P:B ratio >2.6).
NPV: Negative predictive value; P:B: Propionate-to-butyrate; PPV: Positive predictive value.

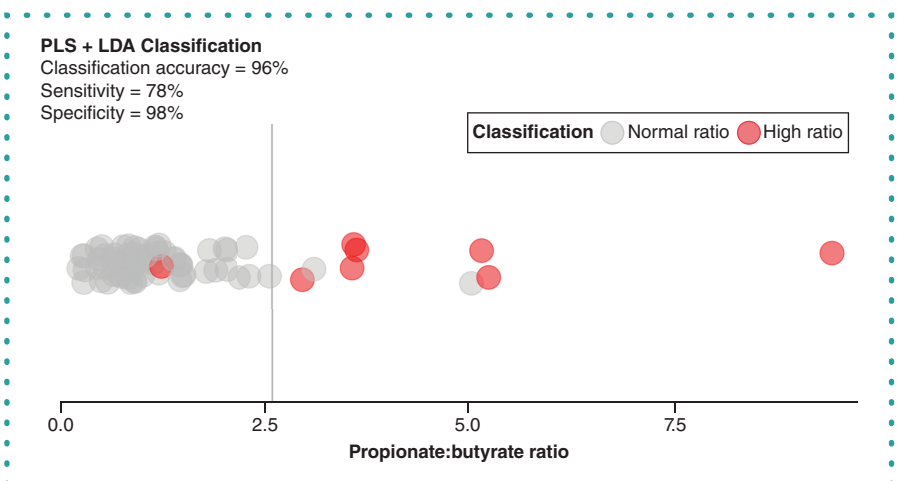

Figure 3. Combined partial least squares and linear discriminant analysis model prediction accuracy with liquid array diagnostics probe signals used as an input. The vertical line placed at 2.6 marks the border between normal and high propionate-to-butyrate ratio. The positioning of each dot shows the real sample ratio (x-axis), whereas the dot color indicates the classification by the model (gray = normal, red = high). Most (seven of nine) samples were correctly classified as having a high ratio (red dots on the right side of the 2.6 border). Only one normal ratio sample was misclassified as a high ratio sample (red dot on the left).
LDA: Linear discriminant analysis; PLS: Partial least squares.

tempting to speculate that this threshold may very well reflect an important biological threshold with a direct implication for the etiology of complex diseases.

Functional & strain resolution associations with P:B ratio

We chose to further analyze 23 randomly selected samples of various P:B ratios (17 normal, six high) by performing whole-genome shotgun sequencing in an attempt to further explore the biological differences between the two classes (i.e., normal and high ratio). On average, samples with a normal P:B ratio displayed 205 species, whereas samples with a high ratio harbored ten fewer species, suggesting a lower diversity in the latter. However, this difference did not exhibit an acceptable significance level ($p > 0.1$).

Looking deeper into the composition, we found that high ratio samples were significantly richer in *Escherichia coli*, *Phocaeicola dorei* (a known propionate producer, formerly named *Bacteroides dorei* [50]), *Enterocloster* sp001517625 (named *Clostridium bouchedurhonense* at National Center for Biotechnology Information), *Blautia_A* sp900066165 and *Anaerotruncus colihominis* (butyrate producer [51]). There was also a tendency for lower richness of *F. prausnitzii_C* (butyrate producer [37,38]) and *Eisenbergiella sp900066775* and higher richness of *Akkermansia muciniphila* (propionate producer [52]) ($p < 0.1$) (Figure 4).

Among these strains, our test is designed to detect both *E. coli* and *F. prausnitzii.* These two species have commonly been found to act as markers in a wide range of diseases [53–55]. Next, we used the sequencing reads to search for genes related to propionate and butyrate production using DIAMOND software [31], and no linear relationship was found (Figure 5).

This finding corroborates those retrieved from PacBio sequencing, where the majority of OTUs correlated to either propionate or butyrate were not known to be producers of such acids. Furthermore, the abundance of bacteria known to produce specific SCFAs was not always in a positive correlation with the fecal levels of such short-chain fatty acids. This was the case with *D. welbionis* and *F. plautii* – both butyrate producers – where relative abundance was found to be negatively correlated with butyrate levels. The latter species was instead found in a positive correlation with propionate levels, and is a target in our assay.

The seemingly complicated relationship between bacterial species and butyrate and propionate levels suggests that levels of SCFAs in fecal samples cannot be inferred by quantifying known acid producers alone, presumably because of the complex cross-feeding mechanisms involved [56]. For example, we believe that the inclusion of *Bifidobacterium adolescentis* (lactate and acetate producer) as a target of our test, is tightly related to cross-feeding between this bacterium and well-described butyrate producers (i.e., the production of butyrate is enhanced by *B. adolescentis* activity) [57,58].

Clinical utility

Currently, it appears that the most relevant clinical application of the P:B ratio would be related to neurodegenerative diseases such as Alzheimer's [20] and Parkinson's [59]. A contributing cause of neurodegenerative disease in elderly individuals is their reduced ability to metabolize propionate as a result of decreased methylmalonyl-CoA mutase activity [60]. This leads to potential accumulation of toxic methylmalonic acid, which has been associated with decreased cognitive function in older adults [61]. By contrast, it has been shown that butyrate, a histone deacetylase inhibitor, can act as a therapeutic agent by reducing levels of abnormally deposited brain amyloid-β [62,63]. Flagging samples with a high P:B ratio in a timely manner would assist clinicians to offer the necessary dietary advice to the elderly.

Other diseases and disorders can also potentially be linked to high P:B levels. An association with a significant propionate increase or butyrate decrease has been reported for the ailments listed in Table 2. The most pronounced association was reported between a high P:B ratio and IBS [18]. In addition to being a biomarker, there could also be a causality between the P:B ratio and IBS severity. Thus, this ratio could potentially have utility in treatment of these patients through, for example, dietary advice.

To the best of our knowledge, detecting samples with high P:B ratios can only be achieved by directly quantifying SCFAs and computing the ratios afterward. Measuring the level of SCFAs in fecal samples is usually accomplished by employing GC, LC, capillary electrophoresis or NMR [68]. However, given the complex sample clean-up and preparation procedure combined with high volatility of these acids, SCFA measurement using today's technology remains a challenging task [23,24,69]. This is why knowledge in the field continues to be derived from fragmented, small-scale studies that are insufficiently standardized across laboratories.

The lack of robust methods for use in clinical settings creates a gap between the state-of-the-art knowledge in the field and its practical utility and application. A simple molecular diagnostic method like the LAD test presented here allows inexpensive, high-throughput screening of fecal samples, bridging this gap. The major benefits of LAD in a clinical setting are related to simplicity and cost as well as the ability to detect the microorganisms underlying the P:B ratio, which in turn can be used in therapeutics.

Our approach offers a solution for at least two problems. First, it focuses on the ratio between propionate and butyrate, ignoring their absolute values, which are known to fluctuate based on the time of day a sample is collected and processed [70]. Second, it circumvents the need to measure SCFA levels, utilizing a robust molecular diagnostic system instead.

We offer an indirect way of detecting both propionate and butyrate levels by identifying biological outliers, that is samples with highest propionate and/or lowest butyrate ratios. The tool we present here is not aimed at replacing other conventional 16S rRNA gene or SCFA analyses; nor does it have the capacity to do so, as it is strictly focused on inferring a narrow segment of microbial functionality. Rather, it represents an applicable solution that integrates both types of methodologies into a single measurement of high clinical utility.

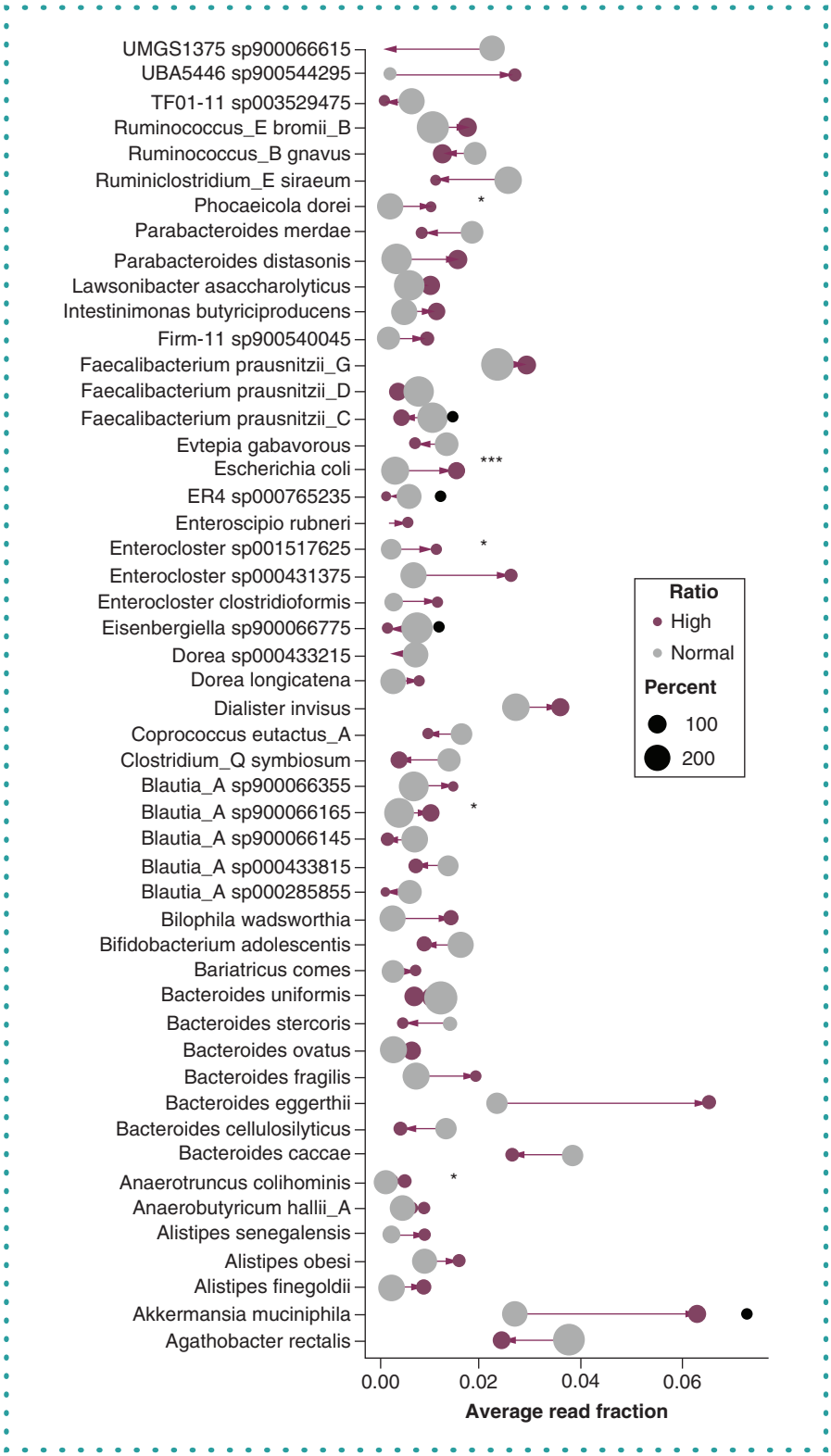

Figure 4. Top 50 species with the greatest difference in average abundance between groups (samples with normal vs high propionate-to-butyrate ratio). Gray circles indicate the average abundance of normal ratio samples, and burgundy represents the abundance of samples with a high ratio. The circle size shows the percentage of samples within the group where the bacterium was found. Arrows point toward samples with a high ratio. Dot and star symbols indicate significant differences as determined by Wilcoxon test.
*$p < 0.05$; ***$p < 0.005$; •$p < 0.1$.

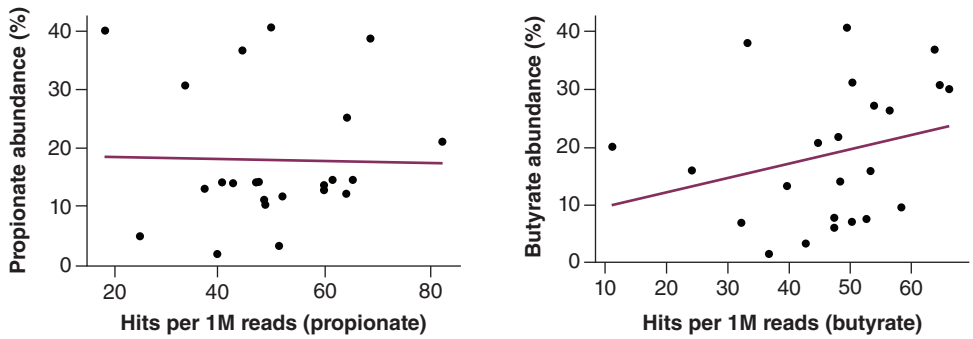

Figure 5. Relationship between relative abundance of propionate and butyrate and corresponding number of reads with a hit (highest bit score e-value ≤1e-05) with respective marker genes. The number of hits was normalized after considering the sequence length of queries and sequencing depth. M: Million.

Table 2. Studies associating diseases with an increase/decrease in fecal short-chain fatty acid levels (high propionate-to-butyrate ratio).

Health disorder	Individuals tested	Significant change compared with controls	Ref.
Obesity	30 lean 35 overweight 33 obese	↑Total SCFAs ↑Propionate	[11]
Type 2 diabetes	952 from LifeLines DEEP cohort	↑Propionate	[19]
NAFLD	27 healthy 32 NAFLD	↑Acetate ↑Propionate	[21]
IBS	25 healthy 25 IBS	↓Butyrate	[18]
Stroke	51 low risk of stroke 54 medium risk of stroke 36 high risk of stroke	↓Butyrate	[22]
ASD	20 healthy 30 ASD	↓Acetate ↓Butyrate ↑Valerate	[64]
CKD	61 healthy 128 CKD	↓Butyrate	[65]
Rett syndrome	29 healthy 50 Rett syndrome	↑Total SCFAs ↑Propionate ↑Isovalerate ↑Isobutyrate	[66]
IBS	26 healthy 26 IBS	↑Total SCFAs ↑Acetate ↑Propionate	[67]
PD	34 healthy 34 PD	↓Total SCFAs ↓Butyrate	[59]

ASD: Autism spectrum disorder; CKD: Chronic kidney disease; IBS: Irritable bowel syndrome; NAFLD: Nonalcoholic fatty liver disease; PD: Parkinson's disease; SCFAs: Short-chain fatty acids.

Conclusion

Here we present a novel qPCR-compatible, single-tube multiplex test that predicts samples with increased ratios of propionate relative to butyrate. Circumventing the need to directly measure the SCFA content in fecal samples, a robust and simple test like LAD will enable high-throughput analysis and regular monitoring of functional dysbiosis in the gut.

Executive summary

- Healthy adult fecal propionate and butyrate levels are expected to be equimolar.
- An increased propionate-to-butyrate ratio has been linked to several health disorders.
- Measurement of levels of short-chain fatty acids is challenging because of their highly volatile nature, presenting a major bottleneck for high-throughput studies.
- The challenges related to short-chain fatty acid measurements create a gap between knowledge acquired in the field and its clinical utility.
- This article presents a method for predicting and classifying samples with significantly elevated propionate-to-butyrate ratios by directly targeting predictor bacteria, circumventing the need to measure short-chain fatty acid levels.
- The method is based on a liquid array diagnostics assay, a quantitative PCR-compatible test capable of detecting multiple targets in a single-tube multiplex reaction.
- The test predicting samples with high propionate-to-butyrate ratios showed 78% sensitivity and 98% specificity (leave-one-out cross-validated).
- The assay presented here has the potential to be utilized in high-throughput studies, validating reported findings in the literature in addition to serving as a robust screening tool for routine diagnostics.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2022-0045

Author contributions

The study was conceived by K Rudi. P Hiseni planned and conducted the lab work. All authors discussed and aided in interpreting the results. P Hiseni wrote the manuscript with equal input from all authors.

Financial & competing interests disclosure

This work was funded by the Norwegian Research Council (research and development grant no. 283783). At the time of writing the manuscript, P Hiseni, K Furu and FT Hegge were employed by Genetic Analysis AS, a licensee of the liquid array diagnostics method. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The collection and handling of fecal samples were performed in accordance with the Norwegian Regional Committee for Medical and Health Research Ethics (reference no. 2010/3209).

Data sharing statement

PacBio and Illumina whole-genome shotgun sequencing data are deposited at Sequence Read Archive (PRJNA820539).

Open access

References

Papers of special note have been highlighted as: • of interest

1. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* 7(3), 189–200 (2016).
2. Donohoe DR, Garge N, Zhang X *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* 13(5), 517–526 (2011).
3. Wang H-B, Wang P-Y, Wang X, Wan Y-L, Liu Y-C. Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein claudin-1 transcription. *Dig. Dis. Sci.* 57(12), 3126–3135 (2012).
4. Willemsen LEM, Koetsier MA, van Deventer SJH, van Tol EAF. Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E_1 and E_2 production by intestinal myofibroblasts. *Gut.* 52(10), 1442–1447 (2003).
5. Chen J, Vitetta L. Inflammation-modulating effect of butyrate in the prevention of colon cancer by dietary fiber. *Clin. Colorectal Cancer* 17(3), e541–e544 (2018).
6. Encarnação JC, Abrantes AM, Pires AS, Botelho MF. Revisit dietary fiber on colorectal cancer: butyrate and its role on prevention and treatment. *Cancer Metastasis Rev.* 34(3), 465–478 (2015).
7. Miyamoto J, Kasubuchi M, Nakajima A, Irie J, Itoh H, Kimura I. The role of short-chain fatty acid on blood pressure regulation. *Curr. Opin. Nephrol. Hypertens.* 25(5), 379–383 (2016).
8. Dalile B, Van Oudenhove L, Vervliet B, Verbeke K. The role of short-chain fatty acids in microbiota–gut–brain communication. *Nat. Rev. Gastroenterol. Hepatol.* 16(8), 461–478 (2019).
9. Ulven T. Short-chain free fatty acid receptors FFA2/GPR43 and FFA3/GPR41 as new potential therapeutic targets. *Front. Endocrinol. (Lausanne)* 3, 111 (2012).
10. Chambers ES, Viardot A, Psichas A *et al.* Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut* 64(11), 1744–1754 (2015).
11. Schwiertz A, Taras D, Schäfer K *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* 18(1), 190–195 (2010).

12. Cook SI, Sellin JH. Review article: short chain fatty acids in health and disease. *Aliment. Pharmacol. Ther.* 12(6), 499–507 (1998).
13. Fernandes J, Su W, Rahat-Rozenbloom S, Wolever TMS, Comelli EM. Adiposity, gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr. Diabetes* 4(6), e121 (2014).
14. Duncan SH, Belenguer A, Holtrop G, Johnstone AM, Flint HJ, Lobley GE. Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. *Appl. Environ. Microbiol.* 73(4), 1073–1078 (2007).
15. Russell WR, Gratz SW, Duncan SH *et al.* High-protein, reduced-carbohydrate weight-loss diets promote metabolite profiles likely to be detrimental to colonic health. *Am. J. Clin. Nutr.* 93(5), 1062–1072 (2011).
16. Walker AW, Duncan SH, Leitch ECM, Child MW, Flint HJ. pH and peptide supply can radically alter bacterial populations and short-chain fatty acid ratios within microbial communities from the human colon. *Appl. Environ. Microbiol.* 71(7), 3692–3700 (2005).
17. Duar RM, Kyle D, Casaburi G. Colonization resistance in the infant gut: the role of *B. infantis* in reducing pH and preventing pathogen growth. *High-Throughput* 9(2), 7 (2020).
18. Farup PG, Rudi K, Hestad K. Faecal short-chain fatty acids – a diagnostic biomarker for irritable bowel syndrome? *BMC Gastroenterol.* 16(1), 51 (2016).
- **Reported a significant decrease in butyrate levels in patients with irritable bowel syndrome (IBS) and the potential use of propionate-to-butyrate ratio as a metric for IBS diagnosis.**
19. Sanna S, van Zuydam NR, Mahajan A *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nat. Genet.* 51(4), 600–605 (2019).
- **Showed an association between type 2 diabetes and increased levels of propionate in fecal samples.**
20. Killingsworth J, Sawmiller D, Shytle RD. Propionate and Alzheimer's disease. *Front. Aging. Neurosci.* 12, 501 (2020).
21. Rau M, Rehman A, Dittrich M *et al.* Fecal SCFAs and SCFA-producing bacteria in gut microbiome of human NAFLD as a putative link to systemic T-cell activation and advanced disease. *United European Gastroenterol. J.* 6(10), 1496–1507 (2018).
- **Reported increased levels of propionate and acetate in samples collected from people suffering from nonalcoholic fatty liver disease.**
22. Zeng X, Gao X, Peng Y *et al.* Higher risk of stroke is correlated with increased opportunistic pathogen load and reduced levels of butyrate-producing bacteria in the gut. *Front. Cell. Infect. Microbiol.* 9(4), 4 (2019).
23. Torii T, Kanemitsu K, Wada T, Itoh S, Kinugawa K, Hagiwara A. Measurement of short-chain fatty acids in human faeces using high-performance liquid chromatography: specimen stability. *Ann. Clin. Biochem.* 47(5), 447–452 (2010).
24. Li M, Zhu R, Song X, Wang Z, Weng H, Liang J. A sensitive method for the quantification of short-chain fatty acids by benzyl chloroformate derivatization combined with GC-MS. *Analyst* 145(7), 2692–2700 (2020).
25. Hiseni P, Wilson RC, Storrø O, Johnsen R, Øien T, Rudi K. Liquid array diagnostics: a novel method for rapid detection of microbial communities in single-tube multiplex reactions. *Biotechniques* 66(3), 143–149 (2019).
26. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13(5), 278–289 (2015).
27. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584 (2016).
28. Aitchison J. The statistical analysis of compositional data. *J. R. Stat. Soc. Series B Stat. Methodol.* 44(2), 139–160 (1982).
29. Mehmood T, Martens H, Sæbø S, Warringer J, Snipen L. A partial least squares based algorithm for parsimonious variable selection. *Algorithms Mol. Biol.* 6(1), 27 (2011).
30. Eggesbø M, Moen B, Peddada S *et al.* Development of gut microbiota in infants not exposed to medical interventions. *APMIS* 119(1), 17–35 (2011).
31. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 18(4), 366–368 (2021).
32. Reichardt N, Duncan SH, Young P *et al.* Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* 8(6), 1323–1335 (2014).
33. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20(1), 257 (2019).
34. Breitwieser FP, Baker DN, Salzberg SL. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* 19(1), 198 (2018).
35. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* 3, e104 (2017).
36. Hiseni P, Rudi K, Wilson RC, Hegge FT, Snipen L. HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome* 9(1), 165 (2021).
37. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 52(Pt 6), 2141–2146 (2002).
38. Lopez-Siles M, Khan TM, Duncan SH, Harmsen HJM, Garcia-Gil LJ, Flint HJ. Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl. Environ. Microbiol.* 78(2), 420–428 (2012).
39. Ahn S, Jin TE, Chang DH *et al. Agathobaculum butyriciproducens* gen. nov. sp. nov., a strict anaerobic, butyrate-producing gut bacterium isolated from human faeces and reclassification of *Eubacterium desmolans* as *Agathobaculum desmolans* comb. nov. *Int. J. Syst. Evol. Microbiol.* 66(9), 3656–3661 (2016).
40. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. *Environ. Microbiol.* 19(1), 29–41 (2017).
41. Zhao R, Sun J, Mo H, Zhu Y. Analysis of functional properties of *Lactobacillus acidophilus*. *World J. Microbiol. Biotechnol.* 23(2), 195–200 (2007).
42. Takada T, Kurakawa T, Tsuji H, Nomoto K. *Fusicatenibacter saccharivorans* gen. nov., sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 63(Pt 10), 3691–3696 (2013).
43. Liu C, Finegold SM, Song Y, Lawson PA. Reclassification of *Clostridium coccoides*, *Ruminococcus hansenii*, *Ruminococcus hydrogenotrophicus*, *Ruminococcus luti*, *Ruminococcus productus* and *Ruminococcus schinkii* as *Blautia coccoides* gen. nov., comb. nov., *Blautia hansenii* comb. nov., *Blautia hydrogenotrophica* comb. nov., *Blautia luti* comb. nov., *Blautia producta* comb. nov., *Blautia schinkii* comb. nov. and description of *Blautia wexlerae* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 58(8), 1896–1902 (2008).
44. Le Roy T, Moens de Hase E, Van Hul M *et al. Dysosmobacter welbionis* is a newly isolated human commensal bacterium preventing diet-induced obesity and metabolic disorders in mice. *Gut* 71(3), 534–543 (2021).
45. Carlier J-P, Bedora-Faure M, K'ouas G, Alauzet C, Mory F. Proposal to unify *Clostridium orbiscindens* Winter *et al.* 1991 and *Eubacterium plautii* (Séguin 1928) Hofstad and Aasjord 1982, with description of *Flavonifractor plautii* gen. nov., comb. nov., and reassignment of *Bacteroides capillosus* to *Pseudoflavonifractor capillosus* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* 60(3), 585–590 (2010).
46. Takahashi K, Nishida A, Fujimoto T *et al.* Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. *Digestion* 93(1), 59–65 (2016).
47. Laserna-Mendieta EJ, Clooney AG, Carretero-Gomez JF *et al.* Determinants of reduced genetic capacity for butyrate synthesis by the gut microbiome in Crohn's disease and ulcerative colitis. *J. Crohns Colitis* 12(2), 204–216 (2018).
48. Pozuelo M, Panda S, Santiago A *et al.* Reduction of butyrate- and methane-producing microorganisms in patients with irritable bowel syndrome. *Sci. Rep.* 5(1), 12693 (2015).
49. Leys C, Ley C, Klein O, Bernard P, Licata L. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49(4), 764–766 (2013).
50. Gutiérrez N, Garrido D. Species deletions from microbiome consortia reveal key metabolic interactions between gut microbes. *mSystems* 4(4), e00185–19 (2019).
51. Lawson PA, Song Y, Liu C *et al. Anaerotruncus colihominis* gen. nov., sp. nov., from human faeces. *Int. J. Syst. Evol. Microbiol.* 54(Pt 2), 413–417 (2004).
52. de Vos WM. Microbe profile: *Akkermansia muciniphila*: a conserved intestinal symbiont that acts as the gatekeeper of our mucosa. *Microbiology (Reading)* 163(5), 646–648 (2017).
53. Khan AA, Khan Z, Malik A *et al.* Colorectal cancer–inflammatory bowel disease nexus and felony of *Escherichia coli*. *Life Sci.* 180, 60–67 (2017).
54. Lopez-Siles M, Martinez-Medina M, Busquets D *et al.* Mucosa-associated *Faecalibacterium prausnitzii* and *Escherichia coli* co-abundance can distinguish irritable bowel syndrome and inflammatory bowel disease phenotypes. *Int. J. Med. Microbiol.* 304(3), 464–475 (2014).
55. Machiels K, Joossens M, Sabino J *et al.* A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis. *Gut* 63(8), 1275 (2014).
56. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de los Reyes-Gavilán CG, Salazar N. Intestinal short chain fatty acids and their link with diet and human health. *Front. Microbiol.* 7(185), 185 (2016).
57. Belenguer A, Duncan SH, Calder AG *et al.* Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl. Environ. Microbiol.* 72(5), 3593–3599 (2006).
58. Rios-Covian D, Gueimonde M, Duncan SH, Flint HJ, de los Reyes-Gavilan CG. Enhanced butyrate formation by cross-feeding between *Faecalibacterium prausnitzii* and *Bifidobacterium adolescentis*. *FEMS Microbiol. Lett.* 362(21), fnv176 (2015).
59. Unger MM, Spiegel J, Dillmann K-U *et al.* Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Relat. Disord.* 32, 66–72 (2016).

60. Elliot JM. Propionate metabolism and vitamin B12. In: *Digestive Physiology and Metabolism in Ruminants: Proceedings of the 5th International Symposium on Ruminant Physiology, Held at Clermont-Ferrand, on 3rd–7th September, 1979*. Ruckebusch Y, Thivend P (Eds). Springer Netherlands, Dordrecht, The Netherlands, 485–503 (1980).
61. McCracken C, Hudson P, Ellis R, McCaddon A. Methylmalonic acid and cognitive function in the Medical Research Council Cognitive Function and Ageing Study. *Am. J. Clin. Nutr.* 84(6), 1406–1411 (2006).
62. Fernando W, Martins IJ, Morici M *et al.* Sodium butyrate reduces brain amyloid-β levels and improves cognitive memory performance in an Alzheimer's disease transgenic mouse model at an early disease stage. *J. Alzheimers Dis.* 74(1), 91–99 (2020).
63. Govindarajan N, Agis-Balboa RC, Walter J, Sananbenesi F, Fischer A. Sodium butyrate improves memory function in an Alzheimer's disease mouse model when administered at an advanced stage of disease progression. *J. Alzheimers Dis.* 26, 187–197 (2011).
64. Liu S, Li E, Sun Z *et al.* Altered gut microbiota and short chain fatty acids in Chinese children with autism spectrum disorder. *Sci. Rep.* 9(1), 287 (2019).
- **Found decreased levels of acetate and butyrate in samples collected from children with autism spectrum disorder.**
65. Wang S, Lv D, Jiang S *et al.* Quantitative reduction in short-chain fatty acids, especially butyrate, contributes to the progression of chronic kidney disease. *Clin. Sci. (Lond.)* 133(17), 1857–1870 (2019).
- **Showed that people suffering from chronic kidney disease had significantly lower levels of butyrate in their fecal samples.**
66. Strati F, Cavalieri D, Albanese D *et al.* Altered gut microbiota in Rett syndrome. *Microbiome* 4(1), 41 (2016).
- **Reported an association between Rett syndrome and increased levels of propionate and total short-chain fatty acid levels.**
67. Tana C, Umesaki Y, Imaoka A, Handa T, Kanazawa M, Fukudo S. Altered profiles of intestinal microbiota and organic acids may be the origin of symptoms in irritable bowel syndrome. *Neurogastroenterol. Motil.* 22(5), 512–519; e114-5 (2010).
- **Found increased levels of acetate and propionate in fecal samples collected from people suffering from IBS.**
68. Primec M, Mičetić-Turk D, Langerholc T. Analysis of short-chain fatty acids in human feces: a scoping review. *Anal. Biochem.* 526, 9–21 (2017).
69. García-Villalba R, Giménez-Bastida JA, García-Conesa MT, Tomás-Barberán FA, Espín JC, Larrosa M. Alternative method for gas chromatography–mass spectrometry analysis of short-chain fatty acids in faecal samples. *J. Sep. Sci.* 35(15), 1906–1913 (2012).
70. Jones J, Reinke SN, Ali A, Palmer DJ, Christophersen CT. Fecal sample collection methods and time of day impact microbiome composition and short chain fatty acid concentrations. *Sci. Rep.* 11(1), 13964 (2021).