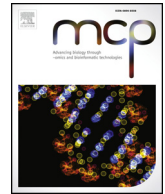




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Development of qPCR platform with probes for quantifying prevalent and biomedically relevant human gut microbial taxa

Irina Kurina^{a,b,*}, Anna Popenko^{a,b}, Natalia Klimenko^{a,c}, Stanislav Koshechkin^a, Liubov Chuprikova^a, Maxim Filipenko^d, Alexander Tyakht^{a,c}, Dmitry Alexeev^{a,b}

^a Knomics LLC, Skolkovo Innovation Center, Bolshoy Bulvar str., building 42, premise 1, room 1639, Moscow, 143026, Russia

^b Atlas Biomed Group, 31 Malaya Nikitskaya str., 121069, Moscow, Russia

^c Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology Russian Academy of Sciences, 34/5 Vavilova str., 119334, Moscow, Russia

^d Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, 8 Lavrentiev ave., Novosibirsk, 630090, Russia

ARTICLE INFO

Keywords:

Quantitative real-time PCR
Gut microbiome
Taxonomic profiling
Microbial biomarkers
Diet
Nutrition assessment technique
Responders

ABSTRACT

Nowadays the advent of innovative high-throughput sequencing allows obtaining high-quality microbiome profiling. However, PCR-based tests are still considered the “golden standard” for many clinical applications. Here, we designed a qPCR-based platform with fluorescent-labeled oligonucleotide probes for assessing human gut microbiome composition. The system allows conducting qualitative and semiquantitative analysis for 12 prokaryotic taxa that are prevalent in the human gut and associated with diseases, diet, age and other factors. The platform was validated by comparing microbiome profile data obtained with two different methods - the platform and high-throughput 16S rRNA sequencing - across 42 stool samples. The test can form the basis for precise and cost-efficient microbiome assay for large-scale surveys including clinical trials with interventions related to diet and disease risks.

1. Introduction

The gut microbiome is virtually an organ of the human body. The gut is colonized with 10^{13} of microbes belonging to thousands of species. Alterations of microbiome composition are linked to lifestyle changes, short- and long-term diet [1], probiotics intake [1,2], medications including antibiotics [3] and diseases including obesity, type 2 diabetes [4], colorectal cancer [5], Crohn's disease [6] and others. Based on the discovered associations, a number of microbial taxa have been proposed as potential biomarkers. For example, the abundance of *Akkermansia muciniphila*, a mucin-degrading bacterium is inversely associated with body fat mass and glucose intolerance [7]. Disrupted gut community structure is also one of the risk factors for cardiovascular diseases [8]. Many of the bacteria belonging to *Clostridiales* order are efficient producers of butyrate and other short-chain fatty acids known to possess anti-inflammatory, immunomodulatory and other important activities including gut epithelium integrity maintenance [9,10].

The task of typing and enumerating gut microbial taxa is highly important. There are a plethora of methods for addressing it, including cultivation-based methods, PCR approaches and high-throughput sequencing. Despite the recent progress with culturomics [11], routine

stool cultivation allows to enumerate a narrow range of gut microbes with low accuracy. The methods based on microbiome sequencing are becoming more popular - the “shotgun” metagenomics as well as amplicon-based surveys (using 16S rRNA gene for prokaryotes and ITS region for yeast, etc.). Unlike the culture-dependent approaches, the sequencing gives a total semiquantitative portrait of the microbiome - capturing the uncultivable diversity as well.

An alternative method for taxonomic and functional profiling of microbial communities are hybridization microarrays that allow high-throughput detection of thousands of targets at a time including prokaryotes, viruses, yeast and protozoa [12–14]. Although microarrays capture low-abundant microbes and can be cost-efficient on a large scale, they only detect a fixed set of microbial targets and require a specialized microarray analysis instrument.

Another category for microbiome profiling methods is based on PCR using oligonucleotide sequences specific to selected microbial clades. The real-time PCR (qPCR) is superior to the above-mentioned methods in terms of generally available equipment, cheaper sample preparation, flexibility in application and robustness for diagnostic applications. Human gut-specific PCR panels have been described before. One of them describes a qPCR test for detecting 4 taxa using SYBR Green:

* Corresponding author. Knomics LLC, Skolkovo Innovation Center, Bolshoy Bulvar str., building 42, premise 1, room 1639, Moscow, 143026, Russia.

E-mail addresses: kurina@atlas.ru (I. Kurina), a.tyakht@gmail.com (A. Tyakht).

Bacteroidetes, *Firmicutes*, *Bacteroides thetaiotaomicron* and *Enterococcus* spp. [15]. In a more extensive manner, a real-time quantitative PCR-based analysis platform termed 'GUt Low-Density Array' (GULDA) with SYBR Green for 27 bacterial taxa was introduced [16]. Detailed clade-focused applications are available - for example, for detecting all known species of human gut-dwelling bifidobacteria [17] using genus- and species-specific primers and probes for qPCR and further proceeding with PCR product restriction for species identification.

In this article we describe a qPCR-based platform for taxonomic profiling of gut microbiome targeting taxa that are prevalent in the gut and/or associated with important diseases, diet, populations, age and other factors. The platform provides a flexible tool for analyzing an individual microbiome composition (e.g., from a stool sample) to assess its general deviation from the reference population distributions, as well as quantify potential microbial biomarkers of dysbiosis - universal as well as disease-specific.

2. Materials and methods

2.1. Selection of targets

The target taxa were selected based on the principles of combining the taxa differentially abundant in subjects with major diseases (obesity, type 2 diabetes, inflammatory bowel diseases [IBD] and atherosclerosis), the taxa enriched in probiotic representatives and the "driver" taxa of the human gut microbiome. We collected a statistical summary of the targets' abundance in the 16S rRNA datasets of stool samples collected from the Russian general population during the OhMyGut crowd-funded project [1] and other studies (see [Supplementary Table 2](#)).

2.2. Primers and probes design

The HITdb v1.00 database was used as a non-redundant curated set of 16S rRNA genes specific for human gut microbiota [18]. In order to identify primers and probes targeting specific groups of gut microbes, the gene sequences were analyzed to identify the regions that were both highly conservative within a given group and sufficiently variable to separate the group from other members of the gut microbiome.

On the first step, nucleotide sequences belonging to a certain taxon were extracted from the database and aligned using the Aligner tool [19]. The obtained alignment was then used to define regions conservative within a given taxon. For this, we split the alignment into overlapping 20-mer blocks (set of columns) and only highly conservative blocks (showing 90% agreement in the consensus sequence with < 3 degenerate positions) were selected.

On the next step, we checked the taxon-specificity of each conservative block. For this, a BLAST database was constructed using all the sequences from HITdb except those for the analyzed taxon. A conservative block was considered taxon-specific if, during a search against the BLAST database, it had aligned to < 1% of non-specific targets with 100% identity and coverage. In case when no such taxon-specific conservative blocks were identified, all the blocks were combined into pairs of blocks separated from each other by 50–300 bases (typical qPCR fragment size). A pair of blocks was considered taxon-specific if the blocks had not aligned to the same non-specific target simultaneously.

2.3. DNA extraction

Stool DNA was extracted as described previously [1]. For facilitating PCR amplification, additional purification from PCR inhibitors was performed using AmpliPrime DNK-Sorb-B (Federal State Institution Central Research Institute of Epidemiology of Rospotrebnadzor, Russia) according to the manufacturer's instructions.

2.4. Cloning and production of recombinant plasmids containing DNA fragments for amplifying target loci

A 1200-1400 bp fragment of 16S rDNA was amplified as follows. The master mix in a reaction volume of 50 μ L included: 1x Taq polymerase buffer (65 mM Tris-HCl [pH 8.9]); 16 mM $(\text{NH}_4)_2\text{SO}_4$; 0.05% Tween 20; 3.5 mM MgCl_2 , 0.2 mM dNTPs, 50 ng genomic bacterial DNA or DNA isolated from human feces, 0.3 μ M solutions of oligonucleotide primers, 1U Taq-polymerase (Biosan), 0.5U Pfu-polymerase (Biosan). Amplification was performed in Tercyc thermal cycler (DNK-Tehnologiya) under the following program: 15 min at 95 °C for initial denaturation, 30 cycles of 10 s at 95 °C for denaturation, 10 s at 60 °C for annealing, 60 s at 72 °C for elongation. The amplification products were hydrolyzed using the respective restriction endonucleases (SibEnzyme) and ligated using pBluscriptII SK(+) vector hydrolyzed with the same endonucleases within 3 h with 100U T4 DNA ligases (Biosan). Competent cells of *E. coli* strain XL1-Blue (Stratagene) were transformed with a ligase mixture. In the plasmid clones selected according to the results of specific PCR, the nucleotide sequence of the insert was determined by Sanger sequencing for confirmation. Plasmid DNA was extracted from 100 mL of overnight culture on LB medium using QIAGEN Plasmid Midi Kit (QIAGEN) according to the manufacturer's instructions.

2.5. Measurement of exact plasmid concentrations, preparation of reference samples for qPCR

Concentrations of the obtained reference plasmid DNA were measured by spectrophotometry and fluorometry (Qubit BR kit, Invitrogen). Two μ g DNA was hydrolyzed by restriction endonuclease SalGI for linearization. The obtained linear standards were diluted to a concentration of 10^7 to 10^0 copies of plasmid DNA per μ L in a sterile buffer containing 10 mM TrisHCl pH 7.6 and poly-A RNA (5 ng per μ L). DNA concentration in the obtained standards was measured using digital PCR on a QX100 Droplet Digital PCR System (Bio-Rad) platform according to the manufacturer's instructions. For this purpose, 20 μ L PCR mixture containing the analyzed DNA (< 66 ng per 20 μ L), 1X PCR mixture (Bio-Rad), 300 nM oligonucleotide primers and 100 nM TaqMan probe were prepared. To obtain microdroplets, 20 μ L of the prepared PCR mixture and 70 μ L of oil to generate drops were placed in open DG8 cartridges. 40 μ L of the obtained microdroplets was transferred to a 96-well PCR plate, sealed with foil and placed in a thermal cycler. Amplification program: 96 °C–10 min then 50 cycles were performed as follows: 96 °C–30 s, 57 °C–60 s with final warming for 10 min at 98 °C. Then the microdroplets were read and analyzed using a Droplet Reader instrument; the obtained data were processed in the QuantaSoft program (Bio-Rad).

2.6. qPCR

Quantitative qPCR was carried out using a CFX96 thermal cycler (BioRad) in 96-well PCR plate (PCR-96M2-HS-C; Axygen). Sequences of primers and TaqMan probes are provided in [Supplementary Tables 2 and 3](#). Amplification for each sample was carried out in two replicates in a total volume of 20 μ L containing: 65 mM Tris-HCl (pH 8.9), 16 mM $(\text{NH}_4)_2\text{SO}_4$, 3.0 mM MgCl_2 , 0.05% Tween 20, 0.2 mM dNTPs (BIOSAN LLC), 0.3 μ M oligonucleotide primers with 0.1 μ M TaqMan probe, 1 U hot start Taq polymerase (ICBFM SB RAS), 2 μ L target DNA or plasmid standard. Amplification was carried out according to the following protocol:

1. 95 °C - 15 min
2. 95 °C - 10 s

60 °C - 30 s FAM/Green, R6G/JOE/HEX/VIC/Yellow and ROX/Orange; Cy5/Red

72 °C - 30 s

Repeat step #2 45 times

To quantify the DNA of the target microbial taxa in a sample, the method of calibration curves was used: successive 10-fold dilutions of plasmid standards were obtained and then amplified. The concentration range of calibration standards is shown in [Supplementary Table 4](#).

Gene amplification levels obtained by PCR were recorded as Ct values that are inversely related to the initial DNA concentration. Quantity Mean was calculated automatically by the CFX96 (BioRad) software control program based on the set calibration curve values. After automatic data processing, the inaccuracies of automatic reading for some of the curves were corrected by manual alignment. The quantity values were exported to a spreadsheet for further analysis and comparison of data with the results of 16S rRNA gene sequencing.

2.7. Comparison of qPCR and 16S rRNA sequencing microbiome profiles

The library preparation for stool DNA samples and Illumina MiSeq high-throughput sequencing was performed as described previously [2] targeting V4 region of the gene with the following modification of 515F–806R primers: GTGBCAGCMGCCGCGGTAA and GACTACNVGG-GTMTCTAATCC. Data analysis has been done using Knomics-Biota system [20]. Reads were denoised using DADA2 algorithm [21] with the variable trimming length (from 251 to 253 bp). The taxonomic classification of the denoised reads has been done using QIIME2 naive-bayes classifier [22] and GreenGenes database [18] with pre-trimmed and clustered with 97% identity sequences using TaxMan [23] and CD-HIT software (version 4.8.1) [24]. The relative abundance of each taxon was calculated by rarefying to 3000 reads per sample and normalizing to 100%. Proportions of genera and families were obtained by summing the levels of their member amplicon-sequencing variants (ASV). The qPCR and 16S rRNA values for target taxa abundance were compared using Spearman's correlation. Multiple comparison adjustment has been done using the Benjamini-Hochberg method. During PERMANOVA analysis, abundance values of PCR targets were used as predictors and Bray-Curtis dissimilarity matrix calculated from 16S rRNA composition - as output.

3. Results

The developed qPCR platform allows semiquantitative detection of major gut microbial taxa using qPCR with oligonucleotide probes. It provides an efficient way to characterize the major components of the human gut microbiome in relation to diseases, dietary interventions as well as for comparison with other factors. The platform includes specific oligonucleotide probes for detection and quantification of the microbes listed in [Table 1](#).

3.1. Selection of targets

The final list included 12 taxa shown in [Table 1](#). The number of taxa was selected to balance the scope of detection with the economical expenses due to the probe design and analysis cost.

3.2. Results of primers and probes design

Following the *in silico* analysis described in the Materials and Methods, the consensus sequences were derived from taxon-specific conservative blocks (or pairs of blocks). Primers and probes were designed based on the consensus sequences using Vector NTI software [54]. The following requirements had to be fulfilled: melting temperature (T_m) — 58–62 °C for a primer and 65–75 °C for a probe, GC content — 40–65%, no hairpins with the $\Delta G < -3$ kcal/mol and > 2 kcal/mol, no primer-dimers with the $\Delta G < -5$ kcal/mol.

The 12 pairs of primers and probes selected in this way are specific to the investigated taxa. The qPCR-based platform contains a pair of

primers and an oligonucleotide probe for each taxon.

3.3. Analytical characteristics of the qPCR platform

The platform characteristics were calculated based on the amplification of consecutive 10-fold dilutions of plasmid standards. An example of amplification for one of the primer sets is shown in [Fig. 1](#). All primer sets worked in the efficiency range of 81–128.9%, the coefficient of determination was high ($R^2 > 0.979$) confirming high reproducibility of the results. Detailed analytical characteristics are listed in [Supplementary Table 5](#). Due to multitude of targeted taxa, the specificity of the qPCR platform was evaluated via comparison with the results of high-throughput 16S rRNA sequencing.

3.4. Evaluating the qPCR platform on stool samples

For validation of the qPCR platform, we selected 42 stool samples from generally healthy Russian population that, according to previously performed 16S rRNA sequencing, were representative of global gut microbiome diversity and contained widely varying proportions of the target taxa of the qPCR platform. Each of the stool DNA samples were amplified using each of the primer pairs and probes (listed in [Supplementary Tables 2 and 3](#)). In each qPCR run, the analysis was performed for each taxon on all samples accompanied with negative control and a series of positive controls. The positive controls were plasmids in 4 dilutions each used to assess PCR efficacy (concentrations are shown in [Supplementary Table 4](#)).

Two repeats were performed for each qPCR experiment to obtain more precise results. Standard curve was drawn by plotting the logarithmic input DNA concentration vs. mean cycle threshold (mCt). Overall, the differences between the replicates for stool samples were significantly lower than between samples ($\Delta Ct 0.54 \pm 0.29$ vs. 5.31 ± 4.63 , $p = 0$; see [Supplementary Fig. S1](#)). In case if the discrepancy between the two replicates is more than 1 cycle it is recommended that the experiment should be repeated and the outlying replicate be removed.

3.5. Comparison with 16S rRNA sequencing microbiome profiles

The same stool DNA samples were also subject to 16S rRNA library preparation (V4 region) and sequencing. The taxonomic classification of each read was used to obtain the total composition of each sample and the relative abundance of each identified taxon. Overall, when the abundance values across all taxa and samples were pooled, the correlation between the qPCR and 16S rRNA sequencing was significant (Spearman's $r = 0.66$, $p = 0$). For 11 of the 12 target taxa, their levels were significantly correlated between the two methods (FDR-adjusted $p < 0.05$) (see [Fig. 2](#) and [Supplementary Tables 6 and 7](#); per-sample comparison is shown in [Supplementary Fig. S2](#)). The remaining taxon [*Ruminococcus*] produced insignificant correlation ($r = 0.03$, FDR-adjusted $p = 0.8414$).

[Fig. 3](#) shows the percentage of variance in community structure according to NGS explained by the level of each PCR target; totally all targets explain 36.5% of it.

4. Discussion

Considerable conformance of our qPCR platform with the results of high-throughput sequencing suggests that the panel is a prototype applicable for profiling major components of the human gut microbiome. The analyzed sample types can range from stool samples - as the non-invasive method of choice in such studies - to swabs and biopsies as well as contents of human intestinal tract simulators. Large-scale datasets obtained via the standard described protocol and using the same DNA extraction protocol can be subsequently subject to statistical analysis like the ones used for NGS microbiome data to conduct exploratory

Table 1
List of the 12 target taxa in the platform.

Taxon	Associations	References
<i>Akkermansia</i>	Inversely associated with metabolic signatures of obesity including insulin resistance and fasting glucose level. Suggested marker for calorie restriction diet efficiency prediction.	[7,25–28].
<i>Bacteroides</i>	One of the dominant genera in adult gut microbiome; linked to “Western diet”. Can change in abundance after high-fiber dietary interventions.	[1,29,30]
<i>Bifidobacterium</i>	Dominates in the gut community of healthy infants. Inversely associated with obesity and Crohn's disease. Abundant in certain Eurasian populations. Contains many probiotic strains. Can increase in abundance after intervention with probiotics.	[31] [2,32–34]
<i>Blautia</i>	One of the dominant genera in adult gut microbiome. Linked to type 2 diabetes.	[4,33]
<i>Christensenellaceae</i>	Inversely associated with obesity and IBD and directly - to Parkinson's disease; members of the family can promote reduction of excessive weight.	[32,35,36]
<i>Enterobacteriaceae</i>	Positively associated with metabolic disorders and IBD.	[6,32,37,38]
<i>Enterococcus</i>	Linked to inflammation, can carry a wide range of virulence factors and antibiotic resistance genes. Its intestinal domination increases the risk of bacteremia in immunocompromised patients. Prevalent in infant microbiome.	[39,40]; [41].
<i>Lactobacillus</i>	Inversely associated with type 2 diabetes, ulcerative colitis. Contains many probiotic strains.	[25,42]
<i>Methanobrevibacter</i>	Linked to eating behavior, anorexia nervosa, body mass index, obesity, IBD.	[43–45]
<i>Prevotella</i>	One of the dominant genera in adult gut microbiome. Increased in rural populations of the world. Linked to diet high in fiber and low levels in animal protein and fat. Maternal carriage during pregnancy is reversely linked to food allergy in the offspring.	[46–48]
[<i>Ruminococcus</i>] (from <i>Lachnospiraceae</i> family)	High abundance is linked to IBD.	[49,50]
<i>Streptococcus</i>	High abundance is associated with liver cirrhosis, as well as to the increased risk of bacteremia in immunocompromised patients.	[51,52]. [53]

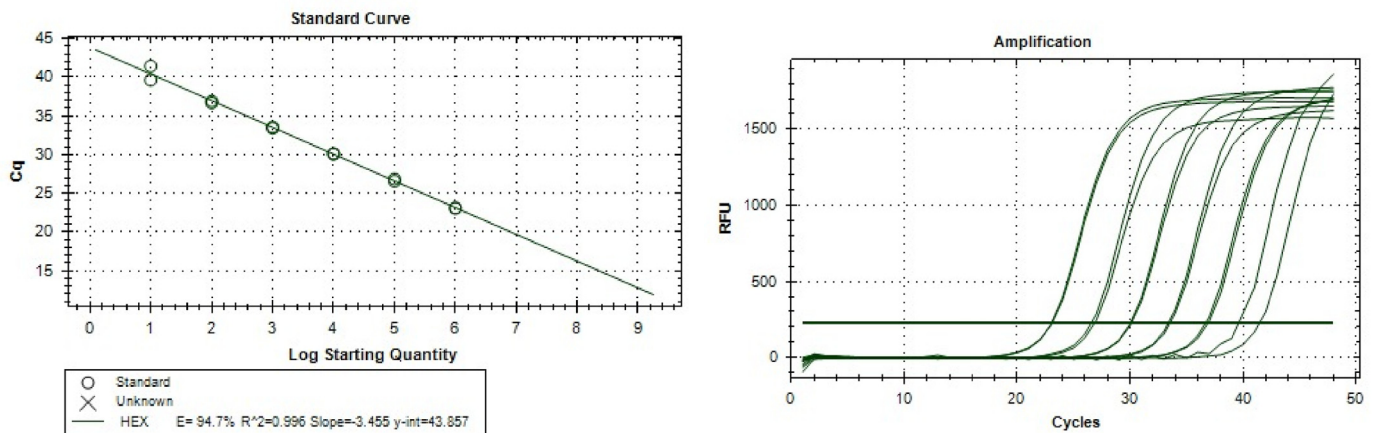


Fig. 1. Example of amplification results (for *Blautia* target).

analysis and reveal associations with factors like diseases, diet and lifestyle [20].

The platform was designed in a way to reach equilibrium between the coverage of the human gut microbiome diversity and cost per analysis. Subsequently, as the gut microbiome is highly complex, many prevalent taxa are not currently included among the targets. According to our comparison with NGS results, only a third of the total composition variance is explained by the platform. Our results show that the other families and genera that are abundant and variable across the test samples according to 16S rRNA sequencing but not targeted by the PCR test include *Lachnospiraceae* spp. ($9.2 \pm 4.3\%$), *Faecalibacterium* ($5.9 \pm 5.0\%$) and *Ruminococcus* ($3.4 \pm 3.0\%$). These along with their members are potential candidates to be included in the prospective extended panel - along with various members of the Firmicutes phylum including *Dorea*, *Roseburia* and others. However, the taxonomy within the phylum especially its Clostridiales order is rather ambiguous hence further elaboration of the primers and probes will be required.

One of the methods for microbiome profiling that can be viewed as an alternative to the proposed panel is Nanopore sequencing. Its advantages includes the possibility to obtain the total microbial composition in an untargeted way. However, although the analysis time (~12 h) takes shorter than for 16S rRNA on Illumina platform, compared to qPCR it takes longer (qPCR: 2–3 h), is less cost-efficient and

scalable and requires more expensive equipment [55].

For some of the taxa, we observed a reduced correlation of qPCR levels with the 16S rRNA sequencing results. At least for some of the taxa they could be due to a lower dynamic range provided by NGS (at given coverage) - so that the values below 1 read could be detected only by qPCR. A low correlation for [*Ruminococcus*] could be due to above-described complexity of Clostridiales phylogeny and need to improve the alignment of taxonomic annotations with the 16S rRNA database.

While our method is cultivation-independent and allows to capture a variety of taxa undetected during a classic stool test, its taxonomic detection accuracy is limited by the level of genus. A single genus can include species and strains with a highly variable gene content resulting in dramatically different phenotypes - from probiotic to human pathogen like in the case of *Escherichia coli* [6] - or correlating to opposing dietary patterns as shown for *Prevotella copri* [56]. Despite this limitation, a general-level profile of the gut microbiome composition provided by the proposed platform can be useful for assessing the general degree of dysbiosis. It can help in more advanced tasks like predicting response to drug therapy (as suggested by the discovered link between cancer immunotherapy response and *Bacteroides* levels [57]) or diet (for example, high-fiber diet [1]), measuring contribution of gut microbes to microbiome of built environments [58] as well as rapidly assessing the gut microbiome dysbiosis in patients with viral infections like COVID-

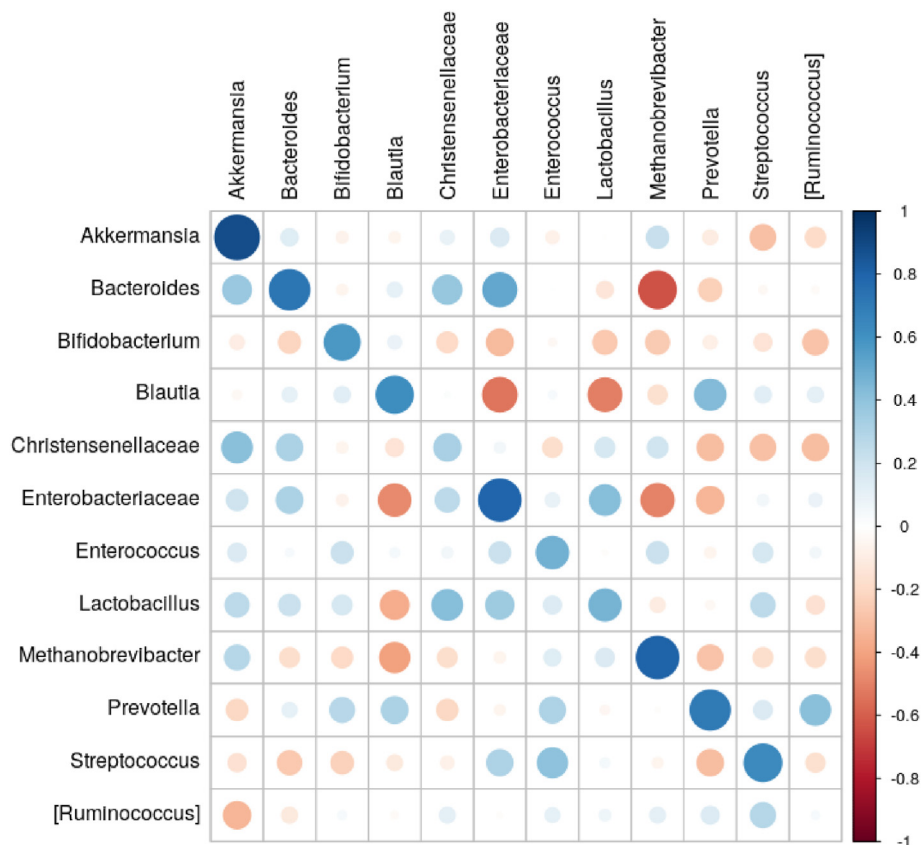


Fig. 2. Correlation between levels of target taxa evaluated using qPCR values and 16S rRNA gene sequencing. Rows - 16S rRNA, columns - PCR data; color shows Spearman's correlation coefficient.

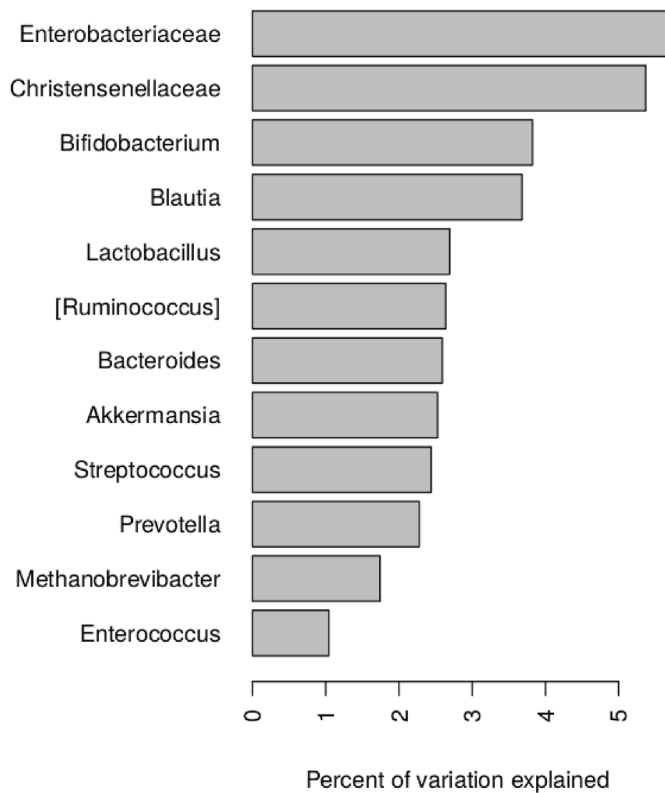


Fig. 3. Percentage of variance of NGS-derived community structure explained by target qPCR taxa (assessed using PERMANOVA).

19 (sometimes characterized by low presence of *Lactobacillus* and *Bifidobacterium* [59]).

One of the problems of the PCR approach is related to normalization. To address this issue, the set could be extended by adding a universal pair of primers (and a probe) targeting total prokaryotic content that can be used for normalization purposes (for example, by dividing signals from other taxa by it). Although this would reflect the microbial concentration in the analyzed DNA sample, this concentration could not directly correspond to the concentration in the subject's stool, as it can change considerably during the extraction. One of the suggested ways to approach absolute normalization is to collect an exactly equal volume of stool each time. Still, our platform allows us to assess the pairwise ratios between the targeted taxa. Consideration of ratios in the context of microbiome data is especially relevant as for many techniques - including metagenomics - they are often inherently compositional implying use of statistical approaches based on ratios, for instance, the balances [60]. The ratios and their derivatives can further be used as microbial community features to explore the associations with clinical status, diet and other factors. Finally, the proposed platform can be used for prioritization purposes: basing on its results for a specific subject, one can opt for performing additional in-depth analysis using other qPCR platforms targeting microbial genes (like virulence factors and antibiotic resistance determinants) or pathogenic taxa.

5. Conclusions

The developed qPCR platform is a perspective technique for rapid evaluation of human gut microbiome composition. Positive correlation with the data obtained using a different method demonstrates that the qPCR method can be applied for larger cohorts during the clinical and observational studies related to diet and diseases. Further larger clinical trials are needed to identify the method specificity and accuracy when

it comes to individual microbiome analysis.

Funding

This work was financially supported by the Skolkovo Foundation for Development of the Center for Elaboration and Commercialization of New Technologies (#G94/16 to Knomics LLC).

Authors' contributions

AD conceived the project. AD, TA and PA managed the project. KI, FM and KS performed the experiments. KN, CL, FM, KI, and KS performed the data analysis. KI, KN, TA, PA and CL wrote the manuscript. All authors read and approved the final manuscript.

CRedit authorship contribution statement

Irina Kurina: Writing - original draft, Visualization, Writing - review & editing, Formal analysis, Investigation, Validation. **Anna Popenko:** Methodology, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration. **Natalia Klimentko:** Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Stanislav Koshechkin:** Validation, Formal analysis, Investigation, Resources, Data curation, Project administration, Writing - review & editing. **Liubov Chuprikova:** Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Data curation, Visualization. **Maxim Filipenko:** Validation, Resources, Data curation, Visualization, Formal analysis, Investigation. **Alexander Tyakht:** Software, Validation, Data curation, Visualization, Methodology, Writing - original draft, Writing - review & editing, Project administration, Supervision. **Dmitry Alexeev:** Project administration, Funding acquisition, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank Maria Kardakova for performing language editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mcp.2020.101570>.

References

- [1] Natalia S. Klimentko, Alexander V. Tyakht, Anna S. Popenko, Anatoly S. Vasiliev, Ilya A. Altukhov, Dmitry S. Ischenko, Tatiana I. Shashkova, et al., Microbiome responses to an uncontrolled short-term diet intervention in the frame of the citizen science project, *Nutrients* 10 (5) (2018), <https://doi.org/10.3390/nu10050576>.
- [2] Olesya Volokh, Natalia Klimentko, Yulia Berezhnaya, Tyakht Alexander, Polina Nesterova, Popenko Anna, Dmitry Alexeev, Human gut microbiome response induced by fermented dairy product intake in healthy volunteers, *Nutrients* 11 (3) (2019), <https://doi.org/10.3390/nu11030547>.
- [3] Evgenii I. Olekhovich, Alexander I. Manolov, Andrey E. Samoilov, Nikita A. Prianichnikov, Maja V. Malakhova, Alexander V. Tyakht, Alexander V. Pavlenko, et al., Shifts in the human gut microbiota structure caused by quadruple eradication therapy, *Front. Microbiol.* 10 (August) (2019) 1902.
- [4] Lilit Egshatyan, Daria Kashtanova, Popenko Anna, Olga Tkacheva, Tyakht Alexander, Dmitry Alexeev, Natalia Karamnova, et al., Gut microbiota and diet in patients with different glucose tolerance, *Endocrine Connect.* 5 (1) (2016) 1–9.
- [5] Jakob Wirbel, Paul Theodor Pyl, Ece Kartal, Konrad Zych, Alireza Kashani, Alessio Milanese, Jonas S. Fleck, et al., Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer, *Nat. Med.* 25 (4) (2019) 679–689.
- [6] Alexander V. Tyakht, Alexander I. Manolov, Alexandra V. Kanygina, Dmitry S. Ischenko, Boris A. Kovarsky, Anna S. Popenko, Alexander V. Pavlenko, et al., “Genetic diversity of *Escherichia coli* in gut microbiota of patients with Crohn's disease discovered using metagenomic and genomic analyses, *BMC Genom.* 19 (1) (2018) 968.
- [7] Maria Carlota Dao, Amandine Everard, Judith Aron-Wisniewsky, Nataliya Sokolovska, Edi Prifti, Eric O. Verger, Brandon D. Kayser, et al., Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology, *Gut* 65 (3) (2016) 426–436.
- [8] Fredrik H. Karlsson, FåK Frida, Intawat Nookaew, Valentina Tremaroli, Björn Fagerberg, Dina Petranovic, Fredrik Bäckhed, Jens Nielsen, Symptomatic atherosclerosis is associated with an altered gut metagenome, *Nat. Commun.* 3 (2012) 1245.
- [9] Kathleen Machiels, Marie Joossens, João Sabino, Vicky De Preter, Ingrid Arijis, Venessa Eeckhout, Vera Ballet, et al., A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium prausnitzii* defines dysbiosis in patients with ulcerative colitis, *Gut* 63 (8) (2014) 1275–1283.
- [10] Serena Sanna, Natalie R. van Zuydam, Anubha Mahajan, Kurilshikov Alexander, Arnau Vich Vila, Urmo Vösa, Zlatan Mujagic, et al., Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases, *Nat. Genet.* 51 (4) (2019) 600–605.
- [11] Jean-Christophe Lagier, Dubourg Grégory, Matthieu Million, Frédéric Cadoret, Melhem Bilen, Florence Fenollar, Levasseur Anthony, Jean-Marc Rolain, Pierre-Edouard Fournier, Didier Raoult, Culturing the human microbiota and culturomics, *Nat. Rev. Microbiol.* 16 (May) (2018) 540–550.
- [12] James B. Thissen, Nicholas A. Be, Kevin McLoughlin, Shea Gardner, Paul G. Rack, Michael H. Shaper, Raymond R.R. Rowland, Slezak Tom, Crystal J. Jaing, Axiom microbiome Array, the next generation microarray for high-throughput pathogen and microbiome analysis, *PLoS One* 14 (2) (2019) e0212045.
- [13] El Kaoutari, Abdessamad, Fabrice Armougom, Quentin Leroy, Bernard Valiettes, Matthieu Million, Didier Raoult, Henrissat Bernard, Development and validation of a microarray for the investigation of the CAZymes encoded by the human gut microbiome, *PLoS One* 8 (12) (2013) e84033.
- [14] Rajilić-Stojanović, Hans G.H. Mirjana, Douwe Molenaar, Kajsa Kajander, Surakka Anu, Hauke Smidt, M. Willem, de Vos, Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults, *Environ. Microbiol.* (2009), <https://doi.org/10.1111/j.1462-2920.2009.01900.x>.
- [15] Martin Iain Bahl, Bergström Anders, Tine Rask Licht, Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis, *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* 329 (2) (2012) 193–197.
- [16] Anders Bergström, Tine R. Licht, Andrea Wilcks, Jens B. Andersen, Line R. Schmidt, Hugo A. Grönlund, Louise K. Vigsnaes, Kim F. Michaelsen, Martin I. Bahl, Introducing GUt low-density Array (GULDA) - a validated approach for qPCR-based intestinal microbial community analysis, *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* (2012), <https://doi.org/10.1111/1574-6968.12004>.
- [17] Koen Venema, Annet J.H. Maathuis, A PCR-based method for identification of bifidobacteria from the human alimentary tract at the species level, *FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Lett.* 224 (1) (2003) 143–149.
- [18] Jarmo Ritari, Jarkko Salojärvi, Leo Lahti, M. Willem, de Vos, Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database, *BMC Genom.* 16 (December) (2015) 1056.
- [19] James R. Cole, Qiong Wang, Jordan A. Fish, Benli Chai, Donna M. McGarrell, Yanni Sun, C. Titus Brown, Andrea Porras-Alfaro, Cheryl R. Kuske, James M. Tiedje, Ribosomal database project: data and tools for high throughput rRNA analysis, *Nucleic Acids Res.* 42 (Database issue) (2014) D633–D642.
- [20] Daria Efimova, Tyakht Alexander, Popenko Anna, Anatoly Vasilyev, Ilya Altukhov, Nikita Dovidchenko, Odintsova Vera, et al., Knomics-biota - a system for exploratory analysis of human gut microbiota data, *BioData Min.* 11 (November) (2018) 25.
- [21] Benjamin J. Callahan, Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, Susan P. Holmes, DADA2: high-resolution sample inference from Illumina amplicon data, *Nat. Methods* 13 (7) (2016) 581–583.
- [22] Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al., Author correction: reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, *Nat. Biotechnol.* 37 (9) (2019) 1091.
- [23] Bernd W. Brandt, Marc J. Bonder, Susan M. Huse, Egija Zaura, TaxMan: a server to trim rRNA reference databases and inspect taxonomic coverage, *Nucleic Acids Res.* 40 (2012) Web Server issue: W82–87.
- [24] Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, Weizhong Li, CD-HIT: accelerated for clustering the next-generation sequencing data, *Bioinformatics* 28 (23) (2012) 3150–3152.
- [25] Kristoffer Forslund, Falk Hildebrand, Trine Nielsen, Gwen Falony, Emmanuelle Le Chatelier, Shinichi Sunagawa, Edi Prifti, et al., Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota, *Nature* 528 (7581) (2015) 262–266.
- [26] Amandine Everard, Clara Belzer, Lucie Geurts, Janneke P. Ouwerkerk, Céline Druart, Laure B. Bindels, Yves Guiot, et al., Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity, *Proc. Natl. Acad. Sci. U.S.A.* 110 (22) (2013) 9066–9071.
- [27] I.G. Macchione, L.R. Lopetusso, G. Ianiro, M. Napoli, G. Gibiino, G. Rizzatti, V. Petito, A. Gasbarrini, F. Scalfarri, Akkermansia muciniphila: key player in metabolic and gastrointestinal disorders, *Eur. Rev. Med. Pharmacol. Sci.* 23 (18)

- (2019) 8075–8083.
- [28] L.K. Brahe, E. Le Chatelier, E. Prifti, N. Pons, S. Kennedy, T. Hansen, O. Pedersen, A. Astrup, S.D. Ehrlich, L.H. Larsen, Specific gut microbiota features and metabolic markers in postmenopausal women with obesity, *Nutr. Diabetes* 5 (June) (2015) e159.
- [29] Lawrence A. David, Corinne F. Maurice, Rachel N. Carmody, David B. Gootenberg, Julie E. Button, Benjamin E. Wolfe, Alisha V. Ling, et al., Diet rapidly and reproducibly alters the human gut microbiome, *Nature* 505 (7484) (2014) 559–563.
- [30] Gary D. Wu, Jun Chen, Christian Hoffmann, Kyle Bittinger, Ying-Yu Chen, Sue A. Keilbaugh, Meenakshi Bewtra, et al., Linking long-term dietary patterns with gut microbial enterotypes, *Science* 334 (6052) (2011) 105–108.
- [31] Fredrik Bäckhed, Josefine Roswall, Yangqing Peng, Qiang Feng, Huijue Jia, Petia Kovatcheva-Datchary, Li Yin, et al., Dynamics and stabilization of the human gut microbiome during the first year of life, *Cell Host Microbe* 17 (6) (2015) 852.
- [32] Floris Imhann, Arnau Vich Vila, Marc Jan Bonder, Jingyuan Fu, Dirk Gevers, Marijn C. Visschedijk, Lieke M. Spekhorst, et al., Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease, *Gut* 67 (1) (2018) 108–119.
- [33] Alexander V. Tyakht, Elena S. Kostryukova, Anna S. Popenko, Maxim S. Belenikin, Alexander V. Pavlenko, Andrey K. Larin, Irina Y. Karpova, et al., Human gut microbiota community structures in urban and rural populations in Russia, *Nat. Commun.* 4 (2013) 2469.
- [34] D.A. Russell, R.P. Ross, G.F. Fitzgerald, C. Stanton, Metabolic activities and probiotic potential of bifidobacteria, *Int. J. Food Microbiol.* 149 (1) (2011) 88–105.
- [35] Julia K. Goodrich, Jillian L. Waters, Angela C. Poole, Jessica L. Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, et al., Human genetics shape the gut microbiome, *Cell* 159 (4) (2014) 789–799.
- [36] V.A. Petrov, I.V. Saltykova, I.A. Zhukova, V.M. Alifirova, N.G. Zhukova, Yu B. Dorofeeva, A.V. Tyakht, et al., “Analysis of gut microbiota in patients with Parkinson's disease, *Bull. Exp. Biol. Med.* 162 (6) (2017) 734–737.
- [37] Na Fei, Liping Zhao, An opportunistic pathogen isolated from the gut of an obese human causes obesity in germfree mice, *ISME J.* 7 (4) (2013) 880–884.
- [38] Claudia Lupp, Marilyn L. Robertson, Mark E. Wickham, Inna Sekirov, Olivia L. Champion, Erin C. Gaynor, B. Brett Finlay, Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of enterobacteriaceae, *Cell Host Microbe* 2 (3) (2007) 204.
- [39] Kelli L. Palmer, Paul Godfrey, Allison Griggs, Veronica N. Kos, Jeremy Zucker, Christopher Desjardins, Gustavo Cerqueira, et al., Comparative genomics of enterococci: variation in *Enterococcus faecalis*, clade structure in *E. faecium*, and defining characteristics of *E. Gallinarum* and *E. Casseliflavus*, *mBio* (2012), <https://doi.org/10.1128/mbio.00318-11>.
- [40] Krista A. Dubin, Deepthi Mathur, Peter T. McKenney, Bradford P. Taylor, Eric R. Littmann, Jonathan U. Peled, Marcel R.M. van den Brink, Taur Ying, Eric G. Pamer, Joao B. Xavier, Diversification and evolution of vancomycin-resistant *Enterococcus faecium* during intestinal domination, *Infect. Immun.* 87 (7) (2019), <https://doi.org/10.1128/IAI.00102-19>.
- [41] S. Fanaro, R. Chierici, P. Guerrini, V. Vigi, Intestinal microflora in early infancy: composition and development, *Acta Paediatr.* (2007), <https://doi.org/10.1111/j.1651-2227.2003.tb00646.x>.
- [42] R.M. Jones, The use of *Lactobacillus casei* and *Lactobacillus paracasei* in clinical trials for the improvement of human health, *Microbiota Gastrointest. Pathophysiol.* (2017), <https://doi.org/10.1016/b978-0-12-804024-9.00009-4>.
- [43] M. Million, E. Angelakis, M. Maraninchi, M. Henry, R. Giorgi, R. Valero, B. Viallettes, D. Raoult, Correlation between body mass index and gut concentrations of *Lactobacillus reuteri*, *Bifidobacterium animalis*, *methanobrevibacter smithii* and *Escherichia coli*, *Int. J. Obes.* 37 (11) (2013) 1460–1466.
- [44] Schwensen, Ferlov Hanna, Carol Kan, Janet Treasure, Niels Høiby, Magnus Sjögren, A systematic review of studies on the faecal microbiota in anorexia nervosa: future Research may need to include microbiota from the small intestine, *Eat. Weight Disord.* EWD 23 (4) (2018) 399–418.
- [45] Catherine A. Mbakwa, John Penders, Paul H. Savelkoul, Carel Thijs, Pieter C. Dagnelie, Monique Mommers, Ilja C.W. Arts, Gut colonization with *methanobrevibacter smithii* is associated with childhood weight development, *Obesity* 23 (12) (2015) 2508–2516.
- [46] De Filippo, Carlotta, Duccio Cavalieri, Monica Di Paola, Matteo Ramazzotti, Jean Baptiste Poullet, Sebastien Massart, Silvia Collini, Giuseppe Pieraccini, Paolo Lionetti, Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural africa, *Proc. Natl. Acad. Sci. U.S.A.* 107 (33) (2010) 14691–14696.
- [47] Stephanie L. Schnorr, Marco Candela, Simone Rampelli, Manuela Centanni, Clarissa Consolandi, Giulia Basaglia, Silvia Turroni, et al., Gut microbiome of the hadza hunter-gatherers, *Nat. Commun.* 5 (April) (2014) 3654.
- [48] Peter J. Vuillermin, O'Hely Martin, Fiona Collier, Katrina J. Allen, Mimi L.K. Tang, Leonard C. Harrison, John B. Carlin, et al., Maternal carriage of *Prevotella* during pregnancy associates with protection against food allergy in the offspring, *Nat. Commun.* 11 (1) (2020) 1452.
- [49] Louise E. Tailford, C. David Owen, John Walshaw, Emmanuelle H. Crost, Jemma Hardy-Goddard, Gwenaelle Le Gall, M. Willem, de Vos, Garry L. Taylor, Nathalie Juge, Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation, *Nat. Commun.* 6 (July) (2015) 7624.
- [50] Chin Wen Png, Sara K. Lindén, Kristen S. Gilshenan, Erwin G. Zoetendal, Chris S. McSweeney, Lindsay I. Sly, Michael A. McGuckin, Timothy H.J. Florin, Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria, *Am. J. Gastroenterol.* 105 (11) (2010) 2420–2428.
- [51] Nan Qin, Fengling Yang, Ang Li, Edi Prifti, Yanfei Chen, Li Shao, Jing Guo, et al., Alterations of the human gut microbiome in liver cirrhosis, *Nature* 513 (7516) (2014) 59–64.
- [52] Ying Taur, Joao B. Xavier, Lauren Lipuma, Carles Ubeda, Jenna Goldberg, Asia Gobourne, Yeon Joo Lee, et al., “Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation.” clinical infectious diseases, *Off. Publ. Infect. Dis. Soc. Am.* 55 (7) (2012) 905–914.
- [53] Veronika B. Dubinkina, Alexander V. Tyakht, Vera Y. Odintsova, Konstantin S. Yarygin, Boris A. Kovarsky, Alexander V. Pavlenko, Dmitry S. Ischenko, et al., Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease, *Microbiome* 5 (1) (2017) 141.
- [54] G. Lu, Vector NTI, a balanced all-in-one sequence analysis suite, *Briefings Bioinf.* (2004), <https://doi.org/10.1093/bib/5.4.378>.
- [55] Thidathip Wongsurawat, Mayumi Nakagawa, Atiq Omar, Hannah N. Coleman, Piroon Jenjaroenpun, James I. Allred, Angela Trammel, Pantakan Puengrang, David W. Ussery, Intawat Nookaew, An assessment of oxford Nanopore sequencing for human gut metagenome profiling: a pilot study of head and neck cancer patients, *J. Microbiol. Methods* 166 (November) (2019) 105739.
- [56] De Filippis, Francesca, Edoardo Pasolli, Adrian Tett, Sonia Tarallo, Alessio Naccarati, Maria De Angelis, Erasmo Neviani, et al., Distinct genetic and functional traits of human intestinal *Prevotella copri* strains are associated with different habitual diets, *Cell Host Microbe* 25 (3) (2019) 444–453 e3.
- [57] V. Gopalakrishnan, C.N. Spencer, L. Nezi, A. Reuben, M.C. Andrews, T.V. Karpinets, P.A. Prieto, et al., Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients, *Science* 359 (6371) (2018) 97–103.
- [58] Natalia S. Klimenko, Alexander V. Tyakht, Stepan V. Toshchakov, Margarita A. Shevchenko, Aleksei A. Korzhenkov, Ebrahim Afshinnekoo, Christopher E. Mason, Dmitry G. Alexeev, Co-occurrence patterns of bacteria within microbiome of moscow subway, *Comput. Struct. Biotechnol. J.* 18 (February) (2020) 314–322.
- [59] Kaijin Xu, Hongliu Cai, Yihong Shen, Qin Ni, Yu Chen, Shaohua Hu, Jianping Li, et al., Management of corona virus disease-19 (COVID-19): the Zhejiang experience, *J. Zhejiang Univ. Med. Sci.* 49 (1) (2020) 0.
- [60] J. Rivera-Pinto, J.J. Egozcue, V. Pawlowsky-Glahn, R. Paredes, M. Noguera-Julian, M.L. Calle, Balances: a New perspective for microbiome analysis, *mSystems* 3 (4) (2018), <https://doi.org/10.1128/mSystems.00053-18>.