ISSN 2304-3415, Russian Open Medical Journal
2022. Volume 11. Issue 4 (December). Article CID e0413
DOI: 10.15275/rusomj.2022.0413

Original article

Difficulties of Enterobacteriaceae genome annotation in deciphering gastrointestinal microbiome datasets obtained by 16S rRNA gene amplicon sequencing

Elizaveta S. Klimenko, Natalya L. Belkova, Anna V. Pogodina, Lubov V. Rychkova, Marina A. Darenskaya

Research Center for Family Health and Problems of Human Reproduction, Irkutsk, Russia

Received 21 February 2022, Revised 14 April 2022, Accepted 26 August 2022

Abstract: Sequencing of the 16S rRNA gene amplicon is the cornerstone of the method for studying diverse bacteria in complex microbial communities. However, its use is complicated by an error rate of 10–17% when annotating 16S rRNA gene sequences. In our study, we examined the degree of accuracy of the taxonomic database of Enterobacteriaceae, compiled using the SILVA 132 reference database and a previously obtained dataset, viz. the microbiome of the gastrointestinal tract in adolescents with normal body weight and obesity.
Material and Methods — In this study, previously obtained 16S rRNA gene amplicon sequencing data were used, and the deciphering was carried out using the QIIME2 2019.4 platform. Phylogenetic analysis was performed using MEGA X software.
Results — Phylogenetic analysis of this family based on the studied V3–V4 fragment was hampered by polyphyly among some genera, and for half of the variants of the amplicon sequences it was not possible to clarify their genus. Statistical analysis did not reveal significant differences between the samples.
Conclusion — Although the average values of bacterial genera in the studied groups intuitively differed from each other, statistical analysis did not reveal significant differences between the samples. However, it can be assumed that a more detailed study of taxonomic diversity, taking into account factors, such as enterotype, duration of breastfeeding and family history, may reveal differences in the frequency distribution.

Keywords: gut microbiome; taxonomic annotation; adolescents; opportunistic microorganisms; 16S amplicon sequencing.

Cite as Klimenko ES, Belkova NL, Pogodina AV, Rychkova LV, Darenskaya MA. Difficulties of Enterobacteriaceae genome annotation in deciphering gastrointestinal microbiome datasets obtained by 16S rRNA gene amplicon sequencing. *Russian Open Medical Journal* 2022; 11: e0413.

Correspondence to Elizaveta S. Klimenko. Address: 16 Timiryazev St., Irkutsk 664003, Russia. Phone: +79501033652. E-mail: klimenko.elizabet@gmail.com.

Introduction

The human microbiome is a complex ecosystem consisting of the genetic material of more than 1 trillion microorganisms living inside the host [1]. Advances in massively parallel DNA sequencing technologies led to a significant increase in knowledge about microorganisms found in the natural environment, food systems, and the human body. In particular, sequencing of the 16S rRNA gene amplicon is a key method for studying the diversity and phylogeny of bacteria. This approach allows the simultaneous identification of most bacteria in complex microbial communities. Although the analysis of 16S rRNA gene diversity presents significant prospects for the study of bacteria from various habitats, the problems of standardizing approaches to sample preparation, DNA sequencing, and data analysis for obtaining reliable information on the composition, structure, and diversity of bacterial populations remain [2].

In studies of the microbiota of most ecosystems or habitats, identification at the species or strain level increases the ecological and/or clinical significance of the results, compared with identification at the genus level. For example, identification at the species level is often critical for host-associated microbial communities, as these communities often include commensal and pathogenic species of the same genus. In addition, some bacterial taxa include species that are specific to several localities and inhabit strictly defined niches of a given environment [3]. High-throughput sequencing of near-full-length 16S rRNA gene fragments (e.g., PacBio single-molecule circular consensus sequences, real-time sequencing, and whole genome sequencing) was expected to improve detection accuracy to species and strain levels. However, due to the greater availability and lower cost of ribosomal amplicon metasequencing, molecular epidemiological studies of the bacterial microbiota of humans, other animals, plants, and the environment are currently conducted on a population scale (i.e., thousands of samples) [4].

Bacterial reference databases with wide phylogenetic diversity, such as SILVA, RDP, and Greengenes, play a key role in data analysis [5-7]. Nevertheless, the taxonomic annotation of 16S rRNA gene sequences is incorrect in 10-17% of cases [8]. SILVA and RDP are regularly updated and represent extensive and complete libraries of 16S rRNA gene sequences from all studied habitats. In contrast, the Greengenes database was last updated in 2013 [7]. Taxonomic assignment using a reference database for large arrays of sequences and pipelines of metagenomic data processing platforms is associated with a certain percentage of misidentifications. Thus, correction of the taxonomic position is required using phylogenetic analysis and sequences of type strains.

Table 1. General characteristics of study participants

Sample code	*Group*	*Gender*	*Age*	*BMI*	*BMI Z-score*	*Obesity grade*	*Obesity grade: numeric value*
D01	Control	Female	15	21.72	0.44	Control	0
D13	Control	Male	14	19.84	0.19	Control	0
D14	Control	Male	12	18.78	0.33	Control	0
D28	Control	Female	13	17.27	−0.81	Control	0
D29	Control	Male	16	17.0	0.31	Control	0
D03	Control	Female	14	19.2	−0.64	Control	0
D30	Control	Female	14	19.5	−0.51	Control	0
D31	Control	Male	15	19.52	−0.34	Control	0
D32	Control	Male	15	19.8	−0.23	Control	0
D33	Control	Female	17	21.16	−0.02	Control	0
D34	Control	Female	15	20.52	0.1	Control	0
D35	Control	Male	15	20.23	0.1	Control	0
D36	Control	Male	17	21.83	0.09	Control	0
D04	Control	Male	16	18.07	−1.12	Control	0
D41	Control	Male	13	17.02	−0.72	Control	0
D42	Control	Male	13	20.51	0.59	Control	0
D43	Control	Female	16	21.36	0.18	Control	0
D44	Control	Male	17	22.01	0.28	Control	0
D45	Control	Female	13	20.52	0.59	Control	0
D46	Control	Male	14	20.89	0.49	Control	0
D47	Control	Male	17	23.28	0.54	Control	0
D05	Control	Female	13	19.29	0.34	Control	0
D11	Obesity	Male	15	36.75	3.34	Severe	3
D12	Obesity	Female	15	37.36	3.16	Severe	3
D15	Obesity	Male	16	33.82	2.84	Moderate	2
D17	Obesity	Female	14	36.1	3.11	Severe	3
D18	Obesity	Female	17	37.97	3.26	Severe	3
D02	Obesity	Female	16	29.6	2.04	Moderate	1
D20	Obesity	Male	16	40.91	3.88	Severe	3
D21	Obesity	Female	15	30.1	2.37	Moderate	1
D22	Obesity	Male	15	26.3	2.36	Moderate	1
D23	Obesity	Female	17	27.9	1.73	Moderate	1
D25	Obesity	Female	15	34.7	2.9	Moderate	2
D26	Obesity	Female	12	31.12	2.89	Moderate	2
D27	Obesity	Male	14	32.7	2.96	Moderate	2
D39	Obesity	Male	13	29.14	2.56	Moderate	2
D40	Obesity	Male	12	31.31	3.03	Moderate	2
D06	Obesity	Male	16	28.82	2.03	Moderate	1
D07	Obesity	Male	15	35.54	3.13	Severe	3
D09	Obesity	Male	13	26.19	2.19	Moderate	1

Our previous studies described the frequency and structure of obesity-associated cardiometabolic risk factors in a cohort of children and adolescents [9], changes in the biochemical status of obese youths [10-12], and analyzed the intestinal microbiota [13, 14]. Species of *Bifidobacterium* are present in the gastrointestinal tract of a healthy person, and a change in the number and composition of their species is a sign of intestinal dysbiosis. Features of the gut microbiome associated with obesity include a decrease in *Bifidobacterium* counts and reduced phylotype diversity. *Bifidobacterium* species were also examined, for which accurate species identification could not be performed using V3-V4 variable regions or standards for amplicon sequencing [15]. In addition, gut microbiome dysbiosis in obese adolescents was associated with altered species spectrum of enteric bacteria. Members of this family include both normal intestinal microbes (*Escherichia coli*) and opportunistic pathogens, such as *Klebsiella spp*. In this regard, accurate species identification is essential for obtaining complete information about the composition of representatives of this family in the intestinal microbiocenosis [16]. The goal of our study was to evaluate the correctness of the taxonomic identification of enteric bacteria by means of using the SILVA 132 reference database. The present study was conducted on samples of patients with normal body weight and obesity, for whom the results of both bacteriological analysis [16] and amplicon sequencing data were available.

Material and Methods

When studying the educational preferences of students studying at the universities of the Stavropol Krai and practicing Islam, 1,500 questionnaires were evaluated, which was a sufficient reference sample.

Brief description of the research material

The study was approved by the Ethics Committee of the Research Center for Family Health and Problems of Human Reproduction (RC FHPHR), Protocol No. 6 of 21 December 2015. The RC FHPHR previously studied the intestinal microbiome of adolescents with normal body weight and obesity [13, 15]. The general characteristics of the patients are presented in *Table* 1. Laboratory studies were carried out using standard operating procedures (SOP), IHMS_SOP 03 V2 and IHMS_SOP 06 V2, developed in the course of implementing the project of the international consortium, *International Human Microbiome Standards*. Amplicon analysis of the V3-V4 variable regions in the 16S rRNA gene was performed at Novogene (China). Primary data were deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SRR11006336-SRR11006339, SRR11006343, and SRR11006351-SRR11006388 (PRJNA604466) [14]. The amplicon libraries were processed using the algorithms of the QIIME2 2019.4 platform [17].

Bioinformatics data processing, phylogenetic and statistical analysis

We used SILVA 132 reference database for taxonomic assignment. To elucidate the phylogeny of amplicon sequence variants (ASVs), identified as *Enterobacteriaceae* sequences, sequences of the complete 16S rRNA gene of type strains of all *Enterobacteriaceae* family species were used.

A total of 63 nucleotide sequences were included in the tree, identified by comparison with the SILVA 132 reference database as belonging to the family *Enterobacteriaceae*, along with typical bacterial strains of this family. Multiple alignment and phylogenetic tree construction were performed using MEGA X software [15]. DNA sequence alignment was originally performed using the MUSCLE algorithm with default settings. The alignment was then visually checked to correct obvious alignment errors and remove areas of questionable alignment. The maximum likelihood method was employed to construct the phylogenetic tree. Statistical support for phylogeny was implemented using bootstrap (1,000 iterations). Bootstrap values ≥85% were considered highly supported, values of 75-84% were classified as moderately supported, and values of 50-74% were categorized as poorly supported. Values <50% were not specified [19].

Results

Basic statistics for library analysis

Molecular genetic analysis performed 2,590,453 reads. The number of reads per sample ranged from 52,945 to 77,290. A total of 2,890 phylotypes (ASVs) were identified, and the range per sample was 342-564. Depth of sequencing evaluation via the

Michaelis-Menten approximation showed that the composition of the microbiome at the ASV level was underestimated by an average of 2.04%.

General characteristics of the representation of Enterobacteriaceae

The *Enterobacteriaceae* content in the total microbiome ranged from 0.76 to 23.45%, and there were no significant differences between the control and obesity groups.

A total of 63 ASVs were assigned to the *Enterobacteriaceae* family. Sensu the taxonomy of the SILVA reference database, this family is represented by the genera *Citrobacter*, *Enterobacter*, *Klebsiella*, *Proteus*, *Raoultella*, *Serratia* and two undifferentiated groups (*Escherichia-Shigella* and *Hafnia-Obesumbacterium*) in the adolescent gut microbiome. In addition to these genera, ASVs have also been examined that could not be assigned to any genus (unidentified *Enterobacteriaceae*). Among all ASVs, only 65be was present in all samples, and this ASV was identified by SILVA as an *Escherichia-Shigella* phylotype (*Figure* 1). None of the ASVs exhibited significant differences in size between the obesity and control groups (Supplementary materials).

Taxonomy of Enterobacteriaceae

At the time of writing, this family was represented by 32 genera and 124 species. The following species had subspecies: *Enterobacter cloacae, Enterobacter hormaechei, Klebsiella pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola*, and *Salmonella enterica*. The December 2017 update of the SILVA reference database contains the genera *Proteus, Serratia, Hafnia,* and *Obesumbacterium*, which are not currently included in *Enterobacteriaceae* according to LPSN (https://lpsn.dsmz.de/). *Proteus* was moved to *Morganellaceae* [17], *Serratia* to *Yersiniaceae*, *Hafnia* and *Obesumbacterium* to *Hafniaceae*, and *Pantoea* to *Erwiniaceae* family.

Phylogenetic analysis revealed that the genera belonging to *Enterobacteriaceae*, according to the studied fragments, were polyphyletic, and they formed mixed clades (*Figure* 2). Only the genera *Cedecea, Gibbsiella, Phytobacter, Pseudocitrobacter, Franconibacter, Mangrovibacter, Izhakiella, Rosenbergiella, Trabulsiella*, and some groups of *Citrobacter, Klebsiella*, and *Kosakonia* were monophyletic and formed clades with good statistical support. Twenty-five ASVs were identified to the generic level, whereas 11 ASVs were assigned to the *Escherichia-Shigella* cluster. The identification matched for 22 ASVs, whereas the reclassification affected 14 ASVs. Twenty-seven ASVs remained identified only at the family level (*Enterobacteriaceae*).

Sequences identified as taxa that do not currently belong to *Enterobacteriaceae* formed separate clades (Nos. 1-4, *Figure* 2). The identification of ASV 5535 as a member of *Proteus* was not confirmed. Clade #1 contains sequences identified as *Hafnia-Obesumbacterium* and unidentified (UI) *Enterobacteriaceae*. The remaining clades included both sequences characterized by SILVA as representatives of reclassified genera and re-identified by phylogeny. Clade #3 contained *Escherichia-Shigella* and *Pantoea* sequences, while clade #4 included *Escherichia-Shigella* and *Serratia* sequences. All of them featured medium to strong bootstrap support. Clade #2 contained sequences identified as *Enterobacter*, *Pantoea*, and UI *Enterobacteriaceae*, and branch nodes had weak bootstrap support. For these sequences, a search for the nearest homologs was performed using BLAST software (*Table* 2). SILVA identification was identical for 11 ASVs. Among the mismatched were representatives of the genera *Erwinia, Pantoea, Serratia* and *Yersinia*.

The reclassified taxa accounted for 0.009-3.6% of the total microbiome. After the taxonomy correction, the content of several genera in the gut microbiome was changed, including *Enterobacter*, *Klebsiella*, and the *Escherichia-Shigella* group. The frequency distribution of other taxa was sporadic (*Figure* 3). Analysis of the overall frequency of ASVs with the same generic identification did not reveal significant differences between obese and control groups in counts for any genus.

Table 2. Search for the closest homology using BLAST

ASV	*Homology (%)*	*Sequence accession No.*	*Identification*
Clade 1			
00a3	98.76	NR_116898	*Hafnia paralvei* ATCC 29927
3491	99.75	NR_116603/NR_112985	*Obesumbacterium proteus* NCIMB 8771/*Hafnia alvei* JCM 1666
4381	100	NR_025334/NR_112985	*Obesumbacterium proteus* 42/*Hafnia alvei* JCM 1666
5bab	99.01	NR_119214/NR_104925	*Raoultella planticola* DSM 3069/*Ewingella americana* CIP 81.94
5d95	99.50	NR_116603/NR_112985	*Obesumbacterium proteus* NCIMB 8771/*Hafnia alvei* JCM 1666
692e	99.01	NR_044152	*Yersinia massiliensis* 50640
Clade 2			
34c5	98.26	NR_041970	*Erwinia amylovora* DSM 30165
768e	99.01	NR_025635	*Klebsiella variicola* F2R9
93d8	99.01	NR_041970	*Erwinia amylovora* DSM 30165
ac23	99.01	NR_148649	*Enterobacter bugandensis* 247BMC
c27b	98.51	NR_104724	*Erwinia aphidicola* X 001
Clade 3			
1e31	100	NR_116755	*Pantoea dispersa* LMG 2603
78e4	99.75	NR_118122	*Pantoea wallisii* LMG 26277
be6c	99.75	NR_116246	*Pantoea eucrina* LMG 2781
dd27	99.75	NR_116755	*Pantoea dispersa* LMG 2603
f963	100	NR_116114	*Pantoea deleyi* LMG 24200
Clade 4			
2614	100	NR_114043	*Serratia marcescens* NBRC 102204
a21f	99.75	NR_044385	*Serratia nematodiphila* DZ0503SBS1
ba84	99.75	NR_036886/NR_114043	*Serratia marcescens* subsp. *sakuensis* KRED/*Serratia marcescens* NBRC 102204
ec2b	100	NR_044385	*Serratia nematodiphila* DZ0503SBS1

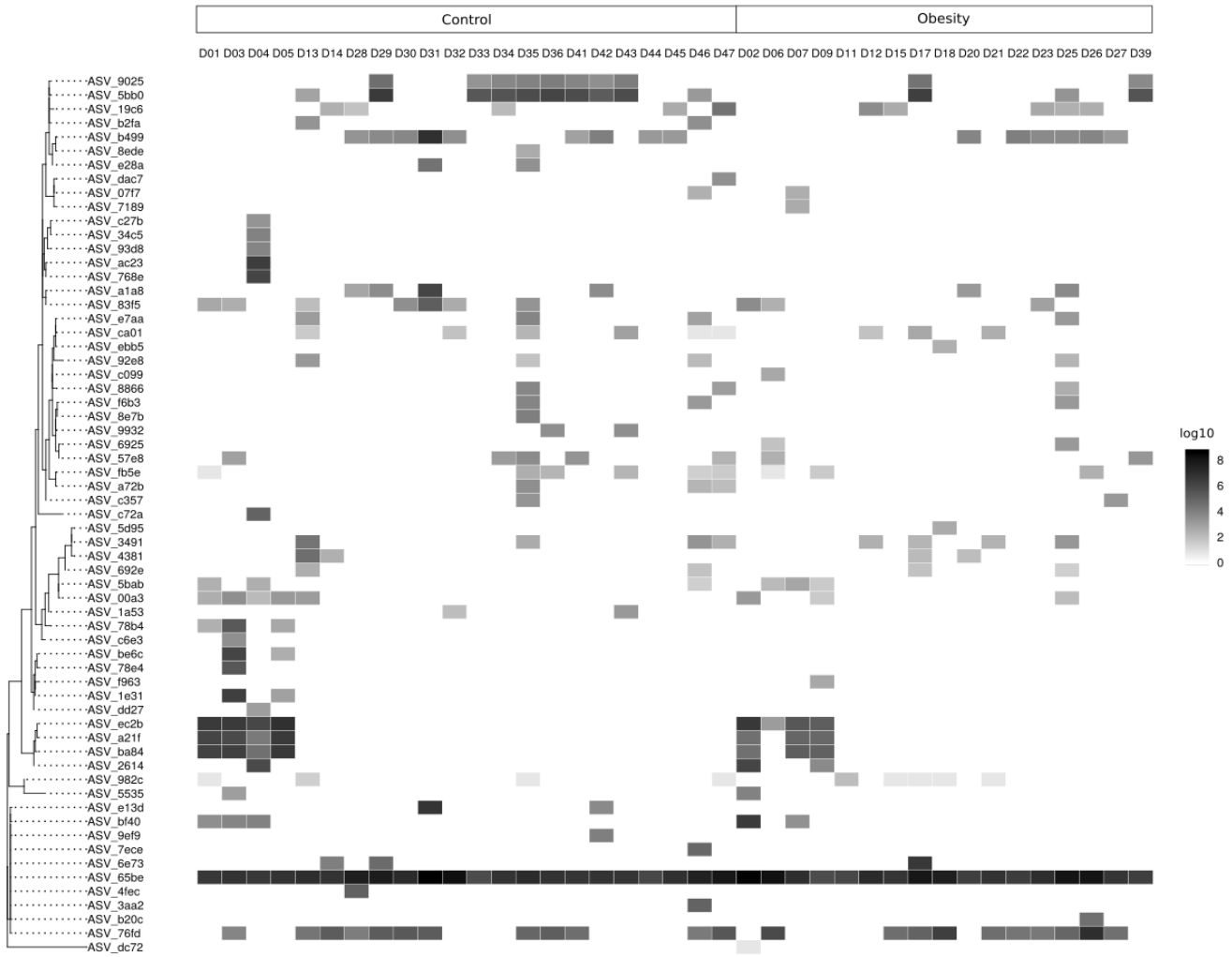

Figure 1. Frequency heatmap of isolated ASVs in gastrointestinal microbiomes in obesity and control groups of youths.

Hence, the phylogenetic analysis of this family for the studied V3-V4 fragment was complicated by the polyphyly of some genera. The genus of half of the ASVs could not be specified.

Discussion

Dysbiosis is mainly associated with an increased number of pathobionts, such as *Escherichia* and *Klebsiella* spp. caused by a reduction in the number of taxa with useful metabolic activity, including lactobacilli and bifidobacteria. Dysbiosis is also associated with a decrease in biodiversity, i.e., a reduction in the number of microbial species present in the microbiome and lower complexity of the microbial community [21]. As for gastrointestinal microbiota in obese and normal-weight children, the former category has fewer counts of *Bifidobacterium* and higher counts of *E. coli*. Studies have shown that a high number of bifidobacteria in infancy and adulthood protects against obesity [22]. It was also revealed that with a decrease in body weight in children achieved by modifying their diet, the numbers of *Bifidobacterium* and *Lactobacillus* increases, while the number of enterobacteria decreases [23]. In inflammatory bowel disease, there is an increase in proteobacteria, viz. intestinal bacteria, including the opportunistic pathogens *E. coli* and *K. pneumoniae*, which increases mucosal inflammation and the risk of infections. Many studies have described a decrease in the number of bifidobacteria and lactobacilli and an increase in the numbers of *Enterobacter* in patients with irritable bowel syndrome (IBS) and diarrhea. Other researchers linked IBS with *Campylobacter, Yersinia, Salmonella, Shigella* and *E. coli*. The heterogeneity of the results is explained by the variety of methods used to determine the microbiota and the different criteria for enrolling patients [21]. While there is controversy as to which types of bacteria are associated with being overweight, some specific genera and species of bacteria appear important.

It is known that even the complete sequence of the 16S rRNA gene has a low low discriminatory power. Branching of genera and species within this family during phylogenesis, based on the 16S rRNA gene, has a significant stochasticity depending on the used algorithms and analyzed bacteria [20]. According to the results of some studies, it can be said that the entire *Enterobacteriales* order is generally characterized by polyphyletic branching and the absence of connected monophyletic groups [20, 24]. The used fragment is not optimal for phylogenetic analysis of this family. Different parts of the genome may have distinct phylogenetic similarities to other taxa. In other words, a group can be monophyletic for some parts of the genome and simultaneously paraphyletic for other parts. In analytical results, this may reflect either analytical ambiguity or actual phylogenetic inconsistencies.

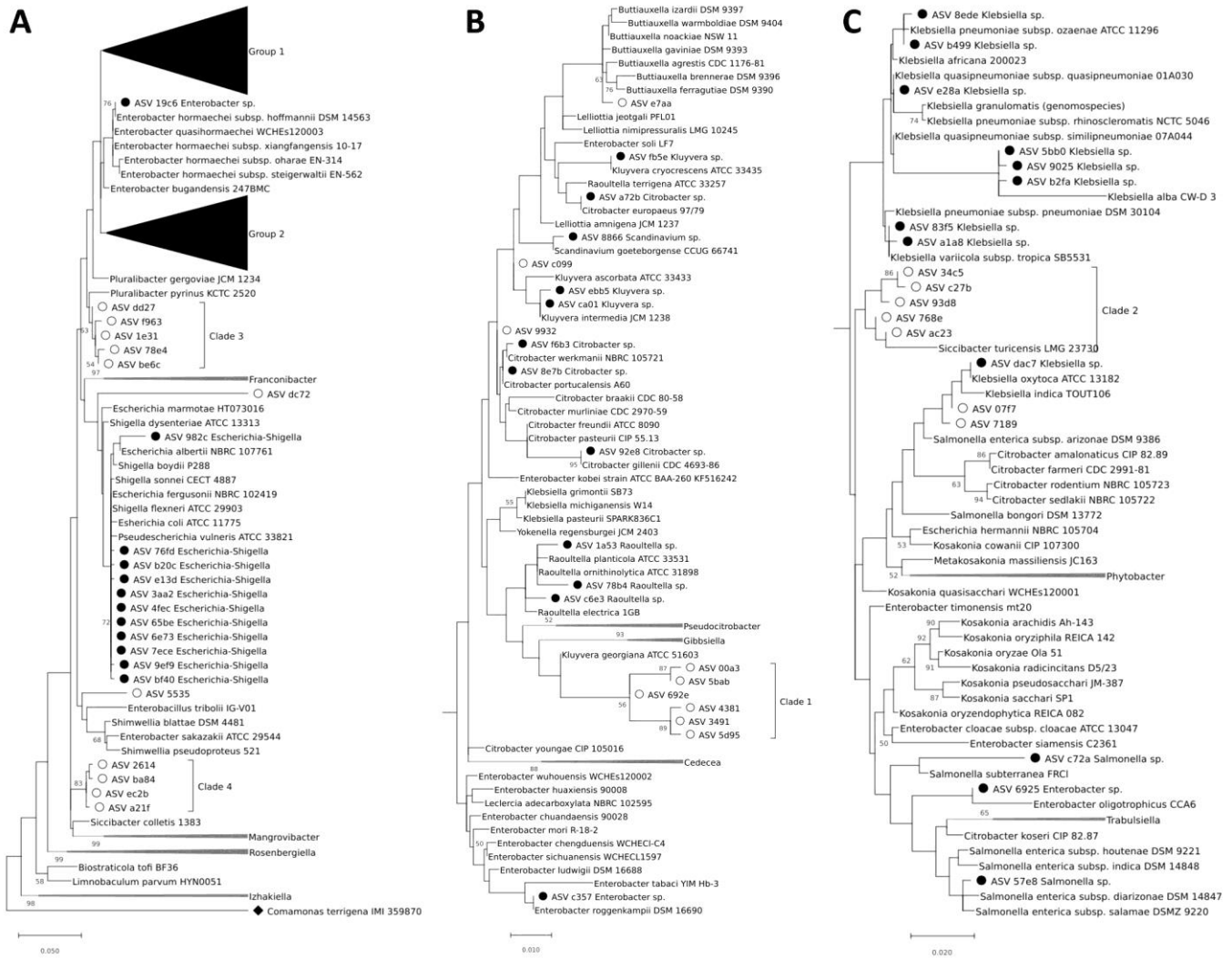

Figure 2. Phylogenetic tree of the studied ASVs and sequences of type strains of *Enterobacteriaceae*. A – The outer group is marked with a diamond-shaped marker. Grey clusters denote monophyletic genera that do not include ASVs. Black markers denote ASVs that were assigned to a specific genus or taxonomic group. White markers denote unidentified ASVs. The scale is five substitutions per 100 base pairs (bp). B – Extended group 1. Gray clusters denote monophyletic genera not including ASV. Black markers identify ASVs that were assigned to a specific genus or taxonomic group. White markers denote unidentified ASVs. The scale is one substitution per 100 bp. C – Extended group 2. Grey clusters denote monophyletic genera that did not include ASVs. Black markers denote ASVs that were assigned to a specific genus or taxonomic group. White markers denote unidentified ASVs. The scale is two substitutions per 100 bp.

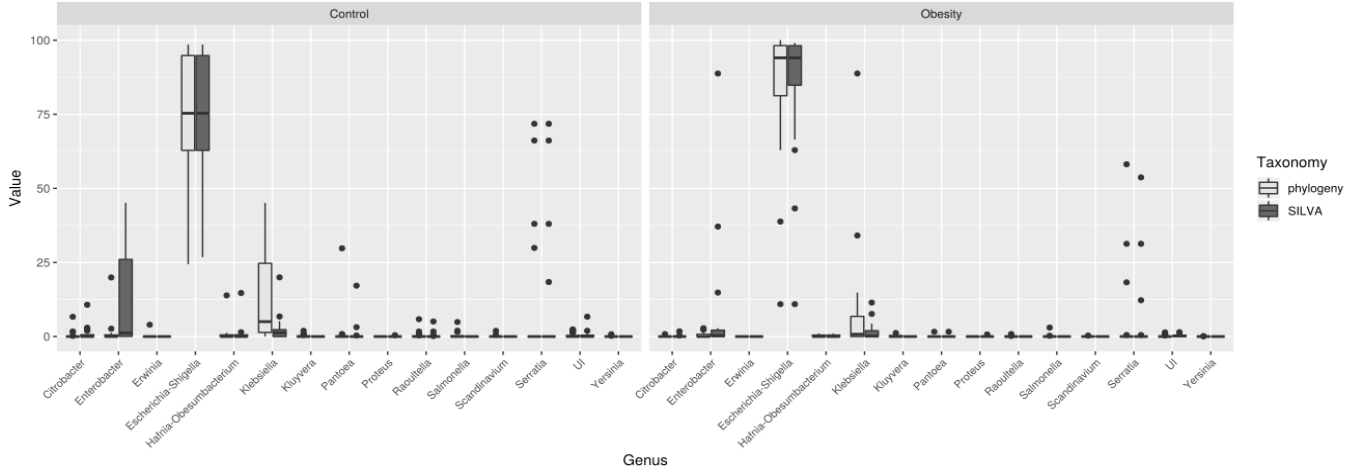

Figure 3. Frequency distribution (% of the total count in the family) of *Enterobacteriaceae* family representatives in control and obesity groups.

The composition of the microbial community depends on such factors as the time of breastfeeding, family history (health status of the mother and other family members), and the dominant component of the microbial community that determines the human enterotype. The health status of the mother during pregnancy (past inflammatory and infectious diseases) can affect early (intrauterine) colonization of the child's body with bacteria, such as *Enterobacter, Enterococcus, Lactobacillus, Photorhabdus* and *Tannerella* [25]. The main stage of colonization of the child's body by symbiotic bacteria occurs at the time of birth. The mode of delivery largely determines the future composition of the microbiome. The microbiome of children born by caesarean section is very different from the microbiome of children born by vaginal delivery. Breastfeeding is the second step in the colonization of the baby's intestines after birth. The method (artificial or natural) and the time of feeding strongly influence the composition of the intestinal microbiome, determining the dominant and minor components of the community. Feeding has the greatest impact on the diversity of representatives of the genus *Bifidobacterium* [25]. The dominant component of the community influences the minor components. There are several approaches to typing the intestinal microbial community according to the dominant component. The first approach involves the use of partitioning around medoids (PAM) and dividing them into three groups (*Bacteroides*, *Prevotella* and *Ruminococcus*). The second approach is based on the Dirichlet multinomial mixtures (DMM) and gives a division into 4 groups (*Ruminococcaceae* [R], *Prevotella* [P], *Bacteroides* 1 [B1] and *Bacteroides* 2 [B2]). In some cases, an additional enterotype is identified with a predominance of representatives of the *Enterobacteriaceae* (H) family – but, as a rule, it is associated with the presence of inflammatory diseases, alcohol dependence, or other ailments. Enterotypes do not depend on gender, age, ethnicity or geography. Rather, they depend on the characteristics of long-term nutrition [26].

Conclusion

Most taxa were characterized by the presence of a single sample; no dependence on division into groups was observed. However, it can be assumed that a more detailed study of taxonomic diversity, taking into account factors, such as enterotype, duration of breastfeeding and family history, may reveal differences in the frequency distribution. Future studies should also include the analysis of these samples using the whole genome sequencing technology due to such type of information retrieval analysis on a larger scale.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding

This study was conducted with the financial support by the Council for Presidential Grants of the Russian Federation (NSh- 3382.2022.1.4).

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

References

1. Stanford J, Charlton K, Stefoska-Needham A, Ibrahim R, Lambert K. The gut microbiota profile of adults with kidney disease and kidney stones: a systematic review of the literature. *BMC Nephrol* 2020; 21(1): 215. https://doi.org/10.1186/s12882-020-01805-w.
2. Xue Z, Kable ME, Marco ML. Impact of DNA sequencing and analysis methods on 16S rRNA gene bacterial community analysis of dairy products. *mSphere* 2018; 3(5): e00410- e00418. https://doi.org/10.1128/msphere.00410-18.
3. Mark Welch JL, Dewhirst FE, Borisy GG. Biogeography of the oral microbiome: The site-specialist hypothesis. *Annu Rev Microbiol* 2019; 73: 335-358. https://doi.org/10.1146/annurev-micro-090817-062503.
4. F Escapa I, Huang Y, Chen T, Lin M, Kokaras A, Dewhirst FE, Lemon KP. Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets. *Microbiome* 2020; 8(1): 65. https://doi.org/10.1186/s40168-020-00841-w.
5. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-Species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 2014; 42(Database issue): D643- D648. https://doi.org/10.1093/nar/gkt1209.
6. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014; 42(Database issue): D633-D642. https://doi.org/10.1093/nar/gkt1244.
7. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 2012; 6(3): 610-618. https://doi.org/10.1038/ismej.2011.139.
8. Edgar R. Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* 2018; 6: e5030. https://doi.org/10.7717/peerj.5030.
9. Pogodina A, Rychkova L, Kravtzova O, Klimkina J, Kosovtzeva A. Cardiometabolic risk factors and health-related quality of life in adolescents with obesity. *Child Obes* 2017; 13(6): 499-506. https://doi.org/10.1089/chi.2016.0330.
10. Darenskaya MA, Kolesnikova LI, Rychkova LV, Kravtsova OV, Semenova NV, Kolesnikov SI. Relationship between lipid metabolism state, lipid peroxidation and antioxidant defense system in girls with constitutional obesity. *AIMS Molecular Science* 2021; 8(2): 117-126. http://doi.org/10.3934/molsci.2021009.
11. Darenskaya MA, Gavrilova OA, Rychkova LV, Kravtsova OV, Grebenkina LA, Osipova EV, et al. The assessment of oxidative stress intensity in adolescents with obesity by the integral index. *International Journal of Biomedicine* 2018; 8 (1): 37-41. http://doi.org/10.21103/Article8(1)_OA5.
12. Kolesnikova LI, Darenskaya MA, Rychkova LV, Grebenkina LA, Semenova NV, Kolesnikov SI. Lipids metabolism and antioxidant status in exogenous constitutional obesity in girls of Buryatia. *Russian Bulletin of Perinatology and Pediatrics* 2021; 66(1): 80-86. Russian. http://doi.org/10.21508/1027-4065-2021-66-1-80-86.
13. Belkova NL, Nemchenko UM, Pogodina AV, Feranchuk SI, Romanitsa AI, Novikova EA, et al. Composition and Structure of Gut Microbiome in Adolescents with Obesity and Different Breastfeeding Duration. *Bull Exp Biol Med* 2019; 167(6): 759-762. http://doi.org/10.1007/s10517-019-04617-7.
14. Belkova N, Klimenko E, Romanitsa A, Pogodina A, Rychkova L. Metagenomic 16S rDNA amplicon datasets from adolescents with normal weight, obesity, and obesity with irritable bowel syndrome from Eastern Siberia, Russia. *Data in Brief* 2020; 32: 106141. http://doi.org/10.1016/j.dib.2020.106141.
15. Klimenko ES, Pogodina AV, Rychkova LV, Belkova NL. The ability of taxonomic identification of bifidobacteria based on the variable regions of 16S rRNA Gene. *Russ J Genet* 2020; 56(8): 926-934. http://doi.org/10.1134/S1022795420080074.
16. Grigorova EV, Belkova NL, Nemchenko UM, Klimenko ES, Pogodina AV, Romanitsa AI, et al. Metasequencing of V3-V4 variable regions of 16S

rRNA gene in opportunistic microbiota and gut biocenosis in obese adolescents. *Bull Exp Biol Med* 2021; 170(3): 321-325. http://doi.org/10.1007/s10517-021-05060-3.

17. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019; 37(8): 852-857. http://doi.org/10.1038/s41587-019-0209-9.
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018; 35(6): 1547-1549. https://doi.org/10.1093/molbev/msy096.
19. Erixon P, Svennblad B, Britton T, Oxelman B. Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst Biol* 2003; 52(5): 665-673. https://doi.org/10.1080/10635150390235485.
20. Adeolu M, Alnajar S, Naushad S, S Gupta R. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': Proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. *Int J Syst Evol Microbiol* 2016; 66(12): 5575-5599. https://doi.org/10.1099/ijsem.0.001485.
21. Nouvenne A, Ticinesi A, Tana C, Prati B, Catania P, Miraglia C, et al. Digestive disorders and Intestinal microbiota. *Acta Biomed* 2018; 89(9-S): 47-51. https://doi.org/10.23750/abm.v89i9-s.7912.
22. Pihl AF, Fonvig CE, Stjernholm T, Hansen T, Pedersen O, Holm JC. The role of the gut microbiota in childhood obesity. *Child Obes* 2016; 12(4): 292-299. https://doi.org/10.1089/chi.2015.0220.
23. Hou YP, He QQ, Ouyang HM, Peng HS, Wang Q, Li J et al. Human gut microbiota associated with obesity in Chinese children and adolescents. *Biomed Res Int* 2017; 2017: 7585989. https://doi.org/10.1155/2017/7585989.
24. Yarza P, Richter M, Peplies J, Uzeby J, Amann R, Schleifer KH, et al. The All-Species Living Tree Project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 2008; 31(4): 241-250. https://doi.org/10.1016/j.syapm.2008.07.001.
25. Yakushin AS, Ukraintsev SE, Denisov MYu. Intestinal Microbiota: Early Formation, Health Effects, and Correction Ways. *Current Pediatrics* 2017; 16(6): 487-492. Russian. https://doi.org/10.15690/vsp.v16i6.1821.
26. Takagi T, Inoue R, Oshima A, Sakazume H, Ogawa K, Tominaga T, et al. Typing of the gut microbiota community in Japanese subjects. *Microorganisms* 2022; 10(3): 664. https://doi.org/10.3390/microorganisms10030664.

Authors:
Elizaveta S. Klimenko – Junior Researcher, Department of Epidemiology and Microbiology, Research Center for Family Health and Human Reproductive Problems, Irkutsk, Russia. https://orcid.org/0000-0003-0979-8816.
Natalya L. Belkova – PhD, Associate Professor, Leading Researcher, Department of Epidemiology and Microbiology, Research Center for Family Health and Human Reproductive Problems, Irkutsk, Russia. https://orcid.org/0000-0001-9720-068X.
Anna V. Pogodina – MD, Lead Researcher, Department of Pediatrics, Research Center for Family Health and Human Reproductive Problems, Irkutsk, Russia. https://orcid.org/0000-0001-8533-3119.
Lubov V. Rychkova – MD, Corresponding Member of Russian Academy of Sciences, Chair of the Department of Pediatrics, Director of Research Center for Family Health and Human Reproductive Problems, Irkutsk, Russia. https://orcid.org/0000-0002-0117-2563.
Marina A. Darenskaya – PhD, Lead Researcher, Department of Personalized and Preventive Medicine, Research Center for Family Health and Human ReproductiveProblems, Irkutsk, Russia. https://orcid.org/0000-0003-3255-2013.