# MODERN METHODS OF STUDYING THE MICROBIOTA OF THE DIGESTIVE TRACT

# СОВРЕМЕННЫЕ МЕТОДЫ ИЗУЧЕНИЯ МИКРОБИОТЫ ПИЩЕВАРИТЕЛЬНОГО ТРАКТА

**УДК 663**

**DOI: 10.24411/2658-4964-2021-10365**

**Нерубенко Е.С.,** Студент, Балтийский федеральный университет имени Иммануила Канта, Россия, г.Калининград

**Nerubenko E. S.,** student, Immanuil Kant Baltic federal university, Kaliningrad, Russia, nerubenko.elena@bk.ru

**Аннотация:** микробиота человека в настоящее время рассматривается как самостоятельный орган. Изучение особенностей микробиома, характерных для пациентов, позволит по-новому взглянуть на этиологию и подходы к диагностике заболеваний. Современные методы исследования микробиоты являются как молекулярно-генетическими, такими как секвенирование геномов микроорганизмов или микробиомов, так и метагеномным анализом. Улучшения в системах секвенирования снизили стоимость процессов секвенирования, что привело к увеличению количества исследований, требующих определения последовательностей нуклеиновых кислот. Такие разработки направлены на увеличение количества технологических операций одновременно для одной реакции секвенирования - мультиплексирования образцов.

**Absract:** human microbiota is currently considered as an independent organ. The study of the microbiome features characteristic of patients will allow us to look in

a new way at the etiology and approaches to the diagnosis of diseases. Modern methods of microbiota research are both molecular genetic, such as sequencing of genomes of microorganisms or microbiomes, and metagenomic analysis. Improvements in sequencing systems reduced the cost of sequencing processes, leading to the increase of studies requiring the determination of nucleic acids. Such developments are aimed at increasing the number of technological operations simultaneously for one sequencing reaction - multiplexing of samples.

**Ключевые слова:** микробиота, пищеварительный тракт, метагеномные исследования, секвенирование.

**Keywords:** microbiota, gastrointestinal tract, metagenomic studies, sequencing.

## Introduction

The normal microflora of the human intestine performs in a number of important functions, providing a high level of body resistance to external influences, starting from birth and throughout a person's life. External factors, such as geography, dietary habits, affect the characteristics of the species composition of the intestinal microbiota [1].

Today, it is important to obtain complete information on the composition of the intestinal microbiota not only for different ethnic populations of people, but also for individual regions of one country. The intestinal flora consists of more than 100 trillion species of microorganisms [2].

The microbiota of the human digestive tract is studied on the basis of morphological, molecular genetic (the most common method for obtaining a profile of the sequences of the gene encoding 16S rRNA of microorganisms. Genome sequencing is also used.) And metabolic characteristics (the profile of metabolites of microorganisms can be obtained by mass spectrometry).

Determination of microorganisms by external characteristics, by morphology, is often impossible due to the fact that for many microorganisms the cultivation conditions are not determined, many microorganisms are able to reproduce only in the community [3]. Modern technologies for the determination of many nucleotide sequences (next generation sequencing), have made it possible to identify many

microorganisms that inhabit the human body and, in particular, the gastrointestinal tract, based on genome analysis. This allowed the accumulation of huge databases, the analysis of which requires the development of bioinformatics tools.

Methods based on amplification (PCR) of DNA are used to detect microorganisms, based on the genomic distribution of repeats of microsatellite sequences of the type (GC) 4, (GTG) 5, on the distribution of conservative regions of tRNA genes, on the distribution of sequences of mobile elements found in the genome of the organism under study. [4]. PCR analysis with primers for target genes (such genes can be housekeeping genes, as well as genes of pathogenicity, etc.) [4]. Methods based on fluorescent in situ hybridization are also used to detect microorganisms, which allows one to identify, locate, and count single microbial cells, as well as their clusters [5]. The method is based on the fact that cells hybridize with fluorescently labeled oligonucleotide probes complementary to specific regions of the gene [5]. However, using the FISH method, bacteria are not identified to species, but only their belonging to larger taxa is determined [5]. One of the modern methods used for the differentiation and identification of microorganisms is gas chromatography-mass spectrometry (GC / MS). It is based on a combination of two analytical methods: capillary gas chromatography and mass spectrometry [6]. The principle of the method is the qualitative and quantitative determination of marker substances (fatty acids (FA), aldehydes, alcohols, sterols, etc.) [6]. It was found that odd, branched and cyclopropane FAs, as well as fatty aldehydes, are found mainly in gram-positive bacteria, and higher fatty β-hydroxy acids are inherent only in gram-negative microorganisms [6].

A marker for identifying microorganisms in next generation sequencing is a gene encoding 16S ribosomal RNA and, more specifically, individual regions of this gene, which are characterized by both conserved and variable regions. For amplification in the polymerase chain reaction of this region, primers homologous to the conserved regions are used. Species specific areas are further used in bioinformatics analysis to identify species. The degree of similarity of species-specific regions very well reflects the evolutionary relationship of different species. Nucleotide sequences of 16S rRNA

of many known bacteria and archaea are available in databases as identified sequences of the studied microorganisms. Taxonomy based on the nucleotide sequence of the 16S rRNA gene makes a significant contribution to the identification of organisms; additions to it are the phenotypic classification based on the appearance of the colonies and the ability to stain according to Gram. 16S rRNA gene sequences for studying bacterial phylogeny and taxonomy is currently a common method for obtaining genetic markers. The 16S rRNA gene is used for a number of reasons: it is present in all bacteria, the function of the 16S rRNA gene has not changed over time, which suggests that random sequence changes are a measure of time (evolution); 16S rRNA gene is 1500 bp and it is long enough for identification [7]. Sequencing of the 16S rRNA gene provides genus identification in most cases (> 90%), which is less true for species (65–83%) [7]. 99% similarity (divergence <1%) of the sequences read may not be sufficient in all cases to guarantee accurate identification. Thus, in the case of Aeromonas veronii, the genome may contain up to six copies of the 16S rRNA gene, which may differ from each other by more than one percent. This implies the intragenomic heterogeneity of the 16S rRNA gene, which will not allow the species to be identified with an accuracy of 99% [7]. Determination of the entire sequence of the 16S rRNA gene would allow more reliable identification of microorganisms. Unfortunately, modern sequencing methods don’t allow massive sequencing of amplicons of the complete 16S rRNA gene. Thus, the question of choosing the most effective hypervariable regions for phylogenetic analysis and taxonomic classification is still under discussion [8]. Several bioinformatics tools have been integrated to create an in silico pipeline for assessing the phylogenetic sensitivity of hypervariable regions compared to corresponding full-length sequences [7]. The full-length sequences of the 16S rRNA gene consist of nine hypervariable regions separated by nine highly conserved regions. The relationship between different subregions showed that V4-V6 were the most reliable regions for representing full-length 16S rRNA sequences in phylogenetic analysis of most bacterial types, while V2 and V8 were the least reliable regions [7]. The results suggest that the V4-V6 subregions may be optimal for the development of universal primers with satisfactory phylogenetic resolution for the bacterial type. The sequence of the

16S rRNA gene was first used in 1985 for phylogenetic analysis. The 16S rRNA gene sequence has become the most widely used marker for profiling bacterial communities, since it contains both highly conserved and hypervariable regions. Therefore, the selection of suitable primers is critical for studying bacterial phylogeny in a variety of environments. Class I, which included V4, V5, and V6, has the highest sensitivity and was proposed to represent the optimal subregions for phylogenetic studies [7]. V3 and V7 class II and have shown moderate sensitivity. Class III, which was introduced by V2 and V8, was not used for phylogenetic resolution at the type level or for phylogenetic analysis of different communities, although class III may still be suitable for phylogenetic analysis and possibly for classification of microbes from the same classes or families [7]. Selected regions of 16S rRNA gene sequences are amplified using PCR and specially selected primers, which may contain several technical sequences simultaneously - barcodes and adapters. In this case, the amplification of the desired region of the sequence, the assignment of individual identifiers to the samples, and the attachment of adapters necessary for the subsequent sequencing of the obtained libraries of the 16S rRNA gene on the selected sequencing platform take place simultaneously. It is possible to obtain libraries in several stages: first, amplification of the region, and then attaching barcodes and adapters. The sequencing results are subjected to bioinformatics processing. First, the obtained sequences are combined into variants of sequences of amplicons (Amplicon sequence variant - ASV's), and then ASV's serve as the basis for classification. The classification is based on nucleotide sequence alignment methods using reference 16S rRNA gene sequence databases. The most common databases in microbiome research include SILVA and Greengenes, which contain a wide variety of prokaryotic sequences. The SILVA database is considered the last updated database (version 132, 2018), while the last Greengenes release was in 2013 (version 13.8).

Considerable attention should be paid to the introduction of techniques such as whole genome sequencing (WGS) analysis. Unlike 16S rRNA sequencing, WGS results allow a more accurate assessment of the genetic makeup of bacteria inhabiting the environment. WGS results provide information on all genes of microorganisms -

virulence factors, phylogenetically informative characters, markers of drug resistance [9]. The existing technologies have a number of disadvantages, which include the short length of single reads, errors in homopolymer regions, the dependence of the reading accuracy on the GC composition of the DNA fragment, and some others [10]. As an alternative, systems for the analysis of nucleic acids based on physical properties have been developed by the PacBio company and nanopore sequencing based on the Oxford Nanopore Technologies (ONT) platform [10]. Both approaches allow for very long single reads. The length of single reads on the PacBio platform is more than 80 thousand bases [10]. For ONTs, the typical single read length is more than 10 thousand bases [10].

High-throughput sequencing technologies have made it possible to accumulate a huge number of read sequences of microorganisms, the ever-increasing power of computers and the development of large database analysis programs make it possible to analyze external data or metagenomic analysis. To do this, you can use the databases available on the website of the Institute for Bioinformatics Research in the USA, NCBI (www.ncbi.nlm.nih.gov). Microbiota can also be investigated based on the data stored in the NCBI, focusing on various aspects, for example, exploring the relationship between the microbiota and various human pathologies, or investigating the relationship of the microbiota of various human organs, etc. The NCBI contains information on DNA, RNA, protein sequences, genomes, taxonomy of various organisms, and literature. NCBI also provides a program for bioinformatics analysis of primary data - BLAST program [11]. The microbiota of the gastrointestinal tract is being actively studied in the world, a lot of information has been obtained.

**Conclusion**

Metagenomic analysis of the microbiota is a potential approach to describing the structure of the bacterial community and understanding its functions, which in the future may turn into a number of diagnostic tools in medicine, including personalized ones. Such metagenomic studies will make it possible to either prevent or correct the development of disorders in the early stages.

**References**


1. Metagenomic 16S rDNA amplicon datasets from adolescents with normal weight, obesity, and obesity with irritable bowel syndrome from Eastern Siberia, Russia / N. Belkova, E. Klimenko, A. Romanitsa, A. Pogodina, L. Rychkova // Data in Brief. - 2020. - V.32. – №1. – P. 106141.
2. Gut–brain axis: how the microbiome influences anxiety and depression. / J. A. Foster, McVey Neufeld // Trends in Neurosciences – 2013. №5. – P.305–312.
3. Dysbiosis (dysbacteriosis) of the intestine: current state of the problem, complex diagnostics and therapeutic correction / M.D. Ardatskaya, S.V. Belmer, V.P. Dobritsa, S.M. Zakharenko, L.B. Lazebnik, O. N. Minushkin, L.S. Oreshko, S.I. Sitkin, E.I. Tkachenko, A.N. Suvorov, A.I. Khavkin, B.A. Shenderov // EiKG. - 2015. - T.117. - No. 5. - S. 13-50.
4. The use of molecular genetic methods for microbiological control of food products / MI Derevshchikova, M. Yu. Syromyatnikov, VN Popov // Technics and technology of food production. - 2018. - T. 48, No. 4. - P. 87–113.
5. Molecular genetic methods for the study of the intestinal microbiome / O.L. Kardymon, A.V. Kudryavtseva // Russian Journal of Gastroenterology, Hepatology, Coloproctology. - 2016. - T.26. - No. 4. - P.4-13.
6. Identification of microorganisms using gas chromatography-mass spectrometry / R.V. Pisanov, E.S. Shipko, O. V. Duvanova, D.I. Simakova // Journal of Microbiology, Epidemiology and Immunobiology. - 2020. - No. 4. - S.356-362.
7. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls / J.M. Janda, S.L. Abbott // Journal of Clinical Microbiology. – 2007. – №9, - P.2761–2764.
8. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis / B. Yang, Y. Wang, PY Qian // BMC Bioinformatics. – 2016. - №135.
9. The use of modern methods of identification of bacteria in the activities of the state collection of pathogenic microorganisms and cell cultures (GKPM-

Obolensk) / A.G. Bogun, A.A. Kislichkina, E.V. Galkina, N.V. Maiskaya and V.I. Solomentsev, T.N. Mukhina, I.G. Shemyakin // Infection and Immunity. - 2016. - T.6. - Number 3. –C.8.

10. Experience of using the technology of monomolecular nanopore sequencing in the study of bacterial strains deposited in the state collection of pathogenic microorganisms and cell cultures "GKPM-Obolensk" / V.I. Solomentsev, A.A. Sizova, Yu.P. Scriabin, A.A. Kislichkina, V.B. Frolov, N.V. Maiskaya, S.A. Ivanov, S.V. Dentovskaya, A.P. Anisimov, A.G. Bohun // Bacteriology. - 2018. - Vol. 3. - No. 4. - S. 60–68.
11. Textbook on bioinformatics [Electronic resource]. - Access mode: http://docplayer.ru/183354307-Uchebnik-n-yu-chasovskih.html (date accessed: 06/10/2021).

**Список литературы**

1. Metagenomic 16S rDNA amplicon datasets from adolescents with normal weight, obesity, and obesity with irritable bowel syndrome from Eastern Siberia, Russia / N. Belkova, E. Klimenko, A. Romanitsa, A. Pogodina, L. Rychkova // Data in Brief. - 2020. - V.32. – №1. – P. 106141.
2. Gut–brain axis: how the microbiome influences anxiety and depression. / J. A. Foster, McVey Neufeld // Trends in Neurosciences – 2013. №5. – P.305–312.
3. Дисбиоз (дисбактериоз) кишечника: современное состояние проблемы, комплексная диагностика и лечебная коррекция / М.Д. Ардатская, С.В. Бельмер, В.П. Добрица, С.М. Захаренко, Л.Б. Лазебник, О.Н. Минушкин, Л.С. Орешко, С.И. Ситкин, Е.И. Ткаченко, А.Н. Суворов, А.И. Хавкин, Б.А. Шендеров // ЭиКГ. – 2015. – Т.117. – №5. – С.13-50.
4. Использование молекулярно-генетических методов для микробиологического контроля пищевой продукции / М. И. Деревщикова, М. Ю. Сыромятников, В. Н. Попов // Техника и технология пищевых производств. – 2018. – Т. 48, № 4. – С. 87–113.
5. Молекулярно-генетические методы для исследования микробиома кишечника / О.Л. Кардымон, А.В. Кудрявцева // Российский журнал

гастроэнтерологии, гепатологии, колопроктологии. - 2016. – Т.26. - №4. – С.4-13.

6. Идентификация микроорганизмов с применением газовой хромато-масс-спектрометрии / Р.В. Писанов, Е.С. Шипко, О.В. Дуванова, Д.И. Симакова // Журнал микробиологии, эпидемиологии и иммунобиологии. - 2020. – №4. – С.356-362.
7. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls / J.M. Janda, S.L. Abbott // Journal of Clinical Microbiology. – 2007. – №9, - P.2761–2764.
8. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis / B. Yang, Y. Wang, PY Qian // BMC Bioinformatics. – 2016. - №135.
9. Использование современных методов идентификации бактерий в деятельности государственной коллекции патогенных микроорганизмов и клеточных культур (ГКПМ-Оболенск) / А.Г. Богун, А.А. Кисличкина, Е.В. Галкина, Н.В. Майская, В.И. Соломенцев, Т.Н. Мухина, И.Г. Шемякин // Инфекция и иммунитет. – 2016. – Т.6. – №3. –С.8.
10. Опыт использования технологии мономолекулярного нанопорового секвенирования при изучении штаммов бактерий, депонированных в государственной коллекции патогенных микроорганизмов и клеточных культур «ГКПМ-Оболенск» / В.И. Соломенцев, А.А. Сизова, Ю.П. Скрябин, А.А. Кисличкина, В.Б. Фролов, Н.В. Майская, С.А. Иванов, С.В. Дентовская, А.П. Анисимов, А.Г. Богун // Бактериология. – 2018. – Т.3. - №4. – С. 60–68.
11. Учебник по биоинформатике [Электронный ресурс]. – Режим доступа: http://docplayer.ru/183354307-Uchebnik-n-yu-chasovskih.html (дата обращения: 10.06.2021).