

METAGENOMIC LIBRARY PREPARATION FOR ILLUMINA PLATFORM

Krasnenko AY¹✉, Eliseev AY², Borisevich DI¹, Tsukanov KY¹, Davydova AI¹, Ilinsky VV¹

¹ Genotek Inc., Moscow, Russia

² Moscow South-West High School No. 1543, Moscow, Russia

Metagenomic sequencing is widely used in both scientific research and clinical practice for characterization of taxonomic profiles including estimation of relative abundance of prokaryotes in microbial communities in various media. Metagenomic sequencing of single marker genes is an excellent tool for studying the human microbiome. Unlike whole-genome sequencing, it targets those genome regions that can be instrumental in identification of microorganism species and genus. The 16S ribosomal RNA (16S rRNA) gene sequence is highly conserved but at the same time there are regions containing species-specific sequences that can discriminate between different bacteria and archaea. These regions can be amplified using universal primers, which makes the whole procedure more cost-effective and less time-consuming. Good primers and protocol design for PCR at the step of library preparation is crucial for achieving high data accuracy. Below we describe how to choose the optimal PCR protocol and universal primers to amplify V3 and V4 regions of the 16S rRNA gene for further sample sequencing using Illumina platform.

Keywords: microbiome, metagenomic sequencing, ribosomal RNA, 16S rRNA, polymerase chain reaction, universal primers, library preparation, double barcode

Acknowledgements: the authors thank Daria Plakhina and Ivan Stetsenok of Genotek for their help and Sergey Glagolev of Moscow South-West High School No. 1543 for his valuable advice and comments.

✉ **Correspondence should be addressed:** Anna Krasnenko
Nastavnicheskiy per. 17, str. 1, pod. 14, Moscow, 105120; annakrasnenko@gmail.com

Received: 12.04.2017 **Accepted:** 24.04.2017

ОСОБЕННОСТИ ПОДГОТОВКИ БИБЛИОТЕК ДЛЯ МЕТАГЕНОМНОГО СЕКВЕНИРОВАНИЯ ОБРАЗЦОВ НА ПЛАТФОРМЕ ILLUMINA

А. Ю. Красненко¹✉, А. Ю. Елисеев², Д. И. Борисевич¹, К. Ю. Цуканов¹, А. И. Давыдова¹, В. В. Ильинский¹

¹ ООО «Генотек», Москва

² Московская гимназия на Юго-западе № 1543, Москва

Метагеномное секвенирование, позволяющее определять таксономическую принадлежность и долю прокариотических организмов в сообществах в разных средах, широко используется не только в научной деятельности, но и в медицинской практике. В настоящее время для изучения микробиома человека применяется так называемое метагеномное секвенирование маркерных генов, при котором секвенируется не весь геном, а лишь те его регионы, по которым можно установить родовую и иногда видовую принадлежность микроорганизмов. Чаще всего для амплификации выбирают участки гена 16S рибосомальной РНК (16S рРНК), последовательность которого, с одной стороны, высококонсервативна, а с другой — содержит вариабельные участки, которые отличаются однонуклеотидными заменами в случае с разными микроорганизмами. При этом для амплификации таких участков генома возможен подбор универсальных праймеров, что значительно снижает стоимость и время исследования. Точность секвенирования в этом случае обеспечивается точностью подбора универсальных праймеров и оптимальностью условий проведения ПЦР на этапе подготовки библиотеки для секвенирования. В статье мы описываем подход к подбору универсальных праймеров и условий ПЦР для амплификации регион-специфической части V3 и V4 участков гена 16S рРНК для дальнейшего секвенирования образцов на платформах Illumina.

Ключевые слова: микробиом, метагеномное секвенирование, рибосомальная РНК, 16S рРНК, полимеразная цепная реакция, универсальные праймеры, библиотеки для секвенирования, двойное баркодирование

Благодарности: авторы благодарят Дарью Плахину и Ивана Стеценку из ООО «Генотек» за помощь в работе, а также Сергея Глаголева из Московской гимназии на Юго-Западе № 1543 — за ценные советы и замечания.

✉ **Для корреспонденции:** Красненко Анна Юрьевна
Наставнический пер., 17, стр. 1, подъезд 14, г. Москва, 105120; annakrasnenko@gmail.com

Статья получена: 12.04.2017 **Статья принята к печати:** 24.04.2017

The human body is home to a number of bacterial communities inhabiting the mouth cavity, gut, genitourinary system, etc. [1]. The totality of microorganisms that have symbiotic relationships with the host is called the microbiome [1]. Study of the human microbiome provides understanding of what bacteria live in healthy and diseased individuals [2].

Sequencing of those regions of bacterial genomes that discriminate between bacterial species and sometimes genera is called metagenomic sequencing, or marker gene sequencing, and is currently widely used in microbiome studies [3]. Regions of the 16S ribosomal RNA (16S rRNA) gene are especially convenient for sequencing. This gene is highly conserved in

prokaryotes enabling the use of universal primers to amplify the target sequence, which makes the whole procedure cost-effective and not so time-consuming [4, 5]. The 16S rRNA gene harbors conserved and variable regions. The latter contain single-base substitutions that are instrumental in identifying microbial species or genera: once these substitutions are detected by sequencing, they can be matched against regularly updated public databases.

One of the core steps of metagenomic sequencing workflow is sample preparation, i.e. converting source nucleic acids into a library of DNA fragments ready to be loaded onto the sequencer. There are a lot of different sequencing platforms, and although their sample preparation strategies are more or less the same, there are some nuances related to the techniques employed for signal detection during sequencing [6]. Sample preparation aims at obtaining DNA fragments that serve to identify a species or a genus of the studied microorganism. In metagenomics accuracy is largely determined by a good choice of primers necessary to produce amplicons by polymerase chain reaction (PCR) [7, 8]. In 2013 Klindworth et al. compiled a list of 512 primer pairs organized into three subgroups based on the next generation sequencing (NGS) technique they can be used for: group 1 consisted of primers for Illumina and Ion Torrent platforms (small amplicons), group 2 consisted of primers for 454 Life Science platform (middle-sized amplicons), and primers included in group 3 were intended for PacBio and other similar platforms (large amplicons) and could be also used to prepare genomic libraries of colonial species [9]. Each group offers a few universal primer pairs for archaea and bacteria instrumental in species/genus identification.

In this work we discuss some aspects of metagenomic sample sequencing with Illumina with a particular focus on library preparation for further sequencing [10]. Illumina-based sequencing takes place in the flow cell coated with single-stranded oligos that are complementary to library adapters ligated to source DNA fragments and enable hybridization. Polymerase lengthens hybridized DNA fragments attached to the flow cell surface. PCR produces multiple copies of a single template molecule forming millions of dense clusters. Clusters are then sequenced in parallel: complementary strands are generated by fluorescently tagged nucleotides, and the emitted signal is recorded after the addition of each nucleotide to the strand. This technology sets certain requirements for sample preparation explained below.

Library preparation for further sequencing with Illumina

Sequencing can be performed on various types of biological material, such as saliva, ear wax, nasal mucosal swabs, etc. We focused on the general aspects of library preparation for sequencing with Illumina regardless of the sample type. Basically, the sequencing workflow includes 5 steps: 1) extracting the intact DNA from the sample; 2) selecting genome

regions for sequencing and choosing primers for further PCR-based amplification (PCR quality is very important because it determines sequencing quality); 3) double barcoding of the obtained libraries; 4) sequencing itself; 5) bioinformatic analysis of the obtained data.

There are a lot of protocols and reagent kits for effective DNA extraction [11] depending on the type of the analyzed sample; therefore, there is no need to discuss this step in more detail here. In our study quality control of the extracted DNA was performed by agarose gel electrophoresis, concentration was measured by Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA) according to the manufacturer's protocol [12].

DNA extraction is followed by PCR amplification of the studied DNA fragments for further sequencing. For this work we chose a few regions of the 16S rRNA gene for the reasons explained above. Quality of the obtained fragments depends on the complementarity of the selected primers to the regions of the 16S rRNA gene [13, 14]. Primers consist of a region-specific sequence complementary to the flanking region of the target fragment and a synthetic sequence non-complementary to the region-specific sequence that will hybridize to the adapter. It is important that at least four 3'-end nucleotides should be non-complementary within and between primers to avoid primer-dimer formation. Even a small mismatch in complementarity of 3–4 nucleotides at the 3'-end of the primer can significantly reduce PCR quality even if the annealing temperature has been adjusted [15, 16].

There are a lot of regularly updated databases containing sequences of the 16S rRNA gene identified for a plethora of microbial species [17, 18], which facilitates rapid selection of universal primers using a special software if necessary [19]. For this work we chose universal primer pairs for the V3 and V4 regions of the 16S rRNA gene [23]. The synthetic sequence of the chosen primers was represented by sequences complementary to Nextera and Truseq adapters (Table 1).

Once the primers are selected, the PCR protocol needs to be optimized (a number of parameters have to be adjusted, such as primer concentration, DNA concentration, annealing temperature, Mg²⁺ concentration, number of cycles, etc.) to obtain a sufficient amount of good quality amplicons for further sequencing. PCR quality control is performed by agarose gel electrophoresis. Negative and positive controls are a must. A negative control is usually a PCR mix without the DNA template. In our case, two DNA samples were used as positive controls: one of *Rhizobium* and another of *Rhodococcus* bacterial genera.

PCR yield can be affected by primer-dimer formation [20]. Primer dimers also occur when performing PCR quality control in the agarose gel (Fig. 1). Diluting primers or adjusting the annealing temperature can be a solution. The optimal annealing temperature determines the purity of the reaction product since it facilitates primer attachment to DNA. State-of-art equipment makes it possible to do a temperature gradient to optimize the annealing temperature in a single run. High yields also depend on Mg²⁺ levels: Mg²⁺ ions bind to dNTP,

Table 1. Universal primer pairs consist of a sequence complementary to the V3 and V4 regions of the 16S rRNA gene and a synthetic sequence complementary to Nextera and Truseq adapters

Adapter	Primer sequence	
Nextera	forward	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
	reverse	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC
Truseq	forward	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTACGGGNGGCWGCAG
	reverse	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGACTACHVGGGTATCTAATCC

primers, DNA template and chelating agents (EDTA) present in the buffer [21]. Polymerase activity is known to grow at high Mg²⁺ concentrations, though polymerase specificity thereby decreases. As a rule, a range of Mg²⁺ concentrations from 1 to

4 mM with a 0.5 mM dilution step is tested to select the optimal concentration for the reaction mix.

When optimizing the PCR protocol, we found out that primers complementary to Nextera adapters were the best for the amplification of the V3 and V4 regions of the 16S rRNA gene. It is probably because the unique Nextera sequence is non-complementary to the regions of the studied bacterial genome, which prevents formation of side products. The optimized PCR protocol is shown in Table 2. PCR was performed using Step One Plus system (Applied Biosystems, USA).

The harvested libraries are dual-indexed (barcoded) in another PCR step. Barcoding is adding an index sequence of 8 nucleotides to the DNA fragment to facilitate further discrimination between different sample sets [22]. There is a wide selection of reagent mixes for barcoding that can be used in Illumina-based sequencing, such as Nextera XT Index kit. We used oligos synthesized by Evrogen, Russia. We ran a few tests with various PCR parameters to discover that barcoding yields did not depend on the purity of the DNA template and required no sample purification. The optimal PCR parameters for Nextera-based barcoding with Nextera primers (Table 3) are shown in Table 4.

It should be reminded that poor library quality control will entail mistakes during sequencing. We usually perform quality control using Agilent Bioanalyzer 2100 (Agilent Technologies, USA) (Fig. 2).

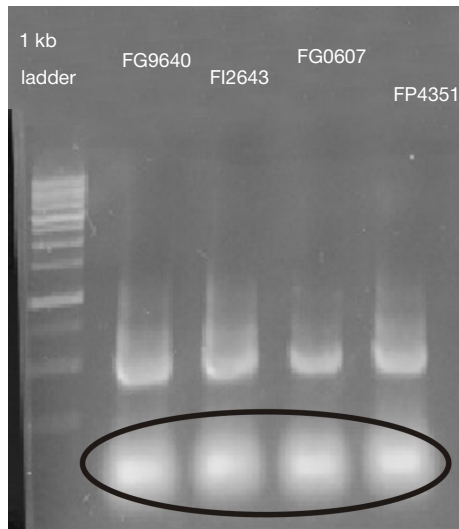


Fig. 1. Agarose gel with PCR products after electrophoresis. The first well contains a DNA-size marker, the rest contain the samples. Primer dimers (PCR by-products) are marked with a black oval

Table 2. Parameters of PCR for amplification of the V3 and V4 regions using universal primers for the Nextera adapter

Composition	V, µl	Protocol		
		T, °C	t, s	Number of cycles
Forward primer (C = 1 µM)	5,0	95	180	1
Reverse primer (C = 1 µM)	5,0	95	30	25
Taq-polymerase	0,5	55	30	
dNTP	1,5	72	30	
PCR buffer	2,5	72	300	1
DNA	2,5			
mQ	8,0			

Table 3. Nextera index primers

Primer name	Primer sequence
i701	CAAGCAGAAGACGGCATAACGAGATTGCGCTTAGTCTCGTGGGCTCGG
i702	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGG
i703	CAAGCAGAAGACGGCATAACGAGATTCTGCCTGTCTCGTGGGCTCGG
i704	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGG
i501	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGTCGGCAGCGTC
i502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTTCGTCGGCAGCGTC
i503	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
i504	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC

Table 4. Parameters of PCR for barcoding with Nextera primers

Composition	V, µl	Protocol		
		T, °C	t, s	Number of cycles
Index 1 (C = 1 µM)	5,0	95	180	1
Index 2 (C = 1 µM)	5,0	95	30	10
Taq-polymerase	0,5	55	30	
dNTP	1,5	72	30	
PCR-buffer	2,5	72	300	1
1st PCR amplicons	5,0			
mQ	5,5			

Sample sequencing

We will not focus on the sequencing step itself in this article, because there are standard sequencing protocols supplied by the vendor [24, 25]. We performed MiSeq paired-end sequencing (Illumina) with 250 b. p. reads according to the standard protocol.

Bioinformatic analysis

The obtained nucleotide sequences are processed and classified as suggested by the Ribosomal Database Project, ver. 11.5 (Michigan University, USA), using the RDPTools ver. 2016-07-21 [26, 27]. Classification confidence threshold (-conf) should be set to 50 % as recommended by [28].

Quality control is essential in sequencing. At least 95 % of sequences in each sample must be high quality, i. e. contain no adapters or contaminating elements that cannot be mapped onto the human genome. The number of reads per sample is especially important in metagenomic sequencing. But on the whole, there is no universal rule here and the number of reads depends mostly on the purpose of the study. If the study aims at identifying dominant bacteria in the sample, the number of reads can be low. For example, 350 reads per each of 22 human gut samples revealed the presence of two dominating bacteria: *Firmicutes* (75 %) and *Bacteroidetes* (18 %) [29]. However, higher read numbers increase chances of discovering microbial “minorities” in the sample and reduces effects of the sampling error. A high-resolution metagenomic analysis requires at least 10,000 reads [29].

The proportion of unclassified sequences and unknown bacterial sequences, the median proportion of sequences for which both genus and species have been reliably identified should be consistent with the results of the 16S rRNA-based microbial metagenomic analysis. The proportion of unclassified sequences should not exceed 20 %, the genus is expected to be identified for at least 70 % sequences, while the species — for at least 50 %. Still, these figures may vary depending on the study.

Normally, a table is compiled based on the analysis, showing a taxonomic hierarchy (domain, phylum, class, etc.) and providing information on the taxonomic tree and relative abundance of taxa in the sample. An example of such table is Table 5.

CONCLUSIONS

Selecting primers for amplification is an important step in metagenomic sequencing of 16S rRNA gene regions. Quality of the obtained amplicons determines accuracy of sequencing. We suggest an approach to designing an optimal PCR protocol for sample preparation that can be used to adjust PCR parameters for library preparation and identified problems that may occur at this step. Although library preparation for metagenomic sequencing has been widely discussed in the literature [30], we have made an attempt to design a well-defined protocol, proposed optimal parameters for the amplification of 16S rRNA V3 and V4 regions using universal primers for further sequencing with Illumina. Of note, the quality of amplification yields depends on many factors, including the purity of the reagents, therefore our PCR protocol is not universal and its PCR conditions may vary depending on the reagents used.

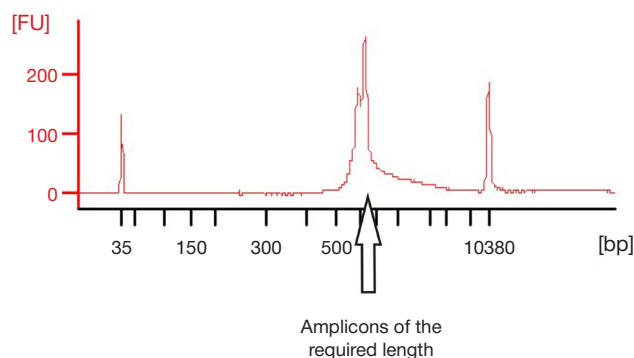


Fig. 2. Sequencing quality control with Agilent Bioanalyzer 2100

Table 5. Bioinformatic analysis: representation of results

Rank	Name	rootrank	domain	phylum	class	order	family	genus	Sample 1	Sample 2	Sample ...
1_domain	Archaea	Root	Archaea								
2_phylum	Crenarchaeota	Root	Archaea	Crenarchaeota							
3_class	Thermoprotei	Root	Archaea	Crenarchaeota	Thermoprotei						
4_order	Acidilobales	Root	Archaea	Crenarchaeota	Thermoprotei	Acidilobales					
5_family	Acidilobaceae	Root	Archaea	Crenarchaeota	Thermoprotei	Acidilobales	Acidilobaceae				
6_genus	Acidilobus	Root	Archaea	Crenarchaeota	Thermoprotei	Acidilobales	Acidilobaceae	Acidilobus			

Note. Rank refers to the elements of the taxonomic hierarchy; Name refers to the name of the taxon. These two tables are used for data filtering. For example, using the Rank column as a filter, one can retrieve information on the species/phyla only, etc. The next 7 columns (rootrank, domain, phylum, class, order, family, genus) provide information about the full hierarchal path. The last columns contain data on the relative abundance of the taxon in the studied sample. Sample names are specified in column titles

References

- Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, et al. Design and Experimental Application of a Novel Non-Degenerate Universal Primer Set that Amplifies Prokaryotic 16S rRNA Genes with a Low Possibility to Amplify Eukaryotic rRNA Genes. *DNA Res.* 2014; 21 (2): 217–27.
- Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, García-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* 2017 Feb 10; 243: 16–24.
- Oulas A, Pavloudi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G., et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights.* 2015 May 5; 9: 75–88.

4. Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* 2009 Jul; 19 (7): 1141–52.
5. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol.* 2014 Sep; 12 (9): 635–45.
6. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif).* 2013; 6: 287–303.
7. Fabrice A, Didier R. Exploring microbial diversity using 16S rRNA high-throughput methods. *J Comput Sci Syst Biol.* 2009; 2 (1): 74–92.
8. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE.* 2011; 6 (12): e27310.
9. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013 Jan 7; 41 (1): e1.
10. Illumina Inc., Web Site [homepage on the Internet]. c2017 [cited 2017 Apr 10]. Next-Generation Sequencing (NGS) [cited 2017 Apr 10]; [about 2 screens]. Available from: <https://www.illumina.com/technology/next-generation-sequencing.html>
11. Bag S, Saha B, Mehta O, Anbumani D, Kumar N, Dayal M, et al. An Improved Method for High Quality Metagenomics DNA Extraction from Human and Environmental Samples. *Sci Rep.* 2016 May 31; 6: 26775.
12. Qubit 3.0 Fluorometer User Guide [file on the Internet]. Revision A.0. US: Thermo Fisher Scientific Inc. c2014 [cited 2017 Apr 10]: 67 p. Available from: https://tools.thermofisher.com/content/sfs/manuals/qubit_3_fluorometer_man.pdf
13. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol.* 1998 Oct; 64 (10): 3724–30.
14. Reysenbach AL, Giver LJ, Wickham GS, Pace NR. Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol.* 1992 Oct; 58 (10): 3417–8.
15. Bru D, Martin-Laurent F, Philippot L. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Appl Environ Microbiol.* 2008 Mar; 74 (5): 1660–3.
16. Wu JH, Hong PY, Liu WT. Quantitative effects of position and type of single mismatch on single base primer extension. *J Microbiol Methods.* 2009 Jun; 77 (3): 267–75.
17. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. CAMERA: a community resource for metagenomics. *PLoS Biol.* 2007 Mar; 5 (3): e75.
18. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013 Jan; 41 (Database issue): D590–6.
19. Mao D-P, Zhou Q, Chen C-Y, Quan Z-X. Coverage evaluation of universal bacterial primers using the metagenomic datasets. *BMC Microbiol.* 2012 May 3; 12: 66.
20. Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, et al. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res.* 1997 Aug 15; 25 (16): 3235–41.
21. Lorenz TC. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *J Vis Exp.* 2012 May 22; (63): e3998.
22. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 2012 Jan; 40 (1): e3.
23. 16S Metagenomic Sequencing Library Preparation. Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System [file on the Internet]. [cited 2017 Apr 10]: 28 p. Available from: https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
24. MiSeq System Denature and Dilute Libraries Guide [file on the Internet]. San Diego: Illumina Inc.; c2016 [cited 2017 Apr 10]: 14 p. Available from: http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-01.pdf
25. Rukovodstvo po sisteme MiSeq [file on the Internet]. San Diego: Illumina Inc.; c2015 [cited 2017 Apr 10]: 106 p. Available from: http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/translations/miseq-system-guide-15027617-01-rus.pdf. Russian.
26. The Ribosomal Database Project (RDP) [database on the Internet]. Release 11. East Lansing, MI: Michigan State University. c1992–2014 [cited 2017 Apr 10]. Available from: <https://rdp.cme.msu.edu/>
27. Collection of commonly used RDP Tools for easy building. In: GitHub [Git repository hosting service]. c2017 [cited 2017 Apr 10]. Available from: <https://github.com/rdpstaff/RDPTools>
28. RDP extensible sequence classifier for fungal, bacterial and archaeal 16S. In: GitHub [Git repository hosting service]. c2017 [cited 2017 Apr 10]. Available from: <https://github.com/rdpstaff/classifier/tree/701e229d4e7cbe53d4261301e23459d91615999d>
29. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A.* 2015 Nov 10; 112 (45): 14024–9.

Литература

1. Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, et al. Design and Experimental Application of a Novel Non-Degenerate Universal Primer Set that Amplifies Prokaryotic 16S rRNA Genes with a Low Possibility to Amplify Eukaryotic rRNA Genes. *DNA Res.* 2014; 21 (2): 217–27.
2. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol.* 2017 Feb 10; 243: 16–24.
3. Oulas A, Pavlouidi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G., et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights.* 2015 May 5; 9: 75–88.
4. Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* 2009 Jul; 19 (7): 1141–52.
5. Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer KH, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol.* 2014 Sep; 12 (9): 635–45.
6. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif).* 2013; 6: 287–303.
7. Fabrice A, Didier R. Exploring microbial diversity using 16S rRNA high-throughput methods. *J Comput Sci Syst Biol.* 2009; 2 (1): 74–92.
8. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE.* 2011; 6 (12): e27310.
9. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013 Jan 7; 41 (1): e1.
10. Illumina Inc., интернет-сайт [официальная веб-страница]. c2017 [дата обращения: 10 апреля 2017 г.]. Next-Generation Sequencing (NGS) [дата обращения: 10 апреля 2017 г.]; [примерно 2 экрана]. Доступно по: <https://www.illumina.com/technology/next-generation-sequencing.html>
11. Bag S, Saha B, Mehta O, Anbumani D, Kumar N, Dayal M, et al. An Improved Method for High Quality Metagenomics DNA

- Extraction from Human and Environmental Samples. *Sci Rep*. 2016 May 31; 6: 26775.
12. Qubit 3.0 Fluorometer User Guide [файл из интернета]. Версия A.O. США: Thermo Fisher Scientific Inc. с2014 [дата обращения: 10 апреля 2017 г.]; 67 с. Доступно по: https://tools.thermofisher.com/content/sfs/manuals/qubit_3_fluorometer_man.pdf
 13. Polz MF, Cavanaugh CM. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol*. 1998 Oct; 64 (10): 3724–30.
 14. Reysenbach AL, Giver LJ, Wickham GS, Pace NR. Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol*. 1992 Oct; 58 (10): 3417–8.
 15. Bru D, Martin-Laurent F, Philippot L. Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. *Appl Environ Microbiol*. 2008 Mar; 74 (5): 1660–3.
 16. Wu JH, Hong PY, Liu WT. Quantitative effects of position and type of single mismatch on single base primer extension. *J Microbiol Methods*. 2009 Jun; 77 (3): 267–75.
 17. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M. CAMERA: a community resource for metagenomics. *PLoS Biol*. 2007 Mar; 5 (3): e75.
 18. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013 Jan; 41 (Database issue): D590–6.
 19. Mao D-P, Zhou Q, Chen C-Y, Quan Z-X. Coverage evaluation of universal bacterial primers using the metagenomic datasets. *BMC Microbiol*. 2012 May 3; 12: 66.
 20. Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, et al. The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Res*. 1997 Aug 15; 25 (16): 3235–41.
 21. Lorenz TC. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *J Vis Exp*. 2012 May 22; (63): e3998.
 22. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res*. 2012 Jan; 40 (1): e3.
 23. 16S Metagenomic Sequencing Library Preparation. Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System [файл из интернета]. [дата обращения: 10 апреля 2017 г.]; 28 с. Доступно по: https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf
 24. MiSeq System Denature and Dilute Libraries Guide [файл из интернета]. Сан-Диего: Illumina Inc.; с2016 [дата обращения: 10 апреля 2017 г.]; 14 с. Доступно по: http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-01.pdf
 25. Руководство по системе MiSeq [файл из интернета]. Сан-Диего: Illumina Inc.; с2015 [дата обращения: 10 апреля 2017 г.]; 106 с. Доступно по: http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/translations/miseq-system-guide-15027617-01-rus.pdf
 26. База данных рибосомальной систематики (RDP) [база данных в интернете]. Релиз 11. Ист-Лансинг, США: Мичиганский университет. с1992–2014 [дата обращения: 10 апреля 2017 г.]. Доступно по: <https://rdp.cme.msu.edu/>.
 27. Collection of commonly used RDP Tools for easy building. На GitHub [хостинг проектов]. с2017 [дата обращения: 10 апреля 2017 г.]. Доступно по: <https://github.com/rdpstaff/RDPTools>
 28. RDP extensible sequence classifier for fungal lsu, bacterial and archaeal 16s. На GitHub [хостинг проектов]. с2017 [дата обращения: 10 апреля 2017 г.]. Доступно по: <https://github.com/rdpstaff/classifier/tree/701e229dde7cbe53d4261301e23459d91615999d>
 29. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A*. 2015 Nov 10; 112 (45): 14024–9.