UDC 579.25:577.21

Original Article

GENETIC BARCODING FOR SURVEILLANCE OF VIRULENCE PLASMIDS IN METAGENOMIC SAMPLES OF MICE GUT MICROFLORA

Shilov S.V.[1], Korotetskiy I.S.[1*], Reva O.N.[2], Kuznetsova T.V.[1], Zubenko N.V.[1], Ivanova L.N.[1], Parenova R.A.[1], Izmailov T.Kh.[3]

[1]JSC "Scientific Center of Anti-Infectious Drugs", al-Farabi ave., 75a, Almaty, 050060, Kazakhstan
[2]Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics, and Microbiology, University of Pretoria, Pretoria, South Africa
[3]LLP "Research and Production Association Kazpharmacom", Khmeleva str., 1/13, Almaty, 050010, Kazakhstan
[*] laeda1@mail.ru

ABSTRACT

Outbreaks of nosocomial infections strike the public health system around the world. Current surveillance systems for infectious agents either take too long to identify pathogens or lack the necessary sensitivity to distinguish the virulent strains from the benign ones in the resident microflora. To address this problem, we propose a method for improving the sensitivity and specificity of detection of virulent clonal lines of bacteria or individual virulence plasmids in metagenomic samples. In this study, we used previously developed computational tools, Barcode Generator and Barcoding 2.0, to analyse gut microflora of laboratory mice infected with a pathogenic multidrug-resistant *S. aureus* and treated with cefazolin, an iodine-containing complex CC-195, and by a combinatorial treatment. We searched the metagenomic samples for representatives of pathogenic microflora using diagnostic genetic barcodes, which were designed based on the whole genome sequences of a collection of clinical isolates. Our results demonstrated a practical applicability of these diagnostic barcodes to monitor clonal lines of pathogens and even individual virulence plasmids in the environment. We also found that the novel drug CC-195 promotes a speedy restoration of the normal gut microflora disturbed by infection and administration of the antibiotic cefazolin.

Keywords: hospital infection; NGS sequencing; metagenome; diagnostics; DNA barcode; gut microflora.

INTRODUCTION

Outbreaks of hospital infections are a well-recognized problem of the public health system around the world. Monitoring these infections is challenging because it can be difficult to distinguish between pathogenic and benign variants of opportunistic bacterial species that may coexist in one sample. The pathogenicity of bacterial strains is rendered by various virulence and antibiotic resistance determinants present in their genomes, which may not always be detectable by the phenotype. Monitoring pathogens in gut microflora is especially challenging because often they compose only a small fraction of the bacterial community. Traditional culture-based methods are labour-intensive, time-consuming and not precise enough to detect individual pathogens within highly diverse bacterial communities. Furthermore, it's difficult to identify virulent variants of the normal resident microflora, which evolved due to an acquisition of virulence plasmids and other mobile genetic elements. It is known that the presence of at least one horizontally acquired virulence gene may suffice to confer pathogenicity to a benign microorganism [1]. Detection of clonal lineages of pathogens by their genotype requires either isolation of the strains and sequencing whole genomes or using specific primers for a diagnostic PCR amplification of marker genomic loci for further sequencing. Both variants are time and labour consuming as much as the traditional culture-based approaches.

Microbial species in metagenomes can be identified by sequencing species-specific marker genes, such as 16S rRNA for bacteria or ITS for fungal species [2]. The sequenced marker genes are then searched through public databases of reference sequences, such as the GreenGene [3] or RDP [2]. However, these phylogenetic markers are too conserved for distinguishing even between related bacterial species.

With the advancement of the Next-Generation Sequencing (NGS) technologies, the shotgun Whole Genome Sequencing (WGS), which generates multiple short random reads from total metagenomic DNA, has become a more popular approach than using the phylogenetic marker genes for metabarcoding. A large collection of short DNA reads generated by the NGS can be binned into individual species by a k-mer based mapping of either DNA fragments or translated amino-acid short reads against reference sequences in the public databases. These algorithms were implemented in Kraken [4] and Kaiju [5] software tools, respectively. However, one limitation of these approaches is that only those species having similar reference sequences deposited in the respective databases can be identified. While this may not be a problem for identification of pathogenic bacteria, which are well represented in the databases, the low sensitivity of the k-mer based methods of binning to the species level leaves many reads unclassified and many closely related species unidentified. The identification based on amino acid k-mers is potentially more sensitive. However, translating short DNA reads to proteins reveals only very short substrings of protein sequences, which may be found in different proteins and different species. This drawback limits the sensitivity of species binning approaches that use read translation to proteins.

One promising approach of surveillance of individual strains of interest in metagenomes consists in creation of diagnostic genetic barcodes based on the whole genome sequences [6]. This approach offers improved sensitivity and specificity enabling detection of barcoded species or individual strains even when they are in a minority in the metagenomic sample. However, this improvement in the sensitivity/specificity is at the cost of inability to identify all the taxa present in a sample but only to monitor the barcoded bacterial lineages. This

approach is well-suited to the problem of monitoring pathogens or individual virulence plasmids, large integrons and virulence-associated genomic islands even in rich environments such as the gut microbiome. Monitoring pathogens and their individual genetic determinants of virulence and antibiotic resistance enables the creation of real-time systems for surveillance of evolution and distribution of pathogens causing hospital infections outbreaks in different countries, regions, and hospitals. This will be of an immense importance for timely prediction of upcoming outbreaks to prevent nosocomial infections in hospitals [7].

An automated design of diagnostic barcodes of bacterial pathogens can be performed using the program Barcode Generator [8]. Provided with a set of whole genome sequences of strains of interest, the program identifies homologous and accessory parts of the genomes to design genome-specific barcode sequences. These barcode sequences can be used as reference sequences for binning NGS reads. A genetic barcode is a combination of sequences that represent genomic loci with an optimal level of conservation and specific adaptive modifications. This allows for unambiguous detection of respective DNA fragments in metagenomic datasets and distinguishing between closely related microorganisms. To improve the specificity of binning, unique sequences found in the reference genomes are also included in the genetic barcodes. The program recommends by default a preferred length of a barcode sequence and composes it with 67 % variable homologous loci and 33 % genome-specific sequences identified in the provided reference genomes. An associated program Barcoding performs binning of metagenomic reads against the generated barcode sequences to identify presence and relative abundance of the searched microorganisms in metagenomic samples [8]. While the capacity of tracking individual plasmids has been declared, this method has not yet been properly validated experimentally.

MATERIALS AND METHODS

Drugs. The iodine-containing complex CC-195 is a compound of crystalline iodine with hydrofluoric acid and glycine, developed at the Scientific Center for Anti-Infective Drugs in Kazakhstan (SCAID). This complex has a broad spectrum of antimicrobial activity, similar to another complex synthesized at SCAID (patent #6039).

Cefazolin is a powder used for the preparation of an injection solution (Santo, Kazakhstan). It belongs to the first generation of cephalosporin antibiotics and has a broad spectrum of antimicrobial activity against both gram-positive and gram-negative microorganisms.

Sample collection. White laboratory mice of both genders were used in this study. The animals were divided into 5 experimental groups (n=3 in each group). The groups were marked as follows: A) Intact mice as a negative control; B) Infected untreated mice as a positive control; C) Infected mice treated daily *per os* with 100.0 mg/kg of a iodine-containing complex CC-195; D) Infected mice treated daily intramuscularly with 25.0 mg/kg of cefazolin; E) Infected mice treated with CC-195 in combination with Cefazolin as described above. The laboratory mice were infected with a clinical isolate *Staphylococcus aureus* SCAID OTT1-2022, characterized with a broad spectrum of resistance to antibiotics (ampicillin; amoxicillin; gentamicin; ceftriaxone; erythromycin; oxacillin, cefepime, cefazolin, levofloxacin, clindamycin, clarithromycin, polymyxin B, colistin, cefixime, aztreonam). Progression of the infection was controlled by plating mouse lung samples on CHROMagar Orientation medium (bioMérieux, France) for overnight cultivation at 37°C and by PCR amplification using the following protocol: 10x buffer (Invitrogen, USA) containing 2 mM mix of dNTPs (Invitrogen, USA), 20.0 pM of forward and reverse primers, $MgCl_2$ (Invitrogen, USA) and 5 U/µL Taq DNA polymerase (Invitrogen, USA) were mixed with 2.0 µl of DNA isolated from 200 µl aliquot of the grinded lung tissue. The following primers, STPYF – 5‘ACGGTCTTGCTGTCACTTATA3‘ and STPYR2 – 5‘TACACATATGTTCTTCCCTAATAA3‘, targeting the strain-specific region of *S. aureus* SCAID OTT1-2022 were designed and tested in the previous work (unpublished results). The target locus was amplified in 35 cycles of melting at 94 °C for 30 sec, annealing at 62 °C for 30 sec and elongation at 72 °C for 45 sec after a pre-heating step at 94 °C for 5 min. When the infection was confirmed in 3 days after injection, the treatment protocols were started. The therapeutic treatment of the infected animals lasted for five days. Then the mice of all the groups were sacrificed for further laboratory studies of the intestinal microbiota.

The intestines of the mice were placed into cuvettes containing 10.0 ml of saline buffer with 0.1% Tween 80. The intestines were then triple washed using a syringe. The resulting suspensions were centrifuged at 500 g for 10 minutes at 4°C to remove the faecal matter and the epithelial cells. The supernatant was then centrifuged again at 16,000 g for 10 minutes at 4°C to obtain the bacterial cells in the pellets. The pellets were subsequently resuspended in 1.0 ml of saline buffer.

DNA extraction was performed using the PureLink Genomic DNA Mini Kit (Invitrogen, USA) as recommended by the manufacturer. Quality control of the DNA samples was performed using the spectrophotometer NanoDrop 2000c (Thermo Scientific, USA) by comparison of the light abortion at 260 and 280 nm.

DNA sequencing and bioinformatics analysis. Six metagenomic samples were prepared for further sequencing from each experimental condition. Total metagenomic DNA samples were sequenced by the Ion Torrent PGM (Life Technologies, USA). The DNA library was created using the Ion Xpress Plus Fragment Library preparation KIT (Life Technologies, USA) after sample barcoding with the Xpress Barcode Adapters (Life Technologies, USA). Sequencing was performed using the chip 318 of the Ion PGM Hi-Q View KIT in accordance with the manufacturer’s instructions. Quality control and trimming of the obtained DNA reads were performed using utilities of UGENE v.44.0 [9] with the program run setting by default.

Creation of reference genomic barcodes

Whole genome sequences of the pathogenic multidrug resistant isolates obtained previously from the Department of Vascular Surgery of Syzganov’s National Scientific Center of Surgery (Almaty, Kazakhstan) [10-12] were used in this study as references (Table 1). Barcodes were generated using the program Barcode Generator [8] with the default parameter setting. Presence of the reference microorganisms in the metagenomic samples was identified by mapping the gener-

ated DNA reads against the barcode sequences using the Barcoding 2.0 tool, which is a part of the program Barcode Generator.

Table 1 – Reference multidrug resistant clinical isolates used for creation of diagnostic barcodes and the accession numbers of the respective replicons at GenBank

Seta of barcodes	Replicons used to create the sets of barcodes	
	Bacterial species	Accessions at GenBank
#1 (chromosomes)	*Burkholderia contaminans* SCAID TST1-2021 (11/270)	CP082807, CP082808, CP082809
	Escherichia coli SCAID WND1-2021 (1/128)	CP082831
	Klebsiella pneumoniae SCAID PHRX1-2021 (13/97)	CP082805
	Pseudomonas aeruginosa SCAID TST-2021 (7/157)	CP082823
	Staphylococcus aureus SCAID OTT1-2022 (150)	CP102945
	Stenotrophomonas maltophilia SCAID WND1-2022 (370)	CP102943
	Staphylococcus epidermidis SCAID OTT1-2021 (597)	CP082816
	Streptococcus pneumoniae SCAID PHRX1-2021	CP082820
#2 (chromosomes)	*Burkholderia contaminans* SCAID TST1-2021 (11/270)	CP082807, CP082808, CP082809
	Escherichia coli SCAID WND1-2021 (1/128)	CP082831
	Escherichia coli SCAID WND1-2022 (119)	CP102947
	Escherichia coli SCAID URN1-2021 (19/278)	CP082824
	Escherichia coli SCAID WND2-2021 (3/145)	CP082827
	Klebsiella pneumoniae SCAID PHRX1-2021 (13/97)	CP082805
	Klebsiella pneumoniae SCAID TST1-2021 (15/226)	CP082799
	Klebsiella pneumoniae SCAID PHRX2-2021 (20/245)	CP082796
	Klebsiella pneumoniae SCAID PND1-2022 (426)	CP102940
	Pseudomonas aeruginosa SCAID TST-2021 (7/157)	CP082823
	Pseudomonas aeruginosa SCAID WND1-2022 (148)	CP102946
	Pseudomonas aeruginosa SCAID PLC1-2021 (16/222)	CP082821
	Pseudomonas aeruginosa SCAID TCT1-2022 (325)	CP102944
	Pseudomonas aeruginosa SCAID WND1-2021 (9/195)	CP082822
	Staphylococcus aureus SCAID OTT1-2022 (150)	CP102945
	Staphylococcus aureus SCAID OTT1-2021 (597/2)	CP082813
	Staphylococcus aureus SCAID WND1-2021 (598)	CP082815
	Stenotrophomonas maltophilia SCAID WND1-2022 (370)	CP102943
	Staphylococcus epidermidis SCAID OTT1-2021 (597)	CP082816
	Streptococcus pneumoniae SCAID PHRX1-2021	CP082820
#3 (plasmids)	*Burkholderia contaminans* SCAID TST1-2021 (11/270) plasmid 1	CP082810
	Burkholderia contaminans SCAID TST1-2021 (11/270) plasmid 3	CP082812
	Escherichia coli SCAID URN1-2021 (19/278) plasmid 1	CP082825
	Escherichia coli SCAID URN1-2021 (19/278) plasmid 2	CP082826
	Escherichia coli SCAID WND1-2021 (1/128) plasmid	CP082832
	Escherichia coli SCAID WND2-2021 (3/145) plasmid 1	CP082828
	Escherichia coli SCAID WND2-2021 (3/145) plasmid 2	CP082829
	Klebsiella pneumoniae SCAID PHRX1-2021 (13/97) plasmid 1	CP082806
	Klebsiella pneumoniae SCAID TST1-2021 (15/226) plasmid 1	CP082800
	Klebsiella pneumoniae SCAID PHRX2-2021 (20/245) plasmid 1	CP082797
	Klebsiella pneumoniae SCAID PHRX2-2021 (20/245) plasmid 2	CP082798
	Klebsiella pneumoniae SCAID PND1-2022 (426) plasmid	CP102941

RESULTS AND DISCUSSION

DNA samples were extracted in six repeats from the laboratory mice used in the five experimental sets and sequenced by Ion Torrent PGM. The obtained DNA reads were quality filtered and trimmed. The general statistics of sequencing the metagenomic samples before and after the quality control is shown in Table 2.

Table 2 – Overview of the results of metagenomic sequencing and quality filtering and trimming

Samples	Initial statistics			Statistics after quality filtering and trimming		
	Number of reads	Total length	Average length of reads	Number of reads	Total length	Average length of reads
A1	1080238	250940528	232	1024948	244622762	238
A2	496940	115702717	232	471361	112795051	239
A3	875856	204030137	232	832621	199075467	239
A4	189871	41006436	215	189860	40227711	211
A5	117291	25853958	220	117283	25358915	216
A6	196707	42924262	218	196707	42144778	214
B1	143587	33640462	234	137303	32889036	239
B2	166565	38136459	228	157237	37102935	235
B3	142142	32002060	225	134980	31211331	231
B4	297987	61171492	205	297982	60191327	201
B5	275112	56696952	206	275109	55686525	202
B6	502538	104620849	208	502531	102845336	204
C1	564646	118624618	210	534008	115409220	216
C2	124769	25742759	206	115941	24872087	214
C3	538710	117521464	218	514900	114800472	222
C4	173512	38130929	219	173510	37472651	215
C5	36438	7664244	210	36434	7505399	205
C6	177499	39479977	222	177495	38812085	218
D1	174467	38441679	220	165499	37459581	226
D2	211830	45990311	217	200323	44752022	223
D3	49960	9530615	190	46329	9181690	198
D4	328919	68352477	207	328917	67152918	204
D5	309256	67743968	219	309250	66559778	215
D6	25581	4617833	180	25581	4543262	177
E1	700454	149088726	212	663506	145073055	218
E2	373262	80708359	216	353306	78560398	222
E3	151142	30837821	204	141259	29835471	211
E4	251507	53525528	212	251505	52584481	209
E5	76491	16947897	221	76490	16636484	217
E6	37523	7373805	196	37522	7235709	192

Three sets of diagnostic barcodes were generated from genome sequences shown in Table 1, which were obtained in our previous studies [10-12]. Several genomes were used repeatedly in the different sets of barcodes; however, their barcode sequences were not the same as the program Barcode Generator creates specific barcodes for every set of the reference sequences to achieve the best distinguishing power to separate these specific organisms. For instance, the set #1 was designed to distinguish between phylogenetically distant pathogens, whereas the set #2 aimed at distinguishing between the most dominant species of the clinical isolates. And the set #3 is a collection of large (> 15,000 bp) virulence plasmids found in the genomes of the selected pathogenic bacteria.

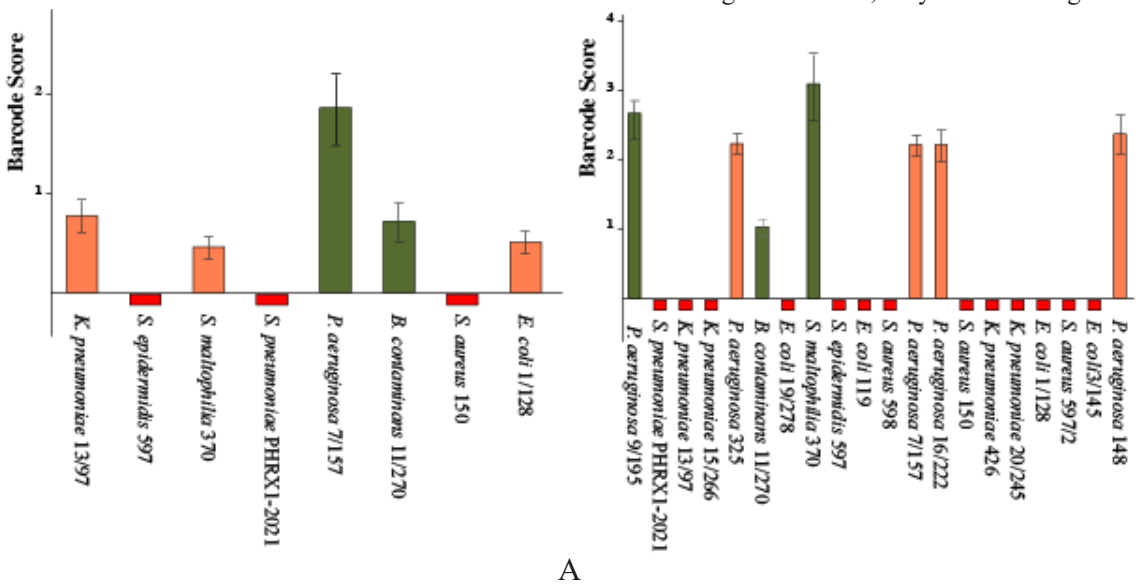
Presence of the reference microorganisms and plasmids in the metagenomic samples was identified by mapping the generated DNA reads against the barcode sequences using the program Barcoding 2.0. The program uses BLASTN alignment of reads against the barcode sequences followed by a statistical analysis of matches in output files. Visualisation of the results of mapping the metagenomic DNA reads against the sets 1 and 2 of the diagnostic barcodes is shown in Figure 1.

It was found that several pathogenic microorganisms, particularly *Stenotrophomonas*, *Burkholderia*, and several clonal lines of *P. aeruginosa*, were present in the microbiome of the healthy mice representing the negative control (Figure 1A). When the mice were infected with *S. aureus* SCAID OTT1-2022, there was a significant change in the gut microflora where enterobacteria *E. coli* and *K. pneumoniae* became dominant (Figure 1B). Treatment of the infected animals with cefazolin did not restore their gut microbiota. It remained essentially the same as in the positive control (Figures 1B and 1C). In contrast to that, application of CC-195 alone (Figure 1D) and in combination with the antibiotic cefazolin (Figure 1E) has stimulated a restoration of the gut microflora of the treated mice to its normal state similar to that in the negative control (Figure 1A).

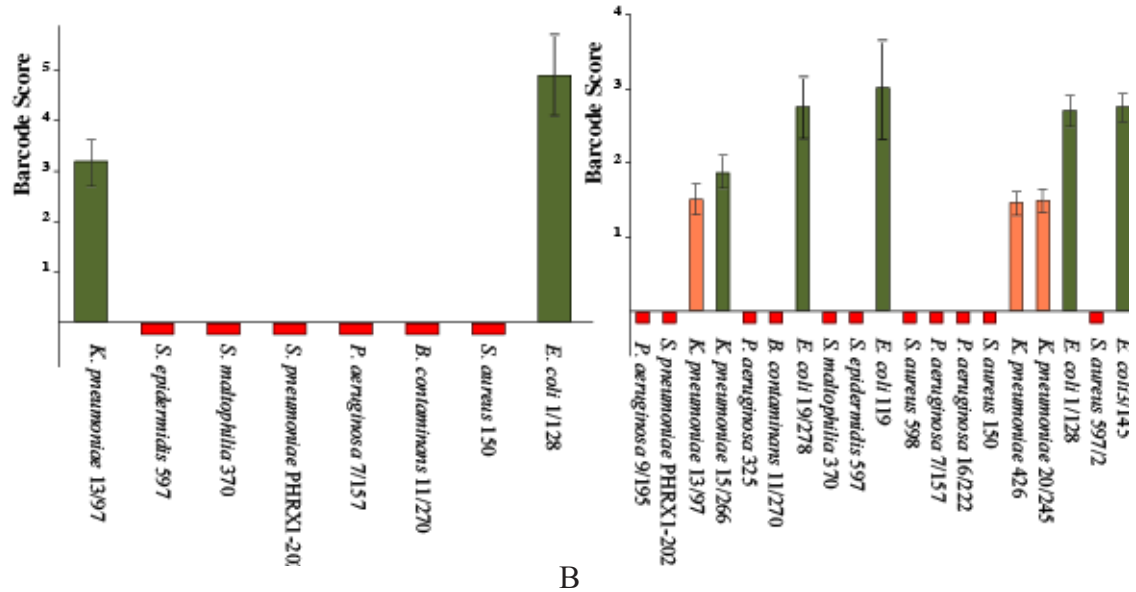
It was of a practical interest to verify the possibility to use this approach to monitor the presence of individual virulence plasmids in metagenomic samples as an acquisition of such a

plasmid by benign resident microorganisms can render them with pathogenicity. The results of application of the plasmid diagnostic barcodes are shown in Figure 2.

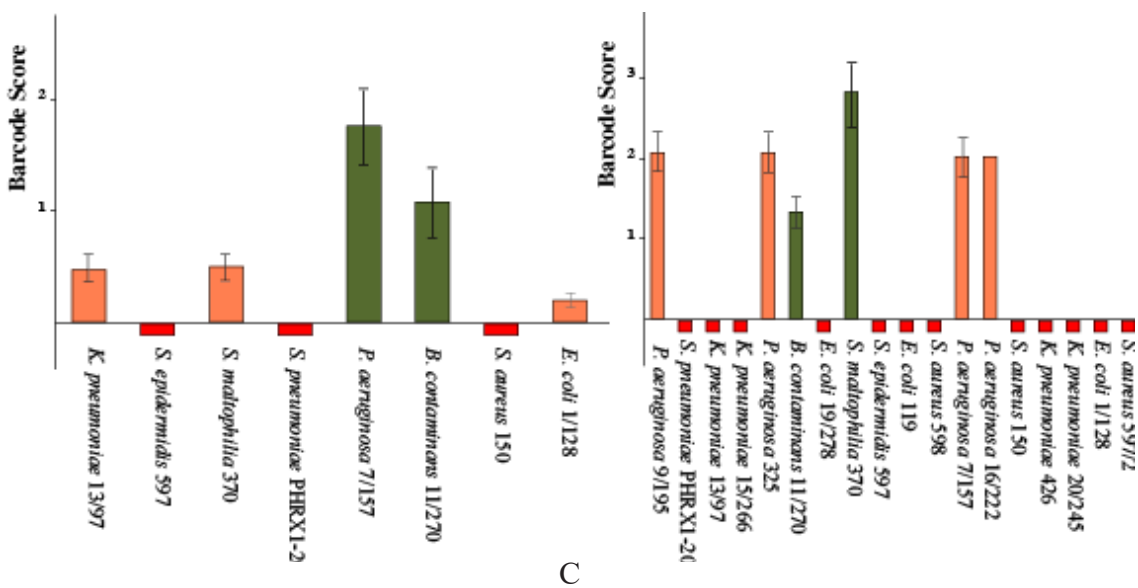
In the negative control, only a few NGS generated DNA



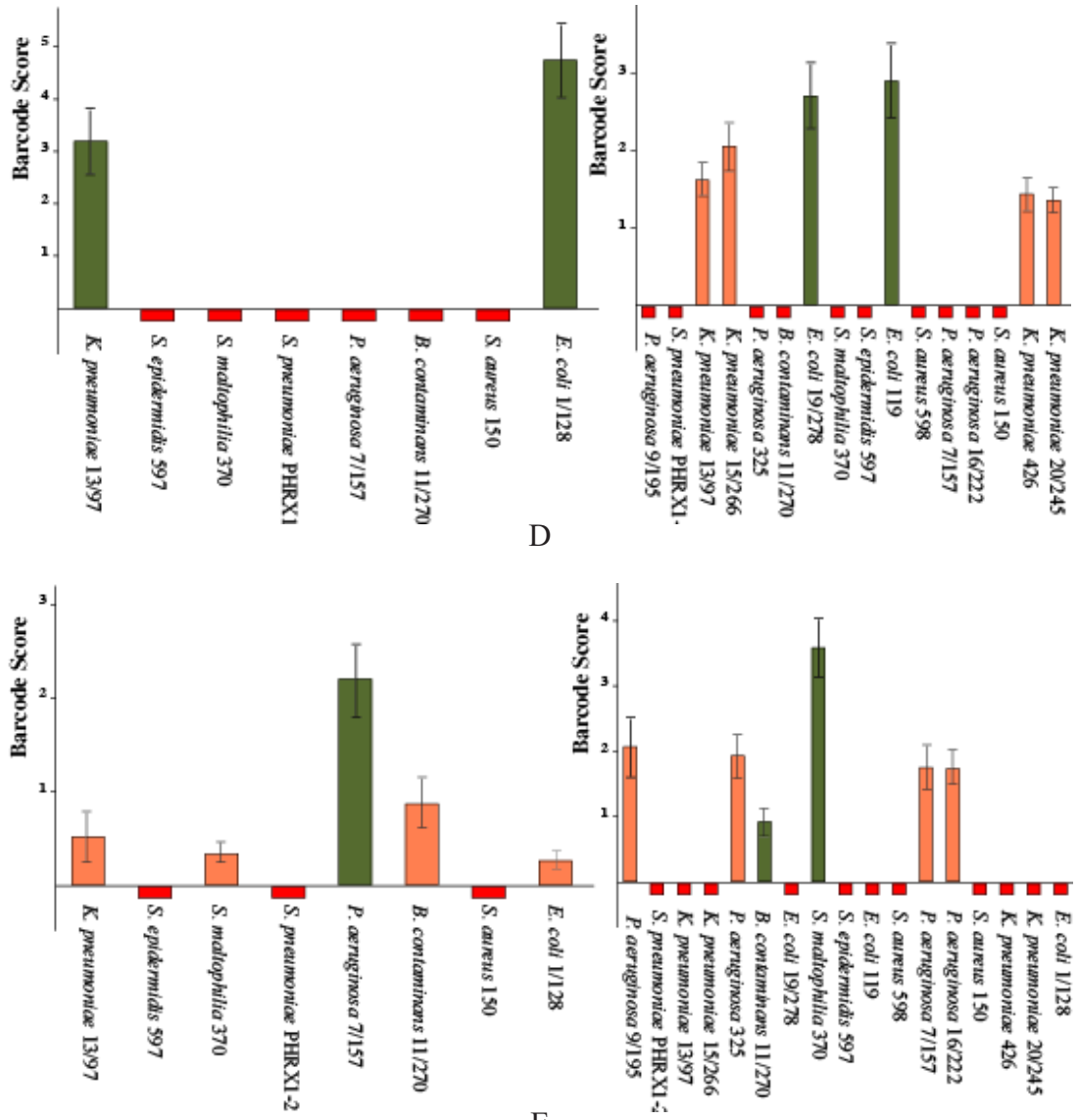
A



B



C



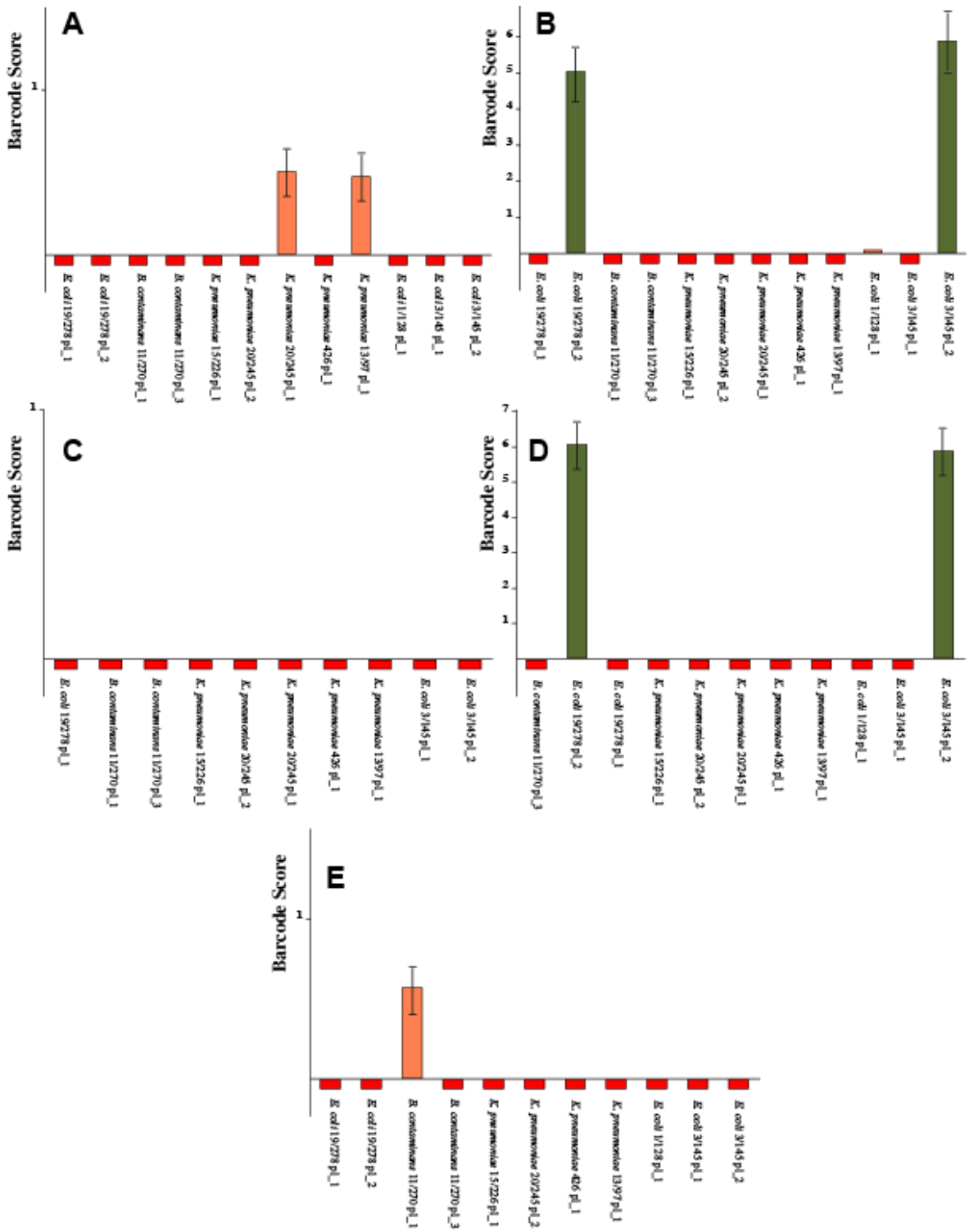
A – negative control; B – positive control (mice infected with *S. aureus* SCAID OTT1-2022); C – infected mice treated with CC-195; D – infected mice treated with cefazolin; E – infected mice treated with cefazolin and CC-195. Green bars depict the presence of the respective reference organisms detected by both the core and accessory parts of the barcodes; orange bars show that the reference strains were detected only by read matches to the core part of the barcodes and short red bars indicate absence of the reference bacteria in the sample or that their abundance in the sample was below the level of sensitivity of the method. Strains are titled by shortened names given in Table 1 in parenthesis.

Figure 1 – Identification of the reference organisms in the metagenomic samples using the barcode sets #1 и #2

fragments were binned to the reference virulence plasmids of *Klebsiella* (Figure 2A). In contrast to that, infecting the mice with *S. aureus* SCAID OTT1-2022 has increased the abundance of bacteria bearing virulence plasmids, which resemble the reference plasmids of the *E. coli* 3/145 and *E. coli* 19/278 (Figure 2B). These two pathogenic strains contain several different plasmids in their genomes (Table 1). Mapping of the metagenomic reads revealed presence of only short plasmids of these pathogens, whereas the large virulence plasmids were missed in the samples.

A promising assumption on a possible positive role of the iodine-containing complex CC-195 on removal of the pathogens bearing virulence plasmids from bacterial communities can be deduced from the analysis of the metagenomic samples of the mice treated solely with this drug (Figure 2C). No virulence plasmids of *E. coli* 3/145 and *E. coli* 19/278 were found in these samples. In contrast to that, these plasmids were found in the group of the mice treated with cefazolin (Figure 2D) basically in the same frequency as in the positive control (Figure 2B). When the infected mice were treated with the combination of cefazolin and CC-195, the virulence plasmids disappeared from the biome except for a few binned reads resembling the plasmids from *Burkholderia* (Figure 2E).

A noteworthy observation from this work is that the results of application of the plasmid-specific diagnostic barcodes agreed with the results of application of the barcodes designed for the bacterial chromosomes. Indeed, the plasmids of enterobacterial pathogens were abundant in those metagenomic samples where the enterobacteria were identified as a dominant component of the microbiomes. This indicates the applicability of the programs Barcode Generator and Barcoding 2.0 to monitor the distribution of not only the pathogens, but individual virulence factors in the environment.



A – negative control; B – positive control (mice infected with *S. aureus* SCAID OTT1-2022); C – infected mice treated with CC-195; D – infected mice treated with cefazolin; E – infected mice treated with cefazolin and CC-195. Green bars depict the presence of the respective plasmids detected by both the core and accessory parts of the barcodes; orange bars show that the reference plasmids were detected only by matches to the core part of the barcodes and short red bars indicate absence of the plasmids in the sample or that their abundance in the sample was below the level of sensitivity of the method. Strains are titled by shortened names given in Table 1 in parenthesis.

Figure 2 – Identification of the virulence plasmids in the metagenomic samples by using the set 3 of the plasmid barcodes

CONCLUSION

This work demonstrated that the programs Barcode Generator and Barcoding 2.0 are useful tools for monitoring clonal lines of pathogens and individual virulence plasmids in environmental and body-associated microbiomes with significant sensitivity and specificity. The method can be used to detect outbreak-causing pathogens and virulence genetic determinants in hospital infection surveillance systems and to evaluate the efficacy and possible side effects of therapeutic regimens administered against multidrug-resistant pathogens. In the future work, the performance of this approach can be further improved by developing additional functions aimed at recognition of individual virulence genes and antibiotic resis-

tance mutations in metagenomic datasets. The practical outcome of this work consists in demonstration of the ability of the iodine-containing complex CC-195 to stabilize and restore the normal gut microflora of the infected mice suffering from a dysbacteriosis caused by the infection with the pathogen *S. aureus* SCAID OTT1-2022 and the antibiotic treatment with cefazolin. A therapeutic effect of administration of CC-195 was promising and requires further studies.

ACKNOWLEDGMENTS

The authors thank the Microbiology Laboratory (SCAID) for maintaining the clinical strain and providing data on antibiotic resistance.

FUNDING

This study was supported by the Ministry of Healthcare of the Republic of Kazakhstan [grant number BR09458960 – "Study of reversion of antibiotic resistance of pathogenic microorganisms for 2021–2023"].

LITERATURE

1. Marcus S.L., Brumell J.H., Pfeifer C.G., Finlay B.B. Salmonella pathogenicity islands: big virulence in small packages //Microbes and Infection. – 2000. – Vol. 2, No 2. – P. 145-156. PMID: 10742687. doi: 10.1016/s1286-4579(00)00273-2.

2. Cole J.R., Wang Q., Fish J.A., Chai B., McGarrell D.M., Sun Y., Brown C.T., Porras-Alfaro A., Kuske C.R., Tiedje J.M. Ribosomal Database Project: data and tools for high throughput rRNA analysis // Nucl. Acids Res. – 2014. – Vol. 42, No D1. – P. D633-D642. PMID: 24288368. doi: 10.1093/nar/gkt1244.

3. McDonald D., Price M.N., Goodrich J., Nawrocki E.P., DeSantis T.Z., Probst A., Andersen G.L., Knight R., Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea // ISME J. – 2012. – Vol. 6, No 3. – P. 610-618. PMID: 22134646. doi: 10.1038/ismej.2011.139.

4. Wood D.E., Salzberg S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments // Genome Biol. – 2014. – Vol. 15, No 3. – P. R46. PMID: 24580807/ doi: 10.1186/gb-2014-15-3-r46.

5. Menzel P., Kim Lee Ng., Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju // Nat. Commun. – 2016. – Vol. 7. – P. 11257. PMID: 27071849. doi: 10.1038/ncomms 11257.

6. Gaytan A., Bergsten J., Canelo T., Perez-Izquierdo C., Santoro M., Bonal R. DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots // Ecol. Evol. – 2020. – Vol. 10, No 19. – P. 10754-10772. PMID: 33072294. doi: 10.1002/ece3.6733.

7. Papaiakovou M., Littlewood D.T.J., Doyle S.R., Gasser R.B., Cantacessi C. Worms and bugs of the gut: the search for diagnostic signatures using barcoding, and metagenomics–metabolomics // Parasites Vectors. – 2022. – Vol. 15, No 1. – P. 118. PMID: 35365192. doi: 10.1186/s13071-022-05225-7.

8. Rotimi A.M., Pierneef R., Reva O.N. Selection of marker genes for genetic barcoding of microorganisms and binning of metagenomic reads by Barcoder software tools // BMC Bioinformatics. – 2018. – Vol. 19, No 1. – P. 309. PMID: 30165813. doi: 10.1186/s12859-018-2320-1.

9. Rose R., Golosova O., Sukhomlinov D., Tiunov A., Prosperi M. Flexible design of multiple metagenomics classification pipelines with UGENE // Bioinformatics. – 2019. – Vol. 35, No 11. – P. 1963-1965. PMID: 30358807. doi: 10.1093/bioinformatics/bty901.

10. Korotetskiy I.S., Jumagaziyeva A.B., Kerimzhanova B., Reva O.N., Shilov S.V., Kuznetsova T., Ivanova L., Korotetskaya N., Bekmukhamedova A., Satylgankyzy G., Klivleyeva N.G. Complete genome sequences of Gram-positive opportunistic pathogens isolated from hospitals in Almaty, Kazakhstan //. Microbiol Resour Announc. – 2022. – Vol. 11, No 4. – P. e00093-22. PMID: 35286173. doi: 10.1128/mra.00093-22.

11. Korotetskiy I., Jumagaziyeva A., Kerimzhanova B., Reva O., Kuznetsova T., Shilov S., Ivanova L., Zubenko N., Parenova R., Iskakbayeva Z., Baimakhanov B., Bekmuhamedova A. Whole genome sequence data of Stenotrophomonas maltophilia SCAID WND1-2022 (370) // Data Brief. – 2022. – Vol. 45. – P. 108694. PMID: 36425995. doi: 10.1016/j.dib.2022.108694.

12. Korotetskiy I.S., Shilov S.V., Kuznetsova T., Kerimzhanova B., Korotetskaya N., Ivanova L., Zubenko N., Parenova R., Reva O.N. Analysis of Whole-Genome Sequences of Pathogenic Gram-Positive and Gram-Negative Isolates from the Same Hospital Environment to Investigate Common Evolutionary Trends Associated with Horizontal Gene Exchange, Mutations and DNA Methylation Patterning // Microorganisms. – 2023. – Vol. 11, No 2. – P. 323. PMID: 36838287. doi: 10.3390/microorganisms11020323.

REFERENCES

1. Marcus S.L., Brumell J.H., Pfeifer C.G., Finlay B.B. Salmonella pathogenicity islands: big virulence in small packages //Microbes and Infection. – 2000. – Vol. 2, No 2. – P. 145-156. PMID: 10742687. doi: 10.1016/s1286-4579(00)00273-2.

2. Cole J.R., Wang Q., Fish J.A., Chai B., McGarrell D.M., Sun Y., Brown C.T., Porras-Alfaro A., Kuske C.R., Tiedje J.M. Ribosomal Database Project: data and tools for high throughput rRNA analysis // Nucl. Acids Res. – 2014. – Vol. 42, No D1. – P. D633-D642. PMID: 24288368. doi: 10.1093/nar/gkt1244.

3. McDonald D., Price M.N., Goodrich J., Nawrocki E.P., DeSantis T.Z., Probst A., Andersen G.L., Knight R., Hugenholtz P. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea // ISME J. – 2012. – Vol. 6, No 3. – P. 610-618. PMID: 22134646. doi: 10.1038/ismej.2011.139.

4. Wood D.E., Salzberg S.L. Kraken: ultrafast metagenomic sequence classification using exact alignments // Genome Biol. – 2014. – Vol. 15, No 3. – P. R46. PMID: 24580807/ doi: 10.1186/gb-2014-15-3-r46.

5. Menzel P., Kim Lee Ng., Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju // Nat. Commun. – 2016. – Vol. 7. – P. 11257. PMID: 27071849. doi: 10.1038/ncomms11257.

6. Gaytan A., Bergsten J., Canelo T., Perez-Izquierdo

C., Santoro M., Bonal R. DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots // Ecol. Evol. – 2020. – Vol. 10, No 19. – P. 10754-10772. PMID: 33072294. doi: 10.1002/ece3.6733.

7. Papaiakovou M., Littlewood D.T.J., Doyle S.R., Gasser R.B., Cantacessi C. Worms and bugs of the gut: the search for diagnostic signatures using barcoding, and metagenomics–metabolomics // Parasites Vectors. – 2022. – Vol. 15, No 1. – P. 118. PMID: 35365192. doi: 10.1186/s13071-022-05225-7.

8. Rotimi A.M., Pierneef R., Reva O.N. Selection of marker genes for genetic barcoding of microorganisms and binning of metagenomic reads by Barcoder software tools // BMC Bioinformatics. – 2018. – Vol. 19, No 1. – P. 309. PMID: 30165813. doi: 10.1186/s12859-018-2320-1.

9. Rose R., Golosova O., Sukhomlinov D., Tiunov A., Prosperi M. Flexible design of multiple metagenomics classification pipelines with UGENE // Bioinformatics. – 2019. – Vol. 35, No 11. – P. 1963-1965. PMID: 30358807. doi: 10.1093/bioinformatics/bty901.

10. Korotetskiy I.S., Jumagaziyeva A.B., Kerimzhanova B., Reva O.N., Shilov S.V., Kuznetsova T., Ivanova L., Korotetskaya N., Bekmukhamedova A., Satylgankyzy G., Klivleyeva N.G. Complete genome sequences of Gram-positive opportunistic pathogens isolated from hospitals in Almaty, Kazakhstan //. Microbiol Resour Announc. – 2022. – Vol. 11, No 4. – P. e00093-22. PMID: 35286173. doi: 10.1128/mra.00093-22.

11. Korotetskiy I., Jumagaziyeva A., Kerimzhanova B., Reva O., Kuznetsova T., Shilov S., Ivanova L., Zubenko N., Parenova R., Iskakbayeva Z., Baimakhanov B., Bekmuhamedova A. Whole genome sequence data of Stenotrophomonas maltophilia SCAID WND1-2022 (370) // Data Brief. – 2022. – Vol. 45. – P. 108694. PMID: 36425995. doi: 10.1016/j.dib.2022.108694.

12. Korotetskiy I.S., Shilov S.V., Kuznetsova T., Kerimzhanova B., Korotetskaya N., Ivanova L., Zubenko N., Parenova R., Reva O.N. Analysis of Whole-Genome Sequences of Pathogenic Gram-Positive and Gram-Negative Isolates from the Same Hospital Environment to Investigate Common Evolutionary Trends Associated with Horizontal Gene Exchange, Mutations and DNA Methylation Patterning // Microorganisms. – 2023. – Vol. 11, No 2. – P. 323. PMID: 36838287. doi: 10.3390/microorganisms11020323.

ГЕНЕТИЧЕСКОЕ БАРКОДИРОВАНИЕ И ОБНАРУЖЕНИЕ ПЛАЗМИД ВИРУЛЕНТНОСТИ В МЕТАГЕНОМНЫХ ОБРАЗЦАХ МИКРОФЛОРЫ КИШЕЧНИКА МЫШЕЙ

Шилов С.В.[1], Коротецкий И.С.[1*], Рева О.Н.[2], Кузнецова Т.В.[1], Зубенко Н.В.[1], Иванова Л.Н.[1], Паренова Р.А.[1], Измайлов Т.Х.[3]

[1]JSC АО «Научный центр противоинфекционных препаратов», пр. аль-Фараби, 75а, Алматы, 050060, Казахстан
[2]Центр биоинформатики и вычислительной биологии, Отдел Биохимии, Генетики и Микробиологии, Университет Претории, Претория, Южная Африка;
[3]ТОО «Научно-производственное объединение Казфармаком», ул. Хмелева, 1/13, Алматы, 050010, Казахстан
[*] laeda1@mail.ru

АБСТРАКТ

Вспышки нозокомиальных инфекций поражают систему здравоохранения во всем мире. Существующие системы эпиднадзора за инфекционными агентами либо требуют слишком много времени для идентификации патогенов, либо не обладают необходимой чувствительностью, чтобы отличить вирулентные штаммы от сопутствующих в резидентной микрофлоре. Для решения этой проблемы мы предлагаем метод повышения чувствительности и специфичности обнаружения, вирулентных клональных линий бактерий или отдельных плазмид вирулентности в метагеномных образцах. В данном исследовании мы использовали, ранее разработанные вычислительные инструменты Barcode Generator и Barcoding 2.0, для анализа микрофлоры кишечника лабораторных мышей, инфицированных патогенным мультирезистентным *S. aureus* и леченных цефазолином, иодсодержащим комплексом СС-195 и совместно. Мы провели поиск представителей патогенной микрофлоры в метагеномных образцах с помощью диагностических генетических баркодов, разработанных на основе полногеномных последовательностей клинических изолятов. Наши результаты продемонстрировали практическую применимость этих диагностических баркодов для мониторинга клональных линий патогенов и отдельных плазмид вирулентности. Мы также обнаружили, что соединение СС-195 способствует быстрому восстановлению нормальной микрофлоры кишечника, нарушенной инфекцией и приемом антибиотика цефазолина.

Ключевые слова: больничная инфекция; метагеном; диагностика; ДНК баркод; микрофлора кишечника.

ГЕНЕТИКАЛЫҚ БАРКОДТАУ ЖӘНЕ ТЫШҚАНДАРДЫҢ ІШЕК МИКРОФЛОРАСЫНЫҢ МЕТАГЕНОМДЫҚ ҮЛГІЛЕРІНДЕГІ ВИРУЛЕНТТІЛІК ПЛАЗМИДАЛАРЫН АНЫҚТАУ

Шилов С.В.[1], Коротецкий И.С.[1*], Рева О.Н.[2], Кузнецова Т.В.[1], Зубенко Н.В.[1], Иванова Л.Н.[1], Паренова Р.А.[1], Измайлов Т.Х.[3]

[1] «Инфекцияға қарсы препараттар ғылыми орталығы» АҚ, аль-Фараби, 75а, Алматы, 050060, Қазақстан
[2]Биоинформатика және есептеу биологиясы орталығы, биохимия, генетика және микробиология бөлімі, Претория университеті, Претория, Оңтүстік Африка
[3]»Қазфармаком ғылыми-өндірістік бірлестігі» ЖШС, Хмелев көшесі, 1/13, Алматы, 050010, Қазақстан
[*] laeda1@mail.ru

АБСТРАКТ

Нозокомиальды инфекциялардың өршуі бүкіл әлемдегі денсаулық сақтау жүйесіне әсер етеді. Қолданыстағы жұқпалы агенттерді қадағалау жүйелері қоздырғыштарды анықтау үшін тым көп уақытты қажет етеді немесе вирулентті штаммдарды резиденттік микрофлорасындағы ілеспе штаммдардан ажырату үшін қажетті сезімталдыққа ие болмайды. Осы мәселені шешу үшін метагеномдық үлгілердегі бактериялардың вирулентті клондық линияларын немесе жеке вируленттілік плазмидаларын анықтау сезімталдығы мен ерекшелігін арттыру әдісін ұсынамыз. Зерттеуде бұрын әзірленген Barcode Generator мен Barcoding 2.0 есептеу құралдарын патогенді мультирезистентті *S. aureus* жұқтырған және құрамында СС-195 иоды бар кешені мен цефазолинмен бірге емделген зертханалық тышқандардың ішек микрофлорасын талдауы үшін қолдандық. Клиникалық изоляттардың толық геномдық тізбегі негізінде жасалған диагностикалық генетикалық баркодтар арқылы метагеномдық үлгілерде патогендік микрофлора өкілдеріне ізденісте болдық. Алынған нәтижелер осы диагностикалық баркодтардың патогендік клондық линиялары мен жеке вируленттілік плазмидаларын бақылау үшін практикалық қолданысын көрсетті. Сондай-ақ СС-195 қосылысы инфекциямен және цефазолин антибиотигімен бұзылған қалыпты ішек микрофлорасының тез қалпына келуіне ықпал ететінін анықтадық.

Негізгі сөздер: аурухана инфекциясы; метагеном; диагностика; баркод ДНҚ-сы; ішек микрофлорасы.