An Algorithm for the Selection of Probes for Specific Detection of Human Disease Pathogens Using the DNA Microarray Technology

E.N. Filatova, PhD, Leading Researcher, Laboratory of Molecular Biology and Biotechnology[1];
A.S. Chaikina, Student[2];
N.F. Brusnigina, MD, PhD, Associate Professor, Head of the Laboratory for Metagenomics
and Molecular Indication of Pathogens[1];
M.A. Makhova, PhD, Senior Researcher, Laboratory for Metagenomics and Molecular Indication of Pathogens[1];
O.V. Utkin, PhD, Head of the Laboratory of Molecular Biology and Biotechnology[1]

[1]Blokhina Scientific Research Institute of Epidemiology and Microbiology of Nizhny Novgorod,
Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rospotrebnadzor),
71 Malaya Yamskaya St., Nizhny Novgorod, 603950, Russia;
[2]Privolzhsky Research Medical University, 10/1 Minin and Pozharsky Square, Nizhny Novgorod, 603005, Russia

The aim of the study was to develop an algorithm for the selection of discriminating probes to identify a wide range of causative agents of human infectious diseases.

Materials and Methods. The algorithm for selecting the probes was implemented in the form of the disprose (DIScrimination PRObe SElection) computer program written in the R language. Additionally, third-party software was used: the BLAST+ and ViennaRNA Package programs. The developed algorithm was tested by selecting specific probes for detecting *Chlamydophila (Chlamydia) pneumoniae* — an atypical bacterial pathogen causing community-acquired pneumonia (CAP). Nucleotide sequences for analysis were downloaded from the NCBI databank.

Results. An algorithm for the selection of specific probes capable of detecting human infectious pathogens has been developed. The algorithm is implemented in the form of the disprose modular program, which allows for performing all stages of the probe selection process: loading the nucleotide sequences and their metadata from available databanks, creating local databases, forming a pool of probes, calculating their physicochemical parameters, aligning the probes and sequences contained in local databases, processing and evaluating the alignment results. The algorithm was successfully tested and its performance was confirmed by selecting a set of probes for the specific detection of *Chlamydophila pneumoniae*. The specificity of the selected probes calculated *in silico* indicated a low risk of their nonspecific binding and a high potential of using them as molecular genetic diagnostic tools (DNA microarrays, PCR).

Conclusion. An algorithm for the selection of specific probes detecting a wide range of human pathogens in clinical biomaterial has been developed and implemented in the form of the disprose modular program. The probes selected using this program can serve as the functional basis of DNA-oriented microarrays able to identify causative agents of polyetiological diseases, such as CAP. Due to the flexibility and openness of the program, the scope of its application can be expanded.

Key words: probe selection algorithm; DNA microarray; DNA microarray design; community-acquired pneumonia; *Chlamydophila pneumoniae.*

Corresponding author: Elena N. Filatova, e-mail: filatova@nniiem.ru

Introduction

In view of the current epidemic situation, the need for diagnostic DNA microarrays capable of detecting bacterial and viral pathogens has become urgent.

The diagnostic value of a DNA microarray depends on the selection of specific oligonucleotide probes. There are difficulties with this selection, which are due to the chemical complexity and species specificity of biological samples; these factors contribute to the high risk of probe cross-hybridization with a non-target DNA, and the appearance of false positive results [1].

The existing algorithms for the probe selection are based on assessing the conformity of candidate sequences to the criteria of specificity and homogeneity [2–5]. As a rule, the software that implements the known algorithms does not allow the user to modify either these criteria or the order of their application and/or parameters. However, the weight of the criteria and their appropriateness may vary depending on the spectrum of detected pathogens, their taxonomic diversity, the biological sample specifics, and the purpose of the examination (identification of pathogenic factors, determination of antibiotic resistance, etc.). Any additional complication in the procedure for probe selection inevitably leads to an increase in the time and labor, as well as an increase in the requirements for computing equipment. In this regard, we suggest that algorithms implemented in the form of modular/modifiable programs are preferable for the selection of specific probes to be used in diagnostic DNA microarrays.

The aim of the study was to develop an algorithm for the selection of discriminating probes to identify a wide range of causative agents of human infectious diseases.

Materials and Methods

In this study, we pursued the following specific goals: to develop an algorithm for the selection of specific probes, to implement it in the form of a computer program, to optimize the performance of the algorithm, and to test it by selecting probes for the detection of *Chlamydophila (Chlamydia) pneumoniae.*

The algorithm for the selection of probes intended for the mass differential detection of bacterial and viral pathogens is implemented in the form of the disprose computer program written in the R programming language and designed as a package of functions. The package is distributed under the GNU GPL-3 license (2007) and is available for downloading from The Comprehensive R Archive Network official international repository (https://CRAN.R-project.org/package=disprose).

The computations were performed using an Intel Xeon 2560 (x2) workstation, 128 GB RAM. The algorithm was tested by searching for probes that allow specific detection of the "atypical" causative agent of community-acquired pneumonia

(CAP) — *C. pneumoniae*. The genetic sequences of *C. pneumoniae* were obtained from the NCBI Nucleotide database [6].

To calculate the minimal folding energy (MFE) of the oligonucleotide sequences in the candidate probes, we used the ViennaRNA Package (version 2.4.14) [7, 8]. The melting point (T_m) was calculated based on the established set of thermodynamic parameters [9] by the nearest neighbor method [10].

Local alignment of nucleotide sequences was carried out using the blastn program from the BLAST+ program package (version 2.10.0) [11]. The search for matches was performed in the full-size downloadable NCBI Nucleotide collection database, as well as in local databases of nucleotide sequences generated using the blastdbcmd program.

This work did not use information that might violate anyone's confidentiality. No human or animal subjects were involved in the study.

Results

Algorithm for the selection of discriminating probes and its implementation in the disprose program

Selection of target nucleotide sequences. Before starting the procedure for selecting the probes, it was necessary to discern between the target nucleotide sequences (to make the probes hybridize) and nonspecific sequences (to avoid hybridization).

The target sequences can be obtained by the researchers themselves or downloaded from the available databanks. To date, the disprose program implements the functions of downloading the sequences and their metadata from several large banks: NCBI (Nucleotide, GenBank, RefSeq databases) and GISAID. Based on the metadata obtained, the researcher can select the sequences of interest from the entire set of available data. The selected nucleotide sequences constitute a local base of target sequences for testing the ability of candidate probes to hybridize to these sequences.

Selection of nonspecific sequences. To date, the downloadable NCBI Nucleotide collection contains more than 71 million sequences, with a total size of 466 gigabases (Gb). Using such a framework to test the probes for nonspecific hybridization is overly time-consuming. To limit the selection process, it is advisable to reduce the volume of nonspecific sequences by including only those sequences that can hypothetically be present in the tested samples.

Using the literature, we have identified more than 40 taxa of microorganisms representing the relevant CAP pathogens and the biological material used for making the diagnosis (sputum, smears from the nasopharynx, oropharynx, etc.). A complete list of the taxa is presented in Appendix 1. For each taxon (using the disprose
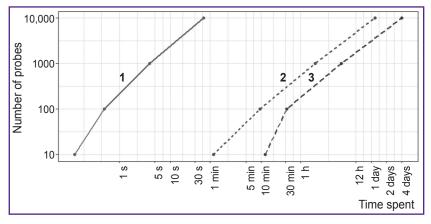
Figure 1. The sequence alignment using the BLAST algorithm: duration of the procedure plotted against the number of probes and the local base size
A specified number of probes designed to detect the sequences of *C. pneumoniae* was aligned with the contents of bases of different sizes: *1* — base of target sequences (0.02 Gb); *2* — base of nonspecific sequences (165 Gb); *3* — base of nonspecific sequences (466 Gb). The time spent on the alignment procedure without processing the results is shown. The logarithmic scale of the axes is used

(for example, at the same hybridization temperature), an important step is to determine the physicochemical properties of candidate probes. The disprose program provides the options for testing four physicochemical parameters allowing to control the conditions of hybridization, stability of the probe secondary structure: percentage of guanine and cytosine nucleotides (GC), the number of homo-repeats, melting temperature (T_m) estimate, and MFE (see the Table). In addition, the program provides the ability to change the computation parameters.

Testing the probe specificity. The algorithm implies a two-step specificity testing procedure. At the first step, the ability of a probe to hybridize to the target sequence is assessed by aligning the probe with the previously selected pathogen sequences using the BLAST algorithm. The results are then processed using specialized functions of the disprose package. During processing, for each probe, the number of target sequences, with which it got aligned under the required conditions (minimal alignment length, percentage of coverage, score, and E-value) is determined.

At the second step, the probes are tested for the specificity by aligning them with the sequences from the nonspecific pool.

As a result, for each probe, a list of target and nonspecific sequences, with which the probe can potentially interact is obtained. By analyzing both the number and nature of the specific and nonspecific interactions (with the help of the disprose functions), it becomes possible to select probes with specificity that is controlled *in silico*.

program), all associated sequences were downloaded from the NCBI Nucleotide collection database; those sequences formed the nonspecific database to be used further (8.7 million sequences with a total size of 165 Gb). Reducing the nonspecific database size made it possible to reduce around 5-fold the time spent later for testing the probe specificity (Figure 1).

Creating the candidate probe pool. The pool of candidate probes was formed by virtual slicing a user-selected "parent" sequence into segments of a specified length. The sequence for slicing is user-defined (a frequent choice is the pathogen reference genome) and can be presented as a FASTA file obtained from any source or downloaded directly from the NCBI bank.

Testing the physical and chemical properties of the probes. Since all probes in a DNA microarray must interact with target sequences under the same conditions

Relevant physicochemical parameters of the probes

Parameter	Impact on probe performance	Values (by default)	References
Probe size	As the length of the probe increases, its discrimination potential decreases, but the efficiency and the hybridization signal increase	24–32 nb	[1, 12, 13]
GC content	Impact on the melting point: a low GC content reduces the hybridization efficiency; a high GC content increases the likelihood of nonspecific hybridization	40–60%	[2, 14–17]
Number of homo-repeats (identical nucleotides repeated in a row)	More than four identical nucleotides in a row increase the likelihood of nonspecific binding	<5 nb	[18]
Minimal folding energy (MFE)	The lower the MFE value, the higher the likelihood of secondary structure stabilization by the probe, leading to a decrease in its sensitivity and the efficiency of hybridization	≥–3 kcal/mol	[7, 8, 13, 19]
Melting temperature	The main condition of the hybridization reaction determines the buffer solutions characteristics. The hybridization temperature is about 5° below the melting point	55–60°C	[10, 13, 16]

E.N. Filatova, A.S. Chaikina, N.F. Brusnigina, M.A. Makhova, O.V. Utkin

The final stage of the analysis. In the event that the selected probes are not specific for all target sequences, the analysis cycle is restarted. In this case, the sequences uncovered in the first cycle make a new target bank of a smaller size; again, a new pool of candidate probes is created based on a new "parent" sequence. The cycles are repeated until probes able to hybridize with the given target sequences of the pathogen are found.

Thus, the proposed algorithm for the selection of discriminating probes includes three main stages, performed sequentially:

1) composing the list of target and nonspecific sequences, creating the local banks of sequences;

2) generating a pool of candidate probes, checking their physicochemical parameters;

3) testing the ability of candidate probes to hybridize with target and nonspecific sequences.

A list of disprose program functions, providing the implementation of this algorithm, and their brief characteristics are presented in Appendix 2.

Additional features of the algorithm and its optimization

In addition to the main algorithm for the selection of specific probes, the disprose program implements additional functions: adding nucleotide adapters to probe sequences and annotating pathogen genome regions interacting with the probes. The possibility of using sequences from the whole genome sequencing projects — WGS (whole genome shotgun) is also available. Optimization of the algorithm performance for its successful implementation is among these additional features.

Working with WGS projects. WGS projects are incomplete assemblies of genomes or chromosomes of prokaryotes and eukaryotes; these sequences are termed "contigs". Including such contigs in the list of target sequences is problematic since each contig is considered an independent unit when aligned by the BLAST algorithm. Therefore, it would necessitate selecting a specific probe for each contig, giving rise to an excessive number of probes and increasing the time of the operation.

The proposed algorithm can be adjusted to WGS projects sequences through specialized functions of the disprose program, which allow considering all contigs of one genome as a single virtual sequence. In this case, probes that match one of the contigs are regarded as specific for the whole genome of the WGS project.

Optimizing the algorithm performance. As the selection of specific probes involves the screening of several millions of candidate sequences, it is of paramount importance to develop a high-performance algorithm. The disprose program utilizes standard techniques for increasing the algorithms' performance, such as storing intermediate data in the SQL database and running most of the functions in parallel (applicable for a multi-core server configuration). However, the main factor that determines the pace of computation is the order in which the program functions are applied.

As shown in Figures 2 and 3, different operations of this algorithm differ in their performance, with the maximum time spent on aligning the probes with sequences from local banks when testing the specificity. That is why we recommend testing the probe specificity at the very last stage when most of the probes have already been eliminated from the candidate list.

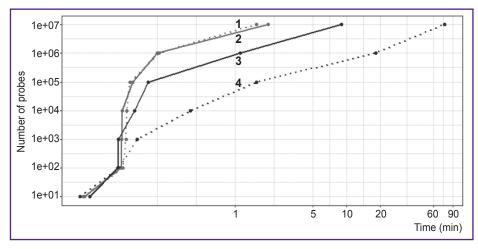The procedures for determining the physicochemical

Figure 2. The time spent plotted against the number of probes and the type of computation

The physicochemical parameters of probes for the detection of *C. pneumoniae* sequences were computed: *1* — test for the presence of homo-repeats; *2* — calculation of the GC percentage; *3* — calculation of the minimal folding energy; *4* — calculation of the melting point. The logarithmic scale of the axes is used
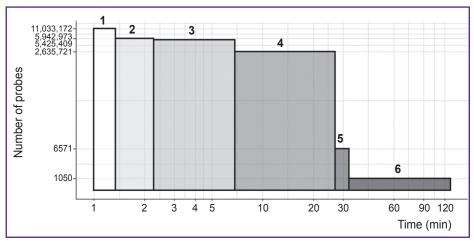
Figure 3. Time spent on the selection of probes for detecting *C. pneumoniae* sequences

The stages of computation are shown: *1* — calculation of the GC percentage; *2* — checking for the presence of nucleotide homo-repeats; *3* — calculating the minimal folding energy; *4* — calculating the melting point; *5* — checking the hybridization, aligning the probe sequences to the target sequence base; *6* — verifying the specificity, aligning the probe sequences to the nonspecific sequence base. When performing the procedures, the number of probes gradually decreased due to the elimination of probes with unacceptable characteristics. For alignment procedures, the times spent on the procedure itself and on the data processing are indicated. The logarithmic scale of the axes is used

parameters of the probes are also relatively slow. Each indicator is calculated at a different speed (see Figure 2). The stages of calculating the percentage of GC and homo-repeats seem to be the most productive and allow one to immediately exclude probes with knowingly unacceptable characteristics. These steps must be executed first. Thus, when selecting probes for the detection of *C. pneumonia,* we used sequential elimination of candidate probes that did not meet the GC presence criteria, the number of homo-repeats, and the MFE levels. This approach allowed us to reduce the number of candidate probes from 11.0 to 2.6 mln before arriving at the stage of calculating the T_m — the slowest operation of the entire process. This reduction saved us the time for calculating the T_m by 3.7-fold (20.1 instead of 75.2 min).

Search for specific probes capable of detecting Chlamydophila pneumoniae. We used the proposed algorithm and the disprose program to search for specific probes that allow the detection of *C. pneumoniae*. To create the list of specific sequences, metadata on the sequences available under the search query "Chlamydia pneumoniae * [organism] OR Chlamydophila pneumoniae * [organism]" were downloaded from the NCBI Nucleotide collection database. The resulting metadata set contained a total of 9062 records. According to further analysis, 17 whole genome sequences and 165 sequences of the WGS project were selected as targets. The target base had a size of 0.02 Gb. The list of target sequences is presented in

Appendix 3. The nonspecific sequence base was made up of previously selected sequences associated with human genetic material, representatives of its normal flora and microbiota.

The sequence "Chlamydia pneumoniae TW-183, complete sequence" (ID number NC_005043 NCBI RefSeq) was chosen as the "parent" sequence to create the candidate probe pool. The parent sequence was sliced into segments of all possible lengths from 24 to 32 nucleotide bases (nb), which constituted the pool of candidate probes (11,033,172 probes).

When testing the physicochemical properties, we used the following selection criteria: the content of G and C in the range of 40–60%, the absence of homo-repeats of 5 nb and longer, and the MEF — not less than 0 kcal/mol. In the end, the T_m value of the probes was calculated. Most probes had T_m values close to 57°C. Due to that, we were able to accelerate the selection process by reducing the number of probes while keeping only those with T_m within 56.97–57.03°C. Thus, the total number of candidate probes selected in such way for the next stage was 6571.

To test the ability of candidate probes to hybridize specifically with target sequences, the probes were aligned with the either target or nonspecific sequences using the BLAST algorithm. To obtain the most specific probes, we established the following conditions for efficient hybridization: for hybridization with the target sequences, the identity should be 100% in the absence of point mismatches and nucleotide gaps; for

E.N. Filatova, A.S. Chaikina, N.F. Brusnigina, M.A. Makhova, O.V. Utkin

hybridization with nonspecific sequences, the identity is required to be 50% or higher.

Of the candidate pool, 6380 probes effectively interacted with all target sequences *in silico.* To shorten the time required for detecting possible nonspecific hybridizations, the number of candidate probes was reduced to 1050 by narrowing the range of acceptable T_m values to 56.994–57.006°C. Probes that efficiently hybridized with at least one nonspecific sequence were excluded from the final pool.

The above operations of the algorithm resulted in 100 specific discriminating probes, which provided the differential detection of *C. pneumoniae* among other pathogens. Due to their high specificity, the selected probes can be used as a functional basis for DNA microarrays designed to identify actual causative agents of CAP. The total time for selecting the probes using the disprose program was 130 min. A list of selected probes and their characteristics is presented in Appendix 4.

Discussion

The proposed algorithm for selecting probes for a DNA microarray involves the identification of target sequences. The specificity and discriminatory potential of the probes (i.e. the ability of a DNA microarray to detect a specific pathogen) directly depend on the conservatism of the target sequence.

Probes for highly conserved regions of the genome, such as the *16S* and *23S* rRNA genes, although less specific, allow one to distinguish between bacteria belonging to different species. Probes for less conservative sequences, for example, bacterial genes *recA*, *gyrB*, *rpoB*, are able to distinguish between strains of bacteria within the same species [1, 12].

If there is no information on the degree of conservatism in the given genome, it is possible to search for these data using a set of target sequences and the method of their multiple alignment. However, multiple alignment of a large number of long sequences (for example, genomes) requires significant computing power and takes a long time [20].

In the disprose program, a different approach to searching for target regions is implemented; that is to generate a large number of short probes and align them into a set of sequences using the BLAST algorithm. This process is less demanding in terms of equipment, it is well implemented using a parallel mode, and it is ten times less time-consuming. Probes aligned to the full target sequence list with 100% coverage are considered to be specific to conserved regions of the sequence pool. Thus, varying the list of target sequences using the disprose program, makes it possible to select probes for designing DNA microarrays with different discriminatory potential.

We tested the developed algorithm by searching for specific probes that would allow for differential detection of *C. pneumoniae* in clinical samples. *C. pneumoniae* is one of the many microorganisms that cause CAP. The cumulative share of this and other "atypical" causative agents of CAP vary from 8 to 30% of cases [21, 22], and their timely detection can accelerate and improve the diagnostic process [23]. This application of the disprose program — the search for probes capable of detecting *C. pneumoniae* — demonstrated the potential of the developed algorithm.

As a result of the algorithm operation, one hundred probes with high specificity for the target pathogen were selected from the large-size pool of candidate probes. After comparing between the sites of origin of the probes and fragments of the annotated reference genome, we found that most of the probes originated from the genes that encoded for enzymes, chaperone proteins, and also regulators of the cell cycle (see Appendix 4). These genes contain regions highly conserved for *C. pneumoniae* and may be of interest not only for the diagnostic purpose but also for phylogenetic studies.

Conclusion

An algorithm has been developed to search for specific probes able to identify human pathogens of bacterial and viral origin. The algorithm is implemented as the disprose computer program written in the R language; its performance has been demonstrated by identifying the probes for detecting *C. pneumoniae*. The algorithm and program for the selection of probes have a number of advantages:

universality — the algorithm is aimed at finding specific areas in a set of sequences of any size and can be easily adapted to solve a wide range of tasks;

modularity — the execution of the algorithm occurs in several stages, their order is determined by the user, while any stage can be skipped or performed using a third-party software product;

openness — the initial code of the disprose package is publicly available and can be modified in accordance with the task;

usability — the algorithm can work with popular databases of genetic information (NCBI, GISAID) and local databases.

The flexibility and openness of the program provide for expansion of the scope of its application.

Conflicts of interest. The authors declare that they have no conflicts of interest.

References

1. Kostić T., Sessitsch A. Microbial diagnostic microarrays for the detection and typing of food- and water-borne (bacterial) pathogens. *Microarrays (Basel)* 2011; 1(1): 3–24, https://doi.org/10.3390/microarrays1010003.

2. Rouillard J.M., Zuker M., Gulari E. OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res* 2003; 31(12): 3057–3062, https://doi.org/10.1093/nar/gkg426.

3. Sung W.K., Lee W.H. Fast and accurate probe selection algorithm for large genomes. *Proc IEEE Comput Soc Bioinform Conf* 2003; 2: 65–74, https://doi.org/10.1109/csb.2003.1227305.

4. Urisman A., Fischer K.F., Chiu C.Y., Kistler A.L., Beck S., Wang D., DeRisi J.L. E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. *Genome Biol* 2005; 6(9): R78, https://doi.org/10.1186/gb-2005-6-9-r78.

5. Watson M., Dukes J., Abu-Median A.B., King D.P., Britton P. DetectiV: visualization, normalization and significance testing for pathogen-detection microarray data. *Genome Biol* 2007; 8(9): R190, https://doi.org/10.1186/gb-2007-8-9-r190.

6. National Center for Biotechnology Information. *Nucleotide*. Bethesda (MD): National Library of Medicine (US); 2021. URL: https://www.ncbi.nlm.nih.gov/nucleotide/.

7. Lorenz R., Bernhart S.H., Höner Zu Siederdissen C., Tafer H., Flamm C., Stadler P.F., Hofacker I.L. ViennaRNA Package 2.0. *Algorithms Mol Biol* 2011; 6(1): 26, https://doi.org/10.1186/1748-7188-6-26.

8. McCaskill J.S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 1990; 29(6–7): 1105–1119, https://doi.org/10.1002/bip.360290621.

9. Junhui L. *TmCalculator: melting temperature of nucleic acid sequences. R package version 1.0.1*. 2020. URL: https://CRAN.R-project.org/package=TmCalculator.

10. SantaLucia J. Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* 1998; 95(4): 1460–1465, https://doi.org/10.1073/pnas.95.4.1460.

11. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T.L. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; 10: 421, https://doi.org/10.1186/1471-2105-10-421.

12. Bodrossy L., Sessitsch A. Oligonucleotide microarrays in microbial diagnostics. *Curr Opin Microbiol* 2004; 7(3): 245–254, https://doi.org/10.1016/j.mib.2004.04.005.

13. Sanguin H., Herrera A., Oger-Desfeux C., Dechesne A., Simonet P., Navarro E., Vogel T.M., Moënne-Loccoz Y., Nesme X., Grundmann G.L. Development and validation of a prototype 16S rRNA-based taxonomic microarray for Alphaproteobacteria. *Environ Microbiol* 2006; 8(2): 289–307, https://doi.org/10.1111/j.1462-2920.2005.00895.x.

14. Maskos U., Southern E.M. A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support. *Nucleic Acids Res* 1993; 21(20): 4663–4669, https://doi.org/10.1093/nar/21.20.4663.

15. Raddatz G., Dehio M., Meyer T.F., Dehio C. PrimeArray: genome-scale primer design for DNA-microarray construction. *Bioinformatics* 2001; 17(1): 98–99, https://doi.org/10.1093/bioinformatics/17.1.98.

16. Wong C.W., Albert T.J., Vega V.B., Norton J.E., Cutler D.J., Richmond T.A., Stanton L.W., Liu E.T., Miller L.D. Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. *Genome Res* 2004; 14(3): 398–405, https://doi.org/10.1101/gr.2141004.

17. Wong C.W., Heng C.L.W., Wan Yee L., Soh S.W.L., Kartasasmita C.B., Simoes E.A.F., Hibberd M.L., Sung W.K., Miller L.D. Optimization and clinical validation of a pathogen detection microarray. *Genome Bio* 2007; 8(5): R93, https://doi.org/10.1186/gb-2007-8-5-r93.

18. Yoo S.M., Keum K.C., Yoo S.Y., Choi J.Y., Chang K.H., Yoo N.C., Yoo W.M., Kim J.M., Lee D., Lee S.Y. Development of DNA microarray for pathogen detection. *Biotechnol Bioprocess Engin* 2004; 9(2): 93–99, https://doi.org/10.1007/bf02932990.

19. Zuker M., Mathews D.H., Turner D.H. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski J., Clark B.F.C. (editors). *RNA biochemistry and biotechnology*. Springer; 1999; p. 11–43, https://doi.org/10.1007/978-94-011-4485-8_2.

20. Pais F.S.M., Ruy P.C., Oliveria G., Coimbra R.S. Assessing the efficiency of multiple sequence alignment programs. *Algorithms Mol Biol* 2014; 9(1): 4, https://doi.org/10.1186/1748-7188-9-4.

21. Rachina S.A., Bobylev A.A. Atypical pathogens of community-acquired pneumonia: epidemiology, diagnosis, and treatment. *Prakticeskaa pul'monologia* 2016; 2: 20–27.

22. Nair G.B., Niederman M.S. Updates on community acquired pneumonia management in the ICU. *Pharmacol Ther* 2021; 217: 107663, https://doi.org/10.1016/j.pharmthera.2020.107663.

23. Zaitsev A.A. Community-acquired pneumonia: diagnostic, treatment and vaccine prevention opportunities in the context of the COVID-19 pandemic. *Prakticeskaa pul'monologia* 2020; 1: 14–20.

E.N. Filatova, A.S. Chaikina, N.F. Brusnigina, M.A. Makhova, O.V. Utkin