

ISSN 2518-1726 (Online),  
ISSN 1991-346X (Print)



«ҚАЗАҚСТАН РЕСПУБЛИКАСЫ  
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ» РҚБ

# Х А Б А Р Л А Р Ы

---

---

**ИЗВЕСТИЯ**

РОО «НАЦИОНАЛЬНОЙ  
АКАДЕМИИ НАУК РЕСПУБЛИКИ  
КАЗАХСТАН»

**N E W S**

OF THE ACADEMY OF SCIENCES  
OF THE REPUBLIC OF  
KAZAKHSTAN

**PHYSICO-MATHEMATICAL SERIES**

**4 (352)**

**OCTOBER – DECEMBER 2024**

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

## **БАС РЕДАКТОР:**

**МУТАНОВ Ғалымқайыр Мұтанұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, ҚР БҒМ ҒК «Ақпараттық және есептеу технологиялары институты» бас директорының м.а. (Алматы, Қазақстан), **Н=5**

## **БАС РЕДАКТОРДЫҢ ОРЫНБАСАРЫ:**

**МАМЫРБАЕВ Өркен Жұмажанұлы**, ақпараттық жүйелер мамандығы бойынша философия докторы (Ph.D), ҚР БҒМ Ғылым комитеті «Ақпараттық және есептеуші технологиялар институты» РМК жауапты хатшысы (Алматы, Қазақстан), **Н=5**

## **РЕДАКЦИЯ АЛҚАСЫ:**

**ҚАЛИМОЛДАЕВ Мақсат Нұрәділұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі (Алматы, Қазақстан), **Н=7**

**БАЙГУНЧЕКОВ Жұмаділ Жанабайұлы**, техника ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Кибернетика және ақпараттық технологиялар институты, Сатпаев университетінің Қолданбалы механика және инженерлік графика кафедрасы, (Алматы, Қазақстан), **Н=3**

**ВОЙЧИК Вальдемар**, техника ғылымдарының докторы (физика), Люблин технологиялық университетінің профессоры (Люблин, Польша), **Н=23**

**БОШКАЕВ Қуантай Авғазыұлы**, Ph.D. Теориялық және ядролық физика кафедрасының доценті, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=10**

**QUEVEDO Nemando**, профессор, Ядролық ғылымдар институты (Мехико, Мексика), **Н=28**

**ЖҮСІПОВ Марат Абжанұлы**, физика-математика ғылымдарының докторы, теориялық және ядролық физика кафедрасының профессоры, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=7**

**КОВАЛЕВ Александр Михайлович**, физика-математика ғылымдарының докторы, Украина ҰҒА академигі, Қолданбалы математика және механика институты (Донецк, Украина), **Н=5**

**РАМАЗАНОВ Тілекқабұл Сәбитұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университетінің ғылыми-инновациялық қызмет жөніндегі проректоры, (Алматы, Қазақстан), **Н=26**

**ТАКИБАЕВ Нұрғали Жабағаұлы**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=5**

**ТИГИНЯНУ Ион Михайлович**, физика-математика ғылымдарының докторы, академик, Молдова Ғылым Академиясының президенті, Молдова техникалық университеті (Кишинев, Молдова), **Н=42**

**ХАРИН Станислав Николаевич**, физика-математика ғылымдарының докторы, профессор, ҚР ҰҒА академигі, Қазақстан-Британ техникалық университеті (Алматы, Қазақстан), **Н=10**

**ДАВЛЕТОВ Асқар Ербуланович**, физика-математика ғылымдарының докторы, профессор, әл-Фараби атындағы Қазақ ұлттық университеті (Алматы, Қазақстан), **Н=12**

**КАЛАНДРА Пьетро**, Ph.D (физика), Нанокұрылымды материалдарды зерттеу институтының профессоры (Рим, Италия), **Н=26**

**«ҚР ҰҒА Хабарлары. Физика және информатика сериясы».**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **№ 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *физика және ақпараттық коммуникациялық технологиялар сериясы.* Қазіргі уақытта: *«ақпараттық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*

Мерзімділігі: *жылына 4 рет.*

Тиражы: *300 дана.*

Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 219 бөл., тел.: 272-13-19*  
*http://www.physico-mathematical.kz/index.php/en/*

## ГЛАВНЫЙ РЕДАКТОР:

**МУТАНОВ Галимкаир Мутанович**, доктор технических наук, профессор, академик НАН РК, и.о. генерального директора «Института информационных и вычислительных технологий» КН МОН РК (Алматы, Казахстан), **H=5**

## ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:

**МАМЫРБАЕВ Оркен Жумажанович**, доктор философии (PhD) по специальности Информационные системы, ответственный секретарь РГП «Института информационных и вычислительных технологий» Комитета науки МОН РК (Алматы, Казахстан), **H=5**

## РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

**КАЛИМОЛДАЕВ Максат Нурадилович**, доктор физико-математических наук, профессор, академик НАН РК (Алматы, Казахстан), **H=7**

**БАЙГУНЧЕКОВ Жумадил Жанабаевич**, доктор технических наук, профессор, академик НАН РК, Институт кибернетики и информационных технологий, кафедра прикладной механики и инженерной графики, Университет Сагпаева (Алматы, Казахстан), **H=3**

**ВОЙЧИК Вальдемар**, доктор технических наук (физ.-мат.), профессор Люблинского технологического университета (Люблин, Польша), **H=23**

**БОШКАЕВ Куантай Авгазыевич**, доктор Ph.D, преподаватель, доцент кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=10**

**QUEVEDO Hemando**, профессор, Национальный автономный университет Мексики (UNAM), Институт ядерных наук (Мехико, Мексика), **H=28**

**ЖУСУПОВ Марат Абжанович**, доктор физико-математических наук, профессор кафедры теоретической и ядерной физики, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=7**

**КОВАЛЕВ Александр Михайлович**, доктор физико-математических наук, академик НАН Украины, Институт прикладной математики и механики (Донецк, Украина), **H=5**

**РАМАЗАНОВ Тлексабул Сабитович**, доктор физико-математических наук, профессор, академик НАН РК, проректор по научно-инновационной деятельности, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=26**

**ТАКИБАЕВ Нурғали Жабағевич**, доктор физико-математических наук, профессор, академик НАН РК, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=5**

**ТИГИНЯНУ Ион Михайлович**, доктор физико-математических наук, академик, президент Академии наук Молдовы, Технический университет Молдовы (Кишинев, Молдова), **H=42**

**ХАРИН Станислав Николаевич**, доктор физико-математических наук, профессор, академик НАН РК, Казахстанско-Британский технический университет (Алматы, Казахстан), **H=10**

**ДАВЛЕТОВ Аскар Ербуланович**, доктор физико-математических наук, профессор, Казахский национальный университет им. аль-Фараби (Алматы, Казахстан), **H=12**

**КАЛАНДРА Пьетро**, доктор философии (Ph.D, физика), профессор Института по изучению наноструктурированных материалов (Рим, Италия), **H=26**

## «Известия НАН РК. Серия физика и информатики».

ISSN 2518-1726 (Online),

ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **№ 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика и информационные коммуникационные технологии.* В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 219, тел.: 272-13-19*

*<http://www.physico-mathematical.kz/index.php/en/>*

#### **EDITOR IN CHIEF:**

**MUTANOV Galimkair Mutanovich**, doctor of technical Sciences, Professor, Academician of NAS RK, acting director of the Institute of Information and Computing Technologies of SC MES RK (Almaty, Kazakhstan), **H=5**

#### **DEPUTY EDITOR-IN-CHIEF**

**MAMYRBAYEV Orken Zhumazhanovich**, Ph.D. in the specialty "Information systems, executive secretary of the RSE "Institute of Information and Computational Technologies", Committee of Science MES RK (Almaty, Kazakhstan) **H=5**

#### **EDITORIAL BOARD:**

**KALIMOLDAYEV Maksat Nuradilovich**, doctor in Physics and Mathematics, Professor, Academician of NAS RK (Almaty, Kazakhstan), **H=7**

**BAYGUNCHEKOV Zhumadil Zhanabayevich**, doctor of Technical Sciences, Professor, Academician of NAS RK, Institute of Cybernetics and Information Technologies, Department of Applied Mechanics and Engineering Graphics, Satbayev University (Almaty, Kazakhstan), **H=3**

**WOICIK Waldemar**, Doctor of Phys.-Math. Sciences, Professor, Lublin University of Technology (Lublin, Poland), **H=23**

**BOSHKAYEV Kuantai Avgazievich**, PhD, Lecturer, Associate Professor of the Department of Theoretical and Nuclear Physics, Al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=10**

**QUEVEDO Hemando**, Professor, National Autonomous University of Mexico (UNAM), Institute of Nuclear Sciences (Mexico City, Mexico), **H=28**

**ZHUSSUPOV Marat Abzhanovich**, Doctor in Physics and Mathematics, Professor of the Department of Theoretical and Nuclear Physics, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=7**

**KOVALEV Alexander Mikhailovich**, Doctor in Physics and Mathematics, Academician of NAS of Ukraine, Director of the State Institution «Institute of Applied Mathematics and Mechanics» DPR (Donetsk, Ukraine), **H=5**

**RAMAZANOV Tlekkabul Sabitovich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Vice-Rector for Scientific and Innovative Activity, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=26**

**TAKIBAYEV Nurgali Zhabagaevich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=5**

**TIGHINEANU Ion Mikhailovich**, Doctor in Physics and Mathematics, Academician, Full Member of the Academy of Sciences of Moldova, President of the AS of Moldova, Technical University of Moldova (Chisinau, Moldova), **H=42**

**KHARIN Stanislav Nikolayevich**, Doctor in Physics and Mathematics, Professor, Academician of NAS RK, Kazakh-British Technical University (Almaty, Kazakhstan), **H=10**

**DAVLETOV Askar Erbulanovich**, Doctor in Physics and Mathematics, Professor, al-Farabi Kazakh National University (Almaty, Kazakhstan), **H=12**

**CALANDRA Pietro**, PhD in Physics, Professor at the Institute of Nanostructured Materials (Monterotondo Station Rome, Italy), **H=26**

#### **News of the National Academy of Sciences of the Republic of Kazakhstan.**

**Series of physics and informatics.**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-ЖК**, issued 14.02.2018  
Thematic scope: *series physics and information technology.*

Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*

Periodicity: *4 times a year.*

Circulation: *300 copies.*

Editorial address: *28, Shevchenko str., of. 219, Almaty, 050010, tel. 272-13-19*

*<http://www.physico-mathematical.kz/index.php/en/>*

УДК 004.4:577.21

©**Y. Golenko**<sup>1\*</sup>, **A. Ismailova**<sup>1</sup>, **K. Kadirkulov**<sup>1</sup>, **R. Kalendar**<sup>2</sup>, 2024.

<sup>1</sup>S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan;

<sup>2</sup> University of Helsinki, Helsinki, Finland.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## **DEVELOPMENT OF AN ONLINE PLATFORM FOR SEARCHING FOR TANDEM REPEATS USING WHOLE GENOME SEQUENCING**

**Golenko Yekaterina** – Ph.D, S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan, E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com), <https://orcid.org/0000-0002-4643-4571>;

**Ismailova Aisulu** – Ph.D., associate professor, Information Systems Department, S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan, E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru), <https://orcid.org/0000-0002-8958-1846>;

**Kadirkulov Kuanysh** – Ph.D, S. Seifullin Kazakh Agrotechnical Research University, Astana, Kazakhstan, E-mail: [kkuanysh@gmail.com](mailto:kkuanysh@gmail.com), <https://orcid.org/0000-0003-0506-4890>;

**Kalendar Ruslan** – Ph.D, Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland, E-mail: [ruslan.kalendar@helsinki.fi](mailto:ruslan.kalendar@helsinki.fi), <https://orcid.org/0000-0003-3986-2460>.

**Abstract.** The article presents the results of developing an online platform for identifying tandem repeats in whole-genome sequencing, the detection of which is critical for many areas of genetics, forensics, and medicine. The development of specialized software for analyzing whole-genome sequencing data is of great importance for biological and medical research, but identifying these repeats is an extremely difficult task due to their high variability and complexity, and traditional methods often encounter difficulties in analyzing tandem repeats. The study provides an analysis and discussion of existing methods for identifying tandem repeats, outlines their advantages and disadvantages, and considers prospects and directions for further research in this area. Along with the description and application of the traditional approach to processing sequencing data to search for tandem repeats, the proposed software uses a specially developed algorithm that allows identifying all types of short tandem repeats. The platform is implemented using a client-server architecture, includes user authorization, pages for setting parameters for calculating tandem repeats, displaying and saving the history of results and viewing them in text and visualized formats.

**Keywords:** tandem repeats, whole genome sequencing, online platform, satellites, microsatellites.

©Е.С. Голенко<sup>1\*</sup>, А.А. Исмаилова<sup>1</sup>, К.К. Кадиркулов<sup>1</sup>, Р.Н. Календарь<sup>2</sup>, 2024.

<sup>1</sup>С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті,  
Астана, Қазақстан;

<sup>2</sup>Хельсинки университеті, Хельсинки, Финляндия.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## ТОЛЫҚ ГЕНОМДЫҚ СЕКВЕНИРЛЕУДЕ ТАНДЕМДІК ҚАЙТАЛАНУЛАРДЫ ІЗДЕУ ҮШІН ОНЛАЙН ПЛАТФОРМАСЫН ӘЗІРЛЕУ

**Голенко Екатерина Сергеевна** – Ph.D., С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан, E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com), <https://orcid.org/0000-0002-4643-4571>;

**Исмаилова Айсулу Абжаппаровна** – Ph.D., қауымдастырылған профессор, Ақпараттық жүйелер кафедрасы, С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан, E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru), <https://orcid.org/0000-0002-8958-1846>;

**Кадиркулов Қуаныш Кайсарович** – Ph.D., С. Сейфуллин атындағы Қазақ агротехникалық зерттеу университеті, Астана, Қазақстан, E-mail: [kkuanysh@gmail.com](mailto:kkuanysh@gmail.com), <https://orcid.org/0000-0003-0506-4890>;

**Календарь Руслан Николаевич** – Ph.D, Хельсинки өмір туралы ғылымдар институты, Хельсинки университеті, Хельсинки, Финляндия, E-mail: [ruslan.kalendar@helsinki.fi](mailto:ruslan.kalendar@helsinki.fi), <https://orcid.org/0000-0003-3986-2460>.

**Аннотация.** Мақалада генетиканың, криминалистиканың және медицинаның көптеген салаларында анықтау өте маңызды болып табылатын тұтас геномды секвенирлеу кезінде тандемді қайталауларды анықтауға арналған онлайн платформаны әзірлеу нәтижелері берілген. Тұтас геномды секвенирлеу деректерін талдауға арналған арнайы бағдарламалық жасақтаманы әзірлеу биологиялық және медициналық зерттеулер үшін үлкен маңызға ие, бірақ бұл қайталануларды анықтау олардың жоғары өзгергіштігі мен күрделілігіне байланысты өте қиын, ал дәстүрлі әдістер тандемді қайталауларды талдау кезінде жиі қиындықтарға тап болады. Зерттеуде тандемді қайталауларды анықтаудың қолданыстағы әдістерін талдау және талқылау, олардың артықшылықтары мен кемшіліктері көрсетілген, сондай-ақ осы саладағы одан әрі зерттеудің перспективалары мен бағыттары талқыланады. Тандемді қайталауларды іздеу үшін секвенирлеу деректерін өңдеудің дәстүрлі тәсілін сипаттау және қолданумен қатар ұсынылып отырған бағдарламалық қамтамасыз ету қысқа тандемді қайталаулардың барлық түрлерін анықтауға мүмкіндік беретін арнайы әзірленген алгоритмді пайдаланады. Платформа клиент-сервер архитектурасын қолдану арқылы жүзеге асырылады, пайдаланушы авторизациясын, қайталанатын есептеулер үшін параметрлерді орнатуға, нәтижелер тарихын көрсететін және сақтауға және оларды мәтіндік және визуалды пішімдерде қарауға арналған беттерді қамтиды.

**Түйін сөздер:** тандемді қайталау, тұтас геномды секвенирлеу, онлайн платформа, сателиттер, микросателиттер.



©Е.С. Голенко<sup>1\*</sup>, А.А. Исмаилова<sup>1</sup>, К.К. Кадиркулов<sup>1</sup>, Р.Н. Календарь<sup>2</sup>, 2024.

<sup>1</sup>Казахский агротехнический исследовательский университет  
им. С. Сейфуллина, Астана, Казахстан;

<sup>2</sup>Университет Хельсинки, Хельсинки, Финляндия.

E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com)

## **РАЗРАБОТКА ОНЛАЙН-ПЛАТФОРМЫ ДЛЯ ПОИСКА ТАНДЕМНЫХ ПОВТОРОВ ПРИ ПОЛНОГЕНОМНОМ СЕКВЕНИРОВАНИИ**

**Голенко Екатерина Сергеевна** – PhD, Казахский агротехнический исследовательский университет им. С. Сейфуллина, Астана, Казахстан, E-mail: [golenko.katerina@gmail.com](mailto:golenko.katerina@gmail.com), <https://orcid.org/0000-0002-4643-4571>;

**Исмаилова Айсулу Абжаппаровна** – PhD, ассоциированный профессор, кафедра Информационных систем, Казахский агротехнический исследовательский университет им. С. Сейфуллина, Астана, Казахстан, E-mail: [a.ismailova@mail.ru](mailto:a.ismailova@mail.ru), <https://orcid.org/0000-0002-8958-1846>;

**Кадиркулов Куаныш Кайсарович** – PhD, Казахский агротехнический исследовательский университет им. С. Сейфуллина, Астана, Казахстан, E-mail: [kkuanysh@gmail.com](mailto:kkuanysh@gmail.com), <https://orcid.org/0000-0003-0506-4890>;

**Календарь Руслан Николаевич** – PhD, Хельсинкский институт наук о жизни, Университет Хельсинки, Хельсинки, Финляндия, E-mail: [ruslan.kalendar@helsinki.fi](mailto:ruslan.kalendar@helsinki.fi), <https://orcid.org/0000-0003-3986-2460>.

**Аннотация.** В статье представлены результаты разработки онлайн-платформы для идентификации tandemных повторов при полногеномном секвенировании, обнаружение которых является критически важным для многих областей генетики, криминалистики и медицины. Разработка специализированного программного обеспечения для анализа данных полногеномного секвенирования имеет огромное значение для биологических и медицинских исследований, однако идентификация этих повторов является крайне сложной задачей из-за их высокой вариабельности и сложности, а традиционные методы часто сталкиваются с трудностями при анализе tandemных повторов. В исследовании приведены анализ и обсуждение существующих методов идентификации tandemных повторов, обозначены их преимущества и недостатки, а также рассмотрены перспективы и направления дальнейших исследований в этой области. Наряду с описанием и применением традиционного подхода к обработке данных секвенирования для поиска tandemных повторов в предлагаемом программном обеспечении используется специально разработанный алгоритм, позволяющий идентифицировать все типы коротких tandemных повторов. Платформа реализована с использованием клиент-серверной архитектуры, включает в себя авторизацию пользователя, страницы задания параметров вычислений повторов, отражение и сохранение истории результатов и их просмотр в текстовом и визуализированном форматах.

**Ключевые слова:** тандемные повторы, полногеномное секвенирование, онлайн-платформа, спутники, микроспутники.

*В качестве направлений будущих исследований планируется дальнейшая разработка предложенной платформы. Настоящая работа проводится при финансовой поддержке Министерства науки и высшего образования Республики Казахстан (AP19678041 «Разработка программного обеспечения для идентификации тандемных повторов при полногеномном секвенировании»).*

**Введение.** В последние годы полногеномное секвенирование (Whole Genome Sequencing – WGS) стало мощным инструментом в области геномики, предоставляя детальную информацию о генетическом материале организмов. Одной из ключевых задач анализа данных WGS является идентификация тандемных повторов (Anisimova, 2015), представляющие собой последовательности, которые состоят из нескольких последовательных повторов одного и того же паттерна в одном направлении. В зависимости от длины этого повторяющегося паттерна тандемные повторы классифицируются на три типа: микроспутники (повторяющийся паттерн менее 6 пар нуклеотидов), также известные как Short Tandem Repeats – STR, миниспутники (повторяющийся паттерн от 7 до 100 пар нуклеотидов) и спутники (повторяющийся паттерн более 100 пар нуклеотидов). Эти повторы играют важную роль в генетической вариации, эволюции и могут быть связаны с различными заболеваниями. Благодаря сложной изменчивости и высокой способности к дискриминации тандемные повторы широко используются в популяционном генетическом анализе (Wang, et al, 2016), судебно-медицинской идентификации (Chiu, et al, 2021) и селекции (Eichler, 2019). Кроме того, известно, что вариации STR связаны с нервными заболеваниями, такими как болезнь Альцгеймера, ожирение и рак, посредством регуляции проксимальной экспрессии генов (Bakhtiari, 2021). Кроме того, тандемные повторы являются основными маркерами в криминалистических приложениях ДНК и используются почти во всех криминалистических базах данных ДНК.

Тандемные повторы могут значительно различаться по длине и структуре, что делает их идентификацию сложной задачей. Традиционные методы анализа последовательностей часто сталкиваются с проблемами при обработке таких повторов из-за их высокой вариабельности и сложности. В этом контексте разработки программного обеспечения, специально предназначенного для идентификации тандемных повторов, являются актуальной задачей. Современные подходы к анализу тандемных повторов включают использование различных алгоритмов и методов машинного обучения, которые позволяют эффективно обрабатывать большие объемы данных, полученные в результате WGS. Однако разработка таких инструментов требует глубокого понимания как биологических аспектов, так и компьютерных методов обработки данных.



Разработка специализированного программного обеспечения для анализа данных полногеномного секвенирования имеет огромное значение для биологических и медицинских исследований. Идентификация tandemных повторов может способствовать выявлению генетических маркеров заболеваний, пониманию эволюционных процессов и улучшению методов диагностики. В данной статье представлены промежуточные результаты разработки программного обеспечения и онлайн-платформы для идентификации tandemных повторов при полногеномном секвенировании. Обсуждаются существующие методы, их преимущества и недостатки, а также перспективы и направления дальнейших исследований в этой области.

Во многих исследованиях и практиках используются длины аллелей tandemных повторов, тогда как подробная последовательность аллелей игнорируется. Это функционально определенное обозначение обусловлено ограничениями традиционных технологий, с помощью которых варианты tandemных повторов обнаруживаются с помощью секвенирования по Сэнгеру или путем измерения длин фрагментов ДНК во время разделения с помощью капиллярного электрофореза. Аллели tandemных повторов одинаковой длины рассматриваются как одни и те же аллели, хотя они могут иметь разные последовательности. Более высокое разрешение аллелей tandemных повторов может быть важно для широкого спектра применений и в настоящее время не полностью учтено. Аллели tandemных повторов можно сообщать как варианты последовательностей или гаплотипы с использованием технологий секвенирования следующего поколения с более высокой достоверностью и меньшими затратами на пару оснований, чем традиционные методы. Для обнаружения гаплотипов из наборов данных последовательностей были разработаны биоинформатические инструменты, такие как STRait Razor (King, et al, 2021), HipSTR (Willems, et al, 2017) и FDSTools (Hoogenboom, et al, 2017). Эти программы могут обнаруживать аллели как на основе длины, так и на основе последовательности. Каждый гаплотип STR (т.е. последовательность) содержит информацию, такую как количество повторов основного мотива и дополнительные точечные мутации, такие как однонуклеотидный полиморфизм и инсерции/делеции, если они присутствуют. Однако прямо или визуально выявить различия между гаплотипами сложно из-за их повторяющегося характера, особенно для сложных гаплотипов. Кроме того, некоторые повторные расширения tandemных повторы, связанных с заболеванием, могут быть очень длинными и содержать несколько типов вариантов, что может еще больше усложнить сравнение. Множественные инструменты выравнивания последовательностей, такие как MAFFT, могут сравнивать очень похожие последовательности и выявлять различия между последовательностями (Nakamura, et al, 2018). Однако эти инструменты обычно разрабатываются для общих целей сравнения.

С другой стороны, недавние достижения в области алгоритмов,

разработанных специально для генотипирования STR, таких как пакет программного обеспечения ExpansionHunter (Dolzhenko, et al, 2020), TexSTRa (Fearnley, et al, 2022), STRetch (Dashnow, et al, 2018) и superSTR (Tang, et al, 2017) показали, что WGS может обеспечить надежное обнаружение tandemных повторов по всему геному с высокой чувствительностью и высокой специфичностью. Например, WGS в сочетании с ExpansionHunter выявил ранее не диагностированные неврологические заболевания, он был первым инструментом, разработанным для поиска по всему геному новых расширений повторов в данных короткого чтения. Эти инструменты могут обрабатывать как выровненные, так и невыровненные чтения, чтобы идентифицировать возможные tandemные повторы в известных и охарактеризованных локусах. Также совсем недавно был разработан STRling с аналогичной новой возможностью обнаружения tandemных повторов (Dashnow, et al, 2022).

Несмотря на успехи в идентификации tandemных повторов, существуют определённые ограничения текущих методов. Одним из основных недостатков является их неспособность учитывать точную последовательность аллелей, фокусируясь в основном на длине повторов. Это приводит к тому, что различные последовательности одинаковой длины рассматриваются как идентичные, что не позволяет выявить потенциальные биологические различия между ними. Более того, сложности возникают при анализе длинных и сложных гаплотипов, что требует разработки более точных и специализированных инструментов для их идентификации и сравнения.

Современные методы, такие как использование капиллярного электрофореза и секвенирование по Сэнгеру, ограничены в разрешении и точности, что требует внедрения технологий следующего поколения. Секвенирование с высоким разрешением и новые алгоритмы генотипирования STR, такие как ExpansionHunter, демонстрируют большие перспективы, однако они также нуждаются в дальнейших улучшениях для повышения достоверности и снижения затрат во времени обработки. Разработка новых инструментов, способных эффективно идентифицировать и анализировать tandemные повторы, остаётся актуальной задачей в геномике, что позволит улучшить диагностику и понимание генетических заболеваний.

**Материалы и методы.** В мировой практике общая схема идентификации tandemных повторов выглядит следующим образом: этап подготовки данных – файл с геномной последовательностью (например, в формате FASTA); этап определения параметров поиска – задание параметров, необходимых для идентификации повторов; этап поиска повторяющихся мотивов – проход по последовательности с заданными параметрами; этап аннотации и вывода – описание найденных повторов с указанием их местоположения, длины, числа копий и других характеристик, а также сохранение или отображение результатов в формате, нацеленном на конечного пользователя (таблица, отчет и т.д.) (Рис. 1).

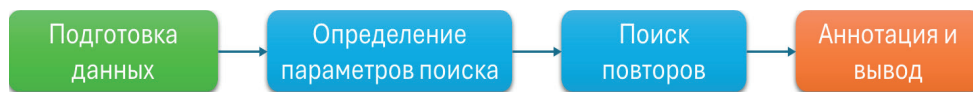


Рисунок 1. Процесс идентификации tandemных повторов

Предлагаемая платформа основывается на представленной выше методике и предназначена для поиска tandemных повторов в белковых и ДНК последовательностях. Общая схема вычислений, работы онлайн-платформы, а также взаимодействия клиентской и серверной части представлена на Рис. 2.

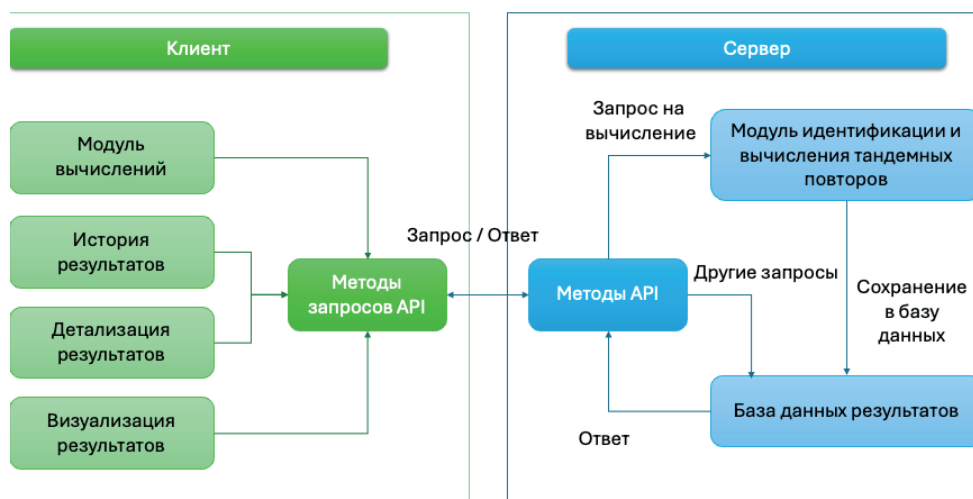


Рисунок 2. Процесс идентификации tandemных повторов

Платформа состоит из двух основных частей – пользовательского (клиентского) интерфейса и серверной части. Серверная часть производит вычисления по поиску tandemных повторов с формированием базы данных результатов. В основе серверной логики заложены методы Rest API, которые позволяют удаленным клиентам производить вычисления. Платформа реализована на базе операционной системы Ubuntu v.22 с использованием языков программирования Java и php. Запросы и результаты формируются в формате JSON.

Клиентская часть представляет собой онлайн систему, которая позволяет вести статистику расчетов пользователей, которые в дальнейшем могут использоваться для сравнений результатов. Онлайн система временно размещена по адресу <https://gene.kkk.kz/>. Доступ осуществляется только с помощью вызова методов калькуляции, через двухфакторную авторизацию и получением токена. Техническая реализация пользовательского интерфейса произведена с использованием HTML5, jQuery, Bootstrap и PHP8 и других web-технологий.

Описываемая платформа используют исходный код идентификации и визуализации перемежающихся и tandemных повторов на уровне всего генома (<https://github.com/rkalendar/Repeater>). Вычисления производятся путем ввода следующих параметров, описанных в Таблице 1.

Таблица 1 – Параметры для анализа

№ п/п	Параметр	Описание
1	ssr	Анализ только локусов SSR/теломеров
2	kmer	Минимальная k-мера, длина подстроки, содержащиеся в биологической последовательности
3	min	Начальная длина повтора
4	sln	Длина строки
5	image	Размерность изображения на выходе вычисления
6	Flanks	Расширение флангов повтора на соответствующую длину
7	Mask	Формирование нового файла с маскирующими повторами
8	Seqshow	Извлечение последовательностей повторов
9	Quick	Признак быстрого анализ повторов, без глубокого анализа и их кластеризации
10	File	Файл для анализа в текстовом формате

Для проведения корректной идентификации повторов критично важно задать правильные параметры, которые позволят с наиболее высокой точностью производить вычисления и выявлять tandemные повторы. Пример задаваемых для расчетов параметров представлен в Таблице 2.

Таблица 2 – Примеры задаваемых параметров для анализа

№ п/п	Примеры параметров
1	kmer=21, min=30
2	ssr=true, seqshow=true, flanks=100
3	kmer=21, min=100, sln=250, image=5000x3000, quick=false, mask=false, seqshow=true

Как видно в представленных примерах, не все параметры анализа являются обязательными для задания их пользователем, однако при использовании более точных настроек платформа демонстрирует лучшие результаты по сравнению с запросами с неопределенными параметрами.

Отличительной особенностью предлагаемой платформы, определяющей ее новизну, является использование специально разработанного алгоритма, который позволяет идентифицировать все типы повторяющихся последовательностей, включая совершенные и несовершенные микросателлитные повторы, а также любые типы коротких tandemных повторов. Эти повторы могут принадлежать к широкому спектру, организованному в повторяющиеся структуры более высокого порядка, такие как крупные сателлитные последовательности и теломеры. Программа представляет собой высокочувствительный и автоматизированный метод для идентификации повторяющихся последовательностей.

**Результаты и обсуждение.** Онлайн-платформа была протестирована на хромосомах животных, насекомых, растений, а также на последовательностях геномов прокариот и гигантских вирусов. Результаты показали, что она является быстрым, эффективной и простой в использовании, предлагая удобный интерфейс. Кроме того, программа позволяет получить информацию о расположении и распределении этих повторов в геноме, что может помочь исследователям выявить потенциальные регуляторные регионы или области, подверженные генетической нестабильности.

Основные модули, входящие в функционал платформы представлены ниже. На Рис. 3 представлен модуль формирования вычислений, в котором пользователь определяет параметры вычисления, ранее описанные в Таблице 1. После определения всех необходимых параметров происходит этап вычисления, после которого пользователь может переходить в историю результатов (Рис. 4).

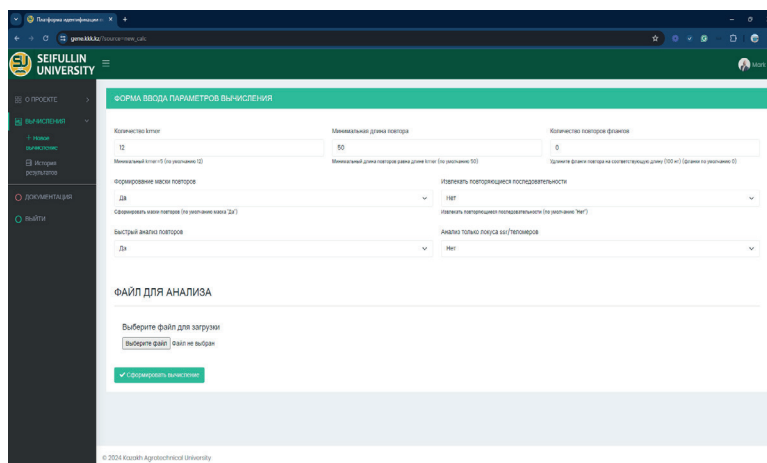


Рисунок 3. Модуль формирования вычислений

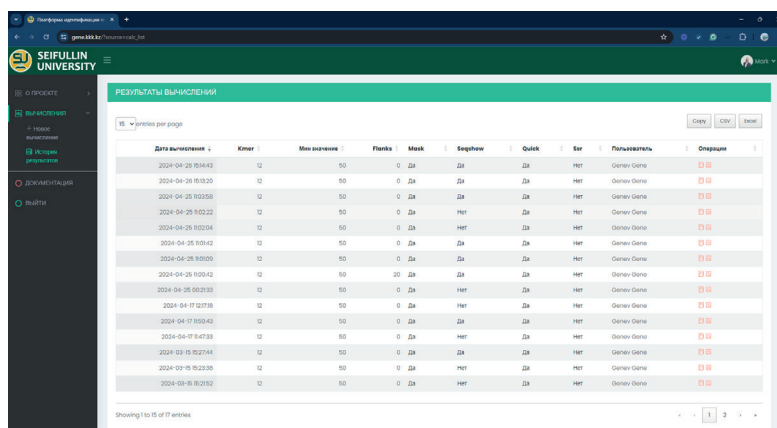


Рисунок 4. История результатов

История результатов отображает все вычисления, произведенные пользователем, что позволяет вести учет расчетов. Каждая история содержит детализированные результаты, отображенные на Рис. 5.

Результаты вычислений

10 entries per page

Copy CSV Excel

Cluster ID	Start position	End position	Length	str_sequence
1	2	126	125	ttaataacttaggcaattctggatcgagctaaagtcctccgagtgccgagcttaagtgcaacctccgaagaagtttaataacttaggcgattc
1	144	197	54	ttaataacttaggcgattctggatcgagctaaagtcctccgagcgcgagcttaag
1	192231	-192172	60	ttaatacttaggcgattctggatcgagctaaagtcctccgagtgfgaactttaggccaca
2	192334	-192197	138	cgagtgccgacgttaacttcagactcgggagaagttgtaacttaggcgattgtgtcgtcgtcgaactcctccgagtgcaaccttagtgcc
2	529001	-528864	138	cgagtgccgacgttagtgcaacctcgggagaaggtgtcacttaggcgattgtgtcgtcgtcgaactcctccgagtgcaaccttagtgcc
3	156299	156351	53	gcgattctggatgatggaaggggttccaatcggggcgtaaatatagggc
3	518716	-518664	53	gcgattctggatgatggaaggggttccaatcggggcgtaaatatagggc
4	4953	5553	601	tgtgttgfgaacattgtgtaaggaccacacatgcacgtgggataaccagccaaagcaatttgcggaagcaattgtcacacaagctc
4	12745	-12145	601	tgtgttgfgaacattgtgtaaggaccacacatgcacgtgggataaccagccaaagcaatttgcggaagcaattgtcacacaagctc
5	198	800	603	gtgttgfgaacattgtgtaaggaccacacatgcacgtgggataaccagccaaagcaatttgcggaagcaattgtcacacaagctc

Showing 1 to 10 of 5,052 entries

Закрывать

Рисунок 5. Детализация результатов

Помимо детализации текстовой части вычислений, предлагаемая платформа дает возможность получать визуальное отображение результатов, где наглядно можно увидеть, как и где были идентифицированы повторы.

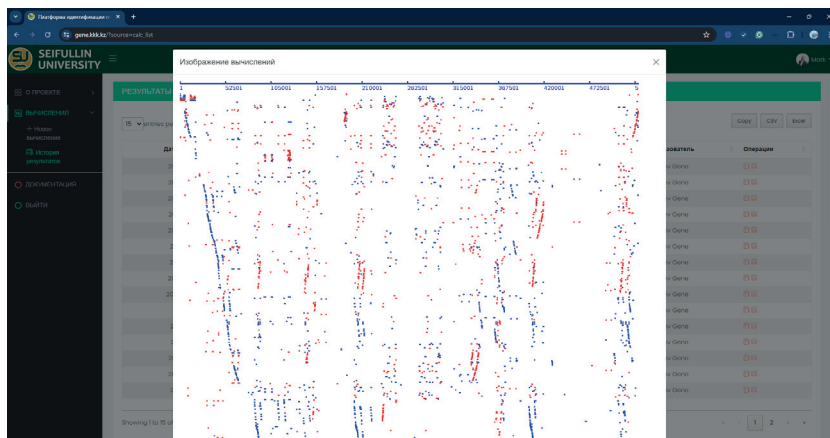


Рисунок 6. Визуализация результатов

Как видно из Рис.6 онлайн-платформа позволяет получить информацию о расположении и распределении тандемных повторов в геноме, что может помочь исследователям выявить потенциальные регуляторные регионы или области, подверженные генетической нестабильности.



**Выводы.** В работе предложена онлайн-платформа для идентификации тандемных повторов при полногеномном секвенировании. Тандемные повторы в биологии играют значительную роль при выявлении генетических маркеров заболеваний, понимании эволюционных процессов и улучшении методов диагностики. Онлайн-платформа для поиска тандемных повторов позволяет обнаруживать элементы прямого и инвертированного повтора, совершенные и несовершенные микросателлитные повторы, а также любые типы коротких и длинных тандемных повторов. Сочетая высокую точность и универсальность, инструмент вносит значительный вклад в понимание сложного ландшафта геномных повторов. Он позволяет исследователям детально анализировать и выявлять различные типы повторяющихся последовательностей, что способствует более глубокому пониманию генетических механизмов и улучшению диагностики генетических заболеваний.

### References

- Anisimova, M., Pecerska, J., Schaper, E. (2015). Statistical Approaches to Detecting and Analyzing Tandem Repeats in Genomic Sequences. *Frontiers in Bioengineering and Biotechnology*. – 3. – 1-6.
- Bakhtiari, M., Park, J., Ding, Y.-C., Shleizer-Burko, S., Neuhausen, S.L., Halldorrsson, B.V., Stefansson, K., Gymrek, M., Bafna, V. (2021). Variable number tandem repeats mediate the expression of proximal genes. *Nature Communications*. – 12(1). – 2075.
- Chiu, R., Rajan-Babu, I.-S., Friedman, J.M., Birol, I. (2021). Straglr: discovering and genotyping tandem repeat expansions using whole genome long-read sequences. *Genome Biology*. – 22(1). – 224.
- Dashnow, H., Lek, M., Phipson, B., Halman, A., Sadedin, S., Lonsdale, A., et al. (2018). STRetch: detecting and discovering pathogenic short tandem repeat expansions. *Genome Biology*. – 19. – 121.
- Dashnow, H., Pedersen, B.S., Hiatt, L., Brown, J., Beecroft, S.J., Ravenscroft, G., et al. (2022). STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci. *Genome Biology*. – 23. – 257.
- Dolzhenko, E., Bennett, M.F., Richmond, P.A., et al. (2020). ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biology*. – 21(1). – 102.
- Eichler, E.E. (2019). Genetic variation, comparative genomics, and the diagnosis of disease. *The New England Journal of Medicine*. – 381(1). – 64-74.
- Fearnley, L., Bennett, M., Bahlo, M. (2022). Detection of repeat expansions in large next generation DNA and RNA sequencing data without alignment. *Scientific Reports*. – 12. – 13124.
- Hoogenboom, J., van der Gaag, K.J., de Leeuw, R.H., Sijen, T., de Knijff, P., Laros, J.F.J. (2017). FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. *Forensic Science International: Genetics*. – 27. – 27-40.
- King, J.L., Woerner, A.E., Mandape, S.N., Kapema, K.B., Moura-Neto, R.S., Silva, R., Budowle, B. (2021). STRait razor online: an enhanced user interface to facilitate interpretation of MPS data. *Forensic Science International: Genetics*. – 52.
- Nakamura, T., Yamada, K.D., Tomii, K., Katoh, K. (2018). Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*. – 34(14). – 2490-2.
- Tang, H., Kirkness, E.F., Lippert, C., Biggs, W.H., Fabani, M., Guzman, E., et al. (2017). Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes. *The American Journal of Human Genetics*. – 101. – 700-715.
- Wang, X., Wang, L. (2016). GMATA: an integrated software package for Genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science*. – 7. – 1350.
- Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., Erlich, Y. Genome-wide profiling of heritable and de novo STR variations. *Natural Methods*. – 14(6). – 590-2.

## CONTENTS

### INFORMATION AND COMMUNICATION TECHNOLOGIES

<b>M. Aitimov, R.U Almenayeva, K.K. Makulov, A.B. Ostayeva, R. Muratkhan</b> APPLICATION OF MACHINE LEARNING METHOD TO ANALYZE AND EXTRACT SEMANTIC STRUCTURES FROM SCIENTIFIC TEXTS.....	5
<b>A.K. Aitim, G.K. Sembina</b> MODELING OF HUMAN BEHAVIOR FOR SMARTPHONE WITH USING MACHINE LEARNING ALGORITHM.....	17
<b>G. Aksholak, A. Bedelbayev, R. Magazov</b> ANALYSIS AND COMPARISON OF MACHINE LEARNING METHODS FOR MALWARE DETECTION.....	29
<b>A.L. Alexeyeva</b> SUBSONIC VIBROTRANSPORT SOLUTIONS OF THE WAVE EQUATION IN SPACES OF DIMENSION $N=1,2,3$ .....	42
<b>K. Bagitova, Sh. Mussiraliyeva, K. Azanbai</b> ANALYSIS OF SYSTEMS FOR RECOGNIZING POLITICAL EXTREMISM IN ONLINE SOCIAL NETWORKS.....	60
<b>A.S. Baegizova, G.I. Mukhamedrakhimova, I. Bapiyev, M.Zh. Bazarova, U.M. Smailova</b> EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING METHODS FOR KEYWORD COVERAGE.....	73
<b>G. Bekmanova, B. Yergesh, G. Yelibayeva, A. Omarbekova, M. Strecker</b> MODELING THE RULES AND CONDITIONS FOR CONDUCTING PRE-ELECTION DEBATES.....	89
<b>M. Bolatbek, M. Sagynay, Sh. Mussiraliyeva</b> USING MACHINE LEARNING METHODS FOR DETECTING DESTRUCTIVE WEB CONTENT IN KAZAKH LANGUAGE.....	99
<b>Y. Golenko, A. Ismailova, K. Kadirkulov, R. Kalendar</b> DEVELOPMENT OF AN ONLINE PLATFORM FOR SEARCHING FOR TANDEM REPEATS USING WHOLE GENOME SEQUENCING.....	112

<b>T. Zhukabayeva, L. Zholshiyeva, N. Karabayev, Sh. Akhmetzhanova</b> A BIBLIOMETRIC ANALYSIS OF EDGE COMPUTING IN INDUSTRIAL INTERNET OF THINGS (IIoT) CYBER-PHYSICAL SYSTEMS.....	123
<b>S.S. Koishybay, N. Meirambekuly, A.E. Kulakaeva, B.A. Kozhakhmetova, A.A. Bulin</b> DEVELOPMENT OF THE DESIGN OF A MULTI-BAND DISCONE ANTENNA.....	138
<b>A. Kydyrbekova, D. Oralbekova</b> SPEAKER IDENTIFICATION USING DISTRIBUTION-PRESERVING X-VECTOR GENERATION.....	152
<b>B. Medetov, A. Nurlankyzy, A. Akhmediyarova, A. Zhetpisbayeva, D. Zhexebay</b> COMPARATIVE ANALYSIS OF THE EFFECTIVENESS OF NEURAL NETWORKS WITHIN THE LOW SNR.....	163
<b>A.A Myrzatay, L.G. Rzaeva, B. Zhumadilla, A.A. Mukhanova, G.A. Uskenbayeva</b> DOUBLE EXPONENTIAL SMOOTHING AND TIME WINDOW METHODS FOR PREDICTIVE LAN MONITORING: ANALYSIS, COMPARISON AND APPLICATION.....	174
<b>L. Naizabayeva, M.N. Satymbekov</b> PREDICTING URBAN SOIL POLLUTION USING MACHINE LEARNING ALGORITHMS.....	194
<b>A.U. Mukhiyadin, U.T. Makhazhanova, A.Z. Alimagambetova, A.A. Mukhanova, A.I. Akmoldina</b> PREDICTING STUDENT LEARNING ENGAGEMENT USING MACHINE LEARNING TECHNIQUES: ANALYSIS OF EDUCATION DATA IN KAZAKHSTAN.....	204
<b>Zh. Tashenova, Zh. Abdugulova, Sh. Amanzholova, E. Nurlybaeva</b> PENETRATION TESTING APPROACHES EMPLOYING THE OPENVAS VULNERABILITY MANAGEMENT UTILITY.....	218
<b>D.B. Tyulemissova, A.K. Shaikhanova, V. Martsenyuk, G.A. Uskenbayeva</b> MODERN APPROACHES TO STUDYING THE DYNAMICS OF INFORMATION FLOW IN SOCIAL MEDIA BASED ON MACHINE LEARNING METHODS.....	231

## МАЗМҰНЫ

### АҚПАРАТТЫҚ-КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР

**М. Айтимов, Р.У Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан**  
ҒЫЛЫМИ МӘТІНДЕРДЕН СЕМАНТИКАЛЫҚ ҚҰРЫЛЫМДАРДЫ  
ТАЛДАУ ЖӘНЕ АЛУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСІН  
ҚОЛДАНУ.....5

**Ә.Қ. Әйтiм, Г.К. Сембина**  
МАШИНАЛЫҚ ОҚУ АЛГОРИТМІН ПАЙДАЛАНЫП СМАРТФОН  
ҮШІН АДАМ МІНЕЗІН МОДЕЛДЕУ.....17

**Г.И. Ақшолақ, А.А. Бедельбаев, Р.С. Мағазов**  
ЗИЯНДЫ БАҒДАРЛАМАЛАРДЫ АНЫҚТАУҒА АРНАЛҒАН  
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ТАЛДАУ ЖӘНЕ САЛЫСТЫРУ.....29

**А.Л. Алексеева**  
N=1,2,3 ӨЛШЕМДІ КЕҢІСТІГІНДЕГІ ТОЛҚЫНДЫҚ ТЕҢДЕУДІҢ  
ДЫБЫСҚА ДЕЙІНГІ ДІРІЛКӨЛІКТІК ШЕШІМДЕРІ.....42

**Қ.Б. Бағитова, Ш.Ж. Мусиралиева, Қ. Азанбай**  
ӘЛЕУМЕТТІК ЖЕЛІЛЕРДЕГІ САЯСИ ЭКСТРЕМИЗМДІ ОНЛАЙН ТАҢУ  
ЖҮЙЕЛЕРІН ТАЛДАУ.....60

**А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова,  
У.М. Смайлова**  
ТҮЙІН СӨЗДЕРДІ ҚАМТУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНІҢ  
ТИІМДІЛІГІН БАҒАЛАУ.....73

**Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова,  
М. Strecker**  
САЙЛАУ АЛДЫНДАҒЫ ПІКІРТАЛАСТАРДЫ ӨТКІЗУ ЕРЕЖЕЛЕРІ  
МЕН ШАРТТАРЫН МОДЕЛЬДЕУ.....89

**М.А. Болатбек, М.Сағынай, Ш.Ж. Мусиралиева**  
ҚАЗАҚ ТІЛІНДЕГІ ДЕСТРУКТИВТІ ВЕБ-КОНТЕНТТІ АНЫҚТАУ ҮШІН  
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ.....99

**Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь**  
ТОЛЫҚ ГЕНОМДЫҚ СЕКВЕНИРЛЕУДЕ ТАНДЕМДІК  
ҚАЙТАЛАНУЛАРДЫ ІЗДЕУ ҮШІН ОНЛАЙН ПЛАТФОРМАСЫН  
ӘЗІРЛЕУ.....112

- Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова**  
ӨНДІРІСТІК ЗАТТАР ИНТЕРНЕТІ (IoT) КИБЕРФИЗИКАЛЫҚ  
ЖҮЙЕЛЕРІНДЕ ШЕТКІ ЕСЕПТЕУЛЕРДІ ҚОЛДАНУҒА  
БИБЛИОМЕТРИЯЛЫҚ ТАЛДАУ.....123
- С.С. Қойшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова,  
А.А. Булин**  
КӨПДИАПАЗОНДЫДИСКОНУСТЫҚАНТЕННАКОНСТРУКЦИЯСЫН  
ӘЗІРЛЕУ.....138
- А.С. Кыдырбекова, Д.О. Оралбекова**  
ТАРАТУДЫ САҚТАЙТЫН Х-ВЕКТОРЛАР ГЕНЕРАЦИЯСЫН  
ПАЙДАЛАНЫП ДАУЫСТЫ ИДЕНТИФИКАЦИЯЛАУ.....152
- Б. Медетов, А. Нурланқызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай**  
СИГНАЛШУЫЛ ҚАТЫНАСЫ ТӨМЕН ЖАҒДАЙДА НЕЙРОНДЫҚ  
ЖЕЛЛЕРДІҢ ТИІМДІЛІГІНЕ САЛЫСТЫРМАЛЫ ТАЛДАУ ЖАСАУ.....163
- А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова,  
Г.А. Ускенбаева**  
ЖЕРГІЛІКТІ ЖЕЛІНІ БОЛЖАМДЫ БАҚЫЛАУҒА АРНАЛҒАН ҚОС  
ЭКСПОНЕНЦИАЛДЫ ТЕГІСТЕУ ЖӘНЕ УАҚЫТ ТЕРЕЗЕЛЕРІНІҢ  
ӘДІСТЕРІ: ТАЛДАУ, САЛЫСТЫРУ ЖӘНЕ ҚОЛДАНУ.....174
- Л. Найзабаева, М.Н. Сатымбеков**  
МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ПАЙДАЛАҢА АРҚЫЛЫ  
ҚАЛА ТОПЫРАҒЫНЫҢ ЛАСТАҢУЫН БОЛЖАУ.....194
- А.Ұ. Мұхиядин, У.Т. Махажанова, А.З. Алимагамбетова, А.А.Муханова,  
А.И. Акмолдина**  
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНА ОТЫРЫП,  
ОҚУШЫЛАРДЫҢ БІЛІМ АЛУҒА ЫҢТАСЫН БОЛЖАУ:  
ҚАЗАҚСТАҢДАҒЫ БІЛІМ БЕРУ ДЕРЕКТЕРІН ТАЛДАУ.....204
- Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева**  
OPENVAS ОСАЛДЫҒЫН БАСҚАРУ УТИЛИТАСЫН ҚОЛДАНА  
ОТЫРЫП, ЕНУДІ ТЕСТІЛЕУ ТӘСІЛДЕРІ.....218
- Д.Б. Тюлемисова, А.К. Шайханова, В.П. Мартценюк, Г.А. Ускенбаева,  
Г.В. Бекешева**  
МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІНЕ НЕГІЗДЕЛГЕН ӘЛЕУМЕТТІК  
ЖЕЛЛЕРДЕГІ АҚПАРАТ АҒЫНЫНЫҢ ДИНАМИКАСЫН ЗЕРТТЕУДІҢ  
ЗАМАНАУИ ТӘСІЛДЕРІ.....231

## СОДЕРЖАНИЕ

### ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ

<b>М. Айтимов, Р.У Альменаева, К.К. Макулов, А.Б. Остаева, Р. Муратхан</b> ПРИМЕНЕНИЕ МЕТОДА МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА И ИЗВЛЕЧЕНИЯ СЕМАНТИЧЕСКИХ СТРУКТУР ИЗ НАУЧНЫХ ТЕКСТОВ.....	5
<b>А.К. Айтим, Г.К. Сембина</b> МОДЕЛИРОВАНИЕ ЧЕЛОВЕЧЕСКОГО ПОВЕДЕНИЯ ДЛЯ СМАРТФОНА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ.....	17
<b>Г.И. Акшолок, А.А. Бедельбаев, Р.С. Магазов</b> АНАЛИЗ И СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ВРЕДОНОСНОГО ПО.....	29
<b>Л.А. Алексеева</b> ДОЗВУКОВЫЕ ВИБРОТРАНСПОРТНЫЕ РЕШЕНИЯ ВОЛНОВОГО УРАВНЕНИЯ В ПРОСТРАНСТВАХ РАЗМЕРНОСТИ $N=1,2,3$ .....	42
<b>К.Б. Багитова, Ш.Ж. Мусиралиева, К. Азанбай</b> АНАЛИЗ СИСТЕМ РАСПОЗНАВАНИЯ ПОЛИТИЧЕСКОГО ЭКСТРЕМИЗМА В СОЦИАЛЬНЫХ СЕТЯХ ОНЛАЙН.....	60
<b>А.С. Баегизова, Г.И. Мухамедрахимова, И.М. Бапиев, М.Ж. Базарова, У.М. Смайлова</b> ОЦЕНКА ЭФФЕКТИВНОСТИ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОХВАТА КЛЮЧЕВЫХ СЛОВ.....	73
<b>Г.Т. Бекманова, Б.Ж. Ергеш, Г.К. Елибаева, А.С. Омарбекова, М. Strecker</b> МОДЕЛИРОВАНИЕ ПРАВИЛ И УСЛОВИЙ ПРОВЕДЕНИЯ ПРЕДВЫБОРНЫХ ДЕБАТОВ.....	89
<b>М.А. Болатбек, М. Сагынай, Ш.Ж. Мусиралиева</b> ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ДЕСТРУКТИВНОГО ВЕБ-КОНТЕНТА НА КАЗАХСКОМ ЯЗЫКЕ.....	99
<b>Е.С. Голенко, А.А. Исмаилова, К.К. Кадиркулов, Р.Н. Календарь</b> РАЗРАБОТКА ОНЛАЙН-ПЛАТФОРМЫ ДЛЯ ПОИСКА ТАНДЕМНЫХ ПОВТОРОВ ПРИ ПОЛНОГЕНОМНОМ СЕКВЕНИРОВАНИИ.....	112



<b>Т. Жукабаева, Л. Жолшиева, Н. Карабаев, Ш. Ахметжанова</b> БИБЛИОМЕТРИЧЕСКИЙ АНАЛИЗ ПРИМЕНЕНИЯ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ В КИБЕРФИЗИЧЕСКИХ СИСТЕМАХ ПРОМЫШЛЕННОГО ИНТЕРНЕТА ВЕЩЕЙ (IIoT).....	123
<b>С.С. Койшыбай, Н. Мейрамбекұлы, А.Е. Кулакаева, Б.А. Кожаметова, А.А. Булин</b> РАЗРАБОТКА КОНСТРУКЦИИ МНОГОДИАПАЗОННОЙ ДИСКОНУСНОЙ АНТЕННЫ.....	138
<b>А.С. Кыдырбекова, Д.О. Оралбекова</b> ИДЕНТИФИКАЦИЯ ГОВОРЯЩЕГО С ИСПОЛЬЗОВАНИЕМ ГЕНЕРАЦИИ X-ВЕКТОРОВ С СОХРАНЕНИЕМ РАСПРЕДЕЛЕНИЯ...152	
<b>Б. Медетов, А. Нурланкызы, А. Ахмедиярова, А. Жетписбаева, Д. Жексебай</b> СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ НЕЙРОННЫХ СЕТЕЙ ПРИ НИЗКОМ ЗНАЧЕНИИ ОТНОШЕНИЯ С/Ш.....	163
<b>А.А. Мырзатай, Л.Г. Рзаева, Б. Жұмаділла, А.А. Муханова, Г.А. Ускенбаева</b> МЕТОДЫ ДВОЙНОГО ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ И ВРЕМЕННЫХ ОКОН ДЛЯ ПРЕДИКТИВНОГО МОНИТОРИНГА ЛВС: АНАЛИЗ, СРАВНЕНИЕ И ПРИМЕНЕНИЕ.....	174
<b>Л. Найзабаева, М.Н. Сатымбеков</b> ПРОГНОЗИРОВАНИЕ ЗАГРЯЗНЕНИЯ ГОРОДСКОЙ ПОЧВЫ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ.....	194
<b>А.У. Мухиядин, У.Т. Махажанов, А.З. Алимагамбетова, А.А. Муханова, А.И. Акмолдина</b> ПРОГНОЗИРОВАНИЕ МОТИВАЦИИ УЧАЩИХСЯ К ОБУЧЕНИЮ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ: АНАЛИЗ ДАННЫХ ОБ ОБРАЗОВАНИИ В КАЗАХСТАНЕ.....	204
<b>Ж.М. Ташенова, Ж.К. Абдугулова, Ш.А. Аманжолова, Э. Нурлыбаева</b> ПОДХОДЫ К ТЕСТИРОВАНИЮ НА ПРОНИКНОВЕНИЕ С ИСПОЛЬЗОВАНИЕМ УТИЛИТЫ УПРАВЛЕНИЯ Уязвимостями OPENVAS.....	218
<b>Д.Б. Тюлемисова, А.К. Шайханова, В. Мартценюк, Г.А. Ускенбаева, Г.В. Бекешева</b> СОВРЕМЕННЫЕ ПОДХОДЫ К ИЗУЧЕНИЮ ДИНАМИКИ ИНФОРМАЦИОННОГО ПОТОКА В СОЦИАЛЬНЫХ МЕДИА НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.....	231

**Publication Ethics and Publication Malpractice  
the journals of the National Academy of Sciences of the Republic of Kazakhstan**

For information on Ethics in publishing and Ethical guidelines for journal publication see <http://www.elsevier.com/publishingethics> and <http://www.elsevier.com/journal-authors/ethics>.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see <http://www.elsevier.com/postingpolicy>), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct ([http://publicationethics.org/files/u2/New\\_Code.pdf](http://publicationethics.org/files/u2/New_Code.pdf)). To verify originality, your article may be checked by the Cross Check originality detection service <http://www.elsevier.com/editors/plagdetect>.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

**[www.nauka-nanrk.kz](http://www.nauka-nanrk.kz)**

**<http://physics-mathematics.kz/index.php/en/archive>**

**ISSN 2518-1726 (Online),**

**ISSN 1991-346X (Print)**

Директор отдела издания научных журналов НАН РК *А. Ботанқызы*

Редакторы: *Д.С. Аленов, Ж.Ш. Әден*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 2.12.2024.

Формат 60x881/8. Бумага офсетная. Печать – ризограф.

16,0 п.л. Тираж 300. Заказ 4.