

УДК 004.9, 579.25  
DOI 10.25205/1818-7900-2019-17-1-90-100

## **Информационно-аналитическая система с алгоритмами геномного анализа патогенов вирусных инфекций**

**В. В. Черненко, Ю. И. Молородов**

*Институт вычислительных технологий СО РАН  
Новосибирск, Россия*

### *Аннотация*

Работа посвящена описанию структуры информационно-аналитической системы «Ixodes», ориентированной на работу с представительной коллекцией иксодовых клещей из разных биотопов, а именно для территорий Алтая, Сибири и Дальнего востока. Показаны варианты применения системы для анализа генетического разнообразия клещей и переносимых ими патогенов при помощи методов статистической обработки в виде круговых и столбчатых диаграмм (гистограмм). Описаны реализованные алгоритмы, позволяющие проводить анализ генетической последовательности исследуемого патогена на основе L-граммного подхода и методами разбиения филогенетического дерева на группы близких последовательностей. При этом для первичной обработки набора геномов используются методы множественного выравнивания последовательностей и метод присоединения соседей, позволяющий выполнить построение филогенетического дерева. Представленные алгоритмы и методы использовались для решения задачи генотипирования вируса клещевого энцефалита (ВКЭ). Представлены результаты апробации для методов разбиения филогенетического дерева и их сравнительный анализ. Описана архитектура информационно-аналитической системы для анализа набора геномов. Система предназначена для анализа множества геномов и их классификации, а именно для анализа генотипов внутри одного вида живых организмов, поскольку методы направлены для выделения различий геномов, имеющих схожую структуру.

### *Ключевые слова*

информационные системы, интеграция данных, классификация, кластеризация, L-граммный анализ, филогенетическое дерево, клещи, секвенирование, энцефалит, штамм

### *Благодарности*

Исследования выполнены при частичной поддержке гранта РФФИ № 18-07-01457, Интеграционного проекта СО РАН № АААА-А18-118022190008-8 (№ 0316-2018-0002) и темы госзадания № АААА-А17-117120670141-7 (№ 0316-2018-0009).

### *Для цитирования*

Черненко В. В., Молородов Ю. И. Информационно-аналитическая система с алгоритмами геномного анализа патогенов вирусных инфекций // Вестник НГУ. Серия: Информационные технологии. 2019. Т. 17, № 1. С. 90–100. DOI 10.25205/1818-7900-2019-17-1-90-100

## **The Information-Analytical System with Using Algorithms Genomic Analysis of Pathogens of Viral Infections**

**V. V. Chernenko, Yu. I. Molorodov**

*Institute of Computational Technologies SB RAS  
Novosibirsk, Russia*

### *Annotation*

The presented paper is to describe the structure of the information-analytical system “Ixodes” working with the collection of Ixodidae ticks from different biotopes, namely from the territories of Altai, Siberia and the Far East. Variants of analyzing the genetic diversity for ticks and pathogens transferred by them have been shown with using statistical methods of building to circular and bar graphs (histograms). The implemented algorithms have been described that allow dealing with the analysis of the pathogen genetic sequence based on the L-gram approach and the methods of partitioning the phylogenetic tree into groups of close sequences. At the same time, for the first processing a set of genomes, methods of multiple sequence alignment and the method of Neighbor-joining allowing to build a phylogenetic tree have been used. The presented algorithms and methods have been used to solve the problem of tick-borne encephalitis virus genotyping. The results of testing for phylogenetic tree partitioning methods and their comparative

© В. В. Черненко, Ю. И. Молородов, 2019

analysis have been presented. The architecture of the information-analytical system for analyzing a set of genomes has been described. The system helps in the analysis of a variety of genomes and their classification, namely, for analyzing genotypes within a single species of living organisms, with the methods to aimed at isolating subtle differences in genomes with a similar structure.

#### Keywords

information systems, data integration, classification, clustering, L-gram analysis, phylogenetic tree, ticks, sequencing, encephalitis, strain

#### Acknowledgements

The work is supported by RFBR (grant no. 18-07-01457), projects no. AAAA-A18-118022190008-8 (no. 0316-2018-0002) and no. AAAA-A17-117120670141-7 (no. 0316-2018-0009).

#### For citation

Chernenko V. V., Molorodov Yu. I. The Information-Analytical System with Using Algorithms Genomic Analysis of Pathogens of Viral Infections. *Vestnik NSU. Series: Information Technologies*, 2019, vol. 17, no. 1, p. 90–100. (in Russ.) DOI 10.25205/1818-7900-2019-17-1-90-100

## Введение

Территория России – один из самых больших в мире ареалов, где распространены около 60 видов иксодовых клещей, для которых мелкие млекопитающие и животные, а также человек являются кормовой базой. При этом опасны не сами клещи, а передаваемые ими при укусах бактериальные инфекции: боррелиоз, анаплазмоз, эрлихиоз, клещевой риккетсиоз. Заболевания вызываются бактериями *Borrelia burgdorferi* и *Borrelia miyamotoi*, вирусом Кемерово и вирусом клещевого энцефалита (ВКЭ). ВКЭ – самый распространенный и тяжелый эпидемический энцефалит на территории России и других стран. Осложнения этой острой инфекции могут завершиться параличом и летальным исходом [1–5].

Ранее была разработана информационно-аналитическая система *Ixodes*<sup>1</sup>, позволяющая хранить данные полевых экспедиций, включая информацию по ареалам расселения, видовому составу насекомых и переносимых ими патогенов инфекционных заболеваний, определяемых путем проведения процедуры секвенирования.

Основой информационной системы (ИС) является интерактивная карта (рис. 1) с отображением мест полевых сборов и информации о них: административное название, биотоп, климат, координаты мест сбора информации.

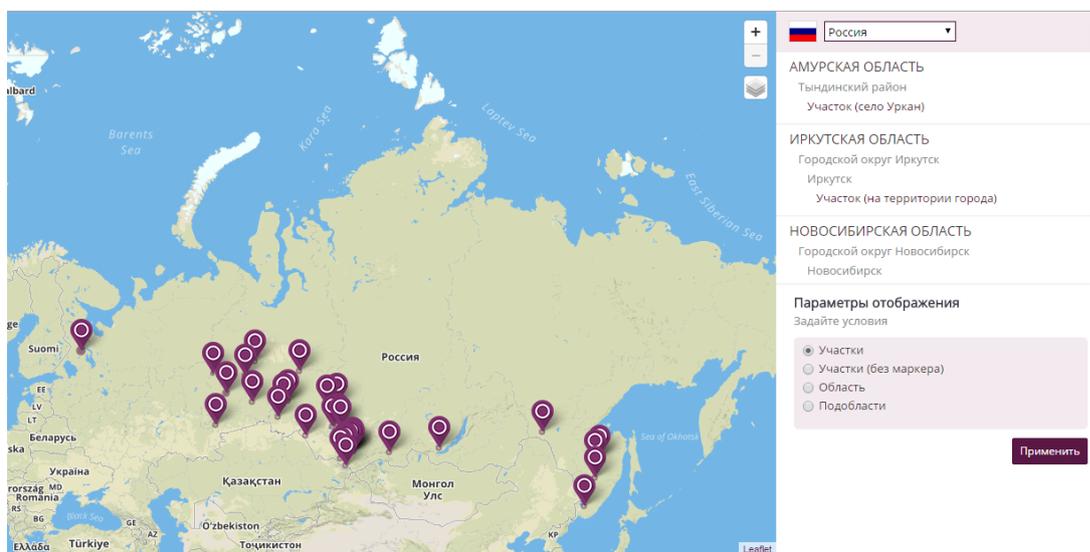


Рис. 1. Картографический интерфейс системы  
Fig. 1. The interface of cartographic system

<sup>1</sup> Ixodes Analysis System. URL: <http://ixodes.ict.nsc.ru> (дата обращения 26.01.2019).

Была реализована возможность просмотра численности клещей по виду, полу, биотопу, генам и инфекциям, связанным с конкретным насекомым (рис. 2). Встроенные алгоритмы статистики позволяли провести сравнительный анализ встречаемости инфекций и генов и построить столбчатые и круговые диаграммы.

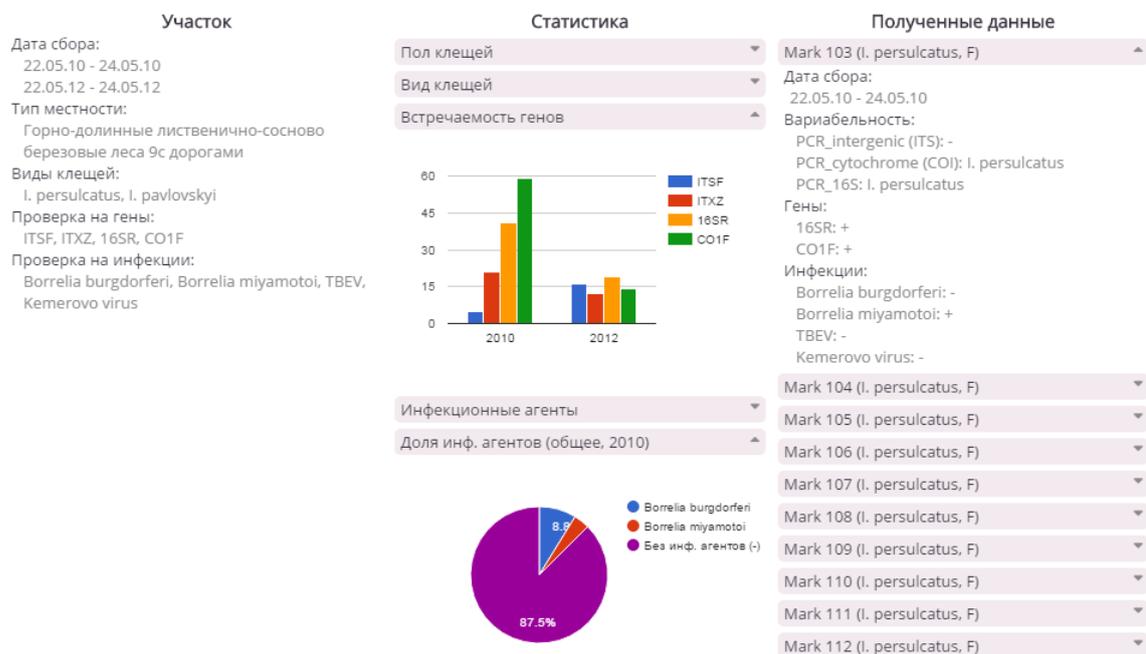


Рис. 2. Представление статистики для патогенов в системе. Левый столбец показывает информацию об участке сбора клещей. Средний столбец представляет гистограммы для отображения различных характеристик клещей. Правый столбец показывает данные по каждому секвенированному образцу на этом участке

Fig. 2. The presentation of statistics for pathogens in the system. The left column shows information about a tick collection site. The middle column presents histograms for displaying of various tick characteristics. The right column shows data for each sequenced sample in this region

Разработанные методы позволяют производить простой анализ, который можно расширить. Для расширения системы были реализованы методы кластеризации и классификации для обработки и проведения анализа набора геномов на основе их последовательностей.

Можно определять родственные отношения между организмами по их геномам при помощи методов филогенетического анализа. Привлечение иерархических методов кластеризации позволяет получить визуальное представление исходных геномов, но при больших исходных данных возникает проблема поиска и анализа визуализированных объектов. Для этого необходимо применить алгоритмы кластеризации. Кластеризация позволяет разбить объекты (геномы) на классы со специфическими характерными признаками. Полученное разбиение на гены решает задачу генотипирования. Под генотипированием мы понимаем задачу отнесения произвольного штамма, представленного полной кодирующей последовательностью, к одному из известных генотипов.

Использование технологии секвенирования [6] (определения последовательности нуклеотидов ДНК всего генома) обеспечивает возможность анализа связи полиморфизма сотен тысяч маркеров однонуклеотидных полиморфизмов (Single Nucleotide Polymorphism, SNP), рассеянных по всему геному, с набором патогенов, переносимых клещами.

Спектр задач, решаемых с помощью молекулярно-генетических маркеров в биологии, является весьма существенным. Это типирование и паспортизация хозяйственно ценных генов, генотипов, индивидов, включая трансгенные растения; коммерческая сертификация; анализ

генетического родства и происхождения особей, сортов, форм, насаждений; исследование генетической структуры популяций и ее динамики; изучение уровня генетического разнообразия видов. Благодаря таким маркерам можно установить филогенетические взаимоотношения видов, решать спорные вопросы таксономии, осуществлять диагностику вирусных, бактериальных и грибных инфекций, построить генетические карты и др. [1].

Чрезвычайно высокая опасность для человека вируса клещевого энцефалита объясняет пристальный интерес биологов к этому феномену и появление большого количества информации по ВКЭ, получаемое методами секвенирования генетического материала насекомых. Стала актуальна задача генотипирования, с помощью которой появилась возможность выделения разнообразных типов ВКЭ, с разной вирулентностью – способностью инфекционного агента заражать живой организм.

Исчерпывающая информация о генотипе содержится в полной кодирующей последовательности генома ВКЭ. Из нее можно извлечь ограниченное количество маркеров генотипирования в виде относительно коротких структурированных фрагментов РНК. В данной работе для апробации разработанных методов в задаче генотипирования ВКЭ мы выделяем структурированные РНК-маркеры [4].

Для формального описания определим конечное множество символов (алфавит), описывающих РНК-маркеры, обозначим  $\Sigma$ . При кодировании генома для РНК это 4 символа: *Аденин (A)*, *Гуанин (G)*, *Цитозин (C)*, *Тимин (T)*, соответствующие азотистым основаниям без *Урацила (U)*, что соответствует используемым геномным данным [7].

### Постановка задачи и исходные данные

В системе разработан функционал, позволяющий решать задачу генотипирования на основе двух подходов: кластеризации и классификации. Каждый из двух подходов базируется на уже известных теоретических основах кластеризации и классификации соответственно и включают новые методики решения этих задач для генотипирования. Для апробации методов используется уже известное разбиение штаммов на типы ВКЭ [8. Прилож. 1]. Геномы для отдельных типов ВКЭ были предоставлены НИИ Биомедицинских технологий Иркутского государственного медицинского университета.

Первый подход направлен на решение задачи генотипирования через кластеризацию на основе филогенетического дерева штаммов. Построение филогенетического дерева относится к задачам иерархической кластеризации [9]. Филогенетическое дерево может строиться на основе различных алгоритмов, таких как «UPGMA» [10], «WPGMA» [10], «Molecular clock» [11] и др. «Neighbor Joining» [12] алгоритм использовался в работе для построения филогенетического дерева. Алгоритм позволяет итеративно построить филогенетическое дерево, присоединяя на каждом шаге итерации наиболее близкие геномы, используя расстояние Хэмминга.

Филогенетическое дерево дает представление отношений между штаммами, но не решает задачу генотипирования. Для этого в работе разработаны алгоритмы кластеризации на основе филогенетического дерева, которые позволяют произвести генотипирование набора штаммов. Некоторые алгоритмы по методике выполнения напоминают уже существующие алгоритмы разбиения графов [13], но направлены на работу с филогенетическими деревьями, решая задачу генотипирования.

Второй подход направлен на решение задачи генотипирования через классификацию. Классификация подразумевает наличие обучающей выборки, которая задает классификатор для каждого генотипа [14]. Возможный подход к выделению РНК-маркеров на основе L-грамм был представлен ранее [15]. В качестве L-грамм рассматриваются последовательности символов, длина последовательности соответствует фиксированному значению L. В системе представлен метод, при котором в качестве классификатора выступает линейный оператор на основе L-граммной характеристики генотипа, позволяющий при помощи L-граммной характеристики нового штамма отнести его к тому или иному типу [8].

### Архитектура системы

Ранее была разработана информационно-аналитическая система «Ixodes», которая представляет собой трехуровневую систему. Трехуровневая архитектура базируется на системе клиент-сервер [16], которая содержит три основных компонента: клиент, сервер и база данных. Клиентом в нашем случае выступает браузер пользователя. Клиент отправляет запрос на сервер, где хранится логика приложения, которая обрабатывает запрос и манипулирует с базой данных. Из базы извлекается нужная информация, снова обрабатывается, и ответ отправляется обратно пользователю.

В системе предусмотрена аутентификация и авторизация. Аутентификация позволяет подтвердить, что клиент является зарегистрированным пользователем сайта. Авторизация позволяет проверять имеет ли пользователь доступ на выполнение определенных действий. Для авторизации в базе данных хранится таблица прав доступа.

На текущий момент уже существующую систему дополнили новыми функциями, которые позволяют производить вычисления для ранее описанных алгоритмов и методов. На сервере выделен отдельный блок, который занимается только сложными вычислениями. Серверная часть разделена на две части. Первая занимается обработкой фактологической информации о патогенах, которая хранится в базе данных. Вторая направлена на вычисление результатов геномного анализа. Усовершенствованная система представима в виде схемы (рис. 3).

Такое разграничение позволило выбрать подходящие инструменты разработки для каждой части. Соответственно для обработки фактологической информации подходит программное обеспечение (ПО), которое работает непосредственно с сервером и быстро обрабатывает простые запросы, например PHP (*PersonalHomePageTools* – скриптовый язык общего назначения, интенсивно применяемый для разработки веб-приложений). Для геномного анализа кроме ПО для работы с сервером необходимо использовать средства для вычислительных задач.

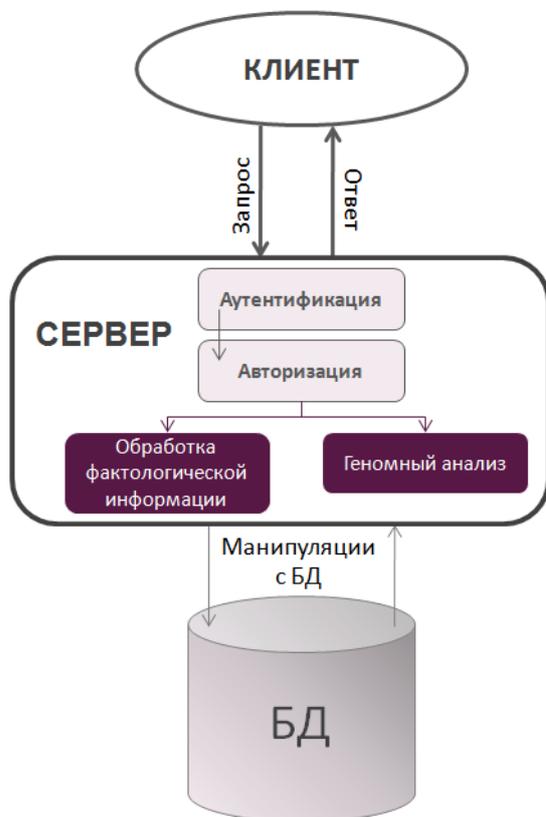


Рис. 3. Схема трехуровневой системы с выделенными блоками для обработки фактологической информации и геномного анализа

Fig. 3. The diagram of the three-level system with dedicated blocks for processing factual information and genomic analysis

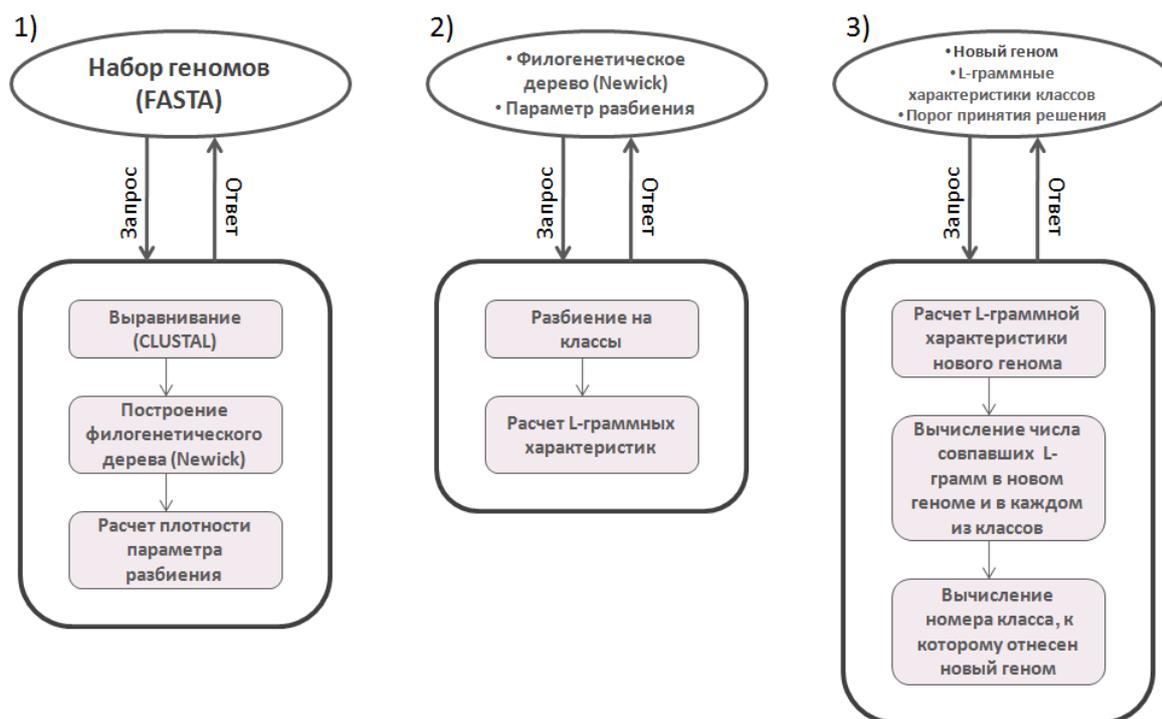


Рис. 4. Схема выполнения геномного анализа:

1 – этап построения филогенетического дерева; 2 – этап кластеризации; 3 – этап классификации

Fig. 4. The scheme of genomic analysis:

1 – the stage of building a phylogenetic tree; 2 – clustering stage; 3 – classification stage

На первом этапе геномного анализа (рис. 4) выполняется запрос на вычисление филогенетического дерева. Сначала геномы, представленные в формате *FASTA*, подвергаются выравниванию с помощью программного обеспечения «MAFFT», написанного на языке *C* (компилируемый, статически типизированный язык программирования общего назначения). Затем по выровненным последовательностям в формате *CLUSTAL* строится филогенетическое дерево методом «NeighborJoining» через программу, написанную нами на *C++*. Полученное дерево записывается в формате *Newick*<sup>2</sup>. Затем рассчитываются гистограммы плотности параметра разбиения. Результат отправляется клиенту для получения следующего запроса. На втором этапе клиент подает на вход филогенетическое дерево в формате *Newick* и параметр разбиения. На сервере в написанной нами программе дерево из формата *Newick* преобразуется в объект, затем по параметру строится разбиение дерева на классы. Для каждого класса рассчитывается L-граммная характеристика, которая вместе с классами отправляется обратно клиенту. На третьем этапе пользователь может определить новый геном и отправить его на сервер вместе с L-граммными характеристиками других классов и порогом принятия решения. На сервере написанная нами программа вычисляет L-граммную характеристику нового генома, решает задачу классификации и возвращает клиенту ответ в виде числа совпавших L-грамм для каждой из характеристик и класса, к которому отнесен новый геном.

<sup>2</sup> The Newick tree format. URL: <http://evolution.genetics.washington.edu/phylip/newicktree.html> (дата обращения 20.01.2019).

### Алгоритмы кластеризации

Геномы для групп ВКЭ были предоставлены НИИ Биомедицинских технологий Иркутского государственного медицинского университета. Кодирующая последовательность символов (CDS) для каждого генома была взята из открытой базы данных генетических последовательностей «GenBank»<sup>3</sup>. В соответствии с официально принятой классификацией различают три основных типа ВКЭ [5]<sup>4</sup>:

- 1) дальневосточный;
- 2) европейский;
- 3) сибирский.

Каждый геном из общего набора нумеруется цифрой от 1 до 3 в зависимости от его типа, которая для входных данных алгоритма эквивалентна группе. Для удобства все геномы нумеруются по порядку внутри группы. Нумерация геномов представлена в приложении 1 [8].

Построенное дерево отображается при помощи библиотеки «*phylotree.js*»<sup>5</sup>. Дерево представлено в радиальной форме, иначе его было бы сложно разметить компактно на одном рисунке. Для лучшего понимания представлен путь от листа до корня дерева (рис. 5).

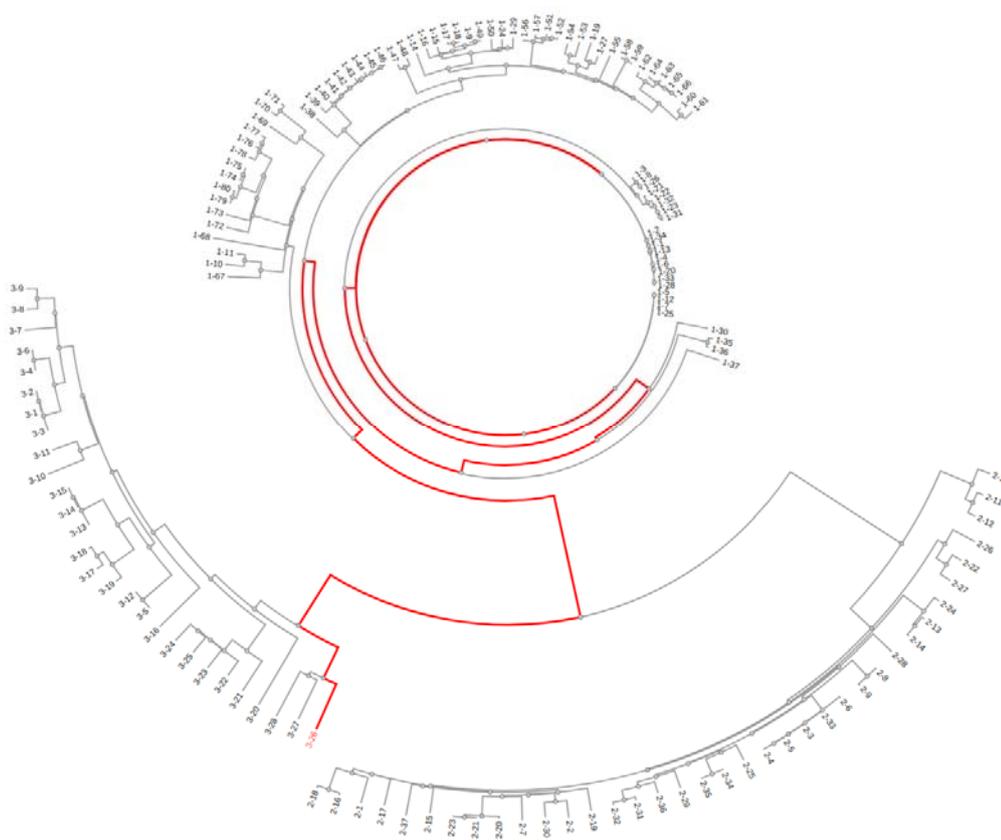


Рис. 5. Филогенетическое дерево для групп ВКЭ. Всего было 145 геномов. Красным помечен путь от генома 3-26 (GU183384.1\_Est54), лежащего в группе 3, соответствующей сибирскому типу ВКЭ

Fig. 5 The phylogenetic tree for groups of TBE. There were 145 genomes in total. Red marked let from the genome 3-26 (GU183384.1\_Est54) lying in group 3, corresponding to the Siberian type of TBE

<sup>3</sup> GenBank Overview. URL: <https://www.ncbi.nlm.nih.gov/genbank/> (дата обращения 20.12.2018).

<sup>4</sup> См. также: GitHub – *veg*/phylotree.js: Interactive viewer of phylogenetic trees. URL: <https://github.com/veg/phylotree.js/tree/master> (дата обращения 10.12.2018).

<sup>5</sup> Ibid.

Для извлечения классов геномов разработаны следующие алгоритмы кластеризации на филогенетическом дереве [8]:

- 1) простой алгоритм разбиения;
- 2) алгоритм последовательного сравнения листьев;
- 3) поуровневый алгоритм.

Для набора групп ВКЭ проведено выделение классов по каждому алгоритму для трех типов:

- 1) разбиение на классы, которые соответствуют группам.
- 2) разбиение на классы меньшие, чем группы.
- 3) разбиение на классы большие, чем группы.

Изначально подобран параметр, чтобы получить разбиение на классы, которые будут соответствовать исходным группам. Затем параметр изменен так, чтобы получить классы большие или меньшие исходных групп, при этом проверяется соответствие классов визуальное на филогенетическом дереве для оценки качества работы алгоритма. Результаты представлены в таблице <sup>6</sup>.

После проведенных вычислений на основе данных ВКЭ можно сделать следующие заключения.

Алгоритм последовательного сравнения листьев строит разбиение на классы, которые больше соответствуют визуальному представлению, чем классы, полученные другими алгоритмами. Поэтому для анализа групп и соответствующих им классов лучше использовать его.

Алгоритм простого разбиения может помочь при анализе дерева на выделенные визуальное узлы, состоящие из одного листа, но он хуже подходит для построения разбиения, которое соответствует крупным визуальным представлениям, чем предыдущий алгоритм.

Предполагалось, что поуровневый алгоритм покажет лучший результат, но из-за того что у корня дерева расположены геномы первой группы, а уже потом из них выходит 2 и 3 группы, алгоритм строит разбиение, которое может не соответствовать визуальному представлению на дереве. Эту проблему можно решить, если начинать проход дерева с узла, где соединяются все три группы, но это можно выполнить только в том случае, если считать филогенетическое дерево неукорененным. В нашем случае рассматривалось укорененное дерево.

Алгоритмы находят непересекающиеся классы, поскольку изначально можно положить, что каждый геном лежит в своем классе. Эти классы попарно не пересекаются. Во время работы алгоритма производится объединение непересекающихся классов, в конце которого получается разбиение. Помимо этого, объединяются всегда класс, состоящий из ранее объединенных геномов, и один новый геном, для которого вершина становится открытой, что включает возможность повторного добавления генома в класс.

При соотношении групп и полученных классов геномы из одного класса могут не находиться в одной и той же группе. Во-первых, если мы получим разбиение на классы, один из которых содержит большее количество геномов, чем соответствующая ему группа, то очевидно, что часть геномов класса будет получена из других групп. Во-вторых, группы геномов устанавливает эксперт, что не гарантирует соответствие результату кластеризации. Для филогенетического дерева ВКЭ в силу соответствия групп и визуального представления мы не получили случая, когда геномы разных групп находятся в одном классе.

---

<sup>6</sup> Результаты проведенных вычислений, представленные в таблицах, а именно разбиения на классы соответствующие группам, большим группам и меньшим группам, представлены в приложениях 2, 3 и 4 в работе [8].

Результаты работы алгоритмов  
The results of the algorithms

Название	Группы	Классы меньшие групп	Классы большие групп
Простой алгоритм разбиения	Дерево содержит выделенный узел, который разбивает дерево на группы в соответствии с визуальным представлением. При неправильном выборе параметра наблюдается выделение отдельных листьев дерева. Такой лист визуально выделен на дереве.	Много классов из одного листа, которые соответствуют визуальному представлению на дереве для отдельного генома. При подборе правильных параметров, можно получить, помимо листьев, выделение классов: 7 (из геномов 9), 10 (из геномов 3), 13 (из геномов 4) [8. Прилож. 2]. Классы соответствуют визуальному представлению на дереве.	Классы большие групп Получили класс, в котором полное присоединение группы 3 к 1. Группа 2 осталась отдельным классом. Объединение групп соответствует визуальному разбиению, поскольку расстояние между визуальным представлением групп 1 и 3 меньше, чем для 1 и 2 [8. Прилож. 2].
Алгоритм последовательного сравнения листьев	3 группы представлены в виде 4 классов, которые совпадают с визуальным представлением. Группа 1 разбита на 2 класса, это связано с тем, что группы 2 и 3 расположены внутри группы 1 [8. Прилож. 3].	Алгоритм выделяет крупные классы и классы, состоящие из одного листа, соответствующие визуальному представлению на дереве. При получении более мелких классов, противоречий с визуальным представлением не возникает. Такая точность, возможно, обусловлена тем, что дерево изначально строится, опираясь на расстояние Хэмминга.	Сначала присоединятся 3 группа к классу для части 1 группы, а потом и 2 группа к этому же классу. Объединение групп соответствует визуальному представлению, поскольку расстояние между визуальным представлением групп 1 и 3 меньше, чем для 1 и 2 [8. Прилож. 3].
Поровневый алгоритм	Дерево обладает выделенным узлом. На уровне этого узла, удалось построить разбиение для 2 и 3 групп в соответствии с визуальным представлением, но отдельно выделились классы для группы 1, которые могут не соответствовать визуальному представлению [8. Прилож. 4].	При правильно выбранном параметре получаем разбиение на классы, среди которых есть как крупные, так и мелкие. Имеются классы, состоящие из листьев, не соответствующие визуальному представлению, – их небольшое количество, что может говорить о том, что параметр разбиения подобран правильно.	Есть уровни, на которых число листьев равно 0. Для этих уровней строились классы. Получаем, что группы 2 и 3 объединяются в класс, захватывая классы группы 1, а группа 1 распадается на классы. Классы группы 1 не соответствуют визуальному представлению [8. Прилож. 4].

### Заключение

Описаны разработанные и реализованные алгоритмы кластеризации. Проведены апробация и сравнительный анализ алгоритмов кластеризации, при помощи апробации алгоритмов на геномах вируса клещевого энцефалита.

Описана архитектура информационно-аналитической системы для анализа набора геномов. Система содержит реализацию разработанных методов и может работать на наборах из любых символьных последовательностей, а не только на наборах геномов. Система помогает в анализе множества геномов и их классификации, а именно для анализа генотипов внутри одного вида живых организмов, поскольку методы направлены на выделение тонких различий геномов, имеющих схожую структуру.

Дополненный новыми функциями электронный ресурс позволит специалистам в области молекулярной биологии проводить анализ геномов, используя новые методы. На данный момент система включает новые функции, которые не использовались раньше в задачах генотипирования для вируса клещевого энцефалита.

### Список литературы / References

1. **Ковалевич А., Падутов В., Баранов О.** Полногеномное секвенирование – новый этап генетических исследований. URL: <https://cyberleninka.ru/article/n/polnogenomnoe-sekvenirovanie-novyy-etap-geneticheskikh-issledovaniy> (дата обращения 10.12.2018).  
**Kovalevich A., Padutov V., Baranov O.** Full genome sequencing – a new stage of genetic research. URL: <https://cyberleninka.ru/article/n/polnogenomnoe-sekvenirovanie-novyy-etap-geneticheskikh-issledovaniy> (in Russ.)
2. **Ружников Г. М. и др.** Современные технологии информационно-аналитической оценки // Бюл. СО РАМН. 2012. Т. 32, № 6.57. С. 55–59.  
**Ruzhnikov G. M. et al.** Modern technologies of information-analytical evaluation. *Bulletin of SB RAMS*, 2012, vol. 32, no. 6.57, p. 55–59. (in Russ.)
3. **Ливанова Н. Н., Боргояков В. Ю., Ливанов С. Г., Фоменко Н. В.** Характеристика природных очагов клещевых боррелиозов Новосибирского научного центра и Новосибирской области // Сибирский медицинский журнал. 2012. Т. 111, № 4. С. 20–23.  
**Livanova N. N., Borgoyakov V. Yu., Livanov S. G., Fomenko N. V.** Characteristics of natural foci of tick-borne borreliosis of Novosibirsk Scientific Center and Novosibirsk Region. *Siberian Medical Journal*, 2012, vol. 111, no. 4, p. 20–23. (in Russ.)
4. **Гусев В. Д., Мирошниченко Л. А., Титкова Т. Н., Джиоев Ю. П., Козлова И. В., Парамонов А. П.** Структурированные РНК-маркеры для генотипирования вируса клещевого энцефалита // Математическая биология и биоинформатика. 2018. Т. 13, № 1. С. 13–37. DOI 10.17537/2018.13.13  
**Gusev V. D., Miroshnichenko L. A., Titkova T. N., Dzhioev Yu. P., Kozlova I. V., Paramonov A. P.** Structured RNA markers for genotyping of tick-borne encephalitis virus. *Mathematical Biology and Bioinformatics*, 2018, vol. 13, no. 1, p. 13–37. (in Russ.) DOI 10.17537/2018.13.13
5. **Дёмина Т. В.** Вопросы генотипирования и анализ генетической вариабельности вируса клещевого энцефалита: Дис. ... д-ра биол. наук. Иркутск, 2013. 248 с.  
**Demina T. V.** Questions of genotyping and analysis of the genetic variability of tick-borne encephalitis virus. Dis. ... Dr. Biol. Sciences. Irkutsk, 2013. 248 p. (in Russ.)
6. **Беликов С. И., Гусев В. Д., Мирошниченко Л. А., Титкова Т. Н.** Сравнительный анализ геномов вируса клещевого энцефалита: дифференциация по степени вирулентности // Докл. IV Междунар. конф. «Математическая биология и биоинформатика» (ICMBB12). Пушино, 2012. С. 52–53.  
**Belikov S. I., Gusev V. D., Miroshnichenko L. A., Titkova T. N.** Comparative analysis of genomes of tick-borne encephalitis virus: differentiation according to the degree of virulence. In: Reports of the IV International Conference “Mathematical Biology and Bioinformatics” (ICMBB12). Pushchino, 2012, p. 52–53. (in Russ.)
7. **Панчин А. Ю.** Сумма биотехнологии. М.: АСТ, 2015. 432 с. ISBN 978-5-17-093602-1

- Panchin A. Yu.** The sum of biotechnology. Moscow, AST, 2015, 432 p. (in Russ.) ISBN 978-5-17-093602-1
8. **Черненко В. В.** Разработка архитектуры информационно-аналитической системы для работы с данными о патогенах, переносимых иксодовыми клещами: дис. магистра математики и компьютерных наук / Новосиб. нац. исслед. гос. ун-т. Новосибирск, 2018.  
**Chernenko V. V.** The development of information and analytical system architecture for working with data on pathogens transferred by ixodid ticks. Dis. Master of Mathematics and Computer Science. Novosibirsk, Novosibirsk State University, 2018. (in Russ.)
  9. **Rokach L., Maimon O.** Clustering methods. In: Data mining and knowledge discovery handbook. Springer US, 2005, p. 321–352.
  10. **Sokal R., Michener C.** A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, 1958, no. 38, p. 1409–1438.
  11. **Zuckermandl E., Pauling L. B.** Molecular disease, evolution, and genic heterogeneity. In: Kasha M., d Pullman B. (eds.). Horizons in Biochemistry. New York, Academic Press, 1962, p. 189–225.
  12. **Saitou N., Nei M.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 1987, vol. 4, iss. 4, p. 406–425.
  13. **Иванов Б. Н.** Дискретная математика. Алгоритмы и программы: Учеб. пособие. М.: Лаборатория Базовых Знаний, 2001 С. 126–130.  
**Ivanov B. N.** Discrete Mathematics. Algorithms and programs: Textbook. Moscow, Laboratory of Basic Knowledge, 2001, p. 126–130. (in Russ.)
  14. **Mohamed A.** Survey on multiclass classification methods. Technical Report, Caltech, 2005.
  15. **Гусев В. Д., Мирошниченко Л. А., Титкова Т. Н.** Сравнительный анализ близких текстов. Выявление «тонких» различий // Материалы Всерос. конф. с междунар. участием «Знания – Онтологии – Теории» (ЗОНТ-2017). Новосибирск, 2017. Т. 1. С. 109–118.  
**Gusev V. D., Miroshnichenko L. A., Titkova T. N.** Comparative analysis of related texts. Identify the “subtle” differences. In: Materials of All-Russia. Conf. with the international participation of “Knowledge – Ontology – Theory” (ZONT-2017). Novosibirsk, 2017, vol. 1, p. 109–118. (in Russ.)
  16. **Коржов В.** Многоуровневые системы клиент-сервер. М.: Открытые системы, 1997.  
**Korzhov V.** Multi-level client-server systems. Moscow, Open Systems Publ., 1997. (in Russ.)

*Материал поступил в редколлегию*

*Received*

*03.12.2018*

#### **Сведения об авторах / Information about the Authors**

**Черненко Владислав Валерьевич**, аспирант, Институт вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия)

**Vladislav V. Chernenko**, PhD student, Institute of Computational Technologies SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation)

vlavlad.chernenko@gmail.com

ORCID 0000-0001-7969-126X

**Молородов Юрий Иванович**, кандидат физико-математических наук, старший научный сотрудник Института вычислительных технологий СО РАН (пр. Академика Лаврентьева, 6, Новосибирск, 630090, Россия)

**Yuri I. Molorodov**, Candidate of Science (Physico-Mathematical), Senior Researcher, Institute of Computational Sciences SB RAS (6 Academician Lavrentiev Ave., Novosibirsk, 630090, Russian Federation)

yumo@ict.sbras.ru

ORCID 0000-0003-3055-0735