

## Быстрый метод распознавания вида флавивируса при секвенировании его генома

Чалей М.Б.<sup>1</sup>, Тюлько Ж.С.<sup>2</sup>, Кутыркин В.А.<sup>3</sup>

<sup>1</sup>*Институт математических проблем биологии РАН – филиал ИПМ РАН им. М.В. Келдыша, Пушчино, Россия*

<sup>2</sup>*ОмГМУ Минздрава России, Омск, Россия*

<sup>3</sup>*МГТУ им. Н.Э. Баумана, Москва, Россия*

[maramaria@yandex.ru](mailto:maramaria@yandex.ru)

Флавивирусы вызывают тяжелые заболевания человека, домашних и диких животных и птиц. Они поражают нервную систему и часто проявляются симптомами геморрагических лихорадок. Выявление в геномах флавивирусов скрытой профильной триплетной периодичности позволило предложить новый метод распознавания вида флавивируса (включая распознавание подтипа) по последовательности его генома. Метод основан на использовании частотных характеристик кодонов аминокислот в полных кодирующих последовательностях полипротеинов. Метод позволяет значительно упростить идентификацию выделенного вируса возбудителя лихорадки после секвенирования его генома, так как не требует процедуры поиска наибольшего сходства с известными последовательностями вирусных геномов в базе GenBank, сопровождаемой выравниванием нуклеотидных последовательностей. Высокая надежность метода подтверждается при распознавании 15 групп геномов флавивирусов различных видов и подтипов, имеющих достаточное количество представителей в GenBank. Рассматриваются десять различных видов флавивирусов, четыре подтипа вируса лихорадки денге и вирус Кунджин, как подтип вируса лихорадки Западного Нила.

*Ключевые слова: геном флавивируса, скрытая профильная триплетная периодичность, частоты кодонов аминокислот, распознавание вида флавивируса.*

## Fast Method to Recognize Flavivirus Species after Sequencing the Viral Genome

Chaley M.B.<sup>1</sup>, Tyulko Zh.S.<sup>2</sup>, Kutyrkin V.A.<sup>3</sup>

<sup>1</sup>*Institute of Mathematical Problems of Biology RAS – the Branch of Keldysh Institute of Applied Mathematics RAS, Pushchino, Russia*

<sup>2</sup>*Omsk State Medical University of Ministry of Healthcare of the Russian Federation, Omsk, Russia*

<sup>3</sup>*Moscow State Technical University n.a. N.E. Bauman, Moscow, Russia*

Flaviviruses give rise to heavy diseases of human, domestic and wild animals and birds. They destroy nervous system and frequently develop the symptoms of hemorrhagic fevers. Latent profile periodicity recognition in the flaviviruses' genomes allowed proposing a new method to detect flavivirus species (including the serotypes detection) while analyzing sequence of the viral genome. The method is based on amino acid codons' frequency characteristics in the whole coding sequences of viral polyproteins. It allows significant simplifying for identification of isolated fever virus after its genome has been sequenced, so the method does not require an alignment procedure with other known viral genomes in the GenBank to find the most similarity to their nucleotide sequences. High reliability of the method is proved while recognizing 15 groups for the genomes of different species and serotypes of flaviviruses which have enough representatives in the GenBank. Ten different species of the flaviviruses, four serotypes of Dengue fever virus and Kunjin virus, as a serotype of West Nile fever virus, are considered in the work.

*Key words: flavivirus genome, latent profile triplet periodicity, frequencies of amino acid codons, flavivirus species recognition.*

## 1. Введение

Многие флавивирuses рода *Flavivirus* из семейства *Flaviviridae* являются опаснейшими патогенами для человека, домашних и диких животных и птиц [1–5]. Они распространяются при укусах насекомых (клещи, комары) и могут вызывать парезы, параличи, энцефалиты и геморрагические лихорадки с высокой летальностью. В настоящее время активизация туризма, деловых связей и миграционных процессов породили проблему завозных случаев вирусных заболеваний из тропических и субтропических стран в неэндемичные районы, в том числе и в Россию. По всему миру распространяются опасные болезни тропического и субтропического пояса, вызываемые такими флавивирuses, как вирусы желтой лихорадки, Зика, Дэнге, Западного Нила, японского энцефалита [6]. В России давно актуальна проблема эпидемического распространения вирусом клещевого энцефалита, Западного Нила, Повассан и др. [7–10].

Для идентификации вида или подтипа флавивируса, вызвавшего заболевание, наряду с традиционными методами иммуноферментного анализа применяется метод обратной транскрипции и полимеразной цепной реакции с дальнейшей процедурой секвенирования cDNA генома вируса [11]. Полученную в результате нуклеотидную последовательность сравнивают с известными геномными последовательностями вирусом в GenBank [12]. При сравнении и изучении эволюции геномов вирусом также используются частотные характеристики их нуклеотидного, динуклеотидного состава, предпочтительного использования кодонов [13–15].

Геном флавивируса представлен одноцепочечной РНК положительной полярности длиной около 11 тыс. нуклеотидов. Геномная РНК кодирует три структурных (С, ргМ/М, Е) и семь неструктурных белков [16]. Последовательно расположенные друг за другом в единой рамке считывания, гены вирусных белков образуют полную кодирующую последовательность полипротеина (СДС полипротеина).

В настоящей работе исследовались 15 групп СДС полипротеинов общим числом 7060. Анализировались полипротеины десяти видов флавивирусов, среди которых вирус Багаза (*Bagaza*), японский энцефалит (*Japanese encephalitis*), энцефалит долины Мюррея (*Murray Valley encephalitis*), Повассан (*Powassan*), энцефалит Сент-Луис (*Saint Louis encephalitis*), утиный вирус Тембусу (*Duck Tembusu*), вирус клещевого энцефалита (*Tick-borne encephalitis*), вирус Усуту (*Usutu*), вирусы лихорадки Западного Нила и желтой лихорадки. Кроме того, отдельно анализировались группы СДС полипротеинов для четырех подтипов вируса лихорадки денге (*Dengue serotype 1*, *Dengue serotype 2*, *Dengue serotype 3*,

*Dengue serotype 4*) и вирус Кунджин (*Kunjin*) как подтип вируса лихорадки Западного Нила, распространенного в Австралии.

Выявление частотных характеристик кодонов аминокислот для каждой группы СДС полипротеинов и идентификация отдельных СДС геномов флавивирусов выполнялись на основе предложенной ранее стохастической модели организации кодирования в последовательностях ДНК (текстовых строках) [17].

### 1.1. Модель организации кодирования в СДС полипротеина

Модель имеет вид случайной строки, полученной на основе последовательных и независимых реализаций случайного кодона размера три. Такой случайный кодон является случайной величиной, имеющей в качестве значений 61 кодон аминокислот и три кодона терминации. Каждое из 64 значений случайного кодона характеризуется соответствующей вероятностью реализации этого значения, так что сумма этих вероятностей равна единице. Стохастическая модель на основе случайного кодона получила название SHOC-модели (*Stochastic Homogeneous Organization of Coding*). Формально SHOC-модель можно описать случайной строкой  $\underbrace{Cdn\ Cdn\ \dots\ Cdn}_{m\ раз}$  длиной  $n = 3m$ , где символ *Cdn*

обозначает упомянутый выше случайный кодон. Этот кодон однозначно характеризуется своим вероятностным распределением  $P = (p_1, p_2, \dots, p_{64})$  64-х значений, где  $p_i$  – вероятность реализации  $i$ -го кодона из списка кодонов генетического кода. С точки зрения SHOC-модели, СДС полипротеина рассматривается как реализация случайной строки, представляющей эту модель. Следовательно, СДС полипротеина определяет выборочное распределение  $\hat{P} = (\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{64})$  кодонов генетического кода, являющееся оценкой вероятностного распределения  $P$  случайного кодона *Cdn*.

Таким образом, все СДС полипротеинов флавивирусов, анализируемых в работе 15 групп, индуцируют множество соответствующих выборочных вероятностных распределений, образующих множество точек в 64-мерном пространстве. В настоящей работе рассматривается задача кластеризации этого множества точек, где отдельный кластер характеризуется только одним из рассматриваемых видов или подтипов флавивирусов. В работе предложен алгоритм, который позволяет по выборочному вероятностному распределению кодонов СДС полипротеина однозначно отнести её к кластеру соответствующего вида или подтипа флавивируса.

Разработанный в настоящей работе алгоритм по распознаванию вида и подтипа флавивирусов основывается на предварительной обучающей

процедуре. Каждая выборка CDS полипротеинов, относящаяся к флавивирусам одной группы, делилась на две численно равные выборки, одна из которых служила обучающей, а вторая рассматривалась как тестируемая (таблица 1). Предложенный в работе алгоритм распознавания вида флавивируса создан на основе этой обучающей процедуры. Этот алгоритм показал практически 100 % эффективность. Результаты применения этого алгоритма демонстрируются в работе на основе полной выборки.

## 2. Материалы и методы

В анализируемой версии GenBank (выпуск 231 от 15.04.2019) содержались геномы порядка сотни различных видов флавивирусов, из которых были отобраны геномы для десяти видов и пяти подтипов вирусов, имеющих наибольшее количество представителей в GenBank.

Для выявления структурно-статистических свойств последовательностей ДНК ранее был предложен спектрально-статистический подход [17]. Согласно этому подходу, последовательность ДНК рассматривается как реализация некоторой случайной строки. Кроме того, этот подход позволяет ввести понятие профильной эквивалентности для широкого класса случайных строк. В реализациях профильно-эквивалентных строк выявляется один и тот же тип скрытой профильной периодичности.

В качестве наиболее общей модели, объясняющей наличие профильной скрытой периодичности, ранее была предложена SHOC-модель [17]. Как отмечалось во введении, эту модель однозначно характеризует случайный кодон, являющийся случайной величиной со значениями в текстовых строках из трех букв заданного алфавита. Для кодирующих районов ДНК в качестве значений такого случайного кодона выступают триплеты генетического кода. Следовательно, в рамках SHOC-модели для кодирующей последовательности ДНК, где наблюдается скрытая триплетная периодичность, можно получить оценку вероятностного распределения случайного кодона в виде набора (вектора) частот кодонов генетического кода, встречающихся в этой последовательности. Поэтому такие последовательности ДНК можно классифицировать по типам распределений случайных кодонов, что и предлагается использовать для распознавания видов флавивирусов на основе анализа распределений частот кодонов аминокислот в CDS их полипротеинов.

В работе рассматривались 15 групп CDS полипротеинов флавивирусов. Виды и подтипы флавивирусов и количество CDS полипротеинов  $N_j$  (в полной и обучающей выборках) в группе представлены в таблице 1, где за каждой группой закрепляется её номер  $j$  ( $j = \overline{1, 15}$ ).

Пусть  $n = \overline{1, N_i}$  – порядковый номер гена полипротеина в группе  $i$ -го вида флавивируса. Тогда  $\mathbf{P}^{(i)}(n)$  – вектор распределения частот кодонов аминокислот этого гена имеет вид:

$$\mathbf{P}^{(i)}(n) = (p_1^{(i)}(n), p_2^{(i)}(n), \dots, p_k^{(i)}(n), \dots, p_{64}^{(i)}(n)). \quad (1)$$

**Таблица 1.** Количественный состав 15 групп анализируемых CDS полипротеинов флавивирусов

$j$	Виды и подтипы флавивирусов	Полная выборка $N_j$	Обучающая выборка $N_j$
1	Bagaza	18	9
2	Dengue serotype 1	1895	947
3	Dengue serotype 2	1412	706
4	Dengue serotype 3	922	461
5	Dengue serotype 4	190	95
6	Japanese encephalitis	300	150
7	Kunjin virus	43	21
8	Murray Valley encephalitis	14	7
9	Powassan	21	10
10	Saint Louis encephalitis	38	19
11	Duck Tembusu	110	55
12	Tick-borne encephalitis	175	87
13	Usutu	147	73
14	West Nile	1652	826
15	Yellow fever	123	61

Обозначим  $e_k^{(j)}$  среднюю частоту  $k$ -го кодона ( $k = \overline{1, 64}$ ) в группе  $j$ -го вида флавивируса, т. е.

$$e_k^{(j)} = \frac{1}{N_j} \sum_{n=1}^{N_j} p_k^{(j)}(n). \quad (2)$$

Тогда, согласно формулам (2), вектор средних частот кодонов вирусов  $j$ -го типа примет вид:

$$\mathbf{E}^{(j)} = (e_1^{(j)}, e_2^{(j)}, \dots, e_k^{(j)}, \dots, e_{64}^{(j)}). \quad (3)$$

Для распознавания векторов, соответствующих одному виду вируса, вводятся величины  $r_{(j)}^{(i)}(n) = r(\mathbf{P}^{(i)}(n), \mathbf{E}^{(j)})$  отклонения («расстояния») вектора частот кодонов  $n$ -й CDS полипротеина вируса  $i$ -го вида (или подтипа) от вектора средних частот кодонов в CDS вируса  $j$ -го вида (или подтипа). Для расчета этого отклонения используется формула:

$$r_{(j)}^{(i)}(n) = r(\mathbf{P}^{(i)}(n), \mathbf{E}^{(j)}) = \gamma \sum_{k=1}^{64} \frac{|p_k^{(i)}(n) - e_k^{(j)}|}{e_k^{(j)}}, \quad (4)$$

где для удобства анализа результатов вычислений выбрано значение  $\gamma = \frac{1}{7}$ .

Кроме того, для демонстрации принципа распознавания вида (подтипа) вируса по вектору распределения частот кодонов в CDS полипротеина в работе вводится величина  $r_{(j)}^{(i)}$ , представляющая среднее отклонение векторов частот кодонов в CDS вирусов  $i$ -го вида (подтипа) от вектора средних частот кодонов в вирусах  $j$ -го вида (подтипа):

$$r_{(j)}^{(i)} = \frac{1}{N_i} \sum_{n=1}^{N_i} r_{(j)}^{(i)}(n). \quad (5)$$

Как показано в работе [18] диагональная компонента  $r_{(i)}^{(i)}$  матрицы  $R = (r_{(j)}^{(i)})_{15}^{15}$ ,  $i = \overline{1,15}$ , таких величин (верхний индекс – строка, нижний – столбец) как для обучающей выборки, так и для полной выборки является минимальной.

Из указанных выше свойств матрицы  $R$  для распознавания вида (подтипа) вируса по CDS полипротеина в работе предлагается следующий принцип. Если  $\mathbf{P} = (p_1, p_2, \dots, p_{64})$  – вектор частот кодонов аминокислот анализируемой CDS, то номер  $m$  распознаваемого вида (подтипа) вируса удовлетворяет условию:

$$r(\mathbf{P}, \mathbf{E}^{(m)}) = \min\{r(\mathbf{P}, \mathbf{E}^{(j)}) : j = \overline{1,15}\}, \quad (6)$$

где, согласно формуле (4):

$$r(\mathbf{P}, \mathbf{E}^{(j)}) = \gamma \sum_{k=1}^{64} \frac{|p_k - e_k^{(j)}|}{e_k^{(j)}} \text{ для } j = \overline{1,15}. \quad (7)$$

Следовательно, для анализируемой CDS полипротеина выбирается тот вид (или подтип) вируса, на котором достигается минимальное отклонение вектора частот кодонов аминокислот CDS полипротеина от среднего вектора частот кодонов аминокислот этого вида (подтипа). Тестирование этого принципа для обучающей выборки показало его практически 100 % эффективность. Поэтому этот же принцип использовался для распознавания по полной выборке.

Кроме того, для подтверждения правильности выбранного вида (подтипа) вируса в работе используются величины:

$$M^{(j)} = \max\{r_{(j)}^{(j)}(n) : n = \overline{1, N_j}\}, j = \overline{1,15}, \quad (8)$$

где  $M^{(j)}$  – максимальное отклонение («расстояние») среди векторов частот кодонов для CDS полипротеинов вирусов  $j$ -го типа от вектора средних частот кодонов вируса этого вида (подтипа). Значения величин  $M^{(j)}$  ( $j = \overline{1,15}$ ) для полной выборки приведены в таблице 2.

Указанный выше принцип, согласно величинам из формулы (8), будет дополнен следующим правилом. Если, согласно принципу, для CDS

полипротеина с вектором частот кодонов аминокислот  $\mathbf{P}$  выбран  $j$ -й вид (подтип) вируса, то этот выбор остается в силе, при выполнении условия (см. формулу (7)):

$$r(\mathbf{P}, \mathbf{E}^{(j)}) \leq M^{(j)}. \quad (9)$$

Если условие (9) нарушено, то считаем, что вид (или подтип) вируса для анализируемой CDS полипротеина не определен.

**Таблица 2.** Значения максимального отклонения  $M^{(j)}$  среди векторов частот кодонов аминокислот для CDS полипротеинов вирусов одного вида (подтипа) от соответствующего им вектора средних частот по каждому из рассматриваемых в работе виду или подтипу ( $j$ -той группе вирусов,  $j = \overline{1,15}$ )

$j$	$M^{(j)}$	$j$	$M^{(j)}$
1	0.558	9	0.552
2	1.261	10	1.241
3	1.714	11	0.941
4	0.882	12	0.881
5	0.785	13	0.946
6	1.529	14	1.655
7	0.429	15	1.402
8	0.768		

### 3. Результаты

Применение предложенного метода распознавания вида или подтипа флавивируса по анализируемому вектору частот кодонов аминокислот в CDS полипротеина показало следующий результат. Для 12 групп вирусов (см. табл. 1), а именно: Bagaza (№ 1), Dengue подтип 1 (№ 2), Dengue подтип 4 (№ 5), Kunjin (№ 7), Murray Valley encephalitis (№ 8), Powassan (№ 9), Saint Louis encephalitis (№ 10), Duck Tembusu (№ 11), Tick-borne encephalitis (№ 12), Usutu (№ 13), Yellow fever (№ 15), ошибок в распознавании вируса не было.

При распознавании остальных трех групп флавивирусов ошибки составляли менее 0.5 % от общего количества CDS полипротеинов этих типов. При этом было всего два случая неправильного определения вида вируса. CDS полипротеина вируса West Nile (GenBank код доступа JN887352) была признана за CDS вируса Kunjin, и CDS полипротеина вируса West Nile (GenBank код доступа FJ159130) была признана за CDS вируса Japanese encephalitis. При распознавании CDS полипротеинов вируса Dengue подтипа 2 вид вируса не был определен в 16 случаях. Аналогично вид вируса не был определен для трех CDS полипротеинов вируса Japanese encephalitis и четырех CDS полипротеинов вируса West Nile.

## 4. Заключение

Практика секвенирования вирусного генома в последнее время используется наряду с методами иммуноферментного анализа для точной идентификации возбудителя заболевания. В работе предложен новый, эффективный и быстрый метод определения вида флавивируса, на основе расшифрованной последовательности его генома.

Поскольку профильная триплетная периодичность является неотъемлемой характеристикой любой достаточно длинной CDS, частотные характеристики кодонов аминокислот, на которые опирается предлагаемый метод, для CDS, выбранной в качестве идентификатора, могут использоваться для определения или подтверждения вида вируса. Так как скрытая профильная триплетная периодичность выявляется во всех анализируемых в работе CDS полипротеинов флавивирусов, предлагаемый метод весьма точно определяет вид флавивируса по его геному. Метод протестирован в распознавании 10 видов и пяти подтипов флавивирусов, геномы которых на сегодняшний день имеют значительное представительство в GenBank. Показано, что для 10 видов и двух подтипов флавивирусов метод показал 100 % надежности. Ошибки распознавания вида вируса не превысили 0.5 %.

Применение предлагаемого метода не ограничивается только идентификацией видов РНК-содержащих флавивирусов. Возможно его использование и для подтверждения вида вирусов других родов и семейств, безотносительно того, представлен геном вируса последовательностью ДНК или РНК.

Предложенный метод существенно упрощает процедуру идентификации вида вируса по последовательности его секвенированного генома, так как для применения метода не требуется знаний о наличии и расположении консервативных участков вирусного генома, которые существенны при сравнении анализируемой последовательности с известными нуклеотидными последовательностями генотипов вирусов, депонированных в международных базах данных.

## 5. Благодарности

Работа была выполнена при поддержке гранта РФФИ № 19-07-00996.

## 6. Список литературы

1. Fernández-Pinero J., Davidson I., Elizalde M., Perk S., Khinich Y., Jiménez-Clavero M.A. Bagaza virus and Israel turkey meningoencephalomyelitis virus are a single virus species. *J. Gen. Virol.* 2014. V. 95. P. 883–887. doi: [10.1099/vir.0.061465-0](https://doi.org/10.1099/vir.0.061465-0).
2. Zhang W., Chen S., Mahalingam S., Wang M., Cheng A. An updated review of avian-origin Tembusu virus: a newly emerging avian

- Flavivirus. *J. Gen. Virol.* 2017. V. 98. P. 2413–2420. doi: [10.1099/jgv.0.000908](https://doi.org/10.1099/jgv.0.000908).
3. Benzarti E., Linden A., Desmecht D., Garigliany M. Mosquito-borne epornitic flaviviruses: an update and review. *J. Gen. Virol.* 2019. V. 100. P. 119–132. doi: [10.1099/jgv.0.001203](https://doi.org/10.1099/jgv.0.001203).
4. Diaz A., Coffey L.L., Burkett-Cadena N., Day J.F. Reemergence of St. Louis encephalitis virus in the Americas. *Emerg. Infect. Dis.* 2018. V. 24. P. 2150–2157. doi: [10.3201/eid2412.180372](https://doi.org/10.3201/eid2412.180372).
5. Clé M., Beck C., Salinas S., Lecollinet S., Gutierrez S., Van de Perre P., Baldet T., Foulongne V., Simonin Y. Usutu virus: A new threat? *Epidemiol. Infect.* 2019. V. 147. Article No. e232. doi: [10.1017/S0950268819001213](https://doi.org/10.1017/S0950268819001213).
6. Holbrook M.R. Historical Perspectives on Flavivirus Research. *Viruses.* 2017. V. 9. No. 5. P. 97. doi: [10.3390/v9050097](https://doi.org/10.3390/v9050097).
7. Локтев В.Б. Вирус клещевого энцефалита. Генетические особенности и его изменчивость в современном мире. *Бюллетень СО РАМН.* 2007. № 4. С. 14–21.
8. Субботина Е.Л., Локтев В.Б. Молекулярная эволюция вируса клещевого энцефалита и вируса Повассан. *Молекулярная биология.* 2012. Т. 46. С. 82–91.
9. Путинцева Е.В., Алексейчик И.О., Чеснокова С.Н., Удовиченко С.К., Бородай Н.В., Никитин Д.Н., Агаркова Е.А., Батулин А.А., Шпак И.М., Фомина В.К. и др. Результаты мониторинга возбудителя лихорадки Западного Нила в Российской Федерации в 2019 г. и прогноз развития эпидемической ситуации на 2020 г. *Проблемы особо опасных инфекций.* 2020. Т. 1. С. 51–60. doi: [10.21055/0370-1069-2020-1-51-60](https://doi.org/10.21055/0370-1069-2020-1-51-60).
10. Колясникова Н.М., Герасимов С.Г., Ишмухаметов А.А., Погодина В.В. Эволюция клещевого энцефалита за 80-летний период: основные проявления, вероятные причины. *Эпидемиология и вакцинопрофилактика.* 2020. Т. 19. № 3. С. 78–88. doi: [10.31631/2073-3046-2020-19-3-78-88](https://doi.org/10.31631/2073-3046-2020-19-3-78-88).
11. *Бюллетень нормативных и методических документов госсанэпиднадзора.* 2016. Т. 3. № 65. С. 25–38.
12. *GenBank.* URL: <https://www.ncbi.nlm.nih.gov/genbank/> (accessed 20.09.2020).
13. Jenkins G., Pagel M., Gould E., de A. Zannotto P.M., Holmes E.C. Evolution of base composition and codon usage bias in the genus Flavivirus. *J. Mol. Evol.* 2001. V. 52. P. 383–390. doi: [10.1007/s002390010168](https://doi.org/10.1007/s002390010168).
14. Belalov I.S., Lukashev A.N. Causes and implications of codon usage bias in RNA viruses. *PLoS One.* 2013. V. 8. № 2. Article No. e56642. doi: [10.1371/journal.pone.0056642](https://doi.org/10.1371/journal.pone.0056642).
15. Di Giallonardo F., Schlub T.E., Shi M., Holmes E.C. Dinucleotide composition in animal

- RNA viruses is shaped more by virus family than by host species. *J. Virol.* 2017. V. 91. Article No. e02381-16. doi: [10.1128/jvi.02381-16](https://doi.org/10.1128/jvi.02381-16).
16. *Руководство по вирусологии: Вирусы и вирусные инфекции человека и животных.* Под ред. Львова Д.К. М.: Медицинское информационное агентство, 2013.
  17. Chaley M., Kutyrkin V. Stochastic Models for Description of Structural-Statistical Properties in DNA Sequences. *J. Theor. Biol.* 2020. V. 496. P. 110126. doi: [10.1016/j.jtbi.2019.110126](https://doi.org/10.1016/j.jtbi.2019.110126).
  18. Чалей М.Б., Тюлько Ж.С., Кутыркин В.А. Распознавание видов флавивирусов на основе кодирующих последовательностей полипротеинов. *Математическая биология и биоинформатика.* 2019. Т. 14. № 2. С. 533–542. doi: [10.17537/2019.14.533](https://doi.org/10.17537/2019.14.533).