

Н.Г. Загоруйко<sup>1,2</sup>, Ю.П. Джгоев<sup>3,4</sup>, И.В. Козлова<sup>3,4</sup>, С.Е. Ткачев<sup>5</sup>, Ю.С. Букин<sup>6</sup>,  
А.И. Парамонов<sup>4</sup>, А.Ю. Борисенко<sup>3</sup>, И.А. Борисова<sup>1,2</sup>, В.И. Злобин<sup>4</sup>, О.А. Кутненко<sup>1,2</sup>

## ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ ПРЕДСКАЗАНИЯ ДЛЯ ВЫЯВЛЕНИЯ ЗАКОНОМЕРНЫХ СВЯЗЕЙ МУТАЦИЙ В БЕЛКАХ ВИРУСА КЛЕЩЕВОГО ЭНЦЕФАЛИТА И ПРОЯВЛЕНИЙ ЕГО ФЕНОТИПИЧЕСКИХ ПРИЗНАКОВ

<sup>1</sup> Институт математики им. С.Л. Соболева СО РАН (Новосибирск)

<sup>2</sup> Новосибирский государственный исследовательский университет (Новосибирск)

<sup>3</sup> Научный центр проблем здоровья семьи и репродукции человека (Иркутск)

<sup>4</sup> Иркутский государственный медицинский университет (Иркутск)

<sup>5</sup> Институт химической биологии и фундаментальной медицины СО РАН (Новосибирск)

<sup>6</sup> Лимнологический институт СО РАН (Иркутск)

Исследуются закономерные связи между аминокислотными последовательностями белков C, prM, E и NS1 штаммов вируса клещевого энцефалита (ВКЭ) и их тремя фенотипическими признаками – нейроинвазивностью, термостабильностью и терморезистентностью. Интенсивность связи оценивается с помощью меры конкурентного сходства. Выбираются подмножества сечений штамма, по особенностям аминокислотного состава которых можно предсказывать значение интересующего фенотипического признака. Показана возможность предсказания пропущенных элементов в данных, как в аминокислотном составе, так и в целевых свойствах. Оценены зависимости между парами фенотипических признаков штаммов.

**Ключевые слова:** вирус клещевого энцефалита, нейроинвазивность, термостабильность, терморезистентность, математические методы предсказания

## APPLICATION OF MATHEMATICAL METHOD PREDICTIONS FOR IDENTIFICATION OF PATTERNS RELATIONS MUTATIONS IN PROTEINS ENCEPHALITIS VIRUS AND A MANIFESTATION OF ITS PHENOTYPIC TRAITS

N.G. Zagoruyko<sup>1,2</sup>, Yu.P. Dzhioev<sup>3,4</sup>, I.V. Kozlova<sup>3,4</sup>, S.E. Tkachev<sup>5</sup>, Yu.S. Bukin<sup>6</sup>,  
A.I. Paramonov<sup>4</sup>, A.Yu. Borisenko<sup>3</sup>, I.A. Borisova<sup>1,2</sup>, V.I. Zlobin<sup>4</sup>, O.A. Kutnenko<sup>1,2</sup>

<sup>1</sup> S. L. Sobolev Institute of Mathematics SB RAS, Novosibirsk

<sup>2</sup> Novosibirsk State Research University, Novosibirsk

<sup>3</sup> Scientific Center for Family Health Problems and Human Reproduction, Irkutsk

<sup>4</sup> Irkutsk State Medical University, Irkutsk

<sup>5</sup> Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk

<sup>6</sup> Limnological Institute SB RAS, Irkutsk

We studied the natural connections between the amino acid sequences of proteins C, prM, E and NS1 virus strains of tick-borne encephalitis (TBE) and their three phenotypic traits – neuroinvasiveness, thermal stability and thermoresistance. Coupling strength is assessed using measures of competitive sequence similarity of each strain with reference strains. For such purposes subsets of strain sections are chosen amino acid composition specifics of which can predict the value of a phenotypic trait of interest. The possibility to predict missing elements in data both in amino acid composition, and in target properties is demonstrated. The relationships between pairs of phenotypic traits of strains were evaluated.

**Key words:** TBE virus, neuroinvasiveness, thermal stability, thermoresistance, mathematical prediction methods

### ВВЕДЕНИЕ

Клещевой энцефалит является одной из наиболее значимых трансмиссивных природно-очаговых вирусных инфекций лесной зоны Евразийского континента. Его возбудителем является вирус клещевого энцефалита (ВКЭ), который относится к семейству *Flaviviridae* [10]. Геномная РНК ВКЭ содержит единую открытую рамку считывания, кодирующую 3 структурных и 7 неструктурных белков [12]. Основными считаются три генотипа ВКЭ – дальневосточный, сибирский и западный, но выявлена циркуляция представителей еще двух предполагаемых генотипов (с прототипными штаммами 886-84 и 178-79) в природных очагах Восточной Сибири [1, 6, 9]. Считается, что наблюдаемые различия в патогенности флавивирусов связаны с точечными мутациями в гликопротеине оболочечного белка Е, которые при-

водят к потере вирулентности (нейровирулентности или нейроинвазивности) и аттенуации флавивирусов [15]. Есть также данные, что мутации в белках NS1 [15], NS3 [8], М [11], NS4В [16] и С [14] могут быть вовлечены в процесс аттенуации вирусов и изменения их вирулентности. Также было продемонстрировано, что даже небольшие делеции в консервативной части 3'-концевой некодирующей области вызывают ослабление их инфекционности, а увеличение размеров таких делеций приводит к нежизнеспособности вирионов [13]. Тем не менее, несмотря на относительную изученность основных функций ряда вирусных белков, вопрос о связи нуклеотидной последовательности геномов ВКЭ (и, соответственно, аминокислотных последовательностей белков) с их биологическими свойствами до сих пор остается малоизученным. Большинство проводимых исследований в данной

области касаются только изучения точечных мутаций в определенных областях геномов, в то время как комплексных исследований последовательностей полноразмерных геномов ВКЭ и их связи с наблюдаемыми биологическими свойствами практически не проводится.

**Цель** данного этапа исследований состояла в том, чтобы показать применимость предлагаемых математических методов для предсказания связи фенотипических признаков – нейроинвазивности, термостабильности и терморезистентности штаммов ВКЭ с определенными мутациями в их белковых структурах, кодируемых структурными и неструктурными генами.

# МАТЕРИАЛЫ И МЕТОДЫ

**Исходные данные и решаемые задачи.** Анализируемая коллекция состояла из 16 изолированных нами на территории Восточной Сибири штаммов ВКЭ, для каждого из которых были определены нуклеотидные последовательности фрагментов генома и выведены аминокислотные последовательности, соответствующие структурным белкам С, рМ, Е и

началу неструктурного белка NS1, и штамма 205, используемого для производства отечественной вакцины Энцевир (в базе данных GenBank) (табл. 1).

Для выбранных штаммов известны значения трех фенотипических признаков ВКЭ (целевых характеристик) (табл. 2):

а) X1: **Термостабильность** (T50) – разница lg титров вируса при заражении культуры клеток вирусосодержащей суспензией, прогретой в течение 20 мин. при температуре 50 °С и непрогретой вирусосодержащей суспензией, с дальнейшим культивированием при 37 °С.

б) X2: **Терморезистентность** (rct42) – это способность вируса к репродукции при супраоптимальной температуре (rct42). Оценивалась по разнице lg титров вируса при культивировании штаммов на культурах клеток при температуре 37 °С и 42 °С. Для определения rct 42 использовали метод титрования вируса на культуре клеток, описанный выше.

в) X3: **Нейроинвазивность** (NI) – разница в lg титров вируса при двух видах заражения мышей: церебральном (интрацеребральном титровании вирусосодержащей суспензии mNc) и подкожном (введении mNsc экстраневрально). Заражали беспло-

Таблица 1

Используемые штаммы вируса клещевого энцефалита

№	Наименование штамма	№ в GenBank	Источник изоляции	Район сбора материала
1	617-90	EU878283	<i>Ix. persulcatus</i>	РБ, Бичурский р-н
2	418-90	Не депон.	<i>I. persulcatus</i>	Читинская обл.
3	886-84	EF469662	Полевка красно-серая	ИО, Эхирит-Булагатский р-н
4	711-84	EU878281	Полевка красно-серая	РБ, Баргузинский р-н.
5	740-84	EU878282	Полевка красно-серая	РБ, Бичурский р-н
6	712-89	JN936350	<i>I. persulcatus</i>	Читинская обл.
7	210-79	JN936336	Сыворотка крови больного	Иркутская область
8	Луковка-3	JN936369	<i>I. persulcatus</i>	ИО, Эхирит-Булагатский р-н
9	Хим-Дым-6	JN936364	<i>I. persulcatus</i>	ИО, Эхирит-Булагатский р-н
10	769-90	JN936360	<i>I. persulcatus</i>	Читинская обл.
11	767-90	JN936359	<i>I. persulcatus</i>	Читинская обл.
12	761-90	JN936357	<i>I. persulcatus</i>	Читинская обл.
13	691-90	Не депон.	<i>I. persulcatus</i>	РБ, Бичурский р-н
14	757-90	JN936356	<i>I. persulcatus</i>	Читинская обл.
15	48-06	JN936330	Полевка-экономка	ИО, Иркутский р-н
<b>Контрольная группа</b>				
16	178-79	EF469661	<i>I. persulcatus</i>	ИО, Боханский р-н
17	205	DQ989336	<i>I. persulcatus</i>	Дальний Восток

Примечание. РБ – Республика Бурятия; ИО – Иркутская область.

Таблица 2

Биофизические свойства штаммов

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
X1 (T50)	2,33	1	3,78	2,36	0,23	1,77	0	1,27	1	0,28	1,33	1	1	2,5	1,67	0,7	0,6
X2 (rct42)	4,3	1,8	0,2	0,5	2,4	1,8	1,7	2,5	2,1	2,3	0,8	2,2	1,8	0,8	3,3	0,7	*
X3 (NI)	2,8	1,5	1,6	2,2	0,8	1,1	2,2	*	*	2,3	2,6	*	1,5	*	3	*	*

Примечание. \* – нет данных.

родных белых мышей массой 5–7 г в мозг по 0,03 мл, наблюдали в течение 14 дней. В опытах титрования на животных титры вируса определяли также по методу Рида и Менча [7] и выражали в lg LD50/мл.

### РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Часто в таблицах экспериментальных данных имеется избыточность в виде похожих друг на друга объектов (в нашем случае – штаммов), а также похожих друг на друга характеристик (в нашем случае – позиций штаммов). Пользуясь этими закономерными зависимостями можно по характеристикам одних объектов предсказать пропущенные характеристики других объектов. Для заполнения пробелов в таблицах с номинальными характеристиками использовался алгоритм WANGA-N [2]. Предварительно каждая позиция в аминокислотной последовательности оценивалась по ее информативности, которую можно определить

по энтропии символов, стоящих на этой позиции у разных штаммов. Чем больше разнообразие символов на данной позиции, тем выше энтропия и тем больший интерес она представляет для анализа. По этому критерию из рассмотрения были исключены позиции, на которых у всех штаммов данной выборки находится один и тот же аминокислотный символ. В результате, не вычеркнутыми остались 53 позиции, так что анализировалась таблица  $\{x_{nm}\}$ ,  $n = 1, \dots, N$ ,  $m = 1, \dots, M$ , из  $M = 17$  объектов (штаммов) и  $N = 53$  признаков (позиций). В результате предобработки в таблице аминокислот пустые клетки были заполнены при помощи алгоритма WANGA-N. После этого были выбраны наиболее информативные позиции, в которых не менее четырех штаммов имели аминокислоты, отличные от аминокислот остальных штаммов, то есть вероятность обнаружить любой символ в этой позиции была меньше 80 %. Таких позиций оказалось  $N = 27$  (табл. 3).

Таблица 3

Наиболее информативные сечения штаммов ВКЭ

Штамм →	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Позиция ↓																	
J = 80	R	R	R	R	R	R	K	K	K	K	K	K	R	K	K	K	K
J = 91	K	K	K	K	K	K	R	R	R	R	R	R	K	R	H	R	R
J = 95	R	R	R	R	R	R	K	K	K	K	K	K	R	K	K	K	K
J = 98	V	V	V	V	V	V	T	T	T	T	T	T	V	T	T	T	T
J = 102	T	T	T	T	T	T	M	M	M	M	M	M	T	M	M	M	M
J = 103	S	S	S	S	S	S	G	G	G	G	G	G	S	G	G	G	G
J = 108	A	A	A	A	A	A	T	T	T	T	T	T	A	T	T	T	T
J = 109	I	I	I	I	I	I	V	V	V	V	V	V	G	I	V	V	V
J = 121	K	K	K	K	K	K	K	R	R	K	K	K	K	K	K	K	R
J = 123	R	R	R	R	R	R	G	G	G	G	G	G	R	G	G	G	G
J = 216	D	D	D	D	D	D	E	E	E	E	E	E	D	E	D	E	E
J = 246	L	L	V	V	V	V	L	L	L	L	L	L	L	L	M	L	L
J = 252	I	I	I	I	I	I	V	V	V	V	V	V	I	V	I	V	V
J = 254	I	I	I	I	I	I	V	V	V	V	V	V	I	V	V	V	V
J = 268	I	V	V	V	V	V	I	V	V	I	I	I	V	I	V	V	V
J = 270	A	A	A	A	A	V	V	V	V	V	V	V	A	V	V	V	V
J = 361	T	T	T	T	T	T	A	T	T	A	A	A	T	A	T	A	T
J = 399	A	A	A	A	A	A	V	A	A	A	V	V	A	V	A	V	A
J = 400	S	S	S	S	S	S	A	A	A	A	A	A	S	A	S	A	A
J = 514	N	N	N	N	N	N	Q	H	H	Q	Q	Q	N	Q	Q	Q	H
J = 593	T	T	T	T	T	T	A	A	A	A	A	A	T	A	A	A	A
J = 597	I	I	I	I	I	T	T	T	T	T	T	T	I	T	T	T	T
J = 611	A	A	A	A	A	A	T	T	T	T	T	T	A	T	T	T	T
J = 623	R	K	R	R	K	R	R	R	R	R	R	R	K	R	H	R	R
J = 688	R	R	R	R	R	R	K	K	K	K	K	K	R	K	K	K	K
J = 728	L	L	L	L	L	L	I	I	I	I	I	I	L	I	L	I	I
J = 735	I	I	I	I	I	I	L	L	L	L	L	L	I	L	L	L	L

По этой таблице требовалось оценить проявления закономерной связи между значением целевой характеристики  $X_i$ ,  $i \in \{1,2,3\}$ , и аминокислотным составом каждого штамма. Для каждой целевой характеристики  $X \equiv X_i$ ,  $i \in \{1,2,3\}$ , задача такого типа решалась отдельно. При этом применялся один и тот же метод решения. Метод обнаружения закономерной связи между фенотипическими свойствами штамма и его аминокислотной структурой состоит в следующем. Для заданной характеристики  $X = \{x_m\}$ ,  $m = 1, \dots, M$ , выбирается два «эталонных» штамма: эталоном № 1 становится штамм с максимальным значением  $x_{\max}$ , эталоном № 2 – штамм с минимальным значением  $x_{\min}$ . Все остальные штаммы располагаются между ними в порядке убывания их значения от самого большого до самого малого. Будем считать расстоянием от  $m$ -го штамма до эталона № 1 величину  $r_{1X} = x_{\max} - x_m$ , а до эталона № 2 – величину  $r_{2X} = x_m - x_{\min}$ .

Каждый штамм описывается аминокислотами, стоящими в его столбце в таблице 3. По этим описаниям можно так же определить расстояния от  $m$ -го штамма до эталонов № 1 и № 2. Эти расстояния определяются по методу Хэмминга: если на  $n$ -й позиции у данных штаммов стоят одинаковые аминокислоты, то расстояние  $r(n) = 0$ , если стоят разные аминокислоты, то  $r(n) = 1$ . Зная  $r(n)$  для всех  $N$  признаков (позиций), можно найти расстояние  $r_{1S}$  между любым штаммом и эталонным штаммом № 1:

$$r_{1S} = \frac{1}{N} \sum_{n=1}^N r(n) \quad (1).$$

Аналогично находится и расстояние  $r_{2S}$  от любого штамма до эталона № 2. Минимальное значение  $r$  равно 0 (при полном совпадении состава аминокислот на всех  $N$  позициях), максимальное значение  $r$  равно 1 (при несовпадении аминокислот по всем  $N$  позициям). В общем случае, зная расстояния  $r_1$  и  $r_2$  от  $m$ -го объекта до двух эталонов, можно вычислить меру сходства этого объекта с эталоном № 1 в конкуренции с эталоном № 2. Для этого используется функция конкурентного сходства (FRIS-функция [3, 4, 5]) следующего вида:

$$F(m, N\#1 | N\#2) = \frac{r_2 - r_1}{r_2 + r_1} \quad (2).$$

Значение функции конкурентного сходства меняется в пределах от 1 до -1:  $F(m, N\#1 | N\#2) = 1$ , если  $m$ -й объект совпадает с эталоном № 1. Если  $m$ -й объект совпадает с конкурирующим эталоном № 2, то  $F(m, N\#1 | N\#2) = -1$ . Если расстояния от  $m$ -го

объекта до эталонов одинаковы, то  $F(m, N\#1 | N\#2) = F(m, N\#1 | N\#2) = 0$ .

По расстояниям  $r_{1X}$  и  $r_{2X}$  можно для всех  $M$  штаммов найти значения  $F_X(m, N\#1 | N\#2)$  сходства их с эталоном № 1 в конкуренции с эталоном № 2 по целевой характеристике  $X$ . Аналогично, по расстояниям  $r_{1S}$  и  $r_{2S}$  находятся значения функции  $F_S(m, N\#1 | N\#2)$  сходства штаммов с эталоном № 1 в конкуренции с эталоном № 2 по аминокислотному составу.

Если из исходного множества  $N$  позиций штамма выбирать разные подмножества из  $n$  позиций ( $n < N$ ), то значения расстояний  $r_{1S}$  и  $r_{2S}$  и, соответственно, функции  $F_S$  будут меняться в зависимости от  $n$  и от того, какие позиции входят в выбранное подмножество. С учетом сказанного, задача обнаружения закономерной связи между значением свойства  $X$  и аминокислотным составом штаммов состоит в следующем: требуется найти такое сочетание из  $n$  позиций ( $n < N$ ), при котором достигается максимум корреляции между двумя рядами значений функций конкурентного сходства: по целевой характеристике  $X (F_X)$  и по аминокислотному составу штаммов ( $F_S$ ).

Известно, что задача выбора наиболее информативного подмножества признаков является NP-трудной задачей. Ее точное решение можно получить только методом полного перебора всех возможных комбинаций признаков. Трудоемкость этого метода неприемлема для решения большинства реальных задач, и по этой причине используются эвристические алгоритмы направленного перебора. В данной работе для нахождения  $F_S$  использовались все  $N = 27$  признаков (позиций).

**Связь между аминокислотной последовательностью белков штаммов и их термостабильностью (Т50).** Значения свойства  $X \equiv X_1$  (термостабильность) были известны для всех 17 штаммов (табл. 2). В качестве эталонов выбраны штаммы № 3 с  $x_{\max} = 3,78$  и № 7 с  $x_{\min} = 0$ . Расстояния между штаммами по их аминокислотному составу вычислялись в метрике Хэмминга. Применение описанной методики к данным по целевому признаку и по структуре дало результаты, показанные в последних четырех строках таблицы 4. На полных данных зависимость между составом и свойством оказалась невысокой: корреляция  $|r| = 0,417$ . После цензурирования четырех штаммов с большими отклонениями от преобладающей закономерности (штаммов 14, 2, 13 и 5) корреляция поднялась до значения 0,855.

**Связь между аминокислотным составом штамма и его терморезистентностью (rct42).** В

Таблица 4

Сходство штаммов со штаммом № 3 в конкуренции со штаммом № 7 по свойству  $X = X_1 (F_X)$  и структуре ( $F_S$ )

Штамм	3	14	4	1	6	15	11	8	2	9	12	13	16	17	10	5	7
$X_1$	3,8	2,5	2,4	2,3	1,8	1,7	1,3	1,3	1	1	1	1	0,7	0,6	0,3	0,2	0
$F_S$	1,00	-1,00	1,00	0,84	0,84	-0,31	-1,00	-0,64	0,85	-0,64	-0,92	0,85	-0,92	-0,64	-0,92	0,93	-1,00
$F_X$	1,00	0,32	0,26	0,21	-0,05	-0,11	-0,32	-0,32	-0,47	-0,47	-0,47	-0,47	-0,63	-0,68	-0,84	-0,89	-1,00
Коэф. $ r  = 0,417$																	
Коэф. $ r  = 0,855$																	

этой задаче проблемы возникли уже на этапе выбора эталонов. Штамм № 1, который имеет максимальное значение свойства, по составу почти не отличается от штамма № 3 с минимальным значением данного свойства. Ясно, что такие эталоны не в состоянии растянуть штаммы по оси значений свойства  $X = X_2$ . В качестве компромисса эталонами были выбраны штаммы № 15:  $x_{\max} = 3,34$  и № 3:  $x_{\min} = 0,18$ . Сравнение значений функции конкурентного сходства  $F_x$  по свойству и значений функции конкурентного сходства  $F_s$  по структуре 16 штаммов показало наличие линейной зависимости между ними. Это позволяет заполнить пробел в значении свойства у штамма № 17. Поясним способ заполнения этого пробела. По аминокислотному составу находим Хэминговы расстояния  $r_{sm}$  от всех  $m$ -х штаммов до штамма № 17. Значением целевой функции свойства у штамма № 17 будем считать средневзвешенное значение этой функции у всех 16 штаммов:

$$x_{17} = \sum_{m=1}^{M-1} x_m \times V_m / \sum_{m=1}^{M-1} V_m \quad (3).$$

Здесь  $V_m = (1 - r_{sm})$  – вес значения свойства у  $m$ -го штамма. В нашем случае предсказанное значение пропущенного элемента оказалось равным  $x_{17} = 1,9$ . Результаты совместного анализа свойства  $X = X_2$  и аминокислотного состава штаммов представлены в таблице 5.

Как и ожидалось, результаты, полученные на всех  $M = 17$  штаммах, оказались невысокими: корреляция между составом  $F_x$  и свойством  $F_s$  была положительной, но небольшой:  $|r| = 0,258$ . Если из выборки исключить 5 штаммов (1, 5, 11, 14, 16) с наиболее сильными отклонениями от закономерности значений  $F_s$ , то корреляция поднимается до величины 0,784. Так что такая зависимость правильно отражает связь между способностью к репродукции при 42 °C и аминокислотным составом на 71 % штаммов.

**Связь между аминокислотным составом штамма и его нейроинвазивностью (NI).** В таблице 2 нейроинвазивность представлена характеристикой  $X = X_3$ . Чем меньше значение данной характеристики, тем выше вирулентность штамма. Эталонными назначены штаммы № 15 и № 5. Значения свойства  $X$  были известны только у 11 штаммов. Пробелы заполнялись по описанному выше методу. Эталонами выбраны штаммы № 5 с  $x_{\min} = 0,8$  и № 15 с  $x_{\max} = 3,0$ . Результаты совместного анализа данного свойства и аминокислотного состава штаммов представлены в таблице 6.

На всех  $M = 17$  штаммах после заполнения пробелов в целевой характеристике корреляция оказалась равной  $|r| = 0,620$ . После исключения двух штаммов с большими отклонениями от преобладающей закономерности (штаммов 4 и 1) корреляция поднялась до значения 0,875. Если вывод подтвердится на большем количестве штаммов, то будет можно предсказать инвазивность штамма на качественном уровне, проанализировав состав его аминокислот в  $N = 27$  указанных сечениях.

#### О зависимостях между целевыми фенотипическими признаками штаммов.

Было интересно исследовать зависимости между тремя изучаемыми целевыми свойствами штаммов. Если бы зависимости оказались сильными, то можно было бы предсказывать одни целевые характеристики по другим, не прибегая к анализу аминокислотного состава штаммов. Парные зависимости изучались на наиболее достоверных данных о целевых свойствах, т.е. на тех штаммах, у которых не наблюдалось больших расхождений между значениями функций  $F_x$  и  $F_s$ . По этой причине для каждой пары свойств выбирались свои подмножества штаммов. Коэффициенты корреляции между соответствующими парами целевых свойств оказались такими:

$$k(X_1, X_2) = -0.609; \quad k(X_1, X_3) = -0.201; \\ k(X_2, X_3) = 0.466.$$

**Таблица 5**  
Сходство штаммов со штаммом № 15 в конкуренции со штаммом № 3 по свойству  $X = X_2 (F_x)$  и структуре ( $F_s$ )

Штамм	1	17	15	8	5	10	12	9	2	6	1,3	7	11	14	16	4	3
$X_2$	4,30	1,90	3,30	2,50	2,40	2,30	2,20	2,10	1,80	1,80	1,80	1,70	0,80	0,80	0,70	0,50	0,20
$F_s$	<b>-0,82</b>	0,44	1,00	0,44	<b>-0,90</b>	0,45	0,39	0,44	-0,81	-0,79	-0,81	0,43	<b>0,43</b>	<b>0,43</b>	<b>0,45</b>	-1,00	-1,00
$F_x$	0,61	0,10	1,00	0,48	0,42	0,35	0,29	0,23	0,03	0,03	0,03	-0,03	-0,61	-0,61	-0,68	-0,81	-1,00
Коеф. $ r _{17} = 0,258$																	
Коеф. $ r _{12} = 0,784$																	

**Таблица 6**  
Сходство штаммов со штаммом № 5 в конкуренции со штаммом № 15 по свойству  $X = X_3 (F_x)$  и структуре ( $F_s$ )

Штамм	5	6	2	13	3	14	17	8	9	12	16	4	7	10	11	1	15
$X_3$	<b>0,8</b>	<b>1,1</b>	<b>1,5</b>	<b>1,5</b>	<b>1,6</b>	1,86	1,86	2,19	2,19	2,19	2,19	<b>2,2</b>	<b>2,2</b>	<b>2,3</b>	<b>2,6</b>	<b>2,8</b>	<b>3</b>
$F_s$	1,00	0,70	0,90	0,90	0,90	-0,44	-0,45	-0,45	-0,45	-0,41	-0,47	<b>0,90</b>	-0,44	-0,47	-0,44	<b>0,74</b>	-1,00
$F_x$	1,00	0,73	0,36	0,36	0,27	0,04	0,04	-0,26	-0,26	-0,26	-0,26	-0,27	-0,27	-0,36	-0,64	-0,82	-1,00
Коеф. $ r _{17} = 0,620$																	
Коеф. $ r _{15} = 0,875$																	



Можно сказать, что термостабильность и терморезистентность (способность к репродукции при 42 °С) штаммов связаны между собой довольно сильной обратной зависимостью. Но надежда на то, что по этим свойствам можно с высокой точностью предсказывать наиболее важное свойство  $X_3$  – нейроинвазивность, не прибегая к анализу аминокислотного состава штаммов, не оправдалась.

## ЗАКЛЮЧЕНИЕ

На имеющихся данных обнаружены сечения в штаммах ВКЭ, по мутации аминокислотного состава которых можно с той или иной точностью предсказывать фенотипические признаки штаммов. Показана возможность предсказания пропущенных элементов в данных, как в аминокислотном составе, так и в целевых свойствах. Оценены зависимости между парами фенотипических признаков штаммов. Гипотеза о возможности предсказывать нейроинвазивность по термостабильности и терморезистентности не получила подтверждения. Подчеркнем, что эти результаты получены на очень малой выборке, пораженной помехами и пробелами. После увеличения выборки обнаруженные закономерности могут измениться: какие-то выводы подтвердятся, другие окажутся опровергнутыми. Цель данного этапа исследований состояла не в том, чтобы гарантированно точно обнаружить реальные закономерности, а в том, чтобы показать применимость предлагаемых математических методов для решения задач такого рода. Можно считать, что эта цель достигнута.

*Исследование выполнено при финансовой поддержке гранта в рамках Программы фундаментальных исследований Президиума РАН № 1 по стратегическим направлениям развития науки на 2014 г. «Фундаментальные проблемы математического моделирования».*

## ЛИТЕРАТУРА REFERENCES

1. Верховина М.М., Злобин В.И., Козлова И.В., Беликов С.И. и др. Молекулярно-эпидемиологическая характеристика региональной популяции вируса клещевого энцефалита Восточной Сибири // Бюлл. ВСНЦ СО РАМН. – 2002. – Т. 2, № 4. – С. 46–49.
2. Verkhovina M.M., Zlobin V.I., Kozlova I.V., Belikov S.I. et al. Molecular-epidemiological characteristics of regional population of tick-borne encephalitis in Eastern Siberia // Bujl. VSNC SO RAMN. – 2002. – Vol. 2, N 4. – P. 46–49. (in Russian)
3. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999. – С. 270.
4. Zaboruyko N.G. Applied methods of data and knowledge analysis. – Novosibirsk: IM SO RAN, 1999. – P. 270. (in Russian)
5. Загоруйко Н.Г., Кутненко О.А. Алгоритм GRAD для выбора признаков // Труды Межд. конференции «Применение многомерного статистического анализа в экономике и оценке качества». – М.: Изд-во МЭСИ, 2006. – С. 81–89.
6. Zaboruyko N.G., Kutnenko O.A. GRAD algorithm for choosing properties // Trudi Mezhd. konferentsii "Primenenie mnogomernogo statisticheskogo analiza v ekonomike i otsenke kachestva". – Moscow: Izd-vo MESI, 2006. – P. 81–89. (in Russian)
7. Загоруйко Н.Г. Интеллектуальный анализ данных, основанный на функции конкурентного сходства // Автометрия. – 2008. – Т. 44, № 3. – С. 31–40.
8. Zaboruyko N.G. Intellectual data analysis, based in function of competitive similarity // Avtometriya. – 2008. – Vol. 44, N 3. – P. 31–40. (in Russian)
9. Загоруйко Н.Г. Когнитивный анализ данных. – Новосибирск: Академическое издательство ГЕО, 2013. – 186 с.
10. Zaboruyko N.G. Cognitive data analysis. – Novosibirsk: Akademicheskoye izdatelstvo GEO, 2013. – 186 p. (in Russian)
11. Злобин В.И., Мамаев Л.В., Джигоев Ю.П., Козлова И.В. Генетические типы вируса клещевого энцефалита // Журнал инфекционной патологии. – 1996. – Т. 3, № 4. – С. 13–17.
12. Zlobin V.I., Mamayev L.V., Dzioyev Yu.P., Kozlova I.V. Genetic types of tick-borne encephalitis // Zhurnal infektsionnoy patologii. – 1996. – Vol. 3, N 4. – P. 13–17. (in Russian)
13. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 295 с.
14. Urbakh V.Yu. Statistical analysis in biological and medical study. – Moscow: Meditsina, 1975. – 295 p. (in Russian)
15. Brault A.C., Huang C.Y., Langevin S.A., Kinney R.M. et al. A single positively selected West Nile viral mutation confers increased virogenesis in American crows // Nat. Genet. – 2007. – Vol. 39 (9). – P. 1162–1166.
16. Demina T.V., Dzhioev Y.P., Verkhozina M.M., Kozlova I.V. et al. Genotyping and characterization of geographical distribution of tick-borne encephalitis virus variants a set of molecular probes // Journal of Medical Virology. – 2010. – Vol. 82, N 6. – P. 965–976.
17. Fauquet C.M., Mayo M.A., Maniloff J., Desselberger U. et al. Virus Taxonomy. – Eighth Report of the International Committee on Taxonomy of Viruses. – San Diego, London: Elsevier Academic Press, 2005. – 1162 p.
18. Maier C.C., Delagrave S., Zhang Z.X., Brown N. et al. A single M protein mutation affects the acid inactivation threshold and growth kinetics of a chimeric flavivirus // Virology. – 2007 Jun. – Vol. 362 (2). – P. 468–474.
19. Mandl C.W., Heinz F.X., Stöckl E., Kunz C. Genome sequence of tick-borne encephalitis virus (Western subtype) and comparative analysis of nonstructural proteins with other flaviviruses // Virology. – 1989. – Vol. 173 (1). – P. 291–301.
20. Mandl C.W., Holzmann H., Meixner T., Rauscher S. et al. Spontaneous and engineered deletions in the 3' noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus // J. Virol. – 1998. – Vol. 72 (3). – P. 2132–2140.
21. Mandl C.W. Steps of the tick-borne encephalitis virus replication cycle that affect neuropathogenesis // Virus Res. – 2005. – Vol. 111 (2). – P. 161–174.

15. McMinn P.C. The molecular basis of virulence of the encephalitogenic flaviviruses // J. Gen. Virol. – 1997. – Vol. 78 (Pt. 11). – P. 2711–2722.

16. Wicker J.A., Whiteman M.C., Beasley D.W., Davis C.T. et al. A single amino acid substitution in the central portion of the West Nile virus NS4B protein confers a

highly attenuated phenotype in mice // Virology. – 2006. – Vol. 349 (2). – P. 245–253.

17. Zagoruiko N.G., Borisova I.A., Dyubanov V.V., Kutnenko O.A. Methods of Recognition Based on the Function of Rival Similarity // Pattern Recognition and Image Analysis. – 2008. – Vol. 18, N 1. – P. 1–6.

#### Сведения об авторах

**Загоруйко Николай Григорьевич** – доктор технических наук, профессор, зав. отделом информатики Института математики им. С.Л. Соболева (630090, г. Новосибирск, просп. им. академика Коптюга, 4; тел.: 8 (3832) 33-32-91; e-mail: zag@math.nsc.ru)

**Джигоев Юрий Павлович** – кандидат биологических наук, ведущий научный сотрудник, руководитель лаборатории молекулярной вирусологии и биотехнологии НИИ биомедицинских технологий Иркутского государственного медицинского университета, старший научный сотрудник лаборатории молекулярной эпидемиологии и генетической диагностики Научного центра проблем здоровья семьи и репродукции человека (e-mail: alanir07@mail.ru)

**Козлова Ирина Валерьевна** – доктор медицинских наук, руководитель лаборатории молекулярной эпидемиологии и генетической диагностики Научного центра проблем здоровья семьи и репродукции человека (e-mail: diwerhoz@rambler.ru)

**Ткачев Сергей Евгеньевич** – научный сотрудник лаборатории микробиологии Института химической биологии и фундаментальной медицины СО РАН (e-mail: sergey.e.tkachev@mail.ru)

**Букин Юрий Сергеевич** – кандидат биологических наук, старший научный сотрудник лаборатории геносистематики Лимнологического института СО РАН (e-mail: bukinyura@mail.ru)

**Парамонов Алексей Игоревич** – младший научный сотрудник лаборатории молекулярной эпидемиологии и генетической диагностики Научного центра проблем здоровья семьи и репродукции человека (e-mail: paramonov\_a.i@mail.ru)

**Борисенко Андрей Юрьевич** – аспирант лаборатории молекулярной вирусологии и биотехнологии НИИ биомедицинских технологий Иркутского государственного медицинского университета (e-mail: 89500720225@mail.ru)

**Борисова Ирина Артемовна** – кандидат технических наук, старший научный сотрудник Института математики им. С.Л. Соболева (e-mail: biamia@mail.ru)

**Злобин Владимир Игоревич** – академик РАН, директор НИИ биомедицинских технологий, заведующий кафедрой микробиологии, вирусологии и иммунологии Иркутского государственного медицинского университета (e-mail: vizlobin@mail.ru)

**Кутненко Ольга Андреевна** – кандидат технических наук, старший научный сотрудник Института математики им. С.Л. Соболева (e-mail: olga@math.nsc.ru)

#### Information about the authors

**Zagoruyko Nikolay Grigoryevich** – Doctor of Technical Sciences, Professor, Head of the Department of Computer Science of Institute of Mathematics named after S.L. Sobolev (630090, Novosibirsk, prospekt akadem. Koptuyga, 4; tel.: +7 (3832) 33-32-91; e-mail: zag@math.nsc.ru)

**Dzhigoev Yuri Pavlovich** – Candidate of Biological Sciences, Senior Research Officer, Head of the Laboratory of Molecular Virology and Biotechnology of Research Institute of Biomedical Technologies of Irkutsk State Medical University, Senior Research Officer at the Laboratory of Molecular Epidemiology and Genetic diagnostics of Scientific Center of Family Health Problems and Human Reproduction (e-mail: alanir07@mail.ru)

**Kozlova Irina Valeryevna** – Doctor of Medical Sciences, Head of the Laboratory of Molecular Epidemiology and Genetic diagnostics of Scientific Center for Family Health Problems and Human Reproduction (e-mail: diwerhoz@rambler.ru)

**Tkachev Sergey Yevgenyevich** – Research Officer of the Laboratory of Microbiology of Institute of Chemical Biology and Fundamental Medicine SB RAS (e-mail: sergey.e.tkachev@mail.ru)

**Bukin Yuri Sergeyevich** – Candidate of Biological Science, Senior Research Officer of the Laboratory of Molecular Genosystematics of Limnological Institute RAS (e-mail: bukinyura@mail.ru)

**Paramonov Alexey Igorevich** – Junior Research Officer of the Laboratory of Molecular Epidemiology and Genetic diagnostics of Scientific Center of Family Health Problems and Human Reproduction (e-mail: paramonov\_a.i@mail.ru)

**Borisenko Andrew Yuryevich** – Postgraduate of the Laboratory of Molecular Virology and Biotechnology of Research Institute of Biomedical Technology of Irkutsk State Medical University (e-mail: 89500720225@mail.ru)

**Borisova Irina Artemovna** – Doctor of Technical Sciences, Senior Research Officer of S.L. Sobolev Institute of Mathematics SB RAS (e-mail: biamia@mail.ru)

**Zlobin Vladimir Igorevich** – Academician of RAS, Director of the Institute of Biomedical Technology, Head of the Department of Microbiology, Virology and Immunology of Irkutsk State Medical University (e-mail: vizlobin@mail.ru)

**Kutnenko Olga Andreevna** – Doctor of Technical Sciences, Senior Research Officer of S.L. Sobolev Institute of Mathematics SB RAS (e-mail: olga@math.nsc.ru)