

## **Сравнение mcl-метода с ранее предложенными методами вычисления эволюционных дистанций на примере анализа мрнк, кодирующих ряд ферментов дыхательной цепи человека, трихинеллы и цианорабдитис**

Проведено сравнение MCL-метода с ранее предложенными методами вычисления эволюционных дистанций (Тамуры-Нея, Тадзимы-Нея, Тамуры) на примере анализа мРНК, кодирующих ряд ферментов дыхательной цепи человека, трихинеллы и цианорабдитис. Значения эволюционных дистанций между изучаемыми мРНК, полученные методом MCL, статистически неразличимы от таковых, полученных методами Тамуры, Тадзимы-Нея и Тамуры-Нея. MCL-метод наиболее сходен с методом Тамуры-Нея, что подтверждает высокий коэффициент корреляции между значениями эволюционных дистанций, полученными данными методами.

Ключевые слова: эволюционная дистанция, MCL-метод, биоинформатика.

Трихинеллез – тяжелое паразитарное заболевание человека и животных, возбудителем которого является трихинелла (*Trichinella spiralis*, Owen, 1835). Коэволюция трихинеллы и ее хозяев хорошо изучена на клеточном, тканевом, организменном и видовом уровнях организации живого. Коэволюционным механизмам в системе “паразит–хозяин”, формирующейся при трихинеллезе на молекулярно-генетическом уровне, посвящены лишь единичные работы. Так, в результате исследований по изучению эволюционных изменений мРНК, кодирующих 12 ферментов дыхательной цепи, доказано сходство стратегий кодирования митохондриальных белков в системе “паразит-хозяин”, формирующейся при трихинеллезе [3, 4].

Потенциальным маркером коэволюции паразитов и их хозяев на молекулярно-генетическом уровне является эволюционная дистанция – среднее число аминокислотных или нуклеотидных замен, приходящихся на пару гомологичных сайтов двух выровненных последовательностей белков или нуклеиновых кислот. К настоящему времени предложено множество методов определения эволюционных дистанций, каждый из которых имеет какие-либо недостатки, что отражается на точности получаемых результатов.

В 2004 г. К. Тамура и соавт. [8] предложили новый метод определения эволюционных дистанций между последовательностями нуклеиновых кислот – MCL-метод (maximum composite likelihood method). Согласно данным разработчиков этот метод является более точным (по сравнению с ранее предложенными методами), так как он основан на одновременном вычислении эволюционных дистанций во всем массиве выровненных последовательностей. При этом точность значений, полученных MCL-методом, возрастает при увеличении количества анализируемых последовательностей. При расчете парных эволюционных дистанций в большом массиве выровненных последовательностей значения стандартных ошибок при использовании MCL-метода значительно меньше, чем при использовании других методов. Это

обусловлено увеличением числа последовательностей и одновременным вычислением для всех последовательностей ряда параметров, таких как соотношение транзиций и трансверсий. Эволюционная дистанция  $d_{ij}$  между последовательностями  $i$  и  $j$  вычисляется по формуле:

$$d_{ij} = 4(g_A g_T k_1 + g_C g_G k_2 + g_R g_Y) \times b_{ij},$$

где  $g_A, g_T, g_C, g_G$  – частота нуклеотидов А, Т, Ц, Г,  $g_R = g_A + g_T, g_Y = g_C + g_G$ . Начальные значения  $k_1, k_2$  и  $b_{ij}$  равны:  $k_1 = (\hat{P}_{1ij}/g_A g_T) / (\hat{Q}_{ij}/g_R g_Y), k_2 = (\hat{P}_{2ij}/g_T g_C) / (\hat{Q}_{ij}/g_R g_Y), b_{ij} = \hat{Q}_{ij} / (4g_R g_Y)$ , где  $\hat{P}_{1ij}, \hat{P}_{2ij}$  и  $\hat{Q}_{ij}$  – наблюдаемые частоты транзиций А↔Т, транзиций Ц↔Г и трансверсий, соответственно. Затем вычисляются средние значения  $k_1$  и  $k_2$ , что необходимо для получения скорректированных значений  $b_{ij}$  путем максимизирования  $L_{ij}$  по формуле:

$$L_{ij} = \hat{P}_{1ij} \ln(P_{1ij}) + \hat{P}_{2ij} \ln(P_{2ij}) + \hat{Q}_{ij} \ln(Q_{ij}) + (1 - \hat{P}_{1ij} - \hat{P}_{2ij} - \hat{Q}_{ij}) \times \ln(1 - P_{1ij} - P_{2ij} - Q_{ij}),$$

где  $P_{1ij}$  – теоретическая частота транзиций А↔Т,  $P_{2ij}$  – теоретическая частота транзиций Ц↔Г,  $Q_{ij}$  – теоретическая частота трансверсий. Значения данных величин равны:

$$P_{1ij} = \frac{2g_A \times g_T}{g_R} \{g_R - \exp[-2(g_R k_1 + g_Y) b_{ij}] + g_Y \exp(-2b_{ij})\},$$

$$P_{2ij} = \frac{2g_T \times g_C}{g_Y} \{g_Y - \exp[-2(g_Y k_2 + g_R) b_{ij}] + g_R \exp(-2b_{ij})\},$$

$$Q_{ij} = 2g_R g_Y [1 - \exp(-2b_{ij})].$$

Эти вычисления позволяют получить скорректированные значения  $k_1$  и  $k_2$  путем максимизирования  $SL$  по формуле:  $SL = \sum_i \sum_{j \neq i} L_{ij}$ . Два последних этапа повторяются до получения постоянных значений  $k_1, k_2$  и  $b_{ij}$ . Вариансы значений MCL-дистанций предпочтительно вычисляются методом bootstrap [2, 8].

Цель исследования: MCL-метод с ранее предложенными методами вычисления эволюционных дистанций при анализе мРНК, кодирующих ряд ферментов дыхательной цепи человека, трихинеллы и цианорабдитис. сравнить Материалы и методы. NCBI National Center for Biotechnology Information; [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) последовательности мРНК, кодирующих 12 митохондриальных белков – ферментов дыхательной цепи (субъединицы 1-6 и 4L НАДН-дегидрогеназы (НАДН-ДГ), цитохром b, субъединицы 1-3 цитохром-с-оксидазы (ЦО), субъединица 6 АТФ-синтетазы) человека (*Homo sapiens*), трихинеллы (*Trichinella spiralis*) и свободноживущего круглого червя (*Caenorhabditis elegans*). Парное выравнивание последовательностей проведено при помощи программы ClustalW Protein [9]. Картина замен нуклеотидов определена методом Монте-Карло на основании значений индекса несоответствия [1, 2]. Эволюционные дистанции вычислены по методам MCL [1, 8], Тамуры (Т) [1, 6], Тамуры-Нея (ТН) [1, 7], Тадзимы-Нея (ТЖН) [1, 5]. Полученные результаты обработаны методами описательной статистики, достоверность различий определена по критерию Стьюдента. Изучены взятые с сервера ( Результаты и обсуждение. Установлено, что картина замен нуклеотидов во всех проведенных попарных сравнениях мРНК является гетерогенной. Полученные значения эволюционных дистанций между изучаемыми мРНК человека и

трихинеллы, а также человека и цианорабдитис представлены в табл. 1 и 2 соответственно.

Таблица 1

Эволюционные дистанции между мРНК, кодирующими ферменты дыхательной цепи человека и трихинеллы

мРНК / метод	MCL	TJN	T	TN
НАДН-ДГ1	0,784±0,045	0,750±0,044	0,740±0,043	0,753±0,044
НАДН-ДГ2	2,103±0,518	1,433±0,093	1,552 ±1,050	1,574±0,664
НАДН-ДГ3	1,052±0,146	1,025±0,110	1,021±0,113	1,033±0,126
НАДН-ДГ4	1,254±0,198	1,366±0,067	1,792±0,307	1,407±0,102
НАДН-ДГ4L	1,029±0,908	1,583±0,242	2,492±1,017	2,021±0,832
НАДН-ДГ5	2,131±0,380	1,443±0,068	2,095±0,798	1,488±0,131
НАДН-ДГ6	2,951±1,274	1,725±0,217	1,760 ±1,075	1,858±0,870
Цитохром b	0,849±0,051	0,828±0,042	0,817±0,041	0,831±0,043
ЦО1	0,557±0,026	0,554±0,025	0,553±0,025	0,557±0,026
ЦО2	0,790±0,558	0,754±0,050	0,751±0,052	0,766±0,057
ЦО3	0,901±0,063	0,839±0,053	0,834±0,054	0,850±0,058
АТФаза б	1,403±0,262	1,501±0,154	1,409±0,156	1,541±0,531

Установлено, что полученные MCL-методом значения эволюционных дистанций статистически неразличимы от дистанций Тамуры-Нея, Тамуры и Тадзимы-Нея (за исключением сравнения мРНК, кодирующих субъединицу 4 НАДН-ДГ человека и трихинеллы,  $p < 0,05$ ). Значения, полученные при использовании MCL-метода, достоверно коррелируют ( $p < 0,001$ ) со значениями, полученными при использовании методов Тадзимы-Нея ( $r = 0,79 \pm 0,130$ ), Тамуры ( $r = 0,66 \pm 0,169$ ) и Тамуры-Нея ( $r = 0,88 \pm 0,104$ ). Более высокий коэффициент корреляции эволюционных дистанций, полученных MCL-методом и методом Тамуры-Нея, свидетельствует об их наибольшем сходстве.

Таблица 2

Эволюционные дистанции между мРНК, кодирующими ферменты дыхательной цепи человека и цианорабдитис

мРНК / метод	MCL	TjN	T	TN
НАДН-ДГ1	0,972±0,103	0,990±0,059	1,039±0,068	0,995±0,062
НАДН-ДГ2	1,163±0,365	1,310±0,091	1,389±0,116	1,312±0,122
НАДН-ДГ3	1,094±0,196	1,125±0,117	1,215±0,170	1,146±0,175
НАДН-ДГ4	0,903±0,048	1,021±0,050	1,061±0,057	1,023±0,052
НАДН-ДГ4L	1,447±1,049	1,079±0,127	1,167±0,223	1,118±0,251
НАДН-ДГ5	0,894±0,240	1,197±0,056	1,274±0,069	1,209±0,059
НАДН-ДГ6	1,674±1,423	1,437±0,176	1,305±0,136	1,318 ±0,562
Цитохром <i>b</i>	0,872±0,044	0,857±0,042	0,900±0,049	0,867±0,046
ЦО1	0,637±0,028	0,626±0,027	0,641±0,029	0,634±0,028
ЦО2	0,993±0,075	0,924±0,065	0,955±0,067	0,943±0,070
ЦО3	0,814±0,052	0,811±0,049	0,838±0,056	0,814±0,051
АТФаза 6	1,259±0,222	1,182±0,096	1,257±0,123	1,192±0,109

При проведении регрессионного анализа значений, полученных данными методами (рис.), получено уравнение  $y=0,85x$  (при условии, что  $y=0$  при  $x=0$ ), где  $y$  – значение дистанции Тамуры-Нея, а  $x$  – значение MCL-дистанции.

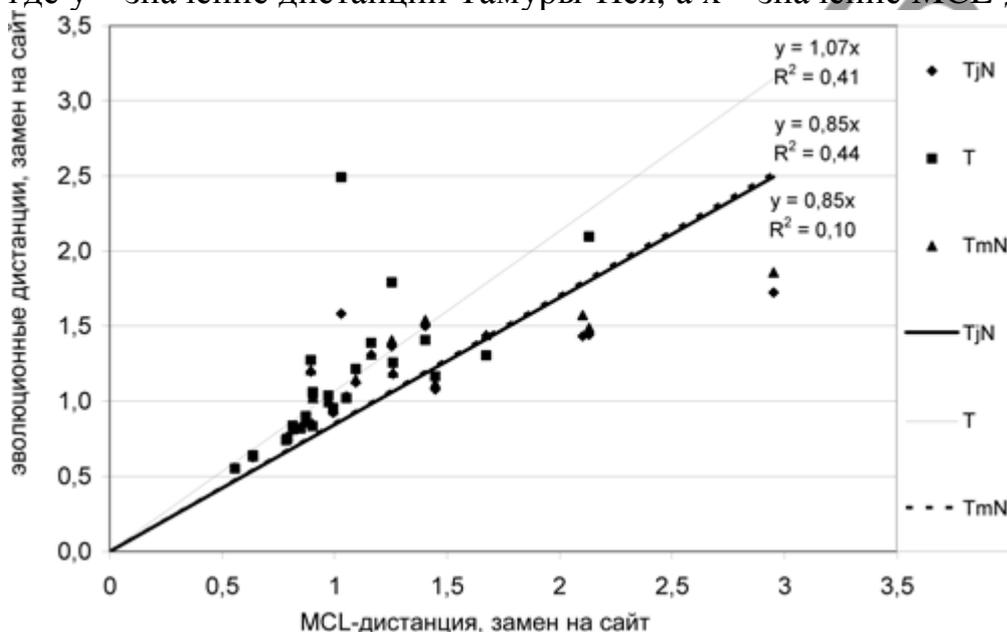


Рисунок. Взаимосвязь значений MCL-дистанции с эволюционными дистанциями Тамуры (T), Тамуры-Нея (TmN) и Тадзимы-Нея (TjN)

При сравнении средних значений (AVR) между изучаемыми мРНК, полученных методами Тадзимы-Нея, Тамуры и Тамуры-Нея (табл. 3) с MCL-дистанциями, установлено, что они также статистически неразличимы в большинстве случаев ( $95,8 \pm 4,18$ ,  $p < 0,001$ ).

Таблица 3

Средние эволюционные дистанции между мРНК, кодирующими ферменты дыхательной цепи человека (H.s.), трихинеллы (T.s.) и цианорабдитис (C.e.)

мРНК / показатель	AVR H.s.-T.s	AVR H.s-C.e.	p AVR H.s.-T.s. / AVR H.s.-C.e.
НАДН-ДГ1	0,748±0,005	1,008±0,019	<0,001
НАДН-ДГ2	1,504±0,070	1,337±0,032	>0,05
НАДН-ДГ3	1,026±0,004	1,162±0,033	<0,001
НАДН-ДГ4	1,522±0,165	1,035±0,016	<0,05
НАДН-ДГ4L	2,038±0,455	1,121±0,031	>0,05
НАДН-ДГ5	1,675±0,258	1,227±0,029	>0,05
НАДН-ДГ6	1,792±0,067	1,394±0,055	<0,01
Цитохром b	0,825±0,005	0,875±0,016	<0,05
ЦО1	0,555±0,002	0,634±0,005	<0,001
ЦО2	0,757±0,006	0,941±0,011	<0,001
ЦО3	0,841±0,006	0,821±0,011	>0,05
АТФаза 6	1,484±0,048	1,210±0,029	<0,001

При рассмотрении эволюционных дистанций как возможного критерия коэволюции обнаружено, что их значения достоверно ( $p < 0,05$ ) отличаются между 8 изученными мРНК человека и трихинеллы от таковых человека и цианорабдитис. Для большей части из них ( $62,5 \pm 18,30\%$ ) свойственно преобладание эволюционных дистанций между мРНК, кодирующими ферменты дыхательной цепи человека и трихинеллы (субъединицы 1 и 3 НАДН-ДГ, цитохром b, субъединицы 1 и 2 цитохром-с-оксидазы).

С нашей точки зрения, наличие такого преобладания лишь для части изученных мРНК можно объяснить функциональной важностью кодируемых ими белков. Так, например, первая субъединица НАДН-дегидрогеназы участвует в переносе электронов и взаимодействует с ингибиторами первого комплекса дыхательной цепи.

Эволюционная дистанция между нуклеотидными последовательностями состоит из двух вносящих в нее неравный вклад составляющих (синонимичной и несинонимичной). По этой причине перспективно отдельное определение синонимичной и несинонимичной дистанций в качестве более чувствительных критериев коэволюции на молекулярно-генетическом уровне в системе “паразит-хозяин”, что и станет объектом наших дальнейших исследований.

Выводы:

1. Значения эволюционных дистанций между изучаемыми мРНК, полученные методом MCL, статистически неразличимы от таковых, полученных методами Тамуры, Тадзимы-Нея и Тамуры-Нея.
2. MCL-метод наиболее сходен с методом Тамуры-Нея, что подтверждается высоким коэффициентом корреляции между значениями эволюционных дистанций, полученными данными методами.
3. Эволюционные дистанции между мРНК, кодирующими субъединицы 1 и 3 НАДН-дегидрогеназы, цитохрома b, субъединицы 1 и 2 цитохром-с-оксидазы человека и трихинеллы, достоверно меньше таковых между мРНК человека и цианорабдитис.

## Литература

1. Барковский, Е. В., Бутвиловский, А. В., Бутвиловский, В. Э., Давыдов, В. В., Хрусталева, В. В., Козюлевич, С. Р. Методы молекулярной эволюции и филогенетики: учеб.-метод. пособие. Минск: БГМУ, 2005. 63 с.
2. Бутвиловский, А. В., Барковский, Е. В., Бутвиловский, В. Э. Базисные методы молекулярной эволюции: учеб.-метод. пособие. Минск: БГМУ, 2006. 36 с.
3. Бутвиловский, В. Э., Барковский, Е. В., Бутвиловский, А. В., Линник, Ю. И. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 2. Картина использования синонимичных кодонов. Содержание аминокислотных групп GARP и FYMINK // Медицинский журнал. Минск, 2007. № 4. С. 39–42.
4. Бутвиловский, В. Э., Линник, Ю. И., Бутвиловский, А. В., Барковский, Е. В. О сходстве стратегий кодирования митохондриальных белков человека и трихинеллы. Часть 1. ГЦ-насыщенность, доля ГЦЗ-кодонов и частота использования претерминальных кодонов // Медицинский журнал. Минск, 2007. № 3. С. 39–41.
5. Tajima, F., Nei, M. Estimation of evolutionary distance between nucleotide sequences // *Mol. Biol. Evol.* 1984. Vol. 1. P. 269–285.
6. Tamura, K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C – content biases // *Mol. Biol. Evol.* 1992. Vol. 9. P. 678–687.
7. Tamura, K., Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in human and chimpanzees // *Mol. Biol. Evol.* 1993. Vol. 10. P. 512–526.
8. Tamura, K., Nei, M., Kumar, S. Prospects for inferring very large phylogenies by using the neighbor-joining method // *PNAS (USA)*. 2004. Vol. 101. P. 11030–11035.
9. Tompson, J.D, Higgins, D.G., Gibson, T.J. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions – specific gap penalties and weight matrix choice // *Nucl. Acids Res.* 1994. Vol. 22. P. 4673–4680.