

*В.Э. Бутвиловский,  
В.В. Давыдов,  
А.В. Бутвиловский*

**Определение вида картины замен в нуклеотидных и аминокислотных последовательностях. Сообщение 3. Вид картины замен в зависимости от положения нуклеотида в кодоне в последовательностях мРНК, кодирующих м-изоферменты креатинкиназ хордовых**

*Белорусский государственный медицинский университет*

Статья посвящена определению вида картины замен в зависимости от положения нуклеотида в кодоне в последовательностях мРНК, кодирующих М-изоферменты креатинкиназ хордовых. Показано, что гомогенная картина замен преобладает во втором и первом положениях кодона, а гетерогенная – в третьем. Учет вида картины замен позволяет получить более точные значения эволюционных дистанций для последующего вычисления на их основе скоростей молекулярной эволюции и построения дендрограмм. Ключевые слова: креатинкиназа, М-изофермент, положение кодона, композиционная дистанция, картина замен, гомогенность, гетерогенность, индекс несоответствия.

Цель исследования: установить вид картины замен в зависимости от положения нуклеотида в кодонах мРНК, кодирующих М-изоферменты креатинкиназ хордовых животных.

Материал и методы

Проанализированы нуклеотидные последовательности мРНК, кодирующие М-изоферменты креатинкиназ хордовых животных: оболочника (*Ciona intestinalis*, С.и., [5]), ланцетника (*Branchiostoma floridae*, В.ф., [5]), рыб (*Ictalurus punctatus*, I.п., [7]), пресмыкающихся (*Zaocys dhumnades*, Z.д., [18]), птиц (*Gallus gallus*, G.г., [11]), грызунов (*Rattus norvegicus*, R.н., [3], *Mus musculus*, М.м., [12]), парнокопытных (*Bos taurus*, В.т., NCBI: AAD30974.1) и приматов (*Homo sapiens*, Н.с., [17]). Выравнивание последовательностей произведено с помощью программ Clustal W DNA и Clustal W Protein [16].

Для указанных последовательностей вычислены композиционные дистанции и индекс несоответствия. Вероятность (Р) отклонения нулевой гипотезы о гомогенной картине замен нуклеотидов на 5%-ном уровне определена методом Монте-Карло при 1000 повторов [9, 10]. Если величина Р больше 0,05, то картина замен считалась гомогенной, а если величина Р меньше 0,05 – гетерогенной. В случае гомогенной картины замещений эволюционные дистанции (ЭД) рассчитаны методами Джукса-Кантора [6], Кимура [8], Тадзима-Ней [13], Тамура [14], Тамура-Ней [15]; в случае гетерогенной – методами Тадзима-Ней [13], Тамура [14], Тамура-Ней [15]. Полученные результаты обработаны методами описательной статистики, а достоверность различий сравниваемых величин определена по критерию t.

Результаты и обсуждение

Композиционные дистанции, значения индекса несоответствия и вероятности отклонения нулевой гипотезы о гомогенной картине замен для первого, второго и третьего положений кодона в изученных последовательностях представлены соответственно в табл. 1, 2 и 3.

Таблица 1

Значения композиционной дистанции, индекса несоответствия и вероятности отклонения нулевой гипотезы о гомогенной картине замен для первого положения кодона в последовательностях мРНК, кодирующих М-изоферменты креатинкиназ хордовых

	H.s.	R.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,0241 (0,0000)	0,0615 (0,0294)	0,0401 (0,0053)	0,0615 (0,0000)	0,0668 (0,0000)	0,2661 (0,1578)	0,4947 (0,2674)	0,2886 (0,0321)
B.t.	1,00		0,0348 (0,0053)	0,0267 (0,0000)	0,0107 (0,0000)	0,1471 (0,0642)	0,3630 (0,2647)	0,5775 (0,3556)	0,4572 (0,2059)
M.m.	0,14	0,38		0,0027 (0,0000)	0,0241 (0,0000)	0,1310 (0,0588)	0,5027 (0,3650)	0,8690 (0,6471)	0,6016 (0,3503)
R.n.	0,35	1,00	1,00		0,0267 (0,0000)	0,1013 (0,0348)	0,4572 (0,3396)	0,7968 (0,5749)	0,5267 (0,2754)
G.g.	1,00	1,00	1,00	1,00		0,1952 (0,0909)	0,4786 (0,3690)	0,7166 (0,4679)	0,6016 (0,3503)
Z.d.	1,00	0,14	0,13	0,22	0,14		0,2219 (0,0882)	0,6016 (0,4358)	0,2594 (0,0000)
I.p.	0,09	0,03	0,01	0,02	0,00	0,18		0,3717 (0,1417)	0,1952 (0,0000)
B.f.	0,07	0,06	0,01	0,01	0,03	0,04	0,20		0,2433 (0,0000)
C.i.	0,35	0,15	0,06	0,10	0,06	1,00	1,00	1,00	

Примечание. В скобках указаны значения индекса несоответствия. Жирным шрифтом выделена вероятность отклонения нулевой гипотезы о гомогенной картине замен. Ячейки со значениями  $P > 0,05$  выделены серым цветом.

Таблица 2 <11>Значения композиционной дистанции, индекса несоответствия и вероятности отклонения нулевой гипотезы о гомогенной картине замен для второго положения кодона в последовательностях мРНК, кодирующих М-изоферменты креатинкиназ хордовых

	H.s.	B.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,0027 (0,0000)	0,0027 (0,0000)	0,0027 (0,0000)	0,0080 (0,0000)	0,0642 (0,0027)	0,0321 (0,0000)	0,2727 (0,1086)	0,1471 (0,0000)
B.t.	1,00		0,0027 (0,0000)	0,0000 (0,0000)	0,0053 (0,0000)	0,0668 (0,0134)	0,0187 (0,0000)	0,2326 (0,0642)	0,1203 (0,0000)
M.m.	1,00	1,00		0,0027 (0,0000)	0,0134 (0,0000)	0,0668 (0,0080)	0,0294 (0,0000)	0,2433 (0,0802)	0,1390 (0,0000)
R.n.	1,00	1,00	1,00		0,0053 (0,0000)	0,0668 (0,0053)	0,0187 (0,0000)	0,2326 (0,0668)	0,1203 (0,0000)
G.g.	1,00	1,00	1,00	1,00		0,0508 (0,0000)	0,0134 (0,0000)	0,2273 (0,0508)	0,0989 (0,0000)
Z.d.	0,35	0,31	0,36	0,39	1,00		0,0535 (0,0000)	0,1872 (0,0187)	0,0825 (0,0000)
I.p.	1,00	1,00	1,00	1,00	1,00	1,00		0,1337 (0,0000)	0,0455 (0,0000)
B.f.	0,18	0,23	0,23	0,25	0,28	0,35	1,00		0,0455 (0,0000)
C.i.	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	

Примечание. В скобках указаны значения индекса несоответствия. Жирным шрифтом выделена вероятность отклонения нулевой гипотезы о гомогенной картине замен. Ячейки со значениями  $P > 0,05$  выделены серым цветом.

Таблица 3

Значения композиционной дистанции, индекса несоответствия и вероятности отклонения нулевой гипотезы о гомогенной картине замен для третьего положения кодона в последовательностях мРНК, кодирующих М-изоферменты креатинкиназ хордовых

	H.s.	B.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,1123 (0,0000)	0,1150 (0,0000)	0,0963 (0,0000)	1,6604 (1,2513)	1,5455 (1,2032)	1,5642 (1,1765)	1,8316 (1,4251)	9,5053 (8,9465)
B.t.	1,00		0,0989 (0,0000)	0,2299 (0,0000)	1,0936 (0,6925)	1,4225 (1,0963)	2,2701 (1,9064)	1,9813 (1,5722)	11,5588 (10,9626)
M.m.	1,00	1,00		0,0829 (0,0000)	1,6497 (1,2674)	1,3235 (0,9545)	1,5267 (1,1230)	1,2112 (0,8102)	10,6872 (10,1096)
R.n.	1,00	1,00	1,00		1,8262 (1,4332)	1,0749 (0,7166)	1,1952 (0,7941)	1,2380 (0,8396)	9,7567 (9,1765)
G.g.	0,02	0,06	0,02	0,01		1,3717 (1,0160)	5,9599 (5,5241)	4,8476 (4,3690)	18,4626 (17,8422)
Z.d.	0,02	0,01	0,03	0,04	0,03		3,3021 (2,9385)	2,3930 (1,9626)	15,4251 (14,8610)
I.p.	0,01	0,00	0,02	0,04	0,00	0,00		0,9251 (0,5374)	5,6150 (5,0455)
B.f.	0,01	0,01	0,04	0,04	0,00	0,00	0,09		10,8503 (10,2834)
C.i.	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	

Примечание. В скобках указаны значения индекса несоответствия. Жирным шрифтом выделена вероятность отклонения нулевой гипотезы о гомогенной картине замен. Ячейки со значениями  $P > 0,05$  выделены серым цветом.

Анализируя полученные данные, установлено, что в проведенных попарных сравнениях наибольшие значения композиционной дистанции характерны для третьего положения кодона, а наименьшие – для второго. Единственным исключением является попарное сравнение последовательностей мРНК, кодирующих М-изоферменты креатинкиназы мыши и крысы, для которых композиционные дистанции по второму и третьему положению равны. Что же касается значений индекса несоответствия, то в 23 сравнениях (63,9±8,0%) они максимальны для третьего положения, а минимальны для второго.

Гомогенная картина замен в первом положении нуклеотида в кодоне наблюдается в 28 проведенных попарных сравнениях (77,8±6,9%), во втором – в 36 сравнениях (100%), а в третьем – в 8 сравнениях (22,2±6,9%). Такие результаты можно объяснить максимальным количеством замен нуклеотидов в третьем положении кодона и минимальным во втором, что, в свою очередь, связано с долей синонимичных замен.

В табл. 4 показаны эволюционные дистанции по третьему положению нуклеотида в кодоне, если картина замен всегда гомогенна (контроль) и, учитывая определенный нами ранее вид картины замен (опытные данные). Выбор третьего положения кодона обусловлен наименьшим процентом гомогенной картины замен.

Таблица 4

Эволюционные дистанции и их ошибка для третьего положения кодона в последовательностях мРНК, кодирующих М-изоферменты креатинкиназы хордовых

	H.s.	B.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,2417± 0,0088	0,3215± 0,0109	0,3539± 0,0125	0,7655± 0,1561	0,5445± 0,0335	0,6352± 0,0353	0,6437± 0,0075	1,0686± 0,0083
B.t.	0,2417± 0,0088		0,3320± 0,0125	0,3831± 0,0167	0,6876± 0,0796	0,5029± 0,0272	0,5519± 0,0177	0,6569± 0,0094	1,3066± 0,0507
M.m.	0,3215± 0,0109	0,3320± 0,0125		0,1135± 0,0016	0,7065± 0,0836	0,6750± 0,0758	0,6678± 0,0300	0,6402± 0,0118	1,1729± 0,0187
R.n.	0,3539± 0,0125	0,3831± 0,0167	0,1135± 0,0016		0,7410± 0,0824	0,6327± 0,0642	0,6684± 0,0133	0,6263± 0,0070	1,2062± 0,0291
G.g.	0,7159± 0,0870	0,6876± 0,0796	0,6642± 0,0688	0,6940± 0,0679		0,5843± 0,0315	0,8083± 0,0629	0,9625± 0,0480	1,5001± 0,1412
Z.d.	0,5259± 0,0301	0,4848± 0,0251	0,6349± 0,0611	0,5985± 0,0507	0,5505± 0,0319		0,6034± 0,0485	0,7954± 0,0445	1,1165± 0,0238
I.p.	0,6192± 0,0296	0,5437± 0,0195	0,6534± 0,0308	0,6498± 0,0310	0,7872± 0,0617	0,5953± 0,0410		0,5839± 0,0167	1,1942± 0,0510
B.f.	0,6268± 0,0194	0,6404± 0,0234	0,6259± 0,0253	0,6120± 0,0210	0,9008± 0,0637	0,7525± 0,0523	0,5839± 0,0167		1,1328± 0,0196
C.i.	1,1073± 0,0340	1,3881± 0,0878	1,2331± 0,0537	1,2828± 0,0680	1,5363± 0,2207	1,2244± 0,0785	1,2172± 0,0534	1,1916± 0,0533	

Примечание. Жирным шрифтом выделены контрольные значения эволюционных дистанций и их ошибка. Для вычисления средних значений ЭД каждого из организмов необходимо учитывать только эволюционные дистанции, полученные при сравнении последовательности, выделенных из этого организма, с последовательностями, выделенными из филогенетически вышестоящих групп организмов [1]. В табл. 5 приведены средние значения ЭД по третьему положению нуклеотида в кодоне для контрольных и опытных данных.

Таблица 5

Средние эволюционные дистанции и их ошибка для третьего положения кодона в последовательностях мРНК, кодирующих М-изоферменты креатинкиназы хордовых (контроль и опытные данные)

	Средняя ЭД ± m, контроль	Средняя ЭД ± m, опыт
B.t.	0,2417±0,0088	0,2417±0,0088
M.m.	0,3268±0,0074	0,3268±0,0074
R.n.	0,3685±0,0206	0,3685±0,0206
G.g.	0,6904±0,0123	0,7252±0,0201
Z.d.	0,5589±0,0296	0,5879±0,0342
I.p.	0,6414±0,0367	0,6558±0,0386
B.f.	0,6775±0,0457	0,7013±0,0541
C.i.	1,2726±0,0503	1,2247±0,0633

Различия между приведенными в табл. 5 контрольными и опытными значениями ЭД не являются статистически достоверными. Несмотря на это следует сопоставить полученные результаты с предполагаемыми временами дивергенции: 500-600 млн. лет для оболочника, 550 млн. лет для ланцетника [2], 405 млн. лет для рыб, 330 млн. лет для пресмыкающихся, 310 млн. лет для птиц, 110 млн. лет для грызунов, 90 млн. лет для парнокопытных [4] (рис. 1).

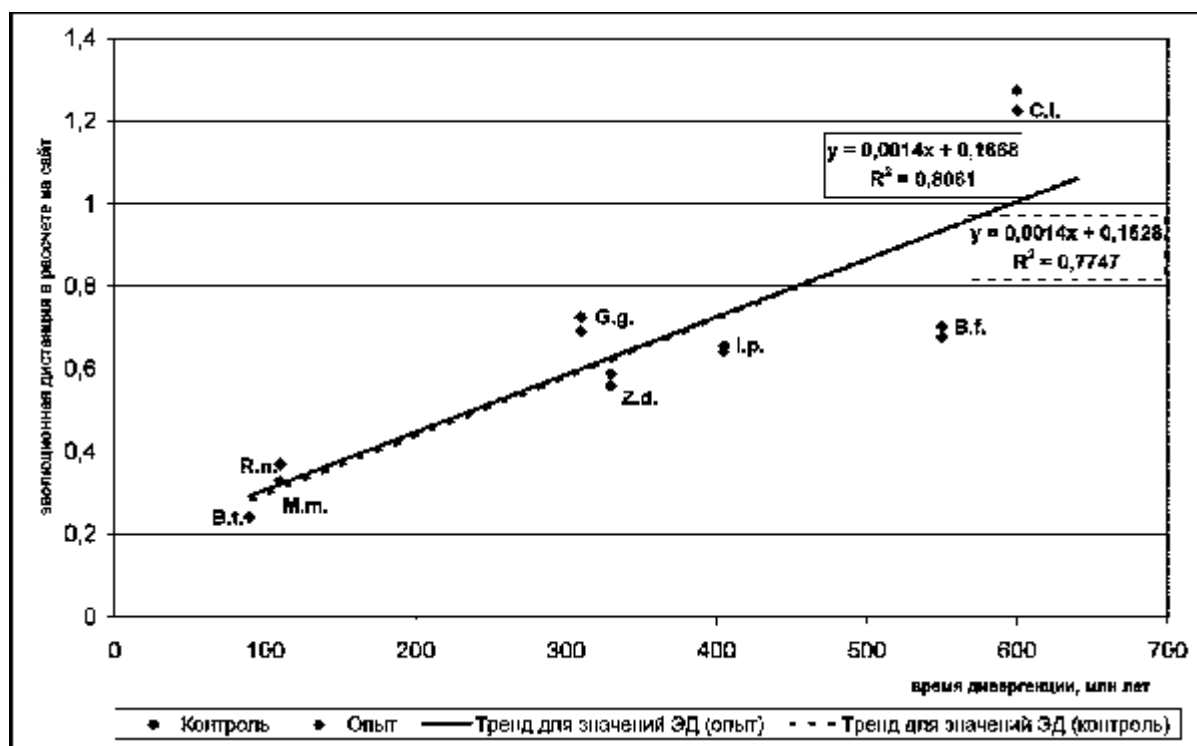


Рис. 1. Зависимость между предполагаемыми временами дивергенции и значениями эволюционных дистанций по третьему положению кодона последовательностей мРНК, кодирующих М-изоферменты креатинкиназы хордовых (контроль и опытные данные).

Средние значения эволюционных дистанций (опыт) более точные, чем значения контроля, о чем свидетельствует коэффициент аппроксимации тренда (0,8061 по сравнению с 0,7747). Поэтому при вычислении эволюционных дистанций, расчете на их основании скоростей молекулярной эволюции и построении дендрограмм следует учитывать вид картины замен. Креатинкиназа, безусловно, является структурно-функционально сформированным белком, на мутации в кодирующей части которого налагаются жесткие ограничения. Однако у других белков, выполняющих менее важные функции и подверженных меньшим ограничениям, вероятно, будет наблюдаться больший процент гетерогенной картины замен. Поэтому для них определение вида картины замен будет иметь большее значение.

Выводы.

- 1) Минимальные значения композиционной дистанции характерны для второго положения нуклеотида в кодоне, а наибольшие – для третьего.
- 2) Во втором и первом положениях кодона преобладает гомогенная картина замен, а в третьем-гетерогенная.
- 3) Учет вида картины замен позволяет получить более точные значения эволюционных дистанций для последующего вычисления скоростей молекулярной эволюции и построения дендрограмм.

### Литература

1. Бутвиловский В.Э., Давыдов В.В., Бутвиловский А.В., Семак В.И., Муравейко Е.Ф. // Медицинский журнал. – Минск, 2005, №3. – С. 43-46.
2. Малахов В.В. // Соросовский образовательный журнал.-1996. – т. 7. – С. 2-9.
3. Benfield P.A., Henderson L., Pearson M.L. // Gene. – 1985. – Vol. 39 (2-3). – P. 263-267.
4. Feng D.-F., Cho G., Doolittle R.F. // Proc. Natl. Acad. Sci. USA. – 1997. – Vol.

94. – P. 13028-13033.

5. Graber N.A., Ellington W.R. //Mol. Biol. Evol.. – 2001. – Vol. 18 (7). – P. 1305-1314.
6. Jukes T.H., Cantor C.R. //In H.N. Munro, ed., Mammalian protein Metabolism. – 1969. – P.21 – 132. Academic Press, New York.
7. Kim S., Li P., Zheng X., Dunham R.A., Liu Z. //Fish Physiol. Biochem.. – 2003. – In press.
8. Kimura M. //J. Mol. Evol. – 1980. – Vol.16. – P.111-120.
9. Kumar S., Gadagkar S.R. //Genetics – 2001. – Vol. 158. – P. 1321-1327.
10. Kumar S., Tamura K., Nei M. //Brief. Bioinform.-2004.-Vol.5. – P.150-163.
11. Orhadl C.P., Evans G.L., Cooper T.A., Kunz G., Perriard J.C. //J. Biol. Chem.. – 1984. – Vol. 259 (24). – P.15224-15227.
12. Shanti K., Kanungo M.S. //Mol. Biol. Rep.. – 2004. – Vol 31 (1).-P. 43-50.
13. Tajima F., Nei M. //Mol. Biol. Evol.. – 1984. – Vol.1. – P.269 – 285.
14. Tamura K. // Mol. Biol. Evol. – 1992. – Vol.9. – P.678-687.
15. Tamura K., Nei M. //Mol. Biol. Evol. – 1993. – Vol.10. – P.512-526.
16. Tompson J.D., Higgins D.G., Gibson T.J. //Nucl. Acids Res.. – 1994. – Vol. 22. – P. 4673 – 4680.
17. Trask R.V., Strauss A.W., Billadello J.J.//J. Biol. Chem. – 1988. – Vol. 263 (32). – P. 17142-17149.
18. Zhang F.-W., Zhao F.-K., Xu G.-J. //Sheng Wu Hua Hsueh Yu Sheng Wu Wu Li Hsueh Pao 32. – 2000.-In press.