

Определение вида картины замен в нуклеотидных и аминокислотных последовательностях. Сообщение 2. Практическое применение (сравнительный анализ М-изоферментов креатинкиназ хордовых животных)

Белорусский государственный медицинский университет

В статье рассмотрены практические аспекты определения вида картины замен в нуклеотидных и аминокислотных последовательностях М-изоферментов креатинкиназ хордовых животных. Установлено, что гомогенная картина замен наблюдается во всех проведенных сравнениях изучаемых аминокислотных последовательностей и в 41,7% сравнений нуклеотидных последовательностей. Для вычисления эволюционной дистанции следует учитывать вид картины замен.

Ключевые слова: креатинкиназа, М-изофермент, композиционная дистанция, картина замен, гомогенность, гетерогенность, индекс несоответствия

Цель исследования: установить картину замен в нуклеотидных последовательностях мРНК и аминокислотных последовательностях М-изоферментов креатинкиназ хордовых и определить эволюционные дистанции между изучаемыми нуклеотидными последовательностями.

Материал и методы

Проанализированы нуклеотидные последовательности мРНК и аминокислотные последовательности М-изоферментов креатинкиназ различных хордовых животных: оболочника (*Ciona intestinalis*, C.i., [5]), ланцетника (*Branchiostoma floridae*, B.f., [5]), рыб (*Ictalurus punctatus*, I.p., [7]), пресмыкающихся (*Zaocys dhumnades*, Z.d., [18]), птиц (*Gallus gallus*, G.g., [11]), грызунов (*Rattus norvegicus*, R.n., [3], *Mus musculus*, M.m., [12]), парнокопытных (*Bos taurus*, B.t., NCBI:AAD30974.1) и приматов (*Homo sapiens*, H.s., [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 (P) отклонения нулевой гипотезы о гомогенной картине замен нуклеотидов на 5%-ном уровне определена методом Монте-Карло при 1000 повторов [9, 10]. Картина замен считалась гомогенной при величине P больше 0,05, в обратном случае – гетерогенной.

В случае гомогенной картины замен 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]. Скорость эволюционных замен вычислена по формуле: $k_{\text{нукл.}} = \text{ЭД}/2T$, где T – число лет, прошедших после эволюционной дивергенции двух цепей от общей для них предковой цепи: множитель 2 в знаменателе соответствует двум ветвям подразумеваемого филогенетического дерева.

Результаты и обсуждение

Композиционные дистанции, полученные для нуклеотидных последовательностей мРНК и аминокислотных последовательностей М-изоферментов креатинкиназ в расчете на сайт, представлены в табл. 1.

Таблица 1

Композиционные дистанции между нуклеотидными последовательностями мРНК и аминокислотными последовательностями М-изоферментов креатинкиназ хордовых

	H.s.	B.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,0375	0,0509	0,0483	0,0938	0,1769	0,1340	0,4611	0,2031
B.t.	0,0793		0,0429	0,0322	0,0697	0,0938	0,0992	0,3458	0,1582
M.m.	0,0597	0,0401		0,0054	0,0912	0,1019	0,1287	0,3378	0,1475
R.n.	0,0775	0,1176	0,0223		0,0938	0,0912	0,1233	0,3378	0,1501
G.g.	0,7602	0,4091	0,5615	0,7121		0,1662	0,1421	0,4048	0,1903
Z.d.	0,3990	0,2513	0,1881	0,1907	0,3111		0,1475	0,1850	0,0885
I.p.	0,3206	0,6105	0,4554	0,3948	1,9902	1,1310		0,2976	0,1260
B.f.	1,1007	0,9180	0,8868	1,0143	1,9082	1,4439	0,9688		0,1743
C.i.	3,1916	4,2308	3,8779	3,5945	6,9955	5,3182	1,8512	4,9955	

Примечание. Жирным шрифтом выделены композиционные дистанции между нуклеотидными последовательностями.

Из приведенных в табл. 1 данных видно, что во всех случаях проведенных сравнений значения композиционной дистанции между нуклеотидными последовательностями достоверно выше таковых для аминокислотных последовательностей ($p < 0,01$).

Следовательно, нуклеотидный состав мРНК, кодирующих М-изоферменты креатинкиназ хордовых варьирует значительно больше по сравнению с их аминокислотным составом. Это можно объяснить тем, что на уровне аминокислотных последовательностей проявляются лишь несинонимичные замены, на которые накладываются большие структурно-функциональные ограничения, а на уровне нуклеотидных последовательностей – несинонимичные и синонимичные замены. Значения индекса несоответствия и вероятность отклонения нулевой гипотезы о гомогенной картине замен для изучаемых нуклеотидных и аминокислотных последовательностей представлены в табл. 2.

Таблица 2

Значения индекса несоответствия и вероятности отклонения нулевой гипотезы о гомогенной картине замен для нуклеотидных последовательностей мРНК и аминокислотных последовательностей М-изоферментов креатинкиназ хордовых

	H.s.	B.t.	M.m.	R.n.	G.g.	Z.d.	I.p.	B.f.	C.i.
H.s.		0,0000 (1,00)	0,0161 (0,12)	0,0080 (0,27)	0,0000 (1,00)	0,0670 (0,06)	0,0054 (0,41)	0,1823 (0,06)	0,0000 (1,00)
B.t.	0,0000 (1,00)		0,0027 (0,42)	0,0000 (1,00)	0,0000 (1,00)	0,0000 (1,00)	0,0000 (1,00)	0,0670 (0,23)	0,0000 (1,00)
M.m.	0,0000 (1,00)	0,0000 (1,00)		0,0000 (1,00)	0,0080 (0,36)	0,0000 (1,00)	0,0000 (1,00)	0,0563 (0,25)	0,0000 (1,00)
R.n.	0,0000 (1,00)	0,0125 (0,33)	0,0000 (1,00)		0,0080 (0,36)	0,0000 (1,00)	0,0000 (1,00)	0,0536 (0,27)	0,0000 (1,00)
G.g.	0,5740 (0,01)	0,2291 (0,07)	0,3922 (0,02)	0,5383 (0,01)		0,0402 (0,20)	0,0080 (0,41)	0,1099 (0,15)	0,0000 (1,00)
Z.d.	0,2291 (0,07)	0,0971 (0,17)	0,0214 (0,33)	0,0276 (0,32)	0,1373 (0,14)		0,0000 (1,00)	0,0000 (1,00)	0,0000 (1,00)
I.p.	0,1257 (0,17)	0,4207 (0,02)	0,2576 (0,07)	0,1970 (0,11)	1,7834 (0,00)	0,9376 (0,00)		0,0214 (0,37)	0,0000 (1,00)
B.f.	0,8351 (0,00)	0,6515 (0,01)	0,6248 (0,02)	0,7522 (0,01)	1,6070 (0,00)	1,1622 (0,00)	0,7094 (0,01)		0,0000 (1,00)
C.i.	2,8601 (0,00)	3,8904 (0,00)	3,5410 (0,00)	3,2558 (0,00)	6,6462 (0,00)	4,9777 (0,00)	1,5258 (0,00)	4,6542 (0,00)	

Примечание. Жирным шрифтом выделены значения для нуклеотидных последовательностей. В скобках указана вероятность отклонения нулевой гипотезы о гомогенной картине замен. Ячейки со значениями $P > 0,05$ выделены серым цветом.

Из данных, показанных в табл. 2 видно, что значения индекса несоответствия для изучаемых нуклеотидных последовательностей достоверно выше таковых для аминокислотных последовательностей ($p < 0,01$). Примечательно, что во всех проведенных попарных сравнениях аминокислотных последовательностей М-

изоферментов креатинкиназ наблюдается гомогенная картина замен, что, вероятно, связано с жесткими структурно-функциональными ограничениями на замены аминокислот. В то же время для нуклеотидных последовательностей гомогенная картина замен обнаруживается лишь в 15 из 36 проведенных сравнений ($41,7 \pm 8,2\%$). Таким образом, определение картины замен для последующего расчета эволюционной дистанции наиболее важно для нуклеотидных последовательностей.

Наличие двух картин замен в нуклеотидных последовательностях мРНК, кодирующих М-изоферменты креатинкиназ хордовых, свидетельствует о необходимости дифференцированного подхода при определении эволюционных дистанций между нуклеотидными последовательностями. Поэтому для попарных сравнений, в которых обнаружена гомогенная картина замен, следует определять средние эволюционные дистанции по пяти вышеназванным методам, учитывающим этот вид картины замен. Для попарных сравнений, в которых наблюдается гетерогенная картина замен, следует вычислять средние эволюционные дистанции лишь по тем трем методам, которые содержат модифицированные формулы для учета гетерогенности. Вычисленные таким способом средние эволюционные дистанции отображены в табл. 3.

Таблица 3

Средние эволюционные дистанции и их стандартная ошибка для нуклеотидных последовательностей 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.		0,0004	0,0004	0,0004	0,0006	0,0006	0,0005	0,0001	0,0003
R.t.	0,0901		0,0006	0,0006	0,0008	0,0005	0,0005	0,0005	0,0007
M.m.	0,1018	0,1090		0,0001	0,0006	0,0009	0,0007	0,0007	0,0005
R.n.	0,1129	0,1164	0,0369		0,0005	0,0007	0,0006	0,0005	0,0005
G.g.	0,2152	0,2105	0,1917	0,1977		0,0005	0,0007	0,0006	0,0005
Z.d.	0,1856	0,1728	0,1903	0,1835	0,1995		0,0012	0,0012	0,0009
I.p.	0,2246	0,2192	0,2291	0,2295	0,2417	0,2214		0,0004	0,0014
B.f.	0,3310	0,3343	0,3261	0,3258	0,3910	0,3592	0,3231		0,0011
C.i.	0,4371	0,4539	0,4473	0,4518	0,4678	0,4540	0,4268	0,4595	

Примечание. Жирным шрифтом выделены значения эволюционных дистанций

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 организмов. При расчете средней ЭД оболочника следует учитывать эволюционные дистанции, полученные при сравнении нуклеотидной последовательности мРНК, кодирующей его М-изофермент креатинкиназы, с последовательностями изучаемого фермента ланцетника, рыб, пресмыкающихся, птиц и четырех видов представителей класса Млекопитающие.

Так, средняя ЭД для *Ciona intestinalis* равна $0,4498 \pm 0,0049$, для *Branchiostoma floridae* – $0,3415 \pm 0,0102$, для *Ictalurus punctatus* – $0,2276 \pm 0,0036$, для *Zoocys dhumades* – $0,1863 \pm 0,0049$, для *Gallus gallus* – $0,2038 \pm 0,0063$, для *Rattus norvegicus* – $0,1147 \pm 0,0025$, для *Mus musculus* – $0,1054 \pm 0,0051$ и для *Bos taurus* – $0,0901 \pm 0,0004$. Интересно сопоставить предполагаемые времена дивергенции (500-600 млн. лет для оболочника, 550 млн. лет для ланцетника [2], 405 млн. лет для рыбы, 330 млн. лет для пресмыкающихся, 310 млн. лет для птиц, 110 млн. лет для грызунов, 90 млн. лет для парнокопытных [4]) с полученными эволюционными дистанциями (рис. 1).

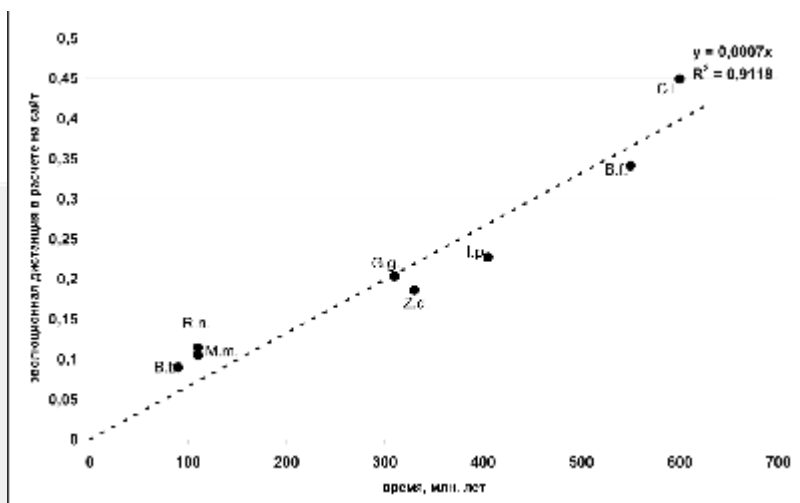


Рис. 1. Зависимость между значениями эволюционных дистанций и временем дивергенции М-изофермента креатинкиназ хордовых животных.

Все точки достаточно хорошо ложатся на тренд, что подтверждает высокая достоверность его аппроксимации ($R^2=0,91$). Наклон тренда равен 0,0065, поэтому $k_{\text{нукл}} = \Delta D / 2T = 0,0065 / 2 \cdot 100 \text{ млн. лет} \cdot 0,33 \cdot 10^{-9}$ замен на нуклеотидный сайт в год, что свидетельствует о приблизительно постоянной скорости эволюции изучаемых нуклеотидных последовательностей. Следует отметить, что вычисленная нами ранее скорость эволюции аминокислотных последовательностей М-изофермента креатинкиназ хордовых животных ($0,18 \cdot 10^{-9}$ замен на аминокислотный сайт в год [1]) почти в два раза меньше таковой для аминокислотных последовательностей, что связано со значительной долей синонимичных замен.

Анализируя полученные результаты можно сделать следующие выводы:

1. Значения композиционной дистанции между нуклеотидными последовательностями достоверно выше таковых для аминокислотных последовательностей М-изоферментов креатинкиназ хордовых.
2. Гомогенная картина замен в изученных аминокислотных последовательностях М-изоферментов креатинкиназ наблюдалась во всех проведенных попарных сравнениях, а в нуклеотидных последовательностях – лишь в 41,7%, что обуславливает необходимость определения вида картины замен на начальном этапе сравнительного анализа последовательностей макромолекул.
3. Скорость эволюции изучаемых нуклеотидных последовательностей, скорректированная с учетом вида картины замен, приблизительно равна $0,33 \cdot 10^{-9}$ замен на сайт в год, что почти в два раза выше таковой для аминокислотных последовательностей.

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