

Положительный отбор в альтернативных областях альтернативно сплайсируемых генов человека

Екатерина Олеговна Ермакова^{1,2} *ermakova8@yandex.ru* ¹ Факультет биоинженерии и биоинформатики МГУ им. М.В. Ломоносова, 119992, г. Москва, ГСП-2, Ленинские горы, МГУ 1, стр. 73
Михаил Сергеевич Гельфанд^{1,2} *gelfand@iitp.ru* ² Учебно-научный центр „Биоинформатика“ ИППИ РАН, 127994, г. Москва, Большой Каретный пер., д.19

У эукариот трансляции мРНК предшествует процессинг, включающий сплайсинг — вырезание интронов и соединение экзонов. Из одной пре-мРНК интроны можно вырезать несколькими способами — у млекопитающих распространён альтернативный сплайсинг. Это позволяет одному гену кодировать несколько белков.

Постоянные и альтернативные области генов эволюционируют по-разному? Да!

Мы рассмотрели выравнивания мРНК, соответствующих ортологичным генам человека и мыши, и сравнили уровень несинонимичных и синонимичных замен в их постоянных и альтернативных кодирующих областях.

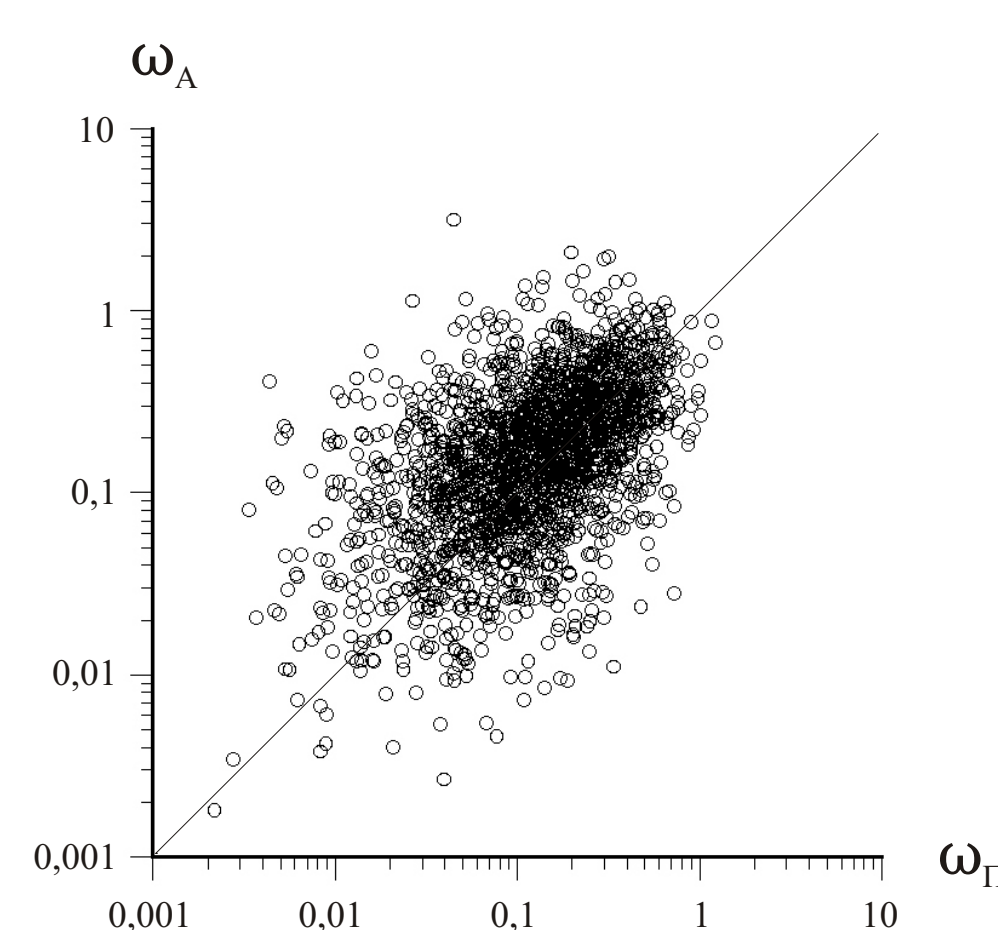
- Мы взяли **12356** выравниваний мРНК, соответствующих ортологичным генам человека и мыши [2], [5].
- Для **5754** генов человека в базе EDAS [6] было найдено несколько различных белков. Белки были выравнены против человеческих мРНК программой ProFrame [3]. Это позволило выделить в выравниваниях постоянные и альтернативные кодирующие участки.
- Для **3029** выравниваний степень сходства превысила 70%, а альтернативные участки транслировались в одной рамке.
- Для **2358** выравниваний длина альтернативных участков превысила 80 п.н..

Эволюция гена состоит в накоплении нуклеотидных замен. При подсчёте замен в кодирующей области число несинонимичных замен обычно нормируют на число несинонимичных позиций, а число синонимичных замен — на число синонимичных позиций.

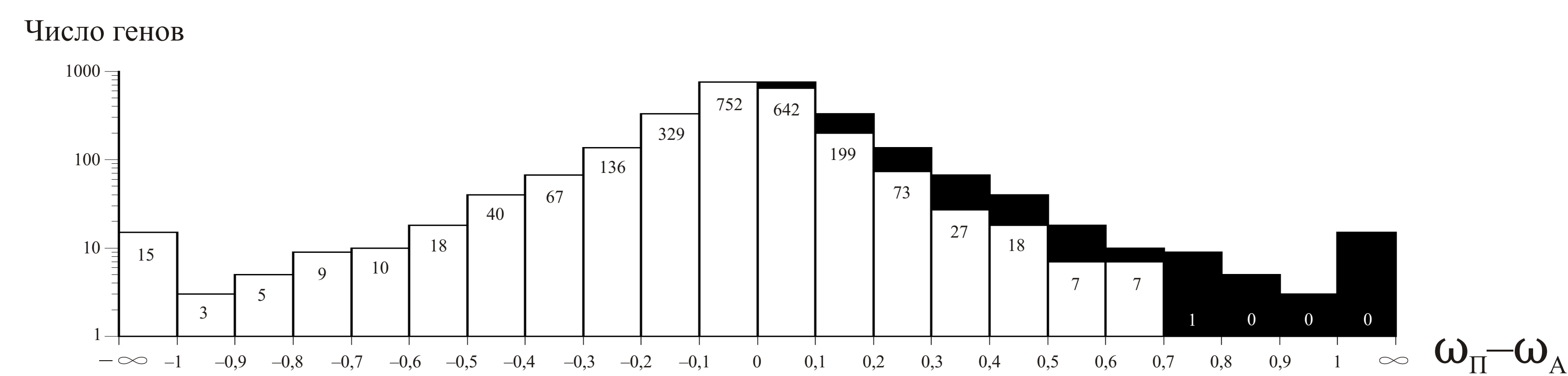
Вторая позиция кодона TCT несинонимичная	Третья позиция кодона TCT синонимичная	Пример несинонимичной замены	Пример синонимичной замены
TTT <i>Phe</i>	TCT <i>Ser</i>	GTA <i>Val</i>	GTA <i>Val</i>
TCT <i>Ser</i>	TCC <i>Ser</i>	↓	↓
TAT <i>Tyr</i>	TCA <i>Ser</i>	GCA <i>Ala</i>	GTG <i>Val</i>
TGT <i>Cys</i>	TCG <i>Ser</i>		

Метод Ины [4] позволяет оценить количество несинонимичных замен на несинонимичную позицию d_N и синонимичных замен на синонимичную позицию d_S , произошедших на кодирующем участке гена с момента его расхождения с мышинным ортологом. $\omega = d_N/d_S$ используют как меру давления отбора.

Мы сравнили отношение скоростей несинонимичных и синонимичных замен ω в постоянных (ω_{Π}) и альтернативных (ω_A) областях 2358 генов человека.



Для большинства генов $\omega_A > \omega_{\Pi}$.



Распределение $\omega_{\Pi} - \omega_A$ асимметрично (тень на графике показывает зеркальное отражение левой части): $\omega_A > \omega_{\Pi}$ для 1384 генов, $\omega_A < \omega_{\Pi}$ для 974 генов. Следовательно, в среднем альтернативные кодирующие участки эволюционируют быстрее, чем постоянные.

Альтернативные области генов эволюционируют быстрее, чем постоянные. Быстрее всего эволюционируют альтернативные области, соответствующие С-концевым участкам белков.

В альтернативных областях генов положительный отбор сильнее и/или отрицательный отбор слабее. Различия заметнее на быстро эволюционирующих генах.

- [1] Iida, K., and H. Akashi. 2000. A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. *Gene*. 261:93-105.
- [2] Jordan, I.K., F.A. Kondrashov, I.B. Rogozin, R.L. Tatusov, Y.I. Wolf, and E.V. Koonin. 2001. Constant relative rate of protein evolution and detection of functional diversification among bacterial, archaeal and eukaryotic proteins. *Genome Biol.* 2:research0053.1-0053.9
- [3] Mironov, A.A., P.S. Novichkov, and M.S. Gelfand. 2001. Pro-Frame: similarity-based gene recognition in eukaryotic DNA sequences with errors. *Bioinformatics*. 17:13-15
- [4] Ina, Y.. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. *J. Mol. Evol.* 40:190-226
- [5] <http://www.ncbi.nlm.nih.gov/RefSeq/>
- [6] <http://www.ig-msk.ru:8005/EDAS>

Мы выделили постоянные и альтернативные кодирующие участки в 3029 выравниваниях альтернативно сплайсируемых генов человека и мыши и вычислили среднюю степень аминокислотного сходства, уровень несинонимичных и синонимичных замен и отношение скоростей несинонимичных и синонимичных замен. Мы также рассмотрели отдельно гены, эволюционирующие медленно, со средней скоростью и быстро. Для этого мы поделили гены на три равные группы, а за меру скорости эволюции взяли степень нуклеотидного сходства — долю идентичных нуклеотидов в выравнивании.

Гены	Участки	Суммарная длина выравнивания, п.н.	Степень аминокислотного сходства	d_N	d_S	ω
Набор 1	постоянные	2822439	0,891	0,068	0,408	0,166
	альтернативные	3081642	0,879	0,077	0,411	0,187
Медленные	постоянные	897471	0,964	0,020	0,323	0,063
	альтернативные	929970	0,960	0,023	0,320	0,072
Средние	постоянные	978984	0,913	0,052	0,418	0,124
	альтернативные	1092459	0,903	0,059	0,419	0,140
Быстрые	постоянные	945984	0,800	0,134	0,485	0,277
	альтернативные	1068213	0,785	0,147	0,489	0,300

Уровень синонимичных замен на альтернативных и постоянных участках примерно одинаков. Уровень несинонимичных замен в альтернативных областях выше, чем в постоянных, а степень аминокислотного сходства — ниже. Следовательно, скорость эволюции альтернативных участков выше. Отношение скоростей несинонимичных и синонимичных замен ω в альтернативных областях также выше. Следовательно, в альтернативных областях положительный отбор сильнее и/или отрицательный отбор слабее, чем в постоянных.

Гены	Участки	Суммарная длина выравнивания, п.н.	Степень аминокислотного сходства	d_N	d_S	ω
Набор 1	постоянные	2822439	0,891	0,068	0,408	0,166
	N-концевые альтернативные	2194521	0,880	0,075	0,405	0,186
	внутренние альтернативные	790026	0,884	0,074	0,421	0,176
	S-концевые альтернативные	97095	0,813	0,132	0,452	0,291
Медленные	постоянные	897471	0,964	0,020	0,323	0,063
	N-концевые альтернативные	679423	0,961	0,022	0,319	0,069
	внутренние альтернативные	230754	0,957	0,025	0,322	0,076
	S-концевые альтернативные	19593	0,940	0,036	0,327	0,111
Средние	постоянные	978984	0,913	0,052	0,418	0,124
	N-концевые альтернативные	751620	0,903	0,059	0,413	0,142
	внутренние альтернативные	313896	0,905	0,057	0,431	0,133
	S-концевые альтернативные	26943	0,887	0,073	0,450	0,162
Быстрые	постоянные	945984	0,800	0,134	0,485	0,277
	N-концевые альтернативные	772278	0,788	0,143	0,481	0,296
	внутренние альтернативные	245376	0,787	0,147	0,511	0,288
	S-концевые альтернативные	50559	0,724	0,207	0,509	0,407

Мы также рассмотрели отдельно альтернативные области, соответствующие N-концевым, С-концевым и внутренним участкам белка. Оказалось, что С-концевые альтернативы вносят наибольший вклад в наблюдаемые различия.

Спасибо

I. King Jordan за выравнивания мРНК человека и мыши,
Рамилю Нуртдинову за белковые изоформы альтернативно сплайсируемых генов человека,
Егору Базыкину, Андрею Миронову и Дмитрию Петрову за плодотворные обсуждения.

Нашу работу поддержали
Российский фонд фундаментальных исследований
Российская академия наук
Howard Hughes Medical Institute
Ludwig Institute for Cancer Research