

РЕКОНСТРУКЦИЯ ПРЕДКОВОГО ГЕНА *Xist* И ПУТИ ЕГО ЭВОЛЮЦИИ У МЛЕКОПИТАЮЩИХ

Е.А. Елисафенко, Н.Н. Колесников, А.И. Шевченко, С.М. Закиян

Институт цитологии и генетики СО РАН, Новосибирск, Россия, e-mail: zakian@bionet.nsc.ru

Xist является регуляторным геном. Его транскрипт, представляющий собой протяженную нетранслируемую ядерную РНК, играет ключевую роль в процессе инактивации X-хромосомы у самок плацентарных млекопитающих. В работе проведен сравнительный анализ гена *Xist* у 10 видов плацентарных млекопитающих. Показано, что *Xist* произошел из белок-кодирующего гена, который в результате экспансии мобильных генетических элементов и тандемных повторов трансформировался в ген, кодирующий протяженную нетранслируемую ядерную РНК, участвующую в регуляции экспрессии генов целой хромосомы. В процессе эволюции современных видов млекопитающих *Xist* аккумулирует видоспецифические отличия, которые, вероятно, необходимы для адаптации РНК гена *Xist* к функционированию в геноме и X-хромосоме каждого конкретного вида.

Введение

Ген *Xist* инициирует процесс инактивации одной из двух X-хромосом, который происходит в раннем развитии у самок плацентарных млекопитающих. *Xist* входит в состав сложного локуса X-хромосомы, названного центром инактивации (X inactivation center, XIC). Во время инициации инактивации некодирующая РНК гена *Xist* распространяется вдоль будущей неактивной X-хромосомы, репрессирует экспрессию генов и запускает каскад модификаций хроматина, поддерживающих ее неактивное состояние (см. обзор Chang *et al.*, 2006).

Ген *Xist* имеет экзоны и интроны, однако не содержит протяженных открытых рамок считывания. РНК гена *Xist* подвергается процессингу, альтернативному сплайсингу, полиаденилированию и не покидает ядра. В составе экзонов гена *Xist* выявлены 6 типов простых тандемных повторов, обозначенных А, В, С, D, Е и F, и мобильные генетические элементы (Brockdorff *et al.*, 1992; Brown *et al.*, 1992; Nesterova *et al.*, 2001). Повтор А, локализующийся в первом экзоне, имеет важное функциональное значение в обеспечении транскрипционного молчания генов X-хромосомы, тогда как другие повторы, возможно, опосредуют связь *Xist*-РНК с хроматином (Wutz *et al.*, 2002).

С помощью комплекса биоинформационных программ и подходов мы сравнили особенности нуклеотидной последовательности и структуры гена *Xist* у 10 видов плацентарных млекопитающих и выяснили происхождение и особенности эволюции этого гена.

Материалы и методы

Сравнительный анализ структуры гена *Xist* проводили у плацентарных млекопитающих: человека (*Homo sapiens*), шимпанзе (*Pan troglodytes*), коровы (*Bos taurus*), лошади (*Equus caballus*), собаки (*Canis familiaris*), кролика (*Oryctolagus cuniculus*), мыши (*Mus musculus*), крысы (*Rattus norvegicus*) и 4 видов полевок (*Microtus rossiaemeridionalis*, *M. arvalis*, *M. kirgisorum* и *M. transcaspicus*). Для сравнительного анализа использовали последовательности из соответствующих баз данных секвенированных геномов на сайте UCSC Genome Bioinformatics и последовательности XIC 4 видов полевок и коровы, зарегистрированные в базе данных под номерами: AJ310127, AJ310130, AJ310129, AY090554 и AJ421481 соответственно.

Экзон-интронная структура гена *Xist* человека, мыши и полевок была определена ранее (Nesterova *et al.*, 2001; Chureau *et al.*, 2002).

Экзон-интронную структуру гена *Xist* собаки, шимпанзе и коровы определяли на основании сравнения с геном *Xist* человека и мыши. Предсказанные границы экзонов и интронов *Xist* коровы и шимпанзе подтвердили с помощью анализа многочисленных транскриптов *Xist*, представленных в базах данных нуклеотидных последовательностей для этих видов.

Поиск гомологичных последовательностей проводили с помощью пакета программ BLAST (Altschul *et al.*, 1990, <http://www.ncbi.nlm.nih.gov/>). Для поиска тандемных повторов и мобильных элементов использовали программы TRF (Benson, 1999) и RepeatMasker (Smit, Green, <http://www.repeatmasker.org>) соответственно. Выравнивания двух и более последовательностей проводили программами FASTA (Pearson, Lipman, 1988) и CLUSTALX (Jeanmougin *et al.*, 1998). Сравнение протяженных районов генома осуществляли программой PipMaker (<http://bio.cse.psu.edu/>), а также с помощью программ, представленных на серверах: <http://genome.ucsc.edu/> и <http://www.ensembl.org/>.

Результаты

Лабильность экзон-интронной структуры гена *Xist*

Межвидовые сравнения показали, что экзон-интронная структура гена *Xist* не является строго консервативной (рис. 1). Экзоны 1, 3, 4 и 8 гена *Xist* входят в состав РНК и совпадают у всех видов. Экзон 2, входящий в РНК *Xist* у мыши, полевки, крысы и коровы, инактивирован у человека, собаки и шимпанзе и является у этих видов частью первого интрона *Xist*. Присутствующий в РНК *Xist* человека, шимпанзе и собаки экзон 2 не обнаруживается у остальных видов. Экзон 5 *Xist* мыши и полевки является фрагментом интрона 4 *Xist* у человека и коровы. Экзоны 6 и 7 *Xist* грызунов соответствуют экзонам 5 и 6 *XIST* человека. Район, гомологичный экзону 7 *XIST* человека, находится в интроне 7 *Xist* у грызунов. Экзон 2 *Xist* у крысы и экзон 7 *Xist* у мыши и полевки укорочены в 3'-концах за счет образования новых 3'-5'-экзон-интронных границ. Таким образом, всего обнаружено 10 экзонов гена *Xist*, но только 6 из них выявляются в составе РНК этого гена у всех видов. Вероятно, предковый ген *Xist* имел

10 экзонов, которые в ходе эволюции плацентарных млекопитающих либо делетировались, либо частично или полностью становились составной частью интронов из-за мутаций в сайтах сплайсинга. Используя последовательности гена *Xist* человека, собаки, коровы и мыши, мы скомпилировали консенсус гена *Xist*, в составе которого выявляются все 10 экзонов, и предположили, что он в определенной степени может отражать последовательность гена предка плацентарных млекопитающих, существовавшего около 100 миллионов лет назад в период начала радиации Eutheria. Мы использовали полученную консенсусную последовательность при дальнейшем анализе эволюции гена *Xist* плацентарных млекопитающих.

Общая длина гена *Xist* варьирует у разных видов от 21 до 37 т.п.н. (табл. 1). Ген короче у представителей отряда Rodentia (21–23 т.п.н.) и длиннее у остальных видов (32–37,5 т.п.н.). Сокращение длины *Xist* у грызунов происходит в основном из-за уменьшения интронов. Внутри отряда грызунов размер гена более консервативен. Различие размера *Xist* у мыши и полевки составляет ~ 1,5 т.п.н., а у близкородственных видов полевок – 0,7 т.п.н. (табл. 1).

Суммарная длина экзонов и интронов *Xist* у представителей разных отрядов существенно отличается (табл. 1). У полевок длина экзонов варьирует от 13,4 т.п.н. до 14,1 т.п.н. Самое большое значение суммарной длины экзонов – 24,5 т.п.н. – выявляется у быка, а самый большой суммарный размер интронов – 16,2 т.п.н. – обнаруживается у собаки.

Из всех экзонов самым вариабельным по размеру оказался экзон 1 (рис. 1). Его длина может различаться в два раза (например, у быка и полевки). Это обусловлено различиями в степени амплификации тандемных повторов (от А до F), находящихся в его составе. Дополнительный вклад в вариабельность размера экзона 1 вносят инсерции видоспецифических SINE (рис. 1). Остальные экзоны гена незначительно варьируют по своим размерам. Например, экзон 3 и экзон 4 у всех видов различаются всего на несколько нуклеотидов (табл. 1).

Интроны гена *Xist* претерпевали в ходе эволюции многократные изменения. Размер соответствующих друг другу интронов у разных видов может отличаться в несколько раз.

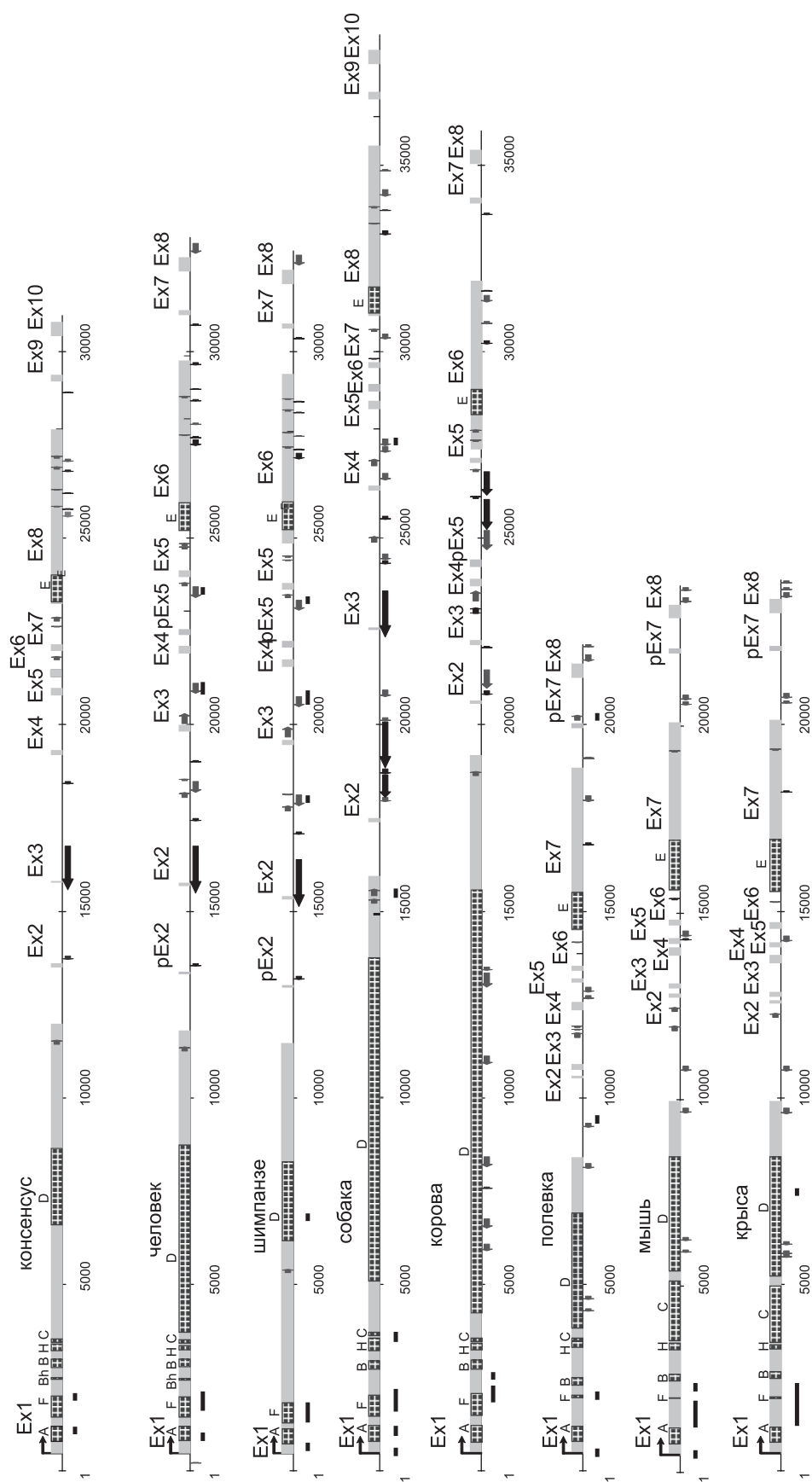


Рис. 1. Сравнение организации гена *Xist* у 7 видов млекопитающих и консенсусная (предковая) последовательность гена *Xist*. Серыми прямоугольниками обозначены экзоны (Ex) и псевдоэкзоны (pEx). Стрелки – LINE и SINE; решетка – простые тандемные повторы A, B, C, D, E, F, H; черные линии – CpG-островки. Для гена *Xist* шимпанзе нет окончательных данных о тандемных повторах B, Bh и C, так как последовательность нуклеотидов в данном районе определена не полностью. Псевдоэкзон (pEx) – часть последовательности интрона, которая у других видов является экзоном.

Таблица 1
Сравнение размеров (в парах нуклеотидов) гена *Xist* и его структурных элементов – экзонов (Ex) и псевдоэкзонов (pEx) у *C. familiaris* (C.f.), *B. taurus* (B.t.), *H. sapiens* (H.s.), *P. troglodytes* (P.t.), *R. norvegicus* (R.n.), *M. musculus* (M.m.), *M. rossiaemeridionalis* (M.r.) и консенсусного гена *Xist* (con.)

Вид	Размер гена	Размер гомологичных экзонов										Суммарный размер интронов		Суммарный размер экзонов
		Ex1	pEx2	Ex2	Ex3	Ex4	pEx5	Ex5	Ex6	Ex7	Ex8	Суммарный размер интронов	Суммарный размер экзонов	
C.f.	37592	15480	96	59	140	211	162	130	4561	195	370	16188	21404	
B.t.	34934	18693	90	–	137	210	164	131	4524	156	372	10457	24477	
H.s.	32063	11333	90	64	137	209	161	164	4543	146	378	14838	17225	
P.t.	32050	11316	90	64	137	209	144	164	4541	146	378	14861	17189	
R.n.	22898	Ex1	Ex2	–	Ex3	Ex4	Ex5	Ex6	Ex7	pEx7	Ex8	7597	15301	
M.m.	22786	9430	83	–	137	213	141	154	4610	146	387	7565	15221	
M.r.	21161	9483	91	–	132	211	147	155	4521	141	340	7671	13490	
anc.	30297	7939	84	–	138	213	103	134	4361	134	384	12736	17561	
con.		11506	93	58	138	209	185	169	4678	151	374			
		Ex1	Ex2	Ex3	Ex4	Ex5	Ex6	Ex7	Ex8	Ex9	Ex10			

Примечание. Псевдоэкзон (pEx) – часть последовательности интрона, которая у других видов является экзоном. (–) – экзон делегирован.

Таблица 3

Характеристика диспергированных повторов в гене *Xist* у *C. familiaris* (C.f.), *B. taurus* (B.t.), *H. sapiens* (H.s.), *P. troglodytes* (P.t.), *R. norvegicus* (R.n.), *M. musculus* (M.m.), *M. rossiaemeridionalis* (M.r.) и консенсусном гене *Xist* (con.)

Вид	Размер гена*, п.н.	Доля повторов разного типа, %							
		Доля GC	SINE	LINE	LTR	ДНК-транспозоны	Простые повторы	Поли-N тракты	Всего повторов
C.f.	38592	38,9	5,77	8,55	–	0,35	1,31	–	15,97
B.t.	35934	40,3	5,96	7,43	0,20	0,74	0,83	–	15,26
H.s.	33063	39,8	6,37	4,55	–	0,52	0,68	0,28	12,40
P.t.	32708	39,6	6,40	4,85	–	0,53	0,36	–	12,13
R.n.	23898	42,4	6,95	–	–	–	0,95	–	7,90
M.m.	23786	41,3	6,95	–	–	0,30	2,18	0,70	10,12
M.r.	22163	42,2	7,64	–	–	0,28	1,76	1,31	10,99
con.	30986	37,7	1,64	0,46	–	0,51	0,84	0,93	4,39

* Анализируемый район у каждого вида включает ген *Xist* и по 500 п.н. 5'- и 3'-фланкирующих последовательностей.

руются компьютерными программами, хотя гомология с предковым геном в этих районах частично сохраняется.

Следует отметить, что для гена *Xist* характерно наличие в составе процессированных транскриптов мобильных элементов. Мобильные элементы либо интегрированы в состав экзона, либо сами формируют экзон-интронные границы. В транскрипте *Xist* обнаруживаются как древние, так и молодые видоспецифичные мобильные элементы. Так, первый экзон *Xist* содержит преимущественно молодые таксоноспецифические мобильные элементы, как правило, *SINE*, и лишь один древний *SINE* – *MIR3*. Во втором по величине экзоне (экзон 6 *XIST* человека и экзон 7 *Xist* мыши) присутствуют исключительно древние мобильные элементы (*MIR*, *L2*, ДНК-транспозоны). Границы экзона 2 *XIST* человека находятся внутри *LINE1*, и при сплайсинге в состав РНК включается фрагмент мобильного элемента размером 64 н. Аналогичная ситуация, по всей видимости, имеет место и у собаки.

Мы обнаружили, что экзоны 2, 3, 4, 8, 9 и 10 консенсусного *Xist* имеют гомологию с мобильными элементами разных классов (табл. 4), которые, вероятно, послужили основой для их формирования.

Тандемные повторы

В предыдущих исследованиях были идентифицированы основные типы тандемных повторов А, В, С, D, Е и F (Brockdorff *et al.*, 1992; Brown *et al.*, 1992; Nesterova *et al.*, 2001). Эти тандемные повторы составляют существенную часть гена *Xist* (табл. 5) и входят в состав его процессированного транскрипта. Они могут составлять почти половину общей длины гена и более 60 % РНК, как это наблюдается у коровы. У разных видов эволюция тандемных повторов идет по-разному. Повторы А, В, D, Е, и F, консервативные и в разной степени амплифицированные, встречаются у всех видов. Некоторые повторы видоспецифичны (С-повтор мыши, В^h-повтор человека) (рис. 1).

Кроме описанных выше, существуют также небольшие видоспецифические тандемные повторы в разных частях гена *Xist*. Однако их вклад в общую долю повторов незначителен.

Анализ консенсусного *Xist* позволяет предположить, что предковый ген уже содержал базовые домены тандемных повторов – А, В, D, Е и F, и их амплификация произошла более 100 млн лет назад. В то же время в предковом гене тандемные повторы составляли существенно меньшую часть (табл. 5). Кроме того, в

Таблица 4

Происхождение структурных элементов гена *Xist* плацентарных млекопитающих

Консенсусный <i>Xist</i>		<i>XIST</i> человека	<i>Xist</i> мыши	Гомологии с <i>Lnx3</i>	Гомологии с мобильными элементами
Pmin c1		Pmin h1	Pmin m1	Экзон 1 Экзон 2	TR, TE
TR экзона 1	A C D F H				Фрагмент ERV Фрагмент ERV Фрагмент ERV Фрагмент ДНК-транспозона LINE, L1
C2		ph2	m2		LINE, L3CR1
C3		h2	–		LINE, LIMC3
C4		h3	m3		LINE, L1
Интрон 3		Интрон 3	Интрон 3	Экзон 3	
C5		h4	m4	Экзон 4	
C6		ph5	m5	Экзон 5	
c7		h5	m6	Экзон 11	
c8		h6	m7		TR, TE
c9		h7	pm7		Фрагмент ERV
c10		h8	m8		Фрагмент ДНК-транспозона

Обозначения: (c1-c10) – экзоны консенсуса *Xist* плацентарных млекопитающих; (h1/m1 – h8/m8) – экзоны *Xist* человека и мыши соответственно; (ph, pm) – псевдоэкзоны *Xist*; Pmin – минимальный промотор; TR – простые тандемные повторы A, C, D, F, H; TE – транспозоны; ERV – эндогенные ретровирусы. Псевдоэкзон (pEx) – часть последовательности интрона, которая у других видов является экзоном.

Таблица 5

Доля тандемных повторов (ТП) в гене *Xist* у *C. familiaris* (C.f.), *B. taurus* (B.t.), *H. sapiens* (H.s.), *P. troglodytes* (P.t.), *R. norvegicus* (R.n.), *M. musculus* (M.m.), *M. rossiaemeridionalis* (M.r.) и консенсусном гене *Xist* (con.)

Вид	Размер гена, п.н.	Размер РНК, н.	Размер экзона 1, п.н.	Суммарный размер ТП в экзоне 1, п.н.	Доля ТП в экзоне 1, %	Суммарный размер ТП, п.н.	Доля ТП в РНК, %	Доля ТП в гене, %
C.f.	37592	21404	15480	5490	35,47	6191	28,92	16,47
B.t.	34934	24477	18693	14637	78,30	15309	62,54	43,82
H.s.	32063	17225	11333	7138	62,98	7832	45,47	24,43
P.t.	32050	17189	11316	3017	26,66	3742	21,77	11,68
R.n.	22898	15301	9430	5350	56,73	6749	44,11	29,47
M.m.	22786	15221	9483	5397	56,91	6748	44,33	29,61
M.r.	21161	13490	7939	4689	59,06	5703	42,28	26,95
con.	30297	17561	11506	3419	29,71	4161	23,69	13,73

консенсусном гене *Xist* выявлены тандемные повторы мономера 33 п.н., которые мы назвали повтором Н. У современных видов мономеры этого повтора дивергировали до неузнаваемости и фрагментарно сохранились только у собаки.

Обнаружено, что мономеры тандемных повторов консенсусного *Xist* имеют гомологию с мобильными элементами разных классов (табл. 4). Слабая гомология с мобильными элементами выявляется также в последовательностях, расположенных между блоками тандемных повторов.

GC-состав и CpG-островки

Ранее показано, что ХИС относительно беден по содержанию G и C нуклеотидов, обогащен повторами и обеднен белок-кодирующими генами. Содержание GC в этом районе составляет

42 % у мыши, 40 % у человека и 39 % у быка (Chureau *et al.*, 2002).

Мы оценили GC-состав самого гена и его 5'- и 3'-фланкирующих районов. Общее содержание GC в гене *Xist* сходно со значениями для целого района ХИС и варьирует от 39,8 до 42,2 % (табл. 6).

В 5'-фланкирующем районе *Xist* у всех сравниваемых видов содержание GC-нуклеотидов выше, чем в самом гене. Вариация незначительна и составляет от 41,3 до 43,2 %. Резко выделяется область минимального промотора, обогащенная GC-нуклеотидами (52,4–56,4 %).

GC-состав 3'-фланкирующей области у разных видов варьирует от 39 до 45 %. У грызунов этот район богаче по GC-составу, чем 5'-область и сам ген, тогда как у человека и быка наблюдается уменьшение содержания GC в 3'-районе

Таблица 6

GC-состав (%) различных элементов гена *Xist* человека (H.s.), собаки (C.f), коровы (B.t.), 4 видов полевок (M.k.), (M.a.), (M.r.), (M.t.), мыши (M.m.) и крысы (R.n.)

Элемент гена	Вид									
	H.s.	C.f.	Элемент гена	B.t.	M.k.	M.a.	M.r.	M.t.	M.m.	R.n.
P _{min}	56,4	55,4		53,2	52,9	52,4	53,3	53,3	54,8	54,0
Ex 1	43,5	34,5		42,5	43,0	43,1	43,1	42,0	42,5	43,9
In 1	34,5	36,5		33,6	42,6	43,4	43,2	43,2	41,6	42,7
Ex 2	48,4	36,2		52,2	41,7	39,8	3,3	38,6	48,3	51,8
In 2	37,5	35,9		40,8	31,8	32,4	31,8	32,4	33,1	43,7
Ex 3	44,5	37,1		43,1	50,0	49,3	49,3	50,0	44,7	48,9
In 3	42,1	39,6		40,5	41,7	41,0	42,2	40,8	40,6	41,9
Ex 4	47,4	49,7		48,6	54,2	54,5	54,0	53,5	54,0	54,5
In 4	39,4	36,2		34,2	42,3	42,7	42,3	43,3	29,2	34,9
			Ex 5	41,6	52,4	52,4	52,4	52,4	44,9	48,9
			In 5	40,0	43,1	38,7	38,2	38,7	39,9	41,7
Ex 5	45,1	40,8	Ex 6	39,7	45,0	44,0	45,5	45,5	43,9	44,8
In 5	38,5	36,5	In 6	35,9	42,6	42,0	41,5	42,4	39,9	42,8
Ex 6	38,6	37,4	Ex 7	36,8	39,6	40,7	39,9	40,0	38,6	39,2
In 6	34,1	31,7	In 7	–	40,8	40,5	41,6	41,1	40,5	40,5
Ex 7	48,6	41,0								
In 7	34,7	33,6								
Ex 8	33,4	34,0		–	44,7	44,0	44,5	45,3	43,0	43,7
3'-flank	44,7	37,4		–	46,0	44,0	44,7	45,7	43,7	45,7

Примечания. P_{min} – минимальный промотор, Ex – экзон, In – интрон, 3'flank – 3'-фланкирующая область гена *Xist*. 4 вида полевок: (M.k.) – *Microtus kirgisorum*; (M.a.) – *Microtus arvalis*; (M.r.) – *Microtus rossiaemeridionalis*; (M.t.) – *Microtus transcasicus*.

по сравнению с геном, и можно отметить некий градиент понижения GC-состава от 5'-района к 3'-району (табл. 6).

В экзонах содержание GC выше, чем в интронах для всех видов. Только у грызунов различия в доле GC между экзонами и интронами не превышают 1 %, а у человека и быка составляют 4,9 и 3,8 % соответственно. Следует отметить, что суммарный GC-состав экзонов, несмотря на их почти двукратные различия в размере между видами, оказался весьма консервативным и колеблется в пределах 41,4–42,6 %, тогда как GC-содержание интронов более вариабельно – от 37,2 до 41,9 % (табл. 6).

Количество CpG-островков и их локализация в гене *Xist* отличаются даже у близкородственных видов (рис. 1). Единственный консервативный CpG-островок ассоциирован с блоком тандемных повторов F. Следует отметить, что как правило CpG-островки ассоциированы с видоспецифическими мобильными элементами или тандемными повторами.

Гомология экзонов *Xist* и белок-кодирующего гена *Lnx3*

2 экзона гена *Xist* гомологичны двум экзонам белок-кодирующего гена *Lnx3*, который располагается в ортологичном локусе у амфибий и птиц (Duret *et al.*, 2006). При сравнении консенсуса гена *Xist* и *Lnx3* курицы (ACC XM_420296.1) мы дополнительно обнаружили сходство промоторного района *Xist* с экзонами 1 и 2 *Lnx3* (табл. 4). Более того, в ортологичной позиции в 5'-области у *Lnx3* и *Xist* находится один и тот же фрагмент древнего LINE, относящегося к семейству L3, которое распространено в геномах у большинства известных видов позвоночных. Гомология между мобильным элементом курицы и консенсусом L3 млекопитающих составляет 64,7 %. Кроме того, мы дополнительно обнаружили гомологию экзонов 3 и 5 *Lnx3* с фрагментом интрона 3 и экзонам 6 консенсусного *Xist* (табл. 4).

Обсуждение

Сравнительный анализ гена *Xist* у представителей разных отрядов плацентарных млекопитающих позволил нам сделать предположения о структуре предкового гена. Вероятно, предковый

ген *Xist* состоял из 10 экзонов и имел размер генного локуса ~30 т.п.н. Его основой послужил белок-кодирующий ген *Lnx3*, который в результате интеграции в него мобильных элементов и амплификации тандемных повторов преобразовался в ген *Xist*. Промоторный район и часть экзона 1, а также экзоны 5, 6 и 7 *Xist* являются, соответственно, дивергировавшими останками экзонов 1, 2, 4, 5 и 11 гена *Lnx3*. Экзоны 2, 3, 4, 8, 9 и 10 предкового *Xist*, по-видимому, сформировались из мобильных элементов. Большой вклад в формирование структуры экзонов предкового гена внесли простые тандемные повторы A, C, D, и F, мономерные единицы которых имеют гомологию с мобильными элементами и, возможно, произошли в результате амплификации их фрагментов.

Нуклеотидная последовательность гена *Xist* не консервативна и изменяется очень быстро. Межвидовые сравнения показали, что замены нуклеотидов (транзиции и трансверсии) вносят менее значительный вклад в дивергенцию последовательностей по сравнению с инсерциями и делециями, которые на порядок величин увеличивают изменения одинаковых структурных элементов гена.

Разные части гена *Xist* изменяются с разной скоростью. Интроны гена *Xist* изменяются быстрее, чем экзоны. Степень дивергенции последовательностей разных экзонов неодинакова: экзоны 4, 5 изменяются медленнее, экзоны 3, 8 – быстрее. Самым консервативным является экзон 4. Парадоксально, что имеющий определенные функции экзон 1, и в частности, район A-повтора, изменяется в эволюции быстрее, чем экзон 4, делеция которого не влияет на инактивацию (см. обзор Chang *et al.*, 2006). К наиболее консервативным, медленно изменяющимся элементам гена следует также отнести район минимального промотора. В целом эволюция гена *Xist* носит мозаичный характер и ее направление различно для разных групп млекопитающих.

Характерной особенностью эволюции гена *Xist* являются интеграция транспозонов и амплификация простых тандемных повторов, обуславливающих межвидовые различия гена *Xist* по размеру РНК и наличию в ее составе видоспецифических экзонов, повторов и мобильных элементов. Различия между видами по длине и структуре экзонов гена *Xist* позволяют

предположить, что размер РНК и, следовательно, последовательности мобильных элементов и повторов, за счет которых различается размер транскрипта *Xist* у разных видов, могут быть либо не существенны для функции гена, либо у каждого вида специфическим образом происходит селективная «адаптация» этих последовательностей в составе зрелого транскрипта к условиям функционирования в геноме и X-хромосоме каждого конкретного вида.

Остается неясным значение интронов в структуре гена *Xist*. В белок-кодирующих генах интроны разделяют структурные и/или функциональные домены белка. В регуляторных генах, продуцирующих ядерные РНК, роль интронов и их отличие от экзонов остаются загадкой, так как ни те, ни другие не кодируют белок. Более того, в процессе эволюции *Xist* наблюдается переход экзонов в интроны и интронов в экзоны. Таким образом, структура гена *Xist* не является строго консервативной и эволюционно стабильной, что, возможно, связано с видоспецифическими особенностями его функционирования.

Несмотря на разительные эволюционные преобразования структуры гена *Xist* у разных видов, GC-состав экзонов на уровне 42 % является одной из консервативных характеристик гена у млекопитающих и, возможно, поддерживается отбором. Интроны содержат меньше GC-пар, чем экзоны, и проявляют значительные различия у представителей разных отрядов. В то же время локализация и структура CpG-островков в гене *Xist* различаются у сравниваемых видов. Выявленные CpG-островки в основном ассоциированы с *SINE*- и *LINE*-повторами. Отсутствие консервативности в локализации CpG-островков и их дивергенция при сходной локализации в гене *Xist* у разных видов могут свидетельствовать о видоспецифических механизмах регуляции его активности.

Характерной особенностью гена *Xist* является наличие в составе процессированного транскрипта древних и молодых (видоспецифических) мобильных элементов. Сходную картину можно наблюдать у других, не кодирующих белок, генов ХИС. Так, у гена *Enox* (*Jpx*) в состав процессированного транскрипта могут входить *SINE*, *LINE* и даже процессированные псевдогены, как, например, у мыши (Johnston *et al.*, 2002). По-видимому, интеграция мобильных элементов в

экзоны сопровождает всю эволюцию этих генов и продолжается у представителей современных Eutheria. Возможно, это является общим механизмом формирования и преобразования регуляторных генов, кодирующих ядерные РНК.

Работа поддержана грантом РФФИ № 06-04-48337.

Литература

- Altschul S.F., Gish W., Miller W. *et al.* Basic local alignment search tool // *J. Mol. Biol.* 1990. V. 215. P. 403–410.
- Benson G. Tandem repeats finder: a program to analyze DNA sequences // *Nucl. Acids Res.* 1999. V. 27. P. 573–580.
- Brockdorff N., Ashworth A., Kay G.F. *et al.* The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus // *Cell.* 1992. V. 71. P. 515–526.
- Brown C.J., Hendrich B.D., Rupert J.L. *et al.* The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus // *Cell.* 1992. V. 71. P. 527–542.
- Chang S.C., Tucker T., Thorogood N.P., Brown C.J. Mechanisms of X-chromosome inactivation // *Front. Biosci.* 2006. V. 11. P. 852–866.
- Chureau C., Prissette M., Bourdet A. *et al.* Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine // *Genome Res.* 2002. V. 12. P. 894–908.
- Duret L., Chureau C., Samain S. *et al.* The *Xist* RNA gene evolved in eutherians by pseudogenization of a protein-coding gene // *Science.* 2006. V. 312. P. 1653–1655.
- Jeanmougin F., Thompson J.D., Gouy M. *et al.* Multiple sequence alignment with Clustal X // *Trends Biochem. Sci.* 1998. V. 23. P. 403–405.
- Johnston C.M., Newall A.E., Brockdorff N., Nesterova T.B. *Enox*, a novel gene that maps 10 kb upstream of *Xist* and partially escapes X inactivation // *Genomics.* 2002. V. 80. P. 236–44.
- Nesterova T.B., Slobodyanyuk S.Y., Elisaphenko E.A. *et al.* Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence // *Genome Res.* 2001. V. 11. P. 833–849.
- Pearson W.R., Lipman D.J. Improved tools for biological sequence comparison // *Proc. Natl Acad. Sci. USA.* 1988. V. 85. P. 2444–2448.
- Wutz A., Rasmussen T.P., Jaenisch R. Chromosomal silencing and localization are mediated by different domains of *Xist* RNA // *Nature Genet.* 2002. V. 30. P. 167–174.

RECONSTRUCTION OF THE ANCESTRAL *Xist* GENE AND ITS EVOLUTIONARY PATH IN MAMMALS

E.A. Elisaphenko, N.N. Kolesnikov, A.I. Shevchenko, S.M. Zakian

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: zakian@bionet.nsc.ru

Summary

Xist is a regulatory gene producing nuclear RNA which play a key role in X-chromosome inactivation process in female eutherians. In the study we undertook comparative analysis of the *Xist* gene in 10 eutherian species. We showed that *Xist* evolved from a protein-coding gene, which due to expansion of mobile elements and tandem repeats has transformed to a gene with a large untranslated RNA controlling expression of whole chromosome. During eutherian evolution *Xist* accumulated species-specific features, which, presumably, adopted *Xist* RNA to the functioning in the genome and X-chromosome of each particular species.