

ТАНДЕМНЫЕ ПОВТОРЫ ГЕНОМОВ МЫШЕВИДНЫХ ГРЫЗУНОВ В БАЗАХ ДАННЫХ И ИХ КАРТИРОВАНИЕ

© Д. И. Остромышенский,¹ И. С. Кузнецова,¹ А. С. Комиссаров,¹
И. В. Картавецца,² О. И. Подгорная¹

¹ Институт цитологии РАН, Санкт-Петербург,

и ² Биолого-почвенный институт ДВО РАН, Владивосток; электронный адрес: necroforus@gmail.com

Тандемно-повторяющиеся последовательности представляют собой уникальный для эукариот класс ДНК. Их содержание в геноме у высших эукариот может достигать десятков процентов. Однако вопросы эволюции этого класса последовательностей остаются малоисследованными. Ранее методами биоинформатики выявлены 62 семейства тандемных повторов (ТП) *Mus musculus*; в настоящей работе 7 из них картировано на метафазных пластинах с помощью флуоресцентной гибридизации in situ (FISH). Те же пробы использовали для двух других видов рода *Mus* — *M. caroli* и *M. spicilegus*. Показано, что наборы ТП совпадают только у близкородственных видов мышей. Но даже у таких видов наблюдаются различия в локализации на хромосомах и интенсивности метки отдельных ТП. Только отдельные семейства ТП остаются общими при увеличении эволюционного расстояния между видами. Применение сочетания биоинформатических и молекулярно-биологических методов для исследования эволюции ТП перспективно, но пока ограничено наличием прочитанных геномов близких видов.

Ключевые слова: тандемные повторы, сателлитные ДНК, грызуны *Mus*.

Принятые сокращения: периЦЕН — перичентромерный район хромосом, сатДНК — сателлитные ДНК, п. н. — пара нуклеотидов, ТП — тандемные повторы, ЦЕН — центромерный район хромосом, SRA — sequence read archive, WGS — полногеномная сборка.

Тандемно-повторяющиеся последовательности представляют собой уникальный для эукариот класс ДНК. Уникальность свойств тандемных повторов (ТП) — следствие их структурной организации: они состоят из многократных тандемно-повторяющихся (голова-к-хвосту) коротких последовательностей (мономеров). В результате поле тандемных повторов приобретает способность к нетривиальной укладке хроматина (Vogt, 1990). ТП есть только у эукариотических организмов. Их содержание в геноме у высших эукариот может достигать десятков процентов. Функциональное значение наличия ТП в геномах эукариот остается невыясненным. Причиной этого является ограниченный арсенал методов их изучения. Первым из представленных в геноме типов ТП описаны и клонированы сателлитные ДНК (сатДНК) (Kit, 1961). Позднее определили их локализацию в ключевых участках хромосом. ТП являются основными компонентами структурно значимых частей хромосом, таких как субтеломеры, центромеры и гетерохроматиновые районы. Предполагается, что огромные массивы ТП являются идеальным субстратом для негомологичных рекомбинаций, способствуя перестановкам, таким как робертс-новские транслокации (Garagna et al., 2001; Kalitsis et al., 2006). СатДНК (тандемные повторы) — максимально изменчивый компонент генома, вплоть до возможной видоспецифичности. Однако в догеномную эру при наличии ограниченного количества клонированных сатДНК

(ТП) невозможно было определить наличие набора ТП у организмов даже одного рода.

Анализ прочитанной части генома дал колоссальный толчок развитию биологии и активно продолжается. Создаются базы данных, содержащие описания отдельных элементов (проект ENCODE). Но ТП остаются за пределами подобных проектов, главным образом по причине отсутствия проработанных подходов к их выявлению и классификации. Из большей части фрагментов с ТП не удастся собрать достаточно длинные контиги и правильно расположить их на хромосомах. Именно поэтому на каждой хромосоме в большинстве собранных геномов для центромерного района оставлен пустой участок размером 3 Mb (Golden Path Gap).

Исторически исходный термин сатДНК недостаточно формализован для работы с базами данных. Мы будем называть тандемными повторами (ТП) большие ТП, максимально сходные с классическими сатДНК. ТП — это фрагмент генома с 1) тандемно повторенными без вставок мономерами, 2) размером мономера не более 2 тыс. п. н. и 3) размером полей в базах не менее 3 тыс. п. н. Такие формальные характеристики позволяют отсеять мини- и микросателлиты.

Удобным объектом для изучения ТП и их эволюции является род *Mus*. Геном *M. musculus* отсекален в 2003 г. двумя независимыми группами (Celera и Mouse Genome Sequencing Consortium), сравнительно хорошо

собранный, и сборка постоянно совершенствуется, есть базы данных некартированных контигов. Анализ ТП в полногеномной сборке генома мыши (WGS) выявил 8 семейств ТП, состоящих из 62 подсемейств, из которых только 2 подсемейства описаны ранее как клонированные сатДНК. Для многих из 60 семейств ТП предсказано положение на хромосомах в эталонном геноме мыши (Komissarov et al., 2011).

Исследованные в нашей работе животные принадлежат к классу Rodentia, надсемейству Muroidea, семейству Muridae, подсемейству Murinae. Роды *Mus*, *Apodemus* и *Sylvaemus* образуют единую кладу, которая разошлась с той, к которой принадлежит род *Rattus*, около 12 млн лет назад. Время разделения предков *Mus* и *Apodemus* и *Sylvaemus* составляет примерно 9.7 млн лет (Michaux et al., 2002). Клада, к которой принадлежит *M. caroli*, является базальной для подрода *Mus* s. str и объединяет восточноазиатских представителей подрода. В ней выделяются три клады несинантропных палеарктических видов *M. spretus*, *M. spicilegus* и *M. macedonicus*, а также синантропный космополитичный вид *M. musculus*. Время дивергенции между *M. musculus* и *M. spicilegus*, согласно данным большинства исследований, в зависимости от метода оценки составляет от 1.1 до 1.7 млн лет (Chevret et al., 2005; Macholan et al., 2007). Иногда время дивергенции между двумя этими видами оценивается как более далекое, вплоть до 2.4 млн лет (Liu et al., 2008). Время дивергенции между восточноазиатской кладой, к которой принадлежит *M. caroli*, и кладой, объединяющей остальных представителей подрода, вычисленное на основании анализа последовательностей митохондриальных ДНК, составляет приблизительно 6 млн лет (Chevret et al., 2005).

В настоящей работе для сравнения набора ТП с *M. musculus* привлечены данные неполного и несобранного генома *M. caroli*. Экспериментальная проверка показала, что пробы, рассчитанные по предсказанным мономерам ТП, пригодны для их картирования методом FISH, и подтвердила, что пробы к найденным в базе данных генома *M. caroli* ТП дают гибридационный сигнал на хромосомах этого вида.

Материал и методика

Объекты исследования: 1) *M. musculus*: линейные мыши, синантропный подвид *M. m. domesticus* (Санкт-Петербург, 2 особи); 2) *M. spicilegus* (Ростовская обл., 5 особей); 3) *M. caroli* (клеточная культура эмбриональных фибробластов с нормальным кариотипом, Вьетнам). Культуры фибробластов выращивали в стандартных условиях — в атмосфере 5 % CO₂ при 37 °C на среде DMEM, содержащей 10 % бычьей эмбриональной сыворотки (Sigma-Aldrich, США).

Препараты метафазных хромосом готовили прямым способом из костного мозга животных и культур клеток фибробластов по общепринятой методике (Макгрегор, Варли, 1986).

Пробы последовательностей сателлитной ДНК *M. musculus*. Использовали клонированный в векторе pBluescript II KS+ фрагмент мажорного сателлита мыши (471 п. н.) и клонированный в векторе pGEM7 фрагмент минорного сателлита мыши (362 п. н.), а также фрагменты MS3 и MS4, клонированные в векторе pUC19 по сайтам BamHI и EcoRI. Для гибридизации in situ и вставки MaCat, MiCat, MS3 и MS4 метили с помощью

прямого и обратного праймеров фага M13 (*M13R*: 5'-GGAAACAGCTATGACCATG-3'; *M13F*: 5'-GTAAA-ACGACGGCCAGT-3') методом ПЦР. Реакционная смесь содержала буфер для Taq-полимеразы (50 mM MgCl₂), 5 нг/мкл плазмидных векторов, по 10 pM олигонуклеотидных праймеров, по 2 mM dATP, dCTP, dGTP и dTTP, 5 ед. Taq-полимеразы (СибЭнзим, Россия), а также биотин-16-dUTP.

Центромерная проба *M. musculus* получена микродиссекцией центромерных и перичентромерных районов (и любезно предоставлена В. А. Трифионовым, Институт цитологии и генетики СО РАН, Новосибирск) нескольких хромосом с последующей амплификацией ДНК с использованием дор-праймера (5'-CCGACTCGAGNNNNNNA-TGTGG-3') (Arneson et al., 2008). Для гибридизации in situ полученный ПЦР-продукт амплифицировали с использованием дор-праймера в присутствии биотин-16-dUTP.

Гибридизация in situ (FISH) двухцепочечных проб. На препараты наносили гибридационный буфер (2-кратный SSC, 70%-ный формамид, 1-кратный декстран), содержащий меченый фрагмент ДНК (2 мкг/мл). Денатурацию проб и препаратов проводили одновременно в течение 4 мин при 72 °C. Гибридизацию проводили во влажной камере при 37 °C в течение ночи. После гибридизации препараты последовательно отмывали при температуре гибридизации в 50%-ном формамиде на 4-кратном SSC в течение 5 мин, затем последовательно в 4- и 2-кратном растворах SSC и трех сменах буфера КСМ (135 mM KCl, 20 mM NaCl, 0.5 mM EDTA, 0.1 % Тритона X-100 и 10 mM Tris-HCl, pH 7.5) при комнатной температуре. Зонд выявляли с помощью стрептавидина, конъюгированного с Cy-3, FITC (Sigma-Aldrich, США) или Alexa 488 (Invitrogen, США). Препараты окрашивали DAPI и заключали в среду VectaShield (Dabco, США).

Гибридизация одноцепочечных проб. Для гибридизации использовали синтезированные олигонуклеотиды (Beagle, Россия), модифицированные биотином по 5'- и 3'-концам. Препараты метафазных хромосом денатурировали в буфере для денатурации (70%-ный формамид и 2-кратный SSC) в течение 2 мин при 65 °C и затем дегидратировали в серии ледяных растворов этанола (70, 90 и 96 %). Препараты высушивали и на них наносили гибридационный буфер (25 % формамида и 4-кратный SSC), содержащий 5 мкг/мл олигонуклеотида. Гибридизацию проводили во влажной камере при 37 °C в течение ночи. Зонд детектировали стрептавидином, конъюгированным с Alexa 488. Препараты окрашивали DAPI и заключали в среду VectaShield. Препараты анализировали на флуоресцентном микроскопе Axioskop-DFC360, Leica DM 6000 B или конфокальном сканирующем микроскопе LEICA TCS SP5.

Компьютерные методы. Содержание ТП в геномах *M. musculus* и *M. caroli* оценивали следующим образом. Для каждого тандемного повтора *M. musculus* брали все найденные в контигах базы данных WGS поля (Komissarov et al., 2011) и с помощью программы bowtie2 (Langmead, Salzberg, 2012) с параметром local выравнивали на эти поля по 4 рана (run) полногеномного секвенирования для обоих видов. Раны полногеномного секвенирования брали из базы DDBJ sequence read archive (SRA) (http://trace.ddbj.nig.ac.jp/dra/index_e.html) для *M. musculus* (ERR008466, ERR008446, ERR008560 и ERR019225) и *M. caroli* (ERR019944, ERR019945, ERR019949 и ERR019951). Содержание в геноме оценивали как отношение количе-

Таблица 1

Последовательности зондов к ТП, использованные в работе

Зонд	Последовательность 5'-3'
TR-24B	GTGCTACAGGTACAGCTGTGCTCT
TR-31A	ATCCCCCTGCCTCAGAAAATCACTCACAGTC
TR-31B	TCACATCCTCTGCTGTGTGTGTGTCAGGGCC
TR-38C	GCCATGGTGACACTCCCCAGAGTTCTGCTGTAAGGCT
TR-40A	CTCCCAGCCTGTGCTCCCAGCCTGTGCTCCCAGCCTGTG
Sat60	GTATATGAGTGAGTTGCACTGAA

ства выровнявшихся более одного раза на поле ридов (reads) к общему числу ридов в ране. Для оценки содержания ТП *M. caroli* брали 4 раза тандемно-повторенный номер повтора и анализировали так же, как поля ТП *M. musculus*.

Пробы для предсказанных *in silico* ТП. Для гибридизации *in situ* подобрали короткие одноцепочечные олигонуклеотидные зонды к каждому из исследуемых подсемейств ТП. Подбор последовательностей осуществляли программой на языке Python, написанной А. С. Комиссаровым по следующей методике: 1) из полногеномной сборки брали все поля, принадлежащие к подсемейству ТП; 2) для этих полей рассчитывали частоты всех возможных *k*-мер с длиной слова (*k*), равной 11; 3) для создания пробы брали наиболее часто встречающиеся *k*-мер; 4) для подбора наиболее оптимальных праймеров увеличивали *k* до 30—60. Контроль качества нуклеотидов (например, отсутствие вторичной структуры) осуществляли с помощью программы Primer 3. Пробы подбирали для максимально гомогенного поля, чтобы охватить все возможные поля ТП. Последовательности зондов приведены в табл. 1.

Результаты

Разрешение FISH не дает возможности определить точную локализацию пробы в области первичной перетяжки — центромерную (ЦЕН) или перицентромерную (периЦЕН). Поэтому для большинства проб мы будем описывать локализацию пробы как ЦЕН в широком смысле, т. е. область первичной перетяжки хромосомы без уточнения позиции. Исключение составляют случаи гибридизаций с зондами к последовательностям, позиция которых на хромосомах хорошо изучена (Kuznetsova et al., 2006).

Проба, полученная в результате диссекции участка хромосомы *M. musculus*, содержит как ЦЕН- так и периЦЕН-районы и все входящие в них последовательности. Поэтому она целиком покрывает область первичной перетяжки на хромосомах *M. musculus*. Количество и размер сигнала на хромосомах варьируют. На хромосомах *M. spicilegus* наблюдаются меньшая степень гибридизации в районе ЦЕН, а также полное отсутствие гибридизационного сигнала на Y-хромосоме (рис. 1, б). На хромосомах *M. caroli* гибридизационный сигнал еще слабее,

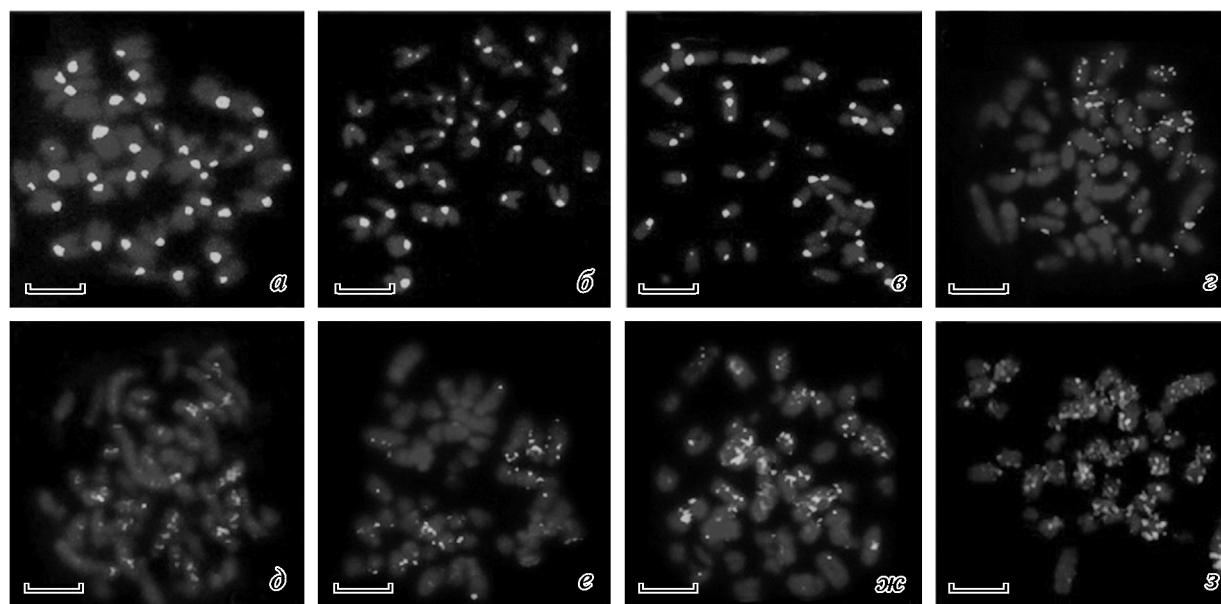


Рис. 1. Локализация пробы панЦЕН *Mus musculus* (светлые пятна) на препаратах метафазных хромосом 8 видов подсемейства мышевидных грызунов Murinae.

a — *M. musculus*, *б* — *M. spicilegus*, *в* — *M. caroli*, *г* — *Sylvaemus uralensis*, *д* — *S. flavicollis*, *е* — *Apodemus agraris*, *ж* — *A. penninslae*, *з* — *Rattus norvegicus coraco*. Хромосомы контрастированы DAPI. Масштабные отрезки — 10 мкм.

Таблица 2

Тандемные повторы *Rattus norvegicus* в базе данных Repbase

Семейство	Длина мономера, п. н.	Наличие в геноме <i>Mus musculus</i>	Комментарий
MurSAT1	120	+	Muridae-специфичный
IMPB02	2171	+	TE-related
91ES81_RN	196	+	
ALSAT_RN	183	—	Alpha, мономер
SAT1_RN	370	—	Alpha, димер
ZP3AR	212	+	TE-related

чем у *M. spicilegus*. У представителей других родов наблюдаются гибридизационный сигнал по плечам хромосом и полное отсутствие гибридизации в районе ЦЕН. На хромосомных пластинках рода *Mus* (рис. 1, а—в) наименьшее количество сигнала наблюдается на половых хромосомах. Таким образом, у представителей рода *Mus* центромерная проба *M. musculus* сохраняет локализацию в области ЦЕН, однако у видов родов *Sylvaemus*, *Apodemus* и *Rattus* панцентромерная проба локализована по плечам хромосом, а не в области ЦЕН.

Несмотря на то что род *Rattus* (крысы) и род *Mus* (мыши) разошлись давно, гибридизация панцентромерной пробы *M. musculus* на хромосомах *R. norvegicus* дала весьма неожиданный результат (рис. 1, з). Гибридизационный сигнал виден по плечам хромосом и, пусть в незначительной мере, но и в области ЦЕН некоторых хромосом. Мы решили проверить наличие в геноме *M. musculus* последовательностей, сходных с ТП *R. norvegicus*.

Для поиска известных сатДНК крысы мы использовали базу данных повторов Repbase, которая для крысы содержит 11 различных сатДНК. 2 из них являются общими для млекопитающих (SVA, основанный на Alu-элементах, и TAR1, связанный с теломерным районом). 3 являются общими для Muridae — MurSAT1, IMPB_02 и ZP3AR. Два последних оказались основанными на фрагментах транспозонов (TE-related). Для крысы известен и хорошо описан центромерный повтор, более похожий на альфа-сателлитную ДНК человека, чем на мажорный (MaCat) или минорный (MiCat) сателлит мыши (Whitney et al., 1983). Ему в Repbase соответствуют два повтора — ALSAT_RN (183 bp, мономер) и SAT1_RN (370 bp, димер). Этого повтора в геноме *M. musculus* не обнаружили.

Из ранее считающихся крысиными сатДНК в геноме мыши найдены RNSAT1a, RNSAT1b и RNSAT1 с длиной мономера 168 bp и 91ES8_RN с мономером 196 bp. Таким образом, последовательности крысы, сходные с 4 ТП, присутствуют в геноме *M. musculus* (табл. 2). Эти данные объясняют картину гибридизации панцентромерной пробы *M. musculus* на хромосомах *R. norvegicus*.

На следующем этапе работы мы сравнивали ТП видов внутри рода *Mus*. В нашей лаборатории с помощью компьютерного анализа генома *M. musculus* выявлено 8 семейств ТП, подразделяемых на 62 подсемейства (Komissarov et al., 2011). Среди 62 подсемейств только 2 совпадают с известными клонированными ТП — MaCat (~76 % от найденных полей ТП) и MiCat (~2 %). Остальные 60 подсемейств не представлены в базе данных повторов Repbase. Для новых ТП предложена следующая классификация. Класс больших ТП разделили на четыре суперсемейства: 1) ЦЕН ТП (MiCat); 2) периЦЕН ТП (MaCat); 3) гетерогенные ТП; 4) ТП, родственные диспергированным повторам. Гетерогенные ТП по положению в эталонном геноме разделены на семейства: 1) мультилокусные (в эталонном геноме найдены более чем в одном локусе); 2) однолокусные (в эталонном геноме найдены только в одном локусе); 3) некартированные (не найдены в эталонном геноме, есть только в базе WGS). Для проб, использованных в настоящей работе, указаны принадлежность к семейству и другие характеристики (табл. 3).

Геном *M. caroli* прочитан, однако не собран даже до контигов. Тем не менее возможна проверка того, присутствуют ли в геноме этого вида те же ТП, что и у *M. musculus*. Анализ генома *M. caroli* из SRA показал, что наибольшее

Таблица 3

ТП, выявленные *in silico* и использованные для расчета олигопроб в настоящей работе

Подсемейство	Длина мономера, п. н.	Семейство	Позиция в эталонном геноме	Число полей в Chromosome Unknown
TR-31A	31	Мультилокусное	7D1, 8C1	3
TR-31B	31	»	7D1, 14A2	2
TR-38C	3824	»	12B1, 13A3.2	0
TR-40C	40	Однолокусное	15A2	0
TR-24B		Некартированное	—	1

Примечание. Позиция в эталонном геноме указана с точностью до хромобенда.

Таблица 4

Доля ТП (%) в геномах *Mus musculus* и *M. caroli* из базы SRA

Семейство	<i>M. musculus</i>	<i>M. caroli</i>
Не ТП	83.6073	98.8742
MaCat	13.5587	0.6627
TR-1590A-MM	2.2426	0.0356
TR-6A-MM	0.2513	0.1548
TR-31B-MM	0.2155	0.1917
TR-107A-MM	0.0392	0.0370
TR-21A-MM	0.0445	0
TR-194A-MM	0.0013	0
TR-57A-MM	0.0038	0.0440
TR-13A-MM	0.0005	0
TR-93A-MM	0.0207	0
TR-30A-MM	0.0004	0
TR-31A-MM	0.0007	0
TR-1149A-MM	0.0011	0
TR-22A-MM	0.0026	0
TR-1526A-MM	0.0004	0
TR-24B-MM	0.0022	0
TR-27A-MM	0.0007	0

Примечание. Полу жирным выделены ТП, картированные экспериментально. Долю ТП меньше 0.0001 % не определяли.

лее представленным из всех исследованных ТП является MaCat (~1 % от всех последовательностей), вторым по количеству в геноме является ТП TR-31B (~0.2 %). Количество другого (TR-31A) в геноме *M. caroli* меньше $1 \cdot 10^{-4}$ %. Из других ТП в геноме *M. caroli* найдены короткий, с длиной мономера 6 н. п. ТП TR-6A, ТП из семейства, основанного на фрагментах диспергированных повторов, TR-1590-A и два ТП из семейства гетерогенных — TR-57A и TR-107-A (табл. 4).

В догеномную эру экспериментально клонированы 4 сатДНК (ТП) *M. musculus* и их положение в области ЦЕН или периЦЕН определено (Kuznetsova et al., 2006). При использовании клонированных проб ТП *M. musculus* с метафазными хромосомами *M. caroli* ЦЕН ТП (МиСат и MS3) не дают сигнала, отличного от шума. Пробы периЦЕН ТП дают хороший сигнал на метафазных хромосомах *M. caroli*. Сигнал MaCat локализован в периЦЕН-участках практически всех хромосом и по плечам 2 хромосом, а сигнал пробы MS4 есть как в периЦЕН, так и по плечам некоторых хромосом (рис. 2).

Клонированные пробы давно используют для маркирования ЦЕН хромосом мыши (Kuznetsova et al., 2006). Для того чтобы использовать ТП, выявленные *in silico*, необходимо проверить, работают ли олигопробы, рассчитанные по предсказанным последовательностям ТП. Геном *M. caroli* содержит лишь незначительное количество ДНК, гомологичной MaCat (Siracusa et al., 1983). Две сатДНК длиной 60 (Sat60) и 79 п. н. обнаружены у *M. caroli* (Kipling et al., 1995). Sat60 обнаружен в периЦЕН на всех аутосомах и Y-хромосоме, в то время как сателлит размером 79 п. н. — в ЦЕН всех аутосом и на X-хромосоме. В составе последнего обнаружен сайт связывания с белком CENP-B (CENP-B-бокс). Обе сателлитные последовательности образуют гомогенные поля длиной более

1 млн п. н. (Kipling et al., 1995). На основании известного мономера ТП в 60 п. н. *M. caroli* синтезировали олигонуклеотиды, как описано (табл. 1).

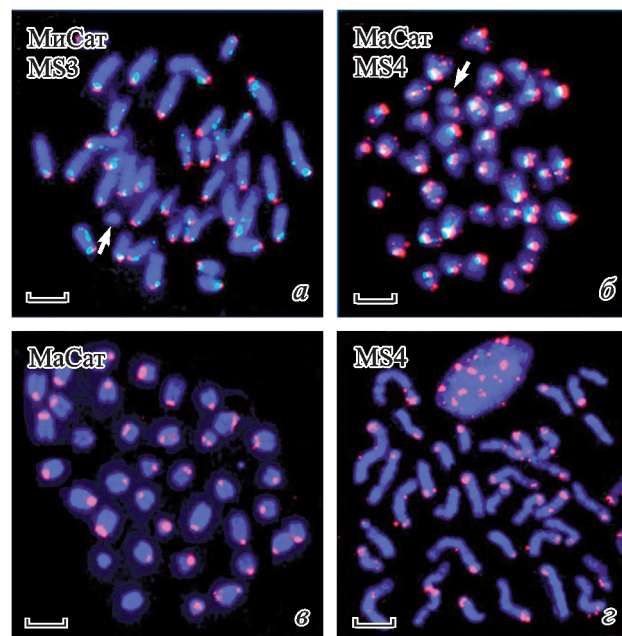


Рис. 2. Локализация проб сатДНК (ТП) *Mus musculus* на хромосомах *M. spicilegus* (а, б) и *M. caroli* (в, з).

а — МиСат (красный цвет) и MS3 (зеленый); б — MaCat (красный) и MS4 (зеленый); в, з — MaCat и MS4 соответственно (красный). Хромосомы контрастированы DAPI (синий). Стрелками показаны Y-хромосомы. Масштабные отрезки — 10 мкм.

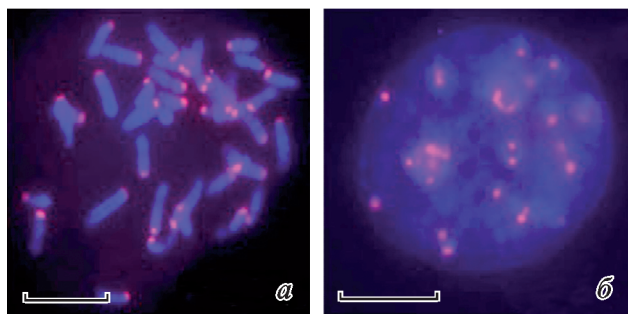


Рис. 3. Гибридизация пробы Sat60 с хромосомами (а) и ядрами (б) *M. caroli*.

Хромосомы контрастированы DAPI (синий). Масштабные отрезки — 10 мкм.

Локализация олигопробы сатДНК *M. caroli* Sat60 совпадает с известной по данным из литературы. Сигнал зонда локализован в области первичной перетяжки всех хромосом *M. caroli* (рис. 3, а). В интерфазных ядрах гибридационный сигнал присутствует только в хромоцентрах, однако не покрывает их целиком (рис. 3, б). На хромосомах *M. musculus* эта проба сигнала не дает.

Гибридизация олигонуклеотидного зонда к семейству TR-31А (рис. 4) дает сигнал на 11 хромосомах *M. musculus*. Для 3 из них характерна субтеломерная локализация сигнала со стороны длинного плеча хромосом. Еще на 4 хромосомах сигнал тоже находится в субтеломерной области, однако кроме того, зонд дает сигнал и в области периЦЕН. На 2 хромосомах сигнал есть в периЦЕН и по плечам хромосом, а на 2 хромосомах сигнал находится в G-позитивных районах по плечам хромосом. На метафаз-

ных пластинках *M. spicilegus* (рис. 4) сильный сигнал есть на 4—5 парах хромосом. Наиболее сильный сигнал принадлежит субтеломерному району 2 пар хромосом. На остальных 2—3 парах есть слабый субтеломерный сигнал, а на нескольких хромосомах — слабый сигнал на плечах хромосом. На метафазных пластинках *M. caroli* (рис. 4) зонд хорошо виден в субтеломерных районах 2 пар хромосом, еще на нескольких хромосомах можно видеть слабый сигнал на плечах хромосом.

При гибридизации олигонуклеотидного зонда к семейству TR-31В (рис. 4) сигнал обнаружен на 13 парах аутосом и X-хромосоме *M. musculus*. На 3 парах аутосом и на X-хромосоме зонд обнаружен в субтеломерных районах большого плеча хромосом. На 2 парах хромосом сигнал расположен в субтеломерном районе и по плечам хромосом, на 1 паре — в субтеломерном районе, по плечам хромосом и в области периЦЕН. На 2 парах сигнал локализован в периЦЕН и по плечам хромосом, на 2 парах — в периЦЕН, а на 3 парах хромосом — по плечам хромосом. На хромосомах *M. spicilegus* сигнал виден на 3 парах аутосом в периЦЕН и на Y-хромосоме (рис. 4). На хромосомах *M. caroli* сигнал присутствует на 3 парах хромосом в основном по большому плечу хромосомы (рис. 4).

Зонд к семейству TR-38С дает гибридационный сигнал на 10 парах аутосом и на половых хромосомах *M. musculus*. На 5 парах аутосом и на X-хромосоме гибридационный сигнал обнаружен в субтеломерной области. На 4 парах аутосом и на Y-хромосоме зонд локализуется на большом плече хромосомы. На первой паре хромосом гибридационный сигнал обнаружен в периЦЕН-районе и по плечам хромосом (рис. 4). На хромосомах *M. spicilegus* сигнал стабильно присутствует в периЦЕН 1 пары аутосом и на Y-хромосоме (рис. 4). На отдельных плас-

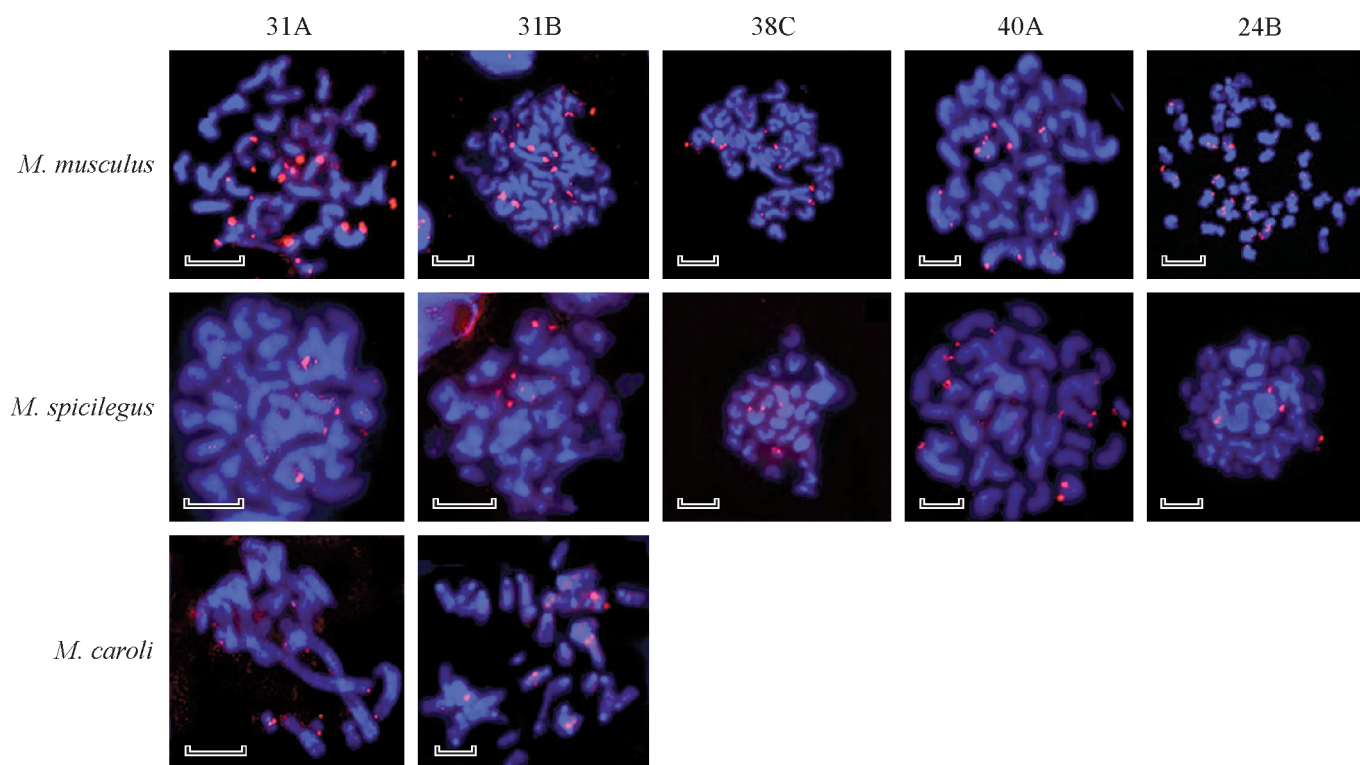


Рис. 4. Гибридизация олигопроб на метафазных пластинках видов рода *Mus*.

Вид указан слева. ТП, на основании которого получены олигопробы, указан над каждой панелью. Хромосомы контрастированы DAPI. Масштабные отрезки — 10 мкм.

тинках сигнал также виден на 1—2 парах хромосом, в основном по плечам. На хромосомах *M. caroli* зонда не обнаружили.

В случае зонда к семейству TR-40С (рис. 4) гибридационный сигнал обнаружен на 10 парах аутосом и X-хромосоме *M. musculus*. На 5 парах аутосом гибридационный сигнал есть только по плечам хромосом, а на 1 паре — по плечам хромосом и в периЦЕН. На 2 парах аутосом сигнал локализован в периЦЕН. На 1 паре хромосом наблюдается субтеломерная локализация сигнала, еще на 1 паре кроме субтеломерной области зонд дает сигнал по плечам хромосом. На X-хромосоме зонд присутствует в периЦЕН, в субтеломерном районах и по плечам хромосом. На хромосомах *M. spicilegus* (рис. 4) сигнал виден в периЦЕН 1 аутосомы и, по всей видимости, на X-хромосоме. На хромосомах *M. caroli* гибридационного сигнала не обнаружено.

При гибридизации зонда к семейству ТП TR-40А (рис. 4) субтеломерная локализация стабильна для 2 пар хромосом *M. musculus*; на некоторых пластинках к ним добавляется еще 1 пара хромосом. На 1 паре хромосом есть сигнал по плечам хромосом. На отдельных пластинках есть периЦЕН-сигнал на 1 хромосоме; по-видимому, это X-хромосома. На хромосомах *M. spicilegus* (рис. 4) на 1—2 парах хромосом обнаружена только субтеломерная локализация сигнала. На отдельных пластинках на 1—2 парах хромосом есть гибридизация по плечам хромосом, а на 1 хромосоме (видимо, X) — периЦЕН-локализация сигнала. На хромосомах *M. caroli* гибридационного сигнала не обнаружено.

Обсуждение

В 2013 г. опубликованы результаты масштабного исследования (Melters et al., 2013), главная цель которого состоит в поиске мажорного по содержанию в геноме ТП для каждого вида с прочитанным геномом. К сожалению, авторы определили мажорный ТП исходя из противоречащих друг другу допущений. Они решили, что мажорным будут считать ТП, во-первых, содержащийся в геноме в максимальном количестве, а во-вторых, содержащий CENP-бок, характерный для ЦЕН ТП. То, что известно о клонированных в догеномную эру и картированных ТП, говорит о том, что мажорными являются ТП, расположенные в периЦЕН-районе (~10 % генома), в то время как CENP-бок, содержащий ЦЕН ТП, составляет не более 1 % (Kuznetsova et al., 2006; Komissarov et al., 2011). Противоречивые допущения авторов привели к закономерному результату — ни один из трех экспериментально найденных ЦЕН ТП не попал в их анализ. Тем не менее работа является наиболее исчерпывающим анализом мажорных по содержанию ТП 282 видов организмов, сделанных методами биоинформатики.

Показано, что мажорные ТП сходны только у филогенетически очень близких видов. У более эволюционно далеких видов мажорные ТП фактически не имеют ничего общего, причем это касается не только собственно последовательности мономера ТП, но и его длины, GC-содержания и содержания мажорного ТП в геномной базе данных. Тем не менее внутри некоторых крупных групп организмов, таких как высшие приматы, цихлидовые рыбы и злаки, есть сходство мажорных ТП у эволюционно достаточно далеких видов. Внутри клады приматов содержание GC в мажорном ТП и длина его мономера кор-

релируют с филогенетическими отношениями видов. Те же признаки у злаков не коррелируют с их филогенетическими отношениями. Возможно, эволюция сатДНК (ТП) в разных группах организмов идет разными путями.

Для поиска мажорных ТП (Melters et al., 2013) использовали самые разные источники. Исходным материалом служили первичные риды, полученные после шотган-секвенирования (полногеномного или отдельных хромосом) почти всеми возможными способами: секвенирование по Сэнгеру, Illumina, 454 и мономолекулярное секвенирование в реальном времени (SMRT, Pacific Bioscience). Первичные риды, полученные с помощью Illumina или 454, сначала собирали геномным ассемблером PRICE, а полученные секвенированием по Сэнгеру и SMRT сразу использовали для поиска ТП с помощью программы tandem repeat finder (TRF; Benson, 1999). Таким образом, для поиска ТП не обязательно нужны собранные геномы. Подобный подход, как показано в настоящей работе, можно использовать не только для поиска мажорного ТП, но и для поиска фактически любых ТП, присутствующих в геноме. Так, найден набор ТП в несобранном геноме *M. caroli* (табл. 3).

Возможно, с учетом библиотечной гипотезы (Fry et al., 1977) минорный по содержанию ТП у одного вида становится мажорным у другого, что можно выявить при сравнении полученных *in silico* наборов ТП каждого генома. Библиотечная гипотеза, которая предложена более 30 лет назад (Fry et al., 1977), предполагает, что в геноме предкового вида есть пул последовательностей, из которых у потомков образуются новые сатДНК путем выборочной амплификации, а последовательности, не подвергшиеся амплификации, остаются в низкокопийном состоянии. У филогенетически близких видов теоретически должны обнаруживаться последовательности, которые у одного вида находятся в высококопийном состоянии, т. е. мажорный ТП (сатДНК), а у других в низкокопийном. Экспериментально библиотечная гипотеза проверена на жуках родов *Tribolium* (Mravinac, Pohl, 2007), *Chrysolina* (Palameque et al., 2005) и *Palorus* (Meštrović et al., 1998). Для позвоночных гипотеза не подтверждена. В пользу гипотезы говорит то, что периЦЕН-повтор в 60 п. н. *M. caroli* мы обнаружили в небольшом (не более $5 \cdot 10^{-6}$ %) количестве в геноме *M. musculus*, однако ЦЕН-сателлит размером 79 п. н. доступными методами поиска не обнаруживается вовсе.

Главный вывод исследования Мелтерса и его коллег (Melters et al., 2013) — мажорные ТП в геномах разных видов различаются практически по всем параметрам. В то же время белки, характерные как для ЦЕН, так и для периЦЕН, весьма консервативны. Понятно, что никакого мотива (бок) для связывания этих белков не существует. Что же тогда детерминирует связывание этих белков? Главный кандидат на эту роль — изогнутость ТП. Показано структурно-специфическое, зависящее от изогнутости ДНК взаимодействие белков р68 и SAF-A с ДНК (Lobov et al., 2001; Erukashvily et al., 2005). Предложена модель (Podgornaya et al., 2003), согласно которой в центромерном районе ТП образуют ряд чередующихся по степени изогнутости последовательностей. Специфические структуры укладки участков хроматина, содержащего ТП, основываются на локус-специфичных структурах ТП и, возможно, стабилизируются связыванием ядерных белков. Код, основанный на ТП, получил название «код укладки хроматина», расширив понятие «хроматиновый код» (Vogt, 1990).

Конститутивный гетерохроматин, основанный на ТП, которые являются мажорным компонентом разных геномов, остается самым загадочным компонентом гетерохроматина. Пренебрежительное отношение к нему как к мусорной (junk) ДНК медленно начинает меняться. Вряд ли живые организмы, существующие с постоянным дефицитом энергии, могут позволить себе поддерживать в геноме 3—10 % ненужной ДНК. Уникальный структурный код, который несут ТП, вероятно, необходим для поддержания высоких уровней архитектоники ядра (Подгорная и др., 2009). ТП являются самым быстро эволюционирующим компонентом геномов и при сравнении наборов ТП могут обеспечить видовую характеристику на уровне ДНК.

В настоящей работе сравнили ТП не аннотированного с точки зрения ТП генома (крыса) и несобранного генома (*M. caroli*) с ТП *M. musculus*. Экспериментальная проверка подтвердила предсказания биоинформатики. Определение функциональной роли ТП как никогда актуально после чтения геномов, но для начала необходимо их выявить и классифицировать. Использованный в настоящей работе подход доказал свою перспективность для решения задачи классификации и сравнения ТП, однако пока он ограничен наличием прочитанных геномов близких видов.

Авторы выражают благодарность В. Г. Маликову и Ф. Н. Голенищеву (Зоологический институт РАН, Санкт-Петербург) за отлов и содержание мышей, а также О. Л. Серову (Институт цитологии и генетики СО РАН, Новосибирск) за предоставленные культуры клеток *Mus caroli*.

Работа выполнена при финансовой поддержке программы президиума РАН «Молекулярная и клеточная биология» (№ 1200955639) и Российского фонда фундаментальных исследований (проект 11-04-01700).

Список литературы

- Макгрегор Г., Варли Дж. 1986. Методы работы с хромосомами животных. М.: Мир. 272 с. (Macgregor H. C., Varley J. M. 1983. Working with animal chromosomes. Wiley, Chichester: John Wiley & Sons Ltd. 250 p.)
- Подгорная О. И., Остромышенский Д. И., Кузнецова И. С., Матвеев И. В., Комиссаров А. С. 2009. Парадоксы организации центромера и гетерохроматина. Цитология. 51 (3): 204—211. (Podgornaya O. I., Ostromishensky D. I., Kuznetsova I. S., Matveev I. V., Komissarov A. S. 2009. Heterochromatin and centromere structure paradox. Tsitologiya. 51 (3): 204—211.)
- Arneson N., Hughes S., Houlston R., Done S. 2008. Whole-genome amplification by degenerate oligonucleotide primed PCR (DOP-PCR). Cold Spring Harbor Protocol.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequence. Nucleic Acids Res. 27 : 573—580.
- Chevret P., Veyrunes F., Britton-Davidian J. 2005. Molecular phylogeny of the genus *Mus* (Rodentia: Murinae) based on mitochondrial and nuclear data. Biol. J. Linnean Soc. 84 : 417—427.
- Enukashvily N., Donev R., Sheer D., Podgornaya O. 2005. Satellite DNA binding and cellular localisation of RNA helicase P68. J. Cell Sci. 118 : 611—622.
- Fry K., Salsler W. 1977. Nucleotide sequences of HS- α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. Cell. 12 : 1069—1084.
- Garagna S., Marziliano N., Zuccotti M., Searle J. B., Capanza E., Redi C. A. 2001. Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. PNAS. 98 : 171—175.
- Kalitsis P., Griffiths B., Choo K. H. 2006. Mouse telocentric sequences reveal a high rate of homogenization and possible role in Robertsonian translocation. PNAS. 103 : 8786—8791.
- Kipling D., Mitchell A. R., Masumoto H., Wilson H. E., Nicol L., Cooke H. J. 1995. CENP-B binds a novel centromeric sequence in the Asian mouse *Mus caroli*. Mol. Cell. Biol. 15 : 4009—4020.
- Kit S. 1961. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. J. Mol. Biol. 3 : 711—716.
- Komissarov A. S., Gavrilova E. V., Demin S. J., Ishov A. M., Podgornaya O. I. 2011. Tandemly repeated DNA families in the mouse genome. BMC Genomics. 12 : 531.
- Kuznetsova I., Podgornaya O., Ferguson-Smith M. A. 2006. High-resolution organization of mouse centromeric and pericentromeric DNA. Cytogenet. Genome Res. 112 : 248—255.
- Langmead B., Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods. 9 : 357—359.
- Liu Y. H., Takahashi A., Kitano T., Koide T., Shiroishi T., Moriwaki K., Saitou N. 2008. Mosaic genealogy of the *Mus musculus* genome revealed by 21 nuclear genes from its three subspecies. Genes Genetic Systems. 83 : 77—88.
- Lobov I. B., Tsutsui K., Mitchell A. R., Podgornaya O. I. 2001. Specificity of SAF-A and lamin B binding *in vitro* correlates with the satellite DNA bending state. J. Cell. Biochem. 83 : 218—229.
- Macholán M., Vyskocilová M., Bonhomme F., Krystufek B., Orth A., Vohralík V. 2007. Genetic variation and phylogeography of free-living mouse species (genus *Mus*) in the Balkans and the Middle East. Mol. Ecol. 16 : 4774—4788.
- Melters D. P., Bradnam K. R., Young H. A., Telis N., May M. R., Ruby J. G., Sebra R., Peluso P., Eid J., Rank D., Fernando Garcia J., Derisi J. L., Smith T., Tobias C., Ross-Ibarra J., Korff L., Chan S. W. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 14 : R10.
- Mestrovic N., Plohl M., Mravinac B., Ugarkovic D. 1998. Evolution of satellite DNAs from the genus *Palorus* — experimental evidence for the «library» hypothesis. Mol. Biol. Evol. 5 : 1062—1068.
- Mravinac B., Plohl M. 2007. Satellite DNA junctions identify the potential origin of new repetitive elements in the beetle *Tribolium madens*. Gene. 394 : 45—52.
- Palomeque T., Muñoz-López M., Carrillo J. A., Lorite P. 2005. Characterization and evolutionary dynamics of a complex family of satellite DNA in the leaf beetle *Chrysolina carnifex* (Coleoptera, Chrysomelidae). Chromosome Res. 13 : 795—807.
- Podgornaya O. I., Voronin A. P., Enukashvily N. I., Matveev I. V., Lobov I. B. 2003. Structure-specific DNA-binding proteins as the foundation for three-dimensional chromatin organization. Int. Rev. Cytol. 224 : 227—296.
- Siracusa L. D., Chapman V. M., Bennett K. L., Hastie N. D., Pietras D. F., Rossant J. 1983. Use of repetitive DNA sequences to distinguish *Mus musculus* and *Mus caroli* cells by *in situ* hybridization. J. Embryol. Exp. Morphol. 73 : 163—178.

TANDEM REPEATS IN RODENTS GENOME AND THEIR MAPPING

D. I. Ostromyshenskii,¹ * I. S. Kuznetsova,¹ A. S. Komissarov,¹ I. V. Kartavtseva,² O. I. Podgornaya¹

¹ Institute of Cytology RAS, St. Petersburg,

and ² Institute of Biology and Soil Sciences, Far East Department of RAS, Vladivostok; * e-mail:necroforum@gmail.com

Tandemly-repeated sequences represent a unique class of eukaryotic DNA. Their content in the genome of higher eukaryotes amounts to tens of percents. However, the evolution of this class of sequences is poorly-studied. In our paper, 62 families of *Mus musculus* tandem repeats are analyzed by bioinformatic methods, and 7 of them are analyzed by fluorescence *in situ* hybridization. It is shown that the same tandem repeat sets co-occur only in closely related species of mice. But even in such species we observe differences in localization on the chromosomes and the number of individual tandem repeats. With increasing evolutionary distance only some of the tandem repeat families remain common for different species. It is shown, that the use of a combination of bioinformatics and molecular biology techniques is very perspective for further studies of the evolution of tandem repeats.

Key words: tandem repeats, satellite DNA, mouse, *Mus*.
