

ТАНДЕМНЫЕ ПОВТОРЫ РАЗНЫХ ПОРОД СВИНЬИ *SUS SCROFA DOMESTICUS*

Н. Г. Иванова, В. Н. Стефанова, И. В. Матвеев, О. И. Подгорная

Институт Цитологии РАН

Санкт-Петербург, Тихорецкий пр.4. 194064, Россия

e-mail: vestefan@mail.ru

Ключевые слова: тандемные повторы, транспозоны, TRF (Tandem Repeat Finder), *Sus scrofa*, олигонуклеотидные пробы, флюоресцентная in situ гибридизация (FISH).

Аннотация. Работа посвящена анализу самой малоизученной и вариабельной части генома, а именно гетерохроматину. Конститутивный гетерохроматин состоит в основном из тандемных повторов (ТП), которые ранее назывались сателлитной ДНК. ТП плохо поддаются клонированию и секвенированию, в результате чего недостаточны представлены в любых геномных сборках. В настоящей работе проанализировали геномные сборки четырех различных пород *Sus scrofa* на наличие ТП. Всего в четырех сборках обнаружено 126 семейств ТП, из которых только были 5 внесены в базу данных повторов RepBase, остальные обнаружены впервые. Обнаружили как общие, так и породоспецифичные семейства ТП. Были сконструировали олигонуклеотидные пробы к некоторым из обнаруженных in silico ТП и с помощью FISH картированы на хромосомах свиньи. Обнаружено много полей семейства ТП, мономеры которых образованы из фрагментов SINE, что, на наш взгляд, составляет характерную особенность генома свиньи и объясняет специфичную картину распределения SINE и LINE повторов на хромосомах свиньи известную из литературы.

TANDEM REPEATS OF DIFFERENT BREEDS PIG *SUS SCROFA DOMESTICUS*

N. G. Ivanova, V. N. Stefanova, I. V. Matveev, O. I. Podgorina

Institute of Cytology of the Russian Academy of Science

Tikhoretsky ave. 4 194064 St-Petersburg, Russia

e-mail: vestefan@mail.ru

Key word: tandem repeats, transposable elements, TRF (Tandem Repeat Finder), *Sus scrofa*, oligo probes, fluorescent in situ hybridization (FISH).

Summary. Our work is dedicated to the analysis of poorly investigated and variable part of the genome – constitutive heterochromatin. Constitutive heterochromatin regions enriched with tandem repeats (TR), initially named satellite DNA. The cloning and sequencing of TR is limited in all sequencing methods that's why TR are under-represented in all genomes' assemblies. Here we report the findings and classifications of TR in genome assemblies of 4 *Sus scrofa* breeds. In 4 assemblies

126 TR families are found, only 5 of them are available from the RepBase repeats database as experimentally cloned, other families are newly found. TR families dataset consists of common and breed-specific TR. Many TR families arrays of Sus scrofa possess monomeres based on SINE fragments, which seems to be specific feature of porcine genome. Results obtained explain strange distribution pattern of LINE and SINE retroelements in porcine chromosomes known from the literature data. Oligonucleotide probes for some of TRs detected in silico were composed and localized on pig chromosomes by fluorescent in situ hybridization (FISH) method.

(Поступила в редакцию 24.06.2016 г.)

Введение. Тандемно повторяющиеся последовательности являются особым классом ДНК эукариот: они состоят из многократно повторяющихся («голова-к-хвосту») коротких последовательностей (мономеров). Их содержание в геноме у высших эукариот может достигать десятков процентов, но функциональное значение тандемных повторов в геномах эукариот остается не выясненным. Тандемные повторы являются основными компонентами структурно значимых частей хромосом, таких как теломеры, центромеры и гетерохроматиновые районы. Первым из представленных в геноме типов тандемных повторов описаны и клонированы сателлитные ДНК (СатДНК) [1]. СатДНК (тандемные повторы) – максимально изменчивый компонент генома, вплоть до возможной видоспецифичности.

Анализ секвенированной части генома дал колоссальный толчок развитию биологии: создаются базы данных, содержащие описания отдельных элементов (проект ENCODE). К сожалению, тандемные повторы остаются за пределами подобных проектов из-за отсутствия отработанных методов их выявления и классификации. Ранее была продемонстрирована возможность поиска тандемных повторов мыши и человека биоинформатическими методами [2].

К настоящему времени у свиньи описаны два мажорных семейства сатДНК, обнаруженные после рестрикции тотальной ДНК и последующем клонировании и картировании фрагментов: это Mc1 семейство располагается на суб/метацентрических хромосомах (1-12) и X хромосоме и Ac2 семейство обнаружено на акроцентрических хромосомах (13-18) [3]. Для семейства Mc1 была показана сильная вариабельность мономеров, были обнаружены мономеры, которые в разной степени гибридизовались к разным суб/метацентрическим хромосомам [4]. Благодаря наличию вариабельности мономеров, были клонированы специфические повторы для хромосомы 1 [3] и для хромосомы 9 [5]. Идентифицировали подсемейство Mc2, обнаруженные клоны которого специфически гибридизовались с 11 хромосомой, также давали более слабый сигнал на некоторых других суб/метацентрических хромосо-

мах. На основании 50 п. н. мономера подобрали специальные праймеры, благодаря которым удалось специфически пометить хромосому 11 [6]. Для суб/метацентриков описали еще несколько семейств сатДНК [4, 5]. Праймеры на основе Mc1 и Ac2 были использованы для мечения центромер групп хромосом *in situ* методом PRINS [7].

Объектом исследования в настоящей работе являются секвенированные геномы свиней четырех пород различной степени собранности, которые присутствуют в открытом доступе. Анализ проводили методами биоинформатики.

Цель работы: установить возможность выявления больших тандемных повторов (ТП) в относительно плохо собранных, по сравнению с геномами мыши и человека, геномах различных пород свиней. В исследовании были поставлены следующие задачи: найти тандемные повторы в пяти сборках геномов свиньи; сравнить их с известными повторами из базы данных Repbase; дать приблизительную оценку количеству найденных семейств тандемных повторов в геномах свиньи с помощью анализа исходных ридов; сопоставить найденные семейства тандемных повторов в разных сборках; (5) проверить полученные данные при помощи FISH.

Материал и методика исследований. Поиск больших тандемных повторов основывался на методике, предложенной в работе Комиссарова с соавт. [2]. Для поиска ТП использовали программу TRF (Tandem repeat finder) [8]. Для обработки результатов работы TRF использовали оригинальную программу, написанную на языке Python (<http://github.com/DmitriiOstr/tandem-repeat-family-finder>). В качестве кандидатов в большие тандемные повторы использовали следующие параметры: длина мономера больше 4 п.н., длина поля больше 2 т. п. н., содержание GC от 20 до 80%, энтропия поля больше 1,76.

Для гибридизации *in situ* подбирали короткие одноцепочечные олигонуклеотидные зонды к каждому из подсемейств ТП.

Для цитогенетического анализа использовали клетки крови, полученные от свиней, разводимых на свинокомплексе. Культивирование клеток крови и флуоресцентную гибридизацию *in situ* проводили стандартными методами [9, 10].

Результаты исследований и их обсуждение. В базе данных геномных сборок <http://www.ncbi.nlm.nih.gov/genome/> присутствуют 5 геномных сборок *Sus scrofa*. Геном свиньи породы Дюрок собран до хромосом. Это репрезентативный геном, уровень сборки предшествующий эталонному (reference) геному.

Одной из главных характеристик качества сборки генома является метрика N50. Значение метрики N50 для скэффолдов или контигов

показывает наибольшую длину скэффолда или контига, такую, что в скэффолдах (контигах) не меньшей длины содержится не меньше 50% суммарной длины скэффолдов (контигов).

Таблица 1 – Геномные сборки свиньи, использованные в работе

WGS	Сборка	Метод сборки	Платформа секвенирования	Порода <i>Sus scrofa</i>
AEMK	Sscrofa10.2	SOAPdenovo v.2.208 and Cortex v1.1	Illumina GALL	Дюрок (Duroc)
AJKK	minipig_v1.0	SOAPdenovo v.2.30	Illumina HiSeq2000	Китайская карликовая (Wuzhishan)
AOCR	SscrofaMinipig	SOAPdenovo v.Sept.2011	Illumina HiSeq	Геттингенская карликовая (Goettingen)
LIDP	Ss10.2_mar2013	GMAP v.DEC-2012; BioScope v.1.2; Newbler v.2.7.	454; SOLiD	Геттингенская карликовая (Goettingen)
AORO	Tibetan_Pig_v1.0	SOAPdenovo v.1.06	Illumina HiSeq	Тибетская (Tibetan)

WGS – название проекта по секвенированию для сборки генома, *Сборка* – название сборки.

Сборка *Sus scrofa* породы Дюрок является наиболее законченной сборкой по сравнению с остальными, но значение N50 для скэффолдов составляет 576.008. Среди сборок, собранных до одного уровня, видно, что наилучшей является сборка минипига Wuzhishan, а сборка тибетской свиньи хоть и хуже, но порядок значения N50 для скэффолдов тот же. Оставшиеся обе сборки минипигов породы Goettingen имеют самые низкие значения N50 для контигов (AOCR – 22.008, LIDP – 17.259). Для сравнения, в таких хорошо собранных геномах, как геном человека, значение N50 для контигов и скэффолдов равно 56,413,054 и 59.364.414 соответственно (сборка GRCh38), а для генома мыши значение N50 для контигов равно 32,273,079, для скэффолдов 52.589.046 (для сборки GRCm38.p4), т.е. величина N50 для этих геномов геномов больше в 1000 раз. Наибольшее количество ТП обнаружено в сборке AJKK, которая характеризуется наивысшим значением N50 среди прочих сборок, у тибетской свиньи со следующим по убыванию значением N50 повторов найдено чуть меньше, но они более разнообразны.

С помощью blastn мы сравнили семейства из различных сборок между собой. Обнаружено 12 семейств, присутствующих во всех сборках. Помимо сателлитов, известных из RepBase, во всех сборках имеется еще 8 общих семейств. В некоторых случаях минимальный мономер семейства отличается в разных сборках. Очень вариабельными мономерами отличается семейство гомологичное центромерному сателлиту

SSAT, семейство гомологичное SScr_AEMK-96A и семейство гомологичное SSRS2.

Наряду с общими, есть и ТП, специфичные для каждой сборки. Найдено 13 специфичных для сборки Дюрок семейств ТП, из которых только одно принадлежит ранее известному сателлитному повтору MSAT-1_SSc, по нашей классификации Sscr_AEMK-38A. В сборке карликовой геттингенской свиньи обнаружили 19 специфичных только для нее семейств, из которых в одном семействе мономеры имеют гомологию к части диспергированного повтора типа LINE1, два семейства родственные SINE и два семейства родственные LTR ERV1. Для сборки карликовой свиньи породы Wuzhishan найдено 20 специфичных семейств, из которых три имеют гомологию к частям LINE. Наибольшее количество специфических семейств ТП найдено в сборке тибетской свиньи – 30 семейств ТП. Всего обнаружено 126 семейств ТП, из них 121 ранее неизвестно. Около 10% семейств ТП присутствуют во всех четырех сборках.

Помимо ТП, основанных на фрагментах LINE, мы обнаружили много полей ТП, основанных на фрагментах SINE, такие ТП еще не были показаны у млекопитающих. Мономеры ТП, основанные на SINE, всегда имеют в своем составе добавочный фрагмент. ТП, основанные на фрагментах ERV3, наблюдали среди ТП мышей [2]. У свиней похожий тип ТП основан на ERV1, и в состав мономера ТП входят фрагменты LTR. Такой тип ТП выявлен впервые.

Известно, что при гибридизации хромосом свиньи с полноразмерными LINE и SINE не удается получить такую четкую картину распределения районов эухроматина и факультативного гетерохроматина, как у человека и мыши [11]. Результаты настоящей работы показывают, что именно необычная организация ТП свиньи может быть ответственна за нечеткость картины.

Для некоторых семейств ТП были подобраны и опробованы олигонуклеотидные зонды для проведения FISH. Первыми использовали пробы к ранее клонированным сателлитным повторам по нашей классификации Sscr_AEMK-15A, Sscr_AEMK-50A и Sscr_AEMK-332A соответственно.

Таблица 2 – Последовательности зондов, использованных для постановки FISH

Семейство ТП	Последовательность зонда	Локализация зонда
1	2	3
Sscr_AEMK-15A	TGCAGTGGATTTTATGCAGC	Цен Ac
Sscr_AEMK-332A	GCAAGCGTTGCTTTCCTGA	Цен Mc, X
Sscr_AEMK-50A	TCGACCAGCTGTCCCGGCC	Цен Mc, X

Продолжение таблицы 2

1	2	3
Sscrf_AEMK-859A	CTGTCAACCAATCGGCCCTTC	Цен Ас
Sscrf_AEMK-243A	AGACTGCAGGTGCTCCAGATGAGCC	Цен Ас
Sscrf_AEMK-39A	AGTCCTCTGGCTCTGGACAACA	Цен Мс

Ас – акроцентрические хромосомы, *Мс* – метацентрические хромосомы, *Цен* – центромера.

Все проверенные нами олигонуклеотидные зонды были картированы в прицентромерных районах акроцентрических или субметацентрических хромосом. Мы надеемся, что результаты пилотных экспериментов, полученные в настоящей работе, будут способствовать созданию работающих олигонуклеотидных проб специфичных к разным хромосомам, а возможно и к разным породам.

Разнообразие генов больше у китайских пород, чем у европейских. Это может быть результатом жесткого одомашнивания в Европе с конца XVIII в., в результате чего многие локальные популяции вымерли или оказались на грани вымирания, или потеряли свою индивидуальность. В Китае действуют законы, ограничивающие перемещение свиней, что способствует стабилизации локальных пород [12].

Во всех проанализированных породах обнаружены семейства ТП, характерные только для них. В этом отношении примечательна сборка генома тибетской свиньи, в которой найдено самое большое количество – 30 специфичных семейств ТП, что составляет приблизительно половину от всего количества найденных семейств ТП в этой свинье. Здесь же обнаружено самое большое количество семейств ТП (62). Это согласуется с данными о генетическом разнообразии китайских свиней. Среди изученных пород именно тибетская свинья является максимальным приближением к дикому типу, обитающему на воле. Гибрид китайской и европейских линий свиньи дает и максимальное многообразие ТП среди четко выявленных пород, т.е. скрещивание приводит к увеличению разнообразия ТП.

Заключение. Обнаружены тандемные повторы (ТП) в пяти геномных сборках *Sus scrofa* четырех различных пород. Выявлено 126 семейств больших ТП, из них 121 ранее не описано. Около 10% семейств (12 семейств) присутствует во всех четырех сборках. Только 4 из этих 12 семейств описаны ранее и внесены в базу Repbase. Таким образом, найдено 8 новых семейств ТП общих для вида. С помощью метода FISH продемонстрировали возможность создания олигонуклеотидных для зондов для картирования ТП. Обнаружены ТП, мономеры которых основаны на частях ретроэлементов, и был проанализирован их состав. Впервые обнаружены ТП, образованные из частей коротких

диспергированных повторов, что, вероятно, является особенностью генома свиньи.

ЛИТЕРАТУРА

1. Kit, S. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues / S. Kit // *J. of Mol. Biol.* – 1961. – V. 3. – P. 711-716.
2. Komissarov, A.S. Tandemly repeated DNA families in the mouse genome / A.S. Komissarov, E.V. Gavrilova, S.J. Demin, A.M. Ishov, O.I. Podgornaya // *BMC Genomics.* – 2011. – V. 12, №531. – P.1-21.
3. Jantsch, M. Meiotic chromosome behaviour reflects levels of sequence divergence in *Sus scrofa domestica* satellite DNA / M. Jantsch, B. Hamilton, B. Mayr, D. Schweizer // *Chromosoma.* – 1990. – V. 99. – P. 330-335.
4. Miller, J.R. A chromosomal basis for the differential organization of a porcine centromere-specific repeat / J.R. Miller, J. Hindkjaer, P.D. Thomsen // *Cytogenet. Cell. Genet.* – 1993. – V.62. – P. 37-41.
5. Janzen, M.A. Characterization of a swine chromosome-specific centromeric higher-order repeat / M.A. Janzen, L.B. Buoen, F. Zhao, C.F. Louis // *Mammalian Genome.* – 1999. – V. 10. – P. 579-584.
6. Riquet, J. Sequence analysis and genetic mapping of porcine chromosome 11 centromeric S0048 marker / J. Riquet, P. Mulsant, M. Yerle, M.S. Cristobal-Gaudy, P. Tissier, D. Milan, J. Gellin // *Cytogenet. Cell. Genet.* – 1996. – V. 74. – P. 127-132.
7. Rogel-Gaillard, C. Construction of a swine YAC library allowing an efficient recovery of unique and centromeric repeat sequences / C. Rogel-Gaillard, N. Bourgeaux, J.C. Save, C. Renard, P. Coullin, P. Pinton, M. Yerle, M. Vaiman, P. Chardon // *Mammalian Genome.* – 1997. – V. 8. – P.186-192.
8. Benson, G. Tandem repeats finder: a program to analyze DNA sequences / G. Benson // *Nucleic acids research.* – 1999. – V. 27, № 2. – P. 573.
9. Moorhead, P.S. Chromosome preparations of leukocytes cultured from human peripheral blood / P.S. Moorhead, P.C. Nowell, W.J. Mellman, D.M. Battips, D.A. Hungerford // *Exp. Cell Res.* – 1960. – V. 20. – P. 613-616.
10. Chowdhary, B.P. FISH on metaphase and interphase chromosomes demonstrates the physical order of the genes for GPI, CRC, and LIPE in pigs / B.P. Chowdhary, C. de la Sena, I. Harbitz, L. Eriksson, I. Gustavsson // *Cytogenet. Cell Genet.* – 1995. – V. 71, №2. – P. 175-178.
11. Thomsen, P.D. Pig genome analysis: differential distribution of SINE and LINE sequences is less pronounced than in the human and mouse genomes / P.D. Thomsen, J.R. Miller // *Mamm Genome.* – 1996. – V. 7, № 1. – P. 42-46.
12. Megens, H.-J. Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication / H.-J. Megens, R. Crooijmans, M. Cristobal, X. Hui, N. Li, M. Groenen // *Genet. Sel. Evol.* – 2008. – V. 40. – P. 103-128.