
ОРИГИНАЛЬНЫЕ
ИССЛЕДОВАНИЯ

УДК 575.181, 577.218

Статья посвящена 120-летию
со дня рождения выдающегося российского генетика,
академика Бориса Львовича Астаурова

**УЧАСТИЕ ПОВТОРЯЮЩЕГОСЯ ЭЛЕМЕНТА ГЕНОМА GGAAA
В ДИФФЕРЕНЦИРОВКЕ ПОЛА У КУРИЦЫ**

© 2024 г. А. Ф. Сайфитдинова^{a, b, *}, А. А. Жукова^{a, c}

^aРоссийский государственный педагогический университет им. А. И. Герцена,
Санкт-Петербург, 191186 Россия

^bСанкт-Петербургский государственный университет,
Санкт-Петербург, 199034 Россия

^cСанкт-Петербургский филиал ФГБНУ ВНИРО “ГосНИОРХ” им. Л. С. Берга,
Санкт-Петербург, 199053 Россия

*e-mail: saifitdinova@mail.ru

Поступила в редакцию 05.12.2024 г.

Окончательная версия 23.12.2024 г.

Принято к публикации 24.12.2024 г.

Тандемные повторяющиеся элементы, образующие протяженные полипуриновые/полипиримидиновые непрерывные последовательности обнаружены в геномах различных видов животных. Особенности их структуры способствуют изгибанию спирали ДНК и переходу к неканоническим формам вторичной структуры ДНК. В современной научной литературе можно встретить множество примеров участия таких элементов в регуляции экспрессии генов и образовании альтернативных транскриптов в клетках разных типов (Matos-Rodrigues et al., 2023). Ранее нами описан повторяющийся элемент (GGAAA)_n курицы (*Gallus gallus domesticus*), который преимущественно локализован на половой хромосоме W и составляет около 1% генома самок (Komissarov et al., 2018). В данной работе мы выявили особенности локализации этого тандемного повтора в геноме курицы в составе аутосом и половой хромосомы Z. Был выявлен ряд генов, содержащих тандемно повторенные элементы (GGAAA)_n в составе некодирующих транскрибируемых регуляторных районов, которые могут влиять на интенсивность экспрессии и образование альтернативных транскриптов. Функциональная характеристика генов, несущих блоки (GGAAA)_n, позволила выдвинуть предположение об участии этих тандемных повторов в регулировании дифференциальной активности генов, важной для развития признаков полового диморфизма у курицы.

Ключевые слова: курица, тандемные повторы, полипуриновые/полипиримидиновые непрерывные последовательности, анализ геномных данных, дифференцировка пола

DOI: 10.31857/S0475145024020015, **EDN:** MDPCTC

ВВЕДЕНИЕ

Для птиц характерно наличие генетического механизма определения пола с системой половых хромосом ZW, при которой самки гетерогаметны, а самцы гомогаметны. В ходе эволюции некодирующие участки W-хромосомы птиц подверглись значительной деградаци и накоплению повторяющихся элементов (Komissarov et al., 2018), однако, в отличие от млекопитающих, гетероморфная половая хромосома у птиц сохранила небольшое число генов, причем все они имеют биаллельную экспрессию в широком спектре взрослых и эмбриональных тканей

и чувствительны к сокращению дозы гена (Bellott et al., 2017). Пол у всех изученных видов птиц определяется генетически, но в составе половых хромосом не найдено конкретных генов — инвариантных переключателей типа половой дифференцировки, подобно *SRY* у млекопитающих (Bellott et al., 2017).

Ранее нами описан повторяющийся элемент (GGAAA)_n, который у курицы локализуется преимущественно в виде двух протяженных блоков на обоих плечах половой хромосомы W и составляет около 1% генома (Komissarov et al., 2018). В эукариотических геномах гомо-

пуриновые последовательности составляют значительную часть простых повторяющихся последовательностей ДНК, они часто встречаются в регуляторных областях генов. Помимо этого, протяженные полипуриновые/полипиримидиновые непрерывные последовательности обнаружены в местах рекомбинации, в точках начала репликации и внутри различных типов мобильных элементов, включая временно встраивающихся в геном последовательности вирусов (Matos-Rodrigues et al., 2023). Особенность организации таких элементов способствует увеличению их копийности в составе нерекombинирующей половой хромосомы. Транскрипция повторяющихся элементов, в состав которых входят комплементарные регуляторным районам последовательности, создает предпосылки для вовлечения их в регуляцию экспрессии генов. Механизм регуляции может быть основан как на связывании белков с неканоническими структурами ДНК и РНК, так и быть опосредован некодирующими РНК, оказывающими влияние на конформационные изменения нуклеиновых кислот (Matos-Rodrigues et al., 2023).

Получены свидетельства того, что локализованный на половой хромосоме W повтор (GGAAA)_n транскрибируется в соматических клетках, однако значение этой транскрипции пока не установлено (Komissarov et al., 2018). Ранее повторяющийся элемент (GGAAA)_n был описан в регулируемом промоторе гена овотрансферрина у фазана (Maroteaux et al. 1983). У курицы наличие тандемного повтора (GGAAA)_n было описано в промоторе гена промежуточных нейрофиламентов, транскрипция которого может изменяться под влиянием стимуляции, причем было показано, что изменение числа повторов в составе промотора влияет на его активацию (Zopf et al., 1990). Мы выдвинули предположение, что повторы (GGAAA)_n, транскрибирующиеся с протяженных блоков на хромосоме W, могут участвовать в половой дифференцировке у курицы посредством взаимодействия с комплементарными короткими блоками тандемных повторов в составе регуляторных областей генов. Мы выявили последовательности тандемных повторов (GGAAA)_n в актуальной версии сборки генома курицы и проанализировали их потенциальное влияние на дифференцировку признаков, специфичных для представителей разного пола.

МАТЕРИАЛЫ И МЕТОДЫ

Для проведения исследования были использованы данные сборки генома курицы (*Gallus gallus*, Linnaeus, 1758) GRCg6a (GenBank accession: GCF_000002315.6). Эта сборка содержит информацию о последовательностях половых хромосом Z и W, а также 33 аутосом из 38 имеющихся в геноме курицы. Для определения локализации повторов на хромосомах мы использовали программное обеспечение Unipro UGENE (Version 35, Unipro, Новосибирск, Россия). Для функциональной характеристики генов использовали данные из базы данных GeneCards (Version 5.22.0, Weizmann Institute of Science, Rehovot, Израиль), дополнительно информация об их белковых продуктах была проанализирована с помощью аннотированной базы данных белков UNIPROT (The UniProt Consortium: EMBL-EBI, SIB, PIR). Для поиска молекулярного взаимодействия и биологических путей использовали плагин StringApp (Version 1.6.0) для платформы Cytoscape (Version: 3.8.1, Cytoscape Consortium).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В геноме курицы (сборка GRCg6a (GCF_000002315.6)) тандемные блоки (GGAAA)_n в количестве 7 и более копий были найдены в составе транскрибируемых некодирующих последовательностей 67 генов (Табл. 1). Из них 55 генов входят в состав аутосом (1, 2, 3, 4, 5, 6, 9, 15, 20, 22, 27, 28, 31) и 12 — в состав половой хромосомы Z. Среди них гены *Robo2*, *Sh3gl2*, *Slit2*, *Zmiz1*, экспрессия которых чувствительна к уровню половых гормонов. Активность генов *Lrrc4c* и *Nrxn1* критична для роста фолликулов в яичниках, а продукты генов *Csmd3*, *Csmd1* и *Rorb* участвуют в формировании яйца. Продукт гена *Grik1* регулирует на уровне центральной нервной системы откладку яиц. Подавляющее большинство интерстициальных коротких блоков повтора (GGAAA)_n были выявлены в составе первых интронов генов, кодирующих трансмембранные белки, которые участвуют в синаптической передаче и вовлечены в процессы обучения, пения и реализации агрессивного поведения. Присутствие протяженных участков полипуриновых/полипиримидиновых непрерывных последовательностей, состоящих из тандемно повторенного мономера (GGAAA)_n, может модулировать активность промоторов или переключать транскрипцию на альтернативные промоторы, приводя к об-

Таблица 1. Локализация тандемного повтора (GGAAA)_n в геноме курицы GRCgba

Хромосома	Ген	Район	Число копий	Характеристика белка
1	<i>Aff3</i>	3 интрон	27	транскрипционный фактор
1	<i>Atp7b</i>	1 интрон	36	медь-транспортирующая АТФаза
1	<i>Bmx</i>	8 интрон	26, 21, 18, 16	тирозиновая протеинкиназа
1	<i>Cd101</i>	10 интрон	11	сигнальный белок мембраны
1	<i>Clybl</i>	7 интрон	44	цитрамалил-КоА-лиаза
1	<i>Dram1</i>	3' область	14	модулятор аутофагии
1	<i>Gja8</i>	2 интрон	40	коннексин
1	<i>Gpc5</i>	6 интрон	34, 15, 14, 13	сигнальный белок мембраны
1	<i>Grik1</i>	1 интрон	30	глутаматергический белок
1	<i>Grm5</i>	2 интрон	31	глутаматергический белок
1	<i>Imp2l</i>	5 интрон	46	пептидаза внутренней мембраны митохондрий
1	<i>Magi2</i>	16 интрон	34, 31, 17	мембранная гуанилат-киназа
1	<i>Mpz1l</i>	4 интрон	11	миелиновый мембранный белок
1	<i>Nbea</i>	30 интрон	51	интегральный белок везикул
1	<i>Pcdh9</i>	2 интрон	38	кадгерин
1	<i>Robo2</i>	1 интрон	26	сигнальный белок мембраны
1	<i>Tmprss15</i>	16 интрон	70	мембранная сериновая протеаза
1	<i>Vwf</i>	21 интрон	22	гликопротеин
1	<i>Zpld1</i>	1 интрон	41	гликопротеин
2	<i>Adarb2</i>	1 интрон	40	аденозиновая деаминаза
2	<i>Amph</i>	16 интрон	34, 30, 21	белок везикулярных мембран
2	<i>Cdh19</i>	2 интрон	29	кадгерин
2	<i>Cntnap2</i>	1 интрон	33	нейрексин
2	<i>Cubn</i>	37 интрон	21	сигнальный белок мембраны
2	<i>Dip2c</i>	4 интрон	31	регуляторный белок
2	<i>Dlgap1</i>	2 интрон	15	компонент глутаматергического синапса
2	<i>Frk</i>	1 интрон	21	тирозиновая киназа
2	<i>Npsr1</i>	1 интрон	21	трансмембранный рецептор
2	<i>Nrn1</i>	2 интрон	45	сигнальный белок мембраны
2	<i>Ppp1r9a</i>	6 интрон	25	мембранная фосфатаза
2	<i>Tsnare1</i>	10 интрон	32	интегральный белок везикул
2	<i>Wisp1</i>	1 интрон	32	сигнальный белок
3	<i>Bmp5</i>	7 интрон	13	сигнальный белок
3	<i>Csmd1</i>	3 интрон	20	сигнальный белок
3	<i>Csmd3</i>	1 интрон 35 интрон	40 9	сигнальный белок

Таблица 1. Окончание

Хромосома	Ген	Район	Число копий	Характеристика белка
3	<i>Grik2</i>	11 интрон	35	глутаматергический белок
3	<i>Msra</i>	6 интрон	25	метионинсульфоксидредуктаза
3	<i>Nrxn1</i>	1 интрон	20	глутаматергический белок
3	<i>Raly1</i>	1 интрон	39	РНК-связывающий белок
3	<i>Sntg2</i>	10 интрон	26	интегральный белок
4	<i>Fhl1</i>	2 интрон	28	интегральный белок
4	<i>Gria3</i>	3' область	33	глутаматергический белок
4	<i>Ppp2r2c</i>	6 интрон	42	мембранная протеаза
4	<i>Sorcs2</i>	3 интрон	13	сигнальный белок
5	<i>Lrrc4c</i>	6 интрон	42	сигнальный белок
6	<i>Gdf10</i>	2 интрон	22, 16	сигнальный белок
6	<i>Zmiz1</i>	2 интрон	8	транскрипционный регулятор
9	<i>Mecom</i>	1 интрон	7	транскрипционный регулятор
15	<i>Tbx1</i>	3 интрон	19	транскрипционный фактор
20	<i>Asip</i>	1 интрон	30, 16	сигнальный белок
22	<i>Nefm</i>	промотор	31	белок цитоскелета нейронов
27	<i>Asic2</i>	2 интрон	19	ионный канал
28	<i>Onecut3</i>	1 интрон	40	транскрипционный фактор
31	<i>Chir-B2</i>	1 интрон	14	сигнальный белок мембраны
31	<i>Chir-B4</i>	1 интрон 2 интрон	15 26	сигнальный белок мембраны
Z	<i>Aldh1a1</i>	1 интрон	19	альдегидная дегидрогеназа
Z	<i>Dnai1</i>	8 интрон	9	компонент моторного белка
Z	<i>Fbxl17</i>	7 интрон	10	регулятор катаболизма
Z	<i>Ptprd</i>	2 интрон 23 интрон	19 20	мембранная тирозиновая фосфотаза
Z	<i>Rab3c</i>	2 интрон	23	малая мембранная ГТФаза
Z	<i>Rasef</i>	1 интрон	21	малая мембранная ГТФаза
Z	<i>Rit2</i>	1 интрон	27	малая мембранная ГТФаза
Z	<i>Rnf165</i>	1 интрон	21	убиквитиновая трансфераза
Z	<i>Rorb</i>	1 интрон 7 интрон	51 30	транскрипционный фактор
Z	<i>Setbp1</i>	2 интрон 3 интрон	27 26	транскрипционный фактор
Z	<i>Sh3gl2</i>	3 интрон	27	транскрипционный регулятор
Z	<i>Trpm3</i>	1 интрон	31	трансмембранный катионный канал

разованию белков разной длины, что особенно важно для регуляции активности трансмембранных рецепторов.

По данным базы данных UNIPROT, продукты более 40% выявленных генов оказались вовлечены во взаимные регуляторные и белок-белковые взаимодействия, а 13 из этих генов образуют общий функциональный кластер. Большая часть генов, содержащих полипуриновые/полипиримидиновые непрерывные последовательности (GGAAA)_n, участвует в синаптической передаче сигнала. Среди таких генов оказались *Grm5*, *Dlgap1*, *Grik1*, *Grik2*, *Gria3*, продукты которых входят в состав глутаматергических синапсов и существенны для синаптической пластичности. Как было показано ранее, эти гены задействованы в обучении и развитии памяти, а для певчих птиц играют существенную роль в обучении пению (Wada et al., 2004). Экспрессия еще трех выявленных нами генов *Lrrc4c*, *Robo2*, *Slit2* также упоминается в литературе в связи с пением и описана в голосовых ядрах мозга у самцов птиц (Lovell et al., 2018). Подавление активности этих генов у самок птиц может быть опосредовано участием транскриптов тандемных повторов (GGAAA)_n, локализованных на половой хромосоме W.

Агрессивное половое поведение самцов *G. gallus*, петухов, непосредственно связано с борьбой за социальное доминирование и является одной из характеристик проявления полового диморфизма у этого вида. При сравнении полученного в ходе анализа списка генов из генома *G. gallus*, несущих в своем составе повторяющийся элемент (GGAAA)_n, с гомологами человека, было выявлено 8 генов, дифференциальная экспрессия которых так или иначе связана с агрессивным поведением. Один из них, а именно *Sorcs2*, был ранее выявлен в ходе скрининга мутаций у пород бойцовых кур, характеризующихся повышенной агрессивностью в поведении петухов (Li et al., 2016). Интересно, что описанная у китайской желтой карликовой породы кур мутация, приводящая к увеличению вероятности реализации альтернативного сайта инициации транскрипции, была локализована в 3-м интроне. Именно в третьем интроне в этом гене в геномной сборке GRCgба локализуется блок тандемных повторов (GGAAA), состоящий из 13 мономеров, который может регулировать реализацию альтернативной транскрипции и увеличивать вероятность продукции более длинного белка с дополнительным интегральным доменом.

ЗАКЛЮЧЕНИЕ

Функциональные особенности генов, несущих в составе некодирующих регуляторных последовательностей тандемный повтор (GGAAA)_n и различия в их экспрессии у самцов и самок, свидетельствуют о том, что полипуриновые/полипиримидиновые непрерывные последовательности (GGAAA)_n могут участвовать в регуляции активности генов, вовлеченных в дифференцировку пола, развитие признаков полового диморфизма и половое поведение у куры.

БЛАГОДАРНОСТИ

Результаты настоящего исследования посвящаются 120-летию со дня рождения выдающегося российского ученого академика Бориса Львовича Астаурова, внесшего существенный вклад в исследование генетики пола. Авторы благодарят Р. В. Четверикову и Е. В. Большакову, выполнявших учебные проекты и внесших вклад в исследование.

ФИНАНСИРОВАНИЕ РАБОТЫ

Исследование не имеет финансовой поддержки.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

При выполнении данного исследования люди и животные не использовались в качестве объектов.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

ВКЛАД АВТОРОВ

Идея исследования принадлежит А. Ф. Сайфитдиновой. Оба автора участвовали в анализе данных, обсуждении результатов и подготовке рукописи.

СПИСОК ЛИТЕРАТУРЫ

- Bellott D.W., Skaletsky H., Cho T.-J. et al. Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators // Nature Genetics. 2017. V. 49. № 387. P. 94.
- Komissarov A.S., Galkina S.A., Koshel E.I. et al. New high copy tandem repeat in the content of the chicken

- W chromosome // *Chromosoma*. 2018. V. 127. № 1. P. 73–83.
- Li Z., Zheng M., Abdalla B.A., Zhang Z. et al. Genome-wide association study of aggressive behaviour in chicken // *Scientific reports*. 2016. V. 6. № . 30981.
- Lovell P.V., Huizinga N.A., Friedrich S.R. The constitutive differential transcriptome of a brain circuit for vocal learning // *BMC Genomics*. 2018. V. 19. № 1.
- Matos-Rodrigues G., Hisey J.A., Nussenzweig A. et al. Detection of alternative DNA structures and its implications for human disease // *Molecular Cell*. 2023. V. 83. № 20. P. 3622–3641.
- Maroteaux L., Heilig R., Dupret D. et al. Repetitive satellite-like sequences are present within or upstream from 3 avian protein-coding genes // *Nucleic Acids Research*. 1983. V. 11. P. 1227–1243.
- Wada K., Sakaguchi H., Jarvis E.D. et al. Differential expression of glutamate receptors in avian neural pathways for learned vocalization // *The Journal of Comparative Neurology*. 2004. V. 476. P. 44–64.
- Zopf D., Dineva B., Betz H., Gundelfinger E.D. Isolation of the chicken middle-molecular weight neurofilament (NF-M) gene and characterization of its promoter // *Nucleic Acids Research*. 1990. V. 18. № 3. P. 521–529.

Participation of Chicken GGAAA Repeat in Sex Differentiation

A. F. Saifitdinova^{1, 2, *}, A. A. Zhukova^{1, 3}

¹Herzen State Pedagogical University of Russia, St. Petersburg, 191186 Russia

²St. Petersburg University, St. Petersburg, 199034 Russia

³Saint Petersburg branch of the Russian Federal Research Institute of Fisheries and oceanography, St. Petersburg, 199053 Russia

*e-mail: saifitdinova@mail.ru

Tandem repeating elements that form extended polypurine/polypyrimidine tract sequences have been found in the genomes of various species. Their structural properties bend the DNA helix and cause a transition to non-canonical DNA secondary structures. The modern scientific literature describes many examples of the involvement of such elements in the regulation of gene expression and the formation of alternative transcripts in cells of different types of differentiation (Matos-Rodrigues et al., 2023). Previously, we described the (GGAAA)_n repeating element of chicken (*Gallus gallus domesticus*), which is predominantly localized on the sex chromosome W and makes up about 1% of the female genome (Komissarov et al., 2018). Here we describe the localization peculiarities of this tandem repeat in the chicken genome within autosomes and the sex chromosome Z. The study identified a number of genes carrying tandemly repeated (GGAAA)_n elements within non-coding transcribed regulatory regions that can influence expression intensity and the formation of alternative transcripts. Functional characterization of genes carrying stacks of (GGAAA)_n allowed us to suggest the involvement of these tandem repeats in regulating the differential activity of genes important for the development of sexual dimorphism traits in chicken.

Keywords: chicken, tandem repeats, polypurine/polypyrimidine tract sequences, genome data analysis, sex differentiation