

Н. Ф. Бакоев

Полногеномное обнаружение и анализ вариаций числа копий у овец сальской породы с использованием массива SNP высокой плотности

Аннотация.

Цель: определить области вариации числа копий (CNV, Copy Number Variation) и определить потенциально специфические CNV у овец сальской породы овец.

Материалы и методы. Исследования проводили на овцах сальской породы в количестве 48 голов. Генотипирования проводили с использованием микрочипов высокой плотности *Ovine SNP 600 K BeadChip array* (Illumina, США). Поиск CNV проводили с использованием программного обеспечения *PenncNV* на основе референсного генома *Oar_v3.1*.

Результаты. Определили 4585 аутосомных CNV у 48 овец сальской породы. Средняя длина CNV составила 92,3 Kb. 21,1 % от общего числа CNV имели размер менее 50 Kb, 72,8 % CNV имели размер от 50 до 100 Kb в среднем, на одно животное определено от 65 до 138 CNV. Было получено 527 аутосомных CNVR (CNV regions), представляющих 54,77 Mb всего генома овцы. Доля хромосом, охваченная CNVR, варьируется в зависимости от хромосомы: от 0,5 % на OAR25 до 9 % на OAR24. Анализ CNVR в системе *ENSEMBL* выявил 42 гена, в которых локализованы CNVR, остальные CNVR расположены в межгенных областях. Сопоставление выявленных CNVR с QTL овец из базы данных *Animal QTL* показало, что 39,24 % генов ответственны за репродуктивные качества, а 20,25 % ответственны за признаки, связанные с шерстью.

Выводы. Проведена функциональная аннотация CNVR, локализованных в генах и межгенных пространствах, и проанализировано общее количество детектированных CNVR в базе QTL. Результаты показали, что CNV локализованы в генах, связанных с селекционно-значимыми признаками овец.

Ключевые слова: сальская порода; CNV; CNVR; полногеномная карта.

Автор:

Бакоев Н. Ф. — кандидат сельскохозяйственных наук; e-mail: nekruz82@bk.ru; ФГБНУ «Федеральный исследовательский центр животноводства – ВИЖ имени академика Л.К. Эрнста»; 142132, Россия, Московская область, Городской округ Подольск, поселок Дубровицы, 60.

Введение. Современные молекулярно-генетические методы позволяют определить продуктивные и фенотипические признаки сельскохозяйственных животных. Одной из форм геномных структурных вариаций, лежащей в основе фенотипического разнообразия живых организмов, является вариация числа копий. CNV представляет собой тип геномного полиморфизма, который может быть важной компонентой фенотипической изменчивости [1, 2]. CNV определяются как дупликации или делеции геномных сегментов, размер которых варьируется от 50 пар оснований (п.н.) до миллионов пар оснований (Мб) и варьируется у разных видов животных [3, 4]. Популяционный анализ с использованием вариации числа копий (CNV) намного сложнее, чем анализ с использованием SNP, из-за разного количества копий и непостоянных границ CNV у разных особей. Эти CNV перекрываются друг с другом, но имеют

разные точки разрыва у разных особей, что затрудняет расчет частот генотипов для CNV, поскольку одно из перекрывающихся событий могло возникнуть независимо от других.

С развитием технологии секвенирования и соответствующими достижениями в области молекулярной биологии геном стали лучше понимать, и было обнаружено, что большое количество генетических вариантов, присутствующих в геноме, оказывают различное влияние на фенотипические признаки [5, 6]. CNV являются важной генетической компонентой генома для многих заболеваний и вносят значимый вклад в эволюцию сельскохозяйственных животных [7].

К примеру, CNV влияют на фенотип пигментации и масти шерсти домашних животных. Уровень экспрессии гена сигнального белка агути (ASIP) был значительно повышен в клетках шерсти у овец, содержащих копию этого гена, кото-

рая располагалась непосредственно после промотора *ITCH* [8, 9]. Так же было обнаружено, что дупликация гена *ASIP* у коз может способствовать появлению белой масти шерсти [8].

В типичных массивах SNP присутствуют два важных показателя интенсивности сигналов SNP, включая log R Ratio (LRR) и частота аллелей В (BAF). LRR является мерой нормализованной общей интенсивности сигнала, а BAF является мерой нормализованного соотношения интенсивностей аллелей [10]. Скрытые модели Маркова (HMM) были успешно применены для идентификации CNV из массивов генотипирования SNP в программах PennCNV [11] и QuantiSNP [12]. PennCNV и QuantiSNP включают LRR и BAF в модель HMM. Кроме того, PennCNV включает информацию о частоте популяционных аллелей каждого SNP и расстояние между соседними SNP. Метод PennCNV показал точность и надежность результатов по поиску CNV и в настоящее время активно применяется в различных крупномасштабных генетических исследованиях [13, 14, 15]. PennCNV определяет шесть скрытых состояний, каждое из которых соответствует различному состоянию номера копии. Для PennCNV требуются файлы ин-

тенсивности сигнала (один файл на образец), файл HMM, файл PFB (частота популяции аллели В) и, при необходимости, файл GCModel в качестве входных файлов. Стандартное обозначение числа копий принято как 0: гомозиготная делеция (потеря 2 копий), 1: гетерозиготная делеция, 3: вставка 1 копии, 4: вставка (\geq вставка 2-х копий).

Цель исследования — определение области CNV (CNVR) и потенциально специфических CNV у овец сальской породы овец.

Материалы и методы. Исследования проводили на овцах сальской породы в количестве 48 голов, биологические образцы были отобраны в одном из хозяйств Ростовской области в рамках выполнения диссертационной работы автора [16]. ДНК выделяли из ушных выщипов с использованием набора ДНК-Экстрас-2 (ЗАО «Синтол», Россия) по стандартному протоколу производителя. Проверку качества и количества выделенной тотальной ДНК осуществляли с использованием приборов Quibit 4.0 (Thermo Fisher Scientific, США) и NanoDrop 8000 (Thermo Fisher Scientific, США), соответственно. Концентрации ДНК довели в пределах 20-70 нм/мкл, по рекомендации производителей микрочипов.

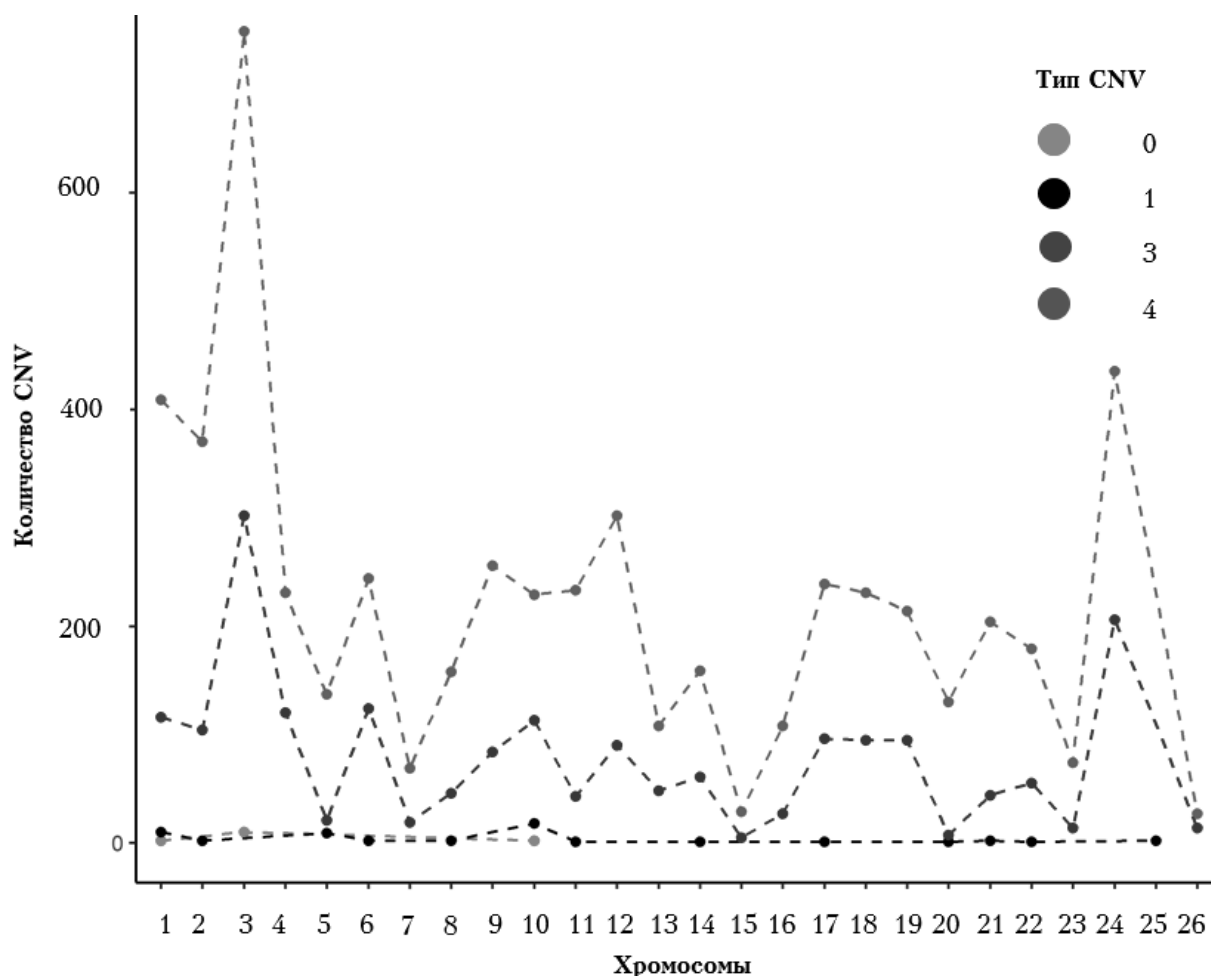


Рис. 1. Распределение делеций и вставок в хромосомах овец сальской породы. Цвета соответствуют видам CNV.

Генотипирования проводили с использованием микрочипов высокой плотности Ovine SNP 600 K BeadChip array (Illumina, США) по стандартному протоколу производителя.

Поиск вариации числа копии (CNV) проводили с использованием программного обеспечения PennCNV на основе референсного генома Oar_v3.1. После определения CNV и областей CNVR они были нормализованы с помощью пакета CNVRanger [17, 18].

Результаты и обсуждение. После применения программы PennCNV и вызова CNV определили 4585 аутосомных CNV (4505 вставок и 80 делеций) у 48 овец сальской породы. Распределения по видам делеций и вставок относительно их локализации по хромосомам представлены на рисунке 1.

Средняя длина CNV составила 92,3 Kb. Было обнаружено, что примерно 21,1 % от общего числа CNV имеют размер менее 50 Kb, 72,8 % CNV имеют размер от 50 до 100 Kb (рис. 2). В среднем, на одно животное определено от 65 до 138 CNV.

После реализации процедуры суммирования вызовов CNV в регионы CNV (CNVR) было получено 527 аутосомных CNVR, представляющих 54,77 Mb всего генома овцы. Средний размер CNVR составил 103,94 Kb и варьируется от 5,2 Kb до 1,21 Mb. CNVR размером от 10 до 500 Kb составляют значительное количество от общего числа CNVR (36,4 %), тогда как CNVR размером более 1 Mb наблюдались редко (1,7%). Количество

событий вставка наблюдается почти в 11 раз больше событий делеция. Оба типа были обнаружены только в семи регионах. Максимальное количество CNVR у одного животного составило 83. На рисунке 3 представлена полногеномная карту CNVR у овец сальской породы (рис 3).

Доля хромосом, охваченная CNVR, варьируется в зависимости от хромосомы: от 0,5 % на OAR25 до 9 % на OAR24. В OAR10, OAR17 и OAR20 наблюдается совокупное присутствие делеций и вставок.

Анализ CNVR в системе ENSEMBL выявил 42 гена, в которых локализованы CNVR, остальные CNVR расположены в межгенных областях. Сопоставление выявленных CNVR с QTL овец из базы данных Animal QTL [19] показало, что 39,24 % генов ответственны за репродуктивные качества, а 20,25 % ответственны за признаки, связанные с шерстью (рис. 4).

Закключение. Таким образом, на основе данных полногеномного генотипирования и метода PennCNV у овец сальской породы определены вариации числа копий (CNV) и регионы CNV (CNVR). Выявлено 4585 аутосомных CNV (4505 вставок и 80 делеций) со средней длиной 92,3 КБ и CNVR, покрывающие около 55 Mb всего генома овцы. Доля хромосом, охваченных CNVR, варьируется от 0,5 % на OAR25 до 9 % на OAR24. Проведена функциональная аннотация CNVR, локализованных в генах и межген-

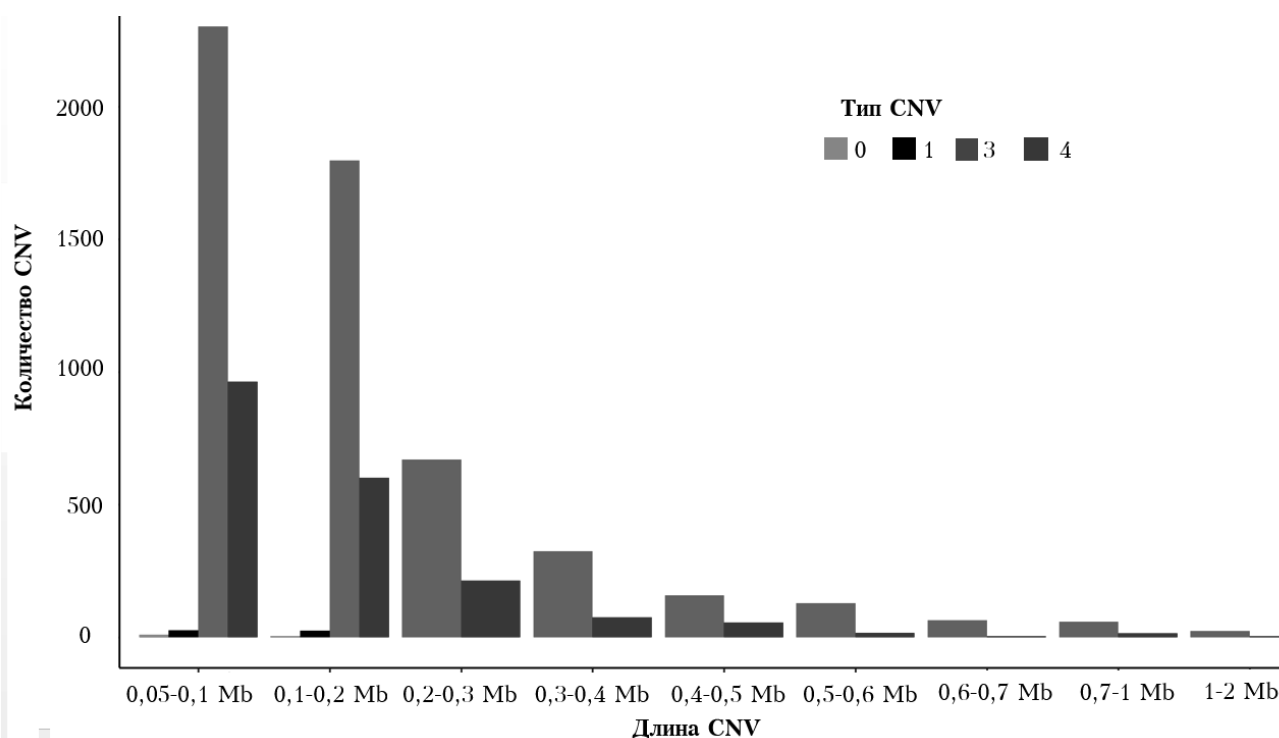


Рис. 2. Распределение CNV по протяженности и по видам.

ных пространствах, и проанализирована представленность общего количества детектированных CNVR в локусах количественных признаков у овец сальской породы. Полученные результаты показали, что CNV локализованы в

генах, связанных с селекционно-значимыми признаками овец. Дальнейшее изучение связи между CNV и продуктивными качествами овец может внести значимый вклад в развитие овцеводства.

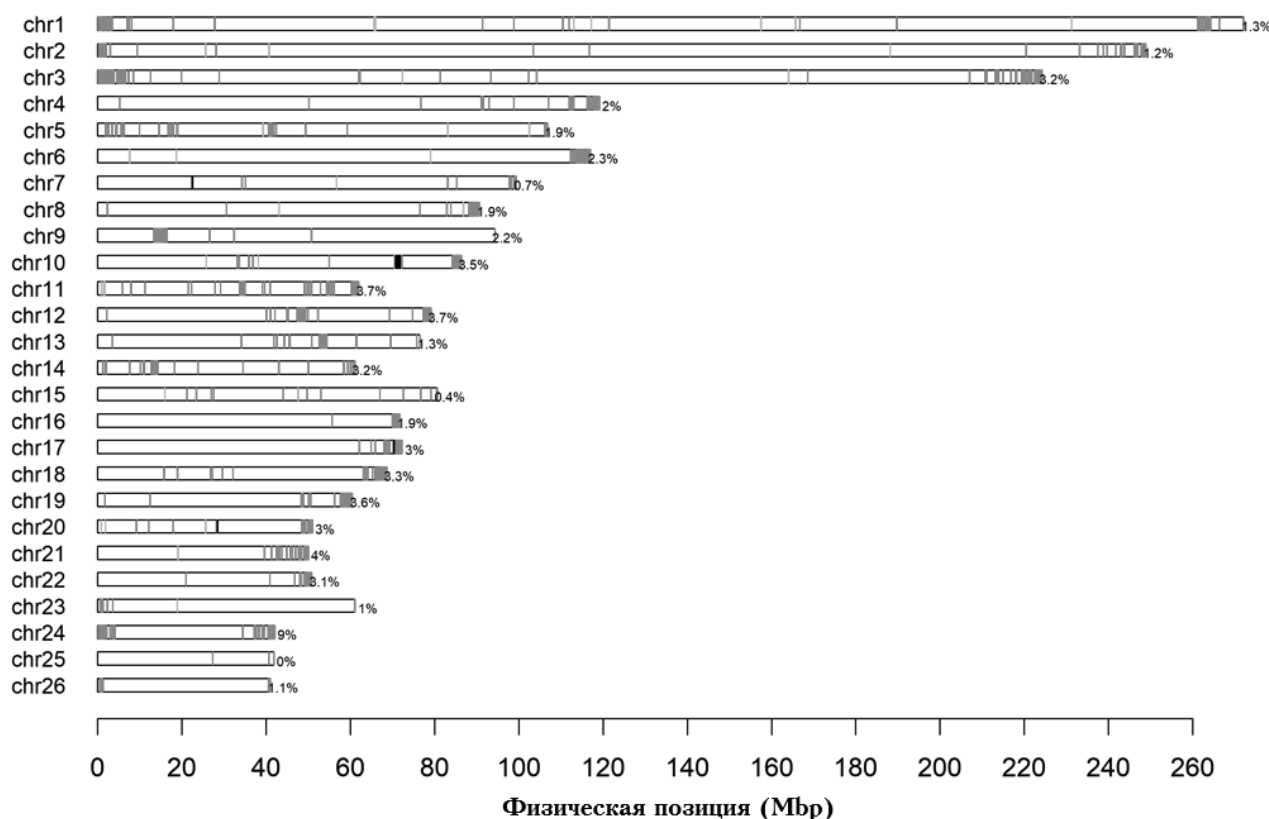


Рис. 3. Полногеномная карта CNVR у овец сальской породы.

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (тема FGGN-2024-0015).

Литература

1. Stranger B. E. Relative impact of nucleotide and copy number variation on gene expression phenotypes / B. E. Stranger, M. S. Forrest et al // Science. — 2007. — Vol. 315 (5813). — P. 848–853. DOI: 10.1126/science.1136678.
2. Кошкина О. А. Вариация числа копий (CNV) как перспективный генетический маркер: распространение, методы валидации и гены-кандидаты в геномах сельскохозяйственных животных (обзор) / О. А. Кошкина, Т. Е. Денискова, Н. А. Зиновьева // Аграрная наука. Евро-Северо-Востока. — 2020. — Т. 21. — № 4. — С. 355–368.
3. Conrad D. Origins and functional impact of copy number variation in the human genome / D. Conrad, D. Pinto et al // Nature. — 2010. — Vol. 464 (7289) — P. 704–12. DOI: 10.1038/nature08516.



Рис. 4. Представленность CNVR в локусах количественных признаков у овец сальской породы

4. Chiang C. / The impact of structural variation on human gene expression / C. Chiang, A. Scott et al // Nat. Genet. — 2017. — Vol. 49(5). — P. 692–699.
5. Bakoev S. Genome-wide analysis of genetic diversity and artificial selection in Large White pigs in Russia / S. Bakoev, L. Getmantseva et al // Peer J. — 2021. — Vol. 9. DOI: 10.7717/peerj.11595.
6. Бакоев С. Ю. Методы оценки инбридинга и подписей селекции сельскохозяйственных животных на основе протяженных гомозиготных областей / С. Ю. Бакоев, Л. В. Гетманцева // Достижения науки и техники АПК. — 2019. — Т. 33. — №11. — С. 63–68. DOI: 10.24411/0235-2451-2019-11114.
7. Sudmant P. Global diversity, population stratification, and selection of human copy-number variation / P. Sudmant, S. Mallick et al // Science. — 2015. — Vol. 349 (6253). DOI: 10.1126/science.aab3761.
8. Fontanesi L. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. / L. Fontanesi, F. Beretti et al // Cytogenet. Genome. Res. — 2009. — Vol. 126 (4). P 333–347. DOI: 10.1159/000268089.
9. Fontanesi L. A first comparative map of copy number variations in the sheep genome / L. Fontanesi, F. Beretti et al // Genomics. — 2011. — Vol. 97(3). — P. 158–165. DOI: 10.1016/j.ygeno.2010.11.005.
10. Peiffer D. A. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping / D. A. Peiffer, J. M. Le et al. // Genome. Res. — 2006. — Vol. 16(9). — P. 1136–1148. DOI: 10.1101/gr.5402306.
11. Wang K. Penn CNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data / K. Wang, M. Li et al // Genome. Res. — 2007. — Vol.17(11). — P. 1665–1674. DOI: 10.1101/gr.6861907.
12. Colella S. Quanti SNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data / S. Colella, C. Yau et al // Nucleic. Acids. Res. — 2007. — Vol. 35(6). — P. 2013–2015. DOI: 10.1093/nar/gkm076.
13. Zhang X. Evaluation of copy number variation detection for a SNP array platform / X. Zhang, R. Du, S. Li, F. Zhang, L. Jin, H. Wang // BMC. Bioinformatics. — 2014. — Vol. 15(50). DOI: 10.1186/1471-2105-15-50.
14. Sanders S. J. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci / S. J. Sanders, X. He et al // Neuron. — 2014. — Vol. 87(6). — P. 1215–1233. <https://doi.org/10.1016/j.neuron.2015.09.016>.
15. Ma Q. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600K SNP array / Q. Ma, X. Liu, J. Pan, L. Ma, Y. Ma, X. He, Q. Zhao, Y. Pu, Y. Li // Sci. Rep. — 2017 — Vol. 7(912). DOI: 10.1038/s41598-017-00847-9.
16. Бакоев Н. Ф. Характеристика генетических и продуктивных особенностей овец тонкорунных пород: диссертация на соискание ученой степени кандидата сельскохозяйственных наук 06.02.07 - Федеральное Государственное бюджетное образовательное учреждение высшего образования «Донской Государственный аграрный университет». — 2022 — 115 с.
17. da Silva V. CNV Ranger: association analysis of CNVs with gene expression and quantitative phenotypes. Bioinformatics / V. da Silva, M. Ramos et al // Bioinformatics. — 2020. — Vol. 36(3) — P. 973–973. DOI: 10.1093/bioinformatics/btz632.
18. Zhou J. Handy CNV: Standardized Summary, Annotation, Comparison, and Visualization of CNV, CNVR and ROH / J. Zhou, L. Liu, T. J. Lopdell, D.J. Garrick, Y. Shi // Front. Genet. — 2021. — Vol. 12 (731355). DOI: 10.3389/fgene.2021.731355.
19. <http://www.animalgenome.org/cgi-bin/QTLdb/OA/browse>.

Bakoev N.

Genome-wide detection and analysis of copy number variations in Sal sheep using a high-density SNP array

Abstract.

Purpose: to identify CNV regions (CNVRs) and identify potentially specific CNVs in Sal sheep.

Materials and methods. The research was carried out on 48 Salsk breed sheep. Genotyping was performed using high-density microarrays Ovine SNP 600 K BeadChip array (Illumina, USA). DNA copy number variation (CNV) searches were performed using PennCNV software based on the Oar_v3.1 reference genome.

Results. We identified 4585 autosomal CNVs in 48 Sal breed sheep. The average CNV length was 92,3 Kb. 21,1 % of the total number of CNVs are less than 50 Kb in size. On average, from 65 to 138 CNVs were identified per animal. 527 autosomal CNVRs were obtained, representing 54,77 Mb of the entire sheep genome. The proportion of chromosomes covered by CNVR varies by chromosome, from 0,5 % on OAR25 to 9 % on OAR24. Analysis of CNVR in the ENSEMBL system revealed 42 genes in which CNVRs are localized; the remaining CNVRs are located in intergenic regions. Comparison of the identified CNVRs with sheep QTLs from the Animal QTL database showed that 39,24 % of the genes were responsible for reproductive traits, and 20,25 % were responsible for wool-related traits.

Conclusions. Functional annotation of CNVRs localized in genes and intergenic spaces was carried out and the total number of detected CNVRs in the QTL database was analyzed. The results showed that CNVs are localized in genes associated with selection-significant traits in sheep.

Key words: Salsk breed; CNV; CNVR; whole genome map.

Author:

Bakoev N. — PhD (Agr. Sci.); e-mail: nekruz82@bk.ru; L. K. Ernst Federal Research Center for Animal Husbandry; Dubrovitsy, 60, Podolsk Municipal District, Moscow Region, 142132, Russia.

References

1. Stranger B. E. Relative impact of nucleotide and copy number variation on gene expression phenotypes / B. E. Stranger, M. S. Forrest et al // Science. — 2007. — Vol. 315 (5813). — P. 848-853. DOI: 10.1126/science.1136678.
2. Koshkina O. A. Copy number variation (CNV) as a promising genetic marker: distribution, validation methods and candidate genes in the genomes of farm animals (review) / O. A. Koshkina, T. E. Deniskova, N. A. Zinovieva // Agrarian science. Euro-North-East. — 2020. — Vol. 21. — № 4. — P. 355–368.
3. Conrad D. Origins and functional impact of copy number variation in the human genome / D. Conrad, D. Pinto et al // Nature. — 2010. — Vol. 464 (7289) — P. 704–712. DOI: 10.1038/nature08516.
4. Chiang C. / The impact of structural variation on human gene expression / C. Chiang, A. Scott et al // Nat. Genet. — 2017. — Vol. 49 (5). — P. 692–699.
5. Bakoev S. Genome-wide analysis of genetic diversity and artificial selection in Large White pigs in Russia / S. Bakoev, L. Getmantseva et al // Peer J. — 2021. — Vol. 9. DOI: 10.7717/peerj.11595.
6. Bakoev S. Yu. Methods for assessing inbreeding and selection signatures of agricultural animals based on extended homozygous regions / S. Yu. Bakoev, L. V. Getmantseva // Achievements of science and technology of the agro-industrial complex. — 2019. — V. 33. — № 11. — P. 63–68. DOI: 10.24411/0235-2451-2019-11114.
7. Sudmant P. Global diversity, population stratification, and selection of human copy-number variation / P. Sudmant, S. Mallick et al // Science. — 2015. — Vol. 349 (6253). DOI: 10.1126/science.aab3761.
8. Fontanesi L. Copy number variation and missense mutations of the agouti signaling protein (ASIP) gene in goat breeds with different coat colors. / L. Fontanesi, F. Beretti et al // Cytogenet. Genome. Res. — 2009. — Vol. 126 (4). P 333–347. DOI: 10.1159/000268089.
9. Fontanesi L. A first comparative map of copy number variations in the sheep genome / L. Fontanesi, F. Beretti et al // Genomics. — 2011. — Vol. 97(3). — P. 158–165. DOI: 10.1016/j.ygeno.2010.11.005.
10. Peiffer D. A. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping / D. A. Peiffer, J. M. Le et al. // Genome. Res. — 2006. — Vol. 16(9). — P. 1136–1148. DOI: 10.1101/gr.5402306.

11. Wang K. Penn CNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data / K. Wang, M. Li et al // *Genome. Res.* — 2007. — Vol.17(11). — P. 1665–1674. DOI: 10.1101/gr.6861907.
12. Colella S. Quanti SNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data / S. Colella, C. Yau et al // *Nucleic. Acids. Res.* — 2007. — Vol. 35(6). — P. 2013–2015. DOI: 10.1093/nar/gkm076.
13. Zhang X. Evaluation of copy number variation detection for a SNP array platform / X. Zhang, R. Du, S. Li, F. Zhang, L. Jin, H. Wang // *BMC. Bioinformatics.* — 2014. — Vol. 15(50). DOI: 10.1186/1471-2105-15-50.
14. Sanders S. J. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci / S. J. Sanders, X. He et al // *Neuron.* — 2014. — Vol. 87(6). — P. 1215–1233. <https://doi.org/10.1016/j.neuron.2015.09.016>.
15. Ma Q. Genome-wide detection of copy number variation in Chinese indigenous sheep using an ovine high-density 600K SNP array / Q. Ma, X. Liu, J. Pan, L. Ma, Y. Ma, X. He, Q. Zhao, Y. Pu, Y. Li // *Sci. Rep.* — 2017 — Vol. 7(912). DOI: 10.1038/s41598-017-00847-9.
16. Bakoev N. F. Characteristics of genetic and productive features of fine-wool sheep breeds: dissertation for the degree of candidate of agricultural sciences 06.02.07 - Federal State Budgetary Educational Institution of Higher Education "Don State Agrarian University". — 2022 — 115 p.
17. da Silva V. CNV Ranger: association analysis of CNVs with gene expression and quantitative phenotypes. *Bioinformatics* / V. da Silva, M. Ramos et al // *Bioinformatics.* — 2020. — Vol. 36(3) — P. 973–973. DOI: 10.1093/bioinformatics/btz632.
18. Zhou J. Handy CNV: Standardized Summary, Annotation, Comparison, and Visualization of CNV, CNVR and ROH / J. Zhou, L. Liu, T. J. Lopdell, D.J. Garrick, Y. Shi // *Front. Genet.* — 2021. — Vol. 12 (731355). DOI: 10.3389/fgene.2021.731355.
19. <http://www.animalgenome.org/cgi-bin/QTLdb/OA/browse>.