

А. А. Кудинов, М. Г. Смарагдов

Выявление QTLs у молочного скота полногеномным ассоциативным анализом

Аннотация. За последние 10 лет в результате полногеномного секвенирования крупного рогатого скота была создана чиповая технология, основанная на однонуклеотидном полиморфизме (SNP) в геноме животных. С помощью полногеномного ассоциативного анализа SNPs — признак нами у голштинского скота Ленинградской области были выявлены гены, которые могут определять потенциальные QTLs. Ассоциативный анализ осуществлен на 280 быках и 500 коровах по признакам удой и выход молочного жира. Все животные были генотипированы чипом BovineSNP50 v/2. Критерии редактирования SNPs были следующие: минорная частота аллелей SNPs не менее 1%; ошибка генотипирования SNPs не более 5%; достоверность отклонения генотипов SNPs от равновесия Харди — Вайнберга ($P < 0.001$). Для племенной оценки животных использовали ANIMAL MODEL. При осуществлении ассоциативного анализа были применены программы Plink 1/9 и EMMAX. Пять SNP преодолели порог достоверности 10⁻⁸ для удоя (признак DYD быков) и 12 SNP для молочного жира (признак YD коров и DYD быков). У быков гены *6-β-N-acetylglucosamintransferase* и *RASA1* идентифицированы как наиболее вероятные претенденты на гены — кандидаты по удою и ген *RASA1* по молочному жиру. Такой результат может быть следствием отрицательной корреляции между удоем и молочным жиром. У коров локализация гена *ZNF704* совпадает с QTL в базе данных Animal Genome, влияющим на выход молочного жира. Ассоциативный анализ племенной ценности с SNP-маркерами подтвердил ожидаемую меньшую достоверность оценки племенной ценности коров по сравнению с быками в случае удоя, но не для молочного жира. Полученные данные свидетельствуют о том, что QTLs, влияющие на выход молочного жира, выявлены как на быках, так и на коровах, причем в разных хромосомах или в разных районах хромосомы 7. Полученные данные следует рассматривать как предварительные, так как выборка животных не превышала 280 быков и 500 коров.

Ключевые слова: QTL, SNPs, WGAS, Animal Model, EMMAX, YD, DYD, молочный жир, удой, молочный скот.

Авторы:

Кудинов Андрей Андреевич — научный сотрудник лаборатории молекулярной организации генома;

Смарагдов Михаил Григорьевич — кандидат биологических наук, старший научный сотрудник лаборатории молекулярной организации генома.

Всероссийский научно-исследовательский институт генетики и разведения сельскохозяйственных животных — филиал Федерального государственного бюджетного научного учреждения «Федеральный научный центр животноводства — ВИЖ имени академика Л. К. Эрнста»; 196601, Россия, Санкт-Петербург, Московское шоссе, 55а.

Введение. Домашние животные, включая крупный рогатый скот, имеют продолжительную историю (тысячи поколений), на протяжении которой были сформированы современные породы. Благодаря искусственной селекции они представляют собой уникальный генетический материал для выявления связи генотип — фенотип. Картирование локусов количественных признаков (QTL) (QTLs — это районы хромосом, которые ответственны за изменчивость количественного признака) у крупного рогатого скота было успешным в отличие от картирования каузальных мутаций, для локализации которых необходима точная локализация каузального SNP. С первого взгляда идентификация таких районов хромосом кажется

простой, учитывая имеющиеся у экспериментаторов геномные и статистические инструменты. К сожалению, эта задача оказалась сложной из-за генетической «архитектуры» QTLs, а именно: наличия малого числа мажорных QTLs и большого числа минорных QTLs с фенотипической вариансой менее 1%.

В восьмидесятые годы XX века была сформулирована методология, позволяющая использовать ДНК — маркеры для картирования локусов количественных признаков [1]. Открытие микросателлитов привело к многочисленным исследованиям по картированию QTLs у крупного рогатого скота. Результаты исследований по выявлению QTLs, ответственных за хозяйственно полезные призна-

ки животных представлены на сайте <http://www.animalgenome.org>.

Недавний прогресс в картировании генов, ответственных за молочную продуктивность, с помощью микросателлитов позволил выявить несколько мажорных генов — *DGAT1*, *GHR*, *ABCG2* [2, 3, 4, 5]. Однако, влияние на количественные признаки большинства генов незначительно и картировать их с помощью микросателлитов невозможно. SNPs маркеры более перспективны для обнаружения QTLs благодаря насыщенности ими генома животных. Полногеномные ассоциированные исследования (GWAS) применяют для идентификации SNPs, маркирующих QTLs. В настоящее время существуют разные подходы для вычисления корреляции SNP — признак. Часто используют алгоритмы для построения непосредственной связи частоты встречаемости SNP со значением признака, например, в программе PLINK [6]. На основе SNP можно сформировать матрицу геномных отношений между животными с последующими расчетами связи SNP — признак с помощью смешанной модели [7]. В своих исследованиях мы использовали оба подхода.

Условия, материалы и методы исследования. Для выявления районов хромосом, влияющих на молочную продуктивность, необходимо оценить племенную ценность коров и быков. В настоящее время ANIMAL MODEL является наилучшей моделью для вычисления племенной ценности быков и коров с максимальной достоверностью. Поэтому мы использовали именно ANIMAL MODEL. Так как вычисленная с помощью ANIMAL MODEL племенная ценность регрессирована к среднему значению, то ее нельзя использовать в ассоциативном анализе. Однако, ANIMAL MODEL позволяет получить данные по отклонению продуктивности Yield Deviation (YD) коров и отклонение продуктивности дочерей Daughter Yield Deviation (DYD) быков, которые не регрессированы к среднему значению.

PTA_{mate} — предсказанная передающая способность.

YD_{prog} — отклонение продуктивности коров.

ДНК быков и коров была выделена фенол — хлороформным методом. Генотипирование 280 быков и 500 коров осуществляли чипом Illumina Bovine SNP 50 V.2 BeadChip. Для редактирования SNPs использовали PLINK 1.9 [6]. В результате были отобраны животные с ошибкой генотипирования менее 5%, MAF < 0.01, с отклонением генотипов SNPs от равновесия Харди-Вайнберга с $P < 10^{-4}$. Окончательное число SNPs составило 37266. Исходя из отобранных SNPs, был осуществлен ассоциативный анализ для коров с помощью пакета программ PLINK 1.9 и для быков EM-MAX [7]. Поиск районов хромосом (генов), где локализуются интересующие нас SNP, проводили на WEB сайте NCBI <https://www.ncbi.nlm.nih.gov/pubmed>.

Таблица 1. Результат ассоциативного анализа SNPs — признак (отклонения дочерей быков по удою (DYD))

Хромосома	SNP	P- value
10	ARS-BFGL-NGS-32212	2.314e-10
2	BTA-47843	1.217e-09
7	BTB-00325965	1.132e-08
19	ARS-BFGL-NGS-279	1.588e-08
19	ARS-BFGL-NGS-25991	5.32e-08

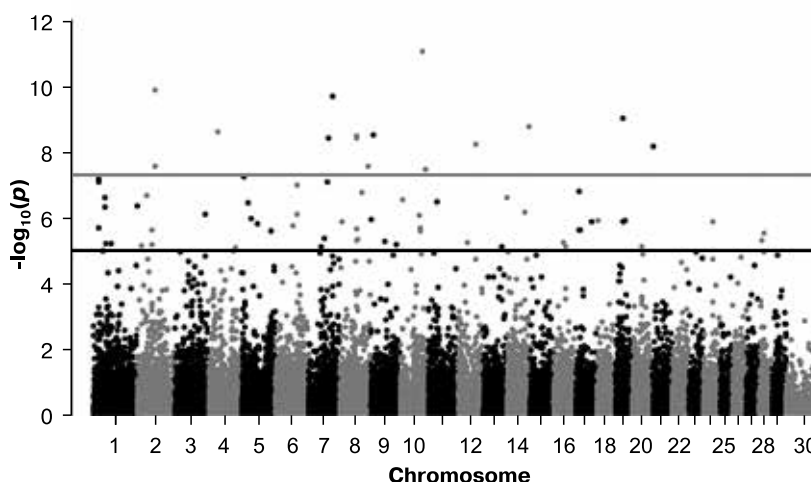


Рис. 1. Визуализация данных ТОП-SNPs согласно таблице 1

$$DYD = \sum q_{prog} w_{2_{prog}} (YD_{prog} - PTA_{mate}) / \sum q_{prog} w_{2_{prog}}$$

q_{prog} — равно 1, если родитель потомства известен, и 2/3, если родитель неизвестен.

$w_{2_{prog}}$ — весовой коэффициент.

Результаты и обсуждение. Наиболее достоверные результаты по ассоциации SNP — удои при использовании признака DYD быков представлены в таблице 1. На рисунке 1 видно, в каких хромосомах быков SNPs преодолевают порог

Таблица 2. Локализация ТОП-SNPs в геноме быков (DYD по удою)

SNP	Хромосома	Ген	Локализация SNP в хромосоме	Ближайший QTL
ARS-BFGL-NGS-32212	10	RGS6Регулятор сигнального пути 6 G-белка	83.8 Mb	90.3 Mb
BTA-47843	2	<i>6-β-N-acetylglucosamintransferase</i>	63.5 Mb	63.9 Mb
BTB-00325965	7	<i>RASA1 GTPase-activating-protein</i>	89.2 Mb	88.9 Mb
ARS-BFGL-NGS-279	19	<i>GAS7 growth arrest-specific 7</i>	29.6 Mb	25.7 Mb
ARS-BFGL-NGS-25991	19	RCVRNrecoverin	29.6 Mb	25.7 Mb

Таблица 3. Результат ассоциативного анализа SNPs — признак (отклонения по удою (YD) коров)

Хромосома	SNP	Достоверность
15	Нармаp40723-BTA-37075	0.0001098
15	ARS-BFGL-NGS-54493	0.0001277
4	ARS-BFGL-NGS-90892	0.0001672
4	ARS-BFGL-NGS-38191	0.0002187
4	ARS-BFGL-NGS-115125	0.0002213
16	BTB-01422500	0.0002788
17	Нармаp57633-rs29027409	0.0002795

Таблица 4. Результат ассоциативного анализа SNPs — признак (отклонения дочерей быков (DYD) по молочному жиру)

Хромосома	SNP	Достоверность
7	BTB-00325965	8.034e-10
19	ARS-BFGL-NGS-25991	2.348e-09
19	ARS-BFGL-NGS-279	2.348e-09
10	ARS-BFGL-NGS-32212	3.638e-09
8	ARS-BFGL-NGS-20552	7.609e-09
8	ARS-BFGL-NGS-27060	8.244e-09

достоверности (горизонтальные линии). В таблице 2 идентифицированы гены, в которых или рядом с которыми локализуются соответствующие таблице 1 SNPs. Гены *6-β-N-acetylglucosamintransferase* и *GT Pase-activating-protein* наиболее близко расположены к QTLs, полученных из базы данных Animal Genome. Следовательно, эти гены являются наиболее вероятными претендентами, обуславливающими эффекты QTLs.

Для удою коров ни один из SNPs не преодолел порог достоверности $P < 10^{-8}$ (Табл. 3). Такой результат является следствием в 3–5 раз меньшей достоверности племенной оценки коров по сравнению с быками.

В таблице 4 представлены данные по ассоциативному анализу SNPs — молочный жир, исходя из DYD быков. На рисунке 2 наглядно видно, в какой хромосоме SNPs преодолевают порог достоверности $P \leq 10^{-8}$. В таблице 5 идентифицированы ге-

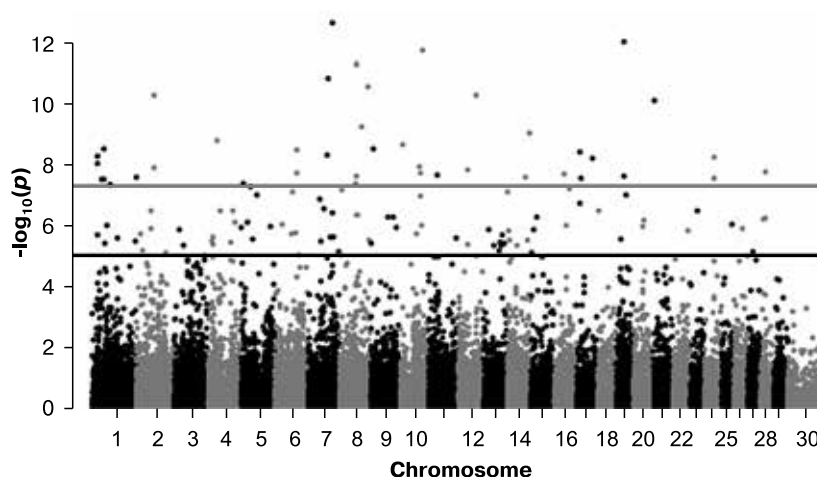


Рис. 2. Визуализация данных ТОП-SNPs согласно таблице 4

Таблица 5. Локализация ТОП-SNPs в геноме быков (DYD по молочному жиру

SNP	Хромосома	Ген	Локализация SNP в хромосоме	Ближайший QTL
BTB-00325965	7	RASA1 <i>GTPase активирующий белок</i>	89.2 Mb	88.9 Mb
ARS-BFGL-NGS-25991	19	<i>RCVRN</i> recoverin	29.6 Mb	25.7 Mb
ARS-BFGL-NGS-279	19	<i>GAS7</i> Ограничивающий рост клеток белок 7	29.6 Mb	25.7 Mb
ARS-BFGL-NGS-32212	10	<i>RGS6</i> Регулятор сигнальных путей 6 для белка G	83.8 Mb	90.3 Mb
ARS-BFGL-NGS-20552	8	<i>FRMPD1</i> Регулятор внутриклеточной локализации активатора G-белка	62.2 Mb	55.1 Mb
ARS-BFGL-NGS-27060	8	F-box protein 10protein-ubiquitin ligases	62.1 Mb	55.1 Mb

ны, в которых локализованы соответствующие таблице 4 SNPs. Для гена *RASA1* наиболее близко расположен QTL, полученный из базы данных Animal Genome, ответственный за молочный жир.

Для молочного жира (YD) коров данные по ассоциации этого признака с SNPs представлены в таблице 6. Графическое их отображение находится на рисунке 3 и гены в таблице 7. Из всех выявленных генов, представленных в таблице 7, только локализация гена *ZNF704* совпадает с QTL из базы данных Animal Genome.

Выводы. Ассоциативный анализ племенной ценности с SNP-маркерами подтвердил ожидаемую меньшую достоверность оценки племенной ценности коров по сравнению

Таблица 6. Результат ассоциативного анализа SNPs признаком YD коров по молочному жиру

Хромосома	SNP	Достоверность
11	ARS-BFGL-NGS-112032	2.18e-29
7	BTB-00299019	6.002e-18
7	BTB-00298967	6.002e-18
14	BTB-00565043	6.002e-18
14	ARS-BFGL-NGS-49187	6.002e-18
14	Нарmap55355-rs29015092	6.002e-18

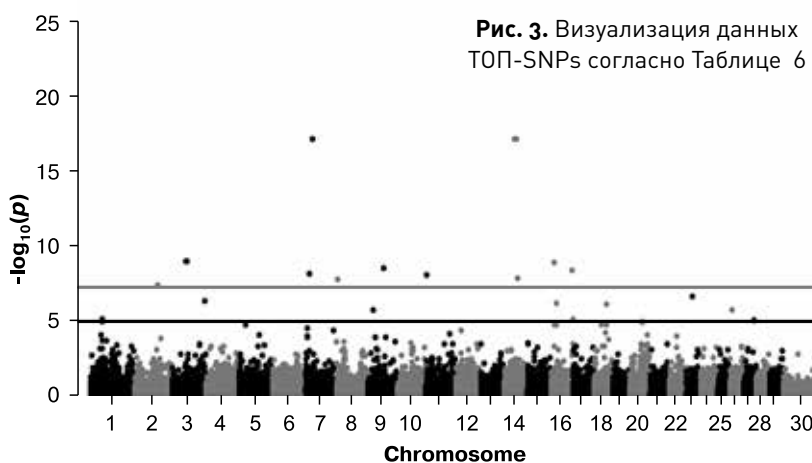


Рис. 3. Визуализация данных ТОП-SNPs согласно Таблице 6

Таблица 7. Локализация ТОП-SNPs в геноме коров (признак — молочный жир)

SNP	Хромосома	Ген	Локализация SNP в хромосоме	Ближайший QTL
ARS-BFGL-NGS-112032	11	<i>LHCGR</i> Luteinizing hormone/chorionadotropin receptor	30.9 Mb	27.5 Mb
BTB-00299019	7	<i>FNIP1</i> Связывающийся с фолликулом белок 1	24 Mb	32 Mb
BTB-00298967	7	<i>RAPGEF6</i> Rap гуанин-нуклеотид обменивающийся фактор 6	24.7 Mb	32 Mb
BTB-00565043	14	<i>UBE2W</i> Убиквитин конъюгирующий фермент E2W	39.2 Mb	46.6
ARS-BFGL-NGS-49187	14	<i>ZNF704</i> Подобный цинковым пальцам белок 704	46 Mb	46 Mb
Нарmap55355rs29015092	14	TCEB1Транскрипционный фактор элонгации В, полипептид 1	39.3 Mb	46 Mb

с быками в случае удоя, но не для молочного жира (Табл. 3 vs. Табл. 6). Полученные данные свидетельствуют о том, что QTLs, влияющие на выход молочного жира, были выявлены как на быках, так и на коровах, причем в разных хромосомах или в разных районах хромосомы 7 (Табл. 5 и 7). Гены *6-β-N-acetylglucosamintransferase* и *RASA1* идентифицированы как наиболее вероятные претенденты на QTL у быков по удою и ген *RASA1* по молочному жиру (Табл. 2 и 5). Такой

результат может быть следствием отрицательной корреляции между удоем и молочным жиром. Локализация генов *6-b-N-acetylglucosamintransferase* и *RASA1* совпадает с QTL идентифицированными J. Cole et al. (Cole et al. 2011) (Табл. 2). Такой же вывод относится к гену *ZNF704* (Табл. 7).

Полученные данные следует рассматривать как предварительные, так как выборка животных не превышала 280 быков и 500 коров.

Работа выполнена при финансовой поддержке ФАНО (Госзадание № АААА-А18-118021590138-1)

Литература

1. Weller J. I. Genomic selection in animals / J. I. Weller // WILEY Blackwell. — 2016. — 174 p.
2. Grisart B. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition / B. Grisart et al. // Genome Research. — 2002. — V. 12. — P. 222–31.
3. Grisart B. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait-nucleotide in affecting milk yield and composition / B. Grisart et al. // Proc. Natl. Acad. Sci. U. S. A. — 2004. — V. 101. — P. 2398–403.
4. Blott S. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition / S. Blott, J. J. Kim, S. Moiso, A. Schmidt-Kuntzel, A. Cornet, P. Berzi et al. // Genetics. — 2003. — V. 163. — P. 253–66.
5. Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loo, J. J., Everts-van der Wind, A., Lee, J. H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I., and Ron, M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle / M. Cohen-Zinder // Genome Research. — 2005. — V. 15. — P. 936–944.
6. Purcell S. PLINK: a tool set for whole genome association and population based linkage analyses / S. Purcell // American Journal of Human Genetics. — 2007. — V. 81. — P. 559–575.
7. Kang H. M., Sul JH, Zaitlen NA, Kong S, Freimer NB, Sabatti C, et al. Variance component model to account for sample structure in genome-wide association studies / H M Kang // Nature Genetics. — 2010. — V. 42. — P. 348–54.
8. Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A, et al. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle / S. Viitala // Genetics. — 2006. — V. 173. — P. 2151–64.
9. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK, Da Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows / JB. Cole // BMC Genomics. — 2011. — 12:408.

Kudinov A. A., Smaragdov M. G.

Identification of QTLs in dairy cattle by wide-genome associative analysis

Abstract. Over the past 10 years, as a result of full genomic sequencing of cattle, a chip technology based on single nucleotide polymorphism (SNP) in the animal genome was created. With the help of a genome — wide comparative analysis of Holstein cattle of the Leningrad Region have been identified genes that can be

responsible for potential QTLs. All animals were genotyped by the BovineSNP50 v.2 chip. The criteria for editing SNPs were as follows: The minor alleles frequency of SNPs was < 1%. The genotyping error of SNPs is not more than 5%. Reliability of SNPs genotypes deviation from Hardy-Weinberg equilibrium was ($P < 0.001$). ANIMAL MODEL was used for breeding values evaluation. In carrying out the associative analysis, the Plink 1.9 and EMMAX programs were applied. Associative analysis was carried out on 280 bulls and 500 cows including traits of milk yield and milk fat yield, which were evaluated with ANIMAL MODEL. Five SNPs overcome the confidence threshold of 10⁻⁸ for milk yield (DYD of the bulls) and 12 SNP for milk fat yield (YD of the cows and DYD of the bulls). In bulls, the genes 6- β -N-acetylglucosamintransferase and RASA1 are identified as the most likely candidates for QTL responsible for milk yield and the gene RASA1 for milk fat yield. It may be the result of a negative correlation between milk yield and milk fat yield. In cows, the localization of the ZNF704 gene coincides with QTL in the Animal Genome database affecting of the milk fat yield. Associative analysis of BV with SNP markers confirmed the expected lower reliability of cows' breeding value evaluation in comparison with bulls ones in case of milk yield, but not for milk fat yield. The obtained data indicate that QTLs affecting the of milk fat yield were detected both in bulls and in cows, and in different chromosomes or in different regions of chromosome 7. The data obtained should be considered as preliminary, since the sample of animals did not exceed 280 bulls and 500 cows.

Keywords: QTL, SNPs, WGAS, Animal Model, EMMAX, YD, DYD, milk fat, milk yield, milk cattle.

Authors:

Kudinov A. A. — researcher of laboratory of molecular organization of the genome; e-mail: kudinov_aa@list.ru;

Smaragdov M. G. — PhD (Biol. Sci.), senior researcher of laboratory of molecular organization of the genome; e-mail: mik7252@yandex.ru.

Russian research institute of farm animal genetics and breeding -branch of the L.K. Ernst Federal science center for animal husbandry; St. Petersburg, Pushkin, Moskovskoe shosse, 55a, 196601

References

1. Weller J. I. Genomic selection in animals / J. I. Weller // WILEY Blackwell. — 2016. — 174 p.
2. Grisart B. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition / B. Grisart et al. // Genome Research. — 2002. — V. 12. — P. 222–31.
3. Grisart B. Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait-nucleotide in affecting milk yield and composition / B. Grisart et al. // Proc. Natl. Acad. Sci. U. S. A. — 2004. — V. 101. — P. 2398–403.
4. Blott S. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosinesubstitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition / S. Blott, J. J. Kim, S. Moiso, A. Schmidt-Kuntzel, A. Cornet, P. Berzi et al. // Genetics. — 2003. — V. 163. — P. 253–66.
5. Cohen-Zinder, M., Seroussi, E., Larkin, D. M., Loo, J. J., Everts-van der Wind, A., Lee, J. H., Drackley, J. K., Band, M. R., Hernandez, A. G., Shani, M., Lewin, H. A., Weller, J. I., and Ron, M. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle / M. Cohen-Zinder // Genome Research. — 2005. — V. 15. — P. 936–944.
6. Purcell S. PLINK: a tool set for whole genome association and population based linkage analyses / S. Purcell // American Journal of Human Genetics. — 2007. — V. 81. — P. 559–575.
7. Kang H. M., Sul JH, Zaitlen NA, Kong S, Freimer NB, Sabatti C, et al. Variance component model to account for sample structure in genome-wide association studies / HM Kang // Nature Genetics. — 2010. — V. 42. — P. 348–54.
8. Viitala S, Szyda J, Blott S, Schulman N, Lidauer M, Maki-Tanila A, et al. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle / S. Viitala // Genetics. — 2006. — V. 173. — P. 2151–64.
9. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK, Da Y. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows / JB. Cole // BMC Genomics. — 2011. — 12:408.