

ISSN 1561-8323 (Print)
ISSN 2524-2431 (Online)

УДК 636.1.082:575(571.56)
<https://doi.org/10.29235/1561-8323-2022-66-3-301-309>

Поступило в редакцию 31.03.2022
Received 31.03.2022

В. Н. Кипень, Е. В. Снытков, М. Е. Михайлова, член-корреспондент Р. И. Шейко

Институт генетики и цитологии Национальной академии наук Беларуси, Минск, Республика Беларусь

ДИФФЕРЕНЦИАЦИЯ ПОРОД ДОМАШНИХ СВИНЕЙ С ИСПОЛЬЗОВАНИЕМ РАСШИРЕННОГО БИОИНФОРМАТИЧЕСКОГО АНАЛИЗА SNP

Аннотация. С использованием методов биоинформатики проведен анализ данных по секвенированию геномов особей вида *Sus scrofa domestica*, которые расположены в базе Sequence Read Archive (NCBI-SRA). Определены *in silico* генотипы для пяти пород домашних свиней – дюрок, ландрас, пьетрен, крупная белая и йоркшир с помощью алгоритма, разработанного на языке программирования Python. На основании двухстадийного биоинформатического анализа определен широкий перечень SNP с высоким потенциалом для дифференциации. Полученные результаты будут использованы при создании экспресс-методов для определения чистопородности свиней данных пород. Расширенный биоинформатический анализ, который включал в себя определение генотипа по 7451 SNP для 248 геномов *Sus scrofa domestica*, позволил выявить суммарно 393 SNP для всех пород, для которых имеется существенная разница в частоте альтернативных аллелей у пород свиней дюрок, ландрас, пьетрен, крупная белая и йоркшир. Обозначены кластеры в пределах хромосом, в которых плотность SNP с высоким дифференцирующим потенциалом наиболее высока. Для свиней породы дюрок нами выявлены 184 SNP, имеющие дифференцирующий потенциал, для 24 из которых показан высокий дифференцирующий потенциал, для свиней породы ландрас – 52 SNP и 7, для свиней породы пьетрен – 39 и 9, для свиней породы крупная белая – 104 и 22, для свиней породы йоркшир – 14 и 5 соответственно.

Ключевые слова: *Sus scrofa domestica*, дифференциация, однонуклеотидный полиморфизм, породоспецифичность, дюрок, ландрас, пьетрен, крупная белая, йоркшир, ROC-анализ, *in silico*

Для цитирования. Дифференциация пород домашних свиней с использованием расширенного биоинформатического анализа SNP / В. Н. Кипень [и др.] // Докл. Нац. акад. наук Беларуси. – 2022. – Т. 66, № 3. – С. 301–309. <https://doi.org/10.29235/1561-8323-2022-66-3-301-309>

Viachaslau N. Kipen, Evgenij V. Snytkov, Mariya E. Mikhailova, Corresponding Member Ruslan I. Sheyko

Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

BREED DIFFERENTIATION OF DOMESTIC PIGS USING SNP – EXTENDED BIOINFORMATICAL ANALYSIS

Abstract. Using the methods of bioinformatics, the analysis of data on sequencing of the genomes of individuals of the species *Sus scrofa domestica*, which are located in the Sequence Read Archive (NCBI-SRA) database, was carried out. Genotypes were determined *in silico* for five breeds of domestic pigs – Duroc, Landrace, Pietrain, Large White and Yorkshire using an algorithm developed in the Python programming language. Based on a two-stage bioinformatics analysis, a wide range of SNPs with a high potential for differentiation was identified. The results obtained will be used to create express methods for determining the purity of pigs of these breeds. Extended bioinformatics analysis, which included genotyping by 7451 SNPs for 248 *Sus scrofa domestica* genomes, revealed a total of 393 SNPs for all breeds for which there is a significant difference in the frequency of alternative alleles in Duroc, Landrace, Pietrain, Large White and Yorkshire pig breeds. Clusters within chromosomes are indicated, in which the density of SNPs with a high differentiating potential is the highest. For Duroc pigs, we identified 184 SNPs with differentiating potential, 24 of which showed a high differentiating potential, for Landrace pigs – 52 SNPs and 7, for Pietrain pigs – 39 and 9, for Large White pigs – 104 and 22, for Yorkshire pigs – 14 and 5, respectively.

Keywords: *Sus scrofa domestica*, differentiation, single nucleotide polymorphism, breed specificity, Duroc, Landrace, Pietrain, Large White, Yorkshire, ROC analysis, *in silico*

For citation. Kipen V. N., Snytkov E. V., Mikhailova M. E., Sheyko R. I. Breed differentiation of domestic pigs using SNP – extended bioinformatical analysis. *Doklady Natsional'noi akademii nauk Belarusi = Doklady of the National Academy of Sciences of Belarus*, 2022, vol. 66, no. 3, pp. 301–309 (in Russian). <https://doi.org/10.29235/1561-8323-2022-66-3-301-309>

Введение. Одна из приоритетных задач для селекции сельскохозяйственных животных – поддержание и совершенствование породных качеств. Это достигается за счет умелого подбора пар для скрещивания, поддержания линий в пределах породы и межлинейных кроссов. Как итог, сохранение генетического разнообразия дает селекционерам возможность совершенствовать

породу в заданном направлении вплоть до создания новых пород. Также чистопородное разведение заводских пород обеспечивает доступность высокоценного улучшающего племенного материала для товарного животноводства. Таким образом, определение чистопородности сельскохозяйственных животных имеет ключевое значение для всей отрасли животноводства.

SNP (Single Nucleotide Polymorphism) – эффективные и высокочувствительные маркеры для выявления генетической подразделенности пород и популяций. Данный тип генетических маркеров используется в качестве инструмента контроля происхождения животных, для выявления степени генетических различий между породами, типами, стадами и генеалогическими группами животных.

Определение чистопородности свиней (*Sus scrofa domestica*) также может быть проведено с использованием SNP. Технология мультиплексирования от компании Illumina[®] позволяет за один запуск прибора iScan[™] System охарактеризовать десятки тысяч SNP одновременно для сотен животных. О возможностях совмещения информации о PorcineSNP60 BeadChip с SRA-файлами, находящимися в открытом доступе, было подробно описано в наших предыдущих работах [1; 2].

Цель данной работы – с использованием современных биоинформатических и статистических методов провести расширенный анализ результатов секвенирования генома *Sus scrofa domestica* и определить SNP со значительным потенциалом для дифференциации пород домашних свиней дюрок, ландрас, пьетрен, крупная белая и йоркшир.

Материалы и методы исследования. *SNP и образцы.* Биоинформатический анализ включал два последовательных этапа. На первом этапе был сформирован максимально широкий перечень потенциально информативных SNP, в который вошли как ранее описанные SNP [2], так и фланкирующие их SNP из Axiom[®] Porcine Genotyping Array (Axiom_PigHDv1) от Affymetrix[®], а также SNP, для которых по состоянию на начало 2021 г. имелась доказанная ассоциация с QTL (Quantitative Trait Locus) и была представлена информация в Pig Quantitative Trait Locus Database (Pig QTLdb (<https://www.animalgenome.org>)). Общее количество SNP в анализе на первом этапе – 7451. Секвенированные нуклеотидные последовательности особей *Sus scrofa domestica*, которые анализировали на данном этапе, были представлены в формате *.fasta и имели статус «aligned» (выровненные) для каждой хромосомы. Общее количество образцов в анализе на первом этапе – 70 (дюрок – 16, ландрас – 22, пьетрен – 7, крупная белая – 25). На втором этапе был сформирован сокращенный перечень SNP, для которых с использованием ROC-анализа (Receiver Operating Characteristic) были выявлены статистически значимые ассоциации ($p < 0,001$). Общее количество SNP в анализе на втором этапе – 261. Секвенированные нуклеотидные последовательности особей *Sus scrofa domestica*, которые анализировали на данном этапе, были представлены в формате SRA. Общее количество образцов в анализе на втором этапе – 178 (дюрок – 69, ландрас – 24, пьетрен – 21, крупная белая – 45, йоркшир – 19). Таким образом, суммарно в исследовании были проанализированы данные 248 геномов особей *Sus scrofa domestica* (дюрок – 85, ландрас – 46, пьетрен – 28, крупная белая – 70, йоркшир – 19) по 7451 SNP.

Данные о секвенированных нуклеотидных последовательностях полных геномов *Sus scrofa domestica* расположены в базе Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>): PRJNA41185 (год регистрации проекта в BioProject – 2009, University of Aarhus, Дания), PRJNA176478 (2012, Uppsala University, Швеция), PRJNA186497 (2013, Novogene, США), PRJEB1683 (2013, Wageningen University & Research, Нидерланды), PRJNA239399 (2014, Agricultural Biotechnology Center, Венгрия), PRJNA260763 (2014, Seoul National University и NICEM – Южная Корея, BGI – КНР), PRJNA255085 (2014, Centre For Research in Agrigenomics, Испания), PRJEB9922 (2015, Wageningen University & Research, Нидерланды), PRJNA309108 (2016, Novogene, США), PRJNA322309 (2016, China Agricultural University, КНР), PRJNA343658 (2016, USDA-ARS-USMARC, США), PRJNA358108 (2016, China Agricultural University, КНР), PRJNA369600 (2017, Universitat Autònoma de Barcelona, Испания), PRJNA378496 (2017, China Agricultural University, КНР), PRJNA393920 (2017, INRA, Франция), PRJNA487172 (2018, Huazhong Agricultural University, КНР), PRJNA506339 (2018, University of Edinburgh, Шотландия), PRJNA507853 (2018, Sichuan Agricultural University, КНР), PRJNA485589 (2018, Agricultural Genome Institute at Shenzhen, КНР), PRJNA488960 (2018,

Huazhong Agriculture University, KHP), PRJNA550237 (2019, Jiangxi Agricultural University, KHP), PRJNA520978 (2019, Centre For Researchin Agrigenomics, Испания), PRJNA553106 (2019, Georg-August-University Goettingen, Германия), PRJNA671763 (2020, Chinese Academy of Agricultural Sciences, KHP), PRJNA626370 (2020, Centre for researchin agricultural Genomics, Испания), PRJNA622908 (2020, ETH Zurich, Швейцария).

Определение генотипа in silico. Для биоинформатического анализа были использованы геномы животных, представленные в открытом доступе в формате SRA, которые дополнительно конвертировали в формат *.fasta с использованием пакета SRA-Toolkit v.2.11. Для автоматизации процесса поиска нуклеотидных последовательностей *in silico*, фланкирующих искомый аллель, использовали программу, написанную на языке программирования Python v.3.10, в среде разработки программного обеспечения Jupyter Notebook. Хромосомная позиция SNP определена в версии сборки генома Sscrofa11.1 (GCF_000003025.6).

Статистический анализ данных. Дифференцирующий потенциал SNP определяли с использованием ROC-анализа в SPSS v.20.0. Количественную интерпретацию ROC дает показатель AUC (area under ROC curve) – площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. При наличии нижней границы асимптотического 95 %-ного доверительного интервала, значение которой более 0,5, для параметра AUC (площадь под кривой) SNP позиционировалось как генетический маркер со значительным дифференцирующим потенциалом, более 0,7 – как генетический маркер с высоким дифференцирующим потенциалом.

Результаты и их обсуждение. Проведенный биоинформатический анализ, направленный на определение генотипа *in silico* для животных вида *Sus scrofa domestica*, позволил рассчитать частоты генотипов у пяти пород свиней – дюрок (выборка «DU»), ландрас («LA»), пьетрен (PI), крупная белая («LW») и йоркшир («YO»). Полученные результаты легли в основу математического анализа классификаций с применением ROC-кривых.

По причине наличия значительного практического материала, полученного в ходе выполнения данного исследования, считаем не возможным подробно предоставить все наши результаты. Подробный анализ для каждой породы свиней будет осуществлен в рамках отдельных публикаций. В то же время в сообщении представлены данные об участках в геноме *Sus scrofa domestica*, где в ходе селекции каждой породы сформировались кластеры (гаплотипы) SNP, молекулярно-генетический анализ которых позволит оценить чистопородность конкретной особи или наличие примесей других пород в исследуемой выборке животных. Сводная информация по наличию таких наиболее информативных кластеров представлена в таблице.

Для свиней породы дюрок нами выявлены 184 SNP со значительным дифференцирующим потенциалом ($p < 0,01$), для 24 из которых показан высокий дифференцирующий потенциал ($p < 7,50E-12$, нижняя граница (НГ) 95 %-ного доверительного интервала (ДИ) $>0,7$). На хромосоме 9 выявлен фрагмент длиной около 50 тыс. п. н., на котором определены четыре SNP (Chr.9:49010671, Chr.9:49034938, Chr.9:49042137 и Chr.9:49046617) с высоким дифференцирующим потенциалом. На данном участке хромосомы или непосредственно рядом с ним расположен ген *SORL1* (sortilin related receptor 1, NCBI Gene ID – 100626812), гомологичный ген у человека относится к семейству рецепторов липопротеинов низкой плотности (LDLR), кодируемый им протеин принимает участие в эндоцитозе. Также в этом фрагменте расположены локусы LOC110255448 и LOC102164776, функциональная роль которых на данный момент неизвестна, а также некодирующие РНК, которые участвуют в посттранскрипционной регуляции экспрессии генов, влияя как на стабильность, так и на трансляцию мРНК – microRNA mir-125b-1, microRNA let-7a-1, microRNA mir-100 и microRNA mir-9804. В нашей предыдущей работе был выявлен фрагмент длиной около 250 тыс. п. н. на хромосоме 14, на котором также определены SNP с высоким дифференцирующим потенциалом для породы свиней дюрок. На данном участке хромосомы или непосредственно рядом с ним расположены гены: *AICF*, *PRKGI*, *ASAH2* и *SGMS1* [1]. В ходе расширенного биоинформатического анализа нами выявлены дополнительные SNP с высоким дифференцирующим потенциалом в данном кластере – Chr.14:107457741, Chr.14:107794160 и Chr.14:107869191.

Расположение высокоинформативных SNP и их кластеров в геноме *Sus scrofa domestica* для дифференциации пород свиней дюрок (DU), ландрас (LA), пьетрен (PI), крупная белая (LW) и йоркшир (YO)

Schematic arrangement of highly informative regions in the *Sus scrofa domestica* genome for differentiation of Duroc (DU), Landrace (LA), Pietrain (PI), Large White (LW) and Yorkshire (YO) pig breeds

Порода Breed	Хромосомная позиция Chromosomal position	AUC	HG 95 % ДИ	ВГ 95 % ДИ	<i>p</i> -уровень <i>p</i> -level	Порода Breed	Хромосомная позиция Chromosomal position	AUC	HG 95 % ДИ	ВГ 95 % ДИ	<i>p</i> -уровень <i>p</i> -level
DU	Chr.4:9676649	0,7953	0,7434	0,8473	8,12E-15	LW	Chr.8:47482649	0,7037	0,6319	0,7755	4,22E-07
DU	Chr.7:106301845	0,8255	0,7710	0,8800	1,13E-17	LW	Chr.9:138661524	0,7090	0,6422	0,7757	2,11E-07
DU	Chr.8:45485535	0,7803	0,7255	0,8352	1,69E-13	LW	Chr.14:73886405	0,6926	0,6194	0,7657	1,74E-06
DU	Chr.8:47482649	0,8231	0,7718	0,8744	1,94E-17	LW	Chr.14:93947383	0,6852	0,6112	0,7593	4,23E-06
DU	Chr.9:25438712	0,7668	0,7122	0,8214	2,28E-12	LW	Chr.14:93968978	0,7153	0,6450	0,7856	9,01E-08
DU	Chr.9:49010671*	0,7958	0,7405	0,8511	7,34E-15	LW	Chr.14:98859586	0,7312	0,6625	0,7998	9,46E-09
DU	Chr.9:49034938	0,7721	0,7149	0,8293	8,39E-13	LW	Chr.14:107298278	0,6879	0,6169	0,7590	3,06E-06
DU	Chr.9:49042137	0,7737	0,7176	0,8298	6,14E-13	LW	Chr.14:107457741	0,6766	0,6041	0,7492	1,15E-05
DU	Chr.9:49046617	0,7729	0,7167	0,8291	7,22E-13	LW	Chr.15:84480210	0,6933	0,6209	0,7657	1,59E-06
DU	Chr.9:138661524	0,7604	0,7047	0,8161	7,50E-12	LW	Chr.15:84509213	0,6850	0,6160	0,7540	4,36E-06
DU	Chr.12:11541028	0,7875	0,7305	0,8445	4,01E-14	LW	Chr.X:24931575	0,7022	0,6317	0,7728	5,12E-07
DU	Chr.13:182876924	0,7892	0,7324	0,8461	2,83E-14	LW	Chr.X:53635886	0,7323	0,6605	0,8042	7,97E-09
DU	Chr.14:73886405	0,7681	0,7109	0,8253	1,80E-12	PI	Chr.3:64788078	0,7409	0,6584	0,8235	2,29E-05
DU	Chr.14:92116337	0,8149	0,7585	0,8713	1,22E-16	PI	Chr.3:70293190	0,6862	0,5756	0,7969	1,06E-03
DU	Chr.14:98859586	0,7540	0,6965	0,8115	2,41E-11	PI	Chr.5:45368146	0,2663	0,1811	0,3514	3,99E-05
DU	Chr.14:107457741	0,7448	0,6880	0,8017	1,21E-10	PI	Chr.7:101257337	0,7418	0,6442	0,8395	2,14E-05
DU	Chr.14:107794160	0,7618	0,7065	0,8172	5,76E-12	PI	Chr.8:41542706	0,7609	0,6647	0,8571	4,54E-06
DU	Chr.14:107869191	0,7434	0,6862	0,8005	1,56E-10	PI	Chr.9:48990007	0,6923	0,5910	0,7937	7,23E-04
DU	Chr.16:32192496	0,8180	0,7671	0,8688	6,20E-17	PI	Chr.9:138675793	0,6945	0,5780	0,8109	6,32E-04
DU	Chr.18:23850297	0,7405	0,6832	0,7978	2,56E-10	PI	Chr.12:11070623	0,6897	0,5947	0,7846	8,58E-04
LA	Chr.6:103109916	0,6635	0,5776	0,7494	4,37E-04	PI	Chr.13:136995166	0,6906	0,5913	0,7899	8,08E-04
LA	Chr.6:121005974	0,7155	0,6375	0,7936	3,56E-06	PI	Chr.13:177262726	0,6918	0,5782	0,8053	7,51E-04
LA	Chr.8:41620747	0,7005	0,6020	0,7990	1,62E-05	PI	Chr.13:196377418	0,7163	0,6100	0,8226	1,44E-04
LA	Chr.8:47482649	0,6800	0,5873	0,7726	1,08E-04	PI	Chr.14:106836471	0,6843	0,5751	0,7936	1,20E-03
LA	Chr.9:25438712	0,7091	0,6312	0,7870	6,90E-06	PI	Chr.14:107794160	0,7198	0,6207	0,8190	1,12E-04
LA	Chr.9:48882095	0,7074	0,6281	0,7866	8,19E-06	PI	Chr.15:79166362	0,6988	0,5862	0,8114	4,75E-04
LA	Chr.9:49034938	0,6578	0,5718	0,7437	6,90E-04	PI	Chr.15:79169250	0,7701	0,6771	0,8631	2,06E-06

LA	Chr.9:59498749	0,6725	0,5935	0,7515	2,07E-04	PI	Chr.15:79172228	0,7143	0,6052	0,8234	1,65E-04
LA	Chr.9:138712218	0,6743	0,5958	0,7528	1,77E-04	PI	Chr.15:84477930	0,6903	0,5989	0,7817	8,23E-04
LA	Chr.12:11541028	0,6902	0,6040	0,7763	4,31E-05	PI	Chr.16:49010636	0,7105	0,5810	0,8400	2,15E-04
LA	Chr.14:11430036	0,6580	0,5682	0,7479	6,77E-04	PI	Chr.16:49283730	0,7201	0,6170	0,8233	1,09E-04
LA	Chr.14:37687463	0,6695	0,5862	0,7528	2,67E-04	PI	Chr.18:23850297	0,7175	0,6207	0,8143	1,32E-04
LA	Chr.14:1121481	0,6602	0,5684	0,7519	5,72E-04	YO	Chr.3:64757843	0,7418	0,6455	0,8382	4,44E-04
LA	Chr.14:107298278	0,6590	0,5880	0,7300	6,25E-04	YO	Chr.3:77687689	0,6808	0,5588	0,8028	8,64E-03
LA	Chr.15:56426211	0,6732	0,6019	0,7444	1,96E-04	YO	Chr.9:138675793	0,7423	0,6271	0,8575	4,34E-04
LA	Chr.15:79166362	0,6841	0,5922	0,7760	7,50E-05	YO	Chr.12:44840898	0,7024	0,5796	0,8253	3,28E-03
LA	Chr.15:79172228	0,6663	0,5726	0,7601	3,47E-04	YO	Chr.13:134572869	0,7162	0,5940	0,8383	1,69E-03
LA	Chr.18:23850297	0,6612	0,5714	0,7510	5,25E-04	YO	Chr.13:196377418	0,7150	0,6148	0,8152	1,79E-03
LA	Chr.X:25039541	0,6894	0,6022	0,7767	4,61E-05	YO	Chr.14:72991761	0,6743	0,5671	0,7815	1,14E-02
LA	Chr.X:59154182	0,6655	0,5880	0,7431	3,70E-04	YO	Chr.14:92116537	0,7148	0,6158	0,8137	1,81E-03
LW	Chr.3:118879246	0,7094	0,6377	0,7810	2,00E-07	YO	Chr.14:98843824	0,6948	0,5614	0,8282	4,68E-03
LW	Chr.4:9676649	0,6886	0,6160	0,7611	2,82E-06	YO	Chr.14:107457741	0,6899	0,5727	0,8070	5,83E-03
LW	Chr.6:146485746	0,6823	0,6079	0,7567	5,99E-06	YO	Chr.14:107869191	0,6885	0,5770	0,8000	6,20E-03
LW	Chr.7:17730613	0,6789	0,6053	0,7525	8,89E-06	YO	Chr.15:79172228	0,6818	0,5663	0,7972	8,30E-03
LW	Chr.7:48837152	0,6818	0,6124	0,7512	6,36E-06	YO	Chr.16:17901382	0,6847	0,5693	0,8002	7,30E-03
LW	Chr.7:52269732	0,7241	0,6546	0,7935	2,64E-08	YO	Chr.16:32192496	0,7291	0,6094	0,8487	8,79E-04
LW	Chr.7:106301845	0,6864	0,6192	0,7536	3,68E-06	YO	Chr.18:3442709	0,7059	0,5863	0,8256	2,78E-03
LW	Chr.8:45485535	0,7016	0,6276	0,7757	5,52E-07						

Примечание: * – отмечены кластеры SNP.
 Note: * – SNP clusters are marked.

Для свиней породы ландрас определены 52 SNP со значительным дифференцирующим потенциалом ($p < 0,01$), для 7 из них НГ 95 %-ного ДИ для AUC $> 0,6$ ($p < 0,001$). На хромосоме 15 в последовательности гена *TTN* (titin, NCBI Gene ID – 100519519) определено два SNP (Chr.15:79166362, Chr.15:79172228) с высоким дифференцирующим потенциалом для этой породы. *TTN* кодирует белок поперечнополосатых мышц, который служит шаблоном для адгезии микрофиламентов при сборке сократительного механизма в мышечных клетках. В немышечных клетках *TTN* играет роль в конденсации хромосом и сегрегации хромосом во время митоза. Для ранее выявленных SNP НГ 95 %-ного ДИ находилась в диапазоне 0,511–0,610 [1], в то время как для ряда новых SNP это значение превышает 0,6 (таблица).

Для свиней породы пьетрен выявлены 39 SNP со значительным дифференцирующим потенциалом ($p < 0,01$), для 9 из них НГ 95 %-ного ДИ для AUC $> 0,6$ ($p < 0,001$). На хромосоме 15 выявлен фрагмент длиной около 10 тыс. п. н., на котором определены три SNP (Chr.15:79166362, Chr.15:79169250 и Chr.15:79172228) с высоким дифференцирующим потенциалом. Непосредственно за этими SNP в направлении 5' $>$ 3' цепи ДНК расположен ген *MAP3K20* (mitogen-activated protein kinase 20, NCBI Gene ID – 100152572), который кодирует белок, опосредующий передачу сигналов гамма-излучения, что приводит к остановке клеточного цикла. Его активность важна в регуляции контрольных точек клеточного цикла.

Для свиней породы крупная белая выявлены 104 SNP со значительным дифференцирующим потенциалом ($p < 0,01$), для 22 из них НГ 95 %-ного ДИ для AUC $> 0,6$ ($p < 0,001$). На хромосоме 14 выявлен фрагмент длиной около 30 тыс. п. н., на котором определены два SNP (Chr.14:93947383, Chr.14:93968978) с высоким дифференцирующим потенциалом для этой породы. Рядом с ним расположен ген *ZWINT* (ZW10 interacting kinetochore protein, NCBI Gene ID – 100623641), который участвует в функционировании кинетохора. Также, согласно данным баз Ensembl на февраль 2022 г., в этом регионе расположено несколько длинных некодирующих РНК (long ncRNAs, lncRNA), функция которых неизвестна – ENSSSCG00000055376, ENSSSCG00000044100. На хромосоме 15 также выявлен фрагмент длиной около 30 тыс. п. н., на котором определены SNP (Chr.15:84480210, Chr.15:84509213), информативные для дифференциации данной породы. Так же как и для свиней породы ландрас, для крупной белой на хромосоме 15 в последовательности гена *TTN* определено несколько SNP с высоким дифференцирующим потенциалом. Таким образом, можно отметить, что для ряда SNP в пределах одного гена выявлены значительные различия по частоте альтернативных аллелей у пород крупная белая и ландрас, при этом каждый из аллелей является информативным. Роль этих SNP для дифференциации пород предстоит оценить в будущем.

Для свиней породы йоркшир определено 14 SNP со значительным дифференцирующим потенциалом ($p < 0,01$), для 5 из них НГ 95 %-ного ДИ $> 0,6$ ($p < 0,002$). На хромосоме 14 выявлены два близко расположенных кластера (расстояние между ними – около 500 тыс. п. н.), на которых определены два SNP (Chr.14:107457741 и Chr.14:107869191) с высоким дифференцирующим потенциалом. Полиморфный локус Chr.14:107457741 расположен в интронной области гена *CC2D2B* (coiled-coiled C2 domain containing 2B, NCBI Gene ID – 100524048), полиморфный локус Chr.14:107869191 – в интронной области гена *TLL2* (tolloidlike 2, NCBI Gene ID – 100157589).

Для ряда SNP со значительным или высоким дифференцирующим потенциалом ранее показана ассоциация с QTL: согласно классификации Pig QTLdb для класса «Exterior» (экстерьерные характеристики) – Chr.13:182876924 [3], Chr.13:196377418 [3], Chr.13:177262726 [4]; для класса «Health» (общее состояние организма, здоровье) – Chr.13:134572869 [5], Chr.13:136995166 [6], Chr.14:107298278 [7], Chr.14:107457741 [7], Chr.14:107794160 [7], Chr.14:107869191 [7], Chr.16:17901382 [8], Chr.18:23850297 [9], Chr.6:146485746 [10], Chr.7:17730613 [7], Chr.8:45485535 [11] и Chr.8:47482649 [12]; для класса «Meat and Carcass» (качество мяса, размеры) – Chr.14:61121481 [13], Chr.16:32192496 [14], Chr.16:49283730 [14], Chr.4:9676649 [15], Chr.7:101257337 [16], Chr.8:41542706 [17], Chr.8:41620747 [17] и Chr.9:25438712 [9]; для класса «Production» (количество товарной продукции) – Chr.12:11541028 [18], Chr.9:59498749 [19]; для класса «Reproduction» (характеристика потомства) – Chr.12:11070623 [20].

Таким образом, выявление SNP, для которых не только определены высокие значения AUC, но и ассоциации с QTL, позволит одновременно решать задачу по оценке чистопородности животных и проявлению хозяйственно полезных признаков.

Заключение. Таким образом, расширенный биоинформатический анализ, который включал в себя определение генотипа по 7451 SNP для 248 геномов *Sus scrofa domestica*, позволил выявить суммарно 393 SNP для всех пород, для которых имеется существенная разница в частоте альтернативных аллелей у свиней дюрок, ландрас, пьетрен, крупная белая и йоркшир. Обозначены кластеры в пределах хромосом, в которых плотность SNP с высоким дифференцирующим потенциалом наиболее высока. Например, для нескольких пород (ландрас и крупная белая) данные кластеры расположены в пределах одного гена – *TTN*, что может быть связано с жесткой селекцией свиней на повышение мясной продуктивности.

Полученные результаты позволяют разработать тест-модели для определения чистопородности свиней и, как следствие, поддержания на оптимальном уровне таких параметров популяции (пород, линий), как уровень гетерозиготности, инбредности, генетической дифференциации и др., что будет способствовать повышению продуктивности животных и определению тактики и стратегии разведения и дальнейшего совершенствования пород свиней.

В целом, данный подход и объем проведенных исследований является на данный момент наиболее полным биоинформатическим анализом геномов *Sus scrofa domestica* для экстраполяции полученных данных на породы, разводимые в Республике Беларусь.

Благодарности. Исследование выполнено в рамках ГПНИ «Биотехнологии-2» (2021–2025 гг.), подпрограмма «Геномика, эпигеномика, биоинформатика», НИР «Разработка системы генетического анализа для определения чистопородности свиней на основе изучения SNP-локусов».

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Acknowledgment. The study was carried out within the framework of the State Scientific Research Program “Biotechnology-2” (2021–2025), subprogram “Genomics, epigenomics, bioinformatics”, research work “Development of a genetic analysis system for determining the purity of pigs based on the study of SNP loci”.

Conflict of interests. The authors declare no conflicts of interest.

Список использованных источников

1. Биоинформатический анализ геномов коммерческих пород домашних свиней для идентификации породоспецифичных SNP / В. Н. Кипень [и др.] // Вес. Нац. акад. навук Беларусі. Сер. аграр. навук. – 2021. – Т. 59, № 4. – С. 464–476. <https://doi.org/10.29235/1817-7204-2021-59-4-464-476>
2. Analysis of HEPH Gene Polymorphism on the X Chromosome for Identification of Wild Boar and Domestic Pig / V. N. Kipen [et al.] // Russ. J. Genet. – 2020. – Vol. 56. – P. 1099–1108. <https://doi.org/10.1134/s1022795420080062>
3. An association and haplotype analysis of porcine maternal infanticide: a model for human puerperal psychosis? / C. R. Quilter [et al.] // Am. J. Med. Genet. B Neuropsychiatr. Genet. – 2012. – Vol. 159(B), N 8. – P. 908–927. <https://doi.org/10.1002/ajmg.b.32097>
4. A study of vertebra number in pigs confirms the association of vertnin and reveals additional QTL / G. A. Rohrer [et al.] // BMC Genet. – 2015. – Vol. 16, N 1. <https://doi.org/10.1186/s12863-015-0286-9>
5. Refined candidate region for F4ab/ac enterotoxigenic *Escherichia coli* susceptibility situated proximal to MUC13 in pigs / T. Goetstouwers [et al.] // PLoS One. – 2014. – Vol. 9, N 8. – Art. e105013. <https://doi.org/10.1371/journal.pone.0105013>
6. A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine / Wei-Xuan Fu [et al.] // PLoS One. – 2012. – Vol. 7, N 3. – Art. e32127. <https://doi.org/10.1371/journal.pone.0032127>
7. Genome-wide association studies for hematological traits in Chinese Sutan pigs / Feng Zhang [et al.] // BMC Genet. – 2014. – Vol. 15, N 1. <https://doi.org/10.1186/1471-2156-15-41>
8. A substitution in the ligand binding domain of the porcine glucocorticoid receptor affects activity of the adrenal gland / E. Murani [et al.] // PLoS One. – 2012. – Vol. 7, N 9. – Art. e45518. <https://doi.org/10.1371/journal.pone.0045518>
9. Genome-wide association study for T lymphocyte subpopulations in swine / Xin Lu [et al.] // BMC Genomics. – 2012. – Vol. 13, N 1. – Art. 488. <https://doi.org/10.1186/1471-2164-13-488>
10. A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants / A. Manunza [et al.] // BMC Genomics. – 2014. – Vol. 15, N 1. – Art. 758. <https://doi.org/10.1186/1471-2164-15-758>
11. Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model / C. Chen [et al.] // BMC Genomics. – 2013. – Vol. 14, N 1. – Art. 848. <https://doi.org/10.1186/1471-2164-14-848>
12. Genome-wide association study reveal constant and specific loci for hematological traits at three time stages in a White Duroc × Erhualian F2 resource population / Z. Zhang [et al.] // PLoS One. – 2013. – Vol. 8, N 5. – Art. e63665. <https://doi.org/10.1371/journal.pone.0063665>

13. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis / S. Ponsuksili [et al.] // *Int. J. Biol. Sci.* – 2014. – Vol. 10, N 3. – P. 327–337. <https://doi.org/10.7150/ijbs.8134>
14. Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations / W. Zhang [et al.] // *Sci. Rep.* – 2016. – Vol. 6, N 1. – Art. 24718. <https://doi.org/10.1038/srep24718>
15. Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs / B. Fan [et al.] // *PLoS One.* – 2011. – Vol. 6, N 2. – Art. e14726. <https://doi.org/10.1371/journal.pone.0014726>
16. SNP- and haplotype-based genome-wide association studies for growth, carcass, and meat quality traits in a Duroc multigenerational population / S. Sato [et al.] // *BMC Genet.* – 2016. – Vol. 17, N 1. – Art. 60. <https://doi.org/10.1186/s12863-016-0368-3>
17. Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population / I. Choi [et al.] // *Meat Sci.* – 2012. – Vol. 92, N 2. – P. 132–138. <https://doi.org/10.1016/j.meatsci.2012.04.023>
18. Genome Wide Association Analysis Reveals New Production Trait Genes in a Male Duroc Population / K. Wang [et al.] // *PLoS One.* – 2015. – Vol. 10, N 9. – Art. e0139207. <https://doi.org/10.1371/journal.pone.0139207>
19. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars / J. T. Howard [et al.] // *BMC Genet.* – 2015. – Vol. 16. <https://doi.org/10.1186/s12863-015-0218-8>
20. Genome-wide prediction of age at puberty and reproductive longevity in sows / J. K. Tart [et al.] // *Anim Genet.* – 2013. – Vol. 44, N 4. – P. 387–397. <https://doi.org/10.1111/age.12028>

References

1. Kipen V. N., Mikhailova M. E., Snytkov E. V., Romanishko E. L., Ivanova E. V., Sheyko R. I. Bioinformatic analysis of genomes of commercial breeds of domestic pigs for identification of breed-specific SNPs. *Vestsi Natsyyanal' nay akademii navuk Belarusi. Seryya agrarnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Agrarian series*, 2021, vol. 59, no. 4, pp. 464–476 (in Russian). <https://doi.org/10.29235/1817-7204-2021-59-4-464-476>
2. Kipen V. N., Ivanova E. V., Snytkov E. V., Verchuk A. N. Analysis of HEPH Gene Polymorphism on the X Chromosome for Identification of Wild Boar and Domestic Pig. *Russian Journal of Genetics*, 2020, vol. 56, no. 9, pp. 1099–1108. <https://doi.org/10.1134/s1022795420080062>
3. Quilter C. R., Sargent C. A., Bauer J., Bagga M. R., Reiter C. P., Hutchinson E. L., Southwood O. I., Evans G., Mileham A., Griffin D. K., Affara N. A. An association and haplotype analysis of porcine maternal infanticide: a model for human puerperal psychosis. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 2012, vol. 159 (B), no. 8, pp. 908–927. <https://doi.org/10.1002/ajmg.b.32097>
4. Rohrer G. A., Nonneman D. J., Wiedmann R. T., Schneider J. F. A study of vertebra number in pigs confirms the association of vertnin and reveals additional QTL. *BMC Genet.*, 2015, vol. 16, no. 1. <https://doi.org/10.1186/s12863-015-0286-9>
5. Goetstouwers T., Van Poucke M., Coppieters W., Nguyen V. U., Melkebeek V., Coddens A., Van Steendam K., Deforce D., Cox E., Peelman L. J. Refined candidate region for F4ab/ac enterotoxigenic *Escherichia coli* susceptibility situated proximal to MUC13 in pigs. *PLoS One*, 2014, vol. 9, no. 8, art. e105013. <https://doi.org/10.1371/journal.pone.0105013>
6. Fu W.-X., Liu Y., Lu X., Niu X.-Y., Ding X.-D., Liu J.-F., Zhang Q. A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine. *PLoS One*, 2012, vol. 7, no. 3, art. e32127. <https://doi.org/10.1371/journal.pone.0032127>
7. Zhang F., Zhang Z., Yan X., Chen H., Zhang W., Hong Y., Huang L. Genome-wide association studies for hematological traits in Chinese Suta pigs. *BMC Genetics*, 2014, vol. 15, no. 1. <https://doi.org/10.1186/1471-2156-15-41>
8. Murani E., Reyer H., Ponsuksili S., Fritschka S., Wimmers K. A substitution in the ligand binding domain of the porcine glucocorticoid receptor affects activity of the adrenal gland. *PLoS One*, 2012, vol. 7, no. 9, art. e45518. <https://doi.org/10.1371/journal.pone.0045518>
9. Lu X., Fu W.-X., Luo Y.-R., Ding X.-D., Zhou J.-P., Liu Y., Liu J.-F., Zhang Q. Genome-wide association study for T lymphocyte subpopulations in swine. *BMC Genomics*, 2012, vol. 13, no. 1, art. 488. <https://doi.org/10.1186/1471-2164-13-488>
10. Manunza A., Casellas J., Quintanilla R., González-Prendes R., Pena R. N., Tibau J., Mercadé A., Castelló A., Aznárez N., Hernández-Sánchez J., Amills M. A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. *BMC Genomics*, 2014, vol. 15, no. 1, art. 758. <https://doi.org/10.1186/1471-2164-15-758>
11. Chen C., Yang B., Zeng Z., Yang H., Liu C., Ren J., Huang L. Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. *BMC Genomics*, 2013, vol. 14, no. 1, art. 848. <https://doi.org/10.1186/1471-2164-14-848>
12. Zhang Z., Hong Y., Gao J., Xiao S., Ma J., Zhang W., Ren J., Huang L. Genome-wide association study reveals constant and specific loci for hematological traits at three time stages in a White Duroc × Erhualian F2 resource population. *PLoS One*, 2013, vol. 8, no. 5, art. e63665. <https://doi.org/10.1371/journal.pone.0063665>
13. Ponsuksili S. Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *International Journal of Biological Sciences*, 2014, vol. 10, no. 3, pp. 327–337. <https://doi.org/10.7150/ijbs.8134>
14. Zhang W., Yang B., Zhang J., Cui L., Ma J., Chen C., Ai H., Xiao S., Ren J., Huang L. Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations. *Scientific Reports*, 2016, vol. 6, no. 1, art. 24718. <https://doi.org/10.1038/srep24718>
15. Fan B., Onteru S. K., Du Z.-Q., Garrick D. J., Stalder K. J., Rothschild M. F. Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS One*, 2011, vol. 6, no. 2, art. e14726. <https://doi.org/10.1371/journal.pone.0014726>

16. Sato S., Uemoto Y., Kikuchi T., Egawa S., Kohira K., Saito T., Sakuma H., Miyashita S., Arata S., Kojima T., Suzuki K. SNP- and haplotype-based genome-wide association studies for growth, carcass, and meat quality traits in a Duroc multigenerational population. *BMC Genetics*, 2016, vol. 17, no. 1, art. 60. <https://doi.org/10.1186/s12863-016-0368-3>

17. Choi I., Bates R. O., Raney N. E., Steibel J. P., Ernst C. W. Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population. *Meat Science*, 2012, vol. 92, no. 2, pp. 132–138. <https://doi.org/10.1016/j.meatsci.2012.04.023>

18. Wang K., Liu D., Hernandez-Sanchez J., Chen J., Liu C., Wu Z., Fang M., Li N. Genome Wide Association Analysis Reveals New Production Trait Genes in a Male Duroc Population. *PLoS One*, 2015, vol. 10, no. 9, art. e0139207. <https://doi.org/10.1371/journal.pone.0139207>

19. Howard J. T., Jiao S., Tiezzi F., Huang Y., Gray K. A., Maltecca C. Genome-wide association study on legendre random regression coefficients for the growth and feed intake trajectory on Duroc Boars. *BMC Genetics*, 2015, vol. 16, no. 1. <https://doi.org/10.1186/s12863-015-0218-8>

20. Tart J. K., Johnson R. K., Bundy J. W., Ferdinand N. N., McKnite A. M., Wood J. R., Miller P. S., Rothschild M. F., Spangler M. L., Garrick D. J., Kachman S. D., Ciobanu D. C. Genome-wide prediction of age at puberty and reproductive longevity in sows. *Animal Genetics*, 2013, vol. 44, no. 4, pp. 387–397. <https://doi.org/10.1111/age.12028>

Информация об авторах

Кипень Вячеслав Николаевич – канд. биол. наук, вед. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: v.kipen@igc.by. ORCID: 0000-0002-7822-0746.

Снытков Евгений Владимирович – мл. науч. сотрудник. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: evsnytkov@gmail.com. ORCID: 0000-0001-7961-1952.

Михайлова Мария Егоровна – канд. биол. наук, заведующий лабораторией. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: M. Mikhailova@igc.by. ORCID: 0000-0001-6087-5069.

Шейко Руслан Иванович – член-корреспондент, д-р с.-х. наук, профессор, директор. Институт генетики и цитологии НАН Беларуси (ул. Академическая, 27, 220072, Минск, Республика Беларусь). E-mail: R.I.Sheyko@igc.by. ORCID: 0000-0001-5442-566X.

Information about the authors

Kipen Viachaslau N. – Ph. D. (Biology), Leading Researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: v.kipen@igc.by. ORCID: 0000-0002-7822-0746.

Snytkov Evgenij V. – Junior Researcher. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: evsnytkov@gmail.com. ORCID: 0000-0001-7961-1952.

Mikhailova Mariya E. – Ph. D. (Biology), Head of the Laboratory. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: M. Mikhailova@igc.by. ORCID: 0000-0001-6087-5069.

Sheyko Ruslan I. – Corresponding Member, D. Sc. (Agrarian), Professor. Institute of Genetics and Cytology of the National Academy of Sciences of Belarus (27, Akademicheskaya Str., 220072, Minsk, Republic of Belarus). E-mail: R.I.Sheyko@igc.by. ORCID: 0000-0001-5442-566X.