

МОДЕЛИРОВАНИЕ ПАНЕЛИ ПОРОДОСПЕЦИФИЧНЫХ SNP-МАРКЕРОВ ДЛЯ ОПРЕДЕЛЕНИЯ ЧИСТОПородНОСТИ ДОМАШНИХ СВИНЕЙ ПОРОДЫ ДЮРОК

КИПЕНЬ Вячеслав Николаевич, *научный сотрудник*
НИЛ молекулярно-биологических исследований
ГУ «Научно-практический центр Государственного комитета
судебных экспертиз Республики Беларусь»

Введение. Дюрок (анг. Duroc) – порода свиней красной масти. Изначально порода была сального направления, но учитывая высокий спрос, со временем направление продуктивности изменилось на мясное. Дюрок – американская порода свиней, выведенная в конце 19 века. В настоящее время широко распространена во всем мире. Дюроки достаточно выносливые и хорошо приспособляются к пастбищному содержанию.

Порода отличается своими высокими мясными качествами. Животные этой породы больших размеров, крепкой конституции. Туловище умеренной длины, глубокое и широкое. По многоплодности свиньи породы дюрок не очень плодовиты. По плодовитости уступают крупной белой и ландрасу. За один опорос свиноматка приносит от 9 до 11 поросят.

Хряки этой породы используются в качестве терминальной линии для производства товарных свиней, характеризующихся высоким качеством мяса и туши. На сегодняшний день порода дюрок

широко распространена в европейских странах. В Республике Беларусь данная порода распространена слабо – доли процента от общей численности поголовья *Sus scrofa domesticus*.

Ранее нами была показана возможность с использованием данных полногеномных сиквенсных проектов коммерческих пород свиней определить наличие породоспецифичных SNP-маркеров для животных некоторых пород свиней [1-3].

Цель и задачи. Смоделировать с использованием MDR-анализа (Multifactor dimensionality reduction [4]) панель генетических маркеров, способную дифференцировать животных породы дюрок от представителей пород крупная белая, ландрас, пьетрен и мейшан, а также охарактеризовать ее с позиций чувствительности, специфичности и общей точности.

Материалы и методы. Поиск породоспецифичных SNP был выполнен с помощью алгоритма SRA Nucleotide BLAST (Sequence Read Archive Nucleotide BLAST) и программы BioEdit v.7.2.5. Количество включенных в анализ SNP – 193 [5]; число полногеномных прочтений для свиней породы дюрок – 28 (BioSample NCBI, [<https://www.ncbi.nlm.nih.gov/biosample>] – SAMEA1557407, SAMEA1557419, SAMEA1557423, SAMEA1557434, SAMEA3497837, SAMEA3497838, SAMEA3497839, SAMEA3497840, SAMN03031126, SAMN03031127, SAMN03031128, SAMN03031129, SAMN03031130, SAMN03031131, SAMN03031132, SAMN03031133, SAMN03031134, SAMN03031135, SAMN03031136, SAMN03031137, SAMN03031138, SAMN03031139, SAMN03031140, SAMN03031141, SAMN03031142, SAMN03031143, SAMN03031144, SAMN03031145), для других пород – 58 (крупная белая – 17 (SAMEA3497789, SAMEA3497790, SAMEA3497853, SAMEA1557412, SAMEA1557415, SAMEA1557431, SAMEA1557413, SAMEA1557402, SAMEA1557389, SAMEA1557435, SAMEA1557422, SAMEA1557406, SAMEA1557404, SAMEA1557383, SAMEA1557425, SAMEA1557427, SAMEA1557399), ландрас – 21 (SAMEA1557405, SAMEA1557416, SAMEA1557436, SAMEA1557390, SAMEA3497847, SAMEA3497848, SAMEA3497850, SAMEA3497851, SAMN03031146, SAMN03031147, SAMN03031148, SAMN03031149, SAMN03031150, SAMN03031151, SAMN03031152, SAMN03031153, SAMN03031154, SAMN03031155, SAMN03031156, SAMN03031157, SAMN03031158), мейшан – 14 (SAMEA3497800, SAMEA3497801, SAMEA3497802, SAMEA3497803, SAMEA3497804, SAMEA3497805, SAMEA3497806, SAMEA3497807, SAMEA3497808, SAMEA3497809, SAMEA1557420, SAMEA1557428, SAMEA1557395, SAMEA1557410), пьетрен – 6 (SAMEA1557430, SAMEA1557432, SAMEA1557392, SAMEA1557408, SAMEA1557397, SAMEA3497860)). Общее количество проанализированных сиквенсов – 32 754 738 518.

Были использованы SRA-данные по полногеномному секвенированию (NGS, Next-Generation Sequencing), размещенные в открытом доступе на облачном сервисе DNAnexus (<http://sra.dnanexus.com/>), а также в SRA-NCBI – high-throughput DNA and RNA sequence read archive (www.ncbi.nlm.nih.gov/sra).

Построение модели взаимодействий SNP (определение минимального и достаточного количества генетических маркеров для решения поставленной задачи) проводилось с использованием биоинформатического метода MDR (<http://www.multifactorialdimensionalityreduction.org/>).

Результаты и обсуждение. В результате проведенного исследования нами было выявлено наличие 32 строго специфичных SNP-маркеров (породоспецифичный аллель отмечен только у представителей данной породы) для породы дюрок: H3GA0041584, ALGA0080273, ALGA0080306, ALGA0121437, ALGA0005730, ALGA0048872, ALGA0066693, ALGA0123578, ASGA0098607, ALGA0025040, ASGA0008074, ALGA0073165, ALGA0019280, ALGA0044392, ASGA0069555, ALGA0078851, ALGA0106902, ASGA0004361, ASGA0050174, ALGA0061248, ALGA0097613, ALGA0100515, ALGA0097844, ALGA0069975, ALGA0093148, ALGA0048730, ALGA0047768, ALGA0066573, ALGA0120198, ASGA0051859, ALGA0112306, ALGA0053392. Частота пороспецифического аллеля варьировала в широких пределах – от 8,0% до 96,4%.

В процессе моделирования нами были использованы высоко консервативные настройки поиска конфигурации модели, которые позволили однозначно дифференцировать наличие/отсутствие статистически значимых эффектов: количество атрибутов (attribute count range) – от 1 до n (где n – количество переменных в модели); воспроизводимость модели (cross-validation count) – 100; анализ топ-моделей (track top models) – 1000; поиск конфигурации модели (search method configuration) – exhaustive; классификация ячеек (ambiguous cell assignment) – unclassified.

В результате проведенного моделирования были определены 5 моделей, отражающие такое сочетание породоспецифичных для породы дюрок SNP, которое позволило наилучшим образом отличить животных этой породы от других пород в рамках данной работы. В частности, каждая мо-

дель включала в себя только два SNP и характеризовалась следующими характеристиками: сбалансированная точность (adj. Balanced accuracy) – не менее 96%, чувствительность (Sensitivity) – не менее 99%, специфичность (Specificity) – не менее 99%, воспроизводимость (Cross Validation Consistency) – 100%.

Закключение. Таким образом, предложена и охарактеризована модель, включающая два SNP-маркера, с помощью которой имеется возможность с высокой точностью (более 96%) отличить чистопородных домашних свиней породы дюрок от особей пород крупная белая, ландрас, пьетрен и мейшан. Для некоторых моделей показана возможность определения генотипа с использованием ПЦР-ПДРФ, т.е. данная схема может быть реализована в молекулярно-генетической лаборатории с минимальным оборудованием.

Полученные результаты могут лечь в основу создания панели SNP-маркеров для определения чистопородности особей породы дюрок подвида *Sus scrofa domesticus*.

Список использованных источников

1. Моделирование панели породоспецифичных SNP-маркеров для определения чистопородности домашних свиней крупной белой породы / В.Н. Кипень // IX Всероссийская научно-практическая конференция с международным участием «Молекулярная диагностика – 2017», категория Ветеринария и селекция в животноводстве: Сборник материалов конф. – Москва, РФ. – 2017. – С. 396-397.

2. Снытков Е.В. Моделирование панели породоспецифичных SNP-маркеров для определения чистопородности домашних свиней породы пьетрен / Е.В. Снытков, В.Н. Кипень // Сборник XVII Всероссийской молодежной научной конференции «Биотехнология в растениеводстве, животноводстве и ветеринарии», 6-7 апреля 2017. – Москва. – 2017. – С. 134-136.

3. Кипень В.Н. Моделирование панели породоспецифичных SNP-маркеров для определения чистопородности домашних свиней породы мейшан / Сборник XVII Всероссийской молодежной научной конференции «Биотехнология в растениеводстве, животноводстве и ветеринарии», 6-7 апреля 2017. – Москва. – 2017. – С. 143-146.

4. Greene, C. Multifactor dimensionality reduction for graphics processing units enables genome-wide testing of epistasis in sporadic ALS // Bioinformatics. – 2010. – P. 694-695.

5. Ramos, A.M. Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing / A.M. Ramos, H.J. Megens, R.P. Crooijmans [et al.] // Anim. Genet. – 2011. – Vol. 42(6). P. 613-620.