

*И.С. Бейшова¹, Д.А. Гриценко², М.Х. Шаменова², А.С. Пожарский²,
Т.В. Ульянова^{1*}, А.М. Ковальчук¹*

¹ НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана», г. Уральск, Республика Казахстан, indira_bei@mail.ru, tatyana.poddudinskaya@gmail.com*, kovalchuk_s89@mail.ru

² РГП на ПХВ «Институт биологии и биотехнологии растений», г. Алматы, Республика Казахстан, d.kopytina@gmail.com, shamekov@gmail.com, aspozarsky@gmail.com

ПОЛНОГЕНОМНЫЙ ПОИСК АССОЦИАЦИЙ С ПРОДУКТИВНЫМИ КАЧЕСТВАМИ У ОТЕЧЕСТВЕННЫХ ПОРОД ЛОШАДЕЙ

Аннотация

В настоящей работе представлены результаты полногеномного поиска ассоциаций однонуклеотидных полиморфизмов (SNP) с продуктивными качествами у отечественных пород лошадей, полученные на основании данных SNP-генотипирования животных, проведенного с помощью набора реагентов Equine 80k HTS («Illumina Inc.», США). Объектом исследований являлись лошади следующих пород и типов: казахской типов джабе (n = 631), адай (n = 303) и найман (n = 158), мугалжарской (n = 584), кушумской (n = 226) и костанайской (n = 116) пород. Сбор информации по фенотипическим данным лошадей проводили до отбора проб, при этом измеряли следующие показатели: высота в холке, косая длина туловища, обхват груди, обхват пясти и живую массу. Общее число SNP, отобранных по результатам контроля качества, составило 60 987. Полногеномный поиск ассоциаций был проведен с помощью линейной регрессии с адаптивным методом перестановок Монте-Карло с коррекцией р-значения для множественных сравнений в программе PLINK1.9. В ходе анализа были выявлены 60 полиморфизмов, сцепленных с генами, участвующими в регуляции процессов развития соединительной ткани и костной системы, нервной системы, регуляции иммунной системы и других процессах. Проведенная работа подтвердила актуальность применения подобного подхода в полногеномно-ассоциативных исследованиях для детектирования единичных SNP, связанных с отдельными признаками.

Ключевые слова: *полногеномный поиск ассоциаций, лошади, тип джабе, адайский тип, найманский тип, мугалжарская порода, кушумская порода, костанайская порода, SNP.*

Введение

В 2007 году полная последовательность генома лошади стала общедоступной, благодаря этому исследованию совершен значительный прорыв в области геномики лошадей [1]. Результаты этой работы послужили ключевым источником для создания коммерческих массивов SNP, что позволило проводить высокопроизводительное генотипирование лошадей. В 2011 году появились массивы ДНК-генотипирования первого и второго поколения, включающие 54 602 и 74 500 SNP-маркеров соответственно [2]. Как следствие, стало возможным проведение полногеномно-ассоциативных исследований (GWAS) у лошадей. Среди наиболее впечатляющих результатов GWAS, полученных к настоящему времени, стоит отметить обнаружение SNP на 18 хромосоме, влияющих на экспрессию гена миостатина (*MSTN*), который связан со спортивными качествами у чистокровных лошадей [3].

Достигнут значительный прогресс в расшифровке менделевских признаков, включая выявление ряда аллельных вариантов, ответственных за разные окрасы шерсти [4], а также ассоциированных с генетическими дефектами [5].

В настоящее время имеется ограниченное количество информации о генетике полигенных количественных признаков у лошадей, включая их продуктивные качества. Исследование количественных признаков осложняется отсутствием информации о данных и

фенотипах животных, а также сложной генетическими механизмами, которые служат основой для этих признаков. Необходимо отметить, что подобные исследования для казахских местных лошадей ранее не проводились, но являются бесспорно актуальными, так как в коневодство в Казахстане является одной из ведущих отраслей животноводства, а изучение генетических особенностей формирования важных признаков у отечественных пород лошадей является важным для сохранения и совершенствования популяции.

Целью настоящей работы было исследование полногеномных ассоциаций SNP с продуктивными качествами в популяциях отечественных пород лошадей.

Методы и материалы

Исследовали биоматериал (волосынные луковицы) казахской породы типов джабе ($n = 631$), адай ($n = 303$) и найман ($n = 158$), мугалжарской ($n = 584$) кушумской ($n = 226$) и костанайской ($n = 116$) пород, отобранных с хозяйств различных регионов Казахстана.

Сбор промеров у лошадей проводили до отбора проб, при этом измеряли следующие показатели: высота в холке, косая длина туловища, обхват груди, обхват пясти и живую массу.

ДНК выделяли с помощью коммерческого набора ДНК-Экстран-2 (ООО «Синтол», Россия) в соответствии с рекомендациями производителя. Перед генотипированием на SNP-чипах полученные препараты проходили контроль качества: измеряли концентрацию двуцепочечной ДНК на флуориметре Qubit 4.0, «Invitrogen, Life Technologies», США), и проверяли её качество с помощью гель-электрофореза. Генотипирование проводили на чипе Equine 80k HTS («Illumina Inc.», США).

Контроль качества генотипирования проводили с помощью программы PLINK1.9 [6]. По его результатам было отобрано 60 987 SNP.

Полногеномный ассоциативный анализ проводили с использованием PLINK1.9. Результаты анализа генетической структуры, полученные ранее [7], были приняты во внимание при отборе образцов для GWAS. Анализ связи был проведен по показателям живой массы и промеров лошадей. Животные в возрасте до трех лет и образцы-выбросы были исключены из анализа. Корреляционный тест Пирсона был использован для обеспечения независимости фенотипических переменных от возраста. Переменная промеров была определена с использованием измерений высоты в холке, косой длины туловища, обхвата груди, обхвата пясти. Эти параметры были нормализованы путем расчета среднего значения и деления на стандартное отклонение, затем был проведен анализ главных компонент (PCA), и первый компонент был выбран в качестве новой переменной промеров животных.

Для выполнения полногеномного поиска ассоциаций применялся тест линейной регрессии с адаптивным методом перестановок Монте-Карло, с коррекцией р-значения для учета множественных сравнений. (команда PLINK '--linear perm'). SNP, находящиеся в состоянии сильного неравновесия связей, были исключены из анализа ($r^2 > 0,7$).

Маркеры с р-значением ниже установленного порога в 0,001 были аннотированы с использованием инструмента VEP (variant effect predictor) [8] и веб-сервера DAVID [9]. В качестве ссылки для аннотации использовалась сборка генома лошади EquCab3.0 (GCA_002863925.1).

Результаты и обсуждение

На сегодняшний день большинство исследований полногеномных ассоциаций у лошадей сосредоточены на изучении скаковых качеств [10] и здоровья животных [11]. Признаки, связанные с продуктивностью, не получили достаточного внимания в геномике лошадей, так как получаемая от них продукция, в основном мясо, остается экзотической во многих странах. Таким образом, изучение генетики таких признаков является новым не только для Казахстана. В данной работе мы проанализировали набор SNP-маркеров на предмет связи с наиболее общими параметрами, характеризующими мясную продуктивность - живой массой и промерами животных.

Полногеномный анализ связи SNP-маркеров с промерами и живой массой лошадей был проведен для всех животных с доступными фенотипическими данными, за исключением лошадей в возрасте до 3 лет и отдаленных генотипов, выявленных с помощью анализа

генетической структуры; всего было исследовано 1533 особи. Для проверки отсутствия влияния возраста в анализируемой выборке, мы провели корреляционный анализ Пирсона для всех фенотипических переменных. Показатели высоты в холке и косой длины туловища показали отсутствие значимой корреляции при пороговом уровне значимости $p = 0,05$ ($p = 0,8388$ и $0,4211$ соответственно). Слабые корреляции были выявлены для обхвата груди ($0,0841$, $p = 0,000507$), обхвата пясти ($0,1011$, $p = 2,9229 \cdot 10^{-5}$) и живой массы ($0,1121$, $p = 3,496 \cdot 10^{-6}$). Показатели промеров были объединены с использованием PCA; первый основной компонент, описывающий 81% общей вариации, был выбран в качестве новой переменной.

GWAS-анализ проводился с использованием алгоритма линейной регрессии, реализованного в программном обеспечении PLINK, и включал адаптивную коррекцию r -значений на основе теста перестановок Монте-Карло. С целью визуализации распределения статистически значимых полиморфных сайтов и их распределения по хромосомам для каждого анализируемого параметра строили Манхэттенские графики (Manhattan plot). По оси Y откладывали значения отрицательного логарифма уровня значимости для каждого полиморфного сайта ($-\log p$). По оси X откладывали хромосомы, в которых локализованы полиморфные сайты. Чем меньше значение p , тем больший уровень значимости имеет показатель и тем выше он расположен относительно начала координат по оси Y.

На рисунке 1 представлены графики, отражающие характер распределения значимых SNP по признакам живой массы и промеров у отечественных лошадей.

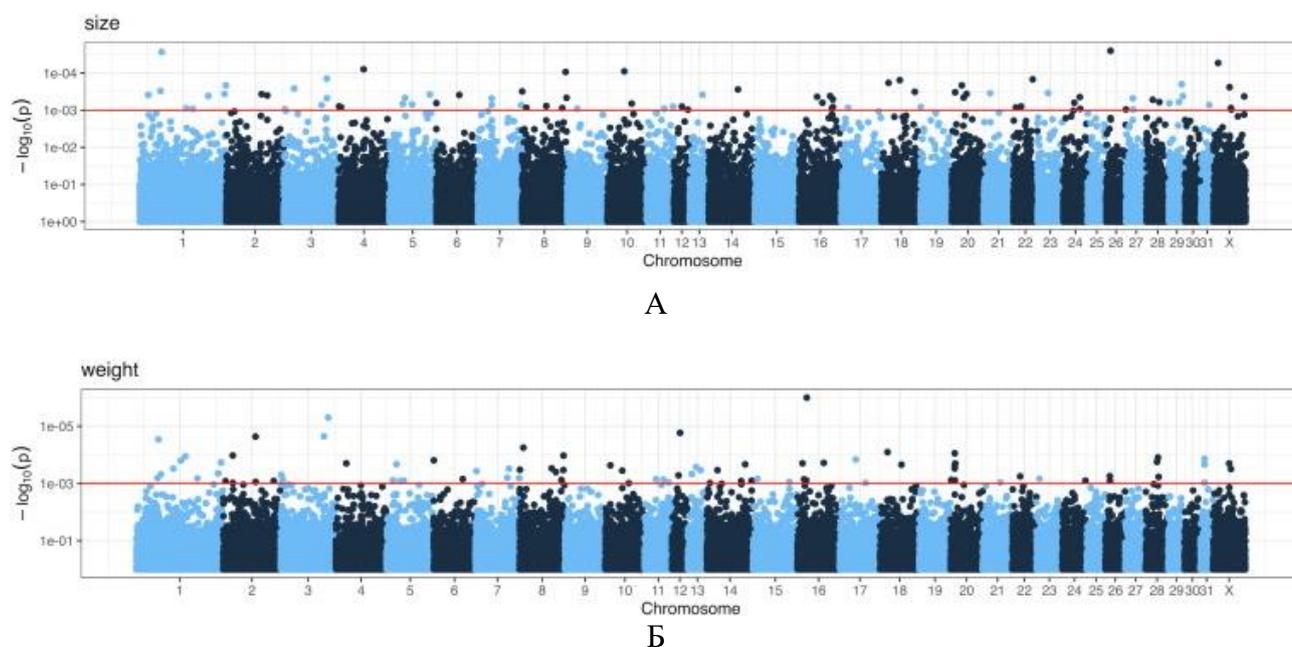


Рисунок 1 - Расположение статистически значимых полиморфных сайтов в хромосомах у лошадей отечественных пород

А – промеры, Б – живая масса. Ось X – хромосома лошади, ось Y – $-\log_{10}(p)$, красная линия – $p = 0,001$.

При выбранном уровне значимости $p = 0,001$ было обнаружено, что 81 и 84 SNP имеют статистически значимые ассоциации с промерами и живой массой, соответственно. Среди выявленных полиморфизмов, 60 SNP были связаны с известными генами лошади с помощью сервера VEP Ensembl и с соответствующими биологическими процессами с помощью DAVID (таблица 1).

Таблица 1 - Гены, обнаруженные в регионе локализации однонуклеотидных полиморфизмов, значимо ассоциированных с живой массой и промерами у отечественных пород лошадей

Название SNP на чипе	Позиция	Ген	P*		GO**	Биологический процесс
			Живая масса	Промеры		
1	2	3	4	5	6	7
BIEC2_7947	1:17461309	<i>ABLIM1</i>	-	0.0003862	GO:0007010	организация цитоскелета
Affx-101142060	31:8550289	<i>ACAT2</i>	0.0009455	-	GO:0006635	β-окисление жирных кислот
AX-104155593	26:41616293	<i>ADARB1</i>	-	0.0009522	GO:0006382	Редактирование аденозина в инозин
BIEC2_901539	5:34710217	<i>AIM2</i>	0.0007969	-	GO:0002218	активация врожденного иммунного ответа
BIEC2_19530	1:41720677	<i>ASAH2</i>	-	0.0003041	GO:0042759	биосинтез длинноцепочечных жирных кислот
BIEC2_516275	20:8331534	<i>BMP6</i>	0.000298	-	GO:0000122	негативная регуляция транскрипции с промотора РНК-полимеразы II
BIEC2_516282	20:8332816		8.805 * 10 ⁻⁵	-		
BIEC2_516343	20:8424937		0.0002289	-		
BIEC2_209409	13:9735624	<i>COL26A1</i>	0.000467	-	GO:0010811	Положительная регуляция адгезии клеток-субстрата
BIEC2_486667	2:69819186	<i>CPE</i>	0.0008825	-	GO:0006518	пептидный метаболический процесс
BIEC2_174862	12:11000582	<i>CREB3L1</i>	0.0005222	-	GO:0006357	регуляция транскрипции с промотора РНК-полимеразы II
BIEC2_900752	5:31687044	<i>DDR2</i>	-	0.0004624	GO:0007169	тирозинкиназный сигнальный путь трансмембранного рецепторного белка
BIEC2_103610	10:9449473	<i>DPF1</i>	0.0002326	-	GO:0045944	положительная регуляция транскрипции с промотора РНК-полимеразы II
Affx-102382610	1:149996219	<i>EIF2AK4</i>	-	0.0004106	GO:0000077	контрольная точка повреждения ДНК
BIEC2_1061054	8:68483265	<i>EPG5</i>	0.000296	-	GO:0006862	транспорт нуклеотидов
BIEC2_252279	14:7810156	<i>ERGIC1</i>	0.0009707	-	GO:0016192	опосредованный везикулами транспорт
BIEC2_515625	20:6838695	<i>F13A1</i>	-	0.0003289	GO:0018149	сшивание пептидов
BIEC2_453812	2:4970163	<i>FAM151A</i>	0.0008247	-	-	-
ТВIEC2_1007439	6:57510178	<i>FGD4</i>	0.0006988	-	GO:0007010	организация цитоскелета
BIEC2_489911	2:75633527	<i>FNIP2</i>	-	0.0003669	GO:0001932	регуляция фосфорилирования белков
BIEC2_645513	24:33029319	<i>FOXN3</i>	-	0.0004419	GO:0045892	негативная регуляция транскрипции, на матрице ДНК
BIEC2-411146	18:38386190	<i>GALNT5</i>	-	0.0001543	GO:0006486	гликозилирование белка
BIEC2_736048	28:25672835	<i>GAS2L3</i>	0.0001233	-	GO:0000226	организация цитоскелета микротрубочек

Продолжение таблицы 1

1	2	3	4	5	6	7
BIEC2_842473	4:814418	<i>GNAT3</i>	-	0.0007865	GO:0007188	сигнальный путь рецептора, связанного с G-белком, модулирующий аденилатциклазу
BIEC2_124717	10:53976398	<i>GRIK2</i>	-	0.0006631	GO:0050804	модуляция синаптической передачи
BIEC2_275332	14:90328895	<i>HEXB</i>	0.0008033	-	GO:0001501	развитие костной системы
		<i>GFM2</i>			GO:0032543	положительная регуляция клеточной пролиферации
UKUL191	1:105140620	<i>IGF1R</i>	0.0001121	-	GO:0008284	положительная регуляция клеточной пролиферации
BIEC2_155534	11:45542428	<i>INPP5K</i>	0.0007336	-	GO:0001933	негативная регуляция фосфорилирования белков
BIEC2_908994	5:53136645	<i>KCND3</i>	-	0.000701	GO:0071805	трансмембранный транспорт ионов калия
BIEC2-144059	11:22985500	<i>MED1</i>	0.0007038	-	GO:0000902	клеточный морфогенез
AX-105000894	7:96769215	<i>METTL15</i>	-	0.0007262	GO:0032259	метилирование
Affx-102902050	5:87153725	<i>NEGR1</i>	-	0.0003762	GO:0010976	положительная регуляция развития проекций нейронов
BIEC2_357099	16:68809463	<i>NEK11</i>	-	0.0008988	GO:0016572	фосфорилирование гистонов
BIEC2_117960	10:39882171	<i>NT5E</i>	0.0003571	$8.929 * 10^{-5}$	GO:0009166	катаболический процесс нуклеотидов
BIEC2_554615	20:29466156	<i>OR12D2N</i>	-	0.0003629	GO:0050911	обнаружение химических стимулов, участвующих в сенсорном восприятии запаха
BIEC2-187196	12:17084040	<i>OR4C269P</i>	$1.7 * 10^{-5}$	0.0007977	-	-
BIEC2_461433	2:21106066	<i>OSCP1</i>	0.000956	-	GO:1990961	трансмембранный экспорт лекарственных средств
BIEC2_553357	21:13990313	<i>PDE4D</i>	-	0.0003479	GO:0007165	сигнальная трансдукция
BIEC2_952801	6:46670386	<i>PIK3C2G</i>	-	0.0003862	GO:0048015	передача сигналов, опосредованная фосфатидилинозитом
BIEC2_583163	22:15520818	<i>PLCB1</i>	0.0005666	0.000792	GO:0000086	переход G2/M митотического клеточного цикла
BIEC2_331272	16:18885332	<i>PPP4R2</i>	0.0007754	-	GO:0006470	дефосфорилирование белков
BIEC2_651775	24:44695504	<i>RCOR1</i>	0.0008017	-	GO:0016575	деацетилирование гистонов
BIEC2_207487	12:29515936	<i>RELA</i>	-	0.0009667	GO:0006357	регуляция транскрипции с промотора РНК-полимеразы II
		<i>SIPA1</i>			GO:0051056	регуляция передачи сигнала, опосредованной малой ГТФазой

Продолжение таблицы 1

1	2	3	4	5	6	7
BIEC2_844810	4:4272449	<i>RELN</i>	-	0.0008289	GO:0001764	миграция нейонов
BIEC2_606629	22:5228365	<i>RIN2</i>	-	0.0008374	GO:0010595	положительная регуляция миграции эндотелиальных клеток
BIEC2_1006740	7:74196933	<i>RRM1</i>	0.0006299	-	GO:0006260	репликация ДНК
AX-1045528930	9:19411659	<i>RRS1</i>	-	0.0009052	GO:0000027	сборка большой субъединицы рибосомы
Affx-101191469	1:78170974	<i>SLC35F3</i>	0.0003044	-	GO:0015888	транспорт тиамина
BIEC2_835239	31:8606400	<i>SOD2</i>	0.0002175	-	GO:0001315	возрастная реакция на активные формы кислорода
BIEC2_878560	31:8606149		0.0001386	-		
BIEC2_154252	11:42866856	<i>SPAG5</i>	-	0.0008988	GO:0051301	деление клеток
BIEC2_736519	28:26656456	<i>SYCP3</i>	0.0005929	-	GO:0051321	мейотический клеточный цикл
BIEC2_1063214	8:76414093	<i>TCF4</i>	0.0003984	-	GO:0006357	регуляция транскрипции с промотора РНК-полимеразы II
BIEC2_504893	2:106480107	<i>TRPC3</i>	0.0008173	-	GO:0070588	трансмембранный транспорт ионов кальция
BIEC2_1028408	8:10248408	<i>TTC28</i>	-	0.0008527	GO:0007346	регуляция митотического клеточного цикла
BIEC2_50291	1:115156781	<i>TUBGCP5</i>	-	0.0009209	GO:0007020	образование микротрубочек
BIEC2_988205	7:26791905	<i>UBE4A</i>	-	0.0004775	GO:0000209	полиубиквитинирование белка
BIEC2_773787	28:26943895	<i>WASHC3</i>	-	0.0006032	-	-
BIEC2_1027876	8:9374074	<i>ZMAT5</i>	5.608*10 ⁻⁵	-	-	-
BIEC2_1066438	8:92652089	<i>ZNF407</i>	-	0.0004628	GO:0010468	регуляция экспрессии генов

Примечания: 1 Only P-values below the selected significance threshold 0.001 are shown.
 2 Only one term per gene is shown. See Table S3 for the complete DAVID annotation report, including Gene Ontology for cellular components (CCs) and molecular processes (MPs).
 ** - Идентификационный номер термина геной онтологий

Из данных таблицы 1 видно, что между двумя наборами маркеров, связанных с соответствующими признаками (промеры и живая масса), почти не было перекрытия. Только два полиморфизма, BIEC2_117960 и BIEC2-187196, показали значительную ассоциацию с обоими признаками. Первый маркер был связан с геном *OR4C269P*, для которого не было доступной геной онтологической аннотации, в то время как последний был связан с геном экто-5'-нуклеотидазы (*NT5E*), который участвует в метаболизме аденозинфосфатов. Идентифицированные гены играют регуляторную или сигнальную роль в разнообразных биологических процессах, охватывая широкий спектр уровней, начиная с клеточного и заканчивая организменным. Гены *BMP6*, *DDR3* и *CREB3L1* участвуют в развитии и метаболизме соединительных тканей, включая костную. Эти гены содержат SNP, которые, согласно нашим данным, ассоциированы с живой массой у отечественных лошадей. Ген *BMP6* содержал три полиморфизма, ассоциированных с продуктивными качествами, что было самым высоким числом полиморфизмов среди всех генов. Ряд генов, включая *DPF1*, *GNAT3*, *NEGR1* и т.д., были аннотированы как участвующие в развитии нервной системы. Так, ген *NEGR1* связан с регуляцией пищевого и двигательного поведения, в то время как ген *GNAT3* влияет на восприятие вкуса. Гены *BMP6*, *RELA1*, *AIM2*, *PDE4D* и *IGF1R* помимо своих других функций, участвуют в регуляции иммунной системы. Ген *EIF2AK4* связан с клеточной реакцией на холодовой стресс и дефицит белков.

Среди всех функционально аннотированных генов можно отметить некоторые определенные аспекты биологических процессов, потенциально связанных с интересующими признаками. Во-первых, развитие соединительных тканей и костной системы, которые имеют решающее значение для поддержания животным своего веса и размеров. Во-вторых, развитие нервной системы: более специфическое влияние генов *GNAT3* и *NEGR1* на предпочтения лошадей в пище и, следовательно, косвенное воздействие на их рост может представлять интересную тему для будущих исследований. В-третьих, регуляция иммунных процессов, которые могут оказывать влияние на рост путем воздействия на общее состояние здоровья.

Выводы

Проведенное полногеномное ассоциативное исследование позволило выявить следующее. Обнаружено, что 60 SNP ассоциированы с одним из двух исследуемых признаков (живая масса и промеры) и связаны с функционально аннотированными генами лошадей. Среди идентифицированных генов были гены, участвующие в различных биологических процессах в качестве регуляторных и сигнальных факторов. Необходимо отметить, что почти все значимые полиморфизмы были независимо связаны с промерами или живой массой несмотря на то, что между этими признаками существует очевидная корреляция. Среди всех функционально аннотированных генов можно отметить некоторые определенные аспекты биологических процессов, потенциально связанных с интересующими признаками. Однако следует иметь в виду, что аннотации генов, выполненные с помощью «Gene ontology», основаны главным образом на данных о человеке и модельных животных. В результате истинная физиологическая роль идентифицированных генов у лошадей может несколько отличаться. Кроме того, возможные ассоциации вариантов, которые пока не были идентифицированы, требуют дополнительного исследования с учетом обновленных аннотационных данных для геномов лошадей.

Благодарность: Работа выполнена в рамках научного проекта грантового финансирования МНВО РК на 2022-2024 гг. ИРН № AP14870614 «Генетическое маркирование продуктивных качеств казахской лошади типа джабе на основе SNP-генотипирования с широким покрытием генома», а также научно-технической программы ПЦФ МСХ РК на 2021-2023 гг. ИРН № BR10764999 «Разработка технологий эффективного управления селекционным процессом и сохранения генофонда в коневодстве».

Список литературы

1. Wade C.M. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse [Text] / C.M. Wade, E. Giulotto, S. Sigurdsson, et al. // Science. – 2009. – V. 326. – P. 865-867.
2. Finno C.J. Applied equine genetics [Text] / C.J. Finno, D.L. Bannasch // Equine Veterinary Journal. - 2014. – V. 46. – P. 538-544.
3. Hill E.W. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses [Text] / E.W. Hill, B.A. McGivney, J. Gu, et al. // BMC Genomics. – 2010. – V. 11. – P. 552-1-552-10.
4. Rieder S. Molecular tests for coat colours in horses [Text] / S. Rieder // J Anim Breed Genet. – 2009. – V. 126. – P. 415-424.
5. Brosnahan M.M. Equine clinical genomics: A clinician's primer [Text] / M.M. Brosnahan, S.A. Brooks, D.F. Antczak // Equine Vet. – 2010. – V. 42. – P. 658-670.
6. Chang C.C. Second-generation PLINK: rising to the challenge of larger and richer datasets [Text] / C.C. Chang, C.C. Chow, L.C. Tellier, et al. // GigaScience. - 2015. – V. 4. – P. 7-1-7-16.
7. Бейшова И.С. Изучение генетического разнообразия отечественных пород лошадей с использованием полногеномного анализа SNP [Текст] / И.С. Бейшова, Д.А. Гриценко, М.Х. Шамекова, А.С. Пожарский, Т.В. Ульянова, А.М. Ковальчук // Izdenister Natigeler. – 2023. - № 3 (99). - С. 48-58.

8. McLaren W. The Ensembl Variant Effect Predictor [Text] / W. McLaren, L. Gil, S.E. Hunt, et al. // Genome Biology. – 2016. – V. 17. – P. 122-1-122-14.

9. Sherman B.T. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update) [Text] / B.T. Sherman, M. Hao, J. Qiu, et al. // Nucleic Acids Research. - 2022. – V. 50. – P. 216-221.

10. Bailey E. Genetics of Thoroughbred Racehorse Performance [Text] / E. Bailey, J.L. Petersen, T.S. Kalbfleisch // Annual Review of Animal Biosciences. - 2022. – V. 10. – P. 131-150.

11. Raudsepp T. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era [Text] / T. Raudsepp, C.J. Finno, R.R. Bellone, J.L. Petersen // Animal Genetics. - 2019. – V. 50. – P. 569-597.

References

1. Wade C.M. Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse [Text] / C.M. Wade, E. Giulotto, S. Sigurdsson, et al. // Science. – 2009. – V. 326. – P. 865-867.

2. Finno C.J. Applied equine genetics [Text] / C.J. Finno, D.L. Bannasch // Equine Veterinary Journal. - 2014. – V. 46. – P. 538-544.

3. Hill E.W. A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses [Text] / E.W. Hill, B.A. McGivney, J. Gu, et al. // BMC Genomics. – 2010. – V. 11. – P. 552-1-552-10.

4. Rieder S. Molecular tests for coat colours in horses [Text] / S. Rieder // J Anim Breed Genet. – 2009. – V. 126. – P. 415-424.

5. Brosnahan M.M. Equine clinical genomics: A clinician's primer [Text] / M.M. Brosnahan, S.A. Brooks, D.F. Antczak // Equine Vet. – 2010. – V. 42. – P. 658-670.

6. Chang C.C. Second-generation PLINK: rising to the challenge of larger and richer datasets [Text] / C.C. Chang, C.C. Chow, L.C. Tellier, et al. // GigaScience. - 2015. – V. 4. – P. 7-1-7-16.

7. Bejshova I.S. Izuchenie geneticheskogo raznoobraziya otechestvennykh porod loshadej s ispol'zovaniem polnogenomnogo analiza SNP [Tekst] / I.S. Bejshova, D.A. Gritsenko, M.KH. SHamekova, A.S. Pozharskij, T.V. Ul'yanova, A.M. Koval'chuk // Izdenister Natigeler. – 2023. - № 3 (99). - S. 48-58.

8. McLaren W. The Ensembl Variant Effect Predictor [Text] / W. McLaren, L. Gil, S.E. Hunt, et al. // Genome Biology. – 2016. – V. 17. – P. 122-1-122-14.

9. Sherman B.T. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update) [Text] / B.T. Sherman, M. Hao, J. Qiu, et al. // Nucleic Acids Research. - 2022. – V. 50. – P. 216-221.

10. Bailey E. Genetics of Thoroughbred Racehorse Performance [Text] / E. Bailey, J.L. Petersen, T.S. Kalbfleisch // Annual Review of Animal Biosciences. - 2022. – V. 10. – P. 131-150.

11. Raudsepp T. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era [Text] / T. Raudsepp, C.J. Finno, R.R. Bellone, J.L. Petersen // Animal Genetics. - 2019. – V. 50. – P. 569-597.

**И.С. Бейшова¹, Д.А. Гриценко², М.Х. Шамекова², А.С. Пожарский²,
Т.В. Ульянова^{1*}, А.М. Ковальчук¹**

¹ «Жәңгір хан атындағы Батыс Қазақстан аграрлық-техникалық университеті» КЕАҚ,
Орал қаласы, Қазақстан Республикасы, indira_bei@mail.ru,
tatyana.poddudinskaya@gmail.com*, kovalchuk_s89@mail.ru

² «Өсімдіктер биологиясы және биотехнологиясы институты» ШЖҚ РМК, Алматы
қаласы, Қазақстан Республикасы, d.kopytina@gmail.com, shamekov@gmail.com,
aspozharsky@gmail.com

ОТАНДЫҚ ЖЫЛҚЫ ТҰҚЫМДАРЫНЫҢ ӨНІМДІЛІК КӨРСЕТКІШТЕРІНІҢ АССОЦИАЦИЯЛАРЫН ТОЛЫҚ ГЕНОМДЫҚ ІЗДЕУ

Аңдатпа

Бұл мақалада Equine 80k HTS («Illumina Inc.», АҚШ) реагент жинағы арқылы талданған SNP генотиптеу деректері негізінде алынған отандық жылқы тұқымдарындағы өнімділік көрсеткіштері бір нуклеотидті полиморфизмдердің (SNPs) ассоциацияларын толық геномдық іздеу нәтижелері берілген. Зерттеу нысаны ретінде келесі жылқылардың тұқымдары мен типтері: қазақ типті жабы (n = 631), адай (n = 303) және найман (n = 158), мұғалжар (n = 584), көшім (n = 226) және қостанай (n = 116) тұқымдары болды. Үлгі алу алдында жылқылардың фенотиптік деректері бойынша мәліметтерді жинау жүргізіліп, келесі көрсеткіштер өлшенді: шоқтық биіктігі, дененің қиғаш ұзындығы, кеуде айналымы, білектің айналымы және тірі салмағы. Сапалық бақылау нәтижесінде таңдалған SNP-дің жалпы саны 60 987-ге тең болды. Ассоциацияларды толық геномдық іздеу PLINK1.9 бағдарламасындағы бірнеше салыстыру нәтижесінде р-мәнінің коррекциясы бар адаптивті Монте-Карло ауыстыруымен сызықтық регрессия арқылы орындалды. Талдау дәнекер тінінің, сүйек, жүйке, иммундық жүйені реттеуге және басқа да процестерге қатысатын гендермен байланысты 60 полиморфизмді анықтады. Жүргізілген жұмыс жеке белгілермен байланысты жалғыз SNP анықтау үшін толық геномдық ассоциациялық зерттеулерде мұндай тәсілді қолдану өзектілігі расталды.

Кілт сөздер: ассоциацияларды толық геномдық іздеу, жылқылар, жабы типті, адай типті, найман типті, мұғалжар тұқымы, көшім тұқымы, қостанай тұқымы, SNP.

*I.S. Beishova¹, D.A. Gritsenko², M.Kh. Shamekova², A.S. Pozharskiy²,
T.V. Ulyanova^{1*}, A.M. Kovalchuk¹*

¹ NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, Kazakhstan, indira_bei@mail.ru, tatyana.poddudinskaya@gmail.com*, kovalchuk_s89@mail.ru

² RSE on the REM «Institute of Plant Biology and Biotechnology», Almaty, Kazakhstan, d.kopytina@gmail.com, shamekov@gmail.com, aspzharsky@gmail.com

A GENOME-WIDE ASSOCIATION STUDY OF PRODUCTIVE TRAITS IN TRADITIONAL HORSE BREEDS

Abstract

The present work presents the results of a genome-wide association study of single nucleotide polymorphisms (SNP) with productive traits in traditional horse breeds, obtained on the basis of data from SNP genotyping of animals, carried out with the reagent set Equine 80k HTS («Illumina Inc», USA). Horses of the following breeds and types were studied Kazakh types Zhabe (n = 631), Adai (n = 303) and Naiman (n = 158), Mugalzhar (n = 584), Kushum (n = 226) and Kostanay (n = 116) breeds. Information on phenotypic data of the horses was collected before sampling and the following parameters were measured: the height at the withers, oblique body length, chest circumference, cannon bone circumference, and body weight. The total number of SNPs selected by quality control was 60,987. The genome-wide association study was performed using a linear regression test with the adaptive Monte-Carlo permutation method of p-value correction for multiple comparisons in PLINK1.9 software. The analysis revealed 60 polymorphisms associated with genes involved in the regulation of connective tissue and bone system, nervous system, immune system regulation, and other processes. This work confirmed the relevance of using such an approach in genome-wide association studies to detect single SNPs associated with individual traits.

Key words: genome-wide association studies, horses, Zhabe type, Adai type, Naiman type, Mugalzhar breed, Kushum breed, Kostanay breed, SNP.