

Животноводство и кормопроизводство. 2024. Т. 107, № 4. С. 53-67.
Animal Husbandry and Fodder Production. 2024. Vol. 107, no 4. P. 53-67.

Научная статья
УДК 636.082 (571.1)
doi:10.33284/2658-3135-107-4-53

Оценка качества генеалогических данных в племенных предприятиях Западной Сибири

Евгений Варисович Камалдинов¹, Алексей Фёдорович Петров², Кирилл Николаевич Нарожных³, Павел Николаевич Пальчиков⁴

^{1,2,3}Новосибирский государственный аграрный университет, Новосибирск, Россия

⁴АО «Новосибирскагроплем», с. Ярково, Новосибирской обл., Россия

¹ekamaldinov@yandex.ru, <https://orcid.org/0000-0002-0341-5055>

²lexluterking@yandex.ru, <https://orcid.org/0000-0002-7402-4107>

³narozhnykh@nsau.edu.ru, <https://orcid.org/0000-0002-1519-697X>

⁴v.2899936@yandex.ru

Аннотация. Анализ генеалогических баз данных, содержащих более 6 миллионов записей, выявил значительное количество пропусков в данных о происхождении животных, особенно по быкам-производителям. Это затрудняет внедрение технологий автоматизированного подбора родительских пар и снижает эффективность контроля уровня инбридинга. Установлено, что в среднем 7 % записей о предках до третьего ряда родословной содержат пропуски, из которых 85,5 % приходится на быков-производителей. Это представляет серьёзную проблему при создании региональных цифровых платформ для автоматизации принятия решений в племенном животноводстве, ограничивает возможности объективной оценки племенной ценности животных и препятствует необходимому давлению целенаправленного искусственного отбора и закрепления производителей за маточным поголовьем. Предложены способы повышения качества первичных генеалогических данных, в том числе с использованием зарубежных источников. Показано, что автоматизация процесса принятия решений в племенном молочном скотоводстве позволяет не только ускорить выявление ошибок, но и значительно сократить время на их исправление благодаря создаваемым агрегаторам данных. Этот процесс может стать основой для крупномасштабной верификации и валидации данных первичного зоотехнического учёта с последующей передачей в Федеральную государственную информационную-аналитическую систему племенных ресурсов (ФГИАС ПР). Кроме того, это повысит эффективность работы зоотехников-селекционеров в крупных племенных и товарных хозяйствах, а также позволит отбирать быков-производителей, генетически совместимых с маточным поголовьем как отдельного региона, так и их совокупности.

Ключевые слова: молочный скот, генеалогические данные, информационно-аналитическая система, племенная ценность, селекция, инбридинг, сельское хозяйство

Для цитирования: Оценка качества генеалогических данных в племенных предприятиях Западной Сибири / Е.В. Камалдинов, А.Ф. Петров, К.Н. Нарожных, П.Н. Пальчиков // Животноводство и кормопроизводство. 2024. Т. 107, № 4. С. 53-67. <https://doi.org/10.33284/2658-3135-107-4-53>

Original article

Assessment of the quality of genealogical data in breeding enterprises of Western Siberia

Evgeny V Kamaldinov¹, Alexey F Petrov², Kirill N Narozhnykh³, Pavel N Palchikov⁴

^{1,2,3}Novosibirsk State Agrarian University, Novosibirsk, Russia

⁴JSC Novosibirskagroplem, Yarkovo settlement, Novosibirsk region, Russia

¹ekamaldinov@yandex.ru, <https://orcid.org/0000-0002-0341-5055>

²lexluterking@yandex.ru, <https://orcid.org/0000-0002-7402-4107>

³narozhnykh@nsau.edu.ru, <https://orcid.org/0000-0002-1519-697X>

⁴v.2899936@yandex.ru

Abstract. Analysis of genealogical databases containing more than 6 million records revealed a significant number of gaps in the data on the origin of animals, especially for sires. This complicates the

implementation of automated parental pair selection technologies and reduces the effectiveness of inbreeding level control. It was found that on average 7% of records on ancestors up to the third row of the pedigree contain gaps, of which 85.5% are for sires. This is a serious problem in the creation of regional digital platforms for the automation of decision-making in breeding livestock, limits the possibilities of an objective assessment of the breeding value of animals and hinders the necessary pressure of targeted artificial selection and assignment of sires to breeding stock.

Methods for improving the quality of primary genealogical data are proposed, including using foreign sources. It is shown that automation of the decision-making process in purebred dairy breeding allows not only to speed up the detection of errors, but also to significantly reduce the time for their correction thanks to the created data aggregators. This process can become the basis for large-scale verification and validation of primary zootechnical accounting data with subsequent transfer to the Federal State Information and Analytical System of Breeding Resources (FGIAS PR). In addition, this will increase the efficiency of zootechnicians-breeders in large breeding and commercial farms, and will also allow selecting breeding bulls that are genetically compatible with the breeding stock of both a separate region and their totality.

Keywords: dairy cattle, genealogical data, information and analytical system, breeding value, breeding, inbreeding, agriculture

For citation: Kamaldinov EV, Petrov AF, Narozhnykh KN, Palchikov PN. Assessment of the quality of genealogical data in breeding enterprises of Western Siberia. *Animal Husbandry and Fodder Production*. 2024;107(4):53-67. (In Russ). <https://doi.org/10.33284/2658-3135-107-4-53>

Введение.

Молочное скотоводство – одна из важнейших отраслей сельского хозяйства, обеспечивающая население ценными продуктами питания. Повышение эффективности этой отрасли – приоритетная задача аграрного сектора, в решении которой ключевая роль принадлежит селекционно-племенной работе, направленной на улучшение продуктивных и племенных качеств животных (Маринченко Т.Е., 2019; Швелёва О.М. и Свяженина М.А., 2023). Современное отечественное животноводство претерпевает ускоренную цифровую трансформацию, открывающую новые возможности для интенсификации селекционно-племенной работы. Одним из перспективных направлений такой трансформации является создание информационно-аналитических систем ИАС для генетической оценки и управления селекционным процессом (Cañas-Álvarez JJ et al., 2023; Cole JB et al., 2020). Инструменты цифровых платформ для племенной работы позволяют хранить, обрабатывать и анализировать большие массивы данных, вести электронный учет родословных, оценивать племенную ценность, проводить подбор родительских пар, формировать оптимальную структуру стада, управлять селекционным процессом, а также контролировать и прогнозировать его эффективность (Nilforooshan MA and Saavedra-Jiménez LA, 2020).

За рубежом активно развивается использование геномной селекции с применением искусственного интеллекта и машинного обучения для прогнозирования продуктивности животных (Schaeffer L, 2006; Wiggans GR et al., 2011; Wiggans GR and Carrillo JA, 2022). Широко используются платформы типа CDN (cdn.ca/home) Select Star SA (selectstar.ch) или Interbull (interbull.org). В России разрабатывается Федеральная государственная информационно-аналитическая система племенных ресурсов (ФГИАС ПР), которая должна стать основой для цифровизации племенной работы. Примеры использования ИИ в российском племенном деле пока немногочисленны и находятся на стадии разработки и внедрения (Васильев Н.П. и др., 2024).

Эффективность подобных систем напрямую зависит от качества входных данных, включая генеалогическую информацию (Martin P et al., 2019; Fioretti M et al., 2020; Toghiani S and Van Raden PM, 2021). Ее неполнота или некорректность неизменно приводит к искажению результатов оценки генетического потенциала (Aguilar I et al., 2020).

На протяжении нескольких десятилетий в России практикуется массовое использование импортного племенного материала, что привело к изменению генеалогической структуры отечественного маточного поголовья чёрно-пёстрого скота, преимущественно за счёт поглотительного

скрещивания. Изолированность и локализация баз данных на предприятиях не дают общего представления о реальной генеалогической структуре и сдерживают реализацию крупномасштабной селекции, затрудняя оценку и контроль уровня инбридинга в популяциях разного размера (Cavani L et al., 2018; Nyman S et al., 2022; Silva МНМА et al., 2019). Выявленные проблемы с качеством генеалогических данных подчеркивают необходимость дальнейшей разработки и совершенствования цифровых платформ для объективной и эффективной оценки генетического потенциала животных и управления селекционным процессом.

Цель исследования.

Оценка качества генеалогических данных племенных предприятий Новосибирской и Кемеровской областей для последующей разработки цифровых платформ, обеспечивающих объективный прогноз генетического потенциала.

Материалы и методы исследования.

Объект исследования. В качестве источников генеалогической информации были использованы базы данных ряда племенных предприятий Новосибирской и Кемеровской областей по разведению молочного скота голштинской породы, открытые данные, публикуемые ФГБНУ ВНИИ-плем (https://vniipleм.ru/l/gisc/bd_ng/db_ng_mol_krs/) и Canadian Dairy Network (CDN) (https://www.cdn.ca/files_ge_datafiles.php) с общим объемом данных, превышающим 6 млн записей.

Схема эксперимента. Для сопоставления рядов родословных отдельно взятых групп с пропущенными записями о происхождении по всем исследуемым базам данных региона было проведено ранжирование сумм частот, где минимальный ранг соответствует минимальной частоте. Также было проведено ранжирование животноводческих предприятий по степени заполненности генеалогической базы данных с применением интегрального показателя количества пропущенных записей о происхождении животных (без учёта группы «быки-осеменители»). Построены генеалогические деревья линий и семейств, представленных в хозяйствах.

Статистическая обработка. Для обработки данных, проведения анализа и построения генеалогических деревьев использовалось программное обеспечение R (Giorgi FM et al., 2022) с подключением специализированных пакетов (dplyr) (Wickham H and Grolemond GR, 2017).

Для оценки сходства хозяйств по качеству генеалогической информации и выявления групп хозяйств со схожими показателями пропущенных записей был проведён кластерный анализ с использованием расстояния Чебышева (Chen TY, 2019) и метода дальнего соседа (Поддубный В.В. и др., 2006). Для данных с рангами абсолютных частот была выбрана евклидова дистанция (Suwanda R et al., 2020) и метод Уорда (Gaikadi S and Kumar SV, 2024) которые показали наилучшие результаты. Для проведения кластерного анализа и визуализации дендрограмм были использованы пакеты factoextra (Irnawati I et al., 2021) и cluster (Bivand R, 2022). В качестве входных данных для кластерного анализа были использованы две матрицы: матрица с процентом пропущенных записей в генеалогии для каждой группы животных (быки, коровы, тёлки) в 18 хозяйствах и матрица с рангами абсолютных частот пропущенных записей о происхождении в группах животных до третьего ряда родословной. Данный метод позволяет визуализировать процесс объединения хозяйств в кластеры на дендрограмме, что делает его более наглядным и интерпретируемым.

Результаты исследований.

Анализ заполненности родословных пробандов племенных предприятий Новосибирской и Кемеровской областей выявил существенную вариабельность качества генеалогической информации в зависимости от источника. Оценка доли пропущенных записей о происхождении животных до третьего ряда родословной показала, что в среднем по региону этот показатель составляет около 7 % (табл. 1). Однако распределение пропущенных данных неоднородно: наибольшее число пропусков наблюдается в третьем ряду родословной по материнской линии (МММ), в то время как записи о происхождении по отцовской линии (О) являются наиболее полными.

Таблица 1. Ранги абсолютных частот пропущенных записей о происхождении в группах животных региональных предприятий
 Table 1. Ranks of absolute frequencies of missed records of origin in animal groups of regional enterprises

Группа/ Group	Ряд родословной / Pedigreeing row																Градусная / Row group
	O/S	OO/ SS	OM/ SD	OOO/ SSS	OOOM/ SSSD	OMO/ SDS	OMOM/ SSSD	M/ D	MM/ DD	MO/ DS	MMOM/ DDD	MMOO/ DDS	MOOM/ DSD	MOOO/ SDD			
Всего / Total	1	2	2	3	3	3	3	3	3	2	2	3	3	3	3	3	0,778
Всего по бычкам-осеменителям / Total for sters	1	3	5	6	11	9	9	13	13	2	8	4	14	12	10	6	0,827
Всего по бычкам / Total for bulls	2	3	5	6	11	9	9	13	13	1	8	4	14	10	12	7	0,827
Всего по коровам / Total for cows	1	3	8	5	11	7	7	13	13	2	10	4	14	9	12	6	0,705
Всего по тёлкам / Total for heifers	2	4	3	7	9	12	12	13	13	1	5	6	14	11	10	8	0,876
Всего по тёлкам / Total for heifers	2	6	3	9	7	13	13	11	11	1	4	5	12	14	8	10	0,876

Такая закономерность может быть объяснена тем, что зоотехники-селекционеры на предпрятых традиционно уделяют большее внимание заполнению отцовских рядов предков (табл. 1). В условиях широкого использования искусственного осеменения информация о быках-производителях, как правило, более доступна и полностью отражается в сопроводительной документации.

Важно отметить, что высокая доля пропущенных записей, особенно в отношении быков-производителей (85,5 % от общего числа пропусков), представляет собой серьезную проблему для создания полноценной региональной ИАС. Неполные родословные затрудняют адекватную оценку племенной ценности животных и ограничивают возможность проведения эффективного подбора родительских пар, направленного на повышение генетического потенциала поголовья.

Дендрограмма, построенная на основе рангов частот пропущенных генеалогических записей до третьего ряда родословной и представленная на рисунке 1, выявила два основных кластера, разделяющих анализируемые группы животных, что указывает на существенные различия в структуре пропусков между ними.

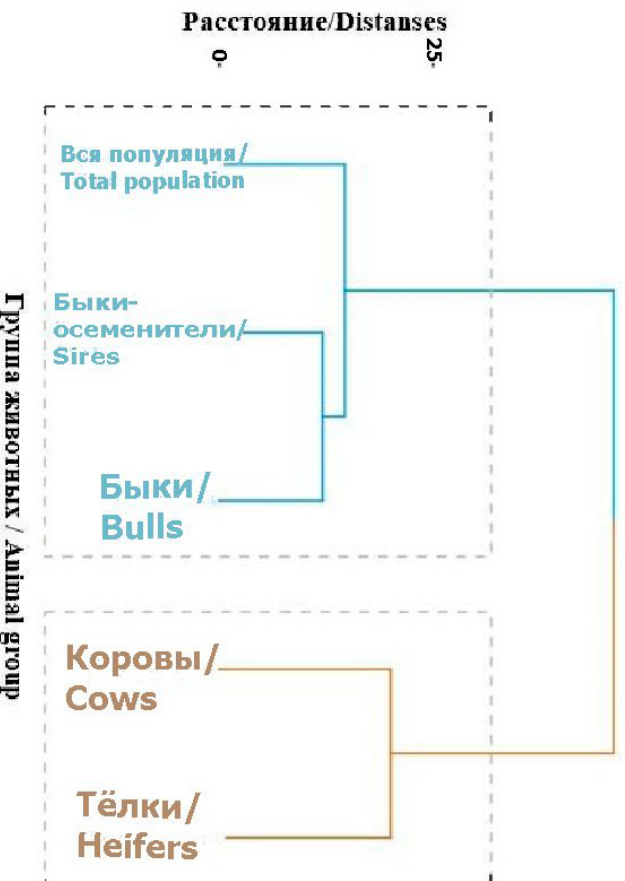


Рисунок 1. Дендрограмма групп животных по рангам пропущенных генеалогических записей
Figure 1. Dendrogram of animal groups by ranks of missing genealogical records

Первый кластер объединяет всю популяцию, быков-осеменителей и быков, свидетельствуя о высокой степени сходства в структуре рангов пропущенных записей в этих группах (дистанция=2). Это означает, что распределение пропусков по разным предкам для быков и быков-осеменителей практически идентично распределению соответствующих пропусков для всей совокупности животных. Такой результат объясняется тем, что быки и быки-осеменители составляют значительную часть популяции, происхождение которых играет ключевую роль в процессе подбора родительских пар. Второй кластер включает коров и тёлочек, что указывает на отличие в структуре рангов пропущенных генеалогических записей для этих групп по сравнению с быками и всей популяцией. Данное различие может быть обусловлено особенностями ведения генеалогического учёта для разных половозрастных групп животных. Информация о происхождении коров и тёлочек заполняется менее тщательно относительно таковой о происхождении быков. Необходимо отме-

тительное расстояние между двумя кластерами на дендрограмме (дистанция=5.5), которое подтверждает существенные различия в структуре пропущенных генеалогических записей. Проведённый анализ был основан на рангах, а не на абсолютных частотах, что позволило снизить влияние выбросов и различий в общем числе пропусков для разных групп животных, фокусируясь на сравнительном анализе структуры пропусков.

Дендрограмма, построенная на основе рангов частот пропущенных генеалогических записей до третьего ряда родословной (рис. 2), выявила пять кластеров, отражающих структуру пропусков информации о происхождении животных. Кластеры формируются группами предков с похожими паттернами отсутствующих данных, что позволяет выявить тенденции в заполнении генеалогической информации.

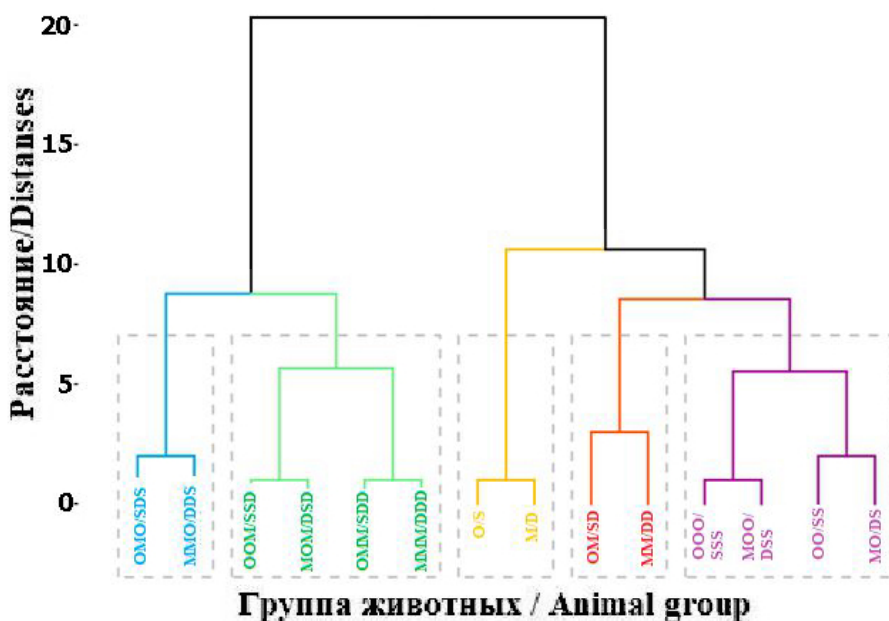


Рисунок 2. Кластеризация комбинаций предков до третьего ряда родословной по рангам частот пропусков

Figure 2. Clustering of ancestral combinations up to the third row of the pedigree by omission frequency rank

Кластер 1 объединяет предков ОМО и ММО (предки третьего ряда родословной с одним материнским и двумя отцовскими предками в каждом случае). Близость этих групп на дендрограмме предполагает наличие схожих причин пропусков информации о происхождении для этих комбинаций предков. Сходство паттернов может быть обусловлено сложностями получения информации о более дальних родственниках, особенно по материнской линии, что объясняет объединение в кластере предков третьего ряда. Кластер 2 включает предков ООМ, МОМ, ОММ и МММ (предки третьего ряда родословной с двумя материнскими и одним отцовским предками, либо с одним материнским и двумя отцовскими, включающие информацию до третьего ряда включительно). В кластер 3 входят предки О и М (предки первого ряда – отец и мать), что указывает на сходную частоту пропусков для непосредственных родителей. Это может быть связано с тем, что информация об отце и матери зачастую записывается одновременно, и если данные об одном из родителей отсутствуют, то с большой вероятностью будут отсутствовать и данные о втором. В кластер 4 входят предки ОМ и ММ (предки второго ряда), близость данного кластера к кластеру 3 связана с тенденцией к более полному заполнению данных о родителях, чем о более дальних предках. Кластер 5 включает ООО, МОО, ОО и МО (предки второго и третьего ряда). Присутствие в

одном кластере предков разных рядов может говорить о том, что на практике заполнение генеалогической информации до второго или третьего ряда родословной обусловлен, скорее всего, практической необходимостью (наличие или отсутствие племенной документации), чем последовательной работой по её заполнению, что объясняет близость предков как второго, так и третьего рядов. Проведённый анализ, основанный на рангах частот пропущенных записей, позволяет сфокусироваться на структуре пропусков и выявить общие тенденции, независимо от абсолютного количества отсутствующих данных в разных группах животных. Дополнительное исследование конкретных рангов внутри кластеров может дать более детальную информацию о причинах возникновения пропусков.

Результаты ранжирования хозяйств по степени заполненности баз данных (табл. 2) подтверждают наличие значительной вариабельности между предприятиями в отношении качества генеалогической информации. Наиболее высокие ранги и, соответственно, наилучшее качество данных демонстрируют хозяйства 16, 18 и 14, в которых ведётся целенаправленная работа по корректировке и актуализации генеалогической информации.

Таблица 2. Ранжирование предприятий по качеству генеалогической информации

Table 2. Ranking of enterprises by quality of genealogical information

Хозяйство/ Farm	Пропуски / Omissions									ранг/ Rank
	всего рядов предков/ Total rows of ancestors	абсолютное количество/ Absolute number			проценты / Percents			итого (без быков) / Total (without bulls)		
		быки /Bulls	коровы /Cows	тёлки/ Heifers	быки /Bulls	коровы /Cows	тёлки/ Heifers	абсолют/ Absolute number	проценты / Percents	
1	84938	5211	561	488	6,14	0,66	0,57	1049	3,46	11
2	123340	3997	666	232	3,25	0,54	0,19	898	1,32	9
3	97146	3997	666	2524	4,13	0,69	2,60	3190	1,80	12
4	178024	3539	530	55	1,99	0,30	0,03	585	1,66	4
5	102718	3650	148	12	3,55	0,14	0,01	160	0,79	6
6	212156	12224	839	922	5,76	0,40	0,43	1761	2,22	10
7	130760	3097	800	922	2,37	0,61	0,71	1722	1,66	7
8	116088	2572	848	317	2,22	0,73	0,27	1165	0,50	6
9	411208	29283	8151	1037	7,12	1,98	0,25	9188	1,11	13
10	450198	46037	3287	965	10,23	0,73	0,21	4252	0,47	14
11	773346	85016	8375	116	10,99	1,08	0,01	8491	0,55	16
12	642180	86987	5699	7431	13,55	0,89	1,16	13130	1,03	17
13	210952	35757	0	154	16,95	0,00	0,07	154	0,04	18
14	1545866	5699	0	5014	0,37	0,00	0,32	5014	0,16	3
15	105588	9043	366	3050	8,56	0,35	2,89	3416	1,62	15
16	540876	0	0	0	0,00	0,00	0,00	0	0,00	1
17	120190	0	2721	119	0,00	2,26	0,10	2840	1,18	5
18	173124	0	0	6	0,00	0,00	0,00	6	0,00	2
T	6018698	336109	32991	23364	5,58	0,55	0,39	56355	0,94	

T – Итого / Total

Дендрограмма, построенная с использованием метрики Чебышева и метода полной связи, иллюстрирует структуру сходства/различия 18 племенных хозяйств по проценту пропущенных генеалогических записей для быков, коров и телок (рис. 3).

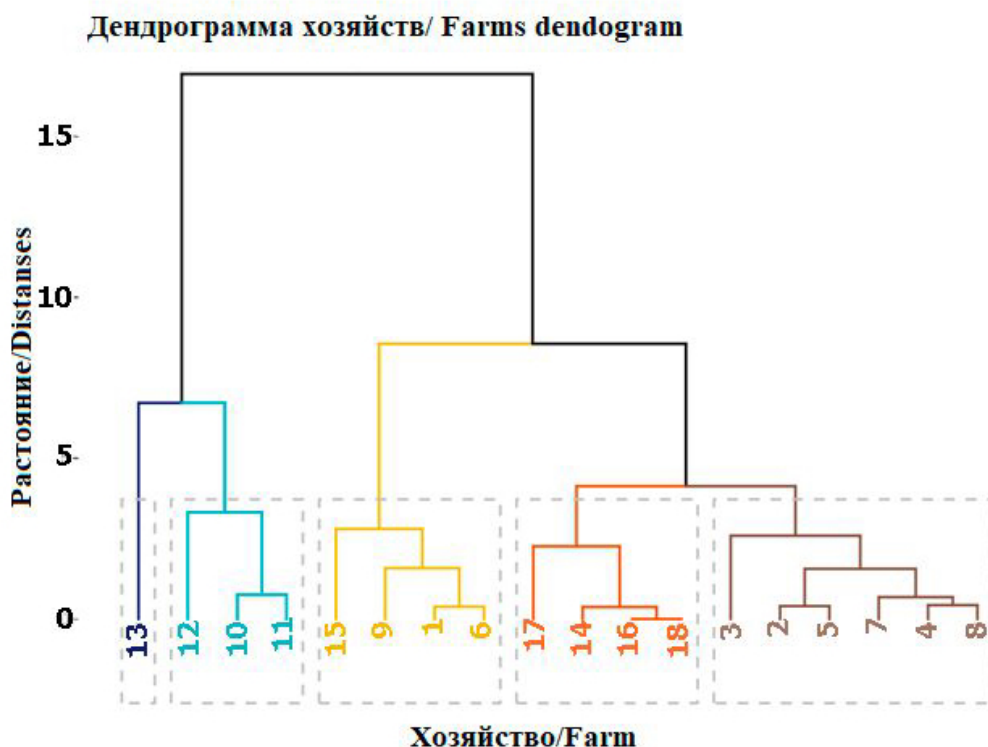


Рисунок 3. Дендрограмма сходства/различия хозяйств по качеству генеалогической информации (1-18, условные обозначения племенных хозяйств)

Figure 3. Dendrogram of similarity/difference of farms according to quality of genealogical information (1-18, designations of breeding farms)

Разбивка на 5 кластеров, выделенных на дендрограмме, позволяет детально проанализировать различия между хозяйствами по качеству генеалогической информации. Кластер 1 обособлен на самом высоком уровне иерархии и представлен исключительно хозяйством 13, характеризующимся экстремально высоким процентом пропусков генеалогических записей по быкам (почти 17 %), что значительно превышает показатели всех остальных хозяйств. Кластер 2 объединяет хозяйства 12, 10 и 11, отличающиеся высоким процентом пропущенных записей по быкам (более 10 % в хозяйствах 11 и 12). Дополнительно, в хозяйстве 12 отмечается повышенный процент пропусков по телкам (более 1 %), в то время как хозяйство 10 демонстрирует средние значения пропусков по коровам и телкам, не превышающие 1 %. Кластер 3 включает хозяйства 15, 9, 1 и 6, демонстрирующие более высокий, чем в среднем по выборке, процент пропущенных записей. Хозяйства 9 и 15 характеризуются повышенным процентом пропусков по коровам (около 2 % для хозяйства 9) и телкам (около 3 % для хозяйства 15) соответственно. В то же время хозяйства 1 и 6 имеют средние значения процента пропусков, находящиеся в интервале 3-6 % для быков и не более 0.7 % для коров и телок. Кластер 4 объединяет хозяйства 17, 14, 16 и 18, демонстрирующие высокое качество генеалогической информации. Хозяйства 16, 14 и 18 характеризуются минимальным количеством пропусков по всем группам животных, близким к нулю. Хозяйство 17, имея аналогичный показатель по быкам (около 0 %), демонстрирует наибольший процент пропусков по коровам внутри данного кластера (более 2 %). Кластер 5 включает хозяйства 3, 2, 5, 7, 4 и 8, демонстрирующие низкий процент пропущенных записей по всем трем группам животных, что указывает на достаточно высокое качество генеалогических данных (рис. 3). Процент пропущенных записей по быкам в этих хозяйствах, как правило, не превышает 5 %, за исключением хозяйства 3, для которого характерен несколько более высокий процент пропусков по телкам. В целом, проведенный кла-

стерный анализ позволяет выявить группы хозяйств со сходными характеристиками в отношении качества генеалогической информации и может служить основой для разработки рекомендаций по улучшению ведения генеалогического учета.

Обсуждение полученных результатов.

Проведённое исследование позволило оценить качество генеалогических и количественных данных племенных предприятий Новосибирской и Кемеровской областей для разработки региональной (ИАС) генетической оценки и селекции молочного скота. Анализ выявил существенную вариабельность качества генеалогической информации между хозяйствами. В среднем по региону доля отсутствующих записей о происхождении животных до третьего ряда родословной составляет около 7 %. Однако распределение пропущенных данных неоднородно: больше всего пропусков наблюдается в третьем ряду родословной по материнской линии, тогда как записи о происхождении по отцовской линии наиболее полные. Это может объясняться тем, что зоотехники-селекционеры традиционно уделяют больше внимания заполнению отцовских рядов предков, особенно при широком использовании искусственного осеменения (Castro-Vásquez R et al., 2023; Nader N et al., 2021; Cole JB et al., 2021). Высокая доля пропущенных записей, особенно по быкам-производителям (85,5 % от общего числа пропусков), представляет серьёзную проблему для создания качественных баз данных регионального и федерального масштабов, поскольку неполные родословные затрудняют оценку племенной ценности и ограничивают возможности эффективного подбора родительских пар.

Оценка качества количественных данных (удой и содержание жира за 305 дней лактации) показала, что распределение этих признаков во многих хозяйствах статистически значимо отличается от гауссовского. Наиболее частые отклонения связаны с асимметрией и двухвершинностью распределений (Камалдинов Е.В. и др., 2022). Причины такого отклонения могут быть различны: генетическая гетерогенность стада, различия в условиях кормления и содержания животных, особенности селекционной работы. Несоответствие распределения количественных признаков нормальному закону может снижать точность оценки племенной ценности, особенно при использовании смешанных линейных моделей.

Для улучшения качества генеалогических данных и повышения эффективности разработки и использования региональной ИАС рекомендуется восстанавливать и актуализировать генеалогическую информацию на племенных предприятиях, уделяя особое внимание заполнению данных о происхождении быков-производителей. Целесообразно также подключить дополнительные источники генеалогической информации, включая зарубежные базы данных, и внедрить инструменты ИАС для автоматизации поиска и исправления ошибок. Важным аспектом является разработка и внедрение системы мотивации племенных предприятий для поддержания высокого качества генеалогической информации (Brito LF et al., 2020; Georges M et al., 2019; Mrode R et al., 2019).

Для корректной оценки племенной ценности и эффективного подбора родительских пар при негауссовском распределении количественных признаков следует рассмотреть применение специальных статистических методов или трансформацию данных для приближения их к нормальному распределению. Перспективным направлением также является разработка и внедрение категориальных систем оценки племенной ценности, учитывающих различия в уровне продуктивности и условиях содержания животных на разных предприятиях.

В настоящее время активно развиваются методы ИИ для контроля качества данных, включая методы обнаружения выбросов (Aggarwal CC, 2015) и методы предсказания пропущенных значений (Little RJA, 1988). Эти методы могут быть перспективными для автоматизации процесса оценки качества генеалогических данных в животноводстве и повышения достоверности информации в ИАС. Однако практическое применение ИИ для этих целей в племенной работе пока ограничено и требует дальнейших исследований. В перспективе разработка и внедрение инструментов на основе ИИ, таких как системы автоматического обнаружения ошибок и аномалий в данных, могут

существенно повысить эффективность контроля качества генеалогических баз данных и обеспечить более высокую достоверность информации, используемой в селекционно-племенной работе.

Реализация этих рекомендаций будет способствовать повышению эффективности разработки и использования региональной ИАС, ускорению генетического прогресса в популяции молочного скота и повышению экономической эффективности молочного скотоводства в регионе.

Заключение.

Исследование, посвящённое оценке качества генеалогических данных молочного скота в племенных предприятиях Западной Сибири, выявило ряд важных закономерностей и проблем. Анализ генеалогических баз данных показал высокую вариабельность качества данных между хозяйствами. В среднем по региону доля отсутствующих записей о происхождении животных до третьего ряда родословной составляет около 7 %, однако в некоторых хозяйствах этот показатель значительно выше. Это может свидетельствовать о недостаточном внимании к ведению генеалогического учёта и необходимости разработки унифицированных стандартов племенной документации. Особую проблему представляет высокая доля пропусков по быкам-производителям (85,5 % от общего числа пропусков). Это затрудняет контроль уровня инбридинга и может привести к его нежелательному повышению, негативно влияющему на продуктивность, воспроизводительные качества и здоровье животных.

Неполнота генеалогической информации серьёзно препятствует созданию полноценной региональной ИАС и может привести к недостоверной оценке племенной ценности и неэффективному подбору родительских пар, замедляя генетический прогресс. Поэтому результаты исследования подчеркивают необходимость разработки и внедрения инструментов для автоматизации поиска и корректировки генеалогических данных, а также системы контроля качества вводимой информации. Создание региональной ИАС, основанной на достоверных и полных генеалогических данных, позволит повысить эффективность селекционно-племенной работы в молочном скотоводстве Западной Сибири и ускорить генетический прогресс в отрасли. В перспективе необходимо продолжить исследования, направленные на совершенствование методов оценки и контроля качества генеалогической информации, а также на разработку методик учёта неоднородности стад при оценке племенной ценности животных.

Список источников

1. Достоверность данных первичного зоотехнического учёта в молочном скотоводстве / Е.В. Камалдинов, А.Ф. Петров, К.С. Шатохин, К.Н. Нарожных, В.Г. Маренков, Т.А. Жигулин, О.В. Богданова, П.Н. Пальчиков, А.А. Плахова // Вестник НГАУ (Новосибирский государственный аграрный университет). 2022. № 2(63). С. 76-83. [Kamaldinov EV, Petrov AF, Shatohin KS, Narozhnyh KN, Marenikov VG, Zhigulin TA, Bogdanova OV, Palchikov PN, Plakhova AA. Reliability of primary zootechnical records in dairy farming. Bulletin of NSAU (Novosibirsk State Agrarian University). 2022;2(63):76-83. (In Russ.)]. doi: 10.31677/2072-6724-2022-63-2-76-83
2. Маринченко Т.Е. Повышение эффективности молочного скотоводства // Вестник всероссийского научно-исследовательского института механизации животноводства. 2019. № 2(34). С. 193-203. [Marinchenko TE. Improving of the of dairy breedings efficiency. Journal of VNIIMZH. 2019;2(34):193-203. (In Russ.)].
3. Поддубный В.В., Шевелев О.Г., Бормашов Д.А. Сравнение качества подходов к кластеризации текстов на основе гипергеометрического критерия // Вестник Томского государственного университета. 2006. № 293. С. 120-125. [Poddubny VV, Shevelyov OG, Bormashov DA. Comparison of texts clusterization methods quality on the base of hypergeometrical criterion. Tomsk State University Journal. 2006;293:120-125. (In Russ.)].
4. Формирование единой цифровой платформы сельского хозяйства региона / Н.П. Васильев, Л.Д. Протопопова, Г.И. Даянова, А.Н. Крылова, Н.Н. Никитина //

Международный сельскохозяйственный журнал. 2024. № 1(397). С. 53-56. [Vasil'ev NP, Protoporova LD, Dayanova GI, Krylova AN, Nikitina NN. Formation of a unified digital platform for the region's agriculture. *Mezhdunarodnyj sel'skohozyajstvennyj zhurnal*. 2024;1(397):53-56. (*In Russ.*)]. doi: 10.55186/25876740_2024_67_1_53

5. Шевелёва О.М., Свяженина М.А. Влияние быков на продуктивные качества потомства // *Животноводство и кормопроизводство*. 2023. Т. 106. № 4. С. 40-56. [Sheveleva OM, Svyazhenina MA. The influence of bulls on the productive qualities of progeny. *Animal Husbandry and Fodder Production*. 2023;106(4):40-56. (*In Russ.*)]. doi: 10.33284/2658-3135-106-4-40

6. Aggarwal CC. Outlier analysis: advanced concepts. In: *Data Mining*. Springer, Cham; 2015:265-83. doi: 10.1007/978-3-319-14142-8_9

7. Aguilar I, Fernandez EN, Blasco A, Ravagnolo O, Legarra A. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J Anim Breed Genet*. 2020;137(4):356-364. doi: 10.1111/jbg.12470

8. Bivand R. R packages for analyzing spatial data: a comparative case study with areal data. *Geographical Analysis*. 2022;54(3):488-518. doi: 10.1111/gean.12319

9. Brito LF, Oliveira HR, McConn BR, Schinckel AP, Arrazola A, Marchant-Forde JN, et al. Large-scale phenotyping of livestock welfare in commercial production systems: a new frontier in animal breeding. *Front Genet*. 2020;11:793. doi: 10.3389/fgene.2020.00793

10. Cañas-Álvarez JJ, Ossa-Saraz GA, Garcés-Blanquiceth JL, Burgos-Paz WO. Genealogical structure of the Colombian Romosinuano Creole cattle. *Trop Anim Health Prod*. 2023;55(5):292. doi: 10.1007/s11250-023-03694-1

11. Castro-Vásquez R, Vásquez-Loaiza M, Cruz-Méndez A, Domínguez-Viveros J, Camacho-Sandoval J, Saborío-Montero A. Genealogical information analysis of Gyr and Nelore cattle from Costa Rica. *Cienc Rural*. 2023;53(10):e20220236. doi: 10.1590/0103-8478cr20220236

12. Cavani L, Silva RM de O, Carreño LOD, Ono RK, Bertipaglia TS, Farah MM, et al. Genetic diversity of Brazilian Brahman cattle by pedigree analysis. *Pesq Agropec Bras*. 2018;53(1):74-79. doi: 10.1590/s0100-204x2018000100008

13. CDN. Canadian Network for Dairy Excellence. [Internet] Available from: <https://www.cdn.ca/home.php> (accessed 2024 November 27).

14. Chen TY. New Chebyshev distance measures for Pythagorean fuzzy sets with applications to multiple criteria decision analysis using an extended ELECTRE approach. *Expert Systems with Applications*. 2019;147(2):113164. doi: 10.1016/j.eswa.2019.113164

15. Cole JB, Dürr JW, Nicolazzi EL. Invited review: The future of selection decisions and breeding programs: What are we breeding for, and who decides? *Journal of Dairy Science*. 2021;104(5):5111-24. doi: 10.3168/jds.2020-19777

16. Cole JB, Eaglen SAE, Maltecca C, Mulder HA, Pryce JE. The future of phenomics in dairy cattle breeding. *Animal Frontiers*. 2020;10(2):37-44. doi: 10.1093/af/vfaa007

17. Fioretti M, Negrini R, Biffani S, Quaglia A, Valentini A, Nardone A. Demographic structure and population dynamics of Maremmana cattle local breed after 35 years of traditional selection. *Livestock Science*. 2020;232:103903. doi: 10.1016/j.livsci.2019.103903

18. Gaikadi S, Kumar SV. Is ward-level calculation of urban green space availability important?—A case study on Vellore city, India, using the histogram-based spectral discrimination approach. *Front Sustain Cities*. 2024;6:1393156. doi: 10.3389/frsc.2024.1393156

19. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2019;20(3):135-156. doi: 10.1038/s41576-018-0082-2

20. Giorgi FM, Ceraolo C, Mercatelli D. The R language: an engine for bioinformatics and data science. *Life (Basel)*. 2022;12(5):648. doi: 10.3390/life12050648

21. Interbull. International Bull Evaluation Service Official Website. [Internet] Available from: <https://interbull.org/index> (accessed 2024 November 27).

22. Irnawati I, Riswanto FDO, Riyanto S, Martono S, et al. The use of software packages of R factoextra and FactoMineR and their application in principal component analysis for authentication of oils. *Indonesian Journal of Chemometrics and Pharmaceutical Analysis*. 2021;1(1):1-10. doi: 10.22146/ijcpa.482
23. Little RJA. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988;83(404):1198-202.
24. Martin P, Baes C, Houlahan K, Richardson CM, Jamrozik J, Miglior F. Genetic correlations among selected traits in Canadian Holsteins. *Can J Anim Sci*. 2019; 99(4):693-704. doi: 10.1139/CJAS-2018-0190
25. Mrode R, Ojango JMK, Okeyo AM, Mwacharo JM. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. *Front Genet*. 2019;9:694. doi:10.3389/fgene.2018.00694
26. Nader N, El-Gamal FEZ, El-Sappagh S, Kwak KS, Elmogy M. Kinship verification and recognition based on handcrafted and deep learning feature-based techniques. *PeerJ Computer Science*. 2021;7:e735. doi: 10.7717/peerj-cs.735
27. Nilforooshan MA, Saavedra-Jiménez LA. ggroups: an R package for pedigree and genetic groups data. *Hereditas*. 2020;157(1):17. doi: 10.1186/s41065-020-00124-2
28. Nyman S, Johansson AM, Palucci V, Schönherz AA, Guldbrandtsen B, Hinrichs D, et al. Inbreeding and pedigree analysis of the European red dairy cattle. *Genet Sel Evol*. 2022;54(1):70. doi: 10.1186/s12711-022-00761-3
29. Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 2006;123(4):218-23. doi: 10.1111/j.1439-0388.2006.00595.x
30. Select Star SA. [Internet] Available from: <https://www.selectstar.ch/de/index.htm>. (accessed: 2024 November 27).
31. Silva MHMA da, Malhado CHM, Kern EL, Daltro D dos S, Cobuci JA, Carneiro PLS. Inbreeding depression in Holstein cattle in Brazil. *R Bras Zootec*. 2019;48:e20170212. doi: 10.1590/rbz4820170212
32. Suwanda R, Syahputra Z, Zamzami EM. Analysis of euclidean distance and manhattan distance in the k-means algorithm for variations number of Centroid K. *Journal of Physics: Conference Series*. 2020;1566(1):012058. doi: 10.1088/1742-6596/1566/1/012058
33. Toghiani S, VanRaden PM. National Index Correlations and Actual vs. Expected Use of Foreign Sires. *Interbull Bulletin*. Leeuwarden, The Netherlands, 2021, April 26-30. 2021;56:52-9.
34. Wickham H, Grolemund G. *R for Data Science*. Gravenstein Highway North, Sebastopol, CA: O'Reilly; 2017: 520 p.
35. Wiggans GR, Carrillo JA. Genomic selection in United States dairy cattle. *Front Genet*. 2022;9(13):994466.
36. Wiggans GR, Vanraden PM, Cooper TA. The genomic evaluation system in the United States: past, present, future. *J Dairy Sci*. 2011;94(6):3202-11. doi: 10.3168/jds.2010-3866

References

1. Kamaldinov EV, Petrov AF, Shatohin KS, Narozhnyh KN, Marenkov VG, Zhigulin TA, Bogdanova OV, Palchikov PN, Plakhova AA. Reliability of primary zootechnical records in dairy farming. *Bulletin of NSAU (Novosibirsk State Agrarian University)*. 2022;2(63):76-83. doi: 10.31677/2072-6724-2022-63-2-76-83
2. Marinchenko TE. Improving of the of dairy breedings efficiency. *Journal of VNIIMZH*. 2019; 2(34):193-203.
3. Poddubny VV, Shevelyov OG, Bormashov DA. Comparison of texts clusterization methods quality on the base of hypergeometrical criterion. *Tomsk State University Journal*. 2006;293:120-125.

4. Vasilev NP, Protopopova LD, Dayanova GI, Krylova AN, Nikitina NN. Formation of a unified digital platform for the region's agriculture. *International agricultural Journal*. 2024;1(397):53-56. doi: 10.55186/25876740_2024_67_1_53
5. Sheveleva OM, Svyazhenina MA. The influence of bulls on the productive qualities of progeny. *Animal Husbandry and Fodder Production*. 2023;106(4):40-56. doi: 10.33284/2658-3135-106-4-40
6. Aggarwal CC. Outlier analysis: advanced concepts. In: *Data Mining*. Springer, Cham; 2015:265-83. doi: 10.1007/978-3-319-14142-8_9
7. Aguilar I, Fernandez EN, Blasco A, Ravagnolo O, Legarra A. Effects of ignoring inbreeding in model-based accuracy for BLUP and SSGBLUP. *J Anim Breed Genet*. 2020;137(4):356-364. doi: 10.1111/jbg.12470
8. Bivand R. R packages for analyzing spatial data: a comparative case study with areal data. *Geographical Analysis*. 2022;54(3):488-518. doi: 10.1111/gean.12319
9. Brito LF, Oliveira HR, McConn BR, Schinckel AP, Arrazola A, Marchant-Forde JN, et al. Large-scale phenotyping of livestock welfare in commercial production systems: a new frontier in animal breeding. *Front Genet*. 2020;11:793. doi: 10.3389/fgene.2020.00793
10. Cañas-Álvarez JJ, Ossa-Saraz GA, Garcés-Blanquiceth JL, Burgos-Paz WO. Genealogical structure of the Colombian Romosinuano Creole cattle. *Trop Anim Health Prod*. 2023;55(5):292. doi: 10.1007/s11250-023-03694-1
11. Castro-Vásquez R, Vásquez-Loaiza M, Cruz-Méndez A, Domínguez-Viveros J, Camacho-Sandoval J, Saborio-Montero A. Genealogical information analysis of Gyr and Nelore cattle from Costa Rica. *Cienc Rural*. 2023;53(10):e20220236. doi: 10.1590/0103-8478cr20220236
12. Cavani L, Silva RM de O, Carreño LOD, Ono RK, Bertipaglia TS, Farah MM, et al. Genetic diversity of Brazilian Brahman cattle by pedigree analysis. *Pesq Agropec Bras*. 2018;53(1):74-79. doi: 10.1590/s0100-204x2018000100008
13. CDN. Canadian Network for Dairy Excellence. [Internet] Available from: <https://www.cdn.ca/home.php> (accessed 2024 November 27).
14. Chen TY. New Chebyshev distance measures for Pythagorean fuzzy sets with applications to multiple criteria decision analysis using an extended ELECTRE approach. *Expert Systems with Applications*. 2019;147(2):113164. doi: 10.1016/j.eswa.2019.113164
15. Cole JB, Dürr JW, Nicolazzi EL. Invited review: The future of selection decisions and breeding programs: What are we breeding for, and who decides? *Journal of Dairy Science*. 2021;104(5):5111-24. doi: 10.3168/jds.2020-19777
16. Cole JB, Eaglen SAE, Maltecca C, Mulder HA, Pryce JE. The future of phenomics in dairy cattle breeding. *Animal Frontiers*. 2020;10(2):37-44. doi: 10.1093/af/vfaa007
17. Fioretti M, Negrini R, Biffani S, Quaglia A, Valentini A, Nardone A. Demographic structure and population dynamics of Maremmana cattle local breed after 35 years of traditional selection. *Livestock Science*. 2020;232:103903. doi: 10.1016/j.livsci.2019.103903
18. Gaikadi S, Kumar SV. Is ward-level calculation of urban green space availability important?—A case study on Vellore city, India, using the histogram-based spectral discrimination approach. *Front Sustain Cities*. 2024;6:1393156. doi: 10.3389/frsc.2024.1393156
19. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2019;20(3):135-156. doi: 10.1038/s41576-018-0082-2
20. Giorgi FM, Ceraolo C, Mercatelli D. The R language: an engine for bioinformatics and data science. *Life (Basel)*. 2022;12(5):648. doi: 10.3390/life12050648
21. Interbull. International Bull Evaluation Service Official Website. [Internet] Available from: <https://interbull.org/index> (accessed 2024 November 27).
22. Irnawati I, Riswanto FDO, Riyanto S, Martono S, et al. The use of software packages of R factoextra and FactoMineR and their application in principal component analysis for authentication of oils. *Indonesian Journal of Chemometrics and Pharmaceutical Analysis*. 2021;1(1):1-10. doi: 10.22146/ijcpa.482

23. Little RJA. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*. 1988;83(404):1198-202.
24. Martin P, Baes C, Houlihan K, Richardson CM, Jamrozik J, Miglior F. Genetic correlations among selected traits in Canadian Holsteins. *Can J Anim Sci*. 2019; 99(4):693-704. doi: 10.1139/CJAS-2018-0190
25. Mrode R, Ojango JMK, Okeyo AM, Mwacharo JM. Genomic selection and use of molecular tools in breeding programs for indigenous and crossbred cattle in developing countries: current status and future prospects. *Front Genet*. 2019;9:694. doi:10.3389/fgene.2018.00694
26. Nader N, El-Gamal FEZ, El-Sappagh S, Kwak KS, Elmogy M. Kinship verification and recognition based on handcrafted and deep learning feature-based techniques. *PeerJ Computer Science*. 2021;7:e735. doi: 10.7717/peerj-cs.735
27. Nilforooshan MA, Saavedra-Jiménez LA. ggroups: an R package for pedigree and genetic groups data. *Hereditas*. 2020;157(1):17. doi: 10.1186/s41065-020-00124-2
28. Nyman S, Johansson AM, Palucci V, Schönherz AA, Guldbrandtsen B, Hinrichs D, et al. Inbreeding and pedigree analysis of the European red dairy cattle. *Genet Sel Evol*. 2022;54(1):70. doi: 10.1186/s12711-022-00761-3
29. Schaeffer L. Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*. 2006;123(4):218-23. doi: 10.1111/j.1439-0388.2006.00595.x
30. Select Star SA. [Internet] Available from: <https://www.selectstar.ch/de/index.htm>. (accessed: 2024 November 27).
31. Silva MHMA da, Malhado CHM, Kern EL, Daltro D dos S, Cobuci JA, Carneiro PLS. Inbreeding depression in Holstein cattle in Brazil. *R Bras Zootec*. 2019;48:e20170212. doi: 10.1590/rbz4820170212
32. Suwanda R, Syahputra Z, Zamzami EM. Analysis of euclidean distance and manhattan distance in the k-means algorithm for variations number of Centroid K. *Journal of Physics: Conference Series*. 2020;1566(1):012058. doi: 10.1088/1742-6596/1566/1/012058
33. Toghiani S, VanRaden PM. National Index Correlations and Actual vs. Expected Use of Foreign Sires. *Interbull Bulletin*. Leeuwarden, The Netherlands, 2021, April 26-30. 2021;56:52-9.
34. Wickham H, Grolemund G. *R for Data Science*. Gravenstein Highway North, Sebastopol, CA: O'Reilly; 2017: 520 p.
35. Wiggans GR, Carrillo JA. Genomic selection in United States dairy cattle. *Front Genet*. 2022;9(13):994466.
36. Wiggans GR, Vanraden PM, Cooper TA. The genomic evaluation system in the United States: past, present, future. *J Dairy Sci*. 2011;94(6):3202-11. doi: 10.3168/jds.2010-3866

Информация об авторах:

Евгений Варисович Камалдинов, доктор биологических наук, доцент, заведующий кафедрой прикладной биоинформатики, Новосибирский государственный аграрный университет, 630039, г. Новосибирск, ул. Добролюбова, 160.

Алексей Фёдорович Петров, заведующий лабораторией прикладной биоинформатики, Новосибирский государственный аграрный университет, 630039, г. Новосибирск, ул. Добролюбова, 160.

Кирилл Николаевич Нарожных, кандидат биологических наук, доцент кафедры прикладной биоинформатики, Новосибирский государственный аграрный университет, 630039, г. Новосибирск, ул. Добролюбова, 160.

Павел Николаевич Пальчиков, директор, АО «Новосибирскагроплем», 630522, НСО, Новосибирский район, МО Яркоковский сельсовет по Ордынской трассе на территории р-на за 1,5 км до с. Ярково.

Information about the authors:

Evgeny V Kamaldinov, Dr Sci. (Biology), Associate Professor, Head of the Department of Applied Bioinformatics, Novosibirsk State Agrarian University, 160 Dobrolyubova str., 630039, Novosibirsk.

Alexey F Petrov, Head of the Laboratory of Applied Bioinformatics, Novosibirsk State Agrarian University, 160 Dobrolyubova str., 630039, Novosibirsk.

Kirill N Narozhnykh, Cand. Sci. (Biology), Associate Professor of the Department of Applied Bioinformatics, Novosibirsk State Agrarian University, 160 Dobrolyubova str., 630039, Novosibirsk.

Pavel N Palchikov, Director, JSC "Novosibirsk Agroplem", 630522, NSO, Novosibirsk district, MO Yarkovsky village Council along the Ordynsky highway in the territory of the district 1.5 km from the village of Yarkovo.

Статья поступила в редакцию 18.10.2024; одобрена после рецензирования 05.12.2024; принята к публикации 16.12.2024.

The article was submitted 18.10.2024; approved after reviewing 05.12.2024; accepted for publication 16.12.2024.