

Анализ показателей разнообразия STR-локусов в выборках производителей красной скандинавской и голштинской пород

© 2024. В. М. Кузнецов✉

ФГБНУ «Федеральный аграрный научный центр Северо-Востока имени Н. В. Рудницкого», г. Киров, Российская Федерация

Методами одно-двухмерной и многомерной статистики анализ и ровалистандтизированные показатели α - и β -разнообразия (y'_{km}), рассчитанные по микросателлитам ДНК (STR) в выборках быков красной скандинавской (RED, $n = 29$) и голштинской (HOL, $n = 45$) пород (в каждой по три субвыборки). Данные представляли две матрицы 11×7 : объекты – 11 STR-локусов (Eth3, Inra23, Tgla227, Tgla126, Sps115, Eth225, Tgla53, Bm2113, Bm1824, Eth10; №№1-11), переменные – три показателя α -разнообразия (число аллелей и эффективных аллелей на локус, гетерозиготность) и четыре – β -разнообразия (индексы: фиксации по Nei, дифференциации по Meirmans-Hedrick, Jost и Shannon-Sherwin). ANOVA по модели фиксированного типа выявил статистически значимое ($p_{value} < 0,02$) влияние на изменчивость y'_{km} факторов «порода» (2 %), «локус» (36,7 %) и их взаимодействие (15,6 %). По модели смешанного типа (фактор «локус» как случайный) статистически значимым установлен только эффект взаимодействия (25,8 %, $p_{value} < 0,0001$). Вероятность несовпадения номеров наугад выбранной пары локусов из RED- и HOL-выборок составила 31 %. Усреднённая евклидова дистанция между двумя выборками, рассчитанная по одномерным локусам, составила $37,8 \pm 5,35$ %. Корреляция Мантеля между матрицами парных межлокусных дистанций в RED- и HOL-выборках была $0,257 \pm 0,130$ ($p_{value} = 0,056$). Ординации локусов и их группировка (структурирование) в пространстве двух главных компонент выборки REL и выборки HOL отличались (прокрустов тест: $m^2 = 0,994$, $m_{12} = 0,747$, $r_{perm} = 0,164$, $r^2_{Proc} = 0,253$). Оценки расстояния между выборками по профилям показателей α - и β -разнообразия локусов не противоречили, в общем, генетическим дистанциям, рассчитанным по аллельным частотам (29–37 %). Для анализа ковариации (общности) многомерных данных RED и HOL выборок использовали двухблочный метод частных наименьших квадратов (2B-PLS). Интегрированные латентные переменные (LV) максимизировали общий квадрат ковариации («*squaredcovar*») = 14,3 %, в котором 83 % приходилось на первую LV с тах «весаами» по показателям α -разнообразия (αLV). На вторую LV приходилось 16,7 % с тах «весаами» по показателям β -разнообразия (βLV). Линейная связь между RED и HOL выборками по αLV составила 0,717 ($p_{value} = 0,013$), по βLV – 0,395 ($p_{value} = 0,229$), усреднённая – 0,56 ($p_{value} = 0,025$). Общность (ко-дисперсия) двух выборок по αLV и βLV оценивалась в 25,0–32,5 %. 2B-PLS-анализ по редуцированным данным (только по α -разнообразию) показал тах «*squaredcovar*» 0,393, в которой 99,9 % приходилось на первую LV (LV1). По LV1 линейная связь между RED- и HOL-выборками оценивалась в 0,659 ($p_{value} = 0,0253$), ко-дисперсия – в 43,4 % (по αLV была 51,4 %). Ординации локусов в пространстве координат выборок RED и HOL по полному (αLV) и редуцированному (LV1) набору данных имели хорошее соответствие (прокрустов тест: $m^2 = 0,0742$, $m_{12} = 0,0728$, $r_{perm} = 0,001$, $r^2_{Proc} = 0,927$). В структуре межвыборочной ковариации выделялись «сгустки» локусов с бутстрэп-вероятностью [группировки] 50, 75 и 100 %. Можно полагать, что RED и HOL выборки имели некоторую согласованность (конгруэнтность) по показателям α -разнообразия одноимённых локусов. Распространение «многомерного» подхода на описательную статистику α -разнообразия 7 пород молочного скота и 11 пород свиней показало достаточно хорошее соответствие результатов (индекс дифференциации, PCA-ординация) с таковыми, полученными при использовании «традиционных» методов (r_{perm} соответствия ординаций 0,054 и 0,004). Рассмотренные подходы и методы расширяют возможности популяционно-генетических [и селекционно-зоотехнических] исследований, в которых многомерные наборы данных являются нормой, а не исключением.

Ключевые слова: быки-производители, микросателлиты, локусы, α - и β -разнообразие, дисперсионный анализ, дистанции, анализ главных компонент, ординация, двухблочные частные наименьшие квадраты, прокрустов тест

Благодарности: работа выполнена при поддержке Минобрнауки РФ в рамках Государственного задания ФГБНУ «Федеральный аграрный научный центр Северо-Востока имени Н. В. Рудницкого» (№ гос. регистрации 1021060407726-4).

Автор благодарит рецензентов за их вклад в экспертную оценку этой работы.

Конфликт интересов: автор заявил об отсутствии конфликта интересов.

Для цитирования: Кузнецов В. М. Анализ показателей разнообразия STR-локусов в выборках производителей красной и голштинской пород. Аграрная наука Евро-Северо-Востока. 2024;25(3):465–482.

DOI: <https://doi.org/10.30766/2072-9081.2024.25.3.465-482>

Поступила: 20.02.2024

Принята к публикации: 29.05.2024

Опубликована онлайн: 26.06.2024

Analysis of the diversity of STR-loci in the samples of bulls of Red Scandinavian and Holstein breeds

© 2024. Vasiliy M. Kuznetsov ✉

Federal Agricultural Research Center of the North-East named N. V. Rudnitsky,
Kirov, Russian Federation

Standardized indicators of α - and β -diversity (y'_{km}) calculated by DNA microsatellites (STR) in samples of bulls of Red Scandinavian (RED, $n = 29$) and Holstein (HOL, $n = 45$) breeds (each with three sub-samples) were analyzed using one-two- and multivariate statistics methods. The data represented two 11×7 matrices: objects – 11 STR loci (Eth3, Inra23, Tgla227, Tgla126, Tgla122, Sps115, Eth225, Tgla53, Bm2113, Bm1824, Eth10; No.1-11), variables – three indicators of α -diversity (number of alleles and effective alleles per locus, heterozygosity) and four – β -diversity (indices: fixation by Nei, differentiation by Meirmans-Hedrick, Jost and Shannon-Sherwin). ANOVA, using a fixed-type model, revealed a statistically significant ($p_{value} < 0.02$) effect on the variability of y'_{km} factors «breed» (2 %), «locus» (36.7 %) and their interaction (15.6 %). According to the mixed-type model (the «locus» factor as random), only the interaction effect was statistically significant (25.8 %, $p_{value} < 0.0001$). The probability of a noncoincidence between the numbers of a randomly selected pair of loci from the RED and HOL samples was 31 %. The average Euclidean distance between the two samples, calculated by analogical loci, was 37.8 ± 5.35 %. The Mantel correlation between the matrices of paired interlocus distances in RED and in HOL samples was 0.257 ± 0.130 ($p_{value} = 0.056$). The ordinations of loci and their grouping (structuring) in the space of the two main components of the REL sample and the HOL sample differed (Procrust test: $m^2 = 0.994$, $m_{12} = 0.747$, $p_{perm} = 0.164$, $r^2_{Proc} = 0.253$). Estimates of the distance between samples based on the profiles of the α - and β -diversity of loci did not contradict, in general, the genetic distances calculated by allelic frequencies (29–37 %). To analyze the covariance (commonality) of multivariate RED and HOL sample data, a two-block partial least squares (2B-PLS) method was used. The integrated latent variables (LV) maximized the total square of covariance («squared covar» = 14.3 %), in which 83 % accounted for the first LV with max «weights» in terms of α -diversity (aLV). The second LV accounted for 16.7 % with max «weights» in terms of β -diversity (β LV). The linear relationship between RED and HOL samples for aLV was 0.717 ($p_{value} = 0.013$), for β LV – 0.395 ($p_{value} = 0.229$), averaged – 0.56 ($p_{value} = 0.025$). The commonality (co-dispersion) of the two samples for aLV and β LV was estimated at 25.0–32.5 %. 2B-PLS analysis based on reduced data (only for α -diversity) showed a max «squared covar» of 0.393, in which 99.9 % accounted for the first LV (LV1). According to LV1, the linear relationship between RED and HOL samples was estimated at 0.659 ($p_{value} = 0.0253$), the co-dispersion was 43.4 % (according to aLV it was 51.4 %). The ordinations of loci in the coordinate space of the RED and HOL samples for the complete (aLV) and reduced (LV1) datasets had a good match (Procrust test: $m^2 = 0.0742$, $m_{12} = 0.0728$, $p_{perm} = 0.001$, $r^2_{Proc} = 0.927$). In the structure of the inter-sample covariance, «clumps» of loci with a bootstrap probability of [grouping] 50, 75 and 100 % were distinguished. It can be assumed that the RED and HOL samples had some consistency (congruence) in terms of the α -diversity of the loci of the same name. The extension of the «multivariate» approach to descriptive statistics of α -diversity of 7 breeds of dairy cattle and 11 breeds of pigs showed a fairly good correspondence of the results (differentiation index, PCA-ordination) with those obtained using «traditional» methods (p_{perm} of matching ordinations 0.054 and 0.004). The approaches and methods considered expand the possibilities of population-genetic [and breeding-zootecnical] studies in which multidimensional data sets are the norm, not the exception.

Keywords: bulls, microsatellites, loci, α - and β -diversity, variance analysis, distances, principal component analysis, ordination, two-block partial least squares, Procrust test

Acknowledgements: the research was carried out under the support of the Ministry of Science and Higher Education of the Russian Federation within the state assignment of the Federal Agricultural Research Center of the North-East named N. V. Rudnitsky (registrations No. 1021060407726-4).

The author thanks the reviewers for their contribution to the peer review of this work.

Conflict of Interest: the author declared no conflicts of interest.

For citations: Kuznetsov V. M. Analysis of the diversity of STR-loci in the samples of bulls of Red and Holstein breeds. *Agrarnaya nauka Evro-Severo-Vostoka* = Agricultural Science Euro-North-East. 2024;25(3):465–482. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2024.25.3.465-482>

Received: 20.02.2024

Accepted for publication: 29.05.2024

Published online: 26.06.2024

При изучении внутри- и межпородной генетической изменчивости по микросателлитам ДНК (Simple Tandem Repeats – простые тандемные последовательности; в дальнейшем STR) наборами данных являются выборки животных, генотипированных по 10–30 и более локусам. Для их анализа используют разные количественные меры популяционной генетики, которые характеризуют те или иные аспекты генетического разнообразия. Показатели разли-

чия/сходства выборок рассчитывают по каждому локусу. Затем объединяют по всем локусам – финальные оценки, на основе которых формируют выводы и рекомендации, например, по сохранению исчезающих пород. В публикациях оценок показателей разнообразия по каждому локусу (полокусных оценок), как правило, нет. Вместе с тем в совокупности они составляют многомерные данные, анализ которых методами традиционной и многомерной статистики может

дать дополнительную полезную информацию о локусах, именно: их различие и сходство, расщепление и структурирование в выборке (-ах).

В нашей предыдущей работе [1] генотипы по 11 локусам STR-маркеров ДНК 84 быков были использованы для полокусной оценки 14 показателей внутри- и межпородного разнообразия (α - и β -разнообразия). Была сформирована матрица данных 11×14 типа «объект-переменная», где «объект» – это локусы, а «переменная» – показатели разнообразия. К этому набору у многомерных данных был применён метод главных компонент. В результате анализа была получена локусная 2D ординация, в которой локусы *Tgla227* и *Tgla53* формировали группу А с бутстрэп-вероятностью 73 %; локусы *Tgla122*, *Eth225* и *Eth10* – группу В с вероятностью 100 %; локусы *Inra23*, *Bm2113* и *Bm1824* – группу С с вероятностью 73 %. Валидация ординации локусов проводилась по редуцированной матрице (11×7) методом неметрического многомерного шкалирования. Согласованность ординаций по прокрустову тесту составила 96 % ($p_{perm} = 0,001^1$), что позволило говорить о неслучайной координатной структуре локусов, характерной для данной выборки.

Цель настоящей работы – рассмотреть возможность использования полокусных показателей STR-разнообразия для сравнительного анализа пород. В частности, интерес представляли исследования: (а) влияния на изменчивость полокусных оценок факторов «порода», «локус» и взаимодействия «порода×локус»; (б) уровня различия/близости между породами по профилям разнообразия локусов; (в) соответствия/отличия ординаций (структурной организации) локусов разных пород; (г) наличия/отсутствия общности (ко-дисперсии) между многомерными наборами полокусных оценок.

Научная новизна – для решения поставленных вопросов в качестве исходных данных использовались оценки показателей аллельного

разнообразия по локусам (породам), которые анализировались методами традиционной и многомерной статистики. Подобных популяционно-генетических исследований в животноводстве ранее не проводилось.

Материал и методы. Использовали результаты генотипирования быков-производителей по 11 STR-локусам ДНК, представленные в открытом доступе на сайте ВНИИплем². Были отобраны данные только по быкам красной [скандинавской] (VikingRed; в дальнейшем – **RED**; $n = 29$ быков, представляющих три субвыборки: по 10 быков айрширской и красной датской пород, 9 – красной шведской) и голштинской (в дальнейшем – **HOL**; $n = 45$ быков, представляющих субвыборки: 13 быков немецкой, 17 – нидерландской и 15 быков американской селекции) породам³.

В обеих выборках по каждому локусу были рассчитаны семь показателей: n_a – фактическое число аллелей на локус; s_{n_e} – число эффективных аллелей на локус по энтропии [2]; H_e – гетерозиготность при равновесии Харди-Вайнберга [3]; $G'_{ST(N)}$ – индекс фиксации по Нею [4] и три индекса дифференциации [субвыборок]: G''_{ST} – по Вейру и Кокерхэму [5] с модификацией по [6]; D_{est} – по Джосту [7] и D' – по Шеннону/Шервину [8, 9]⁴.

Для анализа полученных оценок (переменных) использовали методы традиционной (описательная статистика, корреляционный и дисперсионный анализы) и многомерной (анализ главных компонент, двухблочный метод частных наименьших квадратов, корреляция Мантеля, прокрустовый анализ) статистики. Расчёты проводили с помощью компьютерных программ GenAlEx 6.502 [15, 16, 17], PAST3 [18], KyPlot 6.0 [19], PROTEST [20, 21], STATGRAPHICS®Centurion XVI⁵.

Результаты и их обсуждение. Преобразование данных. Сводные оценки показателей разнообразия/дифференциации по породам (Breed) были:

Breed	n_a	s_{n_e}	H_e	$G'_{ST(N)}$	G''_{ST}	D_{est}	D'
RED	6,7	4,25	0,683	0,005	0,016	0,011	0,184
HOL	7,0	4,54	0,724	0,015	0,058	0,039	0,142

¹ $p_{value}(p_{perm})$ – достигнутый уровень статистической значимости; в скобках – рассчитанный численным пермутационным методом с 999 рандомизированными перестановками данных.

²База генетических данных быков-производителей. Микросателлиты. [Электронный ресурс].

URL: <http://www.vniiplem.ru/rus/files/Database/DNK/mikrosatellity.pdf> (дата обращения: 09.12.2018).

³Из 22 тестов согласия распределений генотипов с равновесием Харди-Вайнберга (PXB) в одном случае имело место статистически значимое отклонение от PXB – по локусу *Eth225* в выборке голштинских быков из Германии.

⁴Методы расчёта были рассмотрены в [10, 11, 12, 13, 14].

⁵STATGRAPHICS® Centurion XVI User Manual. By StatPoint Technologies, Inc. 2010. 297 p.

Первые три переменные характеризовали внутрипородное α -разнообразие, остальные – β -разнообразие. Переменная $G'_{ST(N)}$ указывала на вероятность фиксации аллелей в локусе, переменные G''_{ST} , D_{est} и D' – на уровни дифференциации субвыборок в пределах породных выборок.

Характеризуя те или иные аспекты генетического разнообразия, показатели имели разные шкалы и единицы измерения. Это делало невозможным их статистическую обработку. Поэтому все оценки по локусам-переменным (y_{km}) были преобразованы в безразмерные величины (y'_{km})⁶ с параметрами распределения: среднее = 100, стандартное отклонение = 10 (табл. 1).

Таблица 1 – Преобразованные оценки показателей разнообразия локусов /
Table 1 – Transformed estimates of loci diversity indicators

Breed	Locus	Nº	n_a	s_{n_e}	H_e	$G'_{ST(N)}$	G''_{ST}	D_{est}	D'
RED	Eth3	1	95,5	97,2	99,0	112,9	111,2	111,1	107,1
	Inra23	2	90,4	91,1	86,7	111,4	104,5	103,1	113,9
	Tgla227	3	110,3	112,7	111,4	93,6	92,0	91,8	107,7
	Tgla126	4	95,5	92,5	95,2	80,7	82,2	83,8	85,2
	Tgla122	5	97,7	95,2	97,1	80,7	81,6	83,5	93,2
	Sps115	6	88,1	87,1	79,0	88,6	90,9	92,7	97,1
	Eth225	7	105,1	107,3	107,6	106,4	110,4	111,3	109,4
	Tgla53	8	107,3	105,3	103,8	96,4	96,5	96,7	110,3
	Bm2113	9	100,0	103,3	105,7	98,6	99,2	99,8	104,1
	Bm1824	10	92,6	89,1	85,7	106,4	101,6	100,4	94,4
	Eth10	11	100,0	106,7	108,6	94,3	94,7	94,2	112,7
HOL	Eth3	1	92,6	90,4	94,3	98,6	97,4	97,5	82,4
	Inra23	2	95,5	97,9	102,9	115,7	115,5	117,3	97,7
	Tgla227	3	117,7	115,4	110,5	107,1	114,3	115,8	104,5
	Tgla126	4	85,2	85,7	84,8	95,7	96,2	95,3	83,3
	Tgla122	5	112,5	110,7	111,4	97,9	100,0	99,8	95,8
	Sps115	6	92,6	91,1	91,4	105,0	103,0	100,9	91,2
	Eth225	7	97,7	95,2	97,1	113,6	110,5	109,6	105,2
	Tgla53	8	125,1	122,2	113,3	100,7	107,2	106,7	113,0
	Bm2113	9	102,9	102,6	108,6	84,3	82,6	79,8	87,4
	Bm1824	10	90,4	93,8	97,1	102,1	102,4	100,5	96,2
	Eth10	11	105,1	107,3	108,6	102,9	106,5	106,4	106,7

Данные таблицы 1 представляли не только два набора чисел по породам, к которым применимы методы традиционной (одно-двухмерной) статистики, но и две матрицы многомерных данных типа «объект-переменная» размерностью 11×7. Во втором случае объектами являлись 11 локусов, а переменными/признаками – 7 мер, оценивающих разные грани одного исходного «конструкта» – STR-разнообразия.

Каждый вектор-столбец – это профиль m -го показателя разнообразия по 11 локусам;

каждый вектор-строка – профиль k -го локуса по семи преобразованным оценкам разнообразия. Последние рассматривали в качестве набора «производственных» показателей, характеризующих локус как оценщика STR-разнообразия. Объекты-локусы в семимерном признаковом пространстве без методов многомерной статистики невозможно геометрически (графически) отобразить и выявить наличие каких-либо отклонений от случайного рассеивания (закономерности).

⁶ $y'_{km} = 100 + [(y_{km} - \mu_m) / \sigma_m] \times 10$, где y_{km} – оригинальная оценка m -го показателя по k -ому локусу; y'_{km} – преобразованная оценка; μ_m и σ_m – среднее и стандартное отклонение по m -му показателю. Преобразование обеспечило соизмеримость переменных с различными средними и дисперсиями в рамках их совместной обработки без какой-либо потери информации.

Тесты на нормальность. Данные таблицы 1 были проверены на нормальность распределения. Одномерный тест Жарке-Бера⁷ показал величины достигнутой статистической значимости (p_{value}) по локусам для RED-выборки в диапазоне от 0,588 до 0,877, для HOL-выборки – 0,535–0,806. Проверка на нормальность распределения матриц RED и HOL (каждая с размерностью 11×7) комплексным критерием Доорника-Хансена (Doornik-Hansen) показала величины p_{value} 0,8 и 0,483 соответственно. Все значения p_{value} были больше принятого критического уровня статистической значимости ($\alpha = 0,05$). Следовательно, можно полагать, что данные получены из нормально распределенных совокупностей. Для проверки эквивалентности ковариационных матриц двух многомерных выборок был

использован М-критерий Бокса (Box's M Test). Он составил 58,6 при $p_{value} = 0,118 > \alpha = 0,05$, что указывало на эквивалентность ковариационных матриц. Таким образом, анализируемые наборы данных соответствовали совокупности допущений, позволяющих применять параметрические методы статистики.

Описательная статистика. На рисунке 1 визуализирована описательная статистика по RED- и HOL-выборкам. Различие между средними выборок было небольшим (рис. А). В пределах выборок имела место вариация комплексных средних по локусам (рис. В). При этом средние в HOL-выборке были в 7 случаях из 11 выше. Вместе с тем, относительная изменчивость в обеих выборках отмечена на уровне 7,5 %.

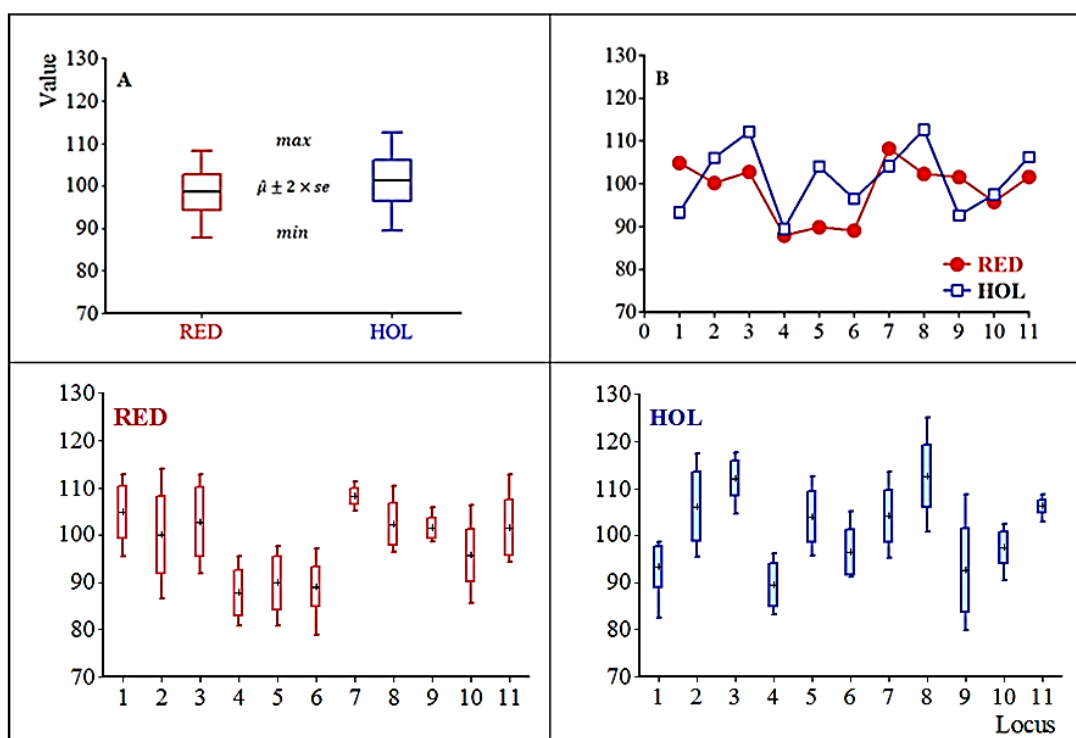


Рис. 1. Графики средних по выборкам (А), локусам (В) и изменчивости полокусных оценок разнообразия в выборках (RED, HOL) /

Fig. 1. Graphs of averages by samples (A), loci (B) and variability of estimates of diversity by loci in samples (RED, HOL)

Диаграммы размаха (рис. 1, RED и HOL) демонстрировали примерно равный разброс преобразованных оценок в пределах локусов, именно: по RED-выборке в диапазоне от 6,2 до 27,2 ед., по HOL-выборке – от 5,7 до 28,8 ед. В то же время имело место наличие существенных различий изменчивости оценок одно-

порядковых локусов. Так, в RED-выборке по локусу 1 размах составил 20,9, в HOL-выборке – 13,2 ед. По локусу 7 соответственно 6,2 и 18,4 (различие в 3 раза), по локусу 9 – 7,1 и 28,8 (в 4 раза), по локусу 11 – 18,5 и 5,7 ед. (в 3,2 раза).

⁷Тест Jarque-Bera (Жарке-Бера, Харке-Бера) – статистический тест, проверяющий ошибки наблюдений на нормальность посредством сверки их третьего момента (асимметрия) и четвёртого момента (эксцесс) с моментами нормального распределения. Если p_{value} больше принятого критического α -уровня, то нулевая гипотеза (H_0) о том, что данные получены из нормально распределённой совокупности, не может быть отвергнута.

Однофакторный дисперсионный анализ (Analysis Of Variance, ANOVA). Цель ANOVA – проверка статистической значимости различий между средними градаций/групп фактора (-ов). ANOVA осуществляет декомпозицию общей суммы квадратов отклонений данных от среднего (Total Sum of Squares, SS_T). Одна часть представляет внутригрупповую изменчивость, которую интерпретируют как «остаток» (Residual) или «случайная ошибка» (SS_e). Вторая – вариацию средних по группам (например, локусам – SS_L). Их деление на соответствующие числа степеней свободы (Df) даёт средние квадраты (Mean Square, MS), которые используют для проверки H_0 -гипотезы по критерию Фишера ($F = MS_L/MS_e$). Если градации изучаемого фактора были *случайно* отобраны,

то рассчитывают varianсы ($\hat{\sigma}_L^2$, $\hat{\sigma}_e^2$) и коэффициент внутриклассовой корреляции (Intraclass Correlation Coefficient, $ICC = \hat{\sigma}_L^2/(\hat{\sigma}_L^2 + \hat{\sigma}_e^2)$), по которому определяют внутригрупповую надёжность переменных и связанность их средних уровней.

В таблице 2 представлены результаты ANOVA по однофакторной статистической модели (фактор «Locus» рассматривался как рандомизированный – локусы отобраны случайно). Тест Фишера показал статистически значимое влияние фактора «Locus» на изменчивость y'_{km} -оценок в обеих выборках ($p_{value} < 0,0001$). Оценки остаточных varianс ($\hat{\sigma}_e^2$) были практически равными, а межлокусная varianса ($\hat{\sigma}_L^2$) в HOL-выборке была на 37,7 % больше, чем в RED-выборке.

Таблица 2 – Однофакторная ANOVA по RED- и HOL-выборкам /

Table 2 – Single-factor ANOVA for RED and HOL samples

RED / Source:	SS	Df	MS	F	p _{value}	$\hat{\sigma}^2$	ICC	95% CI
Locus (L)	3314,33	10	331,43	6,55	<0,0001	40,12	*0,44 **0,85	[0,21, 0,74] [0,65, 0,95]
Residual (e)	3338,38	66	50,58	-	-	50,58		
Total (T)	6652,71	76	-	-	-	-		
HOL / Source:	SS	Df	MS	F	p _{value}	$\hat{\sigma}^2$	ICC	95% CI
Locus (L)	4382,54	10	438,25	8,50	<0,0001	55,24	*0,52 **0,88	[0,28, 0,79] [0,73, 0,96]
Residual (e)	3404,28	66	51,58	-	-	51,58		
Total (T)	7786,82	76	-	-	-	-		

* Надёжность оценки локуса по одному показателю (ICC_{Ind}); ** надёжность усреднённой оценки по семи показателям (ICC_M). Пояснения в тексте /

* Reliability of locus estimation by one indicator (ICC_{Ind}); ** reliability for an average estimate by seven indicators (ICC_M). Explanations in the text.

Индивидуальный коэффициент внутриклассовой корреляции, ICC_{Ind} (в табл. 2 помечены «*»), по RED-выборке составил 0,44, по HOL-выборке – 0,52. «Фундаментальная» интерпретация ICC_{Ind} – это мера доли общей изменчивости, которая относится к объектам оценивания [22], в данном случае – к локусам. 95-процентные доверительные интервалы оценок (95 % CI) перекрывались, что указывало на статистически незначимое различие ICC_{Ind} по выборкам. «Истинные» значения ICC_{Ind} с доверительной вероятностью 95 % находились в диапазоне 0,21...0,79⁸.

Межлокусная varianса ($\hat{\sigma}_L^2$) выражала ту изменчивость, которая являлась общей для индивидуальных оценок одного и того же локуса. Поэтому ICC_{Ind} можно также интерпретировать как меру абсолютного согласия (absolute agreement) или гомогенности (homogeneity) оценок показателей разнообразия по локусам, как надёжность (reliability) индивидуальной (y'_{km}) оценки⁹. По [23] величина $ICC < 0,5$ считается низкой, 0,5–0,75 – средней, 0,75–0,90 – хорошей, >0,90 – отличной. Следовательно, при 95% CI [0,21, 0,79] категория надёжности индивидуальных оценок может быть от «низкой» до «хорошей».

⁸Отметим, что в ANOVA σ_e^2 предполагает остаточную изменчивость, связанную с неизвестными и случайными причинами (ошибку). В действительности σ_e^2 была следствием, главным образом, вариации y'_{km} из-за использования семи мер разнообразия. Поэтому величину $(1-ICC_i)$ можно рассматривать как долю общей дисперсии, обусловленную фактором «метод».

⁹Альтернативой ICC_{Ind} является парная корреляция Пирсона. Усреднённые коэффициенты корреляции y'_{km} -оценок между методами были: по RED-выборке – 0,45, по HOL-выборке – 0,52.

Надёжность средних по семи y'_{km} -оценкам локуса, ICC_M (в табл. 2 помечены «**»), составила по RED-выборке 0,85, по HOL-выборке – 0,88. Различие в ICC_M было статистически незначимо. «Истинная» величина ICC_M с доверительной вероятностью 95 % находилась в диапазоне [0,65, 0,96], т. е. ICC_M может иметь категорию надёжности «средняя» и выше.

Интерес представляли величины ICC_M конкретных локусов. Приблизительные оценки были получены, заменив SS_e на $SS_k \times 11/7$, где SS_k – сумма квадратов по k -ому локусу. Оценки варьировали от 0,60 до 0,95 (рис. 2). По голштинской породе ICC_M были выше, за исключением локусов 3, 8 и 9. Высокие ICC_M локусов в одной выборке не были таковыми в другой.

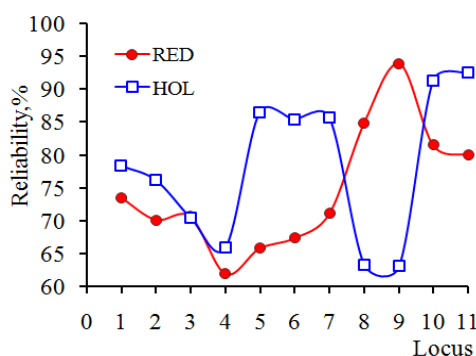


Рис. 2. Вероятный ICC_M по локусам /
Fig. 2. Probable ICC_M by loci

Двухфакторный ANOVA по фиксированной модели. Данные по выборкам были объединены и обработаны двухфакторной моделью ANOVA: $y = \mu + BR + L + I + e$, где $y = y'_{km}$; μ – общее среднее; BR – эффект породы; L – эффект локуса; I – эффект взаимодействия ($Interaction = BR \times L$); e – Residual. Все эффекты, кроме Residual, были фиксированные¹⁰. Результаты представлены в таблице 3.

Влияние каждого фактора было статистически значимым ($p_{value} < 0,02$). Долю (вклад) каждого фактора, $\eta_i^2 = SS_i/SS_T$, также называют «частным коэффициентом детерминации». BR -фактор объяснял 2,0 % общей суммы квадратов, L -фактор – 36,7 %, взаимодействие $BR \times L$ – 15,6 %¹¹. Так как суммы квадратов относились к III типу, то вклад каждого фактора не включал влияние других факторов.

Корень квадратный из $(\eta^2/100)$ даёт оценку силы [нелинейной] связи между фактором и зависимой переменной. Для определения тесноты связи использовали шкалу Чеддока:

Значение	Связь	Значение	Связь
$\eta = 0$	Отсутствует	$0,5 \leq \eta < 0,7$	Заметная
$0 < \eta < 0,2$	Очень слабая	$0,7 \leq \eta < 0,9$	Сильная
$0,2 \leq \eta < 0,3$	Слабая	$0,9 \leq \eta < 1$	Весьма сильная
$0,3 \leq \eta < 0,5$	Умеренная	$\eta = 1$	Функциональная

Таблица 3 – Двухфакторный ANOVA по фиксированной модели (RED + HOL) /
Table 3 – Two-factor fixed-model ANOVA (RED + HOL)

Source (Fix)	SS	η^2 , %	Df	MS	F	p-value
Breed (BR)	296,54	2,0	1	296,54	5,81	0,0174
Locus (L)	5402,15	36,7	10	540,22	10,58	<0,0001
Interaction (BR×L)	2294,71	15,6	10	229,47	4,49	<0,0001
Residual (e)	6742,66	45,7	132	51,08	-	-
Total (T)	14736,06	100,0	153	-	-	-

Между BR -фактором и полокусными оценками $\eta_{BR} = 0,14$, что характеризовало соотношение как «очень слабое». Оценки для η_L и $\eta_{BR \times L}$ были соответственно 0,6 и 0,4, что определяло взаимосвязи как «умеренно-заметные».

Показателем адекватности и надёжности статистической модели ANOVA является коэф-

фициент детерминации ($R^2 = 1 - SS_e/SS_T$, т. е. суммарная доля всех факторов в SS_T). Он составил 54,3 %, а величина $\sqrt{0,543} = 0,74$ представляла собой множественный коэффициент корреляции между полокусными оценками и всеми факторами модели. По шкале Чеддока теснота связи характеризовалась как «сильная».

¹⁰Статистические модели фиксированного типа используют для оценки эффектов не случайно выбранных градаций факторов; выводы распространяются только на анализируемые данные. Случайные (рандомизированные) статистические модели используют для оценки варiances (σ^2); выводы распространяются на популяцию. Смешанные модели используют для одновременной оценки эффектов фиксированных факторов и варiances компонент случайных факторов.

¹¹Взаимодействие – влияние изменения значений одного фактора на величину эффекта, обусловленного воздействием другого фактора (эффект возникает, когда зависимость между фактором В и признаком изменяется под воздействием фактора А).

Межпородные различия по локусам. Влияние породы производителей на вариацию полокусных оценок было статистически значимым ($p_{value} < 0,02$). Однако это не означало, что каждая средняя по локусу в выборке RED значимо отличалась от средней аналогичного локуса в выборке HOL. Были проведены дополнительные тесты. В частности,

в таблице 4 представлены результаты проверки H_0 различий между RED- и HOL-выборками парным тестом Стьюдента. Статистически значимая разность ($p_{value} \leq 0,05$) была выявлена по локусам 1, 3, 5, 6, 8 и 9¹². Средневзвешенное по локусам различие составило -2,3 ед. ($p_{value} = 0,003$) с 95 % доверительным интервалом [-4,2, -0,3].

Таблица 4 – Парный тест Стьюдента по локусам RED- и HOL-выборок /
Table 4 – Student's paired test by loci of RED- and HOL-samples

Locus	RED		HOL		$\Delta = \mu_R - \mu_H$	95% CI для Δ		p_{value} Δ
	μ_R	$\pm s.e.$	μ_H	$\pm s.e.$		lower	upper	
1	104,9	2,6	93,3	1,9	11,5	5,1	17,9	0,004
2	100,2	3,8	106,1	3,4	-5,9	-15,9	4,1	0,267
3	102,8	3,5	112,2	1,7	-9,4	-17,1	-1,7	0,034
4	87,9	2,3	89,5	2,0	-1,6	-7,5	4,3	0,609
5	89,9	2,8	104,0	2,5	-14,2	-21,5	-6,8	0,003
6	89,1	2,0	96,5	2,1	-7,4	-13,1	-1,6	0,027
7	108,2	0,8	104,1	2,5	4,1	-1,1	9,3	0,151
8	102,3	2,0	112,6	3,0	-10,3	-17,4	-3,2	0,015
9	101,5	1,0	92,6	4,1	8,9	0,8	17,1	0,050
10	95,7	2,6	97,5	1,6	-1,8	-7,7	4,2	0,576
11	101,6	2,8	106,2	0,6	-4,6	-10,1	0,9	0,128
μ_w	102,5	0,49	103,7	0,49	-2,3	-4,2	-0,3	0,003

Примечания: μ – среднее; $s.e.$ – стандартная ошибка; 95%CI – 95% доверительный интервал (бутстрэппинг); μ_w – средневзвешенное по локусам с «весами» $1/s.e.^2$.

Notes: μ – the average; $s.e.$ – the standard error; 95% CI is the 95% confidence interval (bootstrapping); μ_w – the weighted average of loci with «weights» of $1/s.e.^2$.

Тест Уилка (Wilks) многомерного дисперсионного анализа (Multivariate Analysis Of Variance, MANOVA)¹³ также показал статистически значимое различие RED- и HOL-выборок ($\lambda = 0,252$, $F = 5,93$, $p_{value} = 0,0024$). Дополнительно был использован непараметрический вариант MANOVA (NPMANOVA или PERMANOVA), если предположения о многомерном нормальном распределении и равных ковариационных матрицах всё же не выполнялись. Равенство средних проверялось пермутационным тестом с 999 повторами и мерой квадрата расстояния Махаланобиса. Тест подтвердил статистически значимое различие выборок ($F = 2,39$, $p_{perm} = 0,0019$). В то же время непараметрический тест Хотеллинга на равенство средних двух парных многомерных наборов данных (Paired Hotelling's test) по локусам со схожими номерами (пара – первая строка RED-выборки и первая строка HOL-выборки табл. 1 и т. д.) показал при критерии $T^2 = 93,4$ $p_{value} = 0,062$.

Двухфакторный ANOVA по смешанной модели. Если принять фактор «локус» как случайный (random factor), то эффект взаимодействия $BR \times L$ также будет случайным, а модель ANOVA будет относиться к смешанному типу (MixedModel, MM). Результаты ANOVA-MM (табл. 5) показали влияние BR-фактора на полокусные оценки разнообразия статистически незначимым ($p_{value} = 0,2821$). По L-фактору p_{value} было тоже больше 0,05, но меньше 0,1. Только эффект $BR \times L$ однофакторно отличался от нуля ($p_{value} < 0,0001$). В таких случаях говорят, что «не доказано отсутствие эффектов [главных факторов]» [24].

¹²С поправкой Бонферрони ($\alpha_{Bonf} = 0,05/11 = 0,0045$) только по локусам 1 и 5.

¹³MANOVA является обобщением обычного одномерного ANOVA и предназначен для выявления различий между группами по совокупности средних значений комплекса признаков. H_0 заключается в предположении о равенстве векторов средних значений для сравниваемых групп наблюдений.

Таблица 5 – Двухфакторный ANOVA-MM объединённой выборки /
Table 5 – Two-factor ANOVA-MM of the combined sample

Source	SS	Df	MS	F	pvalue	$\hat{\sigma}^2$
Breed (BR)	296,54	1	296,54	1,29	0,2821	-
Locus (L)	5402,15	10	540,22	2,35	0,0965	22,20
Interaction (BR×L)	2294,71	10	229,47	4,49	<0,0001	25,48
Residual (e)	6742,66	132	51,08	-	-	51,08
Total (T)	14736,06	153	-	-	-	98,76

При раздельном анализе RED- и HOL-выборок (табл. 2) варианса y'_{km} -оценок, обусловленная L-фактором ($\hat{\sigma}_L^2$), составила в среднем $(40,12 + 55,24)/2 = 47,68$. Анализ объединённой выборки по ANOVA-MM разделил этот компонент изменчивости на части: одна – межлокусная, которая оказалась почти вдвое меньше – 22,20, другая – взаимодействия, равная 25,48. Эффект $BR \times L$ рассчитали из отношения $\frac{\hat{\sigma}_{BR \times L}^2}{\hat{\sigma}_e^2} 100 = \frac{25,48}{98,76} 100 = 25,8 \%$.

Его можно интерпретировать как меру несоответствия профилей разнообразия локусов, т. е. меру различий RED- и HOL-выборок.

Эффекты взаимодействия визуализированы на рисунке 3. Действительное взаимодействие наблюдали в отношениях таких локусов, как 3 и 7, 1 и 5 и др., т. е. когда [условные] ранги оценок в RED-выборке повышались/снижались в HOL-выборке. Так, высокие в RED-выборке оценки по локусам 1, 7 и 9 снизились в HOL-выборке, но значительно повысились оценки локусов 2, 3, 5, 6, 8 и 11 (табл. 4). В случаях, когда различие между оценками локусов в другой породе увеличивалось или снижалось, но последовательность рангов не менялась (локусы 7 и 1, 10 и 4 и т. п.), имела место корреляция «порода-локус».

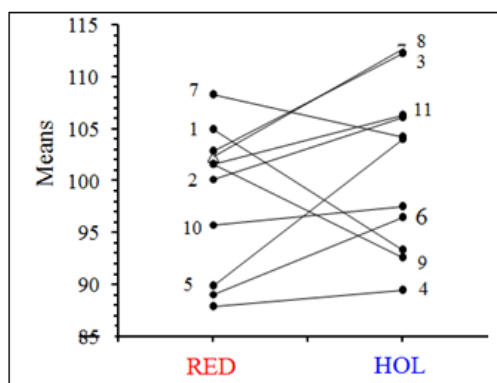


Рис. 3. Эффекты «порода×локус» /
Fig. 3. The effects of «breed×locus»

Ранговая корреляция Кендалла оценок одноимённых локусов между выборками составила $\tau = 0,382$ ($p_{value} = 0,1021$). Корреляция τ имеет интересную интерпретацию: если два объекта выбираются случайно из группы n объектов, то разность между вероятностью того, что они будут иметь одинаковый порядок как по X, так и по Y (P_S) и вероятностью того, что у них будет наблюдаться различие в порядках по X и Y (P_D) равна величине τ [25]. Исходя из отношения $\tau = P_S - P_D$ были рассчитаны вероятности совпадения $P_S = (1 + 0,382)/2 = 0,69$ и не совпадения (различия) $P_D = 1 - 0,69 = 0,31$ номеров *наугад* выбранной пары локусов из RED- и HOL-выборок. Имело место близость оценок $P_D = 0,31$ и взаимодействия $BR \times L = 0,258$ из ANOVA-MM. Как представляется, P_D также можно рассматривать в качестве меры различия породных выборок по профилям разнообразия одноимённых локусов.

Евклидовы дистанции. По каждой выборке были рассчитаны евклидовы дистанции¹⁴ между локусами (табл. 6).

В RED-выборке межлокусные дистанции варьировали от 11,7 до 111,9 %, имели среднее $48,2 \pm 3,32 \%$ и 95 % CI [41,5; 54,8]. В HOL-выборке дистанции были в диапазоне 8,3–162,4 % со средней $63,4 \pm 5,07 \%$ и 95 % CI [53,2; 73,6]. Парный тест Стьюдента показал статистически значимое различие между средними (HOLvsRED), именно: $15,3 \pm 5,3 \%$, $p_{value} = 0,0057$, 95 % CI [4,6; 25,9]. С другой стороны, корреляция Мантеля¹⁵ между треугольными матрицами была $R_M = 0,257 \pm 0,130$ с 95 % CI [-0,009; 0,506] ($p_{value} = 0,056$). Расстояния между профилями локусов в выборке RED были слабо зависимы от расстояний между аналогичными профилями локусов в выборке HOL (и наоборот).

В таблице 7 представлены евклидовы кросс-дистанции между локусами двух выборок.

¹⁴Формула расчёта: $d_{Euclid} = \sqrt{\sum (y'_{km} - y'_{lm})^2}$. Предварительно y_{km} -оценки были выражены в процентах от среднего (μ_m).

¹⁵Метод Мантеля или «соотношение квадратов» (quadratic assignment) проверяет H_0 , что дистанции между объектами в матрице A независимы от дистанций между теми же самыми объектами в матрице B.

Таблица 6 – Дистанции Евклида между локусами в выборке RED (над диагональю) и в выборке HOL (под диагональю) /
Table 6 – Euclidean distances between loci in the RED-sample (above the diagonal) and in the HOL-sample (below the diagonal)

Locus	1	2	3	4	5	6	7	8	9	10	11
1		31,3	67,7	33,0	29,0	52,2	42,9	42,7	27,1	35,9	38,9
2	34,1		94,9	31,9	36,5	23,8	72,3	68,6	55,1	14,4	65,7
3	107,4	80,7		80,7	69,5	111,9	28,3	28,0	41,8	97,5	33,3
4	28,4	58,4	134,1		11,7	37,4	63,8	56,1	44,0	26,6	54,6
5	87,0	63,1	24,7	114,3		46,3	53,7	45,0	34,1	34,2	44,2
6	8,3	31,9	106,0	28,6	86,7		91,7	85,9	73,3	19,0	83,9
7	26,9	15,3	85,1	51,4	67,2	23,2		16,5	21,6	75,8	20,0
8	135,9	110,4	30,4	162,4	50,2	134,6	114,2		20,7	70,8	20,9
9	54,0	41,1	64,2	80,6	40,0	55,7	42,8	89,1		58,4	13,5
10	15,6	24,5	101,2	36,0	81,4	13,9	21,5	129,9	49,9		70,2
11	68,5	41,3	42,1	94,9	23,8	67,2	46,9	70,2	29,8	60,5	

Таблица 7 – Евклидовы дистанции между локусами выборок RED (X) и HOL (Y) /
Table 7 – Euclidean distances between the sample loci RED (X) and HOL (Y)

Locus	X: 1	2	3	4	5	6	7	8	9	10	11
Y: 1	30,5	23,1	89,9	16,9	25,5	30,1	68,9	64,5	50,1	16,0	62,0
2	9,4	38,4	66,3	37,7	33,9	58,5	40,5	43,0	26,3	41,8	38,3
3	83,3	111,5	29,4	100,7	90,3	129,6	41,7	46,2	61,8	113,5	56,6
4	54,0	29,0	116,3	39,7	50,0	15,3	95,1	90,8	76,1	23,3	87,0
5	65,4	93,5	13,2	78,8	68,2	110,5	25,8	26,3	41,2	94,8	36,5
6	26,9	16,0	89,4	21,0	27,5	27,9	67,2	63,3	49,2	11,2	61,2
7	9,2	29,1	70,5	29,6	26,4	48,8	46,3	43,9	31,3	31,2	43,9
8	112,4	139,6	50,4	127,6	116,8	157,1	70,4	72,7	89,5	141,8	82,3
9	41,5	64,6	40,7	42,9	32,7	78,2	33,4	24,8	20,8	64,5	23,8
10	20,3	21,1	83,0	23,3	26,4	37,1	60,7	58,6	41,5	24,1	52,3
11	43,5	72,7	25,9	62,5	52,2	91,6	4,4	14,6	20,1	75,9	17,5

Все дистанции ($n = 121$) находились в диапазоне от 4,4 (RED:L7-HOL:L11) до 157,1 % (RED:L6-HOL:L8). Усреднённая оценка, характеризующая степень различия между двумя случайно выбранными локусами из RED- и HOL-выборок, была $\bar{d} = 54,0 \pm 2,97\%$ с 95 % CI [48,1; 59,9]. Усреднённая дистанция, рассчитанная по одномерным локусам (диагональ), составила $\bar{d}' = 37,8 \pm 5,35\%$ с 95 % CI [27,9; 57,5]. Евклидовы кросс-дистанции также были рассчитаны по редуцированным данным: переменным α -разнообразия (n_a , s_{n_e} и H_e). Усреднённая дистанция по одномерным локусам была $34,1 \pm 6,25\%$ с 95 % CI [20,4; 47,8]. Две последние оценки (в долях – 0,378 и 0,341)

хорошо согласовывались с полученными ранее индексами генетической дифференциации RED- и HOL-выборок – 0,29-0,37 [14]¹⁶.

Анализ главных компонент (Principal Components Analysis, PCA). С помощью процедуры PCA: (а) рассчитали корреляцию между переменными; (б) сгенерировали новые латентные переменные (факторы, компоненты) как линейные комбинации исходных переменных; (в) снизили размерность, выявив два главных компонента, объясняющих максимум общей дисперсии исходных данных; (г) спроецировали ординацию объектов-локусов в пространстве главных компонент.

¹⁶Отметим также, что усреднённая дистанция, рассчитанная только по s_{n_e} , была 0,28, по H_e – 0,15. Последняя оценка близка к значению индекса фиксации аллелей, рассчитанному по AMOVA (Analysis of MOlecular VAriance), 0,11 [13].

На рисунке 4 представлены ординации локусов в системе координат двух главных компонент (PC1 и PC2). В первом ряду даны ординации локусов в RED- и HOL-выборках (близкие точки-локусы обладали более сходными свойствами). Суммарная информативность PC1 и PC2 составила 92,5 и 95,0 % соответственно. Вероятные цепочки отношений между локусами визуализировали некорневые графы, построенные методом «минимального остовного дерева» (Minimum Spanning Tree, MST). MST-метод использовали для (а) корректной группировки (структурирования) локусов и

(б) лучшего восприятия различий. Ординации существенно различались и по конфигурации, и по «сгусткам» локусов, близких по компонентным оценкам (эллипсы).

RED + HOL-контуры (convexhull) визуализировали площади разброса локусов в выборках. Облака точек-локусов двух выборок смешивались. Площадь контура HOL-выборки почти полностью покрывала такую же RED-выборки с некоторым сдвигом по PC1 в «+»-направлении (что соответствовало результатам табл. 2 и 4).

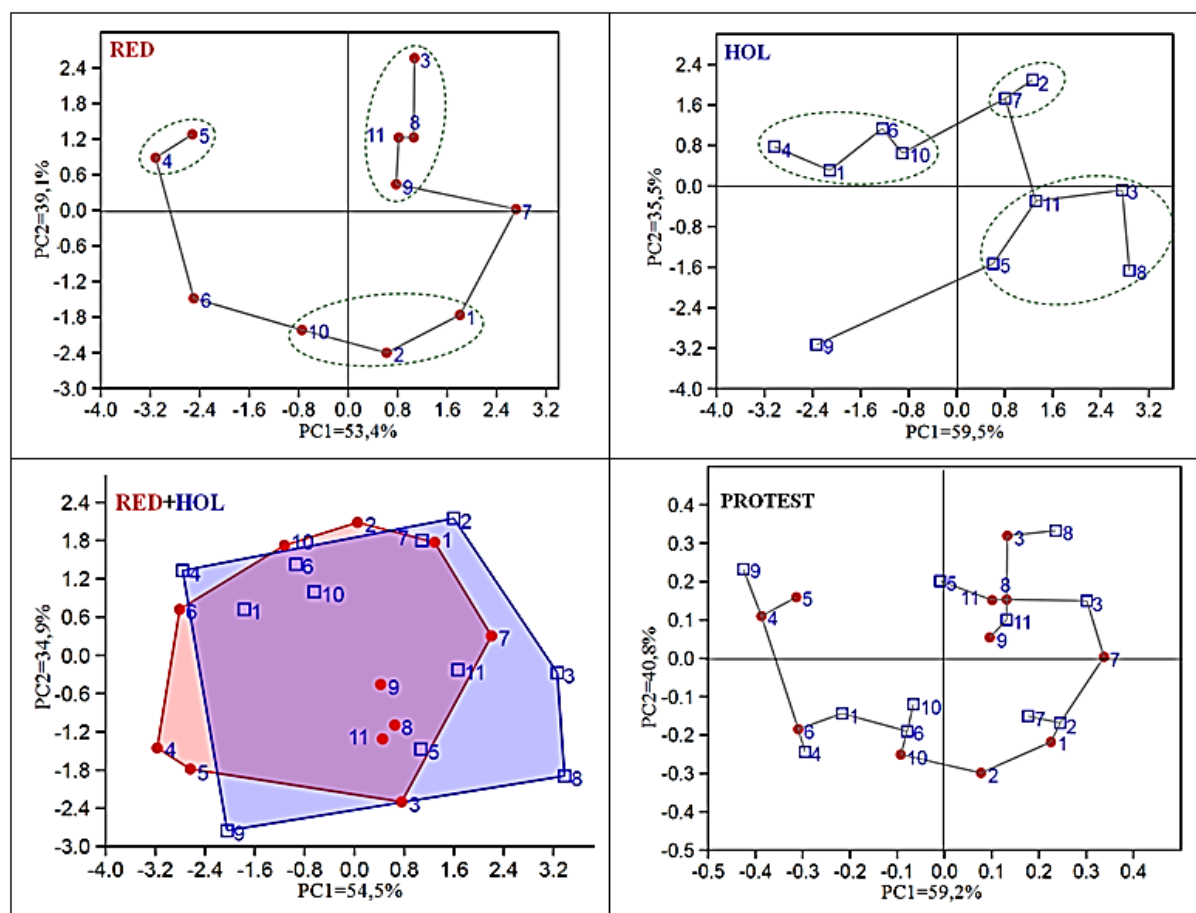


Рис. 4. Ординации локусов в RED и HOL выборках, RED+HOL – контуры выборок в объединённом анализе; PROTEST – ординация локусов после суперналожения /

Fig. 4. Ordinations of loci in RED and HOL samples, RED+HOL – sample contours in the united analysis; PROTEST – ordination of loci after superimposition

Для оценки различия/соответствия двух ординаций локусов использовали прокрустовый анализ (Procrustes analysis)¹⁷; RED-матрица была «референсной», а HOL-матрица – «ротационной». Ординации локусов двух выборок после «суперналожения» представлены

на рисунке 4 (PROTEST). Можно видеть «сгустки» локусов, близких по латентным переменным не только «своей» выборки, но и «чужой». Например, в верхней правой четверти локусы 3, 8 и 11 обеих выборок.

¹⁷Прокрустовый анализ сравнивает наборы многомерных форм, пытаясь преобразовать их в состояние суперналожения [20, 21]. В программе PROTEST это достигается путем минимизации сумм квадратов расстояний между соответствующими точками в каждой форме посредством перемещения, отражения, вращения и масштабирования их координатных матриц, что делает ординации сопоставимыми, помещая их в одну и ту же систему координат.

Суммы квадратов разностей между координатными точками одномерных локусов

Locus	1	2	3	4	5	6	7	8	9	10	11
Невязка	0,20	0,05	0,06	0,13	0,09	0,05	0,05	0,04	0,30	0,02	0,00

Большие «невязки» имели место по локусам 1, 4, 5 и 9, максимальное соответствие – по локусам 10 и 11.

Программа PROTEST минимизировала суммарную «невязку» – m^2 (чем больше величина, тем меньше соответствие двух ординаций), рассчитала m_{12} -статистику (чем меньше величина, тем больше соответствие) и вероятность отвержения нулевой гипотезы (статистически значимое соответствие при $p_{perm} \leq 0,05$). Приблизительную степень соответствия двух ординаций (r_{Proc}^2) можно получить из отношения $m_{12}^2 = 1 - r_{Proc}^2$ [20], именно: $r_{Proc}^2 = 1 - m_{12}^2$. По программе PROTEST с 999 перестановками были получены следующие результаты: $m^2 = 0,994$, $m_{12} = 0,747$, $p_{perm} = 0,164$ ($r_{Proc}^2 = 0,253$, $r_{Proc} = 0,503$). Сопряжённость ординаций локусов RED- и HOL-выборок в пространстве координат двух главных компонент была статистически незначимой.

Двухблочный анализ частных наименьших квадратов (Two-Block Partial Least-Squares, 2B-PLS). Относительно новый 2B-PLS-метод¹⁸ подобен PCA, но для изучения структуры ковариации между двумя блоками многомерных данных, которые обрабатываются одновременно. Метод 2B-PLS даёт возможность выявить степень общности (ко-дисперсию) двух блоков данных. В основе 2B-PLS допущение, что наблюдения (исходные переменные) линейно связаны с небольшим числом латентных переменных (скрытых, непосредственно не наблюдаемых и не измеряемых, явлений, факторов, компонентов, как в PCA). 2B-PLS создаёт ортогональный набор латентных переменных (Latent Variables, LV), которые являются взвешенными линейными комбинациями исходных переменных¹⁹. Линейные комбинации формируются таким образом, чтобы новые интегрированные LV максимизировали квадрат ковариации между двумя многомерными блоками. Такие LV обеспечивают оценку всех возможных межблочных кова-

риаций и количественно определяют, как ассоциативный эффект выражен в анализируемых многомерных блоках.

Если имеется набор данных Y с двумя блоками Y_1 и Y_2 , с числом переменных p_1 и p_2 , то корреляционная(ковариационная) матрица R по Y_1 и Y_2 имеет структуру

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix},$$

где $R_{21} = R_{12}^t$ (R_{12}^t –транспонированная матрица) [28]. В 2B-PLS используется сингулярное разложение²⁰ кросс-корреляционной матрицы R_{12} для получения произведения $F_1 D F_2^t$. Матрица D –диагональная и содержит сингулярные числа (singularvalues) λ_i . Каждый столбец матрицы F_1 содержит веса/нагрузки для линейных комбинаций переменных блока 1, а каждый столбец F_2 – веса/нагрузки для переменных блока 2. Число строк соответствует числу переменных, p_1 и p_2 , в каждом блоке, а число столбцов (размерностей) равно меньшему значению из p_1 и p_2 . Первые k столбцов F_1 и F_2 интерпретируют как «дающие наилучшую аппроксимацию методом наименьших квадратов R_{21} -матрицы». Выражение $\sum \lambda_i^2$ равно сумме квадратов элементов в матрице R_{21} и является мерой общей ковариации между наборами переменных двух блоков. Учитывая, что максимально возможная величина $\sum \lambda_i^2$ равна $p_1 \times p_2$ (которая была бы получена, когда все корреляции равны 1), то отношение $\sum \lambda_i^2 / (p_1 p_2)$ является мерой общего квадрата ковариации («Squaredcovar») между двумя блоками (корреляции, если признаки были стандартизированы).

Результаты 2B-PLS-анализа сведены в таблице 8. Слева представлены сингулярные величины (λ_i) для семи размерностей ($axis$ – ось; их число равно числу переменных) и общий квадрат ковариации между RED- и HOL-выборками по семи переменным – $\sum \lambda_i^2 = 7,0296$.

¹⁸Был предложен Германом Оле Андреасом Вольдом (Herman Ole Andreas Wold) в 1975 году [26]. В современной интерпретации – Projectionto Latent Structure(PLS) – проекция на латентную структуру.

¹⁹В 2B-PLS для блоков A и B с переменными A_i и B_j латентные переменные $LV_A = \sum_{i=1} \alpha_i A_i$ и $LV_B = \sum_{j=1} \beta_j B_j$, где α_i и β_j – веса: $\alpha_i \propto \sum_{j=1} r_{ij} \beta_j$ и $\beta_j \propto \sum_{i=1} \alpha_i r_{ij}$ ($\propto$ – пропорционально; r_{ij} – корреляция), которые рассчитываются так, чтобы $\sum \alpha_i^2 = \sum \beta_j^2 = 1$ [27].

²⁰Сингулярное разложение (Singular Value Decomposition, SVD) – декомпозиция ковариационной матрицы с целью её приведения к каноническому виду, обеспечивающему сокращение вычислений, облегчение анализа и устойчивость всех матричных процедур.

Если бы между всеми переменными коэффициенты корреляции были равны 1, то общий квадрат ковариации был бы равен $7 \times 7 = 49$. Доля общего квадрата ковариации между данными RED и HOL, которая была максимизирована 2B-PLS-анализом, составила $Squared\ covar = 7,0296/49 = 0,143$, или 14,3 % относительно максимально возможного. Из них

0,83 ($= 2,4152^2/7,0296$), или 83 % приходилось на первую ось (столбец «% covar»), также являющейся первой латентной переменной (LV1), и 16,7 % – на вторую латентную переменную (LV2). В сумме LV1 и LV2 объясняли 99,7 % общей ковариации между двумя многомерными наборами данных.

Таблица 8 – Результаты 2B-PLS-анализа RED- и HOL-выборок /

Table 8 – Results of 2B-PLS analysis of RED and HOL samples

Axis	λ_i	% covar	Locus	RED: LV1	HOL: LV1	RED: LV2	HOL: LV2
1	2,4152	82,9790	1	0,3800	2,0561	-2,1895	0,6258
2	1,0822	16,6610	2	1,1945	-0,2361	-2,2852	-2,4530
3	0,1281	0,2335	3	-2,6466	-2,5204	0,8646	-0,9504
4	0,0930	0,1231	4	1,3671	3,0669	2,9019	0,7228
5	0,0145	0,0030	5	0,5451	-1,1794	2,5486	1,0593
6	0,0040	0,0002	6	2,5068	1,5928	0,5493	-0,4410
7	0,0014	0,0000	7	-1,4936	-0,0227	-1,6410	-1,9049
$\Sigma \lambda_i^2$	7,0296	100,0	8	-1,6958	-3,3268	-0,0866	0,3895
Squared covar = $\Sigma \lambda_i^2/p_1p_2 = 7,0296/(7 \times 7) = 0,1435$			9	-0,7005	0,8037	-0,1035	3,5525
			10	2,0891	1,0906	-0,6664	-0,2497
			11	-1,5460	-1,3248	0,1078	-0,3509
Linear correlation, r				0,717 ($p_{value} = 0,013$)		0,395 ($p_{value} = 0,229$)	

Примечания: λ_i = singular value – сингулярная величина; % covar = % covariation – % ковариации; LV – Latent Variable – латентная переменная; Squared covar (covariance) – квадрат ковариации /

Note: λ_i – singular value; % covar – % covariation; LV – Latent Variable; Squared covar – covariance

В таблице 8 справа даны координатные оценки (*scores*) локусов по LV1 и LV2 для RED- и HOL-выборок (RED:LV1 и RED:LV2, HOL:LV1 и HOL:LV2). В обеих выборках наибольшие нагрузки/веса в LV1 имели, в основном, показатели n_a , S_{ne} и H_e , характеризующие внутривыборочное α -разнообразие. Поэтому для LV1 также использовали акроним αLV (алатентная переменная). Для LV2 использовали акроним βLV , т. к. наибольшие нагрузки имели показатели G_{ST} , F'_{ST} , D_J и D'_{β} , характеризующие межсубвыборочное β -разнообразие.

По αLV линейная связь (r) между RED- и HOL-матрицами/блоками оценивалась в 0,717 при $p_{value} = 0,013$ с 95 % CI [0,21, 0,92]. Информационный показатель общности (ко-дисперсия) составил $r^2 = 0,515$, или 51,5 %. По βLV корреляция между матрицами была статистически незначимой (0,395, $p_{value} = 0,229$). Усреднённая оценка составила $\bar{r} = 0,56$ ($p_{value} = 0,025^{21}$), что дало показатель общности $\bar{r}^2 = 31,4$ %.

Схожие результаты были получены, когда векторы координатных оценок локусов

по αLV и βLV были объединены в пределах выборок: $LV_{RED} = \alpha LV_{RED} U \beta LV_{RED}$ и $LV_{HOL} = \alpha LV_{HOL} U \beta LV_{HOL}$, где U – символ объединения. Между LV_{RED} и LV_{HOL} коэффициент конгруэнтности (согласованности) составил $r_c = 0,57$ ($p_{value} = 0,0056, 95$ % CI [0,19, 0,80]), показатель общности выборок $r_c^2 = 32,5$ %. Кроме того, для каждой выборки по k -ому локусу был рассчитан «комплексный» латентный признак: $\gamma LV_{RED(k)} = \alpha LV_{RED(k)} + \beta LV_{RED(k)}$ и $\gamma LV_{HOL(k)} = \alpha LV_{HOL(k)} + \beta LV_{HOL(k)}$. Линейная корреляция между γLV_{RED} и γLV_{HOL} была 0,5 ($p_{value} = 0,114$), показатель общности – 25 %.

На рисунке 5 в верхнем ряду слева представлен разброс локусов в плоскости координат RED- и HOL-выборок по αLV . Облако точек-локусов и линия регрессии говорят о положительной корреляции. Коэффициент детерминации регрессии (R^2), равный 51,5 %, также является информационным показателем общности двух наборов многомерных данных. Справа – результаты кластеринга – дендрограмма (дерево классификации) по алгоритму

²¹Комбинированная статистическая значимость была рассчитана по методу Фишера [29].

невзвешенного парно-группового метода с арифметическим усреднением (Unweighted Pair-Group Method using arithmetic Averages – UPGMA). Кофенетическая корреляция, равная 0,664, показатель того, насколько хорошо характер отношений (сходство/несходство) между локусами представлялось дендрограммой²². В точках ветвления даны бутстрэп-

вероятности объединения локусов в кластеры, близких по αLV . С учётом этих вероятностей на ординации слева овалами выделена возможная группировка локусов – структура ковариации двух многомерных наборов данных (пунктир – бутстрэп-вероятность $\cong 50\%$, тонкая линия – $\cong 75\%$, жирная линия – 100%).

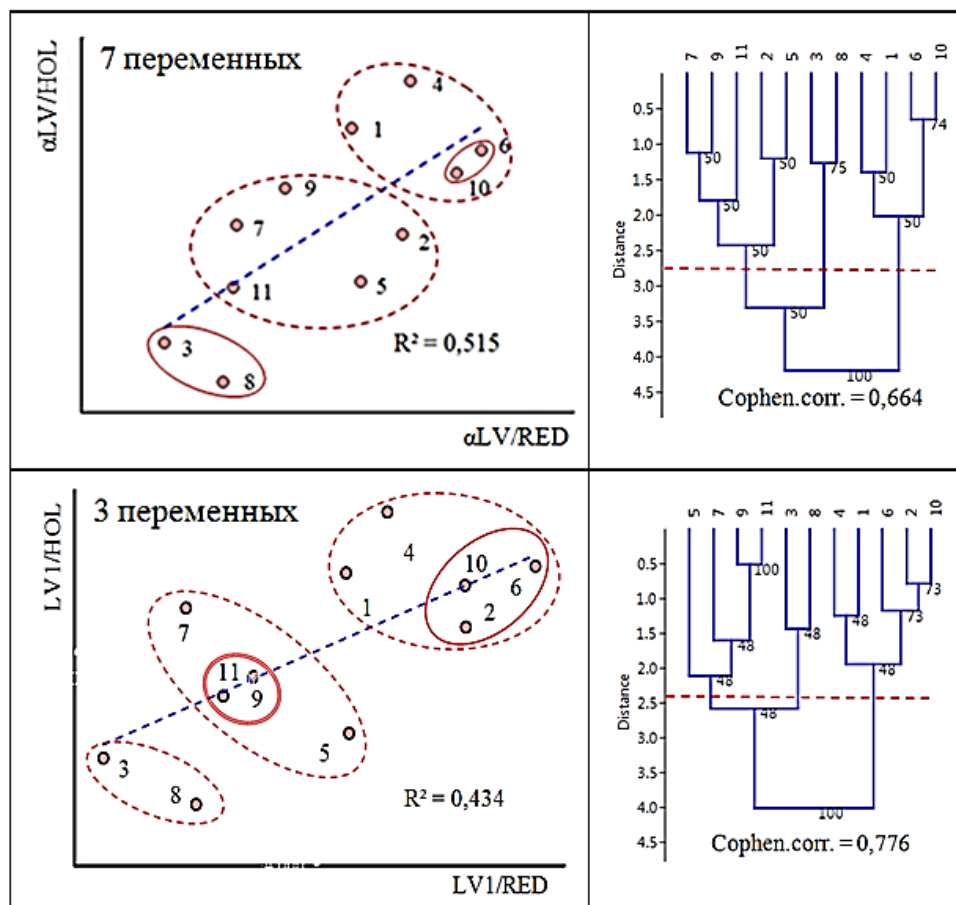


Рис. 5. Структуры ковариации локусов RED- и HOL-выборок (слева) и UPGMA дендрограммы (справа) по αLV (полные данные) и LV1 (редуцированные данные) /

Fig. 5. Covariance structures of the loci in the samples RED and HOL (left) and UPGMA dendrograms (right) by to αLV (complete data) and LV1 (reduced data)

Выше отмечалось, что в βLV наибольшие нагрузки (веса) имели показатели β -разнообразия. Линейная связь двух выборок по βLV была статистически незначимой, что, как представляется, и следовало ожидать между показателями дифференциации субвыборок в RED- и HOL-выборках. Исходя из этого и учитывая небольшой вклад βLV в максимально возможную ковариацию (16,7 %), был проведён

2B-PLS-анализ по *редуцированным* данным – трём переменным α -разнообразия (n_a , S_n и H_e).

Максимальный квадрат ковариации (Squaredcovar) составил 0,393. При этом 99,9 % приходилось на первую латентную переменную (LV1). Линейная связь по LV1 между RED- и HOL-выборками оценивалась в 0,659 при $p_{value} = 0,0253$ с 95 % CI [0,097; 0,902]. Показатель ко-дисперсии (R^2) был 43,4 %.

²²Кофенетическая корреляция вычисляется как линейный коэффициент корреляции Пирсона (или ранговый коэффициент Спирмена) между всеми расстояниями, полученными по дереву иерархии, и элементами исходной матрицы дистанций.

Результаты 2B-PLS анализа и кластеринга представлены на рисунке 5 в нижнем ряду. Ординации и структуры ковариации локусов в пространстве координат RED- и HOL-выборок по полному (αLV) и редуцированному ($LV1$) наборам данных имели хорошее соответствие (прокрустов тест: $m_{12} = 0,0728$, $p_{perm} = 0,001$, $r_{Proc}^2 = 0,927$).

Расширение. Подход к рассмотрению полокусных оценок показателей разнообразия (профилей локусов) выборки, как матрицы многомерного набора данных с последующей нормализацией и обработкой многомерными методами статистики, был применён для выяв-

ления индекса дифференциации и ординации пород. В частности, были использованы данные из публикации чешских учёных [30], в которой обсуждались результаты анализа генетического разнообразия по 11 STR-локусам семи пород крупного рогатого скота Центральной Европы. В таблице 9 воспроизведена описательная статистика разнообразия (Descriptive Diversity Statistics, **DDS**) по породам.

Для оценки дифференциации пород авторы использовали несколько мер, одна из которых была D_S – стандартная дистанция Ней (табл. 10 над диагональю).

Таблица 9 – Описательная статистика разнообразия по породам молочного скота Центральной Европы [30] /
Table 9 – Descriptive statistics of diversity by breeds of dairy cattle in Central Europe [30]

Breed	Code	n_a	H_o	H_e	PIC
Czech Pied	CP	7,90	0,779	0,764	0,757
Slovakian Pied	SP	7,45	0,654	0,650	0,642
Slova Pinzgau	SPg	9,00	0,759	0,751	0,745
Holstein	HO	5,82	0,806	0,697	0,685
Polish Red	PR	8,45	0,748	0,747	0,741
German Red	GR	7,00	0,675	0,716	0,711
Czech Red	CR	7,18	0,789	0,751	0,749

Примечания: H_o – фактическая гетерозиготность; PIC – информационный контент полиморфизма /
Notes: H_o – observed heterozygosity; PIC – polymorphism information content.

Таблица 10 – Треугольные матрицы парных дистанций между породами (над диагональю – D_S [30],
под диагональю – D_{Euclid}) /
Table 10 – Triangular matrices of paired distances between breeds (above the diagonal – D_S [30], below the
diagonal – D_{Euclid})

Breed	CP	SP	SPg	HO	PR	GR	CR
CP		0,102	0,181	0,374	0,198	0,208	0,171
SP	0,429		0,326	0,577	0,319	0,294	0,298
SPg	0,162	0,412		0,267	0,190	0,280	0,182
HO	0,375	0,329	0,466		0,332	0,420	0,332
PR	0,114	0,364	0,076	0,398		0,163	0,091
GR	0,267	0,211	0,316	0,243	0,246		0,151
CR	0,127	0,365	0,246	0,252	0,178	0,193	

Усреднённая D_S -дистанция составила 0,260 с 95 % CI [0,207; 0,312]. Рассчитанная нами по масштабированным DDS-данным треугольная матрица парных евклидовых дистанций (табл. 10 под диагональю) имела среднее 0,275 с 95 % CI [0,224; 0,325]!

По каждой треугольной матрице был сделан PCA. Информативность PCA/ D_S по трём главным компонентам составила $46,5 + 32,0 + 14,7 = 93,2$ %, по PCA/ D_{Euclid} – $61,2 + 20,3 + 13,3 = 94,8$ %. Прокрустов тест на соответствие структур 3D-ординаций пород

(по D_S и D_{Euclid}) дал статистики: $m_{12} = 0,387$, $p_{perm} = 0,054$ (доверительная вероятность 94,6 %), $r_{Proc}^2 = 0,613$ ($r_{Proc} = 0,783$).

Также были использованы данные из работы [31] с результатами анализа полиморфизма 10 STR-локусов и информацией о состоянии генетического разнообразия и структуре 11 пород свиней. По каждой породе DDS-данные включали пять показателей: среднее и эффективное число аллелей на локус, аллельное разнообразие, наблюдаемую и ожидаемую гетерозиготности. По масштаби-

рованным DDS-данным мы рассчитали парные D_{Euclid} между породами. Усреднённая дистанция составила 0,376 с 95 % CI [0,327; 0,425].

Авторы в работе дали треугольную матрицу парных генетических F_{ST} -дистанций. Их усреднённое значение было 0,139 (95 % CI [0,126; 0,152]), что в 2,7 раза ниже средней D_{Euclid} по DDS-данным. Однако было показано [32, 12], что F_{ST} -мера занижает оценки, рассчитанные по полиморфным локусам. После корректировки парных F_{ST} по [33] усреднённая оценка составила 0,358 и была достаточно близка к величине усреднённых парных D_{Euclid} (0,376).

Как и в первом случае, были получены PCA-ординации пород. Информативность PCA/ F_{ST} по трём главным компонентам составила $43,5 + 23,7 + 13 = 80,2$ %, по PCA/ D_{Euclid} – $53,1 + 35,8 + 4,1 = 93$ %. Прокрустов тест на сходство 3D-ординаций пород дал статистики: $m_{12} = 0,489$, $p_{perm} = 0,004$ (высокозначимое соответствие), $r_{Proc}^2 = 0,511$ ($r_{Proc} = 0,715$). Таким образом, оба примера демонстрировали неплохое согласие результатов, полученных по DDS-данным, с «референсными» результатами по «традиционным» методам.

Заключение. Анализ масштабированных оценок трёх показателей α (n_a , s_{n_e} и H_e) и четырёх показателей β ($G'_{ST(N)}$, G''_{ST} , D_{est} , D') разнообразия (y'_{km}) по каждому из 11 STR-локусов двух [подразделённых] выборок производителей красной скандинавской (RED, $n = 29$) и голштинской (HOL, $n = 45$) пород показал следующие результаты.

При относительно равных остаточных вариансах y'_{km} межлокусная вариация в HOL-выборке была на 37,7 % больше, чем в RED-выборке. Гомогенность (надёжность) индивидуальных оценок была ~50 %, усреднённых по профилям локусов ~86 %.

По ANOVA фиксированного типа статистически значимое влияние на изменчивость y'_{km} имели факторы «порода» – 2 %, «локус» – 36,7 % и их взаимодействие – 15,6 % ($p_{value} < 0,02$). По модели смешанного типа статистически значимым было влияние только взаимодействия «порода×локус» – 25,8 % ($p_{value} < 0,0001$). Порода производителей влияла на отношение между локусами и оценками показателей генетического разнообразия.

Рассчитанная из корреляции Кендалла ($\tau = 0,382$) вероятность совпадения порядков [по усреднённым оценкам] наугад выбранной пары локусов из RED- и HOL-выборок оценивалась в 69 %, несовпадения – 31 %. Последнее значение и эффект взаимодействия «порода×локус», равный 25,8 %, можно интерпрети-

ровать как показатели относительного различия породных выборок.

Расстояния между профилями локусов (евклидовы дистанции) в RED-выборке слабо коррелировали с аналогичными расстояниями в выборке HOL ($R_M = 0,257$, $p_{value} = 0,056$). Усреднённая дистанция между профилями локусов в HOL-выборке составила 63,4 % и была на 15,2 процентных пункта больше, чем в RED-выборке ($p_{value} = 0,0057$). Усреднённая дистанция между выборками по одноимённым локусам была 37,8 % с 95 % CI [27,9; 57,5], по редуцированной выборке (только переменные α -разнообразия) – 34,1 % с 95 % CI [20,4; 47,8]. Величины двух последних средних были близки к оценкам генетической дифференциации пород, рассчитанным по частотам аллелей (29–37 %) [14].

При PCA-анализе профилей локусов RED-выборки два первых главных компонента объясняли 92,5 % общей дисперсии y'_{km} , в HOL-выборке – 95 %. Соответствие 2D-ординаций локусов двух выборок было слабым, их конфигурации и структуры существенно отличались (прокрустов тест: $m_{12} = 0,75$, $p_{perm} = 0,164$, $r_{Proc}^2 = 0,25$).

При 2B-PLS-анализе 83 % от максимально возможной ковариации между многомерными данными RED- и HOL-выборок объяснялось первой интегрированной латентной переменной (αLV) и 16,7 % – второй (βLV). По αLV линейная связь между блоками данных оценивалась в 0,717 ($p_{value} = 0,013$), показатель общности выборок составил 51,4 %. По βLV корреляция была статистически незначимой (0,395, $p_{value} = 0,233$). Усреднённая по αLV и βLV оценка корреляции составила 0,56 ($p_{value} = 0,025$), что дало показатель общности 31,4 %. Анализ по редуцированным данным (только переменные α -разнообразия) показал, что 99,9 % возможной ковариации приходилось на первую LV ($LV1$). Линейная связь по $LV1$ между RED и HOL выборками составила 0,659 ($p_{value} = 0,0253$), показатель общности – 43,4 %. По полному (αLV) и редуцированному ($LV1$) наборам данных имело место хорошее соответствие ординаций локусов в пространстве координат двух выборок (прокрустов тест: $m_{12} = 0,0728$, $p_{perm} = 0,001$, $r_{Proc}^2 = 0,927$). В структуре межвыборочной ковариации выделялись кластеры локусов с бутстрэп-вероятностью [группировки] 50, 75 и 100 %. Можно предположить, что RED- и HOL-выборки имели некоторую согласованность (конгруэнтность) по показателям α -разнообразия одноимённых локусов.

Применение «многомерного» подхода к опубликованным в литературе данным описательной статистики разнообразия *пород* показало достаточно высокое соответствие оценок показателей дифференциации и результатов РСА с таковыми, полученными на базе «традиционных» (частотных) методов. Так, прокрустовая корреляция между «многомерной» и «традиционной» структурами семи пород крупного рогатого скота составила 0,783 ($p_{perm} = 0,054$), 11 пород свиней – 0,715 ($p_{perm} = 0,004$).

Итак, анализ полокусных оценок показателей генетического разнообразия, с одной стороны, способствовал извлечению из многомерных наборов данных дополнительной информации (влияние факторов на изменчивость оценок, дистанции между локусами в пределах и между выборками, визуализация

структурной организации локусов, различие и общность выборок), с другой стороны, не противоречил результатам, полученными «традиционными» методами (генетическая дифференциация выборок). Дополнительная информация может использоваться в разведочном анализе генетических данных (для ориентации в первом приближении), при разработке мероприятий по сохранению генофондных стад, в мета-анализе генетической структуры пород по описательной статистике разнообразия, имеющейся, как правило, в многочисленных публикациях. В общем, как представляется, рассмотренные в статье подходы и методы расширяют возможности популяционно-генетических [и селекционно-зоотехнических] исследований, в которых многомерные наборы данных являются нормой, а не исключением.

References

1. Кузнецов В. М. Анализ полокусных оценок аллельного разнообразия STR-маркеров в выборке быков-производителей. *Аграрная наука Евро-Северо-Востока*. 2023;24(5):888–906. DOI: <https://doi.org/10.30766/2072-9081.2023.24.5.888-906> EDN: LCTSP
2. Kuznetsov V. M. Analysis of locus estimates of allelic diversity of STR markers in a sample of breeding bulls. *Agrarnaya nauka Evro-Severo-Vostoka = Agricultural Science Euro-North-East*. 2023;24(5):888–906. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2023.24.5.888-906>
3. Хедрик Ф. Генетика популяций. М.: Техносфера, 2003. 592 с. Khedrik F. Population genetics. Moscow: Tekhnosfera, 2003. 592 p.
4. Nei M. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences*. 1973;70(12):3321–3323. DOI: <https://doi.org/10.1073/pnas.70.12.3321>
5. Weir B. S., Cockerham C. C. Estimating F-statistics for the analysis of population structure. *Evolution*. 1984;38(6):1358–1370. DOI: <https://doi.org/10.2307/2408641>
6. Meirmans P. G., Hedrick P. W. Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*. 2011;11(1):5–18. DOI: <https://doi.org/10.1111/j.1755-0998.2010.02927.x>
7. Jost L. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology*. 2008;17(18):4015–4026. DOI: <https://doi.org/10.1111/j.1365-294X.2008.03887.x>
8. Shannon C. E. A mathematical theory of communication. reprinted with corrections from. *The Bell System Technical Journal*. 1948;27(3):379–423, 623–656. URL: <https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>
9. Sherwin W. B. Entropy and information approaches to genetic diversity and its expression: Genomic geography. *Entropy*. 2010;12(7):1765–1798. DOI: <https://doi.org/10.3390/e12071765>
10. Кузнецов В. М. F-статистики Райта: оценка и интерпретация. *Проблемы биологии продуктивных животных*. 2014;(4):80–104. Режим доступа: <https://www.elibrary.ru/item.asp?id=22833217> EDN: TFRDMN
11. Kuznetsov V. M. Wright's F-statistics: estimation and interpretation. *Problemy biologii produktivnykh zhivotnykh = Problems of Productive Animal Biology*. 2014;(4):80–104. (In Russ.). URL: <https://www.elibrary.ru/item.asp?id=22833217>
12. Кузнецов В. М. Методы Нея для анализа генетических различий между популяциями. *Проблемы биологии продуктивных животных*. 2020;(1):91–110. DOI: <https://doi.org/10.25687/1996-6733.prodanimbiol.2020.1.91-110> EDN: DSEMYO
13. Kuznetsov V. M. Nei's methods for analyzing genetic differences between populations. *Problemy biologii produktivnykh zhivotnykh = Problems of Productive Animal Biology*. 2020;(1):91–110. (In Russ.). DOI: <https://doi.org/10.25687/1996-6733.prodanimbiol.2020.1.91-110>
14. Кузнецов В. М. Сравнение методов оценки генетической дифференциации популяций по микросателлитным маркерам. *Аграрная наука Евро-Северо-Востока*. 2020;21(2):169–182. DOI: <https://doi.org/10.30766/2072-9081.2020.21.2.169-182> EDN: FYQNT
15. Kuznetsov V. M. Comparison of methods for evaluating genetic differentiation of populations by microsatellite markers. *Agrarnaya nauka Evro-Severo-Vostoka = Agricultural Science Euro-North-East*. 2020;21(2):169–182. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2020.21.2.169-182>
16. Кузнецов В. М. Оценка генетической дифференциации популяций молекулярным дисперсионным анализом. *Аграрная наука Евро-Северо-Востока*. 2021;22(2):167–187. DOI: <https://doi.org/10.30766/2072-9081.2021.22.2.167-187> EDN: LGYMFT
17. Kuznetsov V. M. Assessment of genetic differentiation of populations by analysis of molecular variance (analytical review). *Agrarnaya nauka Evro-Severo-Vostoka = Agricultural Science Euro-North-East*. 2021;22(2):167–187. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2021.22.2.167-187>

14. Кузнецов В. М. Информационно-энтропийный подход к анализу генетического разнообразия популяций. *Аграрная наука Евро-Северо-Востока*. 2022;23(2):159–173. DOI: <https://doi.org/10.30766/2072-9081.2022.23.2.159-173> EDN: LSSUYZ
Kuznetsov V. M. Information-entropy approach to the analysis of genetic diversity of populations (analytical review). *Agrarnaya nauka Evro-Severo-Vostoka* = *Agricultural Science Euro-North-East*. 2022;23(2):159–173. (In Russ.). DOI: <https://doi.org/10.30766/2072-9081.2022.23.2.159-173>
15. Peakall R., Smouse P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 2006;6(1):288–295. DOI: <https://doi.org/10.1111/j.1471-8286.2005.01155.x>
16. Peakall R., Smouse P. E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*. 2012;28(19):2537–2539. DOI: <https://doi.org/10.1093/bioinformatics/bts460>
17. Smouse P. E., Whitehead M., Peakall R. An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. *Molecular Ecology Resources*. 2015;15(6):1375–1384. DOI: <https://doi.org/10.1111/1755-0998.12422>
18. Hammer Ø., Harper D. A. T., Ryan P. D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*. 2001;4(1):1–9. URL: https://palaeo-electronica.org/2001_1/past/past.pdf
19. Camúñez L. E. M., Roca C. F., Tornero R. Guía de KyPlot: Programa de análisis de datos en contexto científico. *Facultat de Física- Universitat de València (UVEG)*. 2008. 33 p.
20. Peres-Neto P. R., Jackson D. A. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*. 2001;129(2):169–178. DOI: <https://doi.org/10.1007/s004420100720>
21. Dray S., Chessel D., Thioulouse J. Procrustean co-inertia analysis for the linking of multivariate datasets. *Écoscience*. 2003;10(1):110–119. DOI: <https://doi.org/10.1080/11956860.2003.11682757>
22. McGraw K. O., Wong S. P. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996;1(1):30–46. DOI: <https://doi.org/10.1037/1082-989X.1.1.30>
23. Koo T. K., Li M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. 2016;15(2):155–163. DOI: <https://doi.org/10.1016/j.jcm.2016.02.012>
24. Шеффе Г. Дисперсионный анализ. М.: Наука. Главная редакция физико-математической литературы, 1980. 512 с. Режим доступа: <https://studizba.com/files/show/djvu/3369-1-sheffe-g--dispersionnyy-analiz.html>
- Sheffe G. Analysis of variance. Moscow: *Nauka. Glavnaya re-daktsiya fiziko-matematicheskoy literatury*, 1980. 512 p. URL: <https://studizba.com/files/show/djvu/3369-1-sheffe-g--dispersionnyy-analiz.html>
25. Гласс Дж., Стэнли Дж. Статистические методы в педагогике. Пер. с англ. Л. И. Хайрусовой, общ. ред. Ю. П. Адлера. Послесл. Ю. П. Адлера и А. Н. Ковалева. М.: изд-во «Прогресс», 1976. 495 с.
- Glass Dzh., Stenli Dzh. Statistical methods in pedagogy. *Per. s angl. L. I. Khayrusovoy, obshch. red. Yu. P. Adlera. Poslesl. Yu. P. Adlera i A. N. Kovaleva*. Moscow: *izd-vo «Progress»*, 1976. 495 p.
26. VanRoon P., Zakizadeh J., Chartier S. Partial least squares tutorial for analyzing neuroimaging data. *The Quantitative Methods for Psychology*. 2014;10(2):200–215. DOI: <https://doi.org/10.20982/tqmp.10.2.p200>
27. Sampson P. D., Streissguth A. P., Barr H. M., Bookstein F. L. Neurobehavioral effects of prenatal alcohol: Part II. Partial least squares analysis. *Neurotoxicology and Teratology*. 1989;11(5):477–491. DOI: [https://doi.org/10.1016/0892-0362\(89\)90025-1](https://doi.org/10.1016/0892-0362(89)90025-1)
28. Rohlf F. J., Corti M. Use of two-block partial least-squares to study covariation in shape. *Systematic Biology*. 2000;49(4):740–753. DOI: <https://doi.org/10.1080/106351500750049806>
29. Fisher R. A. Combining independent tests of significance. *American Statistician*. 1948;2(5):30. DOI: <https://digital.library.adelaide.edu.au/dspace/bitstream/2440/15258/1/224A.pdf>
30. Czerneková V., Kott T., Dudková G., Sztankóová Z., Soldát J. Genetic diversity between seven Central European cattle breeds as revealed by microsatellite analysis. *Czech Journal of Animal Science*. 2006;51(1):1–7. DOI: <https://doi.org/10.17221/3902-CJAS>
31. Харзинова В. Р., Зиновьева Н. А. Паттерн генетического разнообразия у локальных и коммерческих пород свиней на основе анализа микросателлитов. *Вавиловский журнал генетики и селекции*. 2020;24(7):747–754. DOI: <https://doi.org/10.18699/VJ20.669> EDN: BJRYAW
- Kharzinova V. R., Zinovieva N. A. The pattern of genetic diversity of different breeds of pigs based on microsatellite analysis. *Vavilovskiy zhurnal genetiki i selektsii*. = *Vavilov Journal of Genetics and Breeding*. 2020;24(7):747–754. (In Russ.). DOI: <https://doi.org/10.18699/VJ20.669>
32. Hedrick P. W. Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution*. 1999;53(2):313–318. DOI: <https://doi.org/10.1111/j.1558-5646.1999.tb03767.x>
33. Hedrick P. W. A standardized genetic differentiation measure. *Evolution*. 2005;59(8):1633–1638. URL: <https://www.jstor.org/stable/3449070>

Сведения об авторе

✉ Кузнецов Василий Михайлович, доктор с.-х. наук, профессор, зав. лабораторией популяционной генетики в животноводстве, ФГБНУ «Федеральный аграрный научный центр Северо-Востока имени Н. В. Рудницкого», ул. Ленина, д. 166а, г. Киров, Российская Федерация, 610007, e-mail: priemnaya@fanc-sv.ru, ORCID: <https://orcid.org/0000-0002-2219-805X>, e-mail: vm-kuznetsov@mail.ru

Information about the author

✉ Vasily M. Kuznetsov, DSc in Agricultural Science, professor, Head of the Laboratory of Population Genetics in Animal Husbandry, Federal Agricultural Research Center of the North-East named N. V. Rudnitsky, Lenin str., 166a, Kirov, Russian Federation, 610007, e-mail: priemnaya@fanc-sv.ru, ORCID: <https://orcid.org/0000-0002-2219-805X>, e-mail: vm-kuznetsov@mail.ru

✉ – Для контактов / Corresponding author