

М. Г. Смарагдов

Оценка межстадного генетического разнообразия голштинизированного черно-пестрого скота Ленинградской области методом главных компонент

Аннотация. Для оценки генетического разнообразия голштинизированного черно-пестрого скота Ленинградской области использован метод главных компонент (PCA). Для анализа отобраны шесть стад из племенных заводов со средними удоями коров от 8100 кг до 11300 кг молока. Из каждого стада отобрана случайная выборка коров в количестве 45–85 голов. Общая численность коров составила 373 головы. Все коровы были генотипированы чипом ILLUMINA BovineSNP50 v.2 (Illumina Inc. USA) в Ирландии (Weatherbys Co. UK). Редактирование SNPs осуществляли по следующим критериям: минорная частота аллелей MAF < 0.01, доля ошибок при генотипировании SNPs менее 5%, достоверность соответствия генотипов SNPs распределению Харди-Вайнберга ($P < 0.0001$). В результате редактирования осталось 39657 SNPs из 54609. При расчете межстадного генетического разнообразия коров методом главных компонент использовали программу EIGENSOFT. Достоверность полученных данных рассчитывали методом ANOVA. Общая достоверность данных для трех собственных векторов была следующей: первый собственный вектор ($P < 0.58$), второй собственный вектор ($P < 3.3 \times 10^{-16}$), третий собственный вектор ($P < 5.9 \times 10^{-06}$). Следовательно, второй и третий собственные вектора можно использовать для оценки полученных данных. Оказалось, что парные межстадные различия максимальны для стада 2 как для собственного вектора 2, так и для собственного вектора 3. Также с высокой достоверностью отличались стада 3_6 для собственного вектора 2 и 3_5 для собственного вектора 3. Коровы в нескольких парных сочетаниях стад генетически не различались. В целом, достоверность вычисленных данных методом PCA зависит от используемого собственного вектора. Таким образом, метод PCA эффективен при изучении межстадного генетического различия молочного скота и может быть рекомендован к использованию для других сельскохозяйственных животных.

Ключевые слова: PCA; генетическое разнообразие; стадо; молочный скот; голштинская порода

Автор:

Смарагдов Михаил Григорьевич — кандидат биологических наук, старший научный сотрудник лаборатории молекулярной организации генома; «Всероссийский научно-исследовательский институт генетики и разведения сельскохозяйственных животных — филиал Федерального государственного бюджетного научного учреждения «Федеральный научный центр животноводства — ВИЖ имени академика Л. К. Эрнста»; 196601, Россия, г. Санкт-Петербург, п. Тярлево, Московское шоссе, 55 а; e-mail: spbvniigen@mail.ru.

Введение. Породы крупного рогатого скота различаются экстерьером, продуктивностью и фитнесом животных. Их изменчивость обусловлена как средовыми, так и генетическими факторами. При разведении молочного скота селекционеры минимизируют инбридинг и максимизируют межстадные генетические различия, используя родословные быков-производителей и матерей коров. Ленинградская область является лидером в молочном животноводстве России. При голштинизации черно — пестрого скота Ленинградской области генетическое разнообразие коров в стадах изменилось. Важно знать степень генетического разнообразия коров в стадах в настоящее время. С появлением полногеномных SNPs данных стало возможным на молекулярном уровне оценивать межстадную генетическую дивергенцию животных.

Следует отметить, что молекулярные методы позволяют более точно оценить межстадные генетические различия, чем при учете родословных.

Среди нескольких методов оценки генетического разнообразия популяций особое место занимает метод главных компонент (PCA) [1]. Суть метода состоит в сведении данных из многомерного пространства к пространству с меньшим числом измерений при условии сохранения основной информации и минимизации информационного шума. Этот метод широко используют для генетической дифференциации пород крупного рогатого скота [2, 3, 4, 5]. Для полногеномных SNPs данных создана программа EIGENSOFT [1] позволяющая вычислить генетические расстояния между популяциями в шкале заданных главных компонент.

Цель исследования состояла в анализе генетического разнообразия коров из шести стад голштинизированного черно-пестрого скота Ленинградской области методом главных компонент.

Условия, материалы и методы исследования. В Ленинградской области были отобраны шесть племенных заводов со средними удоями коров 8100–11300 кг молока. Из них случайной выборкой отобраны коровы 2011–2013 года рождения в количестве 47–74 головы (всего 373 головы). ДНК из образцов крови выделяли фенол-хлороформным методом. Коровы генотипированы чипом Illumina Bovine50 v.2 (Illumina Inc. USA) в Ирландии (Weatherbys Co. UK). Редактирование SNPs осуществляли программой Plink 1.9 [6] по следующим критериям: минорная частота аллелей (MAF) SNPs не менее 1%; ошибка генотипирования SNPs не более 5%; достоверность соответствия генотипов SNPs равновесию Харди-Вайнберга ($P < 0.0001$). В результате редактирования осталось 39657 SNPs из 54609.

Вычисление генетических различий между стадами осуществляли программой EIGENSOFT 1.6 [1]. Статистическая достоверность данных рассчитывалась ANOVA. Рисунки получали в программной R среде [7].

Результаты и обсуждение. С помощью программы EIGENSOFT были рассчитаны общие достоверности оценки полученных данных для трех собственных векторов: первый вектор ($P < 0.58$),

второй вектор ($P < 3.3e-16$), третий вектор ($P < 5.9e-06$). Соответствующие им парные межстадные достоверности представлены в таблице 1. На рисунке 1 находится изображение, на котором каждая из 367 коров расположена в определенном месте в координатной системе первых двух собственных векторов или, другими словами, первых двух главных компонент. На этом изображении невозможно оценить межстадные генетические различия коров из-за взаимного перекрытия месторасположения коров в системе координат. Для того, чтобы оценить межстадные генетические различия были рассчитаны средние значения для каждого стада и соответствующие им достоверности для парных сравнений стад (табл. 1).

Визуально средние генетические расстояния между коровами в стадах можно наблюдать на рис. 2 и рис. 3. На этих рисунках видно, что стадо 2 находится на большом расстоянии от всех других стад. Соответственно для этого стада получены максимальные достоверности для PC 2 и достаточно высокие достоверности для PC 3 за исключением 2_3 и 2_6 (табл. 1). Значительное генетическое отличие стада 2 от других стад объясняется использованием в нем быков-производителей из Голландии, тогда как в других стадах преимущественно использовались быки — производители из США и Канады. Также с высокой достоверностью отличались стада 3_6 для собственного вектора 2 и 3_5 для собственного вектора 3 (табл. 1). В группе близко расположенных

Таблица 1. Достоверность парных межстадных генетических различий, установленных PCA методом

Стада	Достоверность (P)		
	PC 1	PC 2	PC 3
1_2	0.23	1.5e-14	0.006
1_3	0.95	0.037	0.002
1_4	0.34	0.320	0.682
1_5	0.70	0.428	0.034
1_6	0.72	0.104	0.421
2_3	0.24	3.8e-14	0.502
2_4	0.77	1.4e-09	0.008
2_5	0.13	5.99e-12	9.4e-05
2_6	0.37	2.22e-16	0.056
3_4	0.36	0.481	0.002
3_5	0.65	0.264	1.8e-06
3_6	0.76	6.34e-05	0.058
4_5	0.19	0.780	0.141
4_6	0.54	0.017	0.299
5_6	0.46	0.019	0.011

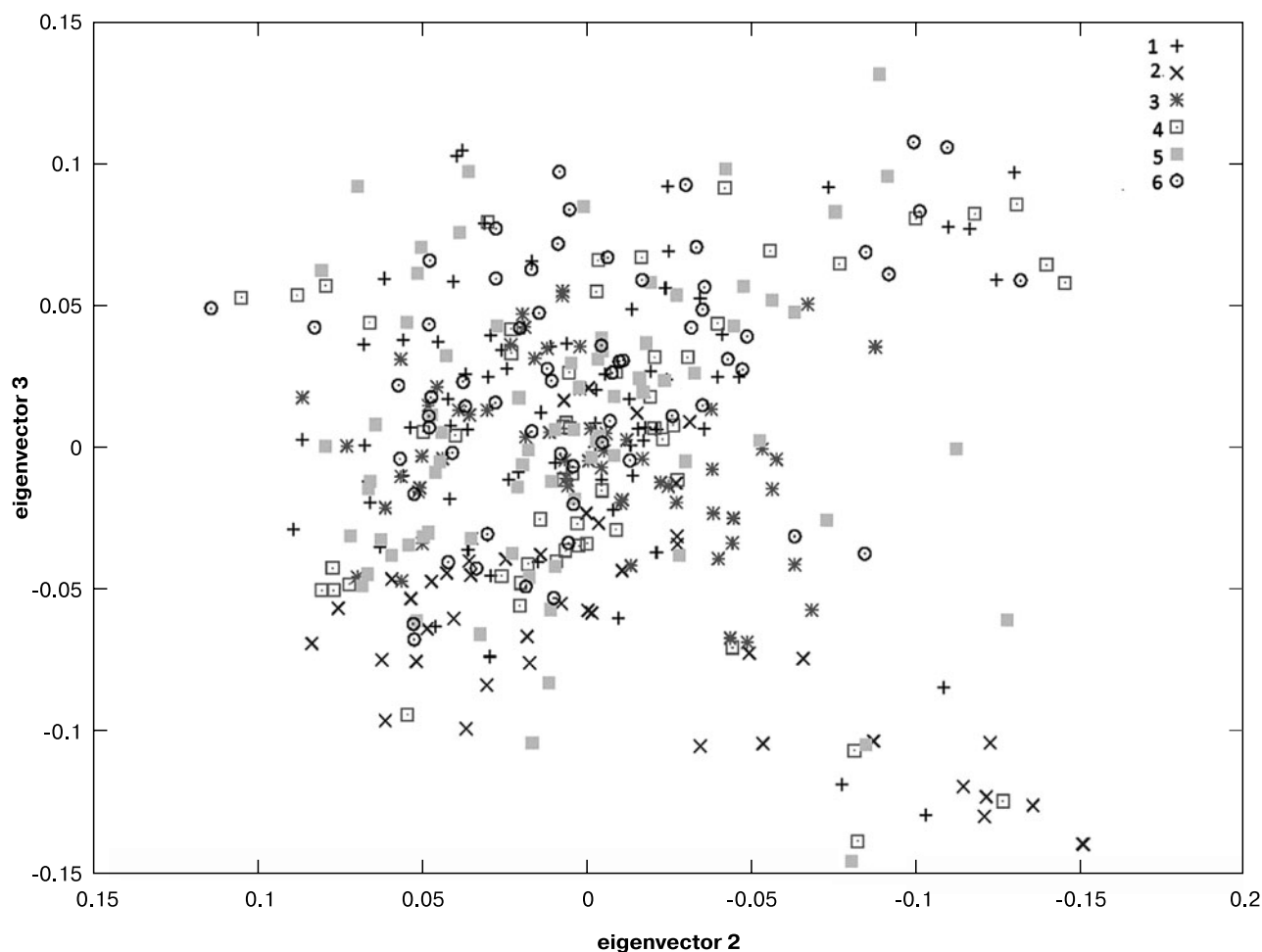


Рис. 1. Взаимное расположение коров из шести стад в координатах РС 2 и РС 3

стад 1–6 только коровы в стадах 1_3, 4_6 и 5_6 различаются с достоверностью $P < 0.05$ для собственного вектора 2 (табл. 1). В той же группе стад коровы в стадах 1_3, 1_5, 3_4 и 5_6 различаются с достоверностью $P < 0.05$ для собственного вектора 3 (табл. 1). Для этих двух собственных векторов достоверные данные совпадают только для пар стад 1_3 и 5_6. Следовательно, межстадные различия коров зависит от использования собственного вектора. Таким образом, РСА метод способен дифференцировать коров из разных стад и его можно использовать для оценки генетического разнообразия коров в популяции только при учете данных для нескольких собственных векторов.

Метод главных компонент часто используют для генетической дифференциации пород крупного

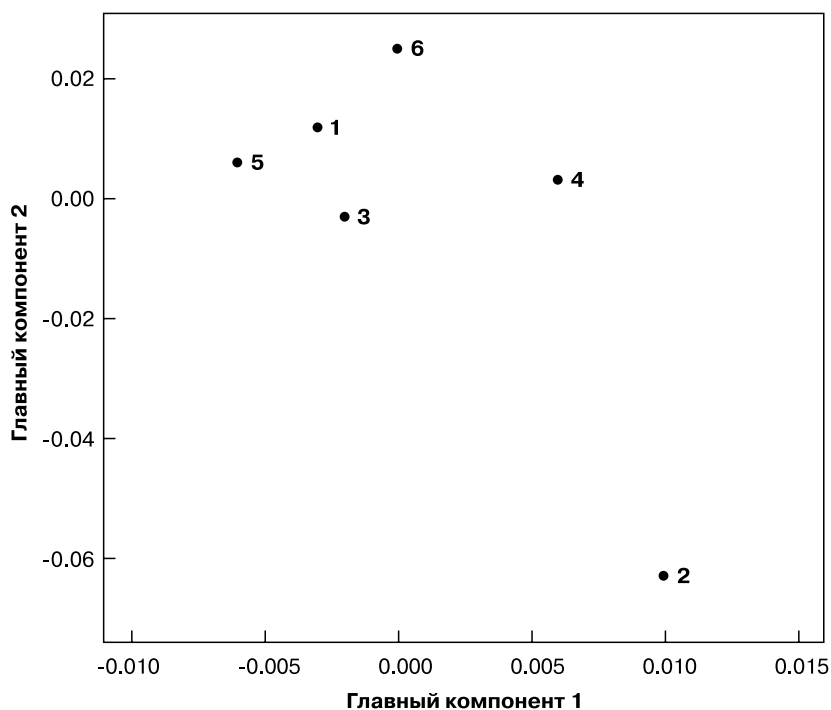


Рис. 2. Расположение средних генетических значений для шести стад в координатах РС 1 и РС 2

рогатого скота. Так, испанские исследователи, исходя из данных PCA анализа, дифференцировали породы мясного скота [2]. Сравнительный PCA анализ популяций, распространенных по всему миру 47 пород крупного рогатого скота, позволил установить сходный паттерн их дифференциации [3]. Установлено, что в Ирландии породы Ангус, Джерси и Герефорд достаточно хорошо генетически различаются PCA методом [4]. Тем же методом незначительные различия обнаружены между быками породы Джерси в США и Новой Зеландии [5]. Коровы породы Джерси из Австралии и США различались ещё меньше, чем быки. Следует также отметить, что межпородные генетические расстояния между животными значительно больше, чем межстадные, и нет необходимости рассчитывать средние значения для каждой породы, так как зоны расположения животных из сравниваемых пород не перекрываются.

Выводы. Метод PCA способен дифференцировать межстадное генетическое разнообразие коров. Достоверность полученных результатов зависит от собственных векторов. Так, все данные для собственного вектора 1 недостоверны, тогда как для собственных векторов 2 и 3 часть данных

высоко достоверна. В целом достоверность межстадных генетических различий зависит от собственного вектора. Из шести проанализированных стад стадо 2 наиболее значительно отличается от других стад из-за использования быков-производителей из Голландии. В целом методом PCA выявлены генетически различающиеся и не различающиеся стада.

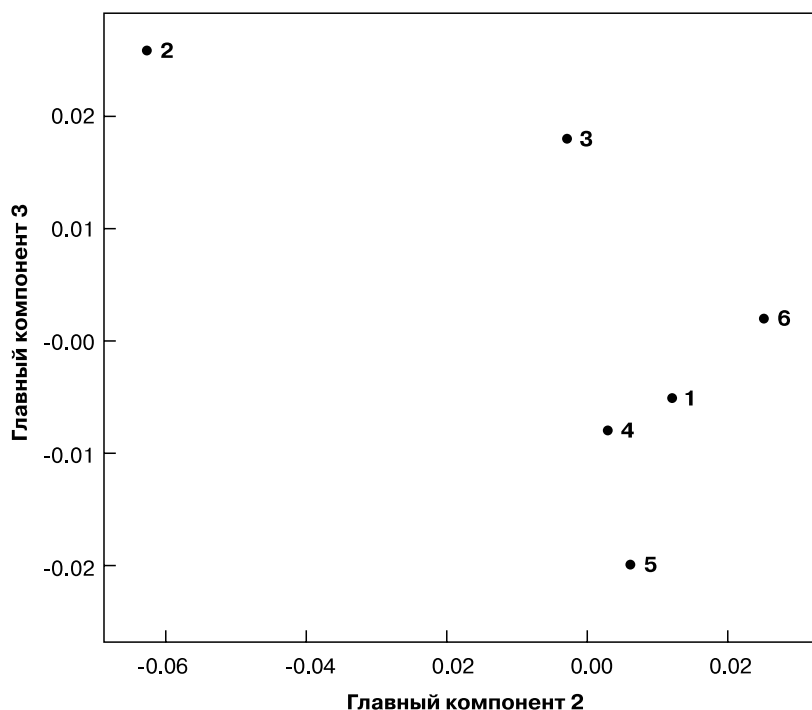


Рис. 3. Расположение средних генетических значений для шести стад в координатах РС 2 и РС 3

Работа выполнена при финансовой поддержке ФАНО (Госзадание № АААА-А18-118021590138-1)

Литература

1. Patterson, N. Population structure and eigen analysis / N. Patterson, A. L. Price, D. Reich // PLoS Genetics. — 2006. — 2:e190.
2. Canas-Alvarez J. J. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high — density SNP chip / J. J. Canas-Alvarez, A. Gonzalez-Rodriguez, S. Munilla et al. // J. Anim. Sci. — 2015. — V. 93. — P. 5164–5174.
3. Gautier M. Insight into the genetic history of French cattle from dense SNP data on 47 worldwide breeds / M. Gautier, D. Laloe, K. Moazami-Goudarzi // PLoS One. — 2010. — V. 5. — e13038.
4. Kelleher M. M. Inference of population structure of purebred dairy and beef cattle using high-density genotype data / M.M. Kelleher, D. P. Berry, J. F. Kearney et al. // Animal. — 2016. — V. 2. — P. 1–9.
5. Howard J. T. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations / J. T. Howard, C. Maltecca, M. Haile-Mariam et al. // BMC Genomics — 2015. — 16 :187.
6. Purcell, S. PLINK: a tool set for whole genome association and population based linkage analyses / S. Purcell // American Journal of Human Genetics. — 2007. — V. 81. — P. 559–575.
7. R Development Core Team R: a language and environment for statistical computing. R foundation for statistical computing. Viena. — 2008. — <http://www.R-project.org>

Smaragdov M.

Estimation between herds genetic diversity of Holsteinized Black-and-White cattle in Leningrad region by Principal Components Analysis

Abstract. *To assess the genetic diversity of Holsteinized Black-and-White cattle in Leningrad Region, Principal Component Analysis (PCA) was used. For the analysis, six herds from Breeding plants with an average cows' milk yield from 8,100 kg to 11,000 kg were selected. From each herd 45–85 cows were selected in random. The total number of cows was 373. All cows were genotyped by the ILLUMINA BovineSNP50 v.2 chip (Illumina Inc. USA) in Ireland (Weatherbys Co. UK). Editing of SNPs was performed according to the following criteria: minor alleles frequency of MAF <0.01, share of errors in genotyping of SNPs less than 5%, reliability of matching of SNPs genotype to Hardy-Weinberg distribution ($P < 0.0001$). As a result of editing there are 41210 SNPs. To calculate between genetic diversity of the cows, EIGENSOFT program was used. The reliability of the obtained data was calculated by ANOVA. The reliability of the data for the three eigenvectors was as follows: the first vector ($P < 0.58$), the second vector ($P < 3.3e-16$), the third vector ($P < 5.9e-06$). Therefore, the second and third eigenvectors can be used to estimate the data obtained. Pairwise between herds differences are maximal for herd 2 both for eigenvector 2 and 3. Highly significant pair of herds 3_6 for eigenvector 2 and 3_5 for eigenvector 3 were obtained. Thus, PCA method is effective in studying between herds genetic diversity of dairy cattle and can be recommended for use on other farm animals.*

Keywords: PCA; genetic diversity; herd; milk cattle; Holstein.

Author:

Smaragdov M. — PhD (Biol. Sci.), senior researcher of laboratory of molecular organization of the genome, Russian research institute of farm animal genetics and breeding, St. Petersburg, Russia, Moscovskoe sh., 55-a; e-mail: mik7252@yandex.ru.

Supported by Federal Agency of Scientific Organizations, №AAAA-A18-118021590138-1

References

1. Patterson, N. Population structure and eigenanalysis / N. Patterson, A. L. Price, D. Reich // PLoS Genetics. — 2006. — 2:e190.
2. Canas-Alvarez J. J. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high — density SNP chip / J. J. Canas-Alvarez, A. Gonzalez-Rodriguez, S. Munilla et al. // J. Anim. Sci. — 2015. — V. 93. — P. 5164–5174.
3. Gautier M. Insight into the genetic history of French cattle from dense SNP data on 47 worldwide breeds / M. Gautier, D. Laloe, K. Moazami-Goudarzi // PLoS One. — 2010. — V. 5. — e13038.
4. Kelleher M. M. Inference of population structure of purebred dairy and beef cattle using high-density genotype data / M.M. Kelleher, D. P. Berry, J. F. Kearney et al. // Animal. — 2016. — V. 2. — P. 1–9.
5. Howard J. T. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations / J. T. Howard, C. Maltecca, M. Haile-Mariam et al. // BMC Genomics — 2015. — 16 :187.
6. Purcell, S. PLINK: a tool set for whole genome association and population based linkage analyses / S. Purcell // American Journal of Human Genetics. — 2007. — V. 81. — P. 559–575.
7. R Development Core Team R: a language and environment for statistical computing. R foundation for statistical computing. Viena. — 2008. — <http://www.R-project.org>