

Berechnung der Populationsstruktur beim Deutschen Warmblut anhand von genomischen Daten

Résumé der Masterarbeit von Wietje Nolte; Universität Rostock

Einleitung und Zielsetzung

Genetische Diversität ist der Grundstein für jegliche Selektion und Zuchtfortschritt. Daher ist detaillierte Kenntnis der gegenwärtigen Populationsstruktur und ihrer Entwicklung fundamental für die erfolgreiche Konzeption und Implementierung von Zuchtprogrammen. Insbesondere der Inzuchtkoeffizient (F) und die effektive Populationsgröße (N_e) geben Einblicke in den Status Quo der genetischen Diversität und die historische Entwicklung einer Rasse.

Das deutsche Warmblutpferd hat in den vergangenen Jahrzehnten eine Nutzungswandlung vom Arbeits- und Militärpferd zum Sport- und Freizeitpartner vollzogen. Die deutsche Pferdezüchtlandschaft ist im Besonderen durch eine Rassevielfalt geprägt, die auf teils jahrhundertelange Tradition zurückblickt. Deutsche Sportpferde sind im internationalen Sport in einer Disziplinvelfalt renommiert und hochehrgefragt. Da die Sportszene heute mehr denn je einem hohen Konkurrenzdruck unterliegt, ist die korrekte Selektion genetisch wertvoller Elterntiere für die nachfolgende Generation von hoher Wichtigkeit.

Während in den vergangenen Jahrhunderten Stammbäume das Hauptinstrument zur strukturellen Untersuchung der Population waren, stehen den Züchtern und Verbänden seit der vollständigen Sequenzierung des Pferdegenoms nun genomische Daten und Untersuchungsmethoden zur Verfügung. Mithilfe kommerziell verfügbarer SNP-Typierungschips sind molekulargenetische Analysen schnell und kostengünstig möglich. Ein Set sogenannter Single Nucleotide Polymorphisms (SNPs), ermöglicht es, das genetische Profil von Individuen zu betrachten und auch intra- bzw. interpopulations-spezifische Untersuchungen durchzuführen.

Die fortlaufende Selektion vorteilhafter Allele oder Allelkombinationen hinterlässt molekulare Muster im Genom¹, sogenannte Selektionssignaturen, die sich z.B. durch ausgedehnte Homozygotieabschnitte (Runs of Homozygosity, ROH) manifestieren. Ein vorteilhaftes Allel, das über mehrere Generationen hinweg künstlich oder natürlich selektiert wird, ist von benachbarten Allelen umgeben, die über Kopplung gemeinsam vererbt werden. Während Demographie und Populationsstruktur gleichmäßig auf das ganze Genom einwirken, beeinflusst positive Selektion nur individuelle Gene und umliegende Regionen^{1,2}. Selektionssignaturen geben folglich Aufschluss über vergangene oder andauernde Selektionsprozesse.

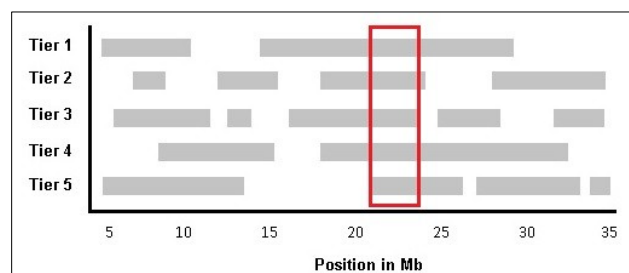


Abbildung 1: Schematische Darstellung eines ROH-Overlaps bei fünf Individuen

Das Ziel dieser Arbeit war es, die Populationsstruktur der vier deutschen Warmblutrassen Trakehner, Holsteiner, Hannoveraner und Oldenburger hinsichtlich ihres Inzuchtgrades und der effektiven Populationsgröße zu analysieren. Zusätzlich wurde mit multiplen Methoden genomweit nach Signaturen positiver Selektion gesucht.

Material und Methoden

Für die genomischen, markerbasierten Untersuchungen standen Daten von 944 Hengsten (*Equus caballus*) aus den fünf deutschen Warmblutrassen Trakehner (N=44), Holsteiner (N=358), Hannoveraner (N=319), Oldenburger (N=221) und Sächsisch-Thüringisches Warmblut (N=2) zur Verfügung, wobei nur die ersten vier Rassen zu intra- und interpopulationsspezifischen Analysen herangezogen wurden. Die Probenahme erfolgte im Rahmen der Körung bzw. Körvorauswahl in den Jahren 2007 und 2008 und inkludiert Tiere aus den Geburtsjahrgängen 1998 und 2001 bis 2006. Demnach handelt es sich bei dem vorliegenden Datensatz um eine vorselektierte Stichprobe. Nach der Genotypisierung mittels Illumina EquineSNP50 BeadChips wurden für Kalkulationen 48.410 autosomale SNPs auf 31 Chromosomen verwendet. Von allen Holsteinern, Hannoveranern und Oldenburgern lagen Abstammungsnachweise mit einer Pedigreetiefe von vier Generationen vor.

Pedigreedaten lieferten mithilfe des Onlinetools Poprep⁴ für jeden Geburtsjahrgang durchschnittliche Inzuchtkoeffizienten, aus denen zusätzlich die Ne abgeleitet wurde. Individuelle F-Werte wurden mit dem R-Paket „pedigree“ (Version 1.4⁵) errechnet.

Mittels SNP-Daten wurden in PLINK 1.09⁶ insgesamt vier verschiedene genomische Inzuchtkoeffizienten nach Vorgehensweisen von Yang, VanRaden und Wright⁷⁻¹⁰ geschätzt. Zusätzlich wurde in Anlehnung an VanRaden⁸ in R 3.2.3 über eine genomische Verwandtschaftsmatrix und abgeleitete Allelfrequenzen ein weiterer individueller F-Wert ermittelt.

Die Berechnung des paarweisen LD zwischen Markern wurde mit vorab in Beagle 4.0 haplotypisierten und imputierten SNP-Daten in Haploview 4.2¹¹ chromosomenweise durchgeführt.

Für die Berechnung der Ne wurden drei grundsätzlich verschiedene Datengrundlagen gewählt: pedigree-, marker- und zensusbasierte Daten. Die demografische Datengrundlage für die populationsspezifische Ne bilden die Anzahlen, die in den Zuchtbüchern Trakehner, Holsteiner, Hannoveraner und Oldenburger eingetragenen Zuchtstuten und -hengste aus dem Jahr 2006¹². Die Berechnung der pedigreebasierten Ne erfolgte auf Basis von vier Generationen mithilfe des Online-Tools PopRep⁴ jahresweise nach Ansätzen von Falconer & Mackay¹³. Für die LD-basierte Ne wurden das paarweise LD zwischen Markern, die genetische Distanz zwischen Loci und die Anzahl betrachteter Tiere nach Hills Modell¹⁴ in Beziehung gesetzt. Es wurden zunächst chromosomenweise Ne-Werte für zurückliegende Generationen generiert, die im Anschluss über alle autosomalen Chromosomen gemittelt wurden.

Zur Aufdeckung von Spuren positiver Selektion wurden Runs of Homozygosity (ROH), der integrated Haplotype Score (iHS), cross population Extended Haplotype Homozygosity (xpEHH) und Tajima's D-Statistik berechnet. Hierbei wurden ROH, iHS und Tajima's D sowohl für alle Tiere gemeinsam kalkuliert, als auch rassespezifisch. Der xpEHH-Wert ergibt sich hingegen aus dem Vergleich je zweier Rassen. Für SNPs von besonderem Interesse wurden Extended Haplotype Homozygosity-Werte (EHH) errechnet. Die Tajima's D-Statistik¹⁵ wurde mit dem Linuxprogramm VCFtools¹⁶ erstellt.

ROHs wurden individuell in PLINK 1.09 ermittelt und auf sogenannte „Overlaps“ gefiltert, d.h. auf konsekutiv homozygote Abschnitte, die ein definierter Mindestprozentsatz von Tieren gemeinsam hat.

Die Größen iHS, EHH und xpEHH wurden mithilfe des R-Pakets „REHH 2.0.0“¹⁷ ermittelt. Dafür ist eine Markerkarte nötig, die angibt, welches Allel ursprünglich und welches neu ist. Für die Erstellung einer solchen Karte wurden SNP-Daten eines Hausesels (*Equus asinus*) herangezogen, die am Pferdereferenzgenom EquCab2.0 ausgerichtet wurden. Dieser Ansatz folgt der Annahme, dass der Hausesel, der sich einen gemeinsamen Vorfahren mit caballischen Pferden teilt, noch eher

ursprüngliche Allele trägt, wohingegen sich beim Pferd durch künstliche Selektion alternative Allele durchgesetzt haben könnten.

Abschließend wurden Regionen mit Selektionssignaturen über die Online-Gendatenbank NCBI (National Center for Biotechnical Information) auf notierte Gene untersucht¹⁸. Die sich daraus ergebende Genliste wurde über das Online-Klassifikationssystem PANTHER¹⁹ auf Zugehörigkeit zu molekularen und biologischen Funktionsgebieten untersucht.

Ergebnisse und Diskussion

Die Interpretation sämtlicher Ergebnisse sollte unter Berücksichtigung des Punktes erfolgen, dass es sich um eine vorselektierte Stichprobe handelt, da nur männliche Tiere, die die potentielle Riege neuer Zuchttiere darstellen, untersucht wurden.

Inzuchtkoeffizienten

Die Pedigreeanalyse ergab in der vorliegenden Arbeit für Holsteiner, Hannoveraner und Oldenburger durchschnittliche F-Werte von 1,9, 0,4 und 0,6 Prozent. Holsteiner weisen damit den höchsten Inzuchtgrad auf. Erwähnenswert ist, dass 11,1 Prozent der untersuchten Tiere dieser Rasse einen Inzuchtkoeffizienten von $F \geq 5$ Prozent haben. Eine frühere an Hannoveranern durchgeführte pedigreebasierte Studie zur Schätzung der genetischen Diversität, ergab einen durchschnittlichen Inzuchtgrad von 1,33 Prozent²⁴. Differenzen können beispielsweise auf die Pedigreetiefe zurückzuführen sein, da es bei vier Generationen zu Unterschätzungen des Inzuchtgrades kommen kann, da Ahnen ohne Abstammungsdaten, sogenannten Foundern, automatisch ein Inzuchtkoeffizienten von $F=0$ zugeschrieben wird. Studien am paternal vererbten Y-Chromosom konnten beim modernen Pferd jedoch nur sechs unterschiedliche väterliche Haplotypen identifizieren²², was die landläufige These stützt, dass das heutige Reitpferd auf wenige Gründerhengste zurückgeht. Beim Englischen Vollblut werden dem Gründerhengst „Darley Arabian“ sogar 95 Prozent der heutigen paternalen Linien zugeschrieben²³. Demnach ist es höchst unwahrscheinlich, dass nur vier Generationen zurückliegende Ahnen nicht ingezüchtet sind. Folglich unterschätzen die hier ermittelten pedigreebasierten F-Werte den realen Inzuchtgrad.

Während pedigreebasierte F-Werte auf einer Skala von 0 bis 1 rangieren, verteilen sich genomische F-Werte im Bereich von -1 bis +1, wobei negative Werte anzeigen, dass Tiere „ausgezüchtet“ sind, d.h. dass sie weniger homozygot sind als erwartet.

Die genomischen Inzuchtkoeffizienten, die über fünf verschiedene Berechnungsmethoden ermittelt wurden, wurden zu einem Durchschnittswert zusammengefasst. Während Holsteiner (0,82%), Hannoveraner (0,48%) und Oldenburger (0,44%) einen Inzuchtkoeffizienten von unter 1 Prozent aufweisen, liegen Trakehner (3,14%) im Mittel deutlich höher. Die fünf verschiedenen Ansätze zur Berechnung der F-Werte liefern innerhalb der Rassen teils deutlich unterschiedliche Ergebnisse. Dass die drei mittels PLINK 1.09 ermittelten F-Werte sich signifikant unterscheiden, wurde bereits an Rindern erwiesen²⁵. Unabhängig von der Methode blieben Hannoveraner, Holsteiner und Oldenburger jedoch im Rassedurchschnitt unter einem kritischen Inzuchtwert von 5 Prozent.

Da sich die Skalen zwischen pedigree- und markerbasierten F-Werten unterscheiden, ist ein direkter Vergleich der Ergebnisse nicht möglich.

Allgemein lässt sich über die ermittelten Inzuchtwerte der vier Rassen sagen, dass sie im Durchschnitt unter dem Wertebereich liegen, der imminente, strikte Maßnahmen zur Reduzierung erfordern würde. Zwecks Erhaltung der genetischen Diversität ist ein Monitoring des populationspezifischen

Inzuchtgrades jedoch prinzipiell ratsam. Erwähnenswert ist, dass in allen Rassen tierindividuelle Maximalwerte zwischen 12 und 16 Prozent festgestellt werden konnten. Es wäre bei der Selektion zukünftiger Zuchttiere überlegenswert einen maximalen F-Wert festzulegen. Im Besonderen bei der Anpaarung sollte auf Verwandtschaftsverhältnisse der Elterntiere geachtet werden.

Effektive Populationsgröße

Die über die Zensusdaten ermittelten N_e -Schätzungen ergaben für Trakehner, Holsteiner, Hannoveraner und Oldenburger jeweils 781, 1.014, 2.132 und 1.066 Tiere. Diese Ergebnisse würden allen vier Rassen hinsichtlich der genetischen Diversität einen sehr guten Status ausweisen, jedoch führt diese Methode gemeinhin zu Überschätzungen. In züchterisch bearbeiteten Reitpferdepopulationen ist das zahlenmäßige Verhältnis von Hengsten und Stuten sehr ungleich; nicht alle registrierten Zuchttiere sind auch aktiv und gerade Vatertiere tragen in sehr differentem Maße zur neuen Generation bei, da der Einsatz eines Zuchthengstes stark von der Nachfrage der Züchter abhängig ist. Die Berechnung der N_e basierend auf Zensusdaten ist schlussendlich wenig aussagekräftig.

Die drei verschiedenen Methoden zur Schätzung der N_e anhand von Abstammungsdaten über PopRep ergaben im Vergleich zu den zensusbasierten Ergebnissen deutlich niedrigere Werte. Die N_e -Werte für den Jahrgang 2006 sind für Holsteiner (49, 57, 104) niedriger als bei Hannoveranern (205, 310, 106) und Oldenburgern (69, 232, 92). Allgemein konnte für alle drei Rassen für den Zeitraum von 1996 bis 2006 eine negative Entwicklung beobachtet werden. Demnach schrumpfen die Rassen in ihrer Diversität.

Ein Abfall der effektiven Populationsgrößen konnte auch mit der genomischen, LD-basierten Betrachtung gezeigt werden (Abbildung 2). Zwar zeigen die Ergebnisse bei Oldenburgern und Hannoveranern seit der 10. zurückliegenden Generation wieder eine Aufwärtstendenz, jedoch sind N_e -Schätzungen für 10 Generationen und jünger wenig zuverlässig. Die großen Abstände zwischen den Loci haben niedrige und sehr variable LD-Werte zur Folge, wodurch Schätzungen ungenau werden²¹. Für weiter zurückliegende Generationen handelt es sich

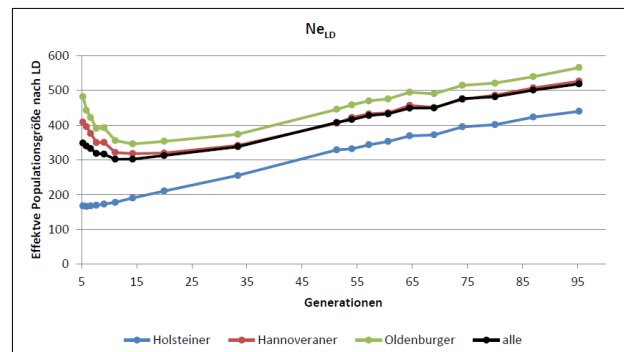


Abbildung 2: LD-basierte effektive Populationsgrößen im historischen Verlauf im Rassevergleich

jedoch um eine bewährte Methode, die u.a. bereits beim Englischen Vollblut angewandt wurde²⁰. Die Ergebnisse für Trakehner finden hier keine Berücksichtigung, da die Stichprobengröße ($N=44$) nicht ausreichend war. So ergaben sich für jüngste Generationen negative N_e -Werte. Für sehr langfristige Betrachtungen können LD-basierte, historische N_e -Werte für Zuchtverbände erkennbar machen, ob eventuelle Strategien zur Diversitätskonservierung erfolgreich waren.

Selektionssignaturen

Bei einer rasseübergreifenden Betrachtung von homozygoten Segmenten von mindestens 10 SNPs Länge, die mindestens 25 Prozent der Tiere gemeinsam haben (ROH-Overlap), konnten 37 solcher Signaturen auf 20 Chromosomen gefunden werden. Hervorzuheben ist eine rassespezifische Signatur von 10 SNPs Länge, die bei Holsteinern auf ECA1 im Bereich 39,7-40,2Mb gefunden werden konnte. Diese Signatur inkludiert den Marker BIEC2-18316, der in einer früheren QTL-Studie mit

Springveranlagung assoziiert werden konnte²⁷. Die springbetonte Zucht der Holsteiner unterstreicht dieses Ergebnis.

Die Identifizierung von ROHs im Genom hängt in ihrer Qualität von der Markerdichte des verwendeten SNP-Chips ab. Höher auflösende SNP-Chips oder Next-Generation-Sequencing-Daten sind zur genaueren Detektion von ROHs besser geeignet. Durch eine deutlich höhere SNP-Zahl lassen sich auch kürzere ROH-Segmente finden²⁵. Zudem kann die korrekte Länge so präziser gemessen werden. Eine frühere Studie zu ROHs an verschiedenen Pferderassen²⁶ brachte rassespezifische Unterschiede, sowohl in Länge als auch Lokalisation, zutage. Zwischen den von Metzger et al. präsentierten ROHs bei Hannoveranern (N = 4) und eigenen Ergebnissen zu dieser Rasse (N = 319) konnten außer auf ECA12 im Bereich 9,1-9,2 Mb keine Kongruenzen gefunden werden.

Die iHS-Methode hat für die Auffindung selektierter Allele, die ein mittleres Frequenzniveau erreicht haben aber noch nicht fixiert sind, die höchste statistische Power¹. Für die Suche nach populationsspezifischen Selektionssignaturen war sie demnach gut geeignet. Der iH-Score wurde erstmalig von Voight²⁸ für die Kartierung von Selektionssignaturen im Humangenom vorgestellt und hat sich seitdem fest etabliert. Bei den vorliegenden Pferderassen konnten beispielsweise rasseübergreifende Signaturen auf ECA1, 4 und 7 gefunden werden (Abbildung 3). Bei einem Vergleich der iH-Scores zwischen den Rassen fielen Hannoveraner und Oldenburger besonders durch die ähnliche Ausgestaltung der Signaturen auf. Auch bei den ROH-Overlaps und bei Tajima's D gab es gemeinsame Signaturen auf ECA3. Anhand der Ergebnisse lässt sich jedoch nicht sagen, ob identische Selektionskriterien kausal für diese Kongruenzen waren oder ob beispielsweise der Einsatz derselben Hengste dazu geführt hat, dass sich dieselben Haplotypen ausgebreitet haben.

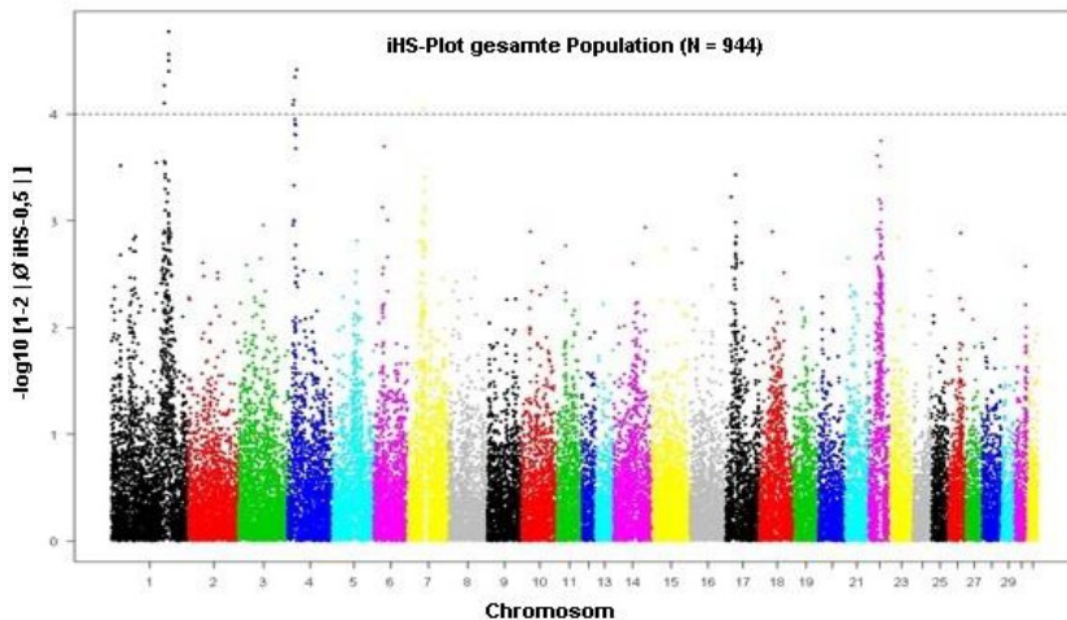


Abbildung 3: genomweiter, rasseübergreifender iHS-Plot zur Detektion von Selektionssignaturen

Die zum Populationsvergleich verwendete Statistik xpEHH hat im Vergleich zu iHS zum Aufdecken kompletterer „selective sweeps“ die doppelte statistische Power und eignet sich sehr gut zur Identifikation selektierter, fixierter Allele in einer Population, die in einer anderen noch polymorph sind¹. Es konnten damit Unterschiede zwischen allen vier Rassen gezeigt werden. Wider Erwarten konnten zwischen Hannoveranern und Oldenburgern, die sich im Vergleich der iHS-Plots noch sehr ähnlich zeigten, die meisten signifikanten Unterschiede gefunden werden.

Die Untersuchung der biologischen funktionalen Hintergründe der Selektionssignaturen, d.h. die Betrachtung der favorisierten Gene in selektierten Regionen, zeigte, dass prozentual die meisten Gene in zelluläre und metabolische Prozesse involviert sind. Anhand der Signaturen und den zugrunde liegenden Genen ist es derzeit nicht möglich zu erkennen, welche davon genau favorisiert wurden und somit ist eine definitive Interpretation der funktionalen Hintergründe so noch nicht möglich. Höher auflösende Chips und ein Abgleich mit phänotypischen Daten könnten hier einen Lösungsansatz bieten.

Literaturverzeichnis

1. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. *Livest Sci.* 2014;166(1):133-143. doi:10.1016/j.livsci.2014.05.003.
2. Biswas S, Akey JM. Genomic insights into positive selection. *Trends Genet.* 2006;22(8):437-446. doi:10.1016/j.tig.2006.06.005.
3. Browning BL. Beagle 4.0. 2015:5. <https://faculty.washington.edu/browning/beagle/beagle.03Mar15.pdf>.
4. Groeneveld E, Westhuizen BD, Maiwashe A, Voordewind F, Ferraz JB. POPREP: a generic report for population management. *Genet Mol Res.* 2009;8(3):1158-1178. doi:10.4238/vol8-3gmr648.
5. Coster A. R-Package "pedigree." 2012. <https://cran.r-project.org/package=pedigree>.
6. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575. doi:10.1086/519795.
7. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011.
8. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91(11):4414-4423. doi:10.3168/jds.2007-0980.
9. Wright S. Genetics of populations. In: *Encyclopaedia Britannica.* ; 1948:10: 111 - A - D - 112.
10. Wright S. Coefficients of inbreeding and relationship. *Am Nat.* 1922;56:330-338.
11. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263-265. doi:10.1093/bioinformatics/bth457.
12. Deutsche Reiterliche Vereinigung (FN). Jahresbericht 2007. 2007:239. doi:10.1553/oehj74s317.
13. Falconer D, Mackay T. *Introduction to Quantitative Genetics.* 4th ed. Essex, U.K.: Longman; 1996.
14. Hill W. Estimation of effective population size from data on linkage disequilibrium. *Genet Res Cambridge.* 1981;38(6428):209-216. doi:10.1017/S0016672300020553.
15. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* 1989;123(3):585-595. doi:PMC1203831.
16. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156-2158. doi:10.1093/bioinformatics/btr330.
17. Gautier AM, Vitalis R, Klassmann A. Package "rehh." 2016:1-19. doi:10.1093/bioinformatics/bts115.
18. Geer LY, Marchler-Bauer A, Geer RC, et al. The NCBI BioSystems database. *Nucleic Acids Res.* 2010;38(Database):D492-D496. doi:10.1093/nar/gkp858.
19. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc.* 2013;8(8):1551-1566. doi:10.1038/nprot.2013.092.
20. Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim Genet.* 2010;41(SUPPL. 2):8-15. doi:10.1111/j.1365-2052.2010.02092.x.
21. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet.* 2012;129(4):257-270. doi:10.1111/j.1439-0388.2012.01003.x.
22. Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, Brem G. Identification of Genetic Variation on the Horse Y Chromosome and the Tracing of Male Founder Lineages in Modern Breeds. Ellegren H, ed. *PLoS One.* 2013;8(4):e60015. doi:10.1371/journal.pone.0060015.
23. Cunningham EP, Dooley JJ, Splan RK, Bradley DG. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim Genet.* 2001;32(6):360-364. doi:10.1046/j.1365-2052.2001.00785.x.
24. Hamann H, Distl O. Genetic variability in Hanoverian warmblood horses using pedigree analysis. *J Anim Sci.* 2008;96(7):1503-1513. doi:10.2527/jas.2007-0382.
25. Ma Y, Wei J, Zhang Q, et al. A genome scan for selection signatures in pigs. *PLoS One.* 2015;10(3):1-18. doi:10.1371/journal.pone.0116850.
26. Metzger J, Karwath M, Tonda R, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics.* 2015;16(1):764. doi:10.1186/s12864-015-1977-3.
27. Schröder W, Klostermann A, Stock KF, Distl O. A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses. *Anim Genet.* 2011;43(2010):392-400. doi:10.1111/j.1365-2052.2011.02265.x.
28. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. *PLoS One.* 2006;4(3):1-14. doi:10.1371/journal.pbio.0040072.