EVOLUTIONARY RELATIONSHIPS AND SIGNATURES OF SELECTION IN CATTLE ESTABLISHED USING GENOME-WIDE SINGLE NUCLEOTIDE POLYMORPHISMS

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

---

by

JARED E. DECKER

Dr. Jeremy F. Taylor, Dissertation Supervisor

MAY 2012

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

EVOLUTIONARY RELATIONSHIPS AND SIGNATURES OF SELECTION IN CATTLE ESTABLISHED USING GENOME-WIDE SINGLE NUCLEOTIDE POLYMORPHISMS

Presented by Jared E. Decker

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Jeremy F. Taylor

_____

Dr. Robert D. Schnabel

_____

Dr. J. Chris Pires

_____

Dr. Gavin C. Conant

_____

Dr. Lori S. Eggert

ACKNOWLEDGEMENTS

I would like to thank Dr. Jeremy Taylor for his extraordinary mentorship and guidance. I am indebted to him for the projects, freedom, and support he provided during my Ph.D. training. Furthermore, I will be privileged to receive Dr. Taylor's continued career support after I graduate. I want to thank Dr. Robert Schnabel for teaching me how to manage large data sets, for handing me the reins of a Linux workstation on which many of these analyses were performed, and for discussing science ideas and funding with me. I want to thank Dr. Chris Pires for all the career advice; I have often referred to him as my "life coach." I have always appreciated the kind words and encouragement from Dr. Lori S. Eggert. Her sentiments have been reassuring at stressful times. I have been privileged to have Dr. Gavin Conant as a teacher and am thankful for the insightful discussions about evolution and population genetics. I fully realize how busy professors are and am truly grateful for the support, instruction, and advice my committee has provided.

I want to thank the many lab mates I have worked with during my time at the University of Missouri. I have enjoyed the relationships that have been formed and I look forward to watching everyone's careers develop. I am very grateful for the extremely long hours spent to generate the data I was fortunate enough to analyze.

I appreciate the numerous collaborators who have provided genotypes and samples. They made this research possible, exciting, and impactful.

I want to thank my family for supporting me through this entire process. As a quantitative geneticist, I realize that I am a successful scientist due to both my genetics and environment. I am thankful for the talents and abilities inherited from my ancestors and the supportive environment provided by my parents, wife, and children. I am glad that I chose the right by coming to the University of Missouri.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Abstract

Although high-throughput single nucleotide polymorphism (SNP) microarray assays were primarily developed for association studies, they are a powerful tool in the study of evolution and population genetics. The applications of SNP genotypes to phylogenomics and population genetics were extended in this dissertation. Using SNP probes designed in a single species, a well-resolved phylogeny of 61 species was produced. Ancestral relationships between cattle breeds were analyzed using parsimony analysis of homozygous genotypes, parsimony analysis of all genotypes, network analysis using F_{ST} estimates, principal component analysis, and admixture analysis. A novel method to identify molecular signatures of selection was deployed. In this method, birth date was analyzed as the dependent variable in a mixed model framework to identify SNP loci which predict birth date. It was shown that predictive loci changed in allele frequency much more than theoretically expected due to genetic drift alone; thus, these loci are in linkage disequilibrium with selected casual variants. In addition to identifying loci under artificial selection, loci putatively responding to natural selection were also identified.

1. INTRODUCTION: UNDERSTANDING ANCESTRAL RELATIONSHIPS AND GENETIC FORCES IN DOMESTIC CATTLE

Nothing in Evolution Makes Sense Except in Light of Population Genetics.

- Michael Lynch

Lynch's quote, a play on the more famous quote of Dobzhansky, captures the spirit and scope of this dissertation, as macro-evolutionary relationships to familial relationships are analyzed in the presented research. The second chapter of this dissertation describes results from a study applying a population level molecular tool, the Illumina BovineSNP50 BeadChip, to study relationships at the macro-evolution level, the phylogeny of species within the infraorder Pecora. Population genetic forces such as selection, drift, migration and admixture are also considered. In addition to providing background and context, this introduction will discuss why it is important to understand ancestral relationships and the genetic forces which have shaped modern cattle. This understanding is important for two central reasons. First, there is an intrinsic link between cattle domestication and human success. Second, this understanding allows us to predict the genetic architecture of economically valuable phenotypic traits.

One of the concerns when utilizing single nucleotide polymorphism markers (SNPs) in evolutionary studies is the bias introduced by the ascertainment of the SNPs (Clark et al. 2005; Rosenblum and Novembre 2007). In the site frequency spectrum, rare variants should be observed more frequently than are common variants. However,

when commercial SNP genotyping assays are developed, SNPs with larger minor allele frequencies are selected to be included on the assay (Matukumalli et al. 2009). This ascertainment of SNPs causes the site frequency spectrum of genotyped SNPs to become uniform with approximately the same proportion of SNPs in each minor allele frequency class. This nonrandom selection of SNPs leads to biases in genetic analyses including the estimation of genetic parameters such as F_{ST} values (Clark et al. 2005). In the phylogenomic analysis of the Pecorans, this bias was avoided by refraining from using analyses which depend on allele frequency estimates. Rather, SNP genotypes were treated as discrete characters and analyzed using a parsimony model. As demonstrated in Chapter 2, this approach proved to be extremely effective leading to a resolved phylogeny with extremely high bootstrap support for almost all branches.

Previous phylogenetic analyses of the Ruminants and Pecorans have yielded unresolved phylogenies (Marcot 2007). Although this is common in phylogenetic analyses (Swenson 2009), the situation was particularly dire for Ruminant systematics. For the rooted relationship between Moschidae, Antilocapridae, Giraffidae, Cervidae, and Bovidae, 11 of the 15 possible phylogenies have been published (Gatesy et al. 1992). Even in recent molecular studies with 3,823 informative characters, many clades, such as Reduncinae, Caprinae, Alcelaphinae, and Antilopinae, remained unresolved. Our analysis was able to resolve these relationships. However, even after the publication of our Pecoran cladogram, questions still remain. Using a different approach, MacEachern et al. proposed different relationships within the Bovini (MacEachern et al. 2009); see Figure 1.1. Among other sources, we point out two

possible causes for this incongruity. First, our data and the data of MacEachern and colleagues were based upon different ascertainment strategies, leading to different possible systematic biases. If there is systematic bias in our data it is most likely to affect the relationship of *Bos taurus* and *Bos indicus* to the rest of the Bos genus. Second, our data had many more informative characters, thus the nodes in our cladogram had much greater support than those in the work of MacEachern and colleagues. As Bibi and Vrba aptly point out, although consensus has been reached for many clades, many important genera and species remain unplaced (Bibi and Vrba 2010). It is likely that increased species and sequence sampling (Wiens 2003; Wiens 1998; Rokas and Carroll 2005), and improved sequence evolution models (Baurain, Brinkmann, and Philippe 2007) will resolve these issues.

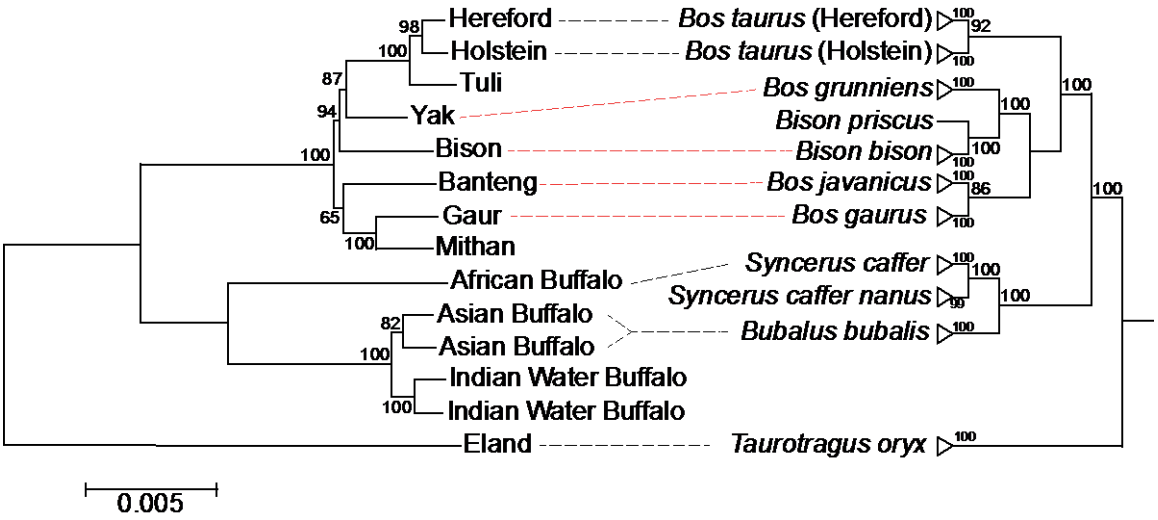

Figure 1.1. Comparison of phylogram from MacEachern et al. on left with cladogram from Decker et al. on right.
Common names are on the left and scientific names on the right. Black lines denote agreement, red lines denote disagreement.

Chapter 2 also highlights the utility of phylogenetic networks in evolutionary and biogeographic studies. In Figure 2.2A one may question the placement of the American Criollo breeds such as Texas Longhorn and Corriente. Likewise, in Figure 2.2B one may question the placement of Jersey cattle. Why are these breeds placed at different locations in the phylogeny when the data are coded differently? The answer lies in the admixed nature of modern cattle populations. Ancestral populations do not generally discretely divide to form new populations; rather, new populations often represent a mixture of multiple ancestral populations. From the previous examples, it is hypothesized that American Criollo cattle received chromosomes from both European and African ancestors and that Jersey cattle received chromosomes from both Iberian and British ancestors. The networks in Figures 2.3 and 2.5 properly place admixed breeds showing their dual ancestry. Though important patterns can be recognized using bifurcating phylogenies, networks provide the most accurate depiction of relationships between admixed populations and hybrid species. It is encouraging that recent work in livestock domestication and population genetics has recognized the utility of networks (Blackburn et al. 2011).

The domestication of plant and animal species enabled the success and growth of human populations (Diamond 2002; Ajmone-Marsan et al. 2010). Contrastingly, domestication events which decreased the genetic diversity among cattle spurred the increase of genetic diversity among humans. To investigate the severity of the bottleneck associated with domestication, researchers have recently used simulations to infer the effective population size (N_e) of animals domesticated in the Fertile Crescent

(Bollongino et al. 2012). Although this research is informative and valuable, these investigators made several questionable assumptions. Most glaring is their assumption of the effective population size of modern cattle. They assumed that the modern cattle N_e was 1,007,170 and denote this as NM_{ef}. They came to this estimate by multiplying the census population counts by 0.806, the proportion of females in herds, and then they divided this number by 10 for an approximation of the ratio of effective population size to census population size. In Chapter 4, we analyze 3,570 Angus animals and estimate an effective population size of contemporary registered Angus to be 94. In 2011, nearly 300,000 Angus cattle were registered (Anon. 2012). With a 5 year generation interval, there may be more than 1,500,000 living registered Angus cattle in the United States. Similar estimates of N_e were found by the Bovine HapMap consortium for 18 other breeds (Gibbs et al. 2009). Thus the approximate census to effective population size ratio of 10 to 1 is simply not appropriate for domestic cattle. The 26 modern mitochondrial sequences used in the Bollongino et al. study also came from animals belonging to only 6 breeds. The maximum effective population size of the breeds analyzed in the Bovine HapMap dataset was 228 (Gibbs et al. 2009). By multiplying 228 by 6 breeds, an NM_{ef} assumption of 1,368 is approximated for the data set analyzed by Bollongino and colleagues. When Bollongino et al. decreased their assumption from 1,007,170 to 100,717 their estimated ND_{ef}, the effective size of cattle at the time of domestication, rose from 80 (95% credible interval of 23 to 452) to 128 (95% credible interval of 44 to 628). Using the analysis published by the Bovine HapMap Consortium, if one assumes that domestication occurred 10,000 years before the

present and the generation interval is 6 years (same assumptions as Bollongino et al., corresponding to 3.22 on the Bovine HapMap Figure 2 x-axis), one would estimate a domestication effective population size of more than 1,500 (Gibbs et al. 2009). The N_e estimate of 1,500 includes diversity introduced by putative introgressions from wild aurochs in Europe, but the effective population size of animals domesticated in the Fertile Crescent was surely much larger than 80. Additionally, the media confused effective and census population sizes when reporting the Bollongino et al. study. Using the ND_{ef} estimate of 80 and reversing the researchers' census size to effective size calculations, we would expect a census size of 992 animals to have been domesticated (80*10/0.806 = 992). Reporting a census population would be much easier for a lay audience to understand. The study of domestication remains an active and contested research area. As more genetic and archeological data are collected, the domestication of cattle will become better understood, although complete resolution may be intractable.

Identifying relationships between populations and individuals has also become important for functional studies. The development of tools which facilitate population-based genome-wide association studies has necessitated the identification of population structure for large cohorts. In an ideal setting, except for affected status, controls would match all of the characteristics of cases, such as race, ancestry, sex, and age. But even mild deviations from this ideal setting can cause inflated test statistics (Devlin, Bacanu, and Roeder 2004). Initially, a method referred to as genomic control was developed to account for this inflation (Devlin and Roeder 1999; Bacanu, Devlin,

and Roeder 2000; Devlin, Bacanu, and Roeder 2004). This method requires a set of neutral markers to estimate the inflation of test statistics, denoted as λ, due to population substructure or cryptic relatedness. Test statistics at all markers are then uniformly adjusted (Price et al. 2006). Another approach to handle population structure is to assign individuals to different clusters (subpopulations) and then test associations within clusters (Pritchard et al. 2000). Initially, identifying clusters was computationally demanding and time consuming, but later implementations such as the computer program ADMIXTURE (Alexander, Novembre, and Lange 2009) reduced this burden. However, identifying the proper number of clusters and handling individuals of mixed ancestry remains a problem for these methods (Price et al. 2006).

If the population structure can be measured or described, the stratification can be fit within the applied statistical model. Geneticists initially used principal component analysis to analyze population allele frequencies to describe population structure (Cavalli-Sforza, Menozzi, and Piazza 1993). Principal component analysis was later applied to the genotypes of individuals and tests for statistical significance of principal components were established (Patterson, Price, and Reich 2006). Significant principal components explicitly describe variation due to population structure for individuals rather than for populations. These principal components are then used to adjust genotypes coded as counts of minor alleles and phenotypes to account for differences in ancestry. This approach removes associations due to confounding population structure; see Figure 1 in Price et al. (2006).

Methods which correct for population stratification generally assume that the sampled individuals are unrelated. This assumption is always violated in many model organisms and agricultural species in which thousands of samples often belong to a single pedigree. In this situation, relationships between all pairs of individuals need to be fit within the statistical model. Building upon the mixed model equations developed by animal breeders decades earlier (Henderson 1963; Quaas and Pollak 1980), statisticians developed two slightly different approaches to account for kinship between samples. Animal breeders were concerned with predicting the genetic merit of individuals and wanted to incorporate DNA markers into the estimation process. It was proposed that dense marker genotypes would be in sufficiently strong linkage disequilibrium with causal variants to predict breeding values for genotyped individuals (Meuwissen, Hayes, and Goddard 2001). In simulations, genomic best linear unbiased prediction (BLUP) and Bayesian methods were found to have the greatest accuracy in predicting genetic merit (Meuwissen, Hayes, and Goddard 2001). This approach soon became known as genomic selection. The development of a high-density genome-wide cattle SNP assay made genomic selection possible (Matukumalli et al. 2009), and in 2008 the first genomic predictions were released (VanRaden et al. 2008). Genomic selection methods treat markers as random effects and predict allele substitution effects (ASEs) at each fit locus. The single locus breeding values of an animal, which are a function of ASEs and allele frequencies, are then summed across all fit loci to estimate the animal's genetic merit. The ASEs at individual loci can be used to map quantitative trait loci; regions containing larger ASEs likely contain causal mutations of large effect.

Unfortunately, because the ASEs are predictions and not parameter estimates, statistical tests of significance are not straightforward (Bolker et al. 2009). However, because all pair-wise relationships are fit in the model, kinship between samples is appropriately modeled.

Rather than predicting the genetic merit of individuals, plant breeders and model organism researchers are interested in identifying the causal genes and mutations underlying the genetic variation in important phenotypes. Thus, SNP effects are usually fit as fixed effects in their mixed model equations, and tests of statistical significance become relatively uncomplicated. Several algorithms, such as TASSELL (Yu et al. 2006), ROADTRIPS (Thornton and McPeek 2010), EMMA (Kang et al. 2008), and EMMAX (Kang et al. 2010), have been implemented to identify genomic regions harboring causal variants while accounting for kinship between genotyped samples.

Mixed model equations have now become more widely used in genome-wide association studies as researchers have realized that most genes have small additive effects. In a news feature (Maher 2008) and a review article (Manolio et al. 2009), authors have made the genomics community aware of the issue coined as "missing heritability." The predicament of "missing heritability" is the inability of markers with statistically significant associations with the analyzed trait to account for a large portion of the trait's heritability. The model trait that revealed this issue is human height. As of 2009, the 40 variants associated with height (which met stringent significance criteria) explained only 5% of the phenotypic variance (Manolio et al. 2009). However, when researchers fit all SNP genotypes simultaneously in a BLUP additive linear model, 45% of

the phenotypic variance was explained (Yang et al. 2010), much closer to the 80% heritability estimated by classical methods and identical by descent haplotype sharing (Manolio et al. 2009; Visscher et al. 2006). It has become clear that much of the genetic variation is not missing, but is due to genes of individually small effect. Thus, we see that classical population genetic theory (Fisher 1918) has been supported by modern genome-wide results (Hill, Goddard, and Visscher 2008).

To handle the issue of small individual additive effects, some researchers have simply lowered the statistical significance threshold required to identify predictive SNPs and have used a Bayesian polygenic framework to model the combined effect of the identified SNPs (Stahl et al. 2012). They found that 65% of the heritability of rheumatoid arthritis, 83-100% of the heritability for celiac disease, 80-100% of the heritability for myocardial infarction and coronary artery disease, and 70 -100% of the heritability for type 2 diabetes could be explained by common SNPs. These heritability estimates were in close agreement with those estimated from genomic linear mixed models. Notably, Stahl et al. also demonstrated that a model with a mixture of common and rare causal variants gave the best fit to the posterior distribution of associated GWAS SNPs, but in this model 536 causal variant loci would be common and 62 causal variant loci would be rare (Stahl et al. 2012). It is important to note that Yang and colleagues and Stahl and colleagues analyzed data sets containing only people of European descent. As has been pointed out, variants that are rare in a species as a whole, can be common within individual populations (Kenny et al. 2012).

What predictions regarding the genetic architecture of economically relevant traits can be made from knowledge of the demographic history of cattle breeds? Of course, genetic architecture varies between traits (Hayes et al. 2010), but population genetic information can be used to predict the genetic architecture of a single trait. For example, population bottlenecks and founder events cause some rare variants in the ancestral population to become common in the new population while many others are lost. Any variants for which the founders are not segregating are lost. Events that have led to decreases in a cattle breed's effective population size, such as domestication, breed formation, and the utilization of artificial insemination, have caused rare variants to either be lost or to increase in frequency. Additionally, new mutations drift to higher frequency or are lost more easily in populations with small effective sizes (Kimura 1983). Animal breeders positively select beneficial mutations causing them to increase in frequency while natural and artificial selection decrease the frequency of detrimental mutations. Furthermore, variants with minor allele frequencies close to 0.5 will have the largest additive genetic variances (Figure 1.2). Conversely, in human populations in which the census and effective population sizes are increasing and selection pressure is decreasing (Lynch 2010), researchers are discovering de novo mutations (i.e., very rare), which putatively cause autism spectrum disorders and epilepsy (Neale et al. 2012; O'Roak et al. 2012; Sanders et al. 2012; Poduri et al. 2012). However, because of drastically different demography, it can be inferred that most of the influential functional variation in production traits within a single cattle breed will be common.

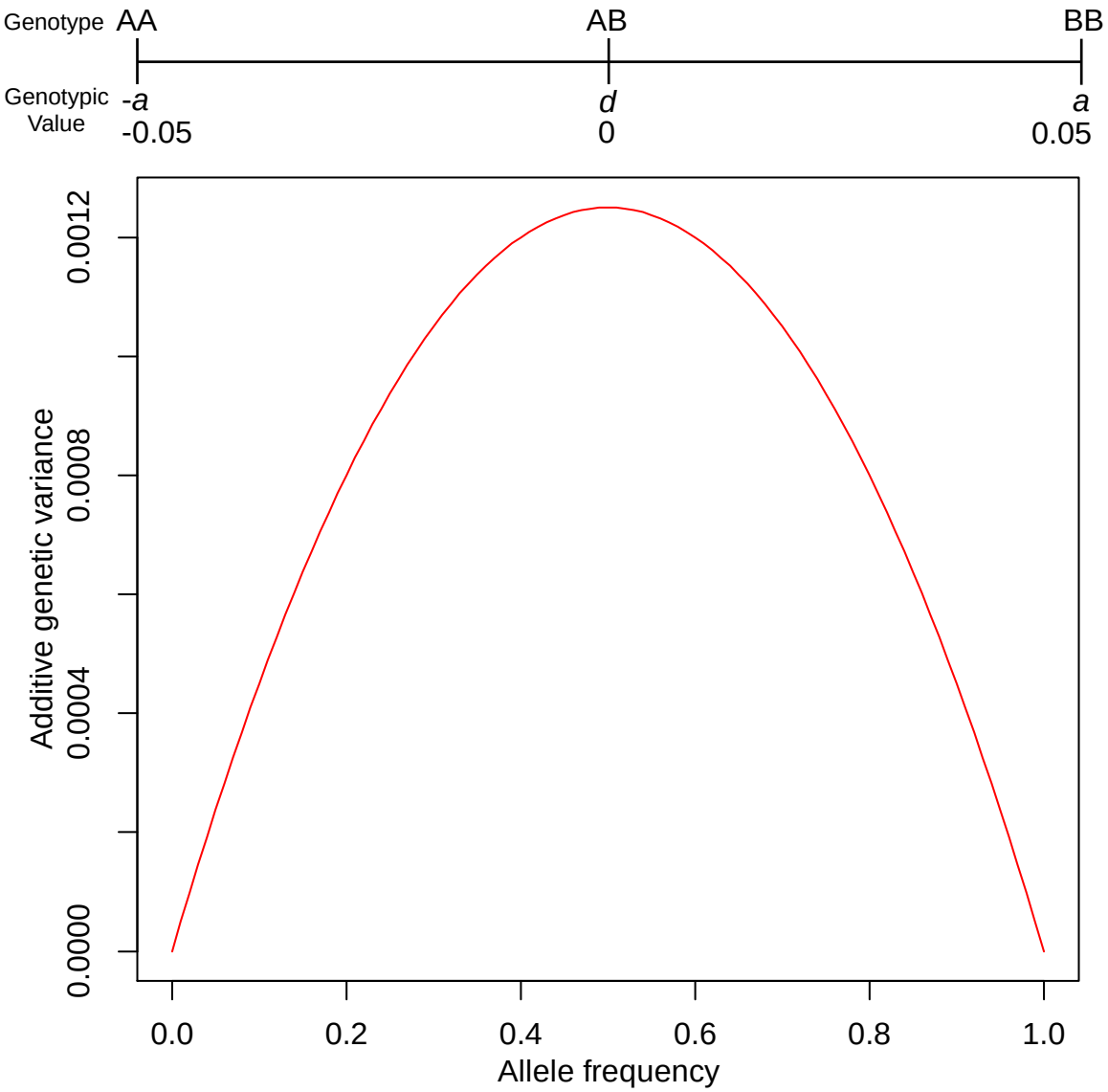

Figure 1.2. Plot of additive genetic variance by allele frequency.
Adapted from (Falconer and Mackay 1996). The additive genetic variance is:

$$2pq[a + d(q-p)]^2$$

Here, the additive effect, a, is set to 0.05 and the dominance deviation, d, is set to 0.

The reductions in effective population sizes that are due to separate domestication events, breed formation, and artificial insemination have caused breeds to diverge. For the 113 *Bos t. taurus* and *Bos t. indicus* breeds sampled in Chapter 4, the F_{ST} values used to create the phylograms and networks ranged from 0.005 to 0.540, with an average of 0.203. Thus, on average, 20% of the genetic diversity among individuals is due to the genetic differentiation among cattle breeds (Holsinger and Weir 2009). For some breeds that share very recent ancestry, such as Angus and Red Angus (F_{ST} = 0.034), across-breed genomic predictions may be feasible if both populations are included in the design of the genomic prediction equations. However, in most instances across-breed genomic predictions will be problematic. Between Holstein and Jersey, two popular dairy breeds, the F_{ST} is 0.157. Even for Angus and Hereford cattle, which are both British breeds and are typically assumed to be similar, the F_{ST} value is 0.143. Thus, about 15% of the genetic diversity is breed specific for these pairs of breeds. Even though these F_{ST} values are biased due to the ascertainment of the SNPs, they reflect the divergence for the data that are used for genomic prediction, because the BovineSNP50 BeadChip is currently the most widely used assay in the design and implementation of genomic selection programs. When prediction equations were trained in Jersey and validated in Holsteins, or vice versa, the predictions had very low accuracies (Hayes, Bowman, Chamberlain, Verbyla, et al. 2009). When a combined reference population was used, accuracies were equivalent to those from the single breed reference populations, even though the reference set was 37% larger (Hayes, Bowman, Chamberlain, Verbyla, et al. 2009). These results do not bode well for across-

breed genomic predictions within taurine cattle and are even less favorable for across-breed genomic predictions among taurine and indicine cattle. However, for variants that segregate in multiple populations, across-breed association studies with dense genotyping will be more effective at identifying regions harboring causal variants, because linkage disequilibrium will extend over shorter distances (Hayes, Bowman, Chamberlain, Verbyla, et al. 2009; Goddard and Hayes 2009; McClure et al. 2012).

Generally, it has been accepted that in cattle most genes are of small effect. Traditional pedigree-based methods for genetic prediction have assumed that genes are of small effect. This assumption is justified by the fact that breeders have successfully used these predictions to substantially change breed means (Hill 2010). Strictly linear genomic BLUP models have been nearly as accurate as nonlinear Bayesian models, providing further support for the conclusion that most genes are of small effect (Hayes, Bowman, Chamberlain, and Goddard 2009). The most effective genomic selection models have fit thousands of SNPs in the prediction equations; if a large number of SNPs are fit then, on average, each must have a small effect (Goddard and Hayes 2009). Furthermore, in addition to the results in Chapter 3, genome-wide scans for causal genes have identified mostly small effects (McClure et al. 2012).

The production of genome-wide SNP microarrays has also fostered the development of statistical methods to identify molecular signatures of selection. Although these methods are quite varied, they can be split into three general classes: those which utilize F_{ST} statistics (Akey et al. 2002; Shriver et al. 2004; Weir et al. 2005), those based on shifts in the site frequency spectrum (Carlson et al. 2005; Kelley et al.

2006), and those which identify haplotype homozygosity (Sabeti et al. 2007; Voight et al. 2006; Wang et al. 2006). Chapter 3 of this dissertation describes a new method to identify selected loci. In this method we analyze birth date as the dependent variable to find SNP markers strongly associated with birth date. Markers strongly associated with birth date have changed in frequency over time and are in linkage disequilibrium with selected causal mutations. Using appropriate statistical models, the method is able to account for population structure and kinship within the sample of genotyped individuals. Selection mapping has a unique property as it identifies genomic regions harboring functional variants even when the selected phenotype is not measured. Thus, selection mapping is complementary to traditional genome-wide association studies of phenotypes.

The research in Chapters 2 and 3 has spawned several questions. To what extent are Iberian cattle admixed between European and African cattle? What is the population structure of cattle in Asia? In 2009, ancient cattle teeth were discovered in the bottom of an old Spanish well in St Augustine, Florida. Are these teeth from ancestors of the modern semi-feral cattle in Florida known as Pineywoods or Florida Crackers? Are the signatures of selection on chromosome 23 observed in Angus cattle common to multiple breeds and are they due to natural selection for disease resistance? Can we identify signatures of selection by performing a genome-wide association study contrasting dairy, dual purpose, and beef cattle breeds? The aims of Chapter 4, where possible, were to answer these questions.

In conclusion, it can be asserted that nothing in bovine genomics makes sense except in the light of population genetics. Sequencing innovations enable the production of a wealth of genomic data, even if some consider this a deluge (Pollack 2011; Editorial 2008). Population genetics will be essential to interpret observations from these genomic data and to identify the processes which produced these observations.

Publication Outline

The following publications are presented as chapters in this dissertation:

2. Decker, Jared E., J. Chris Pires, Gavin C. Conant, Stephanie D. McKay, Michael P. Heaton, Kefei Chen, Alan Cooper, et al. 2009. "Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics." *Proc Natl Acad Sci U S A* 106 (44) (November 3): 18644-18649. doi:10.1073/pnas.0904691106. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2765454&tool=pmcentrez&rendertype=abstract.
3. Decker, Jared E., Daniel A. Vasco, Stephanie D. McKay, Matthew C. McClure, Megan M. Rolf, JaeWoo Kim, Sally L. Northcutt, Stewart Bauck, Brent W. Woodward, Robert D. Schnabel, Jeremy F. Taylor. 2012. A novel analytical method detects response of the Angus (*Bos taurus*) genome to artificial selection on complex traits. *BMC Genomics* (under review).

4. Decker, Jared E., Kefei Chen, Alan Cooper, Carl Halbirt, Allan Roberts, Stephanie D. McKay, Megan M. Rolf, JaeWoo Kim, Antonio Molina, Tad S. Sonstegard, Olivier Hanotte, Anders Götherström, Christopher M. Seabury, Lisa Praharani, Masroor Ellahi Babar, Mehmet Ali Yildiz, Michael P. Heaton, Wansheng Lui, James M. Reecy, Muhammad Saif-Ur-Rehman, Robert D. Schnabel, Jeremy F. Taylor. 2012. Worldwide patterns of exportation, admixture and selection in domesticated cattle. (Under preparation).

2. RESOLVING THE EVOLUTION OF EXTANT AND EXTINCT RUMINANTS WITH HIGH-THROUGHPUT PHYLOGENOMICS

Jared E. Decker[a], J. Chris Pires[b], Gavin C. Conant[a], Stephanie D. McKay[a], Michael P. Heaton[c], Kefei Chen[d], Alan Cooper[d], Johanna Vilkki[e], Christopher M. Seabury[f], Alexandre R. Caetano[g], Gary S. Johnson[h], Rick A. Brenneman[i], Olivier Hanotte[j,k], Lori S. Eggert[b], Pamela Wiener[l], Jong-Joo Kim[m], Kwan Suk Kim[n], Tad S. Sonstegard[o], Curt P. Van Tassell[o], Holly L. Neibergs[p], John C. McEwan[q], Rudiger Brauning[q], Luiz L. Coutinho[r], Masroor E. Babar[s], Gregory A. Wilson[t], Matthew C. McClure[a], Megan M. Rolf[a], JaeWoo Kim[a], Robert D. Schnabel[a], and Jeremy F. Taylor[a,1]

Divisions of [a]Animal Sciences, [b]Biological Sciences, and [h]Department of Veterinary Pathobiology, University of Missouri, Columbia MO 65211

[c]United States Department of Agriculture, Agricultural Research Service, Meat Animal Research Center, Clay Center NE 68933

[d]Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, Adelaide SA 5005, Australia

[e]Agrifood Research Finland MTT, FIN-31600, Jokioinen, Finland

[f]Veterinary Pathobiology, Texas A&M University, College Station TX 77843

[g]Embrapa Recursos Geneticos e Biotecnologia, Brasilia-DF, C.P. 02372, 70770-900, Brasil

[i]Henry Doorly Zoo, Omaha NE 68107

[j]International Livestock Research Institute, P.O. Box 30709, Nairobi 00100, Kenya

[k]School of Biology, University of Nottingham, Nottingham NG7 2RD, United Kingdom

[l]The Roslin Institute, and R(D)SVS, University of Edinburgh, Roslin, Midlothian EH25 9PS, United Kingdom

[m]School of Biotechnology, Yeungnam University, Gyeongsan, Republic of Korea

[n]Department of Animal Science, Chungbuk National University, Cheongju, Republic of Korea

[o]United States Department of Agriculture, Agricultural Research Service, Bovine Functional Genomics Laboratory, Beltsville MD 20705

[p]Department of Animal Sciences, Washington State University, Pullman WA 99164

[q]Animal Genomics, AgResearch, Invermay, PB 50034, Mosgiel, New Zealand

[r]Departamento de Zootecnia, ESALQ-USP, Av. Padua Dias, 11, Piracicaba, SP 13418-900, Brasil

[s]Department of Livestock Production, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

[t]Canadian Wildlife Service, 200 4999 98th Avenue Northwest, Edmonton AB, Canada T6B 2X3

Author contributions: J.E.D. designed research; J.E.D., S.D.M., K.C., A.C., M.C.M., M.M.R., J.W.K., R.D.S., and J.F.T. performed research; M.P.H., K.C., A.C., J.V., C.M.S., A.R.C., G.S.J., R.A.B., O.H., L.S.E., P.W., J.-J.K., K.S.K., T.S.S., C.P.V.T., H.L.N., L.L.C., M.E.B., and

G.A.W. contributed new reagents/analytic tools; J.E.D., J.C.P., G.C.C., J.C.M., R.B., R.D.S., and J.F.T. analyzed data; and J.E.D. and J.F.T. wrote the paper.

Abstract

The Pecorans (higher ruminants) are believed to have rapidly speciated in the Mid-Eocene, resulting in five distinct extant families: Antilocapridae, Giraffidae, Moschidae, Cervidae, and Bovidae. Due to the rapid radiation, the Pecoran phylogeny has proven difficult to resolve, and 11 of the 15 possible rooted phylogenies describing ancestral relationships among the Antilocapridae, Giraffidae, Cervidae, and Bovidae have each been argued as representations of the true phylogeny. Here we demonstrate that a genome-wide single nucleotide polymorphism (SNP) genotyping platform designed for one species can be used to genotype ancient DNA from an extinct species and DNA from species diverged up to 29 million years ago and that the produced genotypes can be used to resolve the phylogeny for this rapidly radiated infraorder. We used a high-throughput assay with 54,693 SNP loci developed for *Bos taurus taurus* to rapidly genotype 678 individuals representing 61 Pecoran species. We produced a highly resolved phylogeny for this diverse group based upon 40,843 genome-wide SNP, which is five times as many informative characters as have previously been analyzed. We also establish a method to amplify and screen genomic information from extinct species, and place *Bison priscus* within the Bovidae. The quality of genotype calls and the placement of samples within a well-supported phylogeny may provide an important test for

validating the fidelity and integrity of ancient samples. Finally, we constructed a phylogenomic network to accurately describe the relationships between 48 cattle breeds and facilitate inferences concerning the history of domestication and breed formation.

Keywords

ancient DNA, Pecorans, domestication

Introduction

The Pecorans are one of the most diverse groups of mammals, ranging in size from the diminutive duiker (adult weight 9–24 kg, shoulder height 0.45–0.51 m) to the giant giraffe (adult weight 500–1,250 kg, shoulder height 4.5–5.8 m). They are indigenous to all continents except South America and Australia (Foss and Prothero 2007) and live in a wide variety of environments. The ruminants are believed to have rapidly radiated in the Mid-Eocene (Foss and Prothero 2007), and due to this rapid radiation, the Pecoran phylogeny has proven difficult to resolve, with 11 of the 15 possible rooted phylogenies describing relationships among the Antilocapridae, Giraffidae, Cervidae, and Bovidae having been argued as representations of the true phylogeny (Gatesy et al. 1992; Marcot 2007). A supermatrix analysis of nucleotide sequence data from 16 genes has resolved some of the nodes within the Pecoran ‘‘Tree of Life (Marcot 2007)’’ and has provided the most strongly supported available phylogeny to which we compare the results of our analyses. However, many of the nodes within this phylogeny either have little support or are completely unresolved

(e.g., the genus Caprinae), and extinct taxa have yet to be phylogenetically placed with confidence (e.g., aurochs). These weakly supported phylogenies have hampered evolutionary studies and conservation efforts for this intriguingly diverse group.

The number and location of prehistoric domestication events for the extinct aurochs (*Bos primigenius*) has also been controversial (Beja-Pereira et al. 2006; Bradley et al. 1996; Gotherstrom et al. 2005; Loftus et al. 1994; Mannen et al. 2004), and the ancestry of many of the derived modern breeds of cattle is unknown. Genome-wide single nucleotide polymorphism (SNP) data captured using high-throughput assays provide a method to perform rapid genomic surveys and have recently been used to resolve the history of human populations (Li et al. 2008; Jakobsson et al. 2008). However, these studies were restricted to a single species, and the remarkable power of these analyses (with 500,000 informative sites) was not fully captured because population relationships depicted using neighbor-joining trees fail to identify multiple ancestral relationships for historically admixed populations. We report an inter-generic, large-scale phylogenomic analysis which applied a genome-wide SNP assay developed for one species to many distantly related species. We also report the application of a genome-wide SNP assay to capture data for ancient DNA samples.

Results

Genotype Fidelity. We have genotyped 16,353 animals representing 61 cattle breeds and 70 species, as divergent from *Bos taurus* as the Savannah elephant (Table 2.S1), with the Illumina BovineSNP50 BeadChip (Van Tassell et al. 2008; Matukumalli et al.

2009) according to Illumina protocols (Steemers et al. 2006). To examine the quality of genotype calls in these outgroup species, we first sequenced the SNP site and flanking regions for rs17871403 in 14 species, with pronghorn the most divergent of the sequenced species (Table 2.S2). This SNP was chosen because it has been well characterized in cattle and is a member of a SNP panel that is widely used for parentage analysis (Heaton et al. 2002). Of the genotypes produced by the BovineSNP50 assay (Illumina) for this SNP in these species, 99.13% were concordant with the sequence when we allowed for genotype ambiguity (i.e., *WW* and *SS*) (see *Methods*). One of the six genotyped North American mountain goats and one of the eight genotyped caribou had discordant BovineSNP50 and sequence-based genotype calls (Table 2.S2). This analysis of a single SNP across multiple species suggests a genotyping error rate for BovineSNP50 loci of only 0.87%.

We next aligned all 40,843 SNP probe sequences, which are 50 bases in length, to the international sheep genomics consortium (www.sheephapmap.org) genome assembly (available at https://isgcdata.agresearch.co.nz/ and in an annotated form at http://www.livestockgenomics.csiro.au/sheep/oar1.0.php) and found that only 26,098 (63.9%) could be uniquely aligned, primarily due to the incomplete status of the assembly. Of these SNP, 829 had an unknown base (*N*) identified at the position of the SNP, and for the remaining 25,269 SNPs, there were 308,518 genotypes called in 17 sheep. Genotype calls were in agreement with the genotype predicted from the respective sequence base for 298,311 genotypes (96.7%). There were 1,834 heterozygous genotypes and 8,373 genotypes that were homozygous for an allele not

predicted by the sequence assembly. This suggests a BovineSNP50 genotyping error rate of between 2.7 and 3.3% in the outgroup species.

Finally, when minor allele frequencies (MAF) averaged over 40,843 SNPs were plotted against average genotype call rates, samples from outgroup species with the lowest call rates had higher than expected MAF (Fig. 2.4). This appears to be indicative of DNA quality issues since, for example, DNA for the *Capra ibex* samples was extracted from irradiated blood samples that had been stored under refrigeration for several years. On removing these samples, there was almost no correlation between MAF and call rate (Fig. 2.4). This indicates that as genetic distance from cattle increases and call rate decreases, spurious heterozygote and alternate homozygote genotype calls rarely arise, indicating support for the quality of these data.

Resolution of the Pecoran Phylogeny. Using genotypes for 40,843 SNPs scored with the BovineSNP50 BeadChip (see *Methods*), we produced a completely bifurcating tree with highly supported nodes for 61 Pecoran species, that contains species that diverged up to 29 million years ago (Fig. 2.1) (Hassanin and Douzery 2003). There were 39,695 parsimony-informative characters using all 678 animals and, remarkably, 21,019 with cattle excluded. Within the Bovidae, only nine nodes had support <100%. We propose 17 relationships and increase the support for 16 previously proposed nodes within the infraorder, when compared to the supermatrix phylogeny of Marcot (Marcot 2007). A striking observation from the phylogeny is that taxonomic classifications of families and subfamilies mirror the topology of the cladogram, since higher taxa form monophyletic

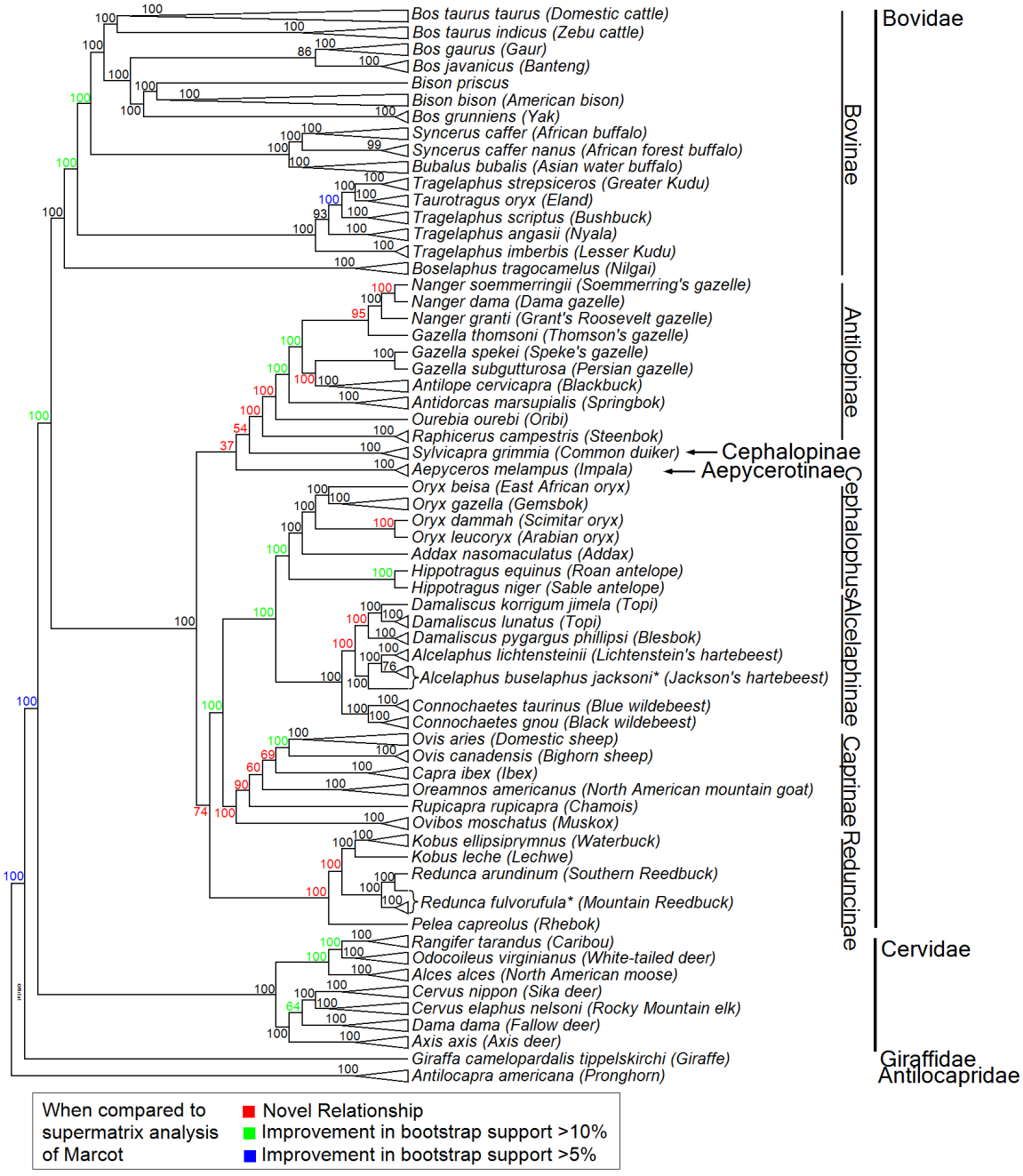

Figure 2.1 Strict consensus cladogram (no branch lengths) of 17 most parsimonious trees based on 40,843 SNP genotypes.

*, Denotes paraphyletic group.

groups. This is an improvement over earlier phylogenies, as previously questionable groupings are now shown to be monophyletic.

Ancient DNA Samples. Currently, PCR-based and non-PCR-based multiple strand displacement amplification (MDA) approaches are used to perform whole genome amplification (Dean et al. 2002; Iwamoto et al. 2007). MDA requires high-quality DNA over 2 Kbp in length and was found to be inefficient for the ancient bison DNA. Consequently, we used a universal linker-based PCR amplification performed with the GenomePlex Whole Genome Amplification kit (Sigma-Aldrich) to amplify the minute amounts of damaged DNA preserved in bone samples from two ancient Russian *Bison priscus* specimens and test whether the Illumina iSelect platform could be used to analyze samples derived from extinct species. The first, sample BS662, was collected from permafrost deposits at Alyoshkina Zaimka, Siberia, and is approximately 20,000 years old (Shapiro et al. 2004). The second, ACAD012, was collected from Sur'ya 5 cave in the Ural Mountains and has been accelerator mass spectrometry radiocarbon dated to 34,460 ± 290 years BP. Due to the low amounts of DNA from the ancient specimens and the short DNA fragment lengths produced in the whole genome amplification of degraded ancient samples, the genotype call rates for these samples were much lower than for modern bison (Table 2.S1). However, when these ancient samples were included in the Bovini phylogeny (Fig. 2.1), BS662 was basal to the modern *Bison bison* clade as expected, but ACAD012 fell within the modern Hereford cattle clade. When we sequenced several overlapping fragments that had been individually amplified from the hypervariable mitochondrial control region of sample ACAD012, we identified variability

within the overlapping regions. This is consistent with the sample having been contaminated with modern DNA or being extremely degraded, as also suggested by our genotype data and consequently the sample was removed from the study. A replicate whole genome amplification (library identification KCMU02) was produced from the *B. priscus* sample used to generate BS662, and when this sample was included in the data set, it was sister to BS662, and both remained sister to modern bison within the phylogeny. However, in the preparation of this library, we avoided the initial DNA fragmentation step within the amplification protocol that appeared to greatly improve the quality and quantity of produced genotypes, as KCMU02 produced a higher genotype call rate (54.9 vs. 45.8%) and far lower heterozygosity (11.5 vs. 39.6%) than did BS662 (Table 2.S3). While only 76.1% of the 12,279 genotypes that were called in both samples were identical, 99.7% of the homozygous genotypes, the only genotype class that has the potential to be phylogenetically informative (see *Methods*), were identical between the replicates.

Relationships Among Cattle Breeds. Phylogenetic relationships were also inferred for 48 cattle breeds (*n* = 372 animals) (Table 2.S1) using parsimony, with most nodes being highly supported (bootstrap values >70%). To accommodate heterozygotes, data were first coded with heterozygotes as polymorphic (noninformative) and then as an independent character state (see *Methods*). When coded as polymorphic, the topology of the cladogram corresponded to the known geographic origins of breeds (Fig. 2.2*A*). Interestingly, however, when heterozygotes were coded as distinct characters, the

topology changed and no longer clearly reflected the biogeography of breed origins (Fig. 2.2*B*).

To further resolve the issue of breed origins, we constructed phylogenetic networks which can reveal conflicting signals in the data (Fig. 2.3 and Fig. 2.5). In Fig. 2.5, *Bos taurus indicus* and *Bos taurus taurus* are distinct groups with long edges between the subspecies. Within *B*. *t*. *taurus*, using the Reynolds et al. (Reynolds, Weir, and Cockerham 1983) distance metric and parsimony cladograms (Fig. 2.2), African taurine cattle were inferred to be more divergent from European cattle than are the Asian *B. t. taurus* breeds, with 100% bootstrap support in cladograms (Fig. 2.2 and Figs. 2.5 and 2.6). Because SNP were almost exclusively discovered from European *B*. *t*. *taurus* samples (Matukumalli et al. 2009), there is a strong ascertainment bias toward SNP common within European *B*. *t*. *taurus* on the BovineSNP50 BeadChip, leading to severe biases in estimates of genetic distance that have prevented us from accurately dating the nodes separating European, African, and Asian cattle (Figs. 2.6 and 2.7). Furthermore, the data were recalcitrant to correction for ascertainment (see *Methods*). The network with individuals at node tips (Fig. 2.3) appears to accurately depict the admixed nature of many populations, for example, the relationship of Belgian Blue to Holsteins and Shorthorns, and Jersey to Iberian and British breeds. The network also reveals pedigree relationships, with sire HO020740 being an interior node to son HO020879.

Discussion

The genotype validation results suggest that BovineSNP50 genotype errors are uncommon, are randomly distributed, and are independent of call rate in the outgroup species. While *Ovis aries* and *B. taurus* are not the most distantly related species surveyed in this study (Fig. 2.1), their most recent common ancestor was at the base of

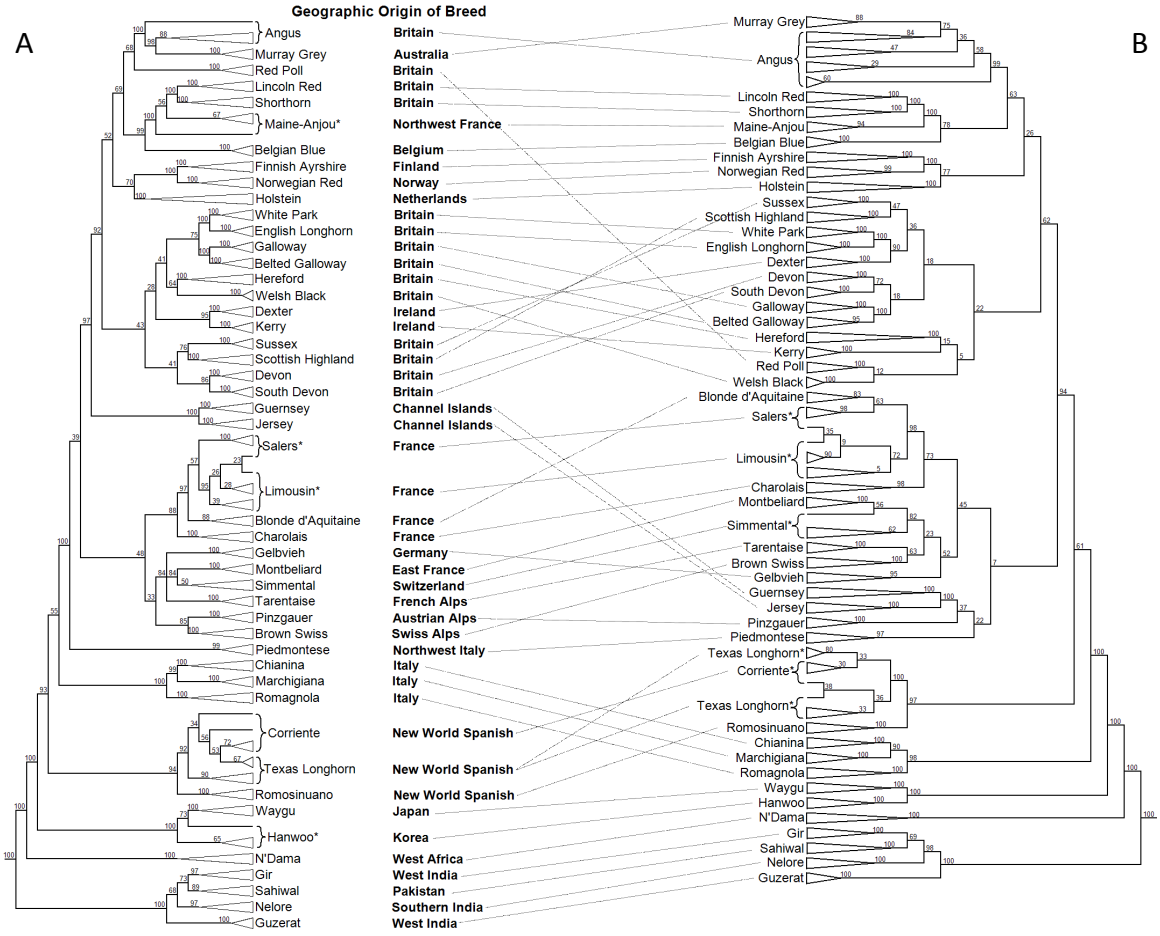

Figure 2.2 Consensus of most parsimonious cladograms of 48 cattle breeds.
(*A*) Most parsimonious cladogram of 48 cattle breeds with heterozygotes coded as polymorphic. Geographic origins were retrieved from the literature (Porter 1991). (*B*) Most parsimonious cladogram of 48 cattle breeds with heterozygotes coded as a third and separate character state. Values at nodes are percent bootstrap support from 1,000 pseudoreplicates. Dotted lines connect clades of a breed between the two cladograms. *B. t. indicus* is represented by the Gir, Sahiwal, Nelore, and Guzerat breeds, with all other breeds being *B. t. taurus* (Table 2.S1). *, Denotes paraphyletic group.

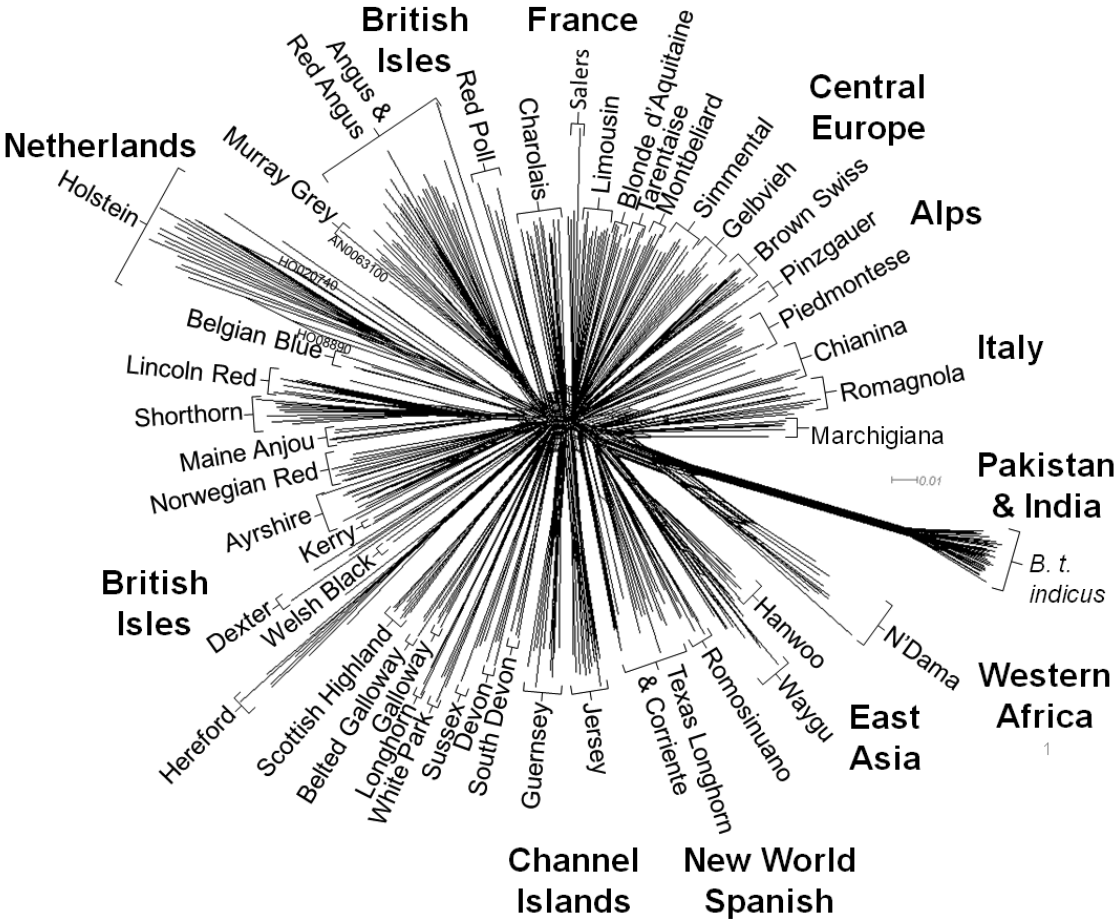

Figure 2.3 Phylogenetic network depicting common ancestry for 372 animals representing 48 cattle breeds.

the Bovidae clade. The use of *O. aries* as a representative for the other species is supported by its 67.2% genotype call rate (Table 2.S1), which was similar to (±7%), or lower than, that for all species and breeds, with the exceptions of Axis deer, Ibex, and Pronghorns, which had call rates <60%.

Despite large amounts of missing data within outgroup species or for the ancient DNA samples, by constructing a larger initial data matrix, which includes more taxa and data than used in previous analyses (Rokas and Carroll 2005; Wiens 1998; Wiens 2003; Heath et al. 2008), we have produced a highly-resolved phylogeny for a rapidly radiated infraorder, which includes extant and extinct species and in which relationships between and within families have been unresolved. Common ancestry can confound studies of speciation and the evolutionary origins and importance of particular traits; the highly resolved phylogeny presented here can control for this issue by allowing the use of phylogenetically independent contrasts (Felsenstein 1985). Further, it facilitates informed conservation efforts, as both ancestral relationships and diversity are clearly defined (Moritz 1995), allowing the identification of species and populations within species to target for preservation. With small data sets, the estimated bootstrap support values can be biased due to the presence of a strong correlation between the samples. Large data sets, such as reported here, accurately estimate the support for internal nodes, since nearly independent pseudosamples can be generated for the construction of bootstrap trees.

We demonstrate that reliable genotypes can be produced from ancient DNA samples, but that more work is needed to optimize amplification and genotyping

protocols. We suspect that the much higher than expected heterozygosities for these samples are due either to template damage or the nonspecific binding of small, possibly exogenous, DNA fragments to the SNP probes. Despite challenges in library optimization, we placed replicate *B*. *priscus* samples as sister to modern bison with strong support and have therefore established the feasibility of high-throughput genotyping of ancient samples. Our results also suggest that the fidelity of the produced genotypes may be assessed by their incorporation into a well-resolved phylogeny and that samples producing unreliable genotypes may be identified and removed from further analysis by this process.

Incongruence between the two breed phylogenies occurred as a result of persistent signatures of admixture, which has been well documented in the histories of several breeds. Thus, the conflicting breed phylogenies oversimplify the complex relationships that exist among populations due to geographic isolation, introgression, migration, and admixture. Networks were effective in revealing both geographic isolation and admixture. There were long branches between *B*. *t*. *taurus* and *B*. *t*. *indicus*, indicating divergence long before domestication. The networks are also consistent with the biogeography of breeds, with European, East Asian, and African taurine cattle forming separate clusters reflecting a predomestication or early postdomestication divergence for these lineages. The West African *B*. *t*. *taurus* N'Dama breed diverges from edges shared with *B*. *t*. *indicus* in Fig. 2.3, and admixture proportions from 0.2–8.6% with African *B*. *t*. *indicus* have previously been estimated for N'Dama populations (MacHugh et al. 1997). Fig. 2.3 also reveals the biogeographical

history of European cattle, which is based upon migrations out of the Fertile Crescent, with domesticated cattle moved sequentially through Turkey, the Balkans, and Italy (Pellecchia et al. 2007), then radiating through Central Europe and France, and finally into the British Isles (Figs. 2.2 and 2.3 and Figs. 2.5 and 2.6). These data also support a second route to the Iberian peninsula by sea from Africa or the Fertile Crescent leading to subsequent admixture with European cattle (Beja-Pereira et al. 2006), as the Spanish breeds found in the New World are basal to German and French breeds (Figs. 2.2 and 2.3). This pattern of geographic dispersal is interrupted only in a few cases in which breed histories document admixture, such as the Belgian Blue, which was formed between 1840 and 1890 by the crossing of local cattle with Friesian and Shorthorn imported from the Netherlands and England, respectively (Porter 1991) (Fig. 2.3). Fig. 2.3 reveals numerous breed relationships, such as the relationship of the Jersey to both Iberian and British breeds (Porter 1991), indicating that many exportations and crossbreeding experiments were performed by early pastoralists. Importantly, this figure reveals that the history of breed formation in cattle has been complicated and has involved bottlenecks, evolution in isolation, coancestry, migration, and admixture.

In all analyses, African cattle were the earliest diverged taurine cattle. Consequently, our results now confine the domestication debate to two distinct hypotheses: (*i*) The occurrence of major domestication events in the Fertile Crescent and Indus Valley (Loftus et al. 1994) were followed by minor captures of aurochs in Africa, East Asia, and Europe (Beja-Pereira et al. 2006; Gotherstrom et al. 2005) or (*ii*) three separate domestication events occurred in the Fertile Crescent, Indus Valley, and

Africa, with a fourth independent domestication in East Asia less likely (Bradley et al. 1996; Mannen et al. 2004).

The largest previous supermatrix analysis of artiodactyls included 3,823 parsimony-informative characters and required several years of data collection (Marcot 2007). We produced 21,019 parsimony-informative characters at a rate of 1,152 samples in 6 days for $100 per sample. Where high-density SNP assays are available for sister species, our approach could affordably be applied to the analysis of other orders and families. Such rapid and inexpensive data generation will transform studies of evolution and domestication through the creation of highly resolved phylogenies, including both extant and extinct species. Genome-wide SNP genotyping assays developed for one species can be used for rapid phylogenomic analysis across a broad taxonomic range and are powerful tools for population and evolutionary studies.

Methods

Whole Genome Amplification of Ancient DNA. Ancient DNA was extracted from fossil bison bone specimens using the standard phenol/chloroform/Amicon Ultra-4 method (Iwamoto et al. 2007). DNA extractions, omniplex library preparations, and PCRs were set-up and performed in a geographically isolated, dedicated ancient DNA facility at the University of Adelaide, Australia. To generate a library of genomic fragments from limited ancient DNA extract, DNA was amplified using the PCR-based GenomePlex Whole Genome Amplification kit (WGA2; Sigma-Aldrich) according to the following protocol: 10 µL DNA were thoroughly mixed with 2 µL library preparation buffer and 1

µL library stabilization solution, and denatured at 95 °C for 2 min. After denaturation, 1 µL library preparation enzyme was added to generate omniplex libraries, followed by a series of incubations at 16 °C for 20 min, 24 °C for 20 min, 37 °C for 20 min, and 75 °C for 5 min in a thermal cycler (Corbett Life Science). The omniplex libraries were next amplified using a limited number of genomic amplification cycles. PCR amplification was conducted in a 75-µL reaction volume containing 14 µL omniplex library, 7.5 µL amplification master mix, 48.5 µL nuclease-free water, and 5 µL WGA DNA polymerase. The PCR amplification conditions were initial denaturation at 95 °C for 3 min, followed by 15 cycles of 94 °C for 15 s and 65 °C for 5 min. GenomePlex-amplified ancient DNA products were finally purified using the GenElute PCR Clean-Up kit (Sigma-Aldrich). Ancient DNA libraries were verified by PCR amplification and sequencing of the hypervariable mtDNA control region before analysis with the BovineSNP50 BeadChip (Illumina). A second amplification, labeled KCMU02, of the sample that produced BS662 was constructed using the same protocol as above, except the genomic fragmentation step within the WGA2 protocol was omitted.

Sample Selection. Table 2.S1 shows the numbers of animals genotyped from each species or cattle breed. In taxa or breeds where <10 animals were genotyped, all animals were sampled. If >10 animals were genotyped, animals with the highest genotype call rates and earliest birth dates were selected. When pedigree information was available, closely related animals were avoided, except in Angus and Holstein where 10 old animals (born in the 1950s, 1960s, and 1970s) and 10 recently born animals (born in the late 1990s and 2000s) were selected. When more than 50 animals within a breed

had call rates of at least 98% and no pedigree information was available, 10 animals were sampled at random. Samples belonging to recently formed crossbred breeds were removed from the analysis, as these samples distort parsimony phylogenies. Genotypes for the two ancient Bison samples were included despite their much lower genotype call rates, which were expected due to DNA degradation and fragmentation, and the use of whole genome amplification, which affect the fidelity of the Infinium assay. The provenance of all samples included in the analyses is provided in Table 2.4.

SNP Selection. The BovineSNP50 BeadChip (Illumina) consists of SNP primarily discovered by the sequencing of reduced representation libraries (Van Tassell et al. 2008), the alignment of random shotgun reads from six cattle breeds to the Hereford assembly, or from the draft assembly of the bovine genome (Matukumalli et al. 2009). To improve genotype quality for *B*. *t*. *indicus* and the outgroup species, we manually adjusted genotype call clusters in Illumina BeadStudio to improve genotype calls. Where pedigree information was available, such as in *O*. *aries* and *B*. *bison*, the rate of misinheritances was minimized. A set of 40,843 SNP was selected from the 54,693 loci queried by the assay. Loci selected for analysis were all located on autosomes, had a call rate of at least 80% in 36 (75%) *B*. *t*. *taurus* breeds, and were not monomorphic in all breeds. This strategy was effective in selecting informative SNP with few genotype errors (Table 2.S5). Data are available at http://animalsciences.missouri.edu/animalgenomics/publications/php.

Genotype Calls in Outgroup Species. Almost 96% of the beads on the BovineSNP50 BeadChip query Infinium II SNP, in which adenine and thymine share a fluorescent

probe and guanine and cytosine share a different fluorescent probe. For samples in which all four bases are present at a single locus, *AA*, *AT*, and *TT* genotypes produce indistinguishable fluorescence intensities, as do *GG*, *GC*, and *CC*. Thus, *A*/*T* or *C*/*G* SNP discovered in *B*. *t*. *taurus* were limited in the assay design (1.8 and 2.2%, respectively, and use Infinium I chemistry). However, in species diverged from *B*. *t*. *taurus* where all four bases could be present, genotypes are *WW* (*W* is the IUPAC code for *A* or *T* bases) for one homozygote class, *SS* (*S* is the IUPAC code for *G or C* bases) for the alternate homozygote, and *NN* (ambiguous) for the heterozygote class. This ambiguity is evident when sequences and genotypes for outgroup species were compared (Table 2.S2). The *WW* and *SS* genotypes were identified in BeadStudio as *AA* and *BB* genotype calls.

Phylogenetic Analysis. Most parsimonious trees were inferred from the genotypes using TNT version 1.1 (Goloboff, Farris, and Nixon 2008). In the analyses involving the outgroup species, phylogenetic signal was obtained only from the homozygous genotypes, and *AA* homozygotes were coded as ‘‘0,’’ *BB* homozygotes were coded as ‘‘1,’’ heterozygotes were coded as a polymorphic character state (i.e., ‘‘[0,1]’’), and missing genotypes were coded as ‘‘?.’’ However, in the analyses of the cattle breeds, an additional data set was created in which heterozygotes were identified by a unique character state (i.e., *AA* = 0, *AB* = 1, *BB* = 2). A heuristic search was conducting using the search technology in TNT, and the search level was initially set to 20. Specifically, we used the SPR-TBR algorithm followed by random sectorial searches, constrained sectorial searches, exclusive sectorial searches, and 10 rounds of tree-drifting. The complete search was replicated 20 times, with 10 rounds of tree fusing at the conclusion

of these 20 replicates. A subset of the samples from the tribe Bovini was independently analyzed along with the ancient bison samples to validate the quality of the data generated from these ancient samples. A data set with 714 samples from all taxon groups was first used to construct the most parsimonious trees. After excluding samples with low quality DNA, low bootstrap support, and/or nonsensical placement in the cladogram (i.e., elephant and horse as sister to *B. taurus*), a final data set with 678 samples was used to construct most parsimonious trees. The cladogram was rooted with *Antilocapra americana.* Using these 678 samples, bootstrap support was calculated using 1,000 pseudoreplicates, and for expediency, the SPR-TBR heuristic search was used.

Allele frequencies were estimated for 40,843 SNP in 22 breeds (Table 2.S6), and these frequencies were used to estimate pairwise Reynolds distances (Reynolds, Weir, and Cockerham 1983) among the breeds (Fig. 2.6). Several attempts were made to correct estimates of genetic distance for SNP ascertainment bias. First, distances were calculated from haplotype frequencies. Haplotypes were inferred for the autosomes of all genotyped animals in our collection within each breed group (Table 2.S6) using fastPhase (Scheet and Stephens 2006). From these haplotyped samples, haplotypes were extracted for the study animals for 885 nonoverlapping loci, each comprising six SNP for which the intermarker distance was <50 Kbp for contiguous SNP. Haplotype frequencies were estimated for each of the 885 loci within each breed group and were used to estimate Reynolds distances between breeds. Next, we formed weighted distances by averaging individual SNP distances weighted according to the frequency of

unascertained SNP (Gibbs et al. 2009) possessing the MAF observed in each of the two populations. Finally, we also subsampled approximately 3,000 or approximately 8,000 SNP such that the resulting MAF distribution conformed to the unascertained distribution of bovine SNP (Gibbs et al. 2009) in Angus or Holstein, respectively. The subsample size was determined by the severity of underrepresentation of SNP within the MAF range 0.005–0.015 and indicates that ascertainment bias was more severe for Angus than for Holstein. Reynolds and Nei genetic distances corrected for sample size (Table 2.S6) were estimated for each subsample and were averaged across 1,000 bootstrap replicates. Distances were used to construct neighbor-joining and UPGMA trees with Phylip (Felsenstein 1989). None of the approaches taken to correct for ascertainment bias were able to establish a tree in which branch lengths were clock-like. Biases in the allele frequency spectrum differ within *B. t. taurus* breeds (Fig. 2.7) causing the distances between breeds to not be clock-like.

Figures of phylogenies and cladograms were produced in MrEnt3 (A. Zuccon and Zuccon 2008), and phylogenetic networks were constructed using SplitsTree version 4.10 (Huson and Bryant 2006). Distances based upon allele frequencies at 40,843 SNP were used to construct a network of 22 breeds. Due to memory limitations in SplitsTree, genotypes at 14,023 SNP were used to construct a network of 372 individuals belonging to 48 breeds. Default settings in SplitsTree were used to construct the networks.

Acknowledgements.

This project was supported by National Research Initiative (NRI) grant nos. 2006–35616-16697, 2008–35205-18864, and 2008–35205-04687 from the U.S. Department of Agriculture Cooperative State Research, Education, and Extension Service (CSREES), 13321 from the Missouri Life Science Research Board and DP0773602 from the Australian Research Council. J.J.K. and K.W.K. were supported by the Technology Development Program for Agriculture and Forestry, Ministry of Agriculture, Forestry and Fisheries, Republic of Korea. We acknowledge the contribution of DNA samples from UK breed societies and cattle breeders as well as the Rare Breeds Survival Trust. We appreciate the critical review and useful comments of Alejandro Rooney. We thank Oliva Handt for help constructing ancient bison libraries. Technical assistance was provided by David Morrice and Karen Troup (Ark Genomics, The Roslin Institute, Edinburgh, UK). We gratefully acknowledge access to Bovine HapMap Project genotypes.

Supplementary Information

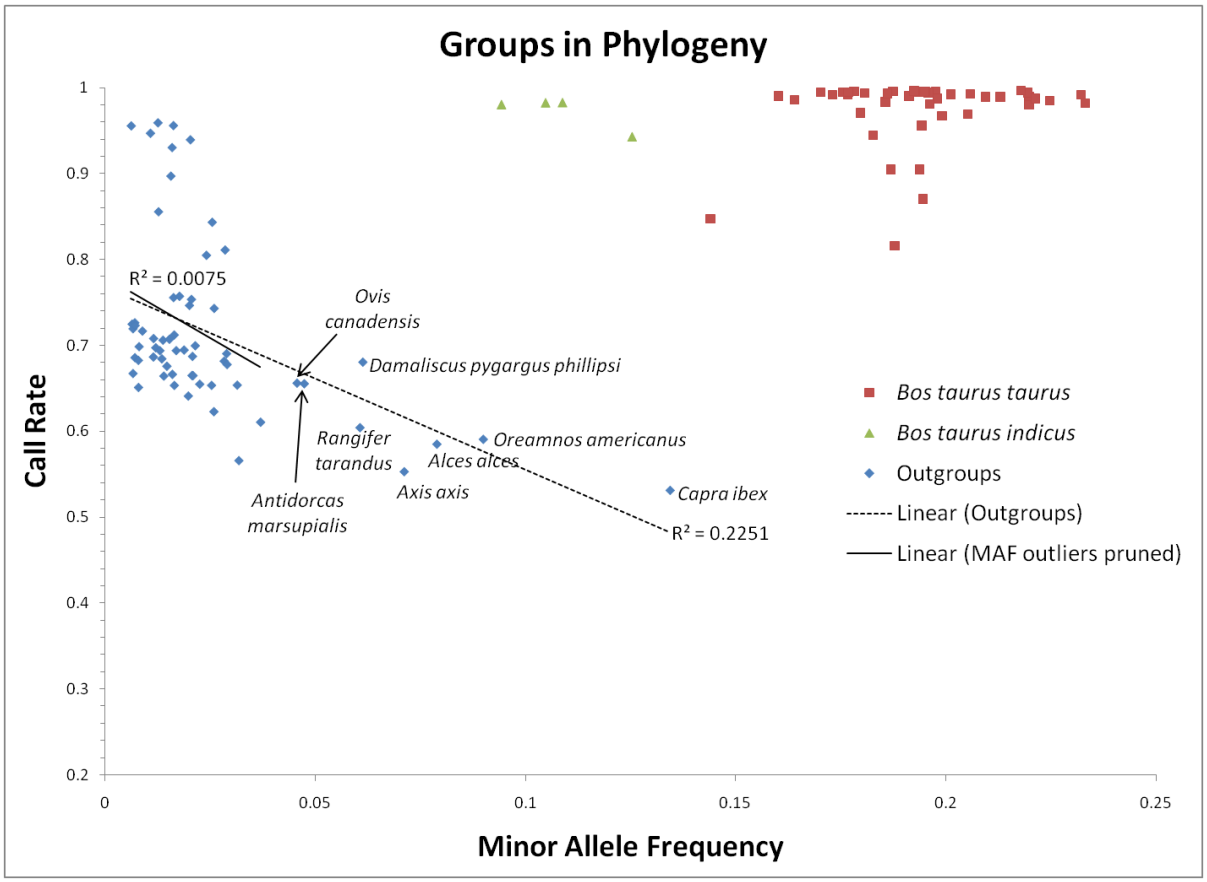

Figure 2.4 Plot of genotype call rate ($CR_{40,843}$) *versus* average minor allele frequency (MAF) averaged over 40,843 SNP for all animals within each group in the phylogeny. A weak linear relationship exists between call rate and MAF for the outgroup species. However, when the 8 outgroups with higher MAF were excluded (labeled with scientific name), almost no relationship exists between call rate and MAF. The lack of a linear relationship among utilized outgroup species supports our conclusion that genotyping errors are few and random, as SNP are assumed to be predominantly monomorphic in these outgroups. This figure also demonstrates the effects of ascertainment bias, with higher MAF in *B. t. taurus* than in *B. t. indicus* breeds.
Fig. S1 in publication.

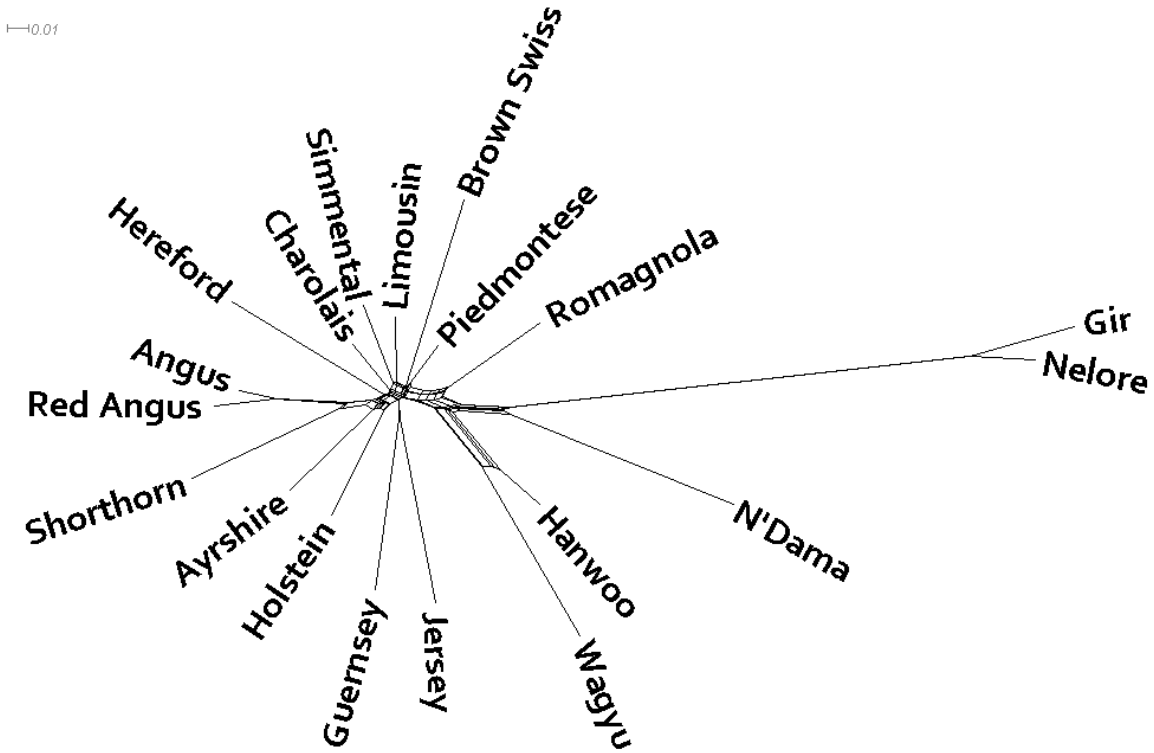

Figure 2.5 Network of 22 breeds using Reynolds genetic distances (Reynolds, Weir, and Cockerham 1983) **estimated from 40,843 SNP for 5,813 animals (Table 2.S5).**
Fig. S2 in publication.

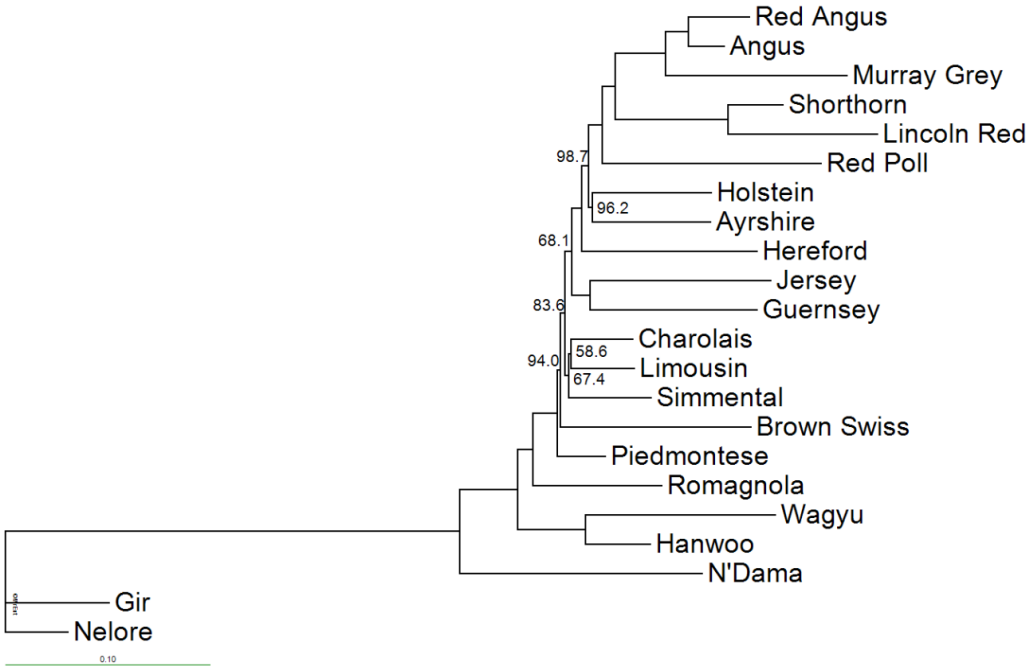

Figure 2.6 Neighbor-joining tree for 22 cattle breeds using Reynolds genetic distances (Reynolds, Weir, and Cockerham 1983) **estimated from allele frequencies for 40,843 SNP. Bootstrap support was estimated from 1,000 pseudo-replicates and was 100% except where indicated.**

Fig. S3 in publication.

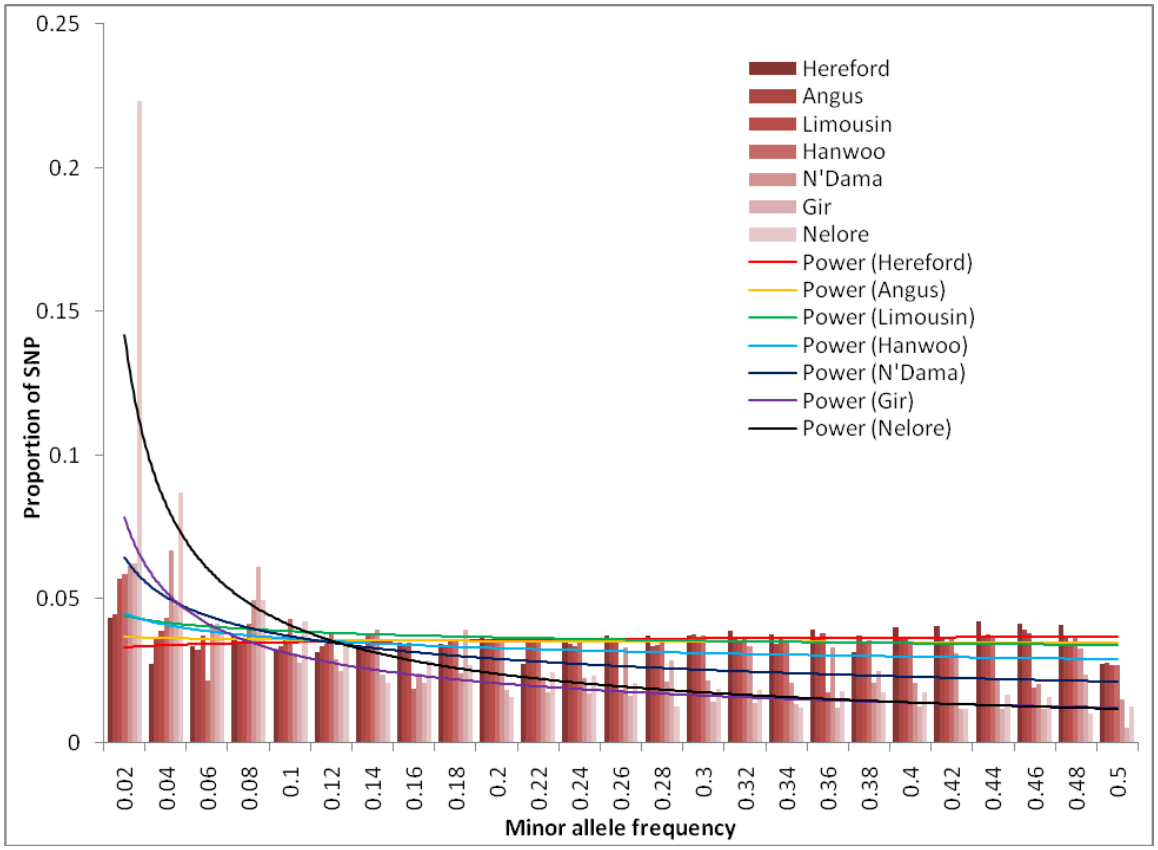

Figure 2.7 Distribution of SNP by MAF illustrates the effects of ascertainment bias. There is a smaller proportion of low MAF SNP in *B. t. taurus* breeds compared to *B. t. indicus* breeds, and there is a higher proportion of high MAF SNP in *B. t. taurus* breeds compared to *B. t. indicus* breeds. Furthermore, there is variation among the frequency spectra within *B. t. taurus* breeds, with Hereford (the sequenced breed) possessing the greatest bias towards high MAF. Trend lines are power functions of the form: SNP proportion = αMAF^{β}.
Fig. S4 in publication.

Table 2.1. List of species and cattle breeds showing the number of genotyped individuals, genotype call rate (total genotypes as a percentage of total possible genotypes) for all animals within a group averaged across all 54,693 SNP ($CR_{54,693}$), the number of animals included in the phylogeny, and genotype call rate for all animals included in the phylogeny within a group averaged across the 40,843 analysed SNP ($CR_{40,843}$).

The two subspecies of bison, *Bison bison* (Plains bison) and *Bison bison athabascae* (Wood bison) were not reciprocally monophyletic and were combined into one *Bison bison* clade in the phylogeny. Species shaded in grey were excluded from the final phylogeny (see text).

Family	Subfamily	Scientific Name	Common Name/ Breed	Number Genotyped (%$CR_{54,693}$)	Number in Phylogeny (%$CR_{40,843}$)
Antilocapridae	Antilocapridae	*Antilocapra americana*	Pronghorn	8 (56.2)	8 (57.3)
Bovidae	Aepycerotinae	*Aepyceros melampus*	Impala	6 (67.8)	5 (69.5)
Bovidae	Alcelaphinae	*Alcelaphus buselaphus jacksoni*	Jackson's hartebeest	3 (70.8)	3 (72.2)
Bovidae	Alcelaphinae	*Alcelaphus lichtensteinii*	Lichtenstein's hartebeest	2 (69.3)	2 (70.7)
Bovidae	Alcelaphinae	*Connochaetes gnou*	Black wildebeest	3 (70.3)	3 (71.6)
Bovidae	Alcelaphinae	*Connochaetes taurinus*	Blue wildebeest	4 (68.6)	4 (70.0)
Bovidae	Alcelaphinae	*Damaliscus korrigum jimela*	Topi	1 (71.2)	1 (72.7)
Bovidae	Alcelaphinae	*Damaliscus lunatus*	Topi	2 (70.2)	2 (71.6)
Bovidae	Alcelaphinae	*Damaliscus pygargus phillipsi*	Blesbok	5 (67.6)	4 (71.0)
Bovidae	Antilopinae	*Antidorcas marsupialis*	Springbok	6 (65.1)	5 (68.0)
Bovidae	Antilopinae	*Antilope cervicapra*	Blackbuck	9 (68.3)	9 (69.6)
Bovidae	Antilopinae	*Gazella dorcas*	Dorcas gazelle	1 (74.6)	0
Bovidae	Antilopinae	*Gazella spekei*	Speke's gazelle	1 (67.8)	1 (68.9)
Bovidae	Antilopinae	*Gazella subgutturosa*	Persian gazelle	1 (67.1)	1 (68.6)
Bovidae	Antilopinae	*Gazella thomsoni*	Thomson's gazelle	1 (66.2)	1 (67.6)
Bovidae	Antilopinae	*Litocranius walleri*	Gerenuk	1 (28.5)	0
Bovidae	Antilopinae	*Nanger dama*	Dama gazelle	1 (68.1)	1 (69.4)
Bovidae	Antilopinae	*Nanger granti*	Grant's gazelle	2 (51.5)	0
Bovidae	Antilopinae	*Nanger granti*	Grant's Roosevelt gazelle	1 (68.2)	1 (69.3)
Bovidae	Antilopinae	*Nanger soemmerringii*	Soemmerring's gazelle	1 (66.3)	1 (67.4)
Bovidae	Antilopinae	*Oreotragus oreotragus*	Klipspringer	1 (67.1)	0
Bovidae	Antilopinae	*Ourebia ourebi*	Oribi	1 (66.0)	1 (67.4)

Bovidae	Antilopinae	*Raphicerus campestris*	Steenbok	2 (66.1)	2 (67.4)
Bovidae	Bovinae	*Bison bison*	Plains Bison	135 (93.8)	25 (94.7)
Bovidae	Bovinae	*Bison Bison athabascae*	Wood Bison	36 (87.9)	30 (94.3)
Bovidae	Bovinae	*Bison sp.*	European Wisent	1 (42.7)	0
Bovidae	Bovinae	*Bison priscus*	Steppe Wisent	1 (44.8)	1 (45.8)
Bovidae	Bovinae	*Bos gaurus*	Gaur	47 (95.1)	10 (97.5)
Bovidae	Bovinae	*Bos grunniens*	Yak	2 (93.4)	2 (96.8)
Bovidae	Bovinae	*Bos javanicus*	Banteng	4 (95.4)	4 (97.2)
Bovidae	Bovinae	*Bos taurus indicus*	Gir	30 (96.0)	10 (99.3)
Bovidae	Bovinae	*Bos taurus indicus*	Guzerat	3 (96.0)	3 (99.2)
Bovidae	Bovinae	*Bos taurus indicus*	Nelore	78 (93.7)	10 (99.4)
Bovidae	Bovinae	*Bos taurus indicus*	Sahiwal	12 (96.2)	10 (99.4)
Bovidae	Bovinae	*Bos taurus taurus*	Angus	6124 (98.3)	20 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Belgian Blue	4 (98.8)	4 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Belted Galloway	4 (98.8)	4 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Blonde d'Aquitaine	5 (98.8)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Brown Swiss	74 (81.8)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Charolais	135 (98.4)	11 (98.8)
Bovidae	Bovinae	*Bos taurus taurus*	Chianina	10 (96.2)	8 (97.0)
Bovidae	Bovinae	*Bos taurus taurus*	Corriente	5 (98.7)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Devon	4 (98.7)	4 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Dexter	4 (96.4)	4 (98.8)
Bovidae	Bovinae	*Bos taurus taurus*	Finnish Ayrshire	444 (98.2)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Galloway	4 (98.5)	4 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Gelbvieh	8 (98.5)	8 (99.8)
Bovidae	Bovinae	*Bos taurus taurus*	Guernsey	23 (97.2)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Hanwoo	48 (96.0)	7 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Hereford	143 (97.7)	10 (96.3)
Bovidae	Bovinae	*Bos taurus taurus*	Holstein	5770 (98.1)	20 (99.7)
Bovidae	Bovinae	*Bos taurus taurus*	Jersey	93 (90.5)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Kerry	3 (98.6)	3 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Limousin	1621 (97.3)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Longhorn	4 (84.1)	3 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Maine Anjou	5 (98.8)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Marchigiana	5 (86.5)	5 (90.2)
Bovidae	Bovinae	*Bos taurus taurus*	Montbeliard	5 (98.6)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Murray Grey	5 (94.9)	5 (96.9)
Bovidae	Bovinae	*Bos taurus taurus*	N'Dama	59 (98.3)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Normande	1 (98.8)	0
Bovidae	Bovinae	*Bos taurus taurus*	Norwegian Red	21 (97.2)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Piedmontese	29 (98.7)	10 (99.9)

Bovidae	Bovinae	*Bos taurus taurus*	Pinzgauer	5 (98.0)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Red Angus	15 (97.2)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Red Poll	5 (97.6)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Romagnola	29 (98.6)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Romosinuano	8 (98.2)	8 (99.6)
Bovidae	Bovinae	*Bos taurus taurus*	Salers	5 (98.8)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Scottish Highland	9 (89.8)	8 (99.0)
Bovidae	Bovinae	*Bos taurus taurus*	Shorthorn/Lincoln Red	108 (97.4)	19 (99.0)
Bovidae	Bovinae	*Bos taurus taurus*	Simmental	777 (97.5)	10 (99.8)
Bovidae	Bovinae	*Bos taurus taurus*	South Devon	4 (93.8)	4 (95.9)
Bovidae	Bovinae	*Bos taurus taurus*	Sussex	4 (98.5)	4 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Tarentaise	5 (98.7)	5 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	Texas Longhorn	32 (98.2)	10 (99.9)
Bovidae	Bovinae	*Bos taurus taurus*	Wagyu	49 (97.9)	10 (99.6)
Bovidae	Bovinae	*Bos taurus taurus*	Welsh Black	2 (98.7)	2 (100.0)
Bovidae	Bovinae	*Bos taurus taurus*	White Park	5 (98.7)	4 (100.0)
Bovidae	Bovinae	*Boselaphus tragocamelus*	Nilgai	8 (74.9)	8 (76.5)
Bovidae	Bovinae	*Bubalus bubalis*	Asian water buffalo	12 (83.9)	10 (86.4)
Bovidae	Bovinae	*Syncerus caffer*	African buffalo	8 (85.1)	8 (86.9)
Bovidae	Bovinae	*Syncerus caffer nanus*	African forest buffalo	3 (79.9)	3 (82.2)
Bovidae	Bovinae	*Taurotragus oryx*	Eland	5 (75.2)	4 (77.0)
Bovidae	Bovinae	*Tragelaphus angasii*	Nyala	6 (74.2)	5 (75.5)
Bovidae	Bovinae	*Tragelaphus imberbis*	Lesser Kudu	2 (65.1)	2 (66.3)
Bovidae	Bovinae	*Tragelaphus scriptus*	Bushbuck	6 (73.8)	5 (76.5)
Bovidae	Bovinae	*Tragelaphus strepsiceros*	Greater Kudu	7 (75.1)	6 (76.6)
Bovidae	Caprinae	*Capra ibex*	Ibex	13 (52.8)	10 (58.9)
Bovidae	Caprinae	*Oreamnos americanus*	North American mountain goat	8 (58.7)	7 (60.6)
Bovidae	Caprinae	*Ovibos moschatus*	Muskox	7 (69.5)	7 (71.0)
Bovidae	Caprinae	*Ovis aries*	Sheep	17 (63.9)	10 (67.2)
Bovidae	Caprinae	*Ovis canadensis*	Bighorn sheep	8 (65.2)	8 (66.6)
Bovidae	Caprinae	*Rupicapra rupicapra*	Chamois	1 (70.4)	1 (72.0)
Bovidae	Cephalophinae	*Sylvicapra grimmia*	Common duiker	3 (69.0)	3 (70.2)
Bovidae	Hippotraginae	*Addax nasomaculatus*	Addax	1 (71.5)	1 (72.9)
Bovidae	Hippotraginae	*Hippotragus equinus*	Roan antelope	1 (69.0)	1 (70.7)
Bovidae	Hippotraginae	*Hippotragus niger*	Sable antelope	1 (72.2)	1 (73.7)
Bovidae	Hippotraginae	*Oryx beisa*	East African oryx	1 (66.1)	1 (67.5)
Bovidae	Hippotraginae	*Oryx dammah*	Scimitar oryx	1 (72.0)	1 (73.5)
Bovidae	Hippotraginae	*Oryx gazella*	Gemsbok	7 (65.0)	7 (66.1)
Bovidae	Hippotraginae	*Oryx leucoryx*	Arabian oryx	1 (71.9)	1 (73.3)

Bovidae	Reduncinae	*Kobus ellipsiprymnus*	Waterbuck	5 (69.1)	5 (70.5)
Bovidae	Reduncinae	*Kobus leche*	Lechwe	1 (64.9)	1 (66.1)
Bovidae	Reduncinae	*Pelea capreolus*	Rhebok	1 (68.1)	1 (69.5)
Bovidae	Reduncinae	*Redunca arundinum*	Southern Reedbuck	1 (69.4)	1 (70.8)
Bovidae	Reduncinae	*Redunca fulvorufula*	Mountain Reedbuck	3 (67.4)	3 (69.0)
Camelidae		*Vicugna pacos*	Alpaca	11 (34.5)	0
Cervidae	Capreolinae	*Alces alces*	North American moose	16 (58.1)	10 (64.7)
Cervidae	Capreolinae	*Odocoileus virginianus*	White-tailed deer	8 (61.9)	8 (62.9)
Cervidae	Capreolinae	*Rangifer tarandus*	Caribou	8 (60.0)	7 (61.2)
Cervidae	Cervinae	*Axis axis*	Axis deer	8 (54.9)	8 (55.8)
Cervidae	Cervinae	*Cervus elaphus nelson*	Rocky mountain elk	8 (64.9)	8 (66.1)
Cervidae	Cervinae	*Cervus nippon*	Sika deer	8 (60.7)	8 (62.0)
Cervidae	Cervinae	*Dama dama*	Fallow deer	8 (63.7)	8 (64.8)
Elephantidae		*Loxodonta africana*	Savanna elephant	4 (52.7)	0
Equidae		*Equus caballus*	feral horse	1 (55.2)	0
Giraffidae		*Giraffa camelopardalis tippelskirchi*	Masai Giraffe	1 (64.7)	1 (66.0)

Table S1 in publication.

Table 2.2. The region harboring SNP rs17871403 was sequenced in 14 species to identify the nucleotides present at the SNP site.

In most cases, the sequenced samples include those that were genotyped with the BovineSNP50 BeadChip.

			Ambiguous Genotypes from BovineSNP50						
Species	**Common Name or Breed**	**Sequence SNP Call**	**Frequency of W Allele**	**Call Rate**	**Number of Genotyped Animals**	***WW***	***NN***	***SS***	**Frequency of Incorrect Genotype Calls**
Bos taurus taurus	Angus	T/G	0.551	0.998	6124	1868	2997	1246	0.000
Bos taurus taurus	Holstein	T/G	0.374	0.999	5769	788	2732	2246	0.000
Bos gaurus	Gaur	T/G	0.234	1.000	47	3	16	28	0.000
Bison bison	Plains bison	G	0.000	0.978	89	0	0	87	0.000
Bos grunniens	Yak	G	0.000	1.000	2	0	0	2	0.000
Bubalus bubalis	Asian water buffalo	A	1.000	0.917	12	11	0	0	0.000
Ovis Canadensis	Bighorn sheep	A	0.000	0.000	8	0	0	0	0.000
Oreamnos americanus	North American mountain goat	A	0.833	0.750	8	5	0	1	0.125
Rangifer tarandus	Caribou	A	0.875	1.000	8	7	0	1	0.125
Odocoileus viginianus	White-tailed deer	A	1.000	1.000	8	8	0	0	0.000
Alces aices	North American moose	A	1.000	0.813	16	13	0	0	0.000
Cervus Nippon	Sika deer	A	1.000	1.000	8	8	0	0	0.000
Dama dama	Fallow deer	A	1.000	1.000	8	8	0	0	0.000
Axis axis	Axis deer	A	1.000	1.000	8	8	0	0	0.000
Antilocapra Americana	Pronghorn	A	1.000	1.000	8	8	0	0	0.000
Outgroup Total					230	79	16	119	0.0087

Table S2 in publication.

Table 2.3. Comparison of genotypes produced from replicate *Bison priscus* ancient DNA libraries BS662 and KCMU02.

For the SNP used in the phylogenetic analyses, BS662 and KCMU02 produced heterozygosities of 39.6% and 11.5%, respectively. When all 54,693 SNPs scored on the BovineSNP50 BeadChip were compared, 74.7% of 15,947 genotypes called in both samples were identical but 99.5% = 100% × (1004 + 8994)/(1004 + 24 +29 + 8994) of the parsimony informative genotypes (homozygotes) were identical between the replicates. When only the SNPs used in the phylogenetic analyses were compared, 76.1% of the 12,279 genotypes called in both samples were identical but 99.7% of the parsimony informative genotypes were identical between the replicates.

All SNP

		KCMU02 genotypes (call rate: 0.542)				
		AA	AB	BB	No Call	Totals
BS662 genotypes (call rate: 0.451)	AA	1004	96	24	865	1989
	AB	818	1915	2561	5102	10396
	BB	29	506	8994	2778	12307
	No Call	3969	2098	7632	16302	30001
	Totals	5820	4615	19211	25047	54693

SNP used in phylogeny

		KCMU02 genotypes (call rate: 0.549)				
		AA	AB	BB	No Call	Totals
BS662 genotypes (call rate: 0.458)	AA	725	42	9	556	1332
	AB	631	1155	1956	3656	7398
	BB	16	279	7466	2201	9962
	No Call	3022	1101	6029	11999	22151
	Totals	4394	2577	15460	18412	40843

Table S3 in publication.

Table 2.4 Provenance for all samples included in the analyses.

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Addax nasomaculatus	Addax	Anas278200	SAZ 19	San Antonio Zoo, Texas, USA	CSM[1]	JEW[2]
Aepyceros melampus	Impala	Amel278230	SAZ 26	San Antonio Zoo, Texas, USA	CSM	JEW
Aepyceros melampus	Impala	Amel329740	200760001	Port Alfred, South Africa	MPH[3]	
Aepyceros melampus	Impala	Amel329750	200760002	Kimberley, South Africa	MPH	
Aepyceros melampus	Impala	Amel330140	200760003	South Africa	MPH	
Aepyceros melampus	Impala	Amel330150	200760004	South Africa	MPH	
Alcelaphus buselaphus jacksoni	Jackson's hartebeest	Abus278150	SAZ 09	San Antonio Zoo, Texas, USA	CSM	JEW
Alcelaphus buselaphus jacksoni	Jackson's hartebeest	Abus329860	200756001	South Africa	MPH	
Alcelaphus buselaphus jacksoni	Jackson's hartebeest	Abus329870	200756002	South Africa	MPH	
Alcelaphus lichtensteinii	Lichtenstein's hartebeest	Slic329800	200757001	Tanzania	MPH	
Alcelaphus lichtensteinii	Lichtenstein's hartebeest	Slic329810	200757002	Tanzania	MPH	
Alces alces	North American moose	Aalc276790	200124003/21	Wyoming, USA	MPH	DAH[4]
Alces alces	North American moose	Aalc276800	200124004/71	Wyoming, USA	MPH	DAH
Alces alces	North American moose	Aalc276810	200124005/72	Wyoming, USA	MPH	DAH
Alces alces	North American moose	Aalc276840	200124008/115	Wyoming, USA	MPH	DAH
Alces alces	North American moose	Aalc277570	200324002/AF7304/UA 53755	Alaska, USA	MPH	JIS[5]
Alces alces	North American moose	Aalc277580	200324003/AF7311/UA 32983	Alaska, USA	MPH	JIS
Alces alces	North American moose	Aalc277590	200324004/AF10475/U A54456	Alaska, USA	MPH	JIS

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Alces alces	North American moose	Aalc277600	200324005/AF10482/UA33007	Alaska, USA	MPH	JIS
Alces alces	North American moose	Aalc277610	200324006/AF18708/UA44530	Alaska, USA	MPH	JIS
Alces alces	North American moose	Aalc277620	200324007/AF30195/UA53788	Alaska, USA	MPH	JIS
Antidorcas marsupialis	Springbok	Amar329890	200751002	Kimberley, South Africa	MPH	
Antidorcas marsupialis	Springbok	Amar329900	200751001	Port Alfred, South Africa	MPH	
Antidorcas marsupialis	Springbok	Amar329940	200751003	South Africa	MPH	
Antidorcas marsupialis	Springbok	Amar329950	200751004	South Africa	MPH	
Antidorcas marsupialis	Springbok	Amar329970	200751006	South Africa	MPH	
Antilocapra americana	Pronghorn	Aame276870	200125001/2	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276880	200125003/80	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276890	200125004/422	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276900	200125005/437	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276910	200125006/648	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276920	200125008/1111	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276930	200325001/1151	Wyoming, USA	MPH	DAH
Antilocapra americana	Pronghorn	Aame276940	200325002/1159	Wyoming, USA	MPH	DAH
Antilope cervicapra	Blackbuck	Acer278070	Bb-s	Texas, USA	CSM	
Antilope cervicapra	Blackbuck	Acer277090	200750001	Kerr County, Texas, USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Antilope cervicapra	Blackbuck	Acer277100	200750002	Kerr County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277110	200750003	Kerr County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277120	200750004	Kerr County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277130	200750005	Kimble County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277140	200750006	Edwards County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277150	200750007	Sutton County, Texas, USA	MPH	
Antilope cervicapra	Blackbuck	Acer277160	200750008	Kerr County, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277170	200728001	Boerne, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277180	200728002	Boerne, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277190	200728003	Boerne, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277200	200728004	Boerne, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277210	200728005	Leakey, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277220	200728006	Real County, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277230	200728007	Sonora, Texas, USA	MPH	
Axis axis	Axis deer	Aaxi277240	200728008	Mountain Home, Texas, USA	MPH	
Bison bison	Plains Bison	BB024060	8110	Arrowhead Ranch, Ohio, USA	RDS[6]	
Bison bison	Plains Bison	BB200700	8001	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB200740	8005	Arrowhead Ranch, Ohio, USA	RDS	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bison bison	Plains Bison	BB200780	8010	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB200790	8011	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB200990	8044	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB201010	8046	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB201110	8072	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB201290	8112	Arrowhead Ranch, Ohio, USA	RDS	
Bison bison	Plains Bison	BB049850	AVID*076*109*828	Custer State Park, South Dakota, USA	RDS	GCB[7]
Bison bison	Plains Bison	BB050640		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB050790		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052300		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052440		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052460		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052470		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052480		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB052510		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains Bison	BB054350		Custer State Park, South Dakota, USA	RDS	GCB
Bison bison	Plains bison	FNBB331730		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM[8]
Bison bison	Plains bison	FNBB331740		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bison bison	Plains bison	FNBB331750		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM
Bison bison	Plains bison	FNBB331760		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM
Bison bison	Plains bison	FNBB331770		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM
Bison bison	Plains bison	FNBB331780		Fort Niobrara Natl Wildlife Refuge, Nebraska, USA	RDS	KMM
Bison bison athabascae	Wood Bison	EINP196490	W83003	Elk Island National Park, Alberta, Canada	GAW[9]	
Bison bison athabascae	Wood Bison	EINP196500	W81026	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196510	W85024	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196520	W85020	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196550	W83018	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196560	W81010	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196570	WR890100061	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196580	WR89005	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196590	WR87029	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	EINP196600	WR87001	Elk Island National Park, Alberta, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196640	Y98	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196650	Y96	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196660	Y86	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196670	Y82	Hook Lake, Northwest Territories, Canada	GAW	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bison bison athabascae	Wood Bison	HL196740	Y68	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196750	Y67	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196760	Y65	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196780	Y55	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196800	Y49	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	HL196830	Y42	Hook Lake, Northwest Territories, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196370	SW18	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196390	SW15	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196410	SW12	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196420	NL110	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196430	SW11	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196440	NL109	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196450	GR108	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196460	NL102	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196470	SW10	Wood Buffalo National Park, Canada	GAW	
Bison bison athabascae	Wood Bison	WBNP196480	SW1	Wood Buffalo National Park, Canada	GAW	
Bison priscus	Steppe Wisent	BB278340	BS662	Alyoshkina Zaimka, Siberia	AC[10]	
Bos gaurus	Gaur	GAUR335269	WG0041998-DNAE12_OGR000001	Henry Doorly Zoo, Nebraska, USA	RAB[11]	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos gaurus	Gaur	GAUR081370	5429	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081390	5443	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081410	4934	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081610	5359	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081650	12915	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081690	11498	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081710		Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081730	12926	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos gaurus	Gaur	GAUR081750	4431	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos grunniens	Yak	YAK339149	WG0042000-DNAG12_OYK000001	USA	MPH	
Bos grunniens	Yak	YAK339159	WG0042000-DNAH12_OYK000002	USA	MPH	
Bos javanicus	Banteng	Bjav334499	WG0042000-DNAC12_OBJ000001	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos javanicus	Banteng	Bjav334509	WG0042000-DNAD12_OBJ000002	Henry Doorly Zoo, Nebraska, USA	RAB	
Bos javanicus	Banteng	Bjav276510	200710003/1848/KB10724	San Diego Zoo, USA	MPH	LGC[12]
Bos javanicus	Banteng	Bjav276520	200710004/1851/KB10623	San Diego Zoo, USA	MPH	LGC
Bos taurus indicus	Gir	GIR335349	WG0041999-DNAB04_GIR000002	Brazil	CAG[13]	ARC[14]
Bos taurus indicus	Gir	GIR335359	WG0041999-DNAC04_GIR000003	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335369	WG0041999-DNAD04_GIR000004	Brazil	CAG	ARC

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus indicus	Gir	GIR335379	WG0041999-DNAE04_GIR000005	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335399	WG0041999-DNAG04_GIR000007	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335409	WG0041999-DNAH04_GIR000008	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335469	WG0041999-DNAF05_GIR000014	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335479	WG0041999-DNAG05_GIR000015	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR335489	WG0041999-DNAH05_GIR000016	Brazil	CAG	ARC
Bos taurus indicus	Gir	GIR333119	227	Brazil	CVT & TSS[15]	LLC[16]
Bos taurus indicus	Guzerat	GUZ333199	1971	Brazil	CVT & TSS	LLC
Bos taurus indicus	Guzerat	GUZ333209	2826	Brazil	CVT & TSS	LLC
Bos taurus indicus	Guzerat	GUZ333219	2827	Brazil	CVT & TSS	LLC
Bos taurus indicus	Nelore	NEL202860	NEL2	Brazil	ARC	
Bos taurus indicus	Nelore	NEL202870	NEL3	Brazil	ARC	
Bos taurus indicus	Nelore	NEL203040	NEL20	Brazil	ARC	
Bos taurus indicus	Nelore	NEL203100	NEL26	Brazil	ARC	
Bos taurus indicus	Nelore	NEL203260	NEL42	Brazil	ARC	
Bos taurus indicus	Nelore	NEL337729	WG0041998-DNAD07_NEL000001	Brazil	CAG	ARC
Bos taurus indicus	Nelore	NEL337789	WG0041998-DNAB08_NEL000007	Brazil	CAG	ARC
Bos taurus indicus	Nelore	NEL337809	WG0041998-DNAD08_NEL000009	Brazil	CAG	ARC

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus indicus	Nelore	NEL337839	WG0041998-DNAG08_NEL000012	Brazil	CAG	ARC
Bos taurus indicus	Nelore	NEL337939	WG0041998-DNAA10_NEL000022	Brazil	CAG	ARC
Bos taurus indicus	Sahiwal	SAHW333289	1675	Qadirabad, Punjab, Pakistan	MEB[17]	
Bos taurus indicus	Sahiwal	SAHW333299	1676	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333309	1683	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333319	1684	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333329	1686	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333349	1689	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333359	1690	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333369	1691	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333379	1693	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus indicus	Sahiwal	SAHW333399	1701	Qadirabad, Punjab, Pakistan	MEB	
Bos taurus taurus	Angus	ANG063100	9.82E+14	USA	JFT[18]	
Bos taurus taurus	Angus	AN047800		USA	JFT	
Bos taurus taurus	Angus	AN044460		USA	JFT	
Bos taurus taurus	Angus	AN046090		USA	JFT	
Bos taurus taurus	Angus	AN208020		USA	JFT	
Bos taurus taurus	Angus	AN018220		USA	JFT	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Angus	AN331000	252	USA	JFT	
Bos taurus taurus	Angus	AN331040	133	USA	JFT	
Bos taurus taurus	Angus	AN331150	8-May	USA	JFT	
Bos taurus taurus	Angus	AN331300	171	USA	JFT	
Bos taurus taurus	Angus	AN331340	167	USA	JFT	
Bos taurus taurus	Angus	AN023660		USA	JFT	
Bos taurus taurus	Angus	AN023750		USA	JFT	
Bos taurus taurus	Angus	AN023780		USA	JFT	
Bos taurus taurus	Angus	AN023790		USA	JFT	
Bos taurus taurus	Angus	AN023850		USA	JFT	
Bos taurus taurus	Angus	AN023990		USA	JFT	
Bos taurus taurus	Angus	AN024030		USA	JFT	
Bos taurus taurus	Angus	AN024040		USA	JFT	
Bos taurus taurus	Angus	AN024050		USA	JFT	
Bos taurus taurus	Belgian Blue	BBLU275070	12	USA	CSM	
Bos taurus taurus	Belgian Blue	BBLU275080	13	USA	CSM	
Bos taurus taurus	Belgian Blue	BBLU275090	14	USA	CSM	
Bos taurus taurus	Belgian Blue	BBLU275510	56	USA	CSM	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Belted Galloway	BGAL328210	BG-5227	Great Britain	PW[19]	
Bos taurus taurus	Belted Galloway	BGAL328220	BG-5230	Great Britain	PW	
Bos taurus taurus	Belted Galloway	BGAL328230	BG-5232	Great Britain	PW	
Bos taurus taurus	Belted Galloway	BGAL328240	BG-5233	Great Britain	PW	
Bos taurus taurus	Blonde d'Aquitaine	BDAQ275100	15	USA	CSM	
Bos taurus taurus	Blonde d'Aquitaine	BDAQ275110	16	USA	CSM	
Bos taurus taurus	Blonde d'Aquitaine	BDAQ275120	17	USA	CSM	
Bos taurus taurus	Blonde d'Aquitaine	BDAQ275130	18	USA	CSM	
Bos taurus taurus	Blonde d'Aquitaine	BDAQ275520	57	USA	CSM	
Bos taurus taurus	Brown Swiss	BSW334769	WG0041997-DNAA01_BSW000001	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334779	WG0041997-DNAB01_BSW000002	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334789	WG0041997-DNAC01_BSW000003	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334799	WG0041997-DNAD01_BSW000004	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334829	WG0041997-DNAG01_BSW000007	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334839	WG0041997-DNAH01_BSW000009	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334899	WG0041997-DNAF02_BSW000015	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334909	WG0041997-DNAG02_BSW000016	USA	CAG	CVT & TSS
Bos taurus taurus	Brown Swiss	BSW334929	WG0041997-DNAA03_BSW000018	USA	CAG	CVT & TSS

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Brown Swiss	BSW334949	WG0041997-DNAC03_BSW000020	USA	CAG	CVT & TSS
Bos taurus taurus	Charolais	CHA335229	WG0041996-DNAF05_CHL000023	Great Britain	CAG	JLW[20]
Bos taurus taurus	Charolais	CHA208210	18	USA	JFT	
Bos taurus taurus	Charolais	CHA208200	17	USA	JFT	
Bos taurus taurus	Charolais	CHA183240	7921	USA	JFT	HDD[21]
Bos taurus taurus	Charolais	CHA183250	3118	USA	JFT	HDD
Bos taurus taurus	Charolais	CHA183270	2995	USA	JFT	HDD
Bos taurus taurus	Charolais	CHA208130	10	USA	JFT	
Bos taurus taurus	Charolais	CHA208100	7	USA	JFT	
Bos taurus taurus	Charolais	CHA208180	15	USA	JFT	
Bos taurus taurus	Charolais	CHA208190	16	USA	JFT	
Bos taurus taurus	Charolais	CHA208280	25	USA	JFT	
Bos taurus taurus	Chianina	CHIA275600	65	USA	CSM	
Bos taurus taurus	Chianina	CHIA275620	67	USA	CSM	
Bos taurus taurus	Chianina	CHIA275630	68	USA	CSM	
Bos taurus taurus	Chianina	CHIA328660	19307293	USA	MPH	
Bos taurus taurus	Chianina	CHIA328670	19307295	USA	MPH	
Bos taurus taurus	Chianina	CHIA328680	19307300	USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Chianina	CHIA328690	19999916	USA	MPH	
Bos taurus taurus	Chianina	CHIA328700	19999923	USA	MPH	
Bos taurus taurus	Corriente	CORR275530	58	USA	CSM	
Bos taurus taurus	Corriente	CORR329110	19202900	USA	MPH	
Bos taurus taurus	Corriente	CORR329120	19202901	USA	MPH	
Bos taurus taurus	Corriente	CORR329130	19202902	USA	MPH	
Bos taurus taurus	Corriente	CORR329140	19202903	USA	MPH	
Bos taurus taurus	Devon	DEV328250	DEV-2089	Great Britain	PW	
Bos taurus taurus	Devon	DEV328260	DEV-2090	Great Britain	PW	
Bos taurus taurus	Devon	DEV328270	DEV-2091	Great Britain	PW	
Bos taurus taurus	Devon	DEV328280	DEV-2093	Great Britain	PW	
Bos taurus taurus	Dexter	DEX328290	DEX-15	Great Britain	PW	
Bos taurus taurus	Dexter	DEX328300	DEX-18	Great Britain	PW	
Bos taurus taurus	Dexter	DEX328310	DEX-19	Great Britain	PW	
Bos taurus taurus	Dexter	DEX328320	DEX-21	Great Britain	PW	
Bos taurus taurus	Finnish Ayrshire	AYR207260	MAHL_88	Finland	JFT	MA[22]
Bos taurus taurus	Finnish Ayrshire	AYR207520	MAHL_135	Finland	JFT	MA
Bos taurus taurus	Finnish Ayrshire	AYR331570	MAHLB21	Finland	JFT	MA

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Finnish Ayrshire	AYR331710	MAHLD4	Finland	JFT	MA
Bos taurus taurus	Finnish Ayrshire	AYR197170	FI_238	Finland	JV[23]	
Bos taurus taurus	Finnish Ayrshire	AYR197900	FI_338	Finland	JV	
Bos taurus taurus	Finnish Ayrshire	AYR198070	FI_136	Finland	JV	
Bos taurus taurus	Finnish Ayrshire	AYR198130	FI_144	Finland	JV	
Bos taurus taurus	Finnish Ayrshire	AYR198960	FI_79	Finland	JV	
Bos taurus taurus	Finnish Ayrshire	AYR199400	FI_16	Finland	JV	
Bos taurus taurus	Galloway	GALL328330	GA-2006	Great Britain	PW	
Bos taurus taurus	Galloway	GALL328340	GA-2162	Great Britain	PW	
Bos taurus taurus	Galloway	GALL328350	GA-2341	Great Britain	PW	
Bos taurus taurus	Galloway	GALL328360	GA-2346	Great Britain	PW	
Bos taurus taurus	Gelbvieh	GEL275540	59	USA	CSM	
Bos taurus taurus	Gelbvieh	GEL275640	69	USA	CSM	
Bos taurus taurus	Gelbvieh	GEL275650	70	USA	CSM	
Bos taurus taurus	Gelbvieh	GEL275660	71	USA	CSM	
Bos taurus taurus	Gelbvieh	GEL335309	WG0042007-DNAF05_GBV000001	USA	MPH	
Bos taurus taurus	Gelbvieh	GEL335319	WG0042007-DNAG05_GBV000002	USA	MPH	
Bos taurus taurus	Gelbvieh	GEL335329	WG0042007-DNAH05_GBV000003	USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Gelbvieh	GEL329320	19929826	USA	MPH	
Bos taurus taurus	Guernsey	GNS335589	WG0042000-DNAB01_GNS000002	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335599	WG0042000-DNAC01_GNS000003	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335609	WG0042000-DNAD01_GNS000004	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335619	WG0042000-DNAE01_GNS000005	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335629	WG0042000-DNAF01_GNS000006	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335639	WG0042000-DNAG01_GNS000007	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335659	WG0042000-DNAA02_GNS000009	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335669	WG0042000-DNAB02_GNS000010	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335689	WG0042000-DNAD02_GNS000012	United Kingdom	CAG	JLW
Bos taurus taurus	Guernsey	GNS335709	WG0042000-DNAF02_GNS000014	USA	CAG	CVT & TSS
Bos taurus taurus	Hanwoo	HANW206790	3	South Korea	JJK[24]	
Bos taurus taurus	Hanwoo	HANW206980	22	South Korea	JJK	
Bos taurus taurus	Hanwoo	HANW207000	24	South Korea	JJK	
Bos taurus taurus	Hanwoo	HANW207020	26	South Korea	JJK	
Bos taurus taurus	Hanwoo	HANW207160	40	South Korea	JJK	
Bos taurus taurus	Hanwoo	HANW207200	44	South Korea	JJK	
Bos taurus taurus	Hanwoo	HANW207220	46	South Korea	JJK	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Hereford	HFD208490	47	USA	JFT	
Bos taurus taurus	Hereford	HFD208440	42	USA	JFT	
Bos taurus taurus	Hereford	HFD183280	1447	USA	JFT	
Bos taurus taurus	Hereford	HFD183290	3020	USA	JFT	
Bos taurus taurus	Hereford	HFD183300	3165	USA	JFT	
Bos taurus taurus	Hereford	HFD208420	39	USA	JFT	
Bos taurus taurus	Hereford	HFD208460	44	USA	JFT	
Bos taurus taurus	Hereford	HFD208450	43	USA	JFT	
Bos taurus taurus	Hereford	HFD182441	41	USA	JFT	
Bos taurus taurus	Hereford	HFD209030	102	USA	JFT	
Bos taurus taurus	Holstein	HO242419		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO242869		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO242889		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO242909		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO244459		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO245109		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO245209		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO246909		USA	CVT & TSS	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Holstein	HO247899		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO247949		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO008890		USA	CVT & TSS	
Bos taurus taurus	Holstein	HO044939		USA	RDS	HDD
Bos taurus taurus	Holstein	HO044940		USA	RDS	HDD
Bos taurus taurus	Holstein	HO045019		USA	RDS	HDD
Bos taurus taurus	Holstein	HO020879		USA	RDS	
Bos taurus taurus	Holstein	HO044959		USA	RDS	
Bos taurus taurus	Holstein	HO203340		USA	RDS	
Bos taurus taurus	Holstein	HO008939		USA	RDS	
Bos taurus taurus	Holstein	HO020740		USA	RDS	
Bos taurus taurus	Holstein	HO020759		USA	RDS	
Bos taurus taurus	Jersey	JER336749	WG0041997-DNAC07_JER000001	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336759	WG0041997-DNAD07_JER000002	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336779	WG0041997-DNAF07_JER000004	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336789	WG0041997-DNAG07_JER000005	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336799	WG0041997-DNAH07_JER000006	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336809	WG0041997-DNAA08_JER000007	USA	CAG	CVT & TSS

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Jersey	JER336849	WG0041997-DNAE08_JER000011	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336929	WG0041997-DNAE09_JER000019	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336939	WG0041997-DNAF09_JER000020	USA	CAG	CVT & TSS
Bos taurus taurus	Jersey	JER336969	WG0041997-DNAA10_JER000023	USA	CAG	CVT & TSS
Bos taurus taurus	Kerry	KERR328410	KE-2112	Great Britain	PW	
Bos taurus taurus	Kerry	KERR328420	KE-2114	Great Britain	PW	
Bos taurus taurus	Kerry	KERR328430	KE-2115	Great Britain	PW	
Bos taurus taurus	Limousin	LM183340	1489	USA	JFT	HDD
Bos taurus taurus	Limousin	LM183360	2717	USA	JFT	HDD
Bos taurus taurus	Limousin	LM183390	1411	USA	JFT	HDD
Bos taurus taurus	Limousin	LM183460	1406	USA	JFT	HDD
Bos taurus taurus	Limousin	LM100770	3877	USA	JFT	
Bos taurus taurus	Limousin	LM127780	10520	USA	JFT	
Bos taurus taurus	Limousin	LM073800		USA	JFT	
Bos taurus taurus	Limousin	LM074261		USA	JFT	
Bos taurus taurus	Limousin	LM074360		USA	JFT	
Bos taurus taurus	Limousin	LM074420		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274500		USA	JFT	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Lincoln Red	SH274510		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274520		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274530		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274540		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274550		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274560		USA	JFT	
Bos taurus taurus	Lincoln Red	SH274710	Lincoln Red	USA	JFT	
Bos taurus taurus	Lincoln Red	SH274720	Tao Lincoln Red	USA	JFT	
Bos taurus taurus	Longhorn	LH328440	LH-1966	Great Britain	PW	
Bos taurus taurus	Longhorn	LH328450	LH-1968	Great Britain	PW	
Bos taurus taurus	Longhorn	LH328470	LH-1973	Great Britain	PW	
Bos taurus taurus	Maine Anjou	MAAN275760	81	USA	CSM	
Bos taurus taurus	Maine Anjou	MAAN275770	82	USA	CSM	
Bos taurus taurus	Maine Anjou	MAAN275780	83	USA	CSM	
Bos taurus taurus	Maine Anjou	MAAN329300	19999902	USA	MPH	
Bos taurus taurus	Maine Anjou	MAAN329310	19999903	USA	MPH	
Bos taurus taurus	Marchigiana	MCHI328760	19360406	USA	MPH	
Bos taurus taurus	Marchigiana	MCHI328770	19360425	USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Marchigiana	MCHI328780	19360430	USA	MPH	
Bos taurus taurus	Marchigiana	MCHI328790	19360441	USA	MPH	
Bos taurus taurus	Marchigiana	MCHI328800	19360459	USA	MPH	
Bos taurus taurus	Montbeliard	MONT328810	19360918	USA	MPH	
Bos taurus taurus	Montbeliard	MONT328820	19360921	USA	MPH	
Bos taurus taurus	Montbeliard	MONT328830	19360932	USA	MPH	
Bos taurus taurus	Montbeliard	MONT328840	19360938	USA	MPH	
Bos taurus taurus	Montbeliard	MONT328850	19360960	USA	MPH	
Bos taurus taurus	Murray Grey	MUGR328860	19204503	USA	MPH	
Bos taurus taurus	Murray Grey	MUGR328870	19204506	USA	MPH	
Bos taurus taurus	Murray Grey	MUGR328880	19204509	USA	MPH	
Bos taurus taurus	Murray Grey	MUGR328890	19204535	USA	MPH	
Bos taurus taurus	Murray Grey	MUGR328900	19204536	USA	MPH	
Bos taurus taurus	N'Dama	NDAM337479	WG0041998-DNAC04_NDA000001	Guinea	CAG	OH[25]
Bos taurus taurus	N'Dama	NDAM337539	WG0041998-DNAA05_NDA000012	Guinea	CAG	OH
Bos taurus taurus	N'Dama	NDAM337599	WG0041998-DNAG05_NDA000021	Guinea	CAG	OH
Bos taurus taurus	N'Dama	NDAM337619	WG0041998-DNAA06_NDA000024	Guinea	CAG	OH
Bos taurus taurus	N'Dama	NDAM337699	WG0041998-DNAA07_NDA000041	Guinea	CAG	OH

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	N'Dama	NDAM080160	ND1	Gambia	OH	
Bos taurus taurus	N'Dama	NDAM080180	ND2	Gambia	OH	
Bos taurus taurus	N'Dama	NDAM080190	ND3	Gambia	OH	
Bos taurus taurus	N'Dama	NDAM080280	ND4	Gambia	OH	
Bos taurus taurus	N'Dama	NDAM080860	ND7	Gambia	OH	
Bos taurus taurus	Norwegian Red	NRC337969	WG0042000-DNAF03_NRC000001	Norway	CAG	SL[26]
Bos taurus taurus	Norwegian Red	NRC337999	WG0042000-DNAA04_NRC000004	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338009	WG0042000-DNAB04_NRC000005	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338019	WG0042000-DNAC04_NRC000006	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338029	WG0042000-DNAD04_NRC000007	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338049	WG0042000-DNAF04_NRC000010	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338069	WG0042000-DNAH04_NRC000012	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338089	WG0042000-DNAB05_NRC000014	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338139	WG0042000-DNAG05_NRC000020	Norway	CAG	SL
Bos taurus taurus	Norwegian Red	NRC338149	WG0042000-DNAH05_NRC000021	Norway	CAG	SL
Bos taurus taurus	Piedmontese	PIED338249	WG0042000-DNAB07_PMT000008	Italy	CAG	PAM[27]
Bos taurus taurus	Piedmontese	PIED338259	WG0042000-DNAC07_PMT000009	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED338269	WG0042000-DNAD07_PMT000010	Italy	CAG	PAM

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Piedmontese	PIED338319	WG0042000-DNAA08_PMT000015	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED338369	WG0042000-DNAF08_PMT000020	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED338379	WG0042000-DNAG08_PMT000021	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED338389	WG0042000-DNAH08_PMT000022	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED338399	WG0042000-DNAA09_PMT000023	Italy	CAG	PAM
Bos taurus taurus	Piedmontese	PIED328910	19310110	USA	MPH	
Bos taurus taurus	Piedmontese	PIED328950	19899802	USA	MPH	
Bos taurus taurus	Pinzgauer	PINZ275840	89	USA	CSM	
Bos taurus taurus	Pinzgauer	PINZ329190	19360705	USA	MPH	
Bos taurus taurus	Pinzgauer	PINZ329200	19360724	USA	MPH	
Bos taurus taurus	Pinzgauer	PINZ329210	19360742	USA	MPH	
Bos taurus taurus	Pinzgauer	PINZ329220	19360769	USA	MPH	
Bos taurus taurus	Red Angus	ANR334049	WG0041998-DNAF10_RGU000003	USA	CAG	RDG[28]
Bos taurus taurus	Red Angus	ANR334059	WG0041998-DNAG10_RGU000004	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334069	WG0041998-DNAH10_RGU000005	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334079	WG0041998-DNAA11_RGU000006	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334089	WG0041998-DNAB11_RGU000007	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334099	WG0041998-DNAC11_RGU000008	USA	CAG	RDG

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Red Angus	ANR334119	WG0041998-DNAE11_RGU000010	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334129	WG0041998-DNAF11_RGU000011	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334139	WG0041998-DNAG11_RGU000012	USA	CAG	RDG
Bos taurus taurus	Red Angus	ANR334149	WG0042007-DNAF06_RGU000013	USA	CAG	RDG
Bos taurus taurus	Red Poll	REDP275910	96	USA	CSM	
Bos taurus taurus	Red Poll	REDP329230	19214256	USA	MPH	
Bos taurus taurus	Red Poll	REDP329240	19214259	USA	MPH	
Bos taurus taurus	Red Poll	REDP329250	19214264	USA	MPH	
Bos taurus taurus	Red Poll	REDP329260	19214265	USA	MPH	
Bos taurus taurus	Romagnola	RMG338469	WG0042000-DNAH09_RMG000006	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338479	WG0042000-DNAA10_RMG000007	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338489	WG0042000-DNAB10_RMG000008	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338499	WG0042000-DNAC10_RMG000009	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338509	WG0042000-DNAD10_RMG000010	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338519	WG0042000-DNAE10_RMG000011	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338529	WG0042000-DNAF10_RMG000012	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338569	WG0042000-DNAB11_RMG000016	Italy	CAG	PAM
Bos taurus taurus	Romagnola	RMG338589	WG0042000-DNAD11_RMG000018	Italy	CAG	PAM

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Romagnola	RMG338619	WG0042000-DNAG11_RMG000021	Italy	CAG	PAM
Bos taurus taurus	Romosinuano	ROMO321660	83	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC[29]
Bos taurus taurus	Romosinuano	ROMO321670	88	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321680	169	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321690	246	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321700	378	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321710	381	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321720	389	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Romosinuano	ROMO321730	390	US, parents imported from Venezuela & Costa Rica	JFT	CCC & SWC
Bos taurus taurus	Salers	SAL275480	53	USA	CSM	
Bos taurus taurus	Salers	SAL275490	54	USA	CSM	
Bos taurus taurus	Salers	SAL329270	19999875	USA	MPH	
Bos taurus taurus	Salers	SAL329280	19999876	USA	MPH	
Bos taurus taurus	Salers	SAL329290	19999880	USA	MPH	
Bos taurus taurus	Scottish Highland	SCHL275570	62	USA	CSM	
Bos taurus taurus	Scottish Highland	SCHL329150	19361120	USA	MPH	
Bos taurus taurus	Scottish Highland	SCHL329160	19361156	USA	MPH	
Bos taurus taurus	Scottish Highland	SCHL329170	19361158	USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Scottish Highland	SCHL328370	HI-2215	Great Britain	PW	
Bos taurus taurus	Scottish Highland	SCHL328380	HI-2221	Great Britain	PW	
Bos taurus taurus	Scottish Highland	SCHL328390	HI-2223	Great Britain	PW	
Bos taurus taurus	Scottish Highland	SCHL328400	HI-2225	Great Britain	PW	
Bos taurus taurus	Shorthorn	SH330660		USA	JFT	
Bos taurus taurus	Shorthorn	SH330700		USA	JFT	
Bos taurus taurus	Shorthorn	SH330720		USA	JFT	
Bos taurus taurus	Shorthorn	SH330780		USA	JFT	
Bos taurus taurus	Shorthorn	SH330800		USA	JFT	
Bos taurus taurus	Shorthorn	SH330840		USA	JFT	
Bos taurus taurus	Shorthorn	SH330850		USA	JFT	
Bos taurus taurus	Shorthorn	SH330880		USA	JFT	
Bos taurus taurus	Shorthorn	SH330900		USA	JFT	
Bos taurus taurus	Shorthorn	SH330980		USA	JFT	
Bos taurus taurus	Simmental	SIM173550	M990926	USA	JFT	
Bos taurus taurus	Simmental	SIM183650	3210	USA	JFT	HDD
Bos taurus taurus	Simmental	SIM183660	3207	USA	JFT	HDD
Bos taurus taurus	Simmental	SIM183680	8298	USA	JFT	HDD

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Simmental	SIM209060	105	USA	JFT	
Bos taurus taurus	Simmental	SIM088430		USA	JFT	
Bos taurus taurus	Simmental	SIM088450		USA	JFT	
Bos taurus taurus	Simmental	SIM088460		USA	JFT	
Bos taurus taurus	Simmental	SIM088470		USA	JFT	
Bos taurus taurus	Simmental	SIM088500		USA	JFT	
Bos taurus taurus	South Devon	SDEV328480	SD-3354	Great Britain	PW	
Bos taurus taurus	South Devon	SDEV328490	SD-4090	Great Britain	PW	
Bos taurus taurus	South Devon	SDEV328500	SD-5290	Great Britain	PW	
Bos taurus taurus	South Devon	SDEV328510	SD-5972	Great Britain	PW	
Bos taurus taurus	Sussex	SUSS328560	SU-2151	Great Britain	PW	
Bos taurus taurus	Sussex	SUSS328570	SU-2152	Great Britain	PW	
Bos taurus taurus	Sussex	SUSS328580	SU-2153	Great Britain	PW	
Bos taurus taurus	Sussex	SUSS328590	SU-2154	Great Britain	PW	
Bos taurus taurus	Tarentaise	TARE329010	19293035	USA	MPH	
Bos taurus taurus	Tarentaise	TARE329020	19293032	USA	MPH	
Bos taurus taurus	Tarentaise	TARE329030	19293033	USA	MPH	
Bos taurus taurus	Tarentaise	TARE329040	19293034	USA	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Tarentaise	TARE329050	19293129	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH275450	50	USA	CSM	
Bos taurus taurus	Texas Longhorn	TXLH329350	19999872	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329390	19330046	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329410	19330055	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329430	19330061	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329500	19330081	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329520	19330086	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329540	19330091	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329590	19330105	USA	MPH	
Bos taurus taurus	Texas Longhorn	TXLH329600	19859870	USA	MPH	
Bos taurus taurus	Wagyu	WAGY206270	686	USA	HLN[30]	
Bos taurus taurus	Wagyu	WAGY206280	1615	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206290	2101	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206300	2126	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206310	2127	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206340	2892	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206360	4210	USA	HLN	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bos taurus taurus	Wagyu	WAGY206370	4803	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206380	4826	USA	HLN	
Bos taurus taurus	Wagyu	WAGY206410	4940	USA	HLN	
Bos taurus taurus	Welsh Black	WEBL328600	WB-5222	Great Britain	PW	
Bos taurus taurus	Welsh Black	WEBL328610	WB-5226	Great Britain	PW	
Bos taurus taurus	White Park	WHPK328620	WP-2061	Great Britain	PW	
Bos taurus taurus	White Park	WHPK328630	WP-2062	Great Britain	PW	
Bos taurus taurus	White Park	WHPK328640	WP-2065	Great Britain	PW	
Bos taurus taurus	White Park	WHPK328650	WP-2066	Great Britain	PW	
Boselaphus tragocamelus	Nilgai	Btra276230	200726001	Raymondville, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276240	200726002	Raymondville, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276250	200726005	Raymondville, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276260	200726007	Raymondville, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276270	200726008	Willacy County, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276280	200726009	Kennedy County, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276290	200726010	Kennedy County, Texas, USA	MPH	
Boselaphus tragocamelus	Nilgai	Btra276300	200726011	Kennedy County, Texas, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub334229	WG0041998-DNAB12_OWB000002	Henry Doorly Zoo, Nebraska, USA	RAB	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Bubalus bubalis	Asian water buffalo	Bbub334239	WG0041998-DNAC12_OWB000003	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub334249	WG0041998-DNAD12_OWB000004	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277670	200713003/643	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277680	200713004/607	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277690	200713005/648	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277710	200713007/641	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277720	200713008/629	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277730	200713009/618	Florida, USA	MPH	
Bubalus bubalis	Asian water buffalo	Bbub277740	200713010/617	Florida, USA	MPH	
Capra ibex	Ibex	IBEX273130		Various locations, Europe	JFT	LFK[31]
Capra ibex	Ibex	IBEX273150		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273170		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273180		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273190		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273200		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273220		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273230		Various locations, Europe	JFT	LFK
Capra ibex	Ibex	IBEX273240		Various locations, Europe	JFT	LFK

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Capra ibex	Ibex	IBEX273250		Various locations, Europe	JFT	LFK
Cervus canadensis	Rocky mountain elk	Ccan276690	200121001/49	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276700	200121002/240	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276710	200121003/284	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276720	200121004/575	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276730	200121007/813	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276740	200121008/858	Wyoming, USA	MPH	DAH
Cervus canadensis	Rocky mountain elk	Ccan276750	200121009	Wyoming, USA	MPH	
Cervus canadensis	Rocky mountain elk	Ccan276760	200121010	Wyoming, USA	MPH	
Cervus nippon	Sika deer	Cnip276430	200729001	Junction, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276440	200729002	Junction, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276450	200729003	Junction, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276460	200729004	Junction, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276470	200729005	Kerr County, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276480	200729006	Kimble County, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276490	200729007	Real County, Texas, USA	MPH	
Cervus nippon	Sika deer	Cnip276500	200729008	Kerr County, Texas, USA	MPH	
Connochaetes gnou	Black wildebeest	Cgno330080	200758005	Mpumalanga, South Africa	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Connochaetes gnou	Black wildebeest	Cgno330090	200758006	South Africa	MPH	
Connochaetes gnou	Black wildebeest	Cgno330100	200758007	South Africa	MPH	
Connochaetes taurinus	Blue wildebeest	Ctau329930	200758001	Kimberley, South Africa	MPH	
Connochaetes taurinus	Blue wildebeest	Ctau330050	200758002	South Africa	MPH	
Connochaetes taurinus	Blue wildebeest	Ctau330060	200758003	Zimbabwe	MPH	
Connochaetes taurinus	Blue wildebeest	Ctau330070	200758004	Tanzania	MPH	
Dama dama	Fallow deer	Ddam276310	200620004	Candor, New York, USA	MPH	
Dama dama	Fallow deer	Ddam276320	200620005	Candor, New York, USA	MPH	
Dama dama	Fallow deer	Ddam276330	200620006	Candor, New York, USA	MPH	
Dama dama	Fallow deer	Ddam276340	200620007	Candor, New York, USA	MPH	
Dama dama	Fallow deer	Ddam276350	200720008	Edwards County, Texas, USA	MPH	
Dama dama	Fallow deer	Ddam276360	200720009	Kerr County, Texas, USA	MPH	
Dama dama	Fallow deer	Ddam276370	200720010	Kerr County, Texas, USA	MPH	
Dama dama	Fallow deer	Ddam276380	200720011	Kerr County, Texas, USA	MPH	
Damaliscus korrigum jimela	Topi	Dkor278160	SAZ 13	San Antonio Zoo, Texas, USA	CSM	JEW
Damaliscus lunatus	Topi	Dlun329910	200754001	South Africa	MPH	
Damaliscus lunatus	Topi	Dlun329990	200754002	Mpumalanga, South Africa	MPH	
Damaliscus pygargus phillipsi	Blesbok	BLES329610	200759001	Port Alfred, South Africa	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Damaliscus pygargus phillipsi	Blesbok	BLES329620	200759002	Kimberley, South Africa	MPH	
Damaliscus pygargus phillipsi	Blesbok	BLES330110	200759003	South Africa	MPH	
Damaliscus pygargus phillipsi	Blesbok	BLES330130	200759005	South Africa	MPH	
Gazella spekei	Speke's gazelle	Gspe278120	SAZ 04	San Antonio Zoo, Texas, USA	CSM	JEW
Gazella subgutturosa	Persian gazelle	Gsub278270	SDZ 12	San Diego Zoo, Texas, USA	CSM	JEW
Gazella thomsoni	Thomson's gazelle	Gtho278110	SAZ 03	San Antonio Zoo, Texas, USA	CSM	JEW
Giraffa camelopardalis tippelskirchi	Masai Giraffe	Gcam278080	KB-3332	San Diego Zoo, Texas, USA	CMS	JEW
Hippotragus equinus	Roan antelope	Hequ330380	200774001	Namibia	MPH	
Hippotragus niger	Sable antelope	Hnig278180	SAZ 16	San Diego Zoo, Texas, USA	CSM	JEW
Kobus ellipsiprymnus	Waterbuck	Kell278100	SAZ 02	San Antonio Zoo, Texas, USA	CSM	JEW
Kobus ellipsiprymnus	Waterbuck	Kell329920	200770001	Port Alfred, South Africa	MPH	
Kobus ellipsiprymnus	Waterbuck	Kell330300	200770002	South Africa	MPH	
Kobus ellipsiprymnus	Waterbuck	Kell330310	200770003	South Africa	MPH	
Kobus ellipsiprymnus	Waterbuck	Kell330320	200770004	Zimbabwe	MPH	
Kobus leche	Lechwe	Klec330410	200776001	South Africa	MPH	
Nanger dama	Dama gazelle	dgaz278220	SAZ 23	San Antonio Zoo, Texas, USA	CSM	JEW
Nanger granti	Grant's Roosevelt gazelle	Ggra278280	SDZ 16	San Diego Zoo, Texas, USA	CSM	JEW
Nanger soemmerringii	Soemmerring's gazelle	Nsoe329700	200753001	Ethiopia	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Odocoileus virginianus	White-tailed deer	Ovir273030		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273040		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273050		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273060		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273070		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273080		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273090		Texas, USA	CSM	
Odocoileus virginianus	White-tailed deer	Ovir273100		Texas, USA	CSM	
Oreamnos americanus	North American mountain goat	Oame276620	200117002	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276630	200117003	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276640	200117004	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276650	200117005	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276660	200117006	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276670	200117007	Wyoming, USA	MPH	DAH
Oreamnos americanus	North American mountain goat	Oame276680	200317001	Montana, USA	MPH	
Oryx beisa	East African oryx	Obei330040	200755007	Ethiopia	MPH	
Oryx dammah	Scimitar oryx	Odam278210	SAZ 21	San Antonio Zoo, Texas, USA	CSM	JEW
Oryx gazella	Gemsbok	Ogaz278130	SAZ 06	San Antonio Zoo, Texas, USA	CSM	JEW

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Oryx gazella	Gemsbok	Ogaz329710	200755001	Port Alfred, South Africa	MPH	
Oryx gazella	Gemsbok	Ogaz329720	200755002	Unknown	MPH	
Oryx gazella	Gemsbok	Ogaz330000	200755003	Namibia	MPH	
Oryx gazella	Gemsbok	Ogaz330010	200755004	South Africa	MPH	
Oryx gazella	Gemsbok	Ogaz330020	200755005	Zimbabwe	MPH	
Oryx gazella	Gemsbok	Ogaz330030	200755006	Zimbabwe	MPH	
Oryx leucoryx	Arabian oryx	Oleu278140	SAZ 07	San Antonio Zoo, Texas, USA	CSM	JEW
Ourebia ourebi	Oribi	Oour330360	200772001	South Africa	MPH	
Ovibos moschatus	Muskox	Omos277390	200314001/AF1191/UA 30652	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277400	200314002/AF1192/UA 30653	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277410	200314003/AF1195/UA 30655	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277420	200314005/AF1243/UA 30657	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277430	200314006/AF12161/UA51073	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277440	200314007/AF27582/UA57951	Alaska, USA	MPH	JIS
Ovibos moschatus	Muskox	Omos277450	200314008/AF34278/UA53847	Alaska, USA	MPH	JIS
Ovis aries	Dorper sheep	OADM339199	5271	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Dorper sheep	OADM339209	5326	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Dorper sheep	OADM339219	5328	Kapiti Plains Estate, Kenya	CVT & TSS	OH

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Ovis aries	Dorper sheep	OADM339229	5337	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339169	3791	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339179	4853	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339189	5020	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339249	5345	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339259	12172	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis aries	Red Maasai sheep	OADM339269	12475	Kapiti Plains Estate, Kenya	CVT & TSS	OH
Ovis canadensis	Bighorn sheep	Ocan277460	200115001/64	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277470	200115002/65	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277480	200115003/66	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277490	200115004/68	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277500	200115005/70	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277510	200115006/72	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277520	200115007/215	Wyoming, USA	MPH	DAH
Ovis canadensis	Bighorn sheep	Ocan277530	200115008/224	Wyoming, USA	MPH	DAH
Pelea capreolus	Rhebok	Pcap330370	200773001	South Africa	MPH	
Rangifer tarandus	Caribou	Rtar276530	200323001/AF322/UA7 3308	Alaska, USA	MPH	JIS
Rangifer tarandus	Caribou	Rtar276540	200323002/AF323/UA7 3309	Alaska, USA	MPH	JIS

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Rangifer tarandus	Caribou	Rtar276550	200323003/AF324/UA73310	Alaska, USA	MPH	JIS
Rangifer tarandus	Caribou	Rtar276560	200323004/AF1204/UA30592	Alaska, USA	MPH	JIS
Rangifer tarandus	Caribou	Rtar276580	200323006/AF12971/UA31862	Alaska, USA	MPH	JIS
Rangifer tarandus	Caribou	Rtar276590	200323007/AF24825/UA69489	Alaska, USA	MPH	JIS
Rangifer tarandus	Caribou	Rtar276600	200323008/AF33766/UA53736	Alaska, USA	MPH	JIS
Raphicerus campestris	Steenbok	Rcam330390	200775001	Namibia	MPH	
Raphicerus campestris	Steenbok	Rcam330400	200775002	South Africa	MPH	
Redunca arundinum	Southern Reedbuck	Raru329880	200771001	Port Alfred, South Africa	MPH	
Redunca fulvorufula	Mountain Reedbuck	Rful330330	200771002	South Africa	MPH	
Redunca fulvorufula	Mountain Reedbuck	Rful330340	200771003	South Africa	MPH	
Redunca fulvorufula	Mountain Reedbuck	Rful330350	200771004	South Africa	MPH	
Rupicapra rupicapra	Chamois	Rrup278090	Rr 6389	San Diego Zoo, Texas, USA	CSM	JEW
Sylvicapra grimmia	Common duiker	Sgri329670	200768001	Namibia	MPH	
Sylvicapra grimmia	Common duiker	Sgri330270	200768002	South Africa	MPH	
Sylvicapra grimmia	Common duiker	Sgri330280	200768003	South Africa	MPH	
Syncerus caffer	African buffalo	Scaf278240	SCM 451	San Antonio Zoo, Texas, USA	CSM	JEW
Syncerus caffer	African buffalo	Scaf338659	WG0042000-DNAE12_OCB000001	Henry Doorly Zoo, Nebraska, USA	RAB	
Syncerus caffer	African buffalo	Scaf338669	WG0042000-DNAF12_OCB000002	Henry Doorly Zoo, Nebraska, USA	RAB	

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Syncerus caffer	African buffalo	Scaf329630	200778001	Zimbabwe	MPH	
Syncerus caffer	African buffalo	Scaf329640	200778002	Zimbabwe	MPH	
Syncerus caffer	African buffalo	Scaf330420	200778003	Zimbabwe	MPH	
Syncerus caffer	African buffalo	Scaf330430	200778004	Tanzania	MPH	
Syncerus caffer	African buffalo	Scaf330440	200778005	Zimbabwe	MPH	
Syncerus caffer nanus	African forest buffalo	Scan278250	SCNM-2-875	San Antonio Zoo, Texas, USA	CSM	JEW
Syncerus caffer nanus	African forest buffalo	SCAN321740	Lennie	San Diego Zoo, USA	LSE[32]	ZSS[33]
Syncerus caffer nanus	African forest buffalo	SCAN321750	WAP06	San Diego Zoo, USA	LSE	ZSS
Taurotragus oryx	Eland	Tory278290	SDZ 18	San Diego Zoo, USA	CSM	JEW
Taurotragus oryx	Eland	Tory330200	200762002	South Africa	MPH	
Taurotragus oryx	Eland	Tory330210	200762003	Namibia	MPH	
Taurotragus oryx	Eland	Tory330220	200762004	Zimbabwe	MPH	
Tragelaphus angasii	Nyala	Tang278260	SDZ 02	San Diego Zoo, USA	CSM	JEW
Tragelaphus angasii	Nyala	Tang329820	200765001	Port Alfred, South Africa	MPH	
Tragelaphus angasii	Nyala	Tang329830	200765002	Kimberley, South Africa	MPH	
Tragelaphus angasii	Nyala	Tang329840	200765003	Zimbabwe	MPH	
Tragelaphus angasii	Nyala	Tang330260	200765005	South Africa	MPH	
Tragelaphus imberbis	Lesser Kudu	Timb278190	SAZ 17	San Antonio Zoo, Texas, USA	CSM	JEW

Cont. Table 2.4 Provenance for all samples included in the analyses

Group	Common Name	UMC Sample ID	Provider Sample ID	Sample Location	Primary Contact	Secondary Contact
Tragelaphus imberbis	Lesser Kudu	Timb330290	200769001	Tanzania	MPH	
Tragelaphus scriptus	Bushbuck	Tscr329650	200761002	Kimberley, South Africa	MPH	
Tragelaphus scriptus	Bushbuck	Tscr329660	200761001	Port Alfred, South Africa	MPH	
Tragelaphus scriptus	Bushbuck	Tscr330170	200761003	Zimbabwe	MPH	
Tragelaphus scriptus	Bushbuck	Tscr330180	200761004	Zimbabwe	MPH	
Tragelaphus scriptus	Bushbuck	Tscr330190	200761005	Zimbabwe	MPH	
Tragelaphus strepsiceros	Greater Kudu	Tstr278170	SAZ 15	San Antonio Zoo, Texas, USA	CSM	JEW
Tragelaphus strepsiceros	Greater Kudu	Tstr329770	200764001	Port Alfred, South Africa	MPH	
Tragelaphus strepsiceros	Greater Kudu	Tstr329780	200763002	Kimberley, South Africa	MPH	
Tragelaphus strepsiceros	Greater Kudu	Tstr330230	200763003	South Africa	MPH	
Tragelaphus strepsiceros	Greater Kudu	Tstr330240	200763004	South Africa	MPH	
Tragelaphus strepsiceros	Greater Kudu	Tstr330250	200763005	Mpumalanga, South Africa	MPH	

Cont. Table 2.4 Provenance for all samples included in the analyses

[1]Christopher M. Seabury, Department of Veterinary Pathobiology, VMS Building Room 206C, Texas A&M University, College Station, Texas 77843-4467, USA

[2]James E. Womack, Department of Veterinary Pathobiology, VMS Building Room 309, Texas A&M University, College Station, Texas 77843-4467, USA

[3]Michael P. Heaton, United States Department of Agriculture, Meat Animal Research Center, P.O. Box 166, U.S. Spur 18D, Clay Center, Nebraska 68933, USA

[4]Deedra A. Hawk, Wyoming Game and Fish Department, 5400 Bishop Boulevard, Cheyenne, Wyoming 82006, USA

[5]Jennifer I. Schmidt, Institute of Arctic Biology, University of Alaska Fairbanks, 902 N. Koyukuk Dr, Fairbanks, Alaska 99775-7000, USA

[6]Robert D. Schnabel, 162 Animal Sciences, University of Missouri, Columbia, Missouri 65211-5300, USA

[7]Gary C. Brundige, Custer State Park, 13329 US Hwy 16A, Custer, South Dakota 57730, USA

[8]Kathy M. McPeak, Fort Niobrara National Wildlife Refuge, HC 14 Box 67, Valentine, Nebraska 69201, USA

[9]Gregory A. Wilson, Canadian Wildlife Service, 200 4999 98th Ave NW, Edmonton, Alberta T6B 2X3, Canada

[10]Alan Cooper, Australian Centre for Ancient DNA, Darling Building, University of Adelaide, North Terrace Campus, South Australia 5005, Australia

[11]Rick A. Brenneman, Center for Conservation and Research, Omaha's Henry Doorly Zoo, 3701 South 10th Street, Omaha, Nebraska 68107, USA

[12]Leona G. Chemnick, Beckman Center for Conservation Research, San Diego Zoo, 15600 San Pasqual Valley Road, Escondido, California 92027, USA

[13]Clare A. Gill, Bovine HapMap Consortium Curator, Animal Genomics, Department of Animal Science, Texas A&M University, College Station, Texas 77843-2471, USA

[14]Alexandre R. Caetano, Embrapa Recursos Geneticos e Biotecnologia, Parque Estacao Biologica, Final Av. W/5 Norte, Brasilia-DF, C.P. 02372 70770-900, Brasil

[15]Curt Van Tassell & Tad S. Sonstegard, United States Department of Agriculture, ARS, Bovine Functional Genomics Laboratory, Beltsville, Maryland 20705, USA

[16]Luiz L. Coutinho, Departamento de Zootecnia, ESALQ-USP, Av. Padua Dias, 11, Piracicaba, SP 13418-900, Brasil

[17]Masroor E. Babar, Department of Livestock Production, University of Veterinary and Animal Sciences, Lahore 54000, Pakistan

[18]Jeremy F. Taylor, S135B Animal Sciences, University of Missouri, Columbia, Missouri 65211-5300, USA

[19]Pamela Wiener, The Roslin Institute, and R(D)SVS, The University of Edinburgh, Roslin, Midlothian EH25 9PS, UK

[20]John L. Williams, Parco Tecnologico Padano, Via Einstein, Polo Universitario, Lodi, Italy

[21]Harvey D. Blackburn, National Animal Germplasm Program, National Center for Genetic Resources Preservation, 1111 South Mason St., Fort Collins, Colorado 80523, USA

[22]Magnus Andersson, Faculty of Veterinary Medicine, Department of Clinical Veterinary Sciences, Saari Unit, Pohjoinen pikatie 800, FIN-04920 Saarentaus, Finland
[23]Johanna Vilkki, Agrifood Research Finland MTT, FIN-31600, Jokioinen, Finland
[24]Jong-Joo Kim, School of Biotechnology, Yeungnam University, Gyeongsan, "Gyeongbuk", Republic of Korea
[25]Olivier Hanotte, School of Biology, University of Nottingham, Nottingham NG7 2RD, United Kingdom
[26]Sigbjørn Lien, Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Box 5003, N-1432 Aas, Norway
[27]Paulo Ajmone-Marsan, Faculty of Agriculture, Università Cattolica del S. Cuore, Piacenza, Italy
[28]Ronnie D. Green, Pfizer Animal Genetics, P.O. Box 486, Sutton, Nebraska 68979, USA
[29]Chad C. Chase & Samuel W. Coleman, USDA ARS SubTropical Agricultural Research Station (STARS), Brooksville, Florida 34601, USA
[30]Holly L. Neibergs, Washington State University, Department of Animal Sciences, P.O. Box 646353, Pullman, Washington 99164, USA
[31]Lukas F. Keller, Zoologisches Museum, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland
[32]Lori S. Eggert, Division of Biological Sciences, University of Missouri, Columbia, Missouri 65211-7400, USA
[33]Beckman Center for Conservation Research, San Diego Zoo, 15600 San Pasqual Valley Road, Escondido, California 92027, USA

Table S4 in publication.

Table 2.5. Average MAF for all 54,693 SNP and the 40,843 SNP used in phylogenomic analysis for all species and breeds (excluding groups with only 1 sample).
The difference shows that we effectively increased the proportion of informative characters within cattle breeds and simultaneously reduced possible genotype errors in species outside of *Bos taurus* using our SNP selection criteria. Species in grey were excluded from the final phylogeny (see text).

Scientific Name	Common Name/Breed	Average MAF (54,693 SNP)	Average MAF (40843 SNP)	Difference
Antilocapra americana	Pronghorn	0.048	0.035	-0.013
Aepyceros melampus	Impala	0.033	0.018	-0.016
Alcelaphus buselaphus jacksoni	Jackson's hartebeest	0.021	0.014	-0.006
Alcelaphus lichtensteinii	Lichtenstein's hartebeest	0.016	0.010	-0.006
Connochaetes gnou	Black wildebeest	0.020	0.014	-0.006
Connochaetes taurinus	Blue wildebeest	0.036	0.027	-0.009
Damaliscus pygargus phillipsi	Blesbok	0.068	0.018	-0.050
Damaliscus korrigum jimela	Topi	0.012	0.007	-0.005
Damaliscus lunatus	Topi	0.018	0.012	-0.006
Antidorcas marsupialis	Springbok	0.053	0.022	-0.031
Antilope cervicapra	Blackbuck	0.027	0.018	-0.009
Gazella dorcas	Dorcas gazelle	0.016		
Nanger granti	Grant's Roosevelt gazelle	0.016	0.012	-0.005
Gazella spekei	Speke's gazelle	0.011	0.007	-0.004
Gazella subgutturosa	Persian gazelle	0.021	0.013	-0.008
Gazella thomsoni	Thomson's gazelle	0.023	0.015	-0.009
Litocranius walleri	Gerenuk	0.249		
Nanger dama	Dama gazelle	0.010	0.006	-0.004
Nanger granti	Grant's gazelle	0.090		
Nanger soemmerringii	Soemmerring's gazelle	0.010	0.006	-0.004
Oreotragus oreotragus	Klipspringer	0.011		
Ourebia ourebi	Oribi	0.021	0.014	-0.007
Raphicerus campestris	Steenbok	0.029	0.021	-0.007
Bison bison	Plains Bison	0.015	0.014	-0.001
Bison Bison athabascae	Wood Bison	0.018	0.009	-0.009
Bison sp.	European Wisent			
Bison priscus	Steppe Wisent			
Bos gaurus	Gaur	0.016	0.013	-0.003
Bos grunniens	Yak	0.006	0.006	0.000
Bos javanicus	Banteng	0.012	0.012	-0.001
Bos taurus indicus	Gir	0.109	0.111	0.002
Bos taurus indicus	Guzerat	0.095	0.098	0.003
Bos taurus indicus	Nelore	0.125	0.111	-0.014

Bos taurus indicus	Sahiwal	0.106	0.111	0.005
Bos taurus taurus	Angus	0.219	0.219	0.000
Bos taurus taurus	Belgian blue	0.193	0.203	0.010
Bos taurus taurus	Belted Galloway	0.178	0.188	0.010
Bos taurus taurus	Blonde d'Aquitaine	0.198	0.207	0.010
Bos taurus taurus	Brown Swiss	0.192	0.186	-0.006
Bos taurus taurus	Charolais	0.232	0.222	-0.010
Bos taurus taurus	Chianina	0.205	0.209	0.004
Bos taurus taurus	Corriente	0.196	0.205	0.010
Bos taurus taurus	Devon	0.177	0.186	0.009
Bos taurus taurus	Dexter	0.180	0.189	0.009
Bos taurus taurus	Finnish Ayrshire	0.209	0.202	-0.007
Bos taurus taurus	Galloway	0.176	0.185	0.009
Bos taurus taurus	Gelbvieh	0.206	0.216	0.010
Bos taurus taurus	Guernsey	0.192	0.192	0.000
Bos taurus taurus	Hanwoo	0.199	0.198	-0.001
Bos taurus taurus	Hereford	0.225	0.213	-0.012
Bos taurus taurus	Holstein	0.221	0.229	0.007
Bos taurus taurus	Jersey	0.187	0.181	-0.006
Bos taurus taurus	Kerry	0.181	0.191	0.010
Bos taurus taurus	Limousin	0.220	0.215	-0.005
Bos taurus taurus	Longhorn	0.144	0.139	-0.005
Bos taurus taurus	Maine Anjou	0.187	0.196	0.009
Bos taurus taurus	Marchigiana	0.195	0.201	0.006
Bos taurus taurus	Montbeliard	0.186	0.195	0.009
Bos taurus taurus	Murray Grey	0.194	0.203	0.009
Bos taurus taurus	N'Dama	0.160	0.160	0.000
Bos taurus taurus	Normande	0.141		
Bos taurus taurus	Norwegian Red	0.218	0.219	0.002
Bos taurus taurus	Piedmontese	0.220	0.224	0.004
Bos taurus taurus	Pinzgauer	0.198	0.207	0.009
Bos taurus taurus	Red Angus	0.209	0.217	0.008
Bos taurus taurus	Red Poll	0.185	0.193	0.008
Bos taurus taurus	Romagnola	0.201	0.205	0.003
Bos taurus taurus	Romosinuano	0.191	0.201	0.010
Bos taurus taurus	Salers	0.195	0.204	0.009
Bos taurus taurus	Scottish Highland	0.193	0.190	-0.003
Bos taurus taurus	Shorthorn	0.196	0.192	-0.004
Bos taurus taurus	Simmental	0.233	0.213	-0.020
Bos taurus taurus	South Devon	0.183	0.190	0.007
Bos taurus taurus	Sussex	0.173	0.181	0.008

Bos taurus taurus	Tarentaise	0.195	0.204	0.009
Bos taurus taurus	Texas Longhorn	0.213	0.217	0.005
Bos taurus taurus	Wagyu	0.164	0.173	0.009
Bos taurus taurus	Welsh Black	0.176	0.184	0.009
Bos taurus taurus	White Park	0.170	0.151	-0.019
Boselaphus tragocamelus	Nilgai	0.024	0.016	-0.008
Bubalus bubalis	Asian water buffalo	0.025	0.013	-0.012
Syncerus caffer	African buffalo	0.014	0.011	-0.003
Syncerus caffer nanus	African forest buffalo	0.026	0.018	-0.008
Taurotragus oryx	Eland	0.021	0.011	-0.009
Tragelaphus angasii	Nyala	0.023	0.011	-0.012
Tragelaphus imberbis	Lesser Kudu	0.028	0.020	-0.008
Tragelaphus scriptus	Bushbuck	0.029	0.013	-0.016
Tragelaphus strepsiceros	Greater Kudu	0.019	0.009	-0.009
Capra ibex	Ibex	0.142	0.104	-0.039
Oreamnos americanus	North American mountain goat	0.100	0.070	-0.030
Ovibos moschatus	Muskox	0.027	0.018	-0.009
Ovis aries	Sheep	0.004	0.004	0.000
Ovis canadensis	Bighorn sheep	0.056	0.044	-0.012
Rupicapra rupicapra	Chamois	0.016	0.009	-0.007
Sylvicapra grimmia	Common duiker	0.022	0.015	-0.007
Addax nasomaculatus	Addax	0.009	0.005	-0.004
Hippotragus equinus	Roan antelope	0.018	0.012	-0.006
Hippotragus niger	Sable antelope	0.009	0.006	-0.004
Oryx beisa	East African oryx	0.031	0.022	-0.009
Oryx dammah	Scimitar oryx	0.008	0.005	-0.003
Oryx gazella	Gemsbok	0.038	0.028	-0.011
Oryx leucoryx	Arabian oryx	0.009	0.006	-0.004
Kobus ellipsiprymnus	Waterbuck	0.024	0.017	-0.007
Kobus leche	Lechwe	0.038	0.027	-0.011
Pelea capreolus	Rhebok	0.019	0.014	-0.005
Redunca arundinum	Southern Reedbuck	0.011	0.008	-0.003
Redunca fulvorufula	Mountain Reedbuck	0.038	0.029	-0.008
Vicugna pacos	Alpaca	0.168		
Alces alces	North American moose	0.088	0.019	-0.069
Odocoileus virginianus	White-tailed deer	0.036	0.027	-0.009
Rangifer tarandus	Caribou	0.068	0.041	-0.027
Axis axis	Axis deer	0.088	0.077	-0.012
Cervus canadensis	Rocky mountain elk	0.022	0.014	-0.008
Cervus nippon	Sika deer	0.049	0.037	-0.012
Dama dama	Fallow Deer	0.027	0.018	-0.010

Loxodonta africana	Savanna elephant	0.107		
Equus caballus	feral horse	0.041		
Giraffa camelopardalis tippelskirchi	Masai Giraffe	0.012	0.007	-0.004

Table S5 in publication.

Table 2.6. List of samples used to estimate allele and haplotype frequencies for the estimation of Reynolds genetic distances.

Breed	Number of Animals
Angus	2056
Red Angus	15
Murray Grey	4
Red Poll	5
Shorthorn	76
Lincoln Red	8
Ayrshire	440
Holstein	1308
Hereford	122
Guernsey	23
Jersey	78
Simmental	78
Brown Swiss	24
Limousin	1210
Charolais	54
Piedmontese	29
Romagnola	29
Hanwoo	48
Wagyu	49
N'Dama	59
Nelore	68
Gir	30
Total	5813

Table S6 in publication.

3. A NOVEL ANALYTICAL METHOD DETECTS RESPONSE OF THE ANGUS (*BOS TAURUS*) GENOME TO ARTIFICIAL SELECTION ON COMPLEX TRAITS

Jared E. Decker[1], Daniel A. Vasco[1,2], Stephanie D. McKay[1], Matthew C. McClure[1,3], Megan M. Rolf[1], JaeWoo Kim[1], Sally L. Northcutt[4], Stewart Bauck[5], Brent Woodward[5], Robert D. Schnabel[1], Jeremy F. Taylor[1]

[1]Division of Animal Science, University of Missouri, Columbia, MO 65211, USA

[2]Biology Department, Duke University, Durham, NC 27708, USA

[3]Bovine Functional Genomics Laboratory, ARS, USDA, Beltsville, MD 20705, USA

[4]American Angus Association, 3201 Frederick Ave, Saint Joseph, MO 64506, USA

[5]Igenity Livestock Business Unit, Merial Limited, Duluth, GA 30096, USA

Abstract

Background

Several methods have recently been developed to identify selective sweeps within genomes. However, recent theoretical and empirical work suggests that polygenic models are required to identify the genomic regions that have responded to selection on complex traits. Using DNA samples from US registered Angus beef bulls born over a 50 year period, we examine the effects of selection on the genome of this breed. We

present results from the application of a quantitative genetic model to identify signatures of recent ongoing selection.

Results

We show that US Angus cattle have been selected to systematically alter their mean additive genetic merit for almost all of the 16 production traits routinely recorded by breeders. We further estimate the time-dependency of allele frequency for 44,817 SNP loci using genomic best linear unbiased prediction, BayesCπ, and generalized least squares. Finally, we reconstruct the primary phenotypes that have historically been exposed to selection from a genome-wide analysis of the 16 production traits and gene ontology enrichment analysis.

Conclusions

We demonstrate that polygenic quantitative genetic models correct for sampling effects which lead to time-dependent pedigree stratification and reveal genomic signatures of ongoing selection. Because multiple traits have historically been simultaneously selected and most quantitative trait loci have small effects, selection has incrementally altered allele frequencies throughout the genome. Two QTL of large effect were not among the most strongly selected loci due to their antagonistic pleiotropic effects on strongly selected phenotypes. Our method may readily be extended to temporally-stratified human or model organism populations.

Keywords

selective sweep, polygenic model, pleiotropy, cattle, phenotype, genotype

Background

Several statistical tests have recently been developed to identify the genomic regions that have been subjected to strong recurrent selection. Most have been based on extreme population differentiation (Akey et al. 2002; Shriver et al. 2004; Weir et al. 2005), the enrichment of rare mutations in the site frequency spectrum (Carlson et al. 2005; Kelley et al. 2006), or patterns of extended haplotype homozygosity (Sabeti et al. 2007; Voight et al. 2006; Wang et al. 2006). See (Akey 2009; Hancock, Alkorta-Aranburu, et al. 2010) for further review. These tests have now been used to detect molecular signatures of selection in cattle (Gautier and Naves 2011; Gibbs et al. 2009; Hayes et al. 2009; Hayes et al. 2008; Qanbari et al. 2011; Qanbari et al. 2010). However, recently, there has been a call to employ polygenic models to simultaneously identify loci responding to selection but that do not fit the typical "hard sweep" paradigm (Pritchard and Di Rienzo 2010; Pritchard, Pickrell, and Coop 2010).

Concurrent with the development of new approaches for the detection of selective sweeps, the statistical models employed for genome-wide association studies have been improved. Some of the refinements deal with the effects of population structure and kinship between sampled individuals, since not accounting for these effects can significantly increase the number of false positive associations; see (Price et al. 2010) for review. Furthermore, there has also been a shift toward the application of

polygenic models for the identification of genetic risk factors and variants associated with complex phenotypes (McClure et al. 2012; Yang et al. 2010).

In this study, we merge the search for loci responding to selection with advanced genome-wide association models to quantify the genome-wide response to selection in US registered Angus cattle. By so doing, we answer the call for the application of polygenic models for the detection of genomic imprints of selection.

Results

Evidence of selection

Deregressed estimated breeding values (EBV) (Garrick, Taylor, and Fernando 2009) for 16 production traits (see Supplementary Information for definitions) were regressed on birth date (measured as a continuous variable with month and day converted to a decimal fraction of a year) for 3,570 registered Angus cattle (Table 3.7, Figure 3.1 and Figures 3.7-3.20). For traits that can easily be appraised and for which expected progeny differences (EPD, one half of the EBV) were implemented earlier in the development of the breed (e.g., growth and stature), selection has significantly changed the breed additive genetic mean over time. For traits for which increased production has been consistently desirable, such as weaning weight, yearling weight, and milk, additive genetic means have increased linearly (Table 3.7; Figures 3.1b, 3.8, and 3.14). However, additive genetic means for birth weight, yearling height, mature weight, and mature height increased until the mid-1980s when breeders recognized the detrimental effects of large birth weights on calving difficulty and large mature size on cow

maintenance requirements and fertility, and these traits were subsequently selected to decrease (Figures 3.1a, 3.9, 3.15, and 3.16). For these traits, the quadratic regression models have a much smaller Akaike Information Criteria (AIC), larger adjusted R^2 values, and smaller p-values (Table 3.7). Traits for which EPDs have only recently been developed and made available to breeders, such as docility and heifer pregnancy rate, show little change in mean additive genetic merit over time (Figures 3.11 and 3.12). Docility and heifer pregnancy rate had among the smallest of R^2 values of all the fitted linear and quadratic regression models. Mean additive genetic merit for all of the growth traits (WW, YW, and CW) and for the incidence of unassisted births (CED and CEM) has increased annually. Weaning weight has increased, on average, by 2.81 pounds per year and the rate of unassisted births (CED) has increased by 0.56% per year – remarkable achievements by Angus breeders considering the 50 year span of these data.

Signatures of Selection

Here we introduce a novel method for identifying loci that are responding to ongoing selection. Selection induces changes in allele frequency for the selected mutation, as well as for neighbouring loci which hitchhike along with the selected loci due to the presence of linkage disequilibrium between the loci. Accordingly, individual allele frequencies (1, 0.5, or 0 for *AA*, *AB*, and *BB* individuals) could simply be regressed on birth date to identify loci that have rapidly changed in frequency over time. However, in the presence of any sampling bias that manifests as a nonrandom ascertainment of

family members in time, this approach suffers from a high false-positive rate of detection of loci subject to selection (data not shown). The bias results from a pedigree-based stratification in the depth of sampling of DNA on individuals within different families and differences in allele frequencies between families such that the differences in allele frequencies between families are partially confounded with differences in allele frequencies in time. In other words, this approach is confounded by pedigree relationships and the nonrandom sampling of families at different time points. In our approach, rather than regressing allele frequencies on birth year, we invert the relationship and fit birth year to a model that includes individual additive genetic merits and simultaneously estimate allele substitution effects (ASEs) for all fitted single nucleotide polymorphisms (SNPs) (McClure et al. 2012). With the analysis framed from this perspective, we can identify the SNPs most strongly associated with differences in birth date while accounting for kinship within the sample through the use of the genomic relationship matrix in the mixed model equations (Price et al. 2010).

We estimated ASEs for 45,073 SNPs that predict birth date for 3,570 registered Angus cattle using genomic best linear unbiased prediction (GBLUP) (McClure et al. 2012; VanRaden 2008; Saatchi et al. 2011), but do not report results for the 256 SNPs that map to unassigned contigs in the UMD3.1 reference assembly (Zimin et al. 2009). Allele substitution effects were converted to estimates of additive genetic variance associated with each SNP (See Methods). Although selection has caused only small changes in allele frequency at most loci, some loci, in particular on chromosomes 1, 10,

23 and 29, have achieved strong selection responses (Figure 3.2). The two peaks on chromosome 23 contain the MHC (BoLA) and numerous olfactory receptors.

We also used the generalized least squares mixed model framework implemented in EMMAX to analyse birth date. EMMAX estimates SNP ASEs as fixed effects for each marker individually for which p-values can also be estimated, whereas GBLUP simultaneously fits all markers as random effects and does not estimate p-values. This analysis also predicted that most of the loci had small responses to selection. However, after correcting for multiple testing using FDR procedures (Storey and Tibshirani 2003), loci on chromosomes 1, 2, 3, 6, 20, 21, 22, 23, 24 and 29 were significant at FDR < 0.1 and loci on chromosomes 1, 2, 5, 7, 10, 14, 25, 26, 27, 28 and X were suggestive at FDR < 0.25 in their statistical significance for response to selection (Figure 3.3).

Finally, we used GenSel (Fernando and Garrick 2012) to fit a non-linear BayesCπ model (Habier et al. 2011) in which the parameter π estimates the proportion of SNPs that are not associated with the trait. We estimated π to be 0.978856, and thus 2.11% (948) of the SNPs were predictive of birth date and therefore putatively exposed to strong selection. BayesCπ employs a MCMC approach in which 1-π of the SNPs are sampled for inclusion in the model in each chain and estimated SNP ASEs are finally shrunken according to the proportion of times each SNP is retained in the selected model. Thus, SNPs that are rarely retained in the model have their ASEs strongly

regressed towards zero. This analysis revealed large peaks on chromosomes 1, 2, 3, 7, 8, 10, 11, 14, 15, 17, 19, 22, 23, 24, 25, 26, 28, 29 and X (Figure 3.4).

The SNP ASEs estimated with GBLUP and with EMMAX were quite similar with a Spearman correlation of 0.9238 and a Pearson correlation of 0.7782. The difference in ASE magnitude identified by the Pearson correlation reflects the difference in fitting SNPs as random or fixed effects. The SNP ASEs estimated by GBLUP and BayesCπ also ranked similarly with a Spearman correlation of 0.8507, but had a Pearson correlation of only 0.5553 likely due to the strong shrinkage of small effect SNPs in BayesCπ. The EMMAX and BayesCπ ASEs had a Spearman correlation of 0.8280 and Pearson correlation of 0.4624. The GBLUP and EMMAX analyses estimated the heritability of birth date to be 0.5336 and 0.5314, respectively, using restricted maximum likelihood estimation of variance components. The BayesCπ analysis estimated the heritability of birth date to be 0.7169.

Effective Population Size and Drift

To demonstrate that drift has a very limited effect on allele frequency changes in time within the artificially selected US Angus breed, we estimated the effective population size, under the neutral model, and modelled the effects of drift on neutral loci. Using a pedigree of up to 63 generations and which comprised 91,001 Angus animals including the 3,570 genotyped cattle and all known ancestors, we estimated the generation interval for US Angus sires to be 4.99 years, which was the average age of sires born between 1941 and 1990 at the birth of their male and female registered progeny. From

this pedigree, we also estimated inbreeding coefficients (denoted as *F*) for all animals from which we estimated effective population size from the regression of *F* on generation. From a principal component analysis of the SNP genotypes, we identified two distinct subgroups within our sample. In Figure 3.21, we identified the Wye Angus herd (Anon.) which was formed from an importation of bulls from the British Isles and then closed to new germplasm in 1958 as a group that was distinct from the remaining US registered Angus cattle. The inbreeding effective population size for the Wye herd was estimated to be 36.41±0.03, whereas the effective population size for the remaining North American Angus was estimated to be 267.59± 0.02 using animals born after 1930 and 116.15± 0.04 using animals born after 1980 (Figure 3.5a and Table 3.1). For each of the 44,817 SNPs, we constructed a test (see Methods) to determine whether the observed change in allele frequency could be explained by drift or was due to selection. From this analysis, we found that the observed allele frequency changes exceeded the likely changes due to drift for 84.60% of the 44,817 SNPs.

We also compared genomic with pedigree estimates of *F*. The realized genomic *F* have a larger variance (s^2= 0.0023) than the pedigree *F* (s^2= 0.0014) and the two measures of *F* were only weakly correlated with a Pearson correlation of 0.648 (Figure 3.5b) which is consistent with the underestimation of pedigree *F* due to the assumption that *F* is zero for all animals in the base generation, pedigree errors, and incomplete pedigree information. We regressed pedigree *F* on genomic *F*, and found the slope of the regression to be 0.49±0.01. Separate regressions for the Wye and North American

Angus revealed pedigree and genomic *F* coefficients to be more similar for the Wye herd than for the remaining North American Angus cattle (see Figure 3.5b and Table 3.2).

Because genomic *F* estimates require fewer assumptions, we also calculated N_e using the genomic *F* coefficients of North American Angus born after 1980. This resulted in a N_e of 94.18±0.10 (Table 3.1). Using this N_e in our drift test, we estimated that allele frequency changes exceeded the likely changes due to drift for 82.41% of the 44,817 SNPs.

Connecting Selected Phenotype to Selected Genotype

Using GBLUP, additive genetic variation is partitioned into contributions from individual loci accounting for the extent of linkage disequilibrium between the loci. We analysed deregressed EBVs in a weighted analysis (Garrick, Taylor, and Fernando 2009) for 16 production traits (Supplementary Information) using data provided by the American Angus Association (AAA) under an animal model which incorporated a genomic relationship matrix and from which we estimated the proportion of additive genetic variance explained by the SNP markers (Table 3.3). Genetic correlations between traits were estimated as the correlations between SNP ASEs for pairs of traits (Table 3.4). With the exception of two QTL on chromosomes 7 and 20, most genes influencing variation in growth traits in Angus cattle are of small effect (Figures 3.6, 3.22-3.36). The most likely location of the pleiotropic QTL on chromosome 7 was found to be at 93.22 Mbp in the GBLUP analyses. However, the largest effect birth date QTL on this chromosome was found at 99.02 Mbp (Figure 3.2) by GBLUP, at 100.02 Mbp by BayesCπ

(Figure 3.4), and a small, but not significant, birth date QTL was found at 99.02 Mbp in the EMMAX results (Figure 3.3). The SNP at 93.22 Mbp on chromosome 7 was ranked 11,224 out of the 44,817 SNP effects for birth date (74^{th} sample percentile, i.e. the ASE for this SNP was larger than 74% of all SNP ASEs). The most likely location of the pleiotropic QTL on chromosome 20 was estimated to be at 4.62 Mbp and QTL signals were detected at 5.1 Mbp in the GBLUP and at 5.9 Mbp in the EMMAX analyses of birth date. The SNP at 4.62 Mbp on chromosome 20 was ranked 5,168 of the 44,817 SNP effects for birth date (88^{th} ASE percentile).

To assess the identity of the trait or combination of traits that have historically been under selection in Angus cattle and that produced the molecular signals of selection, we simultaneously regressed the SNP ASEs for birth date on the corresponding standardized SNP ASE multiplied by the allele frequencies ($pqASE/\sigma_{ASE}$) for all 16 production traits for which the AAA routinely produces EPDs (Table 3.5). This model yielded an adjusted R^2 of 0.3148 and the partial regression coefficients yield estimates of relative selection intensities for which those for weaning weight, calving ease direct, and milk were the largest. The relative selection intensities for mature weight, mature height, fat thickness, and ribeye muscle area were not significant (Bonferroni corrected P> 0.0029). We also fit this model for the 948 SNPs with the largest birth date ASEs, and which yielded an adjusted R^2 of 0.7268 (Table 3.6). Again, weaning weight, milk, and calving ease direct had the largest relative selection intensities. However, the partial regression coefficients for yearling weight, fat thickness, carcass weight, scrotal circumference, heifer pregnancy, mature weight,

mature height, docility, and ribeye area were not significant (Bonferroni corrected P> 0.0029). Table 3.4 shows that ASEs for weaning weight, yearling weight, milk, calving ease maternal, carcass weight, marbling and calving ease direct had the largest pair-wise correlations with ASEs for birth date.

Finally, to elucidate the biological processes associated with the genes located in the genomic regions detected to be under selection, we estimated the gene ontology term enrichment for the annotated genes within these regions (See Additional Files 2 and 3). From the GBLUP results we queried 4,216 genes within 250 Kbp of the top ranked 948 SNPs for their birth date ASEs, and from the BayesCπ results we queried 4,033 genes within 250 Kbp of the top 948 SNPs. There were 1,223 genes shared between the two lists. Various biological processes appear to be under selection based on the GBLUP results - notably growth and metabolic processes at level 1; regulation of cellular component biogenesis, biosynthetic processes, organ growth, cell proliferation, and molting cycle at level 2; and organic acid metabolic processes, protein metabolic processes, and vitamin metabolic processes at level 3; antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, lymphocyte activation, and leukocyte activation at level 4 (Additional File 2 contains the complete list). The GBLUP results also found the MHC class II protein complex to be an enriched cellular component, olfactory receptor activity to be an enriched molecular function, and olfactory transduction to be an enriched KEGG pathway. The BayesCπ analysis found the enriched biological processes to include developmental processes, cellular processes, and biological regulation at level 1; embryonic development, anatomical

structure development, and anatomical structure morphogenesis at level 2; lipid transport, response to oxidative stress, embryonic morphogenesis, appendage morphogenesis, palate development, and cellular response to stress at level 3; and spermatid development, and eating behaviour at level 4. The MHC class II protein complex and PML body (viral infections induce changes in PML (Anon.)) were also found to be enriched cellular components, somatotropin/prolactin gene family members were enriched, and placenta and fetal muscle were found to be enriched tissues. The following heat shock genes were inferred to be under selection: *HSP90AB1*, *HSPA12A*, *HSPA1A*, *HSPA1B*, *HSPA1L*, *HSPA4L*, *HSPB3*, *HSPD1*, *HSPE1*, and *HSPBAP1*.

Discussion

Most of the phenotypes routinely recorded in Angus cattle have been under directional selection in the recent history of the breed. Artificial selection has increased the weights at which cattle are marketed either at weaning or yearling ages (Figures 3.2, 3.8, and 3.17) while simultaneously decreasing the incidence of assisted births (Figures 3.7 and 3.13). Larger birth weights and yearling heights are both strongly associated with increased calving difficulty (Table 3.4) and genetic trend in both traits increased until about the mid-1980s, after which both began to decrease (Figures 3.2 and 3.9). Birth weight was not directly selected by breeders to increase, but increased as a correlated response to selection for increased weaning and yearling weights. However, yearling height was actively selected to increase by some breeders to make Angus cattle more comparable in frame size to the Continental European breeds which were imported into

the US during the 1970s. Once the undesirable correlated response in calving ease became appreciated by breeders, selection was practised to increase weaning and yearling weights while maintaining birth weight and yearling height constant.

Using EMMAX, only eleven loci were found to be significantly associated with birth date, but all loci simultaneously explained 53% of the variation in birth date. BayesCπ estimated that 2.11% of the SNPs were associated with birth date, and produced an estimated heritability of 0.72. The difference in heritability estimates between the GBLUP and EMMAX analyses compared to the BayesCπ analysis is likely due to the fact that GBLUP and EMMAX assume the infinitesimal model in which all SNP ASEs are drawn from a distribution with constant variance and, thus, regress all effects equally towards the mean of zero. On the other hand, BayesCπ begins with a distribution with constant variance but shrinks variance for SNPs rarely fit in the model. This results in much less of a regression for large ASEs which may lead to larger estimates of the additive genetic variance – as was found here - when there are large effect loci underlying variation. In the absence of selection, genotype should be independent of time provided that the effects of drift are negligible. In this case, the infinitesimal model should apply with all SNP ASEs being small leading to a small estimate of heritability. However, this was not the case in US registered Angus cattle and we conclude that a few variants are rapidly responding to selection (our results suggest 2.11%) and that most of the genome (82.4% from the drift analysis) is responding more slowly to selection. Therefore, these analyses strongly support the

infinitesimal model under which selection is expected to produce small changes in allele frequency at a large number of loci, all of small individual effect.

The SNP ASEs for the 16 analysed traits indicate that, with the exception of the two large effect QTL on BTA7 and 20, the vast majority of QTL underlying quantitative traits in beef cattle are of small effect. Of considerable interest, neither of these QTL was found to be under strong selection and this seems to be because of their large antagonistic pleiotropic effects on growth and calving difficulty. When multiple traits are simultaneously selected, the genetic architecture of the population defined by the chromosomal organization of QTL alleles constrains both the phenotypic and genotypic response to selection.

For selection to be effective, the selection intensity and effective population size must be sufficiently large to overcome the effects of genetic drift. We demonstrate that US registered Angus cattle have a sufficiently large effective population size to enable successful artificial selection, but more importantly, that large generational changes in allele frequency are unlikely to occur due to drift alone. Furthermore, we found a considerable disparity between pedigree and genomic estimates of inbreeding coefficients. While others have argued that genomic relationship matrices should be adjusted to more closely resemble pedigree relationship matrices (Powell, Visscher, and Goddard 2010), we assert that genomic relationship matrices provide a more accurate description of the realized relationships among individuals which result from the Mendelian sampling of parental gametes and selection. The use of genomic relationship matrices in place of pedigree relationship matrices avoids the assumption of neutrality

of loci both in the estimation of inbreeding coefficients and the mean value of gametes inherited by progeny - both of which are assumed for the computation of the numerator relationship matrix (Quaas, Anderson, and Gilmour 1984). The disagreement between genomic *F* and pedigree *F* is likely to be due to the assumption that base animals are not inbreed, errors in the pedigrees, and missing pedigree information likely due to the large-scale importation of Canadian Angus cattle in the 1940s and 1950s which were not carriers of dwarfism alleles which had been driven to high frequency due to selection at the time (Baker, Blunn, and Plum 1951). This is supported by the closer agreement between pedigree and genomic *F* coefficients for the Wye herd derived from British stock and with more complete pedigree records than the remaining US registered Angus animals (Table 3.2 and Figure 3.5b).

We attempted to identify the relative selection intensities placed on each selected trait via the imprints that multi-trait selection had left on the Angus genome. Although this analysis assumed no change in relative selection intensities in time, an assumption that is clearly violated in view of the genetic trends in birth weight and yearling height, we were able to confirm that growth traits have historically been under the most strongly selected in US registered Angus cattle. Because Angus is considered to be a maternal breed (i.e., motherly, used as dams in commercial beef production), it is logical that loci which influence calving ease, growth to weaning and milking ability should have been found to be under selection. Angus breeders have successfully selected to increase calving ease and body weight by selecting for body shapes that allow a calf's easy passage through the dam's pelvis. This is supported by the finding of

an enrichment of gene ontology terms related to skeletal development, skeletal morphogenesis, limb development, limb morphogenesis, and palate development within regions of the genome detected as responding to selection. We note that palate development is closely related to face and skeletal system morphogenesis (Anon. 2011). It has previously been shown that calving ease is negatively correlated with several body measures, such as head circumference, head width, hip width, hip height, heart girth, and cannon bone circumference (Bureš et al. 2008; Nugent, Notter, and Beal 1991; Wall et al. 2005). We also observed that the somatotropin/prolactin family of genes was enriched, due to selection for increased growth and milk production (Additional File 3). Conversely, traits such as fat thickness, mature height, ribeye muscle area, docility, and heifer pregnancy rate have not been as intensely selected as growth traits, probably due to the breeding objectives of beef producers, genetic antagonisms constraining selection response in these traits, and the historic difficulty in collecting field data to allow the development of EPDs for these traits.

There is also strong evidence that natural selection has occurred in this population. The regression of birth date ASEs on $pqASE/\sigma_{ASE}$ coefficients for each trait suggests that natural selection was responsible for 27% of the variation in the response of the 948 loci most rapidly changing SNPs. Hair growth and shedding are under selection in Angus cattle, likely due to their introduction into new climates (Anon.). The gene ontology enrichment results also indicated that genes affecting immune response, such as the major histocompatibility complex, have strongly responded to selection (Figures 3.2, 3.3, and 3.4, Additional Files 2 and 3), presumably due to the exposure of

Angus cattle to novel pathogens following their introduction to the US in 1873 (Anon.) and a continuous co-evolutionary "arms race" with bovine pathogens (Stavrinides, McCann, and Guttman 2008; Walker and Roberts 2009). We also observed that olfactory receptor loci have strongly responded to selection. The Bovine Genome Sequencing and Analysis Consortium (Elsik et al. 2009) found that olfactory receptors were commonly duplicated in the bovine genome. Our results suggest that many of these retained duplications remain under direct selection. Furthermore, the Bovine HapMap Consortium (Gibbs et al. 2009) found that *ZNF187*, which is expressed in olfactory tissues, and the MHC had some of the lowest F_{st} values in the entire genome when compared between breeds. Our BayesCπ results identify *ZNF187* as a positively selected gene, and all analyses identified the MHC as being under selection. Thus, the response to selection on BTA23 may be common across cattle populations, causing small F_{st} values between breeds (convergence) but large changes in allele frequency over time (divergence).

Furthermore, natural selection may be buffering against the deleterious effects of inbreeding. We found that seminal plasma proteins, spermatid development and related gene ontology terms were enriched within the strongly selected regions of the genome (Additional Files 3). Seminal plasma proteins have been associated with bull fertility (Killian, Chapman, and Rogowski 1993). Selection may have acted on these loci to offset inbreeding depression in fertility. Genes involved in response to oxidative stress were also identified (Kristensen et al. 2010). We inferred that 10 heat shock proteins are under selection in Angus. It is hypothesized that heat shock proteins help

the organism cope with protein instability and misfolding caused by nonsynonymous mutations which occur as homozygotes at elevated rates due to inbreeding (Kristensen et al. 2010; Ayroles et al. 2009; K. S. Pedersen, Kristensen, and Loeschcke 2005; Sorensen, Kristensen, and Loeschcke 2003; Cheng et al. 2006).

One of the difficulties encountered in identifying genomic signatures of selection is in distinguishing changes that have occurred due to demographic forces as opposed to selective forces. Our mixed model approach specifically accounts for pedigree relationships and explicitly deconvolutes their confounding effects on time-dependent allele frequency changes, which are due to the fact that not all pedigrees are sampled equally deeply in terms of the numbers of genotyped individuals. However, one of the limitations of our approach is the requirement of a temporally stratified sample of genotyped individuals. This will currently limit its application in human populations due to a lack of preserved samples across multiple generations. However, this limitation may be alleviated as it becomes more practical to extract quality DNA from formalin-fixed, paraffin-embedded tissue section samples and ancient remains. Nevertheless, the approach is clearly most easily applied to model organisms for which temporally stratified samples are available. In addition to birth date, environmental variables such as diet composition, latitude, rainfall and temperature measures, could be fit as the dependent variable in a mixed model analysis to identify loci associated with environmental adaptation (Hancock, Alkorta-Aranburu, et al. 2010; Hancock, Witonsky, et al. 2010; Hancock et al. 2008). Phenotypes similar to those used in the works of

Hancock *et al.* could be analysed, but the statistical model would differ and populations would need to be more closely related.

Using the estimated ASEs as informative priors in the development of genomic selection programs (Eggen 2012) is another interesting application of our method. The loci with smaller ASEs for birth date are either of small effect on the selection objective or have undesirable pleiotropic effects. Loci that have larger ASEs for birth date have larger effects on the selection objective which are less constrained by antagonistic pleiotropic effects or closely linked loci with antagonistic phase relationships allowing them to more rapidly respond to selection.

Conclusions

We demonstrate that selection on polygenic traits that approximate the infinitesimal model leaves detectable signatures of selection in the genome that also are polygenic and infinitesimal in nature. If genes with large antagonistic pleiotropic effects exist, they respond to selection as if they were of small effect. By relating the detected signatures of selection to phenotype, we infer that artificial selection in US registered Angus cattle has historically focused primarily on growth and maternal traits including calving ease, weaning weight and milking ability. This result is directly confirmed by the response to selection in these traits estimated directly using EPDs estimated by the AAA. Finally, we show that natural selection has acted in this domesticated population to increase immunity and possibly buffer against inbreeding depression.

Methods

DNA extraction and SNP genotyping

Cryopreserved semen was obtained from semen distributors, the National Animal Germplasm Program, and individual Angus breeders including the University of Maryland who own the Wye herd. DNA was extracted using a proteinase K digestion, Phenol:Chloroform alcohol extraction, and ethanol precipitation (Sambrook, Fritsch, and Maniatis 1989). Single nucleotide polymorphisms were assayed using the Illumina BovineSNP50 BeadChip (Matukumalli et al. 2009) and genotyped using the Illumina GenomeStudio software. Genotypes were filtered using a SNP call rate cut off of 90%, animal call rate cut off of 95%, and minor allele frequency threshold of 0.01. Autosomal and pseudoautosomal SNPs that had a Hardy-Weinberg Chi-square statistic with 1 degree of freedom greater than 300 were also filtered – primarily to remove polymorphisms detected in CNVs rather than remove loci that were under selection (Saatchi et al. 2011). Filtered data were processed through FastPHASE version 1.4.0 (Scheet and Stephens 2006) to impute the 0.49% of missing genotypes. Parameter values were set at T=10, K=20, with -eo flags set. The resulting dataset consisted of genotypes for 45,073 SNPs scored in 3,570 animals with no missing values.

Response to selection

Expected progeny differences for 16 production traits along with their accuracies were provided by the AAA for 103,816 animals including the 3,570 genotyped bulls and all identified ancestors in their pedigrees. These values were doubled to obtain estimated

breeding values that were deregressed for the 3,570 bulls as previously described (Garrick, Taylor, and Fernando 2009). The deregression of estimated breeding values removes parent average information and converts the information available on the individual back to the scale of the underlying phenotype – that is, it removes the "shrinkage" that was applied to convert phenotypes into estimated breeding values. In the statistical package R (R Development Core Team 2011), trait breeding values were plotted against birth date measured as a continuous variable. Linear and quadratic regressions were fit for each trait.

Principal component analysis of Angus genotypes

We used the *smartpca* program, part of EIGENSOFT (Patterson, Price, and Reich 2006), for principal component analysis of the Angus genotypes. We plotted principal component 1 by principal component 2 to visualize the largest elements of population substructure. Figure 3.21 revealed that the primary substructure detected in the population was the largest families – the linearly related members of the Wye herd and the ancestors and sons of N Bar Emulation EXT, a popular bull within the breed that generated numerous sons who were employed in AI.

Estimation of effective population size

The inbreeding effective population size N_e was estimated from the regression of inbreeding coefficients on pedigree generation number using the individual animal data. This requires inverting the relationship $\Delta F = 1/2N_e$, in which ΔF is the increase in mean inbreeding coefficient between adjacent generations (Falconer and Mackay 1996) and is

estimated as the slope of the regression across all generations if N_e is assumed constant in time. A Taylor series expansion leads to an estimate of the standard error of N_e as $SE(N_e) \approx 2N_e SE(\Delta F)$, in which $SE(\Delta F)$ is the standard error of the estimated slope of the regression. Because the depth of available pedigree information varied substantially for the 3,570 sampled Angus animals (animals within the pedigree that were assigned to generation 0 varied in birth year from 1838 to 1954) we considered the estimates of pedigree generation to be unreliable from the perspective of estimating N_e. Accordingly, we estimated generation number for each of the 3,570 genotyped animals by subtracting 1950 from birth year and dividing by the generation interval of 5 years. Because of the closed nature of the Wye herd and complete pedigree information back to foundation animals, we fit separate models for the Wye and remaining North American Angus animals. For the North American Angus subset we fit two models using generation number estimated from birth year for animals born after 1930 and for animals born after 1980 where there appeared to be an inflection in the rate of increase in inbreeding. This corresponds to the point in time at which the uptake in use of AI became significant within the breed.

For each of the 44,817 SNPs, we directly estimated the change in allele frequencies that occurred between the 460 individuals assigned to pedigree generation 58 and the 450 individuals assigned to pedigree generation 59 using PLINK (Purcell 2009; Purcell et al. 2007). These pedigree generations were chosen because they represent the individuals with the deepest pedigrees which are therefore not significantly

influenced by missing pedigree information and also because they are among the most recent generations and are likely to represent all of the families present within the sample. Furthermore, the sample sizes for these generations are sufficiently large to obtain precise estimates of allele frequencies. We compared the allele frequency changes between generations 58 and 59 to the bounds of the 99.999999% ($-\log_{10}(p\text{-value}) = 8$) confidence interval for the change in allele frequency due to drift (estimated under the assumption of normality assuming a mean of 0 and the drift variance for the *ith* SNP to be $p_i(1-p_i)/2N_e$, for p_i the frequency of the A_i allele in generation 58 and N_e = 116.15 (Falconer and Mackay 1996)). For SNP on the X chromosome, the drift variance for the *ith* SNP is $p_i(1-p_i)/1.5N_e$. Loci for which the allele frequency change exceeded the boundaries of the confidence interval were concluded to be changing in frequency due to selection rather than drift.

Genomic BLUP of phenotypic traits

In a weighted analysis using deregressed EBVs as previously described (Garrick, Taylor, and Fernando 2009), genomic BLUP (VanRaden 2008) was used to estimate ASEs for 16 different traits using 45,073 SNPs genotyped in 3,570 animals. Allele substitution effects were converted to additive genetic variances by squaring the ASE and multiplying by $2p_iq_i$, in which p_i and $q_i = 1 - p_i$ are the base generation allele frequencies at the i^{th} SNP (VanRaden 2008). Base generation allele frequencies at each SNP were estimated using the 59 animals born between 1955 and 1974, excluding animals from

the Wye herd. Results are presented only for the 44,817 SNPs that mapped to autosomes or the X chromosome in the UMD3.1 bovine assembly.

Signatures of selection analysis

Genomic BLUP was also used to estimate ASEs for birth date. SNPs with the greatest changes in allele frequency over time will have the largest ASEs for birth date. The ASE reflects the amount of response to selection realized by the genomic region tagged by a SNP.

Genome-wide associations with birth date were also analysed using EMMAX (Kang et al. 2010). A Balding-Nichols matrix (Balding and Nichols 1995) was computed and used in EMMAX as the kinship matrix. Resulting p-values were adjusted to q-values using the method of Benjamini and Hochberg (Benjamini and Hochberg 1995) as implemented by the GenABEL package in R (Aulchenko et al. 2007).

GenSel (Fernando and Garrick 2012) was used to fit a BayesCπ model (Habier et al. 2011) to estimate π and the allele substitution effects for each SNP. The additive genetic variance and residual variance estimated from GBLUP were used as priors in GenSel. The π prior was set at 0.9. GenSel was run for 160,000 iterations, with 1,000 iterations as burn-in.

From Falconer and MacKay (Falconer and Mackay 1996), the change in allele frequency resulting from selection is $\Delta q = -ipqa/\sigma_p$, where i is the selection intensity, a is one half the phenotypic difference between homozygote mean phenotypes, σ_p is the trait variance, and p and q are allele frequencies. Assuming the dominance deviation is

zero, the ASE α is equal to the genotypic value *a*. Thus, we use the ASE as a proxy for *a* which we then scaled as $pq\text{ASE}/\sigma_{ASE}$ to form the independent variables for each of the 16 production traits which were jointly regressed on the birth date ASEs to provide estimates of the relative selection intensity *i* for each trait (the sign is included in the realized estimate). For each trait the ASEs were standardized by dividing by the ASE standard deviation (σ_{ASE} in the equation above). A model was also fit which included birth date ASEs for only the top ranked 948 SNPs.

Functional annotation

Due to the significant extent of linkage disequilibrium within the bovine genome (Gibbs et al. 2009; McKay et al. 2007), we identified all genes within 250 Kbpof the 948 SNP (top 2.11% of the 44,817 SNP, equal to 1-π estimated from GenSel) with the largest additive genetic variances for birth date estimated from the linear GBLUP and nonlinear BayesCπ analyses. We used the DAVID bioinformatics resources (Huang, Sherman, and Lempicki 2009b; Huang, Sherman, and Lempicki 2009a) to identify enriched GO terms in the list of 4,214 genes identified from the GBLUP results and 4,033 genes from the BayesCπ results. We used annotations from *Bos taurus*, *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Canis lupus*, *Pan troglodytes*, *Macaca mulatta*, *Equus caballus*, *Pongo abelii*, *Sus scrofa*, *Ovis aries*, and *Oryctolagus cuniculus* for GO enrichment analysis.

List of Abbreviations

List of Abbreviations are included in the Supplementary Material or are defined in text.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JFT and JED designed the experiment. DAV, JED and JFT analyzed data. SDM, MCM, MMR, JWK, and RDS extracted DNA. MMR and SDM prepared samples for genotyping, SDM ran the Illumina assay, and RDS genotyped samples and managed the genotype database. SLN provided pedigree and estimated genetic merit data and SB and BW provided genotypes on about 900 Angus bulls. JED and JFT wrote the manuscript and other authors provided feedback.

Acknowledgements

We gratefully acknowledge the provision of semen samples from Angus breeders and semen distributors. This project was supported by the University of Missouri, National Research Initiative grants number 2008-35205-04687 and 2008-35205-18864 from the USDA Cooperative State Research, Education and Extension Service and grant number 2009-65205-05635 from the USDA NIFA Agriculture and Food Research Initiative. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Figures

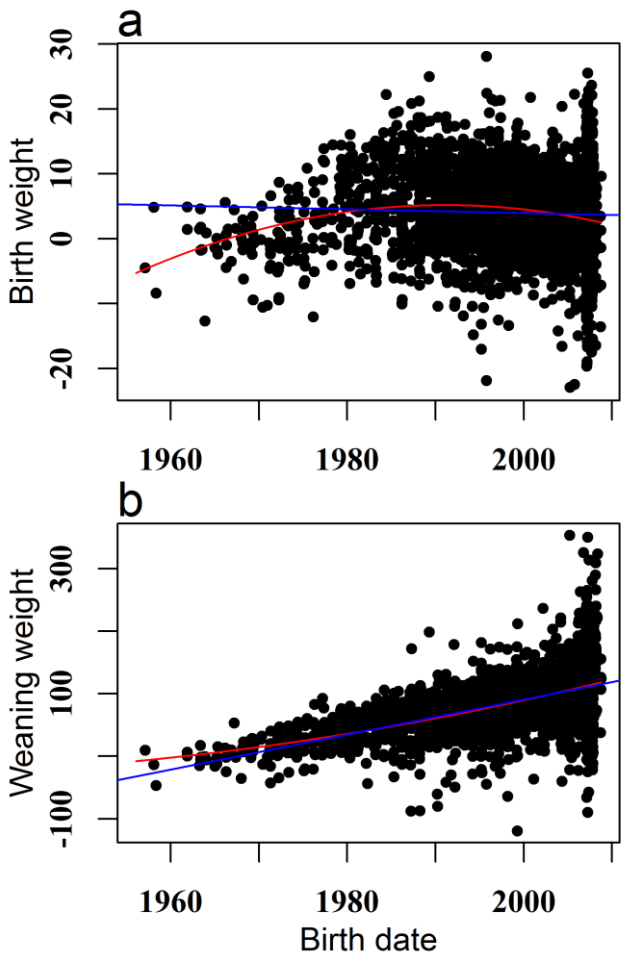

Figure 3.1 Deregressed estimated breeding values for birth and weaning weight plotted against birth date.

Deregressed estimated breeding values plotted against birth date for 3,570 Angus animals. The blue lines represent linear and red lines represent quadratic regressions.
a. Deregressed birth weight EBV.
b. Deregressed weaning weight EBV.

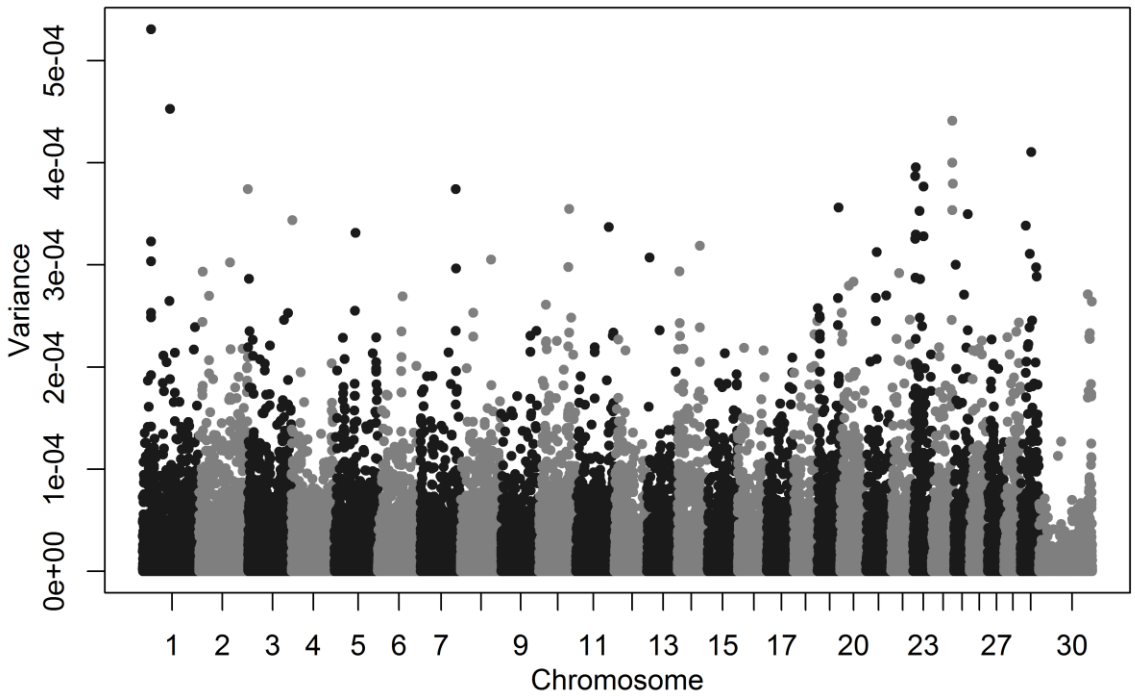

Figure 3.2 Manhattan plot of additive genetic variances explained by each SNP estimated from the GBLUP analysis of birth date.
For each SNP $2p_i(1-p_i)\alpha_i^2$ is plotted where p_i is allele frequency and α_i is the ASE for birth date for the *ith* SNP.

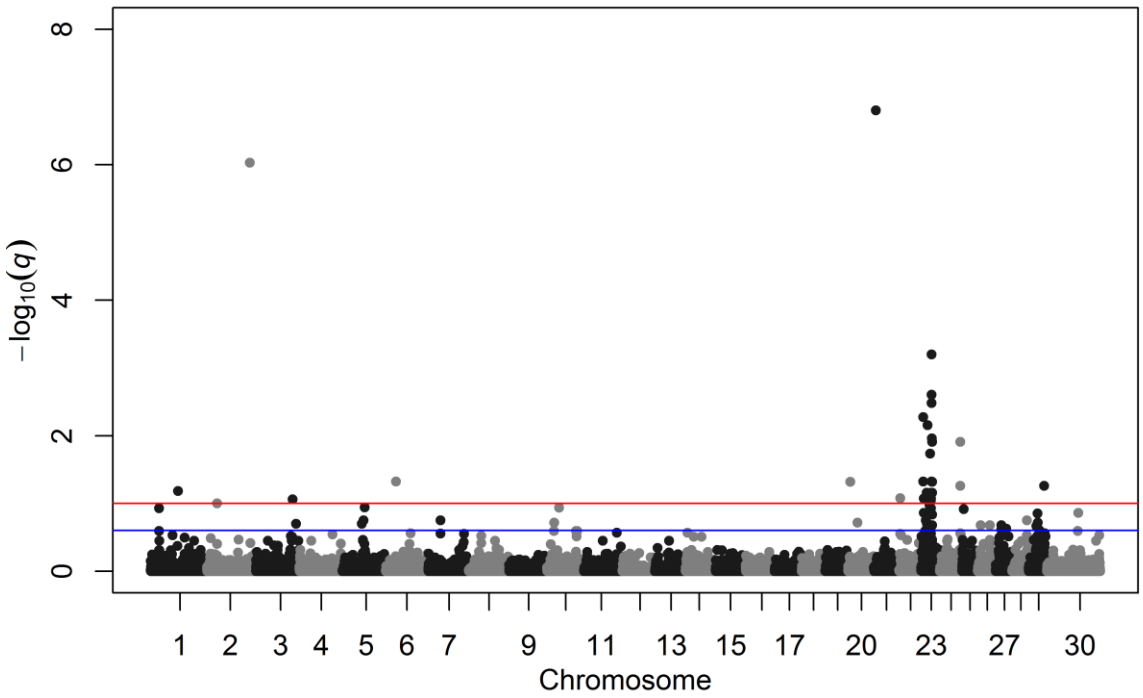

Figure 3.3 Manhattan plot of $-\log_{10}(q\text{-values})$ for SNP effects on birth date estimated in the EMMAX analysis.
Each *q*-value is the expected proportion of false positives among all SNP effects that are at least as extreme as that observed for the current SNP.

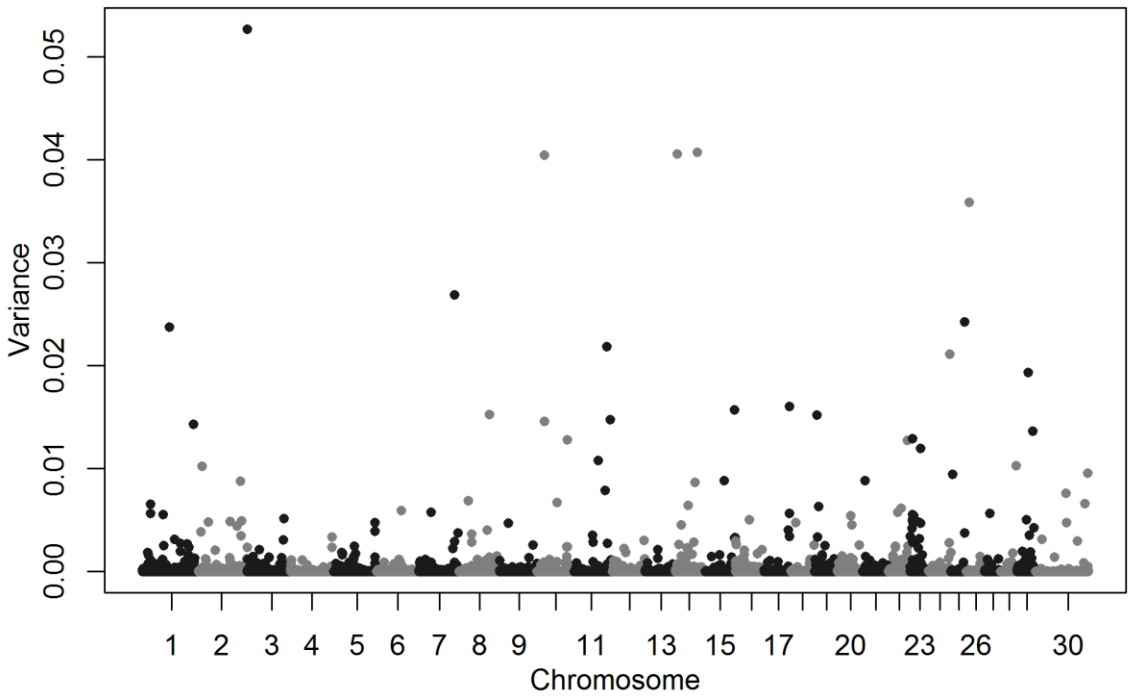

Figure 3.4 Manhattan plot of additive genetic variances explained by each SNP estimated from the BayesCπ analysis of birth date.

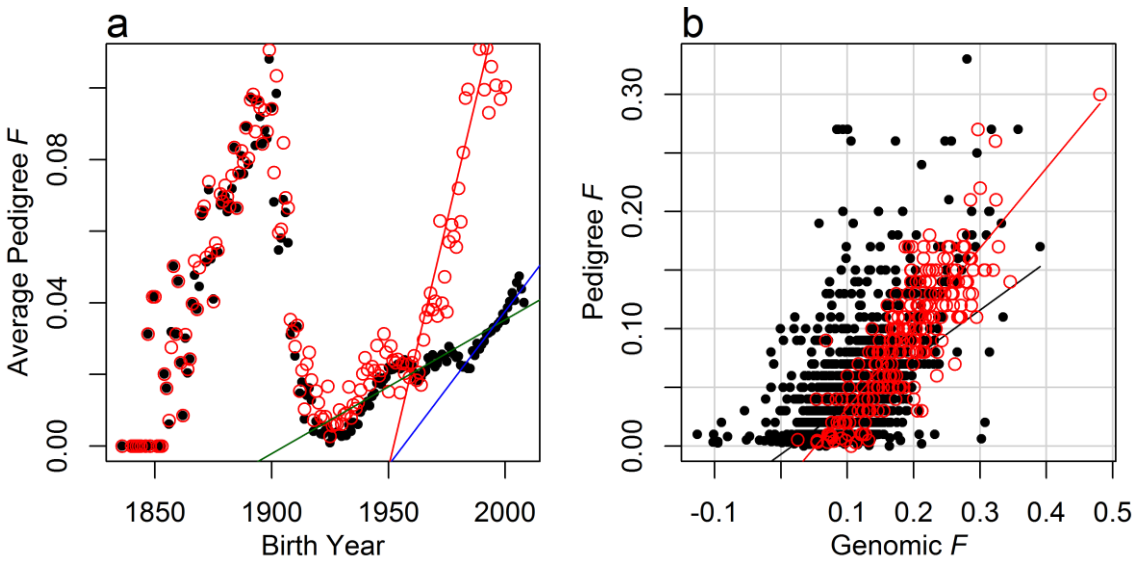

Figure 3.5 Analysis of inbreeding coefficients.
a. Plot of average pedigree *F* by birth year for 91,001 animals in the pedigree of the 3,570 genotyped bulls. Averages for the Wye herd animals and their ancestors are in red; averages for the remaining North American Angus and their ancestors and are in black. The red line represents the regression of pedigree *F* on birth year for Wye pedigree animals born after 1950. The green line is the regression of pedigree *F* on birth year for animals in the North American pedigree born after 1930. The blue line is the regression of pedigree *F* on birth year for animals in the North American pedigree born after 1980. See Table 3.1 for regression parameter estimates.
b. Plot of pedigree against genomic *F* coefficients. Wye herd animals are plotted in red, all other North American animals are plotted in black. The red line represents the regression of pedigree *F* on genomic *F* for Wye herd animals and the black line is for the remaining North American animals. See Table 3.2 for regression parameter estimates.

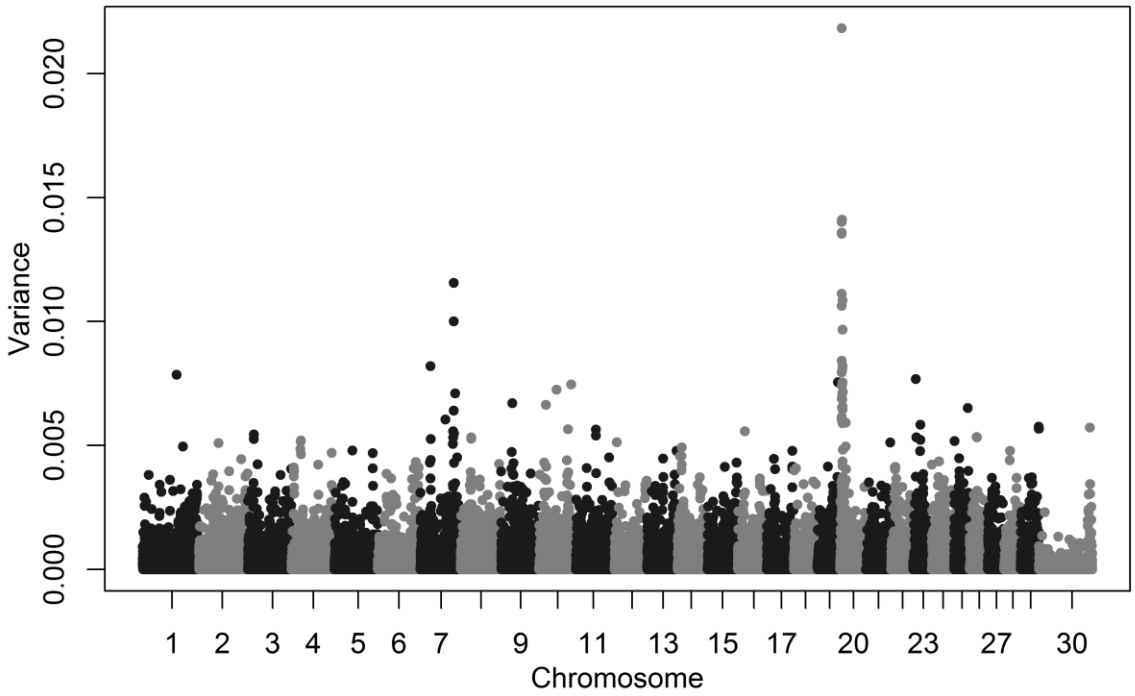

Figure 3.6 Manhattan plot of additive genetic variances explained by each SNP estimated from the GBLUP analysis of deregressed weaning weight EBVs.

For each SNP $2p_i(1-p_i)\alpha_i^2$ is plotted where p_i is allele frequency and α_i is the ASE for weaning weight for the *ith* SNP.

Tables

Table 3.1. Estimates of inbreeding effective population size for registered Angus cattle.

Birth year generations were calculated by subtracting 1950 from the birth year and dividing by the generation interval of 5.

Birth year generation = (birth year – 1950)/5.

Data set	Intercept	*ΔF/generation*	N_e
Wye pedigree, pedigree F	-0.00594	0.01373 ± 0.00046	36.41836 ± 0.03378
North American pedigree born after 1930, pedigree F	0.00724	0.00187 ± 3.51153e-05	267.59478 ± 0.01879
North American pedigree born after 1980, pedigree F	-0.02694	0.00430 ± 0.00015	116.14951 ± 0.03528
North American pedigree born after 1980, genomic F	0.05291	0.00523 ± 0.00048	94.18147 ± 0.09546

Table 3.2. Regression of pedigree *F* on genomic *F*

Sample	Adjusted R^2	Model p-value	Parameter	Estimate	Standard Error	t-value	p-value
All	0.4178	< 2.2e-16	Intercept	-0.01446	0.00116	-12.46	<2e-16
			Slope	0.491216	0.00971	50.61	<2e-16
Wye	0.7077	< 2.2e-16	Intercept	-0.03538	0.00597	-5.93	1.18e-08
			Slope	0.681299	0.02963	23.00	< 2e-16
North American	0.2829	< 2.2e-16	Intercept	-0.00699	0.00127	-5.528	3.48e-08
			Slope	0.410011	0.01128	36.364	<2e-16

Table 3.3 – Summary statistics for deregressed estimated breeding values (EBV) and accuracies (r^2) produced by the American Angus Association for the 3,570 genotyped animals

Trait[1]	Units	Heritability[2]	No Observations[3]	Mean EBV±SD[4]	Mean Acc±SD	C_{max}[5]	C[6]	V_g[7]
Birth Weight	lb	0.42	3241	4.03±5.95	0.78±0.24	0.796	0.77	23.42
Weaning Weight	lb	0.2	3229	86.69±45.98	0.68±0.32	0.822	0.704	690.86
Maternal Milk	lb	0.14	2067	33.79±30.01	0.70±0.27	0.862	0.709	373.15
Yearling Weight	lb	0.49	2776	154.03±78.15	0.69±0.29	0.827	0.78	1961.6
Yearling Height	in	0.45	2250	0.74±1.22	0.70±0.25	0.796	0.796	0.6165
Carcass Weight	lb	0.4	2457	30.93±84.08	0.41±028	0.914	0.627	1438.9
Marbling	units	0.45	3237	0.64±1.14	0.43±0.25	0.913	0.913	0.3542
Ribeye Muscle Area	in^2	0.51	3269	0.16±1.04	0.47±0.23	0.914	0.914	0.3775
Fat Thickness	in	0.34	3189	0.027±0.162	0.40±0.23	0.914	0.914	0.0072
Mature Weight	lb	0.55	1321	67.28±135.26	0.64±0.25	0.849	0.559	5818.8
Mature Height	in	0.82	1291	1.08±2.25	0.64±0.25	0.843	0.56	1.504
Scrotal Circumference	in	0.43	2479	0.55±1.83	0.69±0.25	0.818	0.698	1.641
Calving Ease Direct	%	0.18	3217	8.30±19.77	0.62±0.26	0.868	0.706	154.7
Calving Ease Maternal	%	0.12	1966	12.14±23.77	0.59±0.27	0.903	0.421	146
Docility	%	0.37	698	15.52±21.44	0.48±0.27	0.927	0.343	126.94
Heifer Pregnancy	%	0.13	1366	15.81±47.64	0.50±0.27	0.905	0.712	711.45
Birth Date	yr	0.53	3570	1998.93±8.98	1.00±0.00	NA	NA	25.83

Cont. Table 3.3 – Summary statistics for deregressed estimated breeding values (EBV) and accuracies (r^2) produced by the American Angus Association for the 3,570 genotyped animals

[1]See Supplementary Information for trait definitions.
[2]Narrow sense heritability used by the American Angus Association to compute estimates of additive genetic merit or as estimated from the birth date data.
[3]Number of breeding values that could successfully be deregressed or birth dates.
[4]Deregressed estimated breeding values or birth dates.
[5]Largest possible value of C imposed by the constraint $(1+F_i)/r_i^2 > C \times G_{ii}$ which ensures that weights for all animals' deregressed EBV are strictly positive. F_i is the pedigree inbreeding coefficient, r_i^2 is the accuracy of the deregressed breeding value, and G_{ii} is the diagonal of the genomic relationship coefficient matrix for the i^{th} animal.
[6]Proportion of additive genetic variation explained by 45,073 SNPs computed as V_m/V_g.
[7]Estimated additive genetic variance from the analysis of deregressed breeding values or birthdates.

Table 3.4 – Genetic correlations between traits estimated from correlations between allele substitution effects for 45,073 SNP.

Trait	WW	Milk	YW	YH	CW	MARB	REA	FT	MW	MH	SC	CED	CEM	DOC	HP	Birth Date
Birth Weight (BW)	0.321	-0.134	0.312	0.395	0.305	-0.024	0.084	-0.091	0.268	0.266	0.024	-0.835	-0.35	0.021	0.029	0.01
Weaning Weight (WW)		0.083	0.885	0.255	0.48	0.088	0.277	0.016	0.264	0.217	0.191	-0.188	0.009	0.102	-0.031	0.463
Maternal Milk (Milk)			0.125	-0.047	0.031	0.085	0.015	0.052	-0.118	-0.088	0.134	0.209	0.154	0.016	0.028	0.278
Yearling Weight (YW)				0.294	0.572	0.112	0.292	0.064	0.304	0.249	0.218	-0.171	0.009	0.106	-0.042	0.436
Yearling Height (YH)					0.284	-0.01	0.012	-0.1	0.434	0.551	0.081	-0.342	-0.082	0.031	0.035	-0.025
Carcass Weight (CW)						0.05	0.469	0.144	0.306	0.277	0.118	-0.252	-0.101	0.069	-0.012	0.177
Marbling (MARB)							-0.02	0.173	-0.035	-0.012	0.047	0.069	0.082	-0.016	0.008	0.158
Ribeye Muscle Area (REA)								-0.027	0.057	0.043	0.038	-0.059	-0.032	0.015	-0.007	0.119
Fat Thickness (FT)									-0.104	-0.096	0.054	0.1	0.06	0.001	-0.001	0.063
Mature Weight (MW)										0.811	0.021	-0.23	-0.076	0.015	0.014	0.042
Mature Height (MH)											0.032	-0.222	-0.031	0	0.053	0.014
Scrotal Circumference (SC)												0.023	0.041	0.029	-0.004	0.144
Calving Ease Direct (CED)													0.537	-0.003	-0.046	0.151
Calving Ease Maternal (CEM)														-0.032	-0.068	0.218
Docility (DOC)															0.022	0.085
Heifer Pregnancy (HP)																-0.041

Table 3.5 Relative selection intensities for 16 production traits estimated from the regression of birth date ASEs on standardized SNP ASE coefficients.

ASEs were standardized by conversion to coefficients of $pqASE/\sigma_{ASE}$. The *F*-statistic for the model was 1,288 on 16 and 44800 degrees of freedom (p-value < 2.2e-16), with an R^2 of 0.3151 and an adjusted R^2 of 0.3148.

Model Term	Estimate	Std. Error	t-value	p-values
Intercept	-0.0002202	0.0039108	-0.056	0.95510
Weaning weight (WW)	1.8158248	0.0418758	43.362	< 2e-16
Calving ease direct (CED)	1.1348654	0.0402923	28.166	< 2e-16
Maternal milk (Milk)	0.8445815	0.0200755	42.07	< 2e-16
Birth weight (BW)	0.8199081	0.03781	21.685	< 2e-16
Yearling height (YH)	-0.5286214	0.0245651	-21.519	< 2e-16
Calving ease maternal (CEM)	0.4618728	0.0231311	19.968	< 2e-16
Yearling weight (YW)	0.3618465	0.0453856	7.973	1.59e-15
Marbling (MARB)	0.3617341	0.0194656	18.583	< 2e-16
Heifer Pregnancy (HP)	0.1743628	0.0190981	9.13	< 2e-16
Carcass weight (CW)	-0.1119094	0.0267417	-4.185	2.86e-05
Mature weight (MW)	0.0961186	0.0334404	2.874	0.00405
Scrotal Circumference (SC)	0.0957338	0.019623	4.879	1.07e-06
Docility (DOC)	-0.0586487	0.0189689	-3.092	0.00199
Mature height (MH)	-0.0341907	0.0352666	-0.969	0.33230
Fat thickness (FT)	0.0339062	0.0200513	1.691	0.09085
Ribeye area (REA)	0.0035452	0.0221618	0.16	0.87291

Table 3.6. Relative selection intensities for 16 production traits estimated from the regression of the top 948 birth date ASEs on standardized SNP ASE coefficients.
The 948 SNPs with the largest absolute birth date ASEs were fit in the model. ASEs were standardized by conversion to coefficients of $pqASE/\sigma_{ASE}$. The *F*-statistic for the model was 158.5 on 16 and 931 degrees of freedom (p-value < 2.2e-16), with an R^2 of 0.7314 and an adjusted R^2 of 0.7268.

Model Term	Estimate	Std. Error	t-value	p-values
Intercept	0.00722	0.0171	0.422	0.67301
Weaning weight (WW)	2.16434	0.24754	8.744	< 2e-16
Calving ease direct (CED)	1.25094	0.18103	6.91	8.97E-12
Maternal milk (Milk)	0.92228	0.09211	10.013	< 2e-16
Birth weight (BW)	0.90145	0.16459	5.477	5.56E-08
Yearling Weight (YW)	0.59972	0.27136	2.21	0.02734
Yearling Height (YH)	-0.48315	0.09853	-4.903	1.11E-06
Calving Ease Maternal (CEM)	0.45156	0.10451	4.321	1.72E-05
Marbling (MARB)	0.25746	0.0833	3.091	0.00206
Fat thickness (FT)	0.23892	0.08309	2.875	0.00413
Carcass Weight (CW)	-0.21418	0.12389	-1.729	0.08418
Scrotal Circumference (SC)	0.17662	0.08355	2.114	0.03478
Heifer Pregnancy (HP)	0.14334	0.07967	1.799	0.07231
Mature Weight (MW)	0.12468	0.14685	0.849	0.3961
Mature Height (MH)	0.07	0.14968	0.468	0.64012
Docility (DOC)	-0.04387	0.07843	-0.559	0.57604
Ribeye area (REA)	0.00954	0.09876	0.097	0.92306

Supplementary Material

Supplementary Information

The following definitions and abbreviations include excerpts from: http://www.angus.org/Nce/Definitions.aspx.

Expected Progeny Difference (EPD). Expected performance of future progeny relative to the progeny of other animals. EPDs are one half of the Estimated Breeding Values (EBVs) of each animal and are predicted in mixed linear model analyses which incorporate numerator relationship matrices determined by pedigree information. EPDs are expressed in the units of measurement for the trait.

Accuracy (ACC). The American Angus Association reports accuracy as *ACC* = 1 - $\sqrt{1-r_{TI}^2}$ where r_{TI}^2 is squared correlation between predicted breeding value and true breeding value. These values were transformed in this study to obtain the r_{TI}^2 values necessary to obtain deregressed EBVs and weights for mixed model analyses.

Birth Weight (BW). Birth weight in pounds of a bull's progeny.

Weaning Weight (WW). Weaning weight in pounds of progeny at ~305 d of age.

Maternal Milk (MILK). Bull's genetic merit for the milk and mothering ability of his daughters. It is that part of a calf's weaning weight in pounds that is attributed to milk and mothering ability.

Yearling Weight (YW). Weight in pounds of progeny at 12 months of age.

Carcass Weight (CW). Hot carcass weight in pounds of progeny when slaughtered at ~15 mo of age.

Mature Weight (MW). Mature weight in pounds of a bull's daughters.

Yearling Height (YH). Height in inches of a bull's progeny measured at the hip at 12 months of age.

Mature Height (MH). Mature height in inches of a bull's daughters measured at the hip.

Fat Thickness (FAT). External fat thickness measured between the 12th and 13th ribs. Expressed in inches.

Marbling (MARB). Intramuscular fat content of the *longissimus dorsi* muscle measured between the 12th and 13th ribs.

Ribeye Muscle Area (RE). *Longissimus dorsi* cross-sectional area measured between the 12th and 13th ribs. Expressed in square inches.

Calving Ease Direct (CED). Percentage of unassisted births, with a higher value indicating greater calving ease in first-calf females. It predicts the average ease with which a bull's calves will be born when he is bred to first-calf females.

Calving Ease Maternal (CEM). Percentage of unassisted births with a higher value indicating greater calving ease in first-calf daughters. It predicts the average ease with which a bull's daughters will calve as first-calf heifers.

Scrotal Circumference (SC). Bull's scrotal circumference used as an indirect measure of female fertility. Expressed in centimeters.

Heifer Pregnancy Rate (HP). Percentage of a bull's daughters expected to become pregnant during a breeding season.

Docility (DOC). Percentage differences between bulls' progeny in temperament with higher values being more docile.

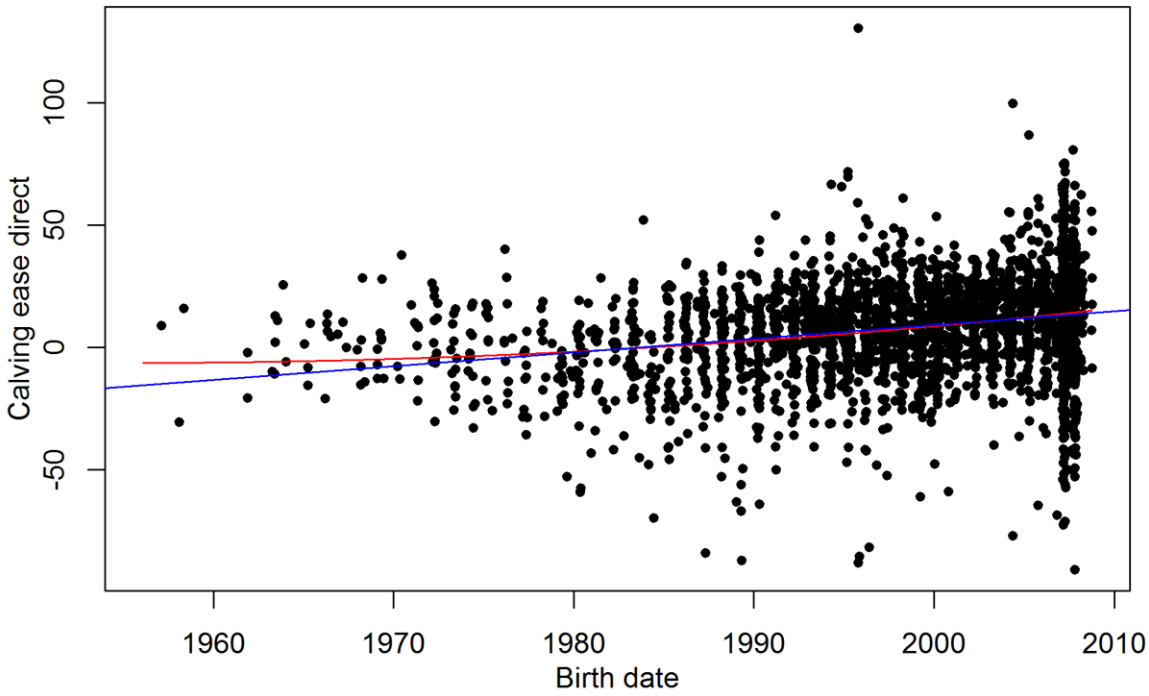

Figure 3.7 Deregressed calving ease direct EBV by birth date.
Figure S1 in publication.

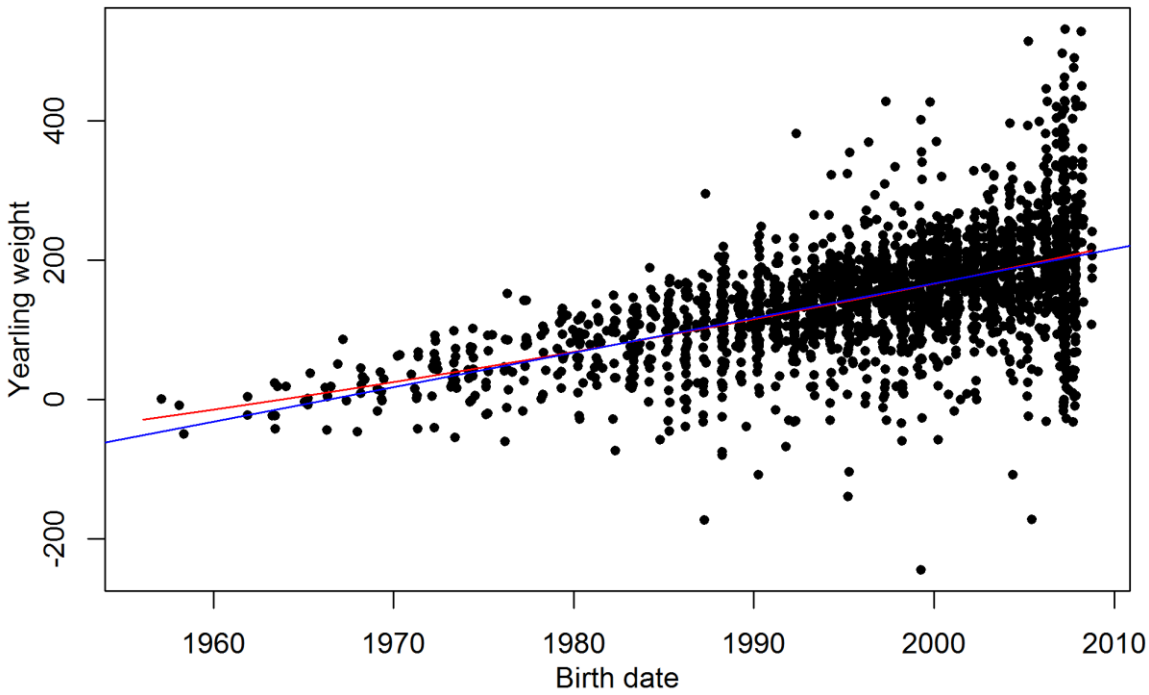

Figure 3.8 Deregressed yearling weight EBV by birth date.
Figure S2 in publication.

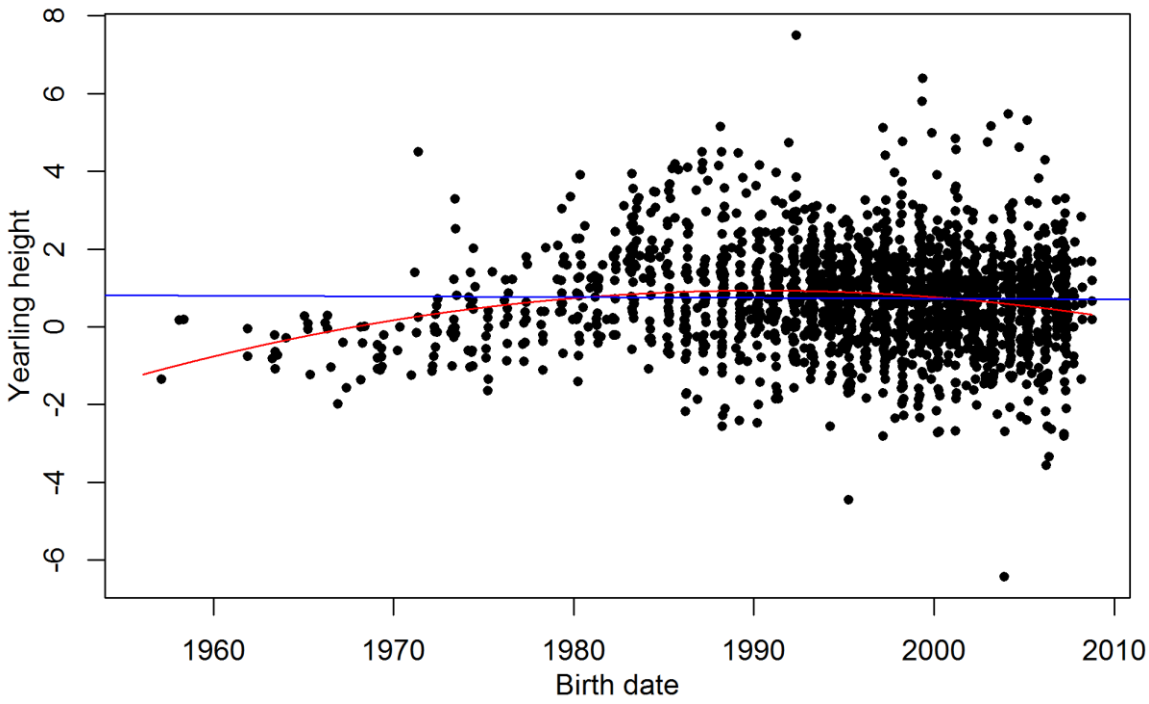

Figure 3.9 Deregressed yearling height EBV by birth date.
Figure S3 in publication.

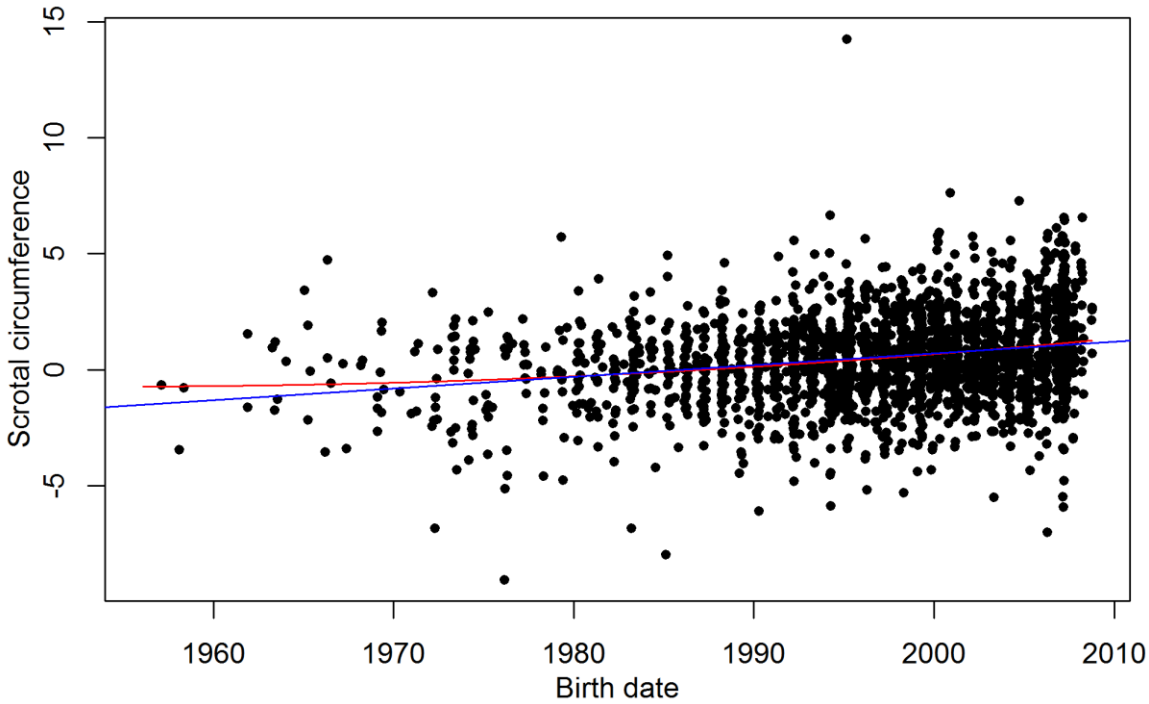

Figure 3.10 Deregressed scrotal circumference EBV by birth date.
Figure S4 in publication.

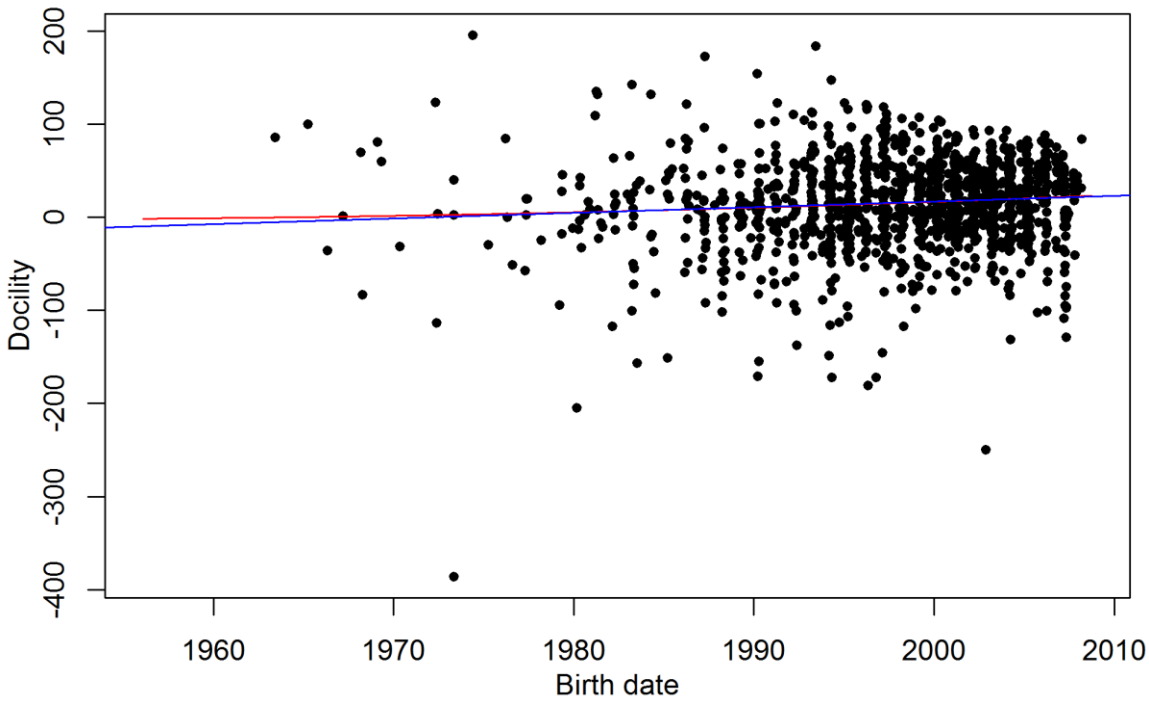

Figure 3.11 Deregressed docility EBV by birth date.
Figure S5 in publication.

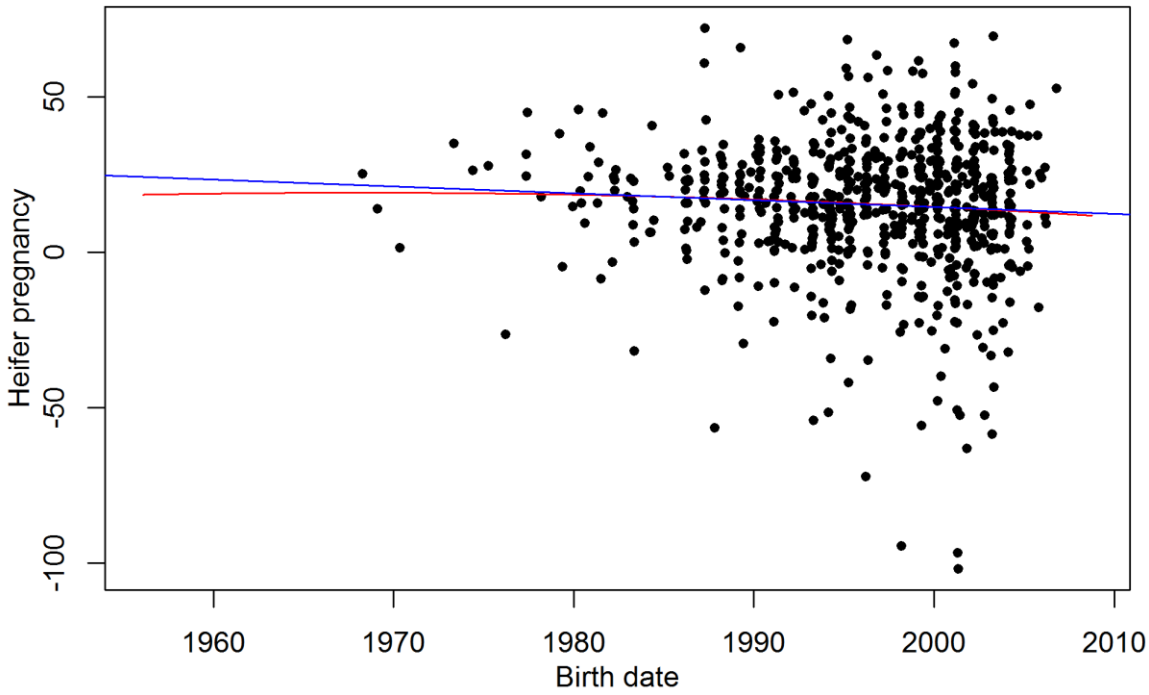

Figure 3.12 Deregressed heifer pregnancy EBV by birth date.
Figure S6 in publication.

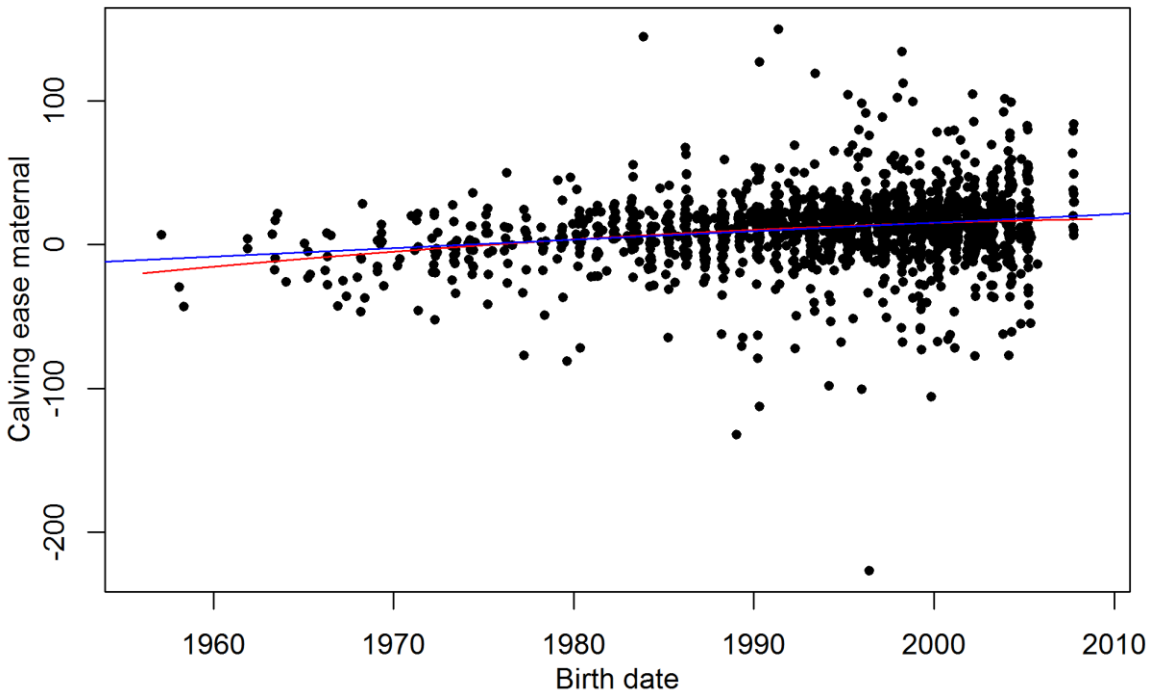

Figure 3.13 Deregressed calving ease maternal EBV by birth date.
Figure S7 in publication.

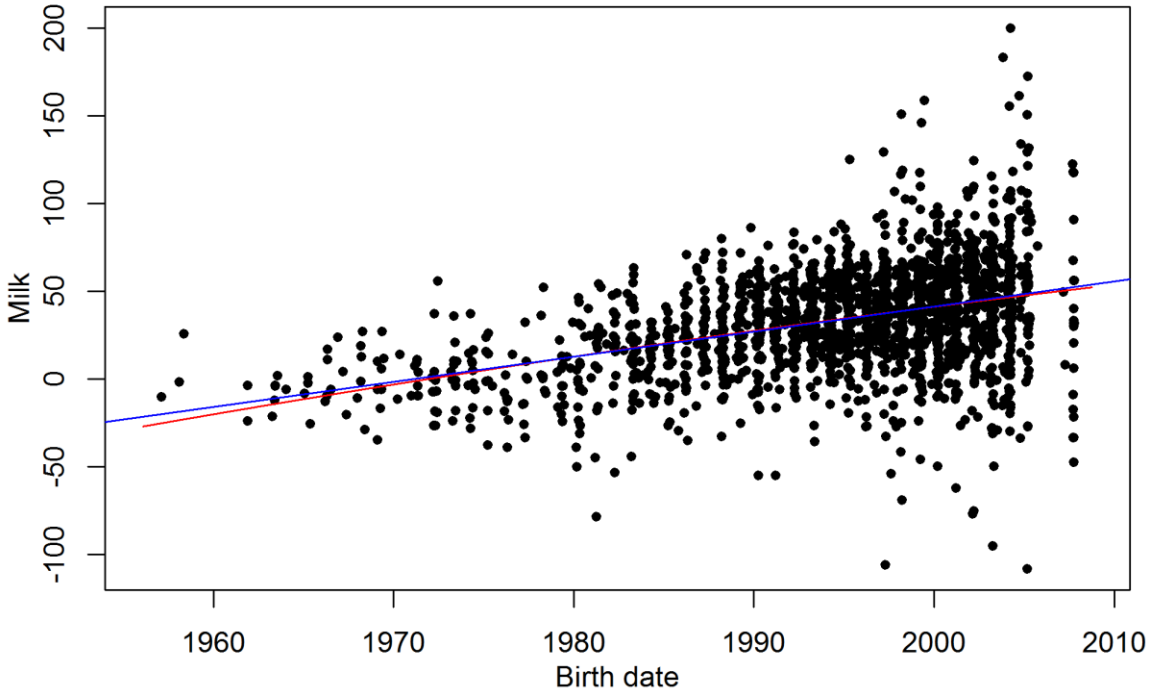

Figure 3.14 Deregressed maternal milk EBV by birth date.
Figure S8 in publication.

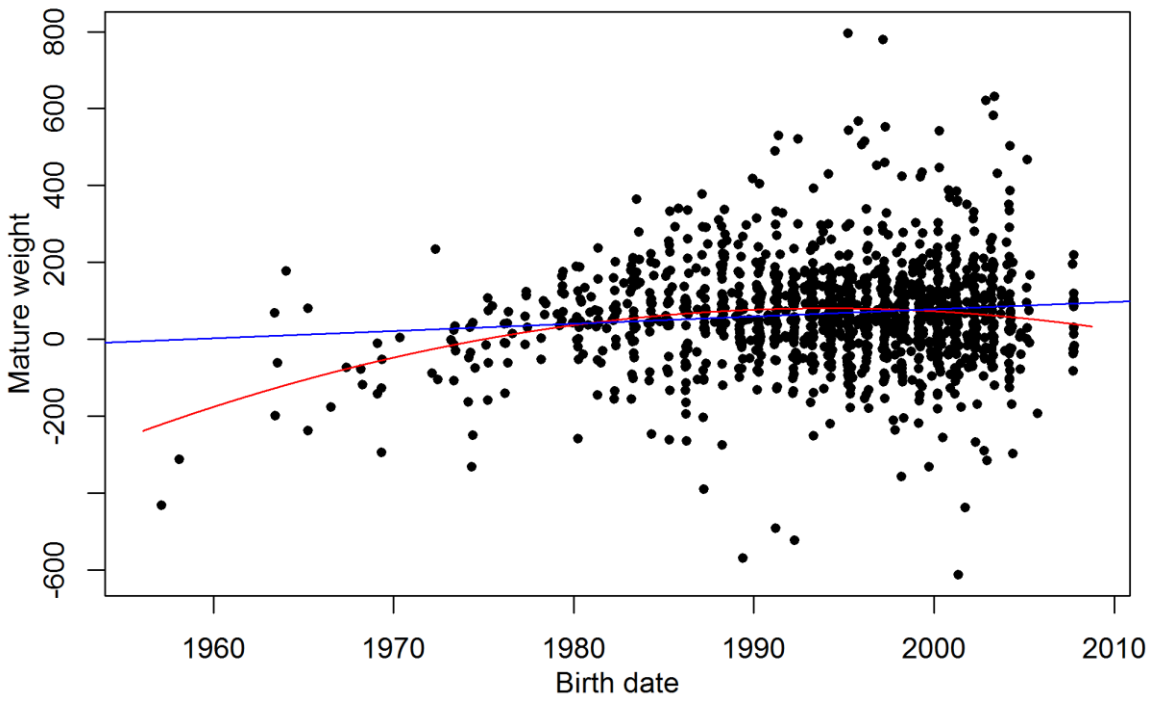

Figure 3.15 Deregressed mature weight EBV by birth date.
Figure S9 in publication.

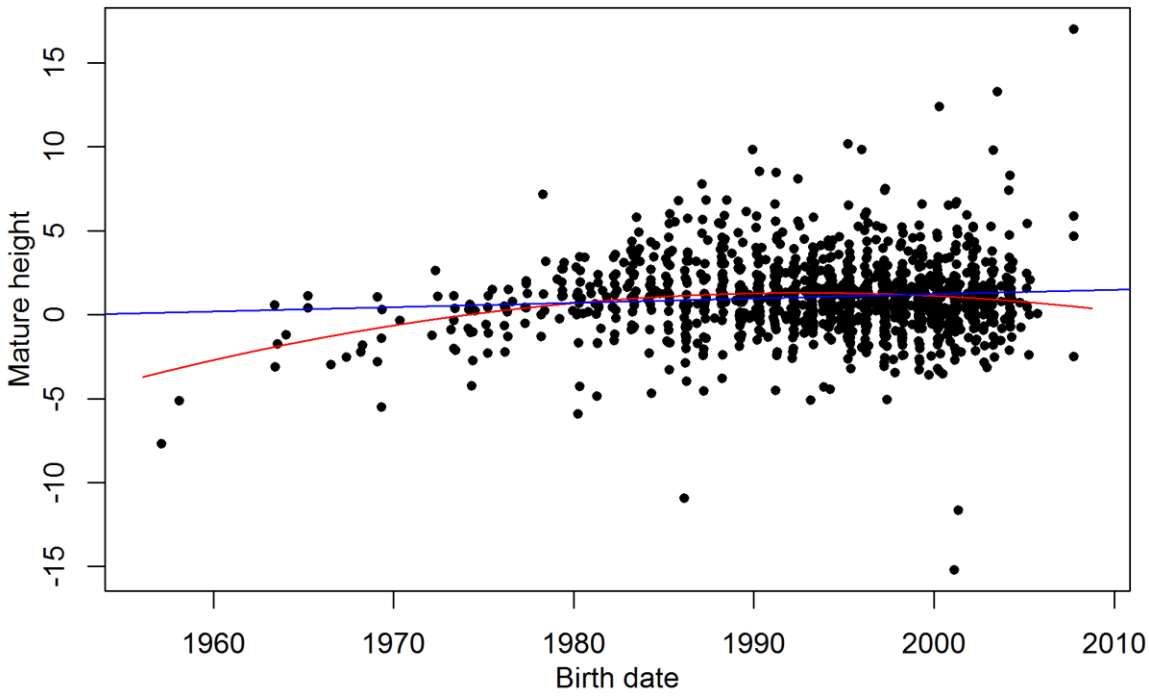

Figure 3.16 Deregressed mature height EBV by birth date.
Figure S10 in publication.

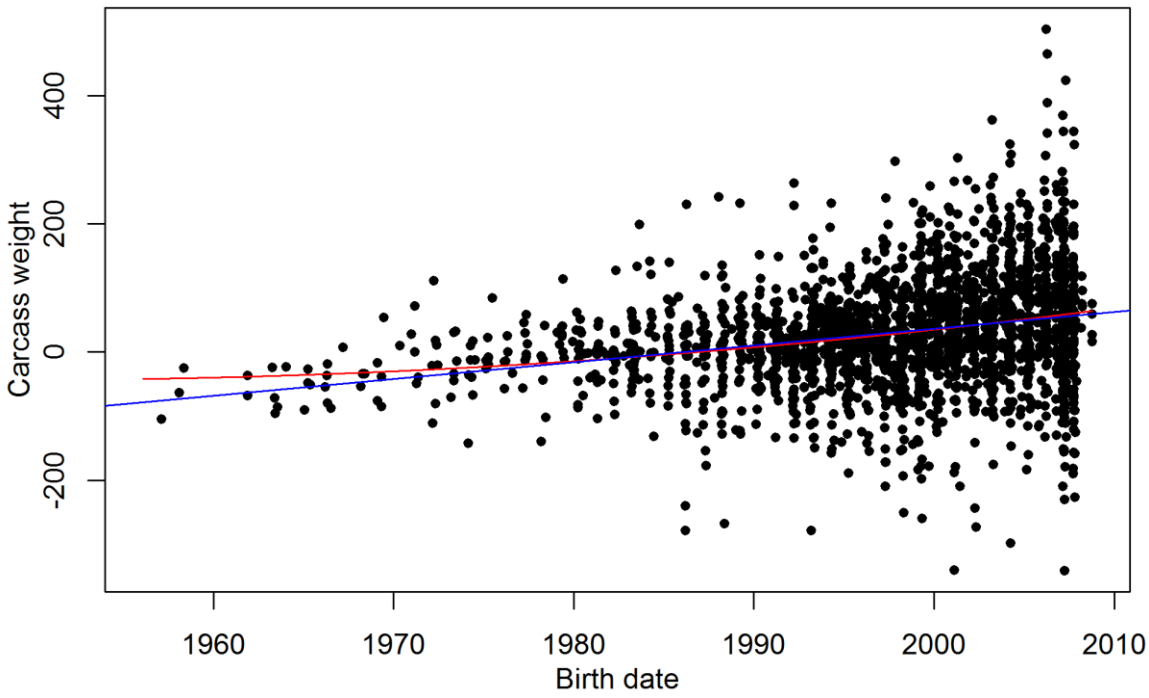

Figure 3.17 Deregressed carcass weight EBV by birth date.
Figure S11 in publication.

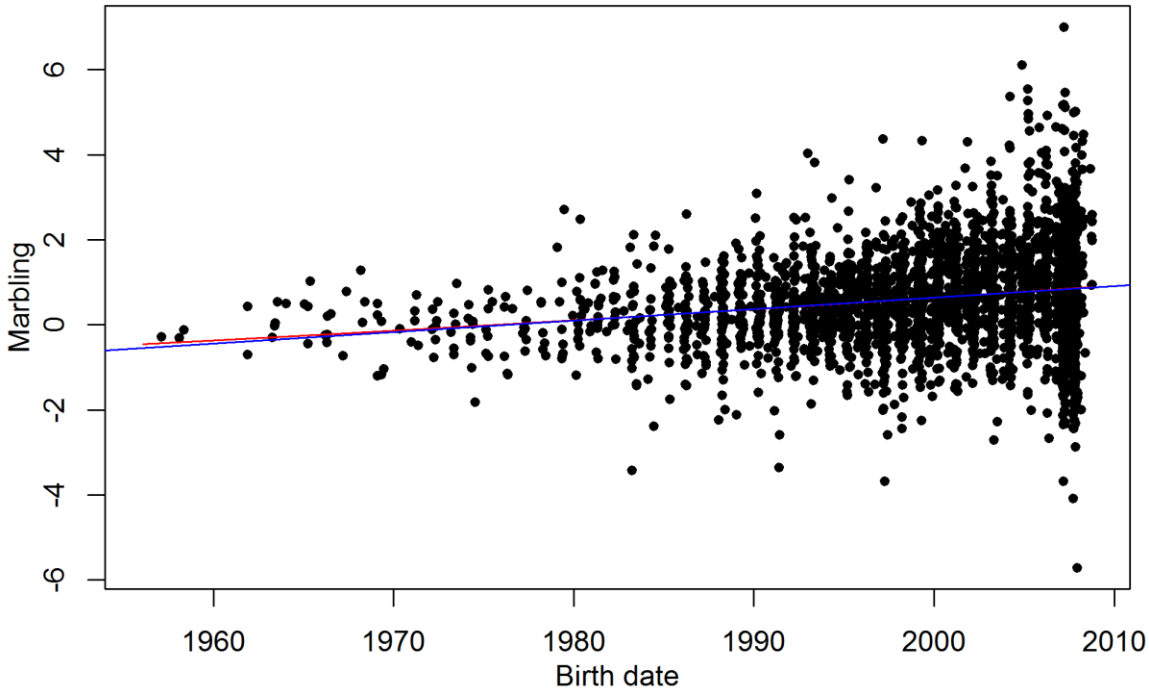

Figure 3.18 Deregressed marbling EBV by birth date.
Figure S12 in publication.

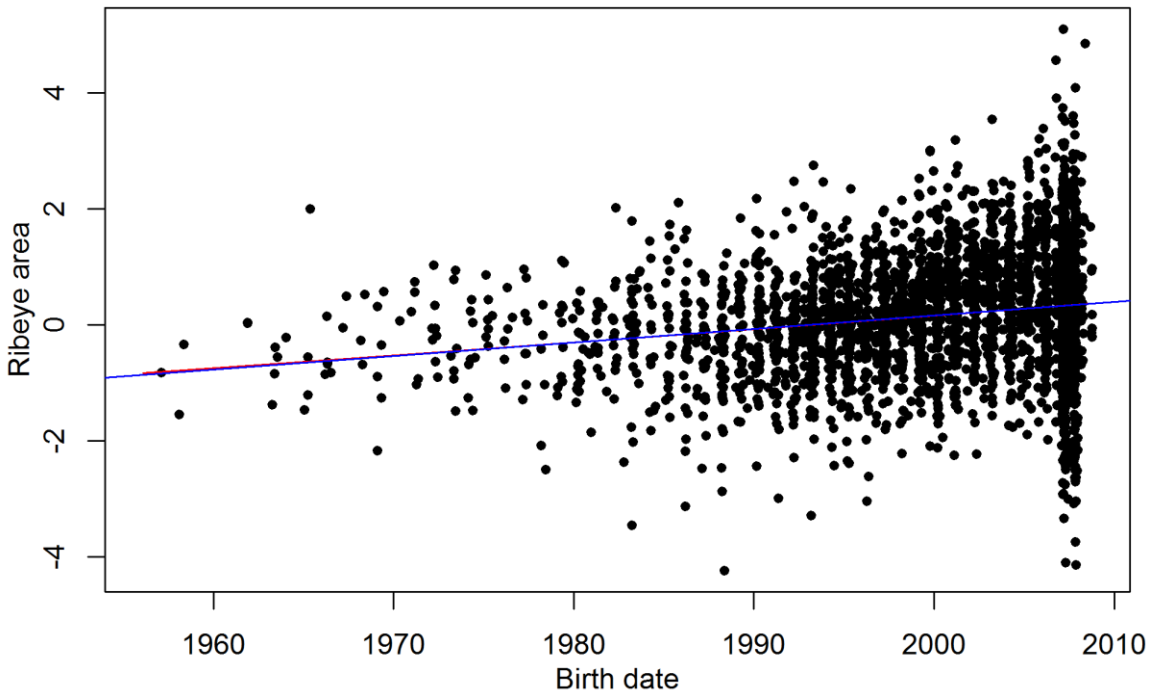

Figure 3.19 Deregressed ribeye area EBV by birth date.
Figure S13 in publication.

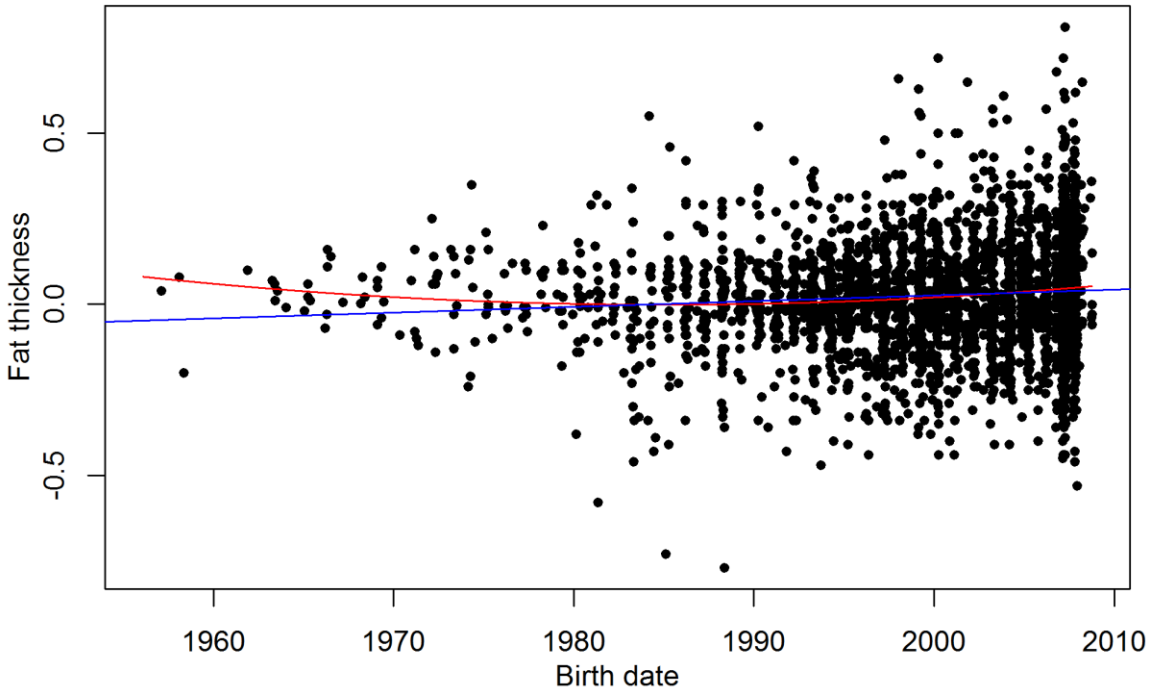

Figure 3.20 Deregressed fat thickness EBV by birth date.
Figure S14 in publication.

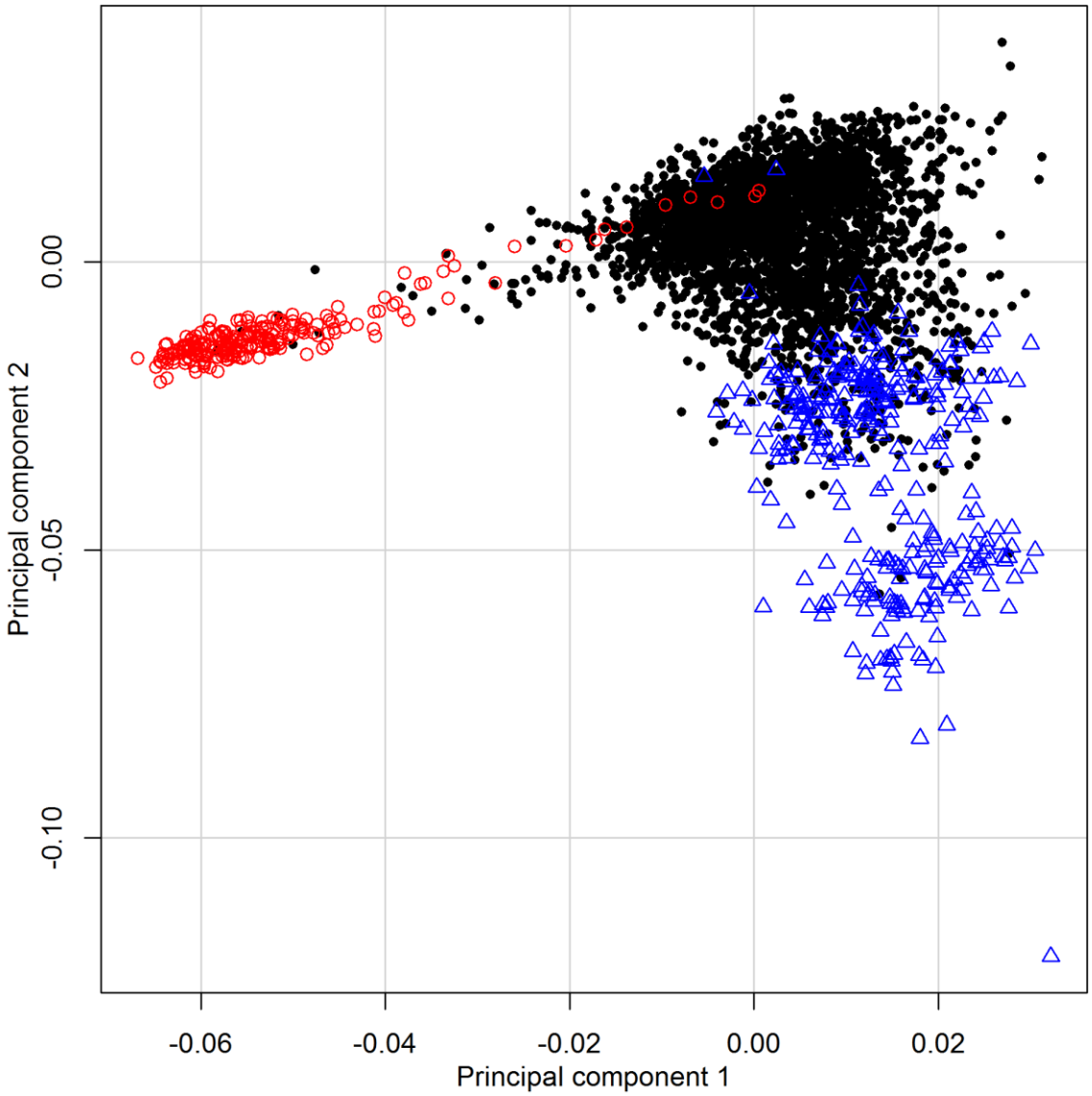

Figure 3.21 Principal component analysis of Angus animal genotypes.

From this analysis we identified two subgroups within our data. The first, denoted by red, is the Wye herd developed from imports from the British Isles and managed as a closed herd. The second is the rest of North American Angus. The blue triangles are a prominent AI sire (lower right corner), his sire, grandsire, progeny, and grandprogeny. Principal components 3 through 3570 break apart family structure in a similar fashion to principal component 2. We correct for population structure and kinship by utilizing a genomic relationship matrix in our analyses.
Figure S15 in publication.

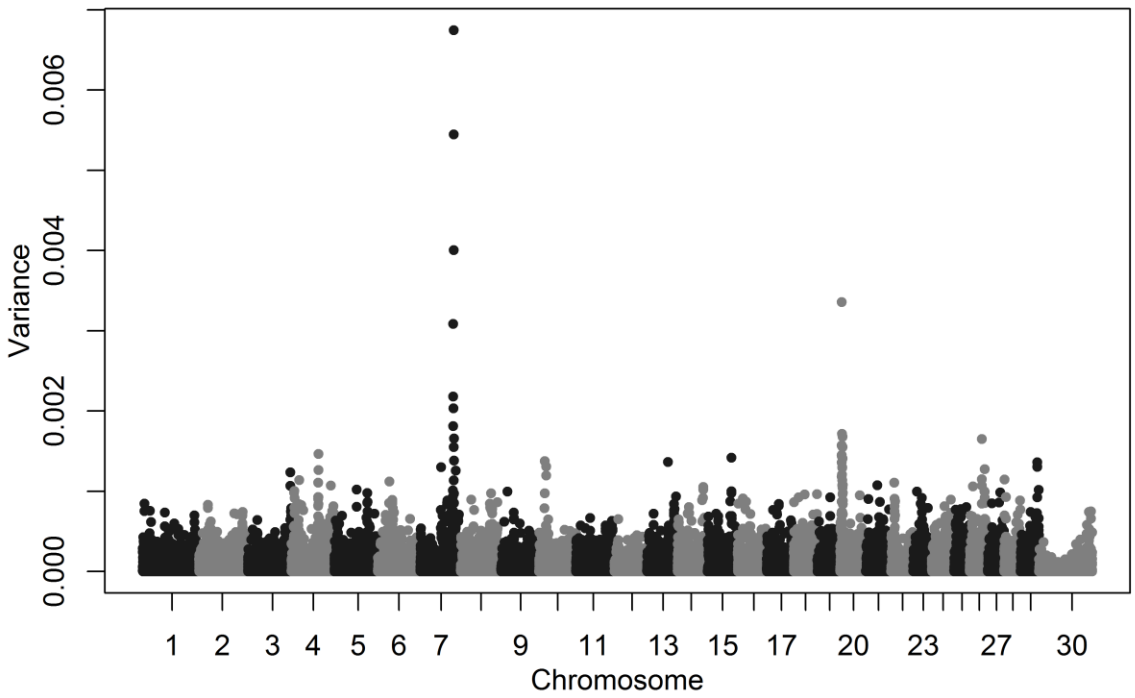

Figure 3.22 Manhattan plot of SNP variances for calving ease direct.
Figure S16 in publication.

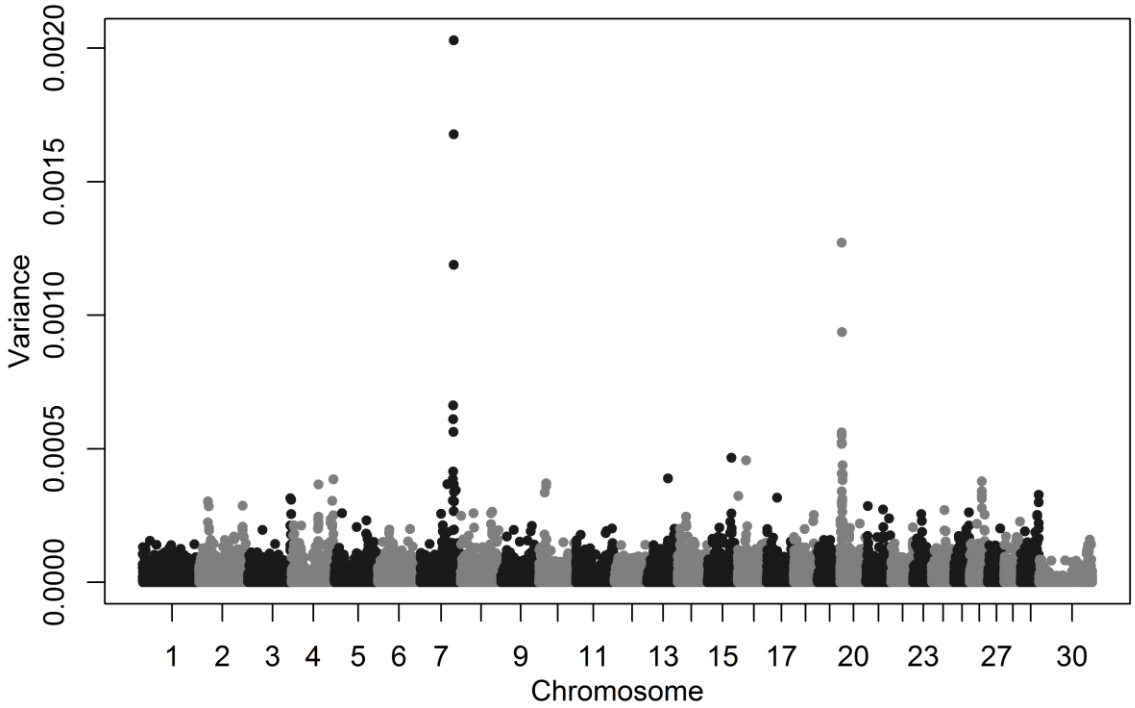

Figure 3.23 Manhattan plot of SNP variances for birth weight.
Figure S17 in publication.

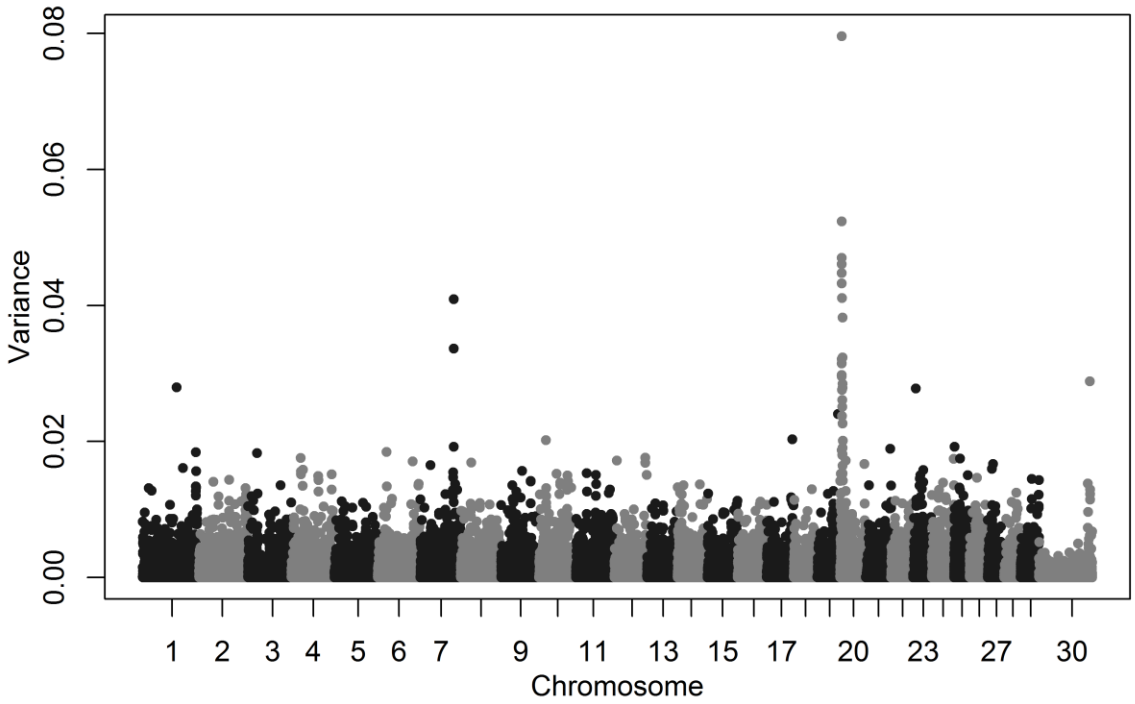

Figure 3.24 Manhattan plot of SNP variances for yearling weight.
Figure S18 in publication.

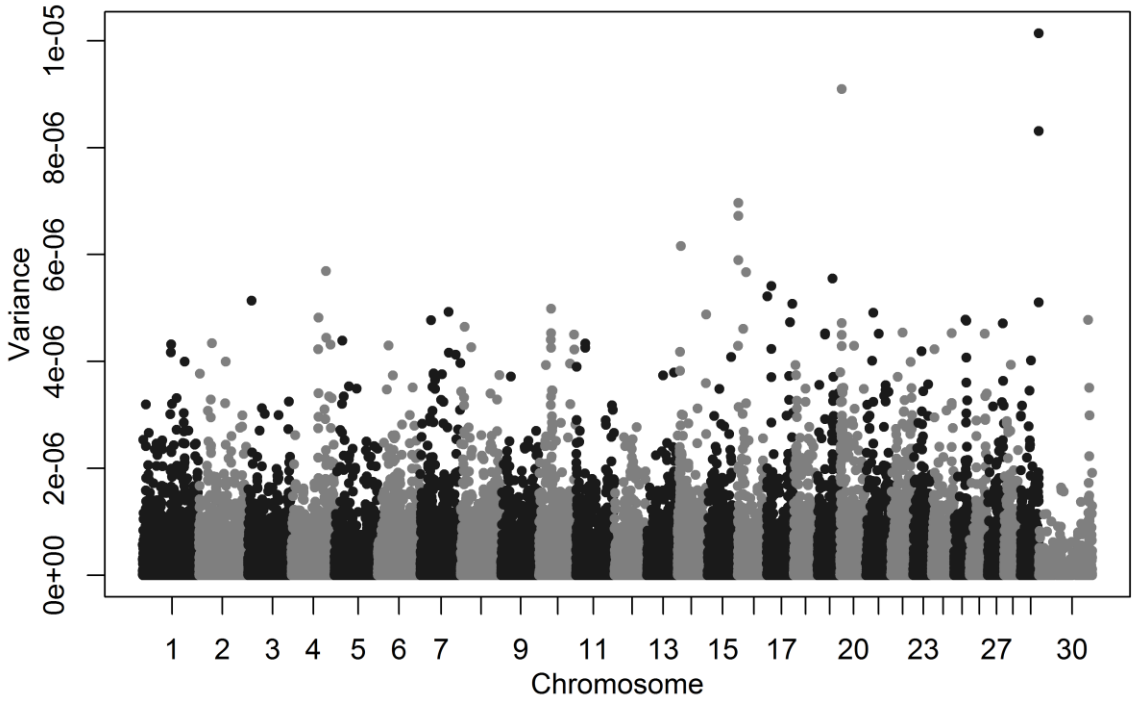

Figure 3.25 Manhattan plot of SNP variances for yearling height.
Figure S19 in publication.

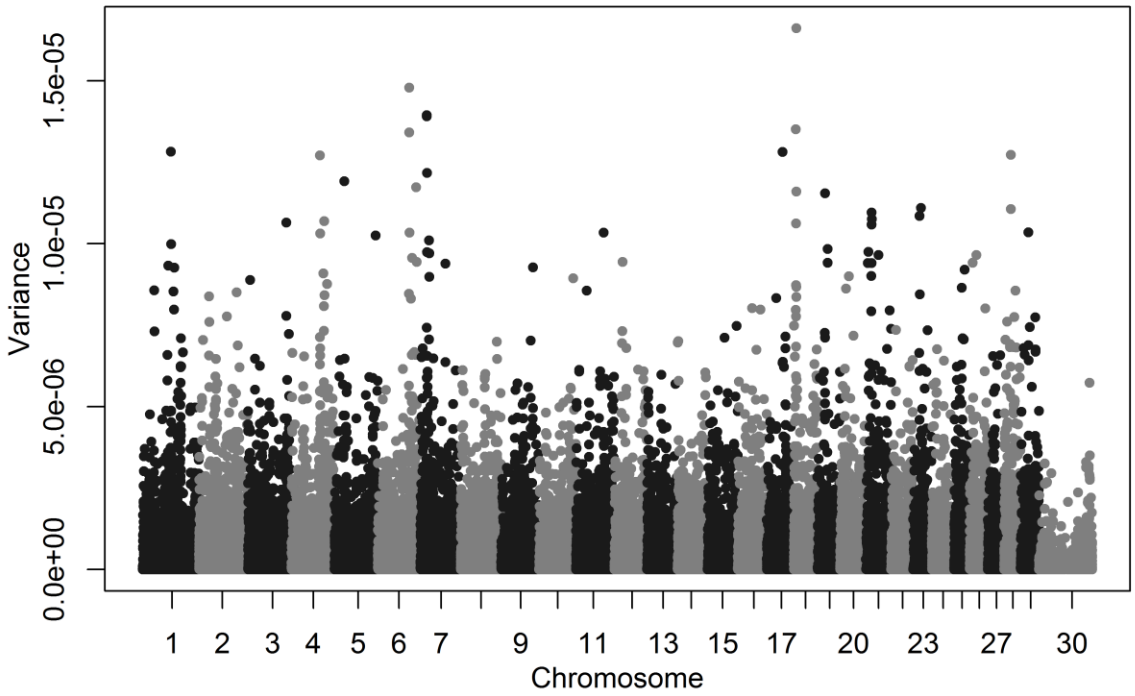

Figure 3.26 Manhattan plot of SNP variances for scrotal circumference.
Figure S20 in publication

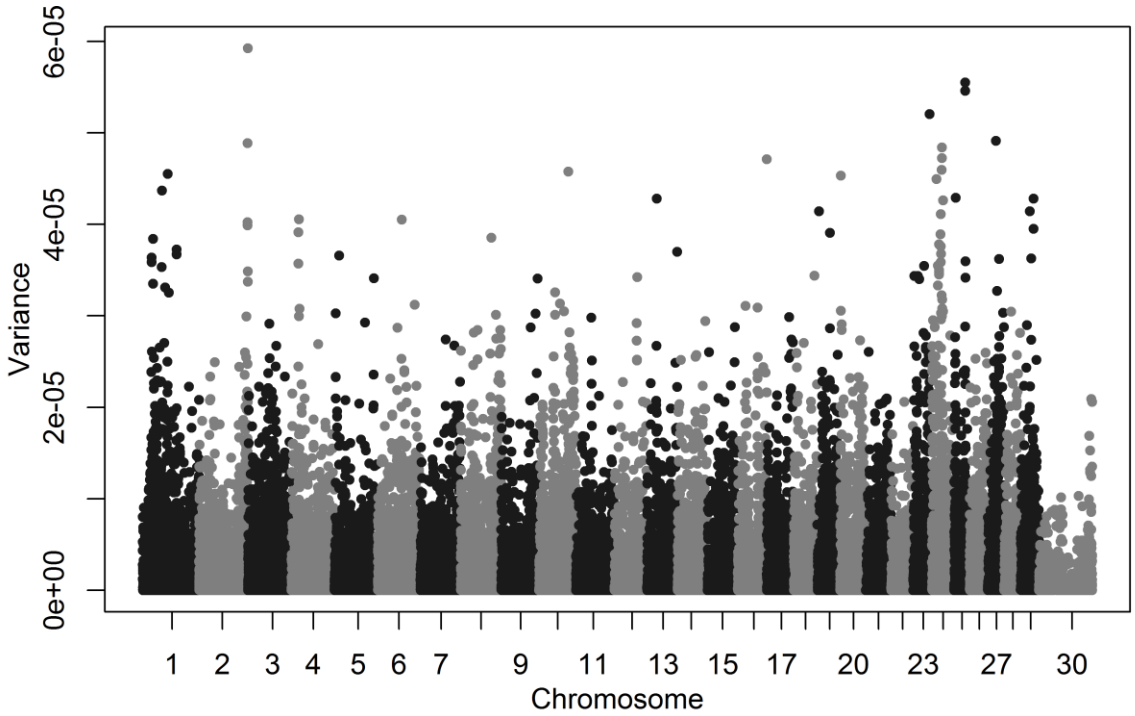

Figure 3.27 Manhattan plot of SNP variances for docility.
Figure S21 in publication.

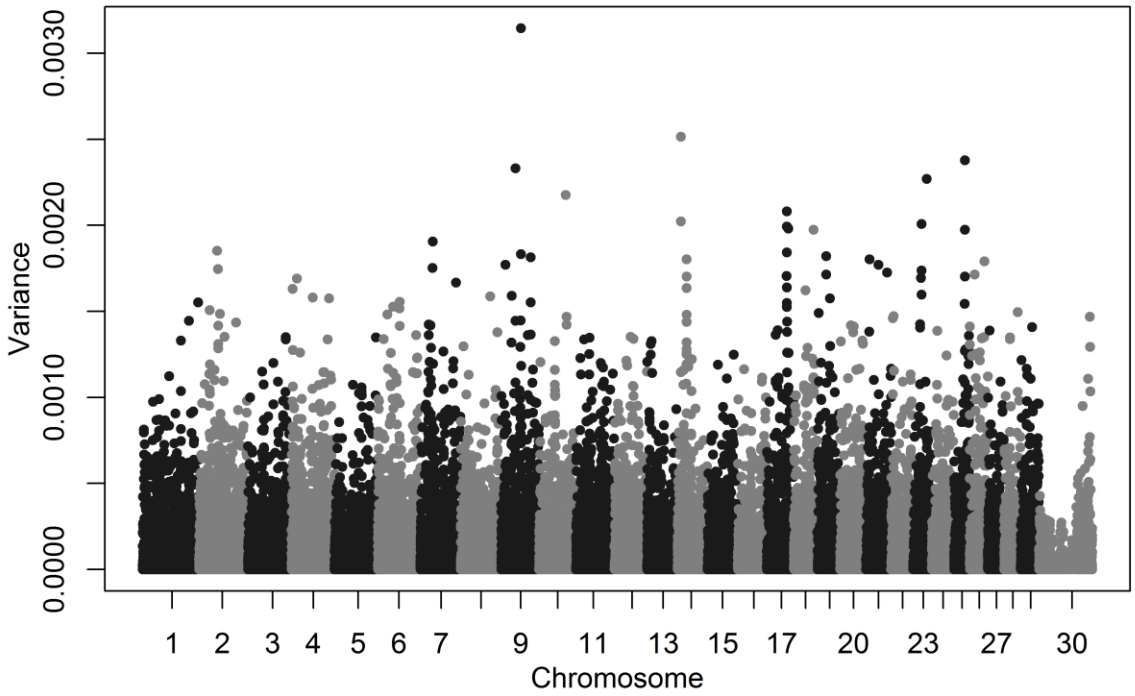

Figure 3.28 Manhattan plot of SNP variances for heifer pregnancy.
Figure S22 in publication.

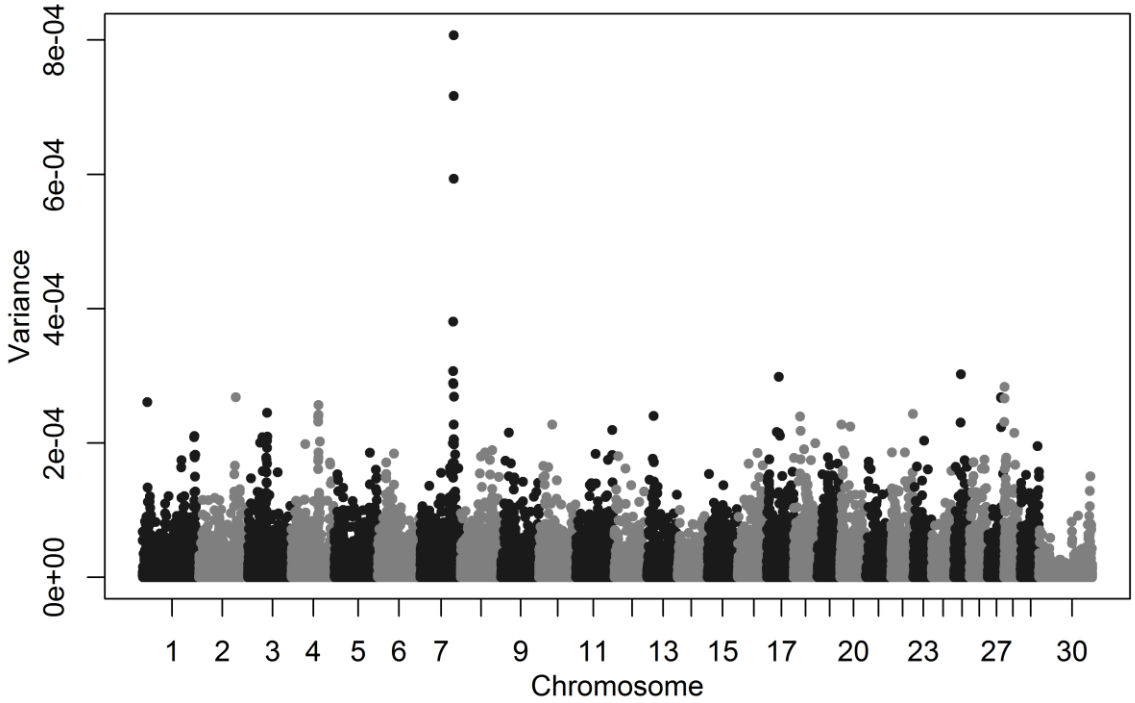

Figure 3.29 Manhattan plot of SNP variances for calving ease maternal.
Figure S23 in publication.

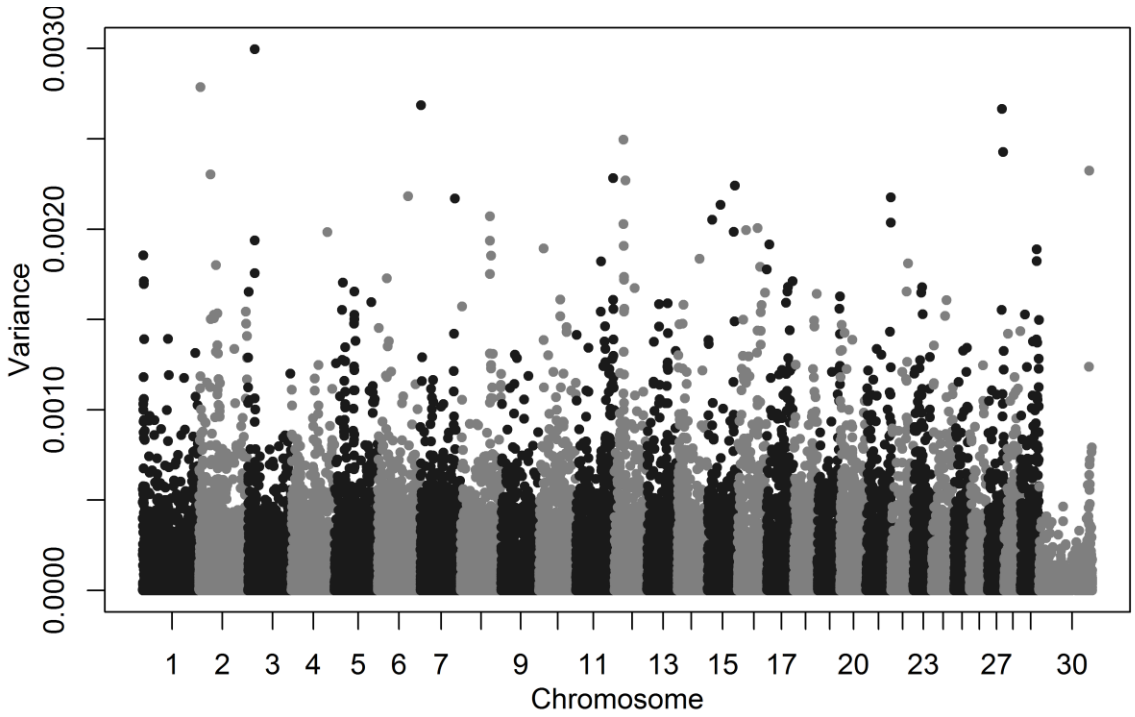

Figure 3.30 Manhattan plot of SNP variances for milk.
Figure S24 in publication.

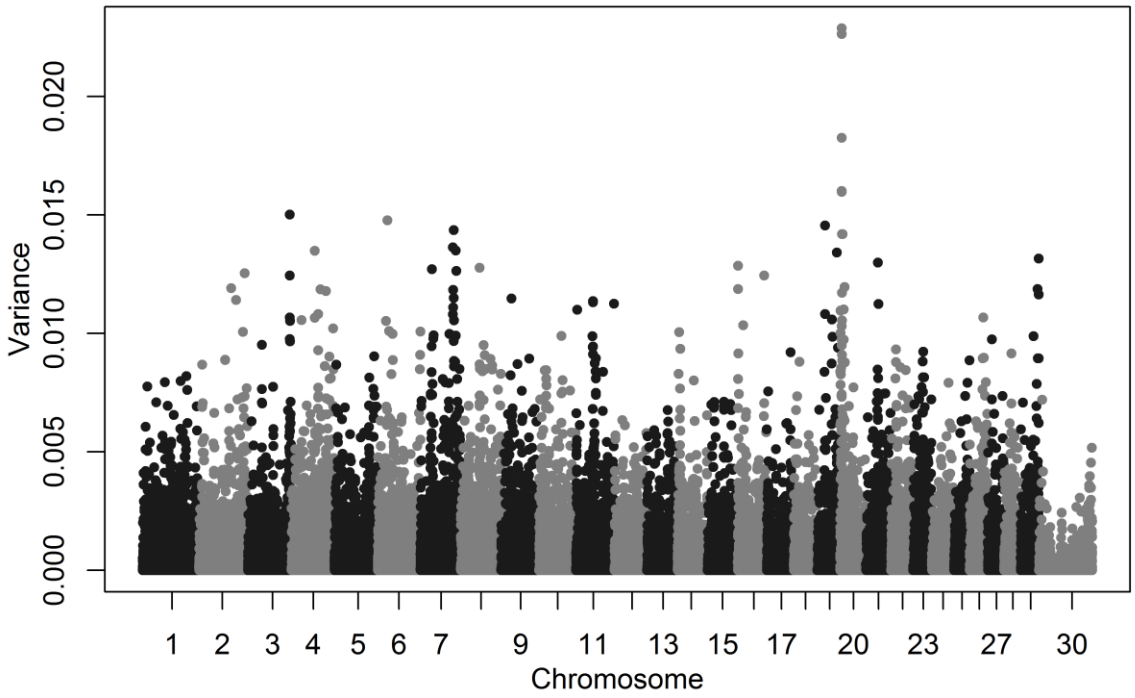

Figure 3.31 Manhattan plot of SNP variances for mature weight.
Figure S25 in publication.

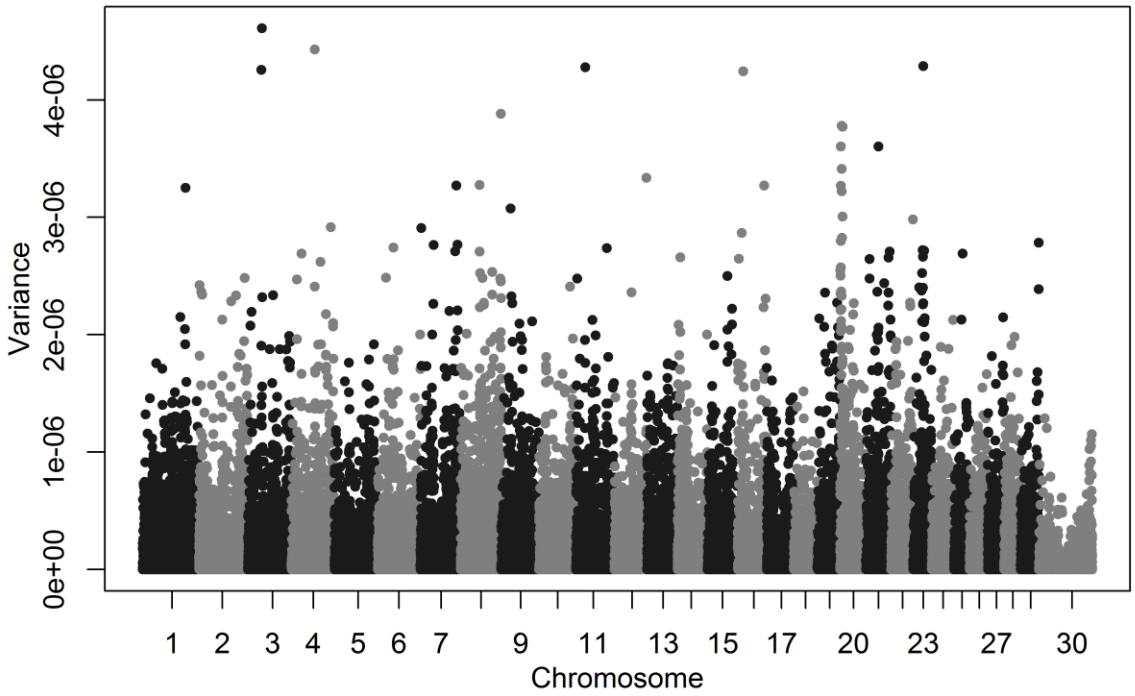

Figure 3.32 Manhattan plot of SNP variances for mature height.
Figure S26 in publication.

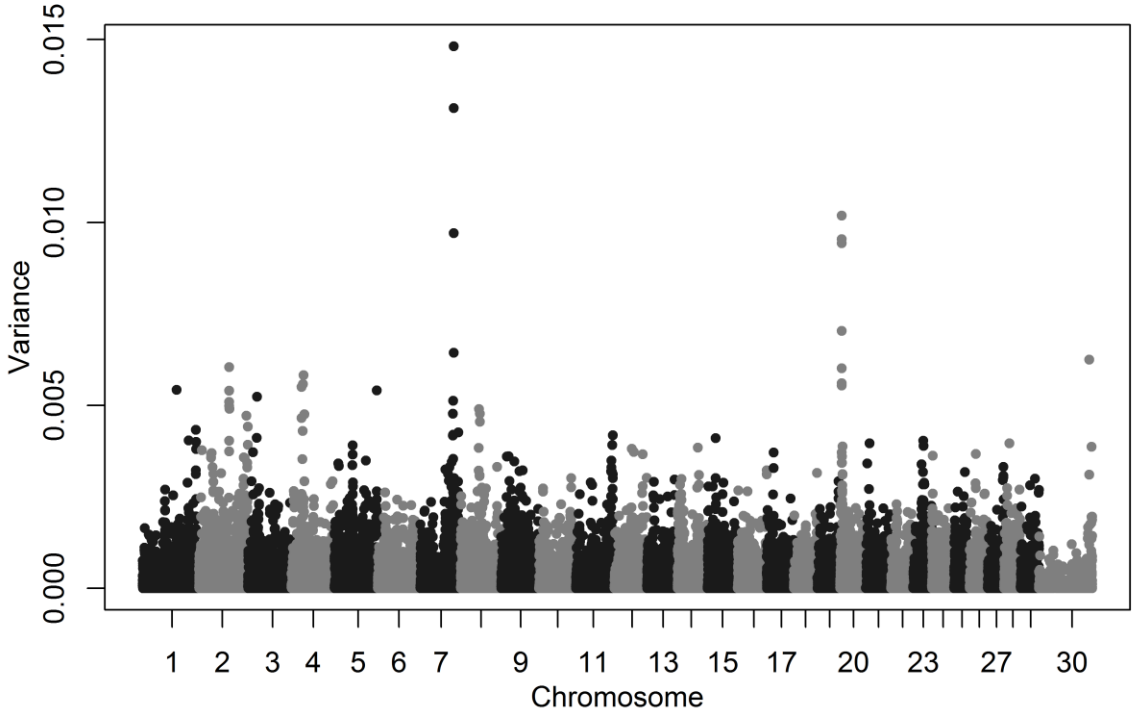

Figure 3.33 Manhattan plot of SNP variances for carcass weight.
Figure S27 in publication.

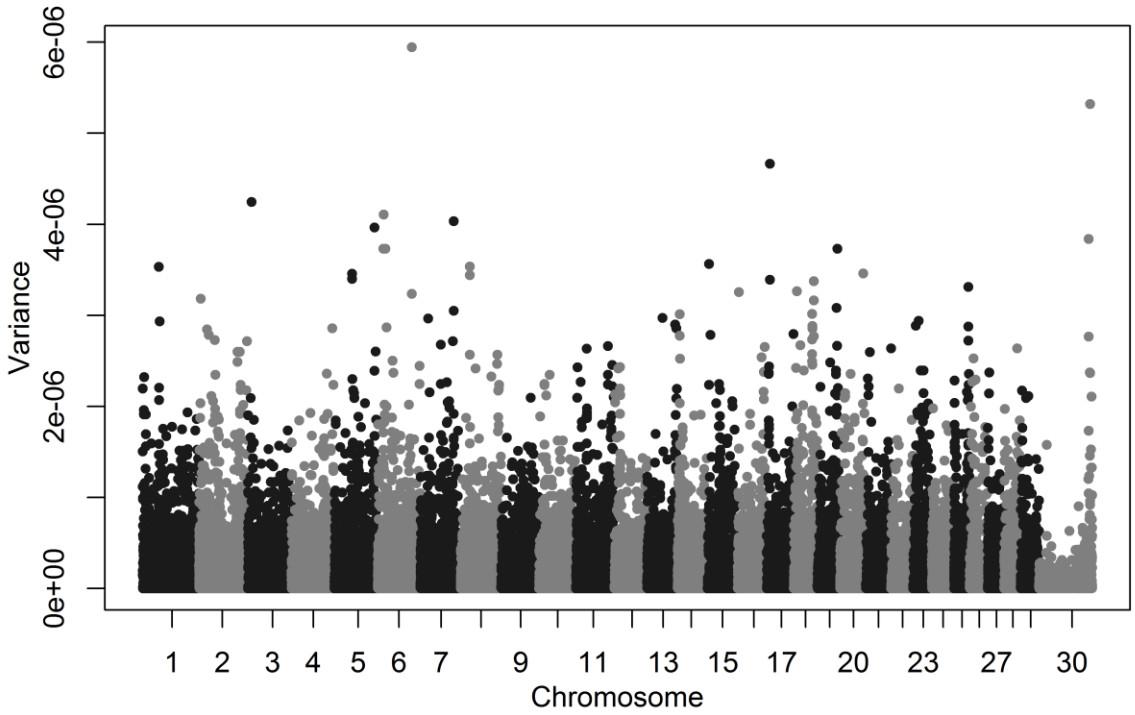

Figure 3.34 Manhattan plot of SNP variances for marbling.
Figure S28 in publication.

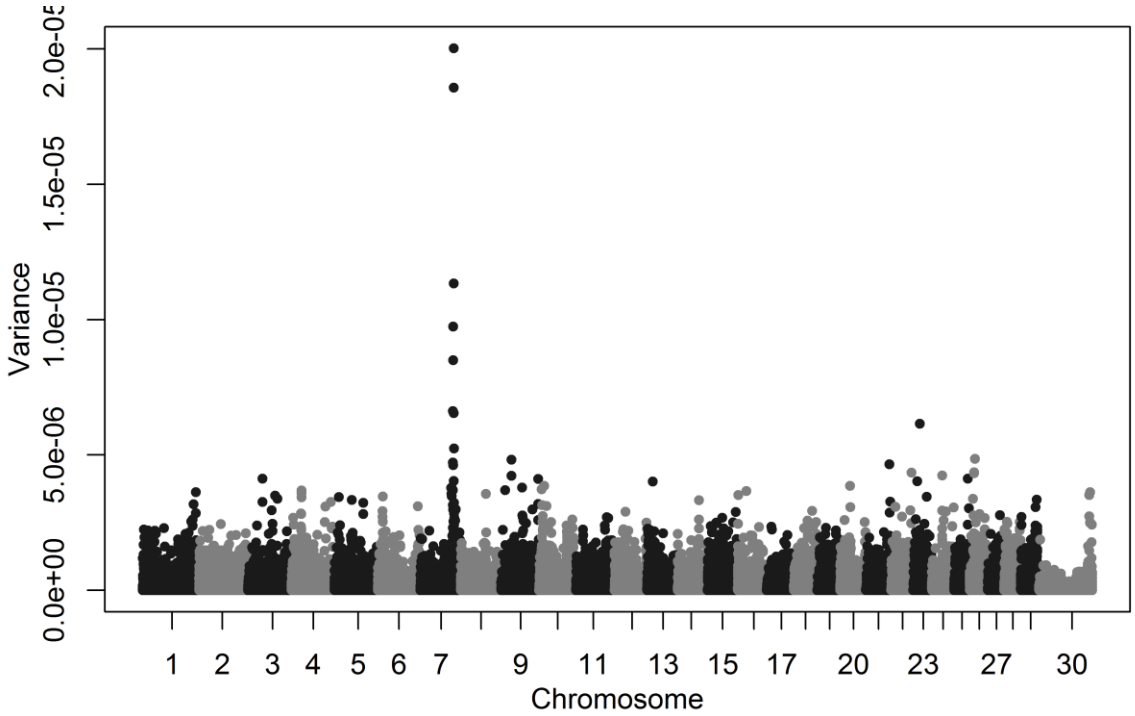

Figure 3.35 Manhattan plot of SNP variances for ribeye area.
Figure S29 in publication.

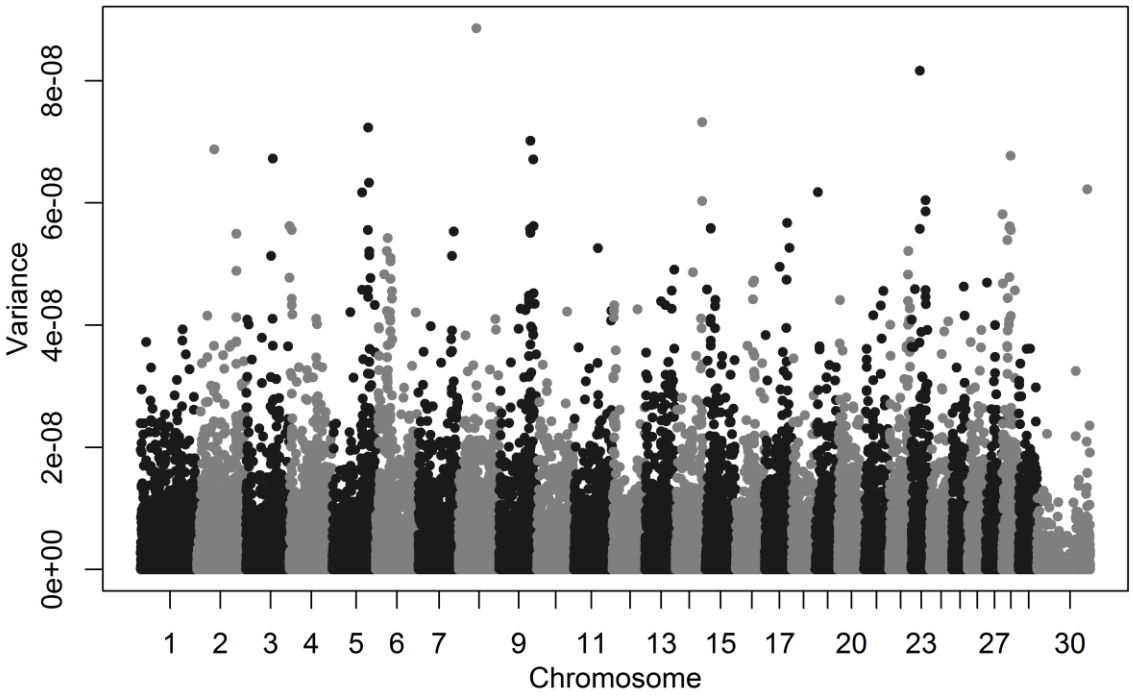

Figure 3.36 Manhattan plot of SNP variances for fat thickness.
Figure S30 in publication.

Table 3.7. Regression of deregressed EBV on birth date for 16 production traits

Trait	Model Type	AIC	Adjusted* R^2	Model p-value	Term	Estimate	Std. Error	t-value	p-value
CED	Linear	28124.6	0.0627	<2.2e-16	Int	-1113	76.3	-14.6	<2e-16
					BD	0.5611	0.0382	14.7	<2e-16
	Quadratic	28120.2	0.0643	<2.2e-16	Int	28375.6	11609.6	2.44	0.0146
					BD	-29.04	11.65	-2.49	0.0128
					BD^2	0.007426	0.00292	2.54	0.0111
BW	Linear	20751.9	0.0017	0.0125	Int	63.03	23.61	2.67	0.0076
					BD	-0.02952	0.01181	-2.5	0.0125
	Quadratic	20664.1	0.0286	<2.2e-16	Int	-33791.9	3549.5	-9.52	<2e-16
					BD	33.95	3.56	9.53	<2e-16
					BD^2	-0.00853	0.00089	-9.54	<2e-16
WW	Linear	32780.7	0.2909	<2.2e-16	Int	-5518.5	154	-35.8	<2e-16
					BD	2.805	0.077	36.4	<2e-16
	Quadratic	32771.5	0.2931	<2.2e-16	Int	73231.1	23447.4	3.12	0.0018
					BD	-76.23	23.53	-3.24	0.0012
					BD^2	0.01983	0.0059	3.36	0.0008
YW	Linear	31056.9	0.3085	<2.2e-16	Int	-9752.4	281.4	-34.7	<2e-16
					BD	4.9598	0.1409	35.2	<2e-16
	Quadratic	31055.9	0.309	<2.2e-16	Int	64038.7	42599.6	1.5	0.1329
					BD	-69.13	42.77	-1.62	0.1061
					BD^2	0.0186	0.01074	1.73	0.0833
YH	Linear	7284.51	-0.0003	0.5274	Int	4.4965	5.9446	0.76	0.4495
					BD	-0.00188	0.00298	-0.63	0.5274
	Quadratic	7221.19	0.0279	5.59e-15	Int	-7278.5	895.2	-8.13	6.97e-16
					BD	0.315	0.899	8.13	6.81e-16
					BD^2	-0.00184	0.00023	-8.14	6.69e-16

Cont. Table 3.7 – Regression of deregressed EBV on birth date for 16 production traits

Trait	Model Type	AIC	Adjusted R^2	Model p-value	Term	Estimate	Std. Error	t-value	p-value
SC	Linear	9875.76	0.0548	<2.2e-16	Int	-99.82	8.34	-11.96	<2e-16
					BD	0.0503	0.0042	12.03	<2e-16
	Quadratic	9873.41	0.0561	<2.2e-16	Int	2556.3	1273.2	2.01	0.0448
					BD	-2.617	1.278	-2.05	0.0408
					BD^2	0.00067	0.00032	2.09	0.0371
DOC	Linear	14425.1	0.0079	0.0006	Int	-1202.2	352.6	-11.96	0.0007
					BD	0.6095	0.1765	3.45	0.0006
	Quadratic	14426.9	0.0073	0.0024	Int	23515.7	59954.3	0.39	0.695
					BD	-24.19	60.14	-0.4	0.6876
					BD^2	0.006218	0.01508	0.41	0.6802
HP	Linear	6261.8	0.0029	0.0821	Int	455.72	252.83	1.8	0.0719
					BD	-0.2205	0.1267	-1.74	0.0821
	Quadratic	6263.68	0.0016	0.2087	Int	-17220.1	52512.4	-0.33	0.7431
					BD	17.52	52.7	0.33	0.7397
					BD^2	-0.00445	0.01322	-0.34	0.7365
CEM	Linear	17952.7	0.0436	<2.2e-16	Int	-1167	123.9	-9.42	<2e-16
					BD	0.5912	0.0621	9.52	<2e-16
	Quadratic	17951.2	0.0448	<2.2e-16	Int	-38160	19856.5	-1.92	0.0548
					BD	37.79	19.97	1.89	0.0586
					BD^2	-0.00935	0.00502	-1.86	0.0626
MILK	Linear	19571.3	0.1602	<2.2e-16	Int	-2819.6	143.6	-19.64	<2e-16
					BD	1.4304	0.072	19.88	<2e-16
	Quadratic	19572.4	0.1601	<2.2e-16	Int	-24577.3	22918.4	-1.07	0.2837
					BD	23.31	23.04	1.01	0.3119
					BD^2	-0.0055	0.00579	-0.95	0.3425

Cont. Table 3.7 – Regression of deregressed EBV on birth date for 16 production traits

Trait	Model Type	AIC	Adjusted R^2	Model p-value	Term	Estimate	Std. Error	t-value	p-value
MW	Linear	16702.55	0.0114	6.00e-05	Int	-3729.2	943	-3.96	8.08e-05
					BD	1.9039	0.4729	4.03	6.00e-05
	Quadratic	16672.11	0.0346	3.00e-11	Int	-876824	152533.8	-5.75	1.12e-08
					BD	879.5	153.3	5.74	1.20e-08
					BD^2	-0.2205	0.0385	-5.72	1.29e-08
MH	Linear	5752.4	0.0072	0.0013	Int	-50.52	16.01	-3.16	0.0016
					BD	0.0259	0.008	3.22	0.0013
	Quadratic	5724.3	0.0294	1.71e-09	Int	-14675.8	2653.3	-5.53	3.85e-08
					BD	14.73	2.67	5.52	4.05e-08
					BD^2	-0.0037	0.00067	-5.51	4.27e-08
CW	Linear	28587.12	0.0658	<2.2e-16	Int	-5169.4	394.3	-13.11	<2e-16
					BD	2.6028	0.1974	13.19	<2e-16
	Quadratic	28585.05	0.0669	<2.2e-16	Int	112383.2	58298.6	1.93	0.054
					BD	-115.4	58.5	-1.97	0.0487
					BD^2	0.0296	0.0147	2.02	0.0439
MARB	Linear	9903.56	0.0389	<2.2e-16	Int	-53.54	4.72	-11.35	<2e-16
					BD	0.0271	0.0024	11.48	<2e-16
	Quadratic	9905.41	0.0386	<2.2e-16	Int	223.7	716.3	0.31	0.7548
					BD	-0.251	0.7186	-0.35	0.7269
					BD^2	0.00007	0.00018	0.39	0.6988
RE	Linear	9365.74	0.0356	<2.2e-16	Int	-46.58	4.24	-10.98	<2e-16
					BD	0.0234	0.0021	11.02	<2e-16
	Quadratic	9367.72	0.0353	<2.2e-16	Int	53.95	646.5	0.08	0.9335
					BD	-0.0775	0.6485	-0.12	0.9049
					BD^2	0.000025	0.000163	0.16	0.8764

Cont. Table 3.7 – Regression of deregressed EBV on birth date for 16 production traits

Trait	Model Type	AIC	Adjusted R^2	Model p-value	Term	Estimate	Std. Error	t-value	p-value
FAT	Linear	-2720.33	0.0074	7.22e-07	Int	-3.32	0.67	-4.93	8.81e-07
					BD	0.001676	0.000338	4.97	7.22e-07
	Quadratic	-2732.61	0.0115	3.73e-09	Int	381.9	101.9	3.75	0.0002
					BD	-0.3848	0.1022	-3.76	0.0002
					BD^2	0.000097	0.000026	3.78	0.0002

Table S1 in publication.

Additional file 2 – GBLUP_DAVID.xls

Chart of enriched GO terms in Excel xls format. We included charts for DAVID's GOTERM_BP_FAT, GOTERM_BP_ALL, GOTERM_BP_1, GOTERM_BP_2, GOTERM_BP_3, GOTERM_BP_4, GOTERM_BP_5, GOTERM_CC_FAT, GOTERM_MF_FAT, KEGG_PATHWAY, UP_TISSUE, SP_comment, and SP_PIR_KEYWORDS with each as an individual tab in the file. We supplied the DAVID resources with a list of 4,216 genes annotated in the UMD 3.1 assembly. See the publication's online supporting material for this file.

Additional file 3 – GenSel_DAVID.xls

Chart of enriched GO terms in Excel xls format. We included charts for DAVID's GOTERM_BP_FAT, GOTERM_BP_ALL, GOTERM_BP_1, GOTERM_BP_2, GOTERM_BP_3, GOTERM_BP_4, GOTERM_BP_5, GOTERM_CC_FAT, GOTERM_MF_FAT, KEGG_PATHWAY, UP_TISSUE, SP_comment, and SP_PIR_KEYWORDS with each as an individual tab in the file. We supplied the DAVID resources with a list of 4,033 genes annotated in the UMD 3.1 assembly. See the publication's online supporting material for this file.

4. WORLDWIDE PATTERNS OF EXPORTATION, ADMIXTURE AND SELECTION IN DOMESTICATED CATTLE

Jared E Decker, Kefei Chen, Alan Cooper, Carl Halbirt, Allan Roberts, Stephanie D McKay, Megan M Rolf, JaeWoo Kim, Antonio Molina, Tad S Sonstegard, Olivier Hanotte, Anders Götherström, Christopher M Seabury, Lisa Praharani, Masroor Ellahi Babar, Mehmet Ali Yildiz, Michael P. Heaton, Wansheng Lui, James M Reecy, Muhammad Saif-Ur-Rehman, Robert D Schnabel, Jeremy F Taylor

Abstract

Using genotypes from 47,282 autosomal single nucleotide polymorphism markers, we evaluate the population structure of 114 domesticated bovid breeds. Patterns of geographic dispersal resulting from cattle exportation are recognizable in phylograms and phylogenetic networks calculated from F_{ST} values. Hybridization occuring after exportation is evident from principal component and admixture analysis. We also identify a cline of *Bos taurus taurus*/*Bos taurus indicus* hybridization in Asia. Iberian, Anatolian, Italian, and East Asian cattle are shown to have introgression from African taurine. Additionally, we demonstrate that three 400 year old teeth found in a Spanish well in St. Augustine, Florida come from an animal with Iberian ancestry and that are closely related to American Criollo breeds. Finally, we show that selection has acted on

the same seven genomic regions in four separate beef and dairy breeds. We argue that exportation, admixture, and selection have all been important forces in shaping bovine genomic variation.

Keywords

migration, admixture, domestication, selection, cattle

Introduction

High-throughput genotyping assays have allowed population geneticists to use genome-wide markers to analyze the histories of many species, including human (Jakobsson et al. 2008; Li et al. 2008), cattle (Gibbs et al. 2009; Decker et al. 2009), sheep (Kijas et al. 2009), dog (Vonholdt et al. 2010), grape (Myles et al. 2011), and horse (McCue et al. 2012). In previous work, we described the structure of domestic bovine populations using their genetic variation inferred from a sample of roughly 41,000 single-nucleotide polymorphisms (Decker et al. 2009). Although we sampled 48 cattle breeds in this previous work, we did not have samples from key areas such as China and Southeast Asia, Anatolia, the Baltic States, southern Africa, and the Iberian Peninsula. As a result of those gaps in geographic sampling, several questions remained. What is the population structure of cattle in central and Southeast Asia? Are Iberian cattle admixed with introgression from African cattle? Also, how has selection shaped the genomes of domestic cattle?

Further, we report some of the earliest archaeological and genetic evidence of cattle in the Americas. Essentially, we have discovered the direct remnants of Iberian

cattle transported across the Atlantic Ocean by the early Spanish settlers. We evaluate the relationship of ancient cattle teeth to modern Criollo cattle using mitochondrial DNA sequences.

We also have assembled a genomic dataset which represents the largest population sampling of any mammalian species. This allows for an extremely detailed description of the population structure of domesticated cattle worldwide. Using this data set, we accurately establish the patterns of exportation, admixture, and selection for domesticated cattle.

Results and Discussion

We describe the phylogenetic relationships between 114 breeds of domesticated bovids (Figures 4.1, 4.2, and 4.3 and Table 4.1). These breeds split into three main clades corresponding to three domesticated (sub)species: *Bos javanicus*, *Bos taurus indicus* and *Bos taurus taurus*. The principal source of SNP genotype variation is between *Bos t. taurus* and *Bos t. indicus* breeds (Figures 4.4 and 4.5). This split corresponds to the cattle which originated from the two separate major centers of domestication in the Fertile Crescent and Indus Valley. The second principal component splits Bali and Shorthorn cattle versus N'Dama, an African taurine breed. Although *Bos javanicus* has a more distant common ancestor compared to *Bos t. indicus* and *Bos t. taurus* (Decker et al. 2009), the uneven sample sizes and SNP ascertainment (McVean 2009) cause the *Bos t. indicus*/*Bos t. taurus* split to be the main source of variation.

Despite our deep sampling, it remains unclear whether cattle were independently domesticated in western Africa. Principal component (Figure 4.4) and admixture analyses (Figure 4.6) show that African taurine animals (the N'Dama breed in our dataset) represent the most divergent of the taurine populations. N'Dama also have distinctive patterns of linkage disequilibrium, causing them to have different ancestral effective population size estimates than either *Bos t. taurus* or *Bos t. indicus* breeds (Gibbs et al. 2009). However, the admixture results for the Anatolian breeds (Figures 4.5, 4.6, 4.7, and 4.8) complicate the interpretation of the African taurine results. Anatolian animals share a large portion of ancestry with African taurine. Is this because cattle domesticated in the Fertile Crescent are the source of African *Bos t. taurus*? Or, were there separate domestications in the Fertile Crescent, the Indian subcontinent *and* Africa, with modern Anatolian breeds being a mixture of cattle domesticated in these three regions? The placement of Anatolian breeds along principal component 1 in Figure 4.4 (McVean 2009) and their extremely short branch lengths in Figure 4.3 lead us to believe that modern Anatolian breeds are indeed admixed. However, because this issue is not conclusively resolved, we will refer to three ancient population centers (regardless of whether they were domestication centers) as being India, Anatolia, and Africa.

Early farmers expanded their range because of the advantage of a reliable supply of food and likely displaced the indigenous hunter-gatherer populations by introducing new diseases (Diamond 2002). The genomes of modern cattle reflect the history of animal exportations out of ancient cattle population centers by these migratory farmers.

In addition to the routes previously described from the Fertile Crescent to Europe (Decker et al. 2009), here we show imprints resulting from exportations out of the Indian subcontinent to China and southeast Asia, India to Africa, India to the Americas, Africa to the Iberian peninsula and Italy, and Europe to the Americas (Figure 4.9). Subsequent to these initial exportations, there have been countless exportations and importations of cattle worldwide. When domesticated cattle were present and new germplasm was imported, the new cattle were often crossed with the local cattle resulting in an admixed population. Admixed populations are most readily identified when *Bos t. indicus* and *Bos t. taurus* animals were crossed, such as in China, Africa, and the Americas (crosses in Figures 4.4 and 4.9).

In the late 18th and 19th centuries, cattlemen began forming closed herds termed breeds (Felius 1995). Because breeds are typically discrete units with little or no interbreeding, as we continue to subdivide the data with our admixture analysis (e.g., increasing *K*), the cross-validation error estimates continued to decrease (Table 4.2). This reflects the large differences in allele frequencies between breeds due to separate domestication events, geographic dispersal and isolation, breed formation and artificial insemination.

Cattle breeds in Asia were derived from cattle imported from the Fertile Crescent and the Indian subcontinent. In addition to typical cattle domesticated from aurochs (*Bos primigenius*), bovids were also domesticated from water buffalo (*Bubalus bubalis*), gaur (*Bos gaurus*), and banteng (*Bos javanicus*), which comprise the Bali breed, which we have sampled (Cockrill 1974; Felius 1995). Although other bovid species have

been domesticated in Asia, most cattle are of *Bos t. taurus* or *Bos t. indicus* descent. Cattle in the north and northeast are primarily of *Bos t. taurus* ancestry (Figure 4.5; HANW, WAGY, and MG). Cattle in Pakistan, India, southern China and Indonesia are predominantly *Bos t. indicus* (Figure 4.5; ONG, MAD, BRE, HN, ACE, PES, ACH, HAR, BAG, GUZ, SAHW, GBI, CHO, GIR, KAN, THA, RSIN, HIS, LOH, ROJ, DHA, and DAJ). Cattle located between these two geographical regions are *Bos t. taurus*/*Bos t. indicus* hybrids (Figures 4.5 and 4.9; QC and LX).

Cattle in Africa also have a gradient of indicine ancestry. N'Dama cattle in the west range from 0% to 17.6% indicine ancestry. Moving from west to east and from south to central, the percent indicine rises from 19.6% to 71.8% (Figure 4.6, AFR, TULI, ANKW, SHK, ZEB, and BOR). These figures likely result from two known waves of indicine importations to Africa: the first occurring in the second millennium BC and the second during and after the Islamic conquests (Ajmone-Marsan et al. 2010; Felius 1995; Hanotte et al. 2002).

Cattle were imported into Europe from the southeast to the northwest. Breeds formed in the British Isles separate into two clades (Figure 4.1), probably representing two distinct importations into the British Isles, one from the south and one from the north (Figure 4.9). The descendants of Durham Shorthorns are the most distinct group of European cattle as they cluster at the extremes of principal component 2 (upper left hand corner of Figure 4.3), and they form a separate cluster when $K = 4$ and $K = 5$ (Figures 4.6 and 4.7). This is for two reasons. First, as shown in Figure 4.7, many breeds share ancestry with Shorthorn cattle, including Milking Shorthorn, Beef Shorthorn,

Lincoln Red, Maine-Anjou, Belgian Blue, Holstein, Charolais, Santa Gertrudis, and Beefmaster (Felius 1995). Second, Shorthorn cattle have higher levels of inbreeding causing them to have longer branch lengths in Figures 4.1, 4.2, and 4.3.

From the previous placement of American Criollo breeds such as Romosinuano, Texas Longhorn, and Corriente, we hypothesized that Iberian cattle were admixed as a result of introgression of cattle from Africa with the local European cattle (Decker et al. 2009). Genotyping individuals from 11 Spanish breeds supported, but clarified, this hypothesis. On average, Spanish cattle have 8.7% of African ancestry, with a minimum of 0% and a maximum of 15.8%. These results are supported by previous analyses of mitochondrial DNA (Mirol et al. 2003; Liron et al. 2006). The remaining Spanish cattle had no indicine ancestry. In addition, Italian cattle (PIED, MCHI, CHIA, and RMG) share ancestry with African taurine (See Figure 4.6). These data show that the reason the American Criollo breeds were found to be more basal in our previous work is their higher percentage of indicine ancestry. For the 5 sampled American Criollo breeds, they had, on average, 4.6% African ancestry (minimum of 0% and maximum of 12.1%) and 7.0% indicine ancestry (minimum of 0% and maximum of 19.6%).

In 2009, three teeth were found inside an old Spanish well in St. Augustine, Florida. They were subsequently identified as cattle teeth, and, because the teeth were found together, we assume they are from the same individual. These teeth are a remarkable finding in the United States due to their antiquity. They were dated to approximately 1600 AD using pottery and other debris found with the samples and samples from all three teeth were radiocarbon dated. Two of the teeth had extremely

wide calibrated date confidence intervals, but the third tooth was dated as originating in 1564 (see Figure 4.10). This date is remarkably close to the settlement of St. Augustine by Pedro Menéndez de Avilés in 1565 (Deagan 1985; Bushnell 1978). The animal from which the teeth originated could have belonged to the *hacienda de la chua* operated by the Menéndez Marquéz family, who were related to Pedro Menéndez de Avilés (Bushnell 1978). Regardless of the animal's true identity, in essence, we have caught Spanish settlers "in the act" of exporting cattle to Florida.

To determine if this ancient animal is related to modern Florida Crackers, we compared its DNA to that from 9 extant Florida Cracker samples. These modern Florida Cracker samples come from a single herd with a closed pedigree that can be traced back for over 100 years with no evidence of introgression from other breeds during this time. In a clean, dedicated ancient DNA laboratory, DNA was extracted from each of the three teeth. Primers were designed to sequence an approximately 600 base pair fragment of the mitochondrial D-loop, resulting in ancient DNA sequences, the longest of which was a 627 base pair fragment from the third sample. All three teeth possessed identical D-loop sequences supporting our conjecture that they originated from the same animal. We also attempted to genotype six ancient DNA libraries (two from each tooth) with the BovineSNP50 BeadChip, but, due to wet and warm environment in Florida, the DNA was too degraded to produce reliable genotypes. For the modern Florida Cracker samples, a 930 base pair fragment from the mitochondrial D-loop was sequenced. In addition to the mtDNA sequences for each of the 9 Florida Cracker samples and the ancient teeth, we downloaded 327 D-loop sequences primarily from Spanish, Portuguese, and

American Criollo breeds from GenBank (Table 4.3) and aligned these sequences using CLUSTALW and CodonCode (Figure 4.11). We calculated a median joining network for these sequences in NETWORK 4.5.16 (Figures 4.12 and 4.13). The ancient teeth haplotype clustered with the most common haplotype, which contained mostly Iberian and Criollo samples. Four of the Florida Cracker haplotypes were one step away in the second most common haplotype, and three more were two steps away. Another two Florida Cracker samples clustered with Canaria, Criollo Poblano, Criollo Nayarit, Criollo Chihuahua, and Ramo Grande samples. The final Florida Cracker sample clustered with two Shorthorn samples and one Holstein sample. This along with their position in Figure 4.1 indicates that modern Florida Cracker cattle have been stongly influenced by crossing with British cattle prior to efforts 100 years ago to conserve the breed.

To further interrogate the relationship between the ancient St. Augustine teeth and modern Florida Crackers, we analyzed the mitochondrial sequences with PhyML to produce a maximum likelihood phylogeny. Figure 4.14 shows the clade containing the ancient teeth haplotype. This clade also contains 3 of the Florida Cracker samples and several Criollo samples. These results suggest that the ancient St. Augustine and Florida Cracker cattle share a very recent ancestor, but the small amount of available DNA data makes it difficult to determine if Florida Cracker cattle are linear descendants of the animal from which the ancient teeth originated.

We have shown that patterns of exportation and admixture have shaped the genomes of domesticated cattle. However, how exactly does selection modify the genome? After cattle are exported to a new environment does natural selection

strongly favor animals that are better suited to that environment? From our genome-wide analysis of 3,570 Angus cattle, we hypothesized that natural selection has altered the innate immunity of Angus cattle. To investigate whether other breeds are experiencing selection at the same loci, we performed a genome-wide association analysis with birth date (measured as a continuous variable with month and day expressed as a fraction of a year) as the dependent variable in a joint data set of 811 Herefords, 1,209 Holsteins, 2,228 Limousins, and 921 Simmental, all belonging to their respective North American breed registries. This analysis identifies genomic loci which are predictive of birth date and that have changed rapidly in frequency due to selection. To identify significant associations with birth date in this across-breed analysis the selected variants must be segregating in most, if not all, of the four breeds, loci must be exposed to similar selection pressures due to the breed being exposed to the same selection criteria and must respond similarly in all 4 breeds, and SNP markers must be in close proximity to the selected variants as linkage disequilibrium extends only over short distances across breeds (Goddard and Hayes 2009; Gibbs et al. 2009).

To account for the population structure between the sampled breeds we identified the principal components for the sample genotype covariance matrix using SMARTPCA (Patterson, Price, and Reich 2006). The first three principal components identified differences between breeds (see Figure 4.15); subsequent principal components identified variation within breeds. Thus, we choose the first three principal components to fit as covariates in our statistical model. To account for pedigree relationships within each of the breeds we fit a Balding-Nichols matrix in a generalized

least squares model as implemented in EMMAX (Kang et al. 2010). Hereford birth years ranged from 1953 to 2008, Holstein birth years ranged from 1952 to 2004, Limousin birth years ranged from 1968 to 2006, and Simmental birth years ranged from 1978 to 2008. The birth date heritability estimated by EMMAX was 0.615, indicating that 61.5% of the variation in birth dates could be explained by the SNP genotypes. As seen in the Q-Q plot in Figure 4.16, our approach effectively accounted for population stratification and relatedness within the data set. In Figure 4.17, we identified significant associations on chromosomes 4, 8, 11, 16, 20, 21, and 23; Table 4.4 contains a list of candidate genes within 100 Kbp of each of these significant SNPs. One of the suggestive SNPs on chromosome 23 at 6,760,915 base pairs is 250 Kbp away from the BOLA-DYA gene, which is part of the major histocompatibility complex in cattle.

The immune system is responding to natural selection in these breeds as mutations in genes involved in phagocytosis and inflammatory response pathways are changing in frequency (Table 4.4). However, loci are also responding to common artificial selection criteria as processes such as limb development, palate development and skeletal muscle tissue development, and possibly appetite are under selection. Limb and palate development may be under artificial selection to decrease the incidence of dystocia, an important trait in both beef and dairy production. As producers seek to increase the production of meat and milk, cattle are required to consume more feed, explaining the putative molecular signal of selection for appetite.

To identify loci that have been selected to significantly different allele frequencies in beef versus dairy cattle, we performed a genome-wide association

analysis contrasting beef and dairy cattle. We coded beef cattle as 0, multiple-purpose breeds as 1, and dairy cattle as 2. We were unable to identify breed type in the literature for 6 of the breeds and we excluded Bali cattle since they were domesticated from *Bos javanicus*. Finally, we analyzed 1,064 samples in EMMAX and fit the first 2 principal components (Figure 4.4) as covariates. We identified only one significant region on chromosome 18 comprising SNPs at 14,401,871 and 14,503,218 base pairs associated with differences between beef and dairy cattle (Figure 4.18). There are probably several reasons why we found only a single locus that was predictive of beef versus dairy type. For a genome-wide association study of a complex trait, our sample size was relatively small. Further, within a type, breeds are exposed to selection on diverse criteria. Within beef breeds, some breeds have been selected for increased intramuscular fat (marbling) while other breeds have been selected for extreme muscling. Within dairy breeds, some breeds have been selected for increased milk fat percentage while others have been selected for increased total milk volume. Lastly, there may be very little causal trait variation segregating in common among these 107 cattle breeds due to independent bottlenecks including separate domestication events and breed formations (Figure 4.1, F_{ST} values range from 0.005 to 0.540). To evaluate the influences of common versus private variation, we split the data set into those samples with more than 50% taurnine ancestry and those with more than 50% indicine ancestry and repeated the genome-wide association analysis in these two subsets. In the taurine subset, again the significant association on chromosome 18 was observed and an additional significant association on chromosome 16 was detected (Figure 4.19). In the

indicine subset, significant associations on chromosomes 16 and 18 were not observed, but a significant association on chromosome 14 was identified (Figure 4.20). These results support the hypothesis that little causal trait variation is segregating among breeds and subspecies. However, the genomic region on chromosome 18 identified in the beef/dairy genome-wide association analysis contains biologically relevant variation. The candidate gene, *ANKRD11*, is involved in bone development and skeletal system morphogenesis, reflecting the different body types selected in beef and dairy cattle.

Conclusions

Domestication, exportation, admixture, and breed formation have had tremendous impacts on the variation present within and between cattle breeds. In Asia, Africa, North and South America, cattle breeders have crossbred *Bos t. taurus* and *Bos t. indicus* cattle to produce hybrids which are well suited to the environment and production system. The ancient teeth found in St. Augustine, Florida can only be of Iberian descent and are closely related to the modern Criollo breeds. Even when selected to produce different products, breeds in the same environment are exposed to similar natural and artificial selection pressures. When common variation exists among breeds, common signatures of selection may be identified. However, selection criteria are diverse within a production type and a moderate amount (on average 20%) of variation is breed specific.

Methods

Sample selection

We used 236 samples from Decker et al. (2009), see Table 4.1. We selected samples that had fewer than 10% of missing genotypes and for breeds with fewer than 20 genotyped samples, we used all available samples which passed the genotype call rate filter. For breeds for which we had no pedigree information, the 20 samples with the highest genotype call rates were selected. For breeds for which we had pedigree information, we filtered any animals whose sire or dam was genotyped. For identified half-siblings, we sampled the sibling with the highest genotype call rate. After removing related genotyped animals, we selected the 20 animals with the highest genotype call rate for that breed.

For the birth date analysis we selected all registered animals from within the sampled breeds with genotype call rates higher than 90%.

Genotyping

Samples from both the worldwide breed and the birth date selection analyses were genotyped with the Illumina BovineSNP50 BeadChip. Because the pseudoautosomal region is not well defined in cattle, we defined from 137 Mbp to the end of the X chromosome to represent the pseudoautosomal region of the bovine sex chromosomes. We filtered all SNPs which mapped to "chromosome unknown" of the UMD 3.1 assembly (Zimin et al. 2009). In PLINK (Purcell et al. 2007; Purcell 2009), we removed SNPs with greater than 10% of missing genotypes and with minor allele frequencies less

than 0.0005 (1/[Number of Samples*2]=0.00044). The average total call rate in the remaining individuals was 0.990383.

Principal component analysis and F_{ST} calculation

The sample covariance matrix was decomposed using SMARTPCA, part of EIGENSOFT 4.2 (Patterson, Price, and Reich 2006). To limit the effects of linkage disequilibrium on the estimation of principal components, for each SNP the residual of a regression on the previous two SNPs was input to the principal component analysis (see EIGENSOFT POPGEN README). We also estimated F_{ST} values using SMARTPCA. A neighbor-joining tree was constructed using the NEIGHBOR program of PHYLIP 3.69 (Felsenstein 1989). A NeighborNet network was also created from F_{ST} values using SPLITSTREE 4.12.3 (Huson and Bryant 2006).

Admixture analysis

ADMIXTURE 1.21 was used to evaluate ancestry proportions for *K* ancestral populations (Alexander, Novembre, and Lange 2009). We ran ADMIXTURE using penalized-likelihood with cross-validation for values of *K* from 1 through 26.

Ancient teeth analysis

Three ancient cattle teeth were discovered in a Spanish well in St. Augustine, Florida along with 75 European pottery fragments that weighed 845.8 grams. Three dates were estimated for the trash deposit using the counts and weights for the potshards. A ceramic date of 1613 was estimated when ceramic counts were considered. A mean ceramic date of 1596 was estimated when ceramic weight was the variable. Using a

ceramic weight to number ratio (wt./no.) a date of 1600 was estimated. Tooth samples were sent to the NSF Arizona AMS Facility at the University of Arizona for radiocarbon dating. Radiocarbon dates were calibrated using OxCal version 4.1.5 (Christopher Bronk Ramsey 2009; P. J. Reimer et al. 2009). Ancient DNA was extracted from teeth using the standard phenol/chloroform/Amicon Ultra-4 method (Iwamoto et al. 2007). DNA extractions, PCRs, and Sanger sequencing were set-up and performed in a geographically isolated, dedicated ancient DNA facility at the University of Adelaide, Australia. A ~600 base-pair (bp) fragment of the mitochondrial control region was amplified in one to four (overlapping) fragments, depending on the quality of the specimen. Two-step multiplex PCR amplifications were performed using primers designed for the bovid mitochondrial control region. Multiplex primer sets A and B were set up separately (see Table 4.5). Multiplex PCR was performed in a final volume of 25 µl containing 2 µl of aDNA extract, 1 mg/ml rabbit serum albumin (RSA; Sigma, fraction V), 6 mM MgSO4, 0.2 µM of each primer, 500 µM of each dNTP, 2 U Platinum Taq Hi-Fidelity and 1 × PCR buffer (Invitrogen Ltd., UK). Multiplex PCR conditions were initial denaturation at 95 °C for 2 min, followed by 35 cycles of 94 °C for 15 sec, 55 °C for 20 sec and 68 °C for 30 sec, and a final extension at 68 °C for 10 min at the end of the 35 cycles. Multiplex PCR products were then diluted to 1:10 as template for the second step of simplex PCR. The second step simplex PCR using Amplitaq Gold (Applied Biosystems) or Hotmaster™ Taq DNA polymerase (5Prime, Milton, Qld) was conducted in a final volume of 25 µl containing 1 µl of diluted multiplex PCR product, 2.5 mM $MgCl_2$, 0.4 µM of each primer, 200 µM of each dNTP, 1 U Amplitaq Gold/ Hotmaster Taq

polymerase and 1 × PCR buffer. The second step simplex PCR conditions were initial denaturation at 95 °C for 2 min, followed by 35 cycles of 94 °C for 20 sec, 55 °C for 15 sec and 72 °C for 30 sec, and a final extension at 72 °C for 10 min at the end of the 35 cycles. Samples were also independently replicated.

One-step simplex PCR amplifications using Platinum *Taq* Hi-Fidelity polymerase were performed on a DNA Engine Tetrad2 Peltier Thermal Cycler (Bio-Rad) in a final volume of 25 µl containing 1 µl of aDNA extract, 1mg/ml rabbit serum albumin (RSA; Sigma, fraction V), 2 mM MgSO4, 0.6 µM of each primer, 250 µM of each dNTP, 1.25 U Platinum *Taq* Hi-Fidelity and 1 × PCR buffer (Invitrogen Ltd., UK). The conditions for PCR amplification were initial denaturation at 95 °C for 2 min, followed by 50 cycles of 94 °C for 20 sec, 55 °C for 20 sec and 68 °C for 30 sec, and a final extension at 68 °C for 10 min at the end of the 50 cycles. Negative extraction controls as well as non-template PCR controls were used throughout all experiments. PCR products were then checked by electrophoresis on 3.5-4.0% agarose TBE gels, and visualized after ethidium bromide staining on an UV transilluminator. PCR amplicons were purified using AMPure magnetic beads (Agencourt®, Beckman Coulter) according to manufacturer's instruction.

To sequence the hypervariable mitochondrial control region in extant Cracker cattle, DNA was extracted from blood of 9 extant Cracker cattle (Sambrook and Fritsch 1989). The PCR reactions contained 10-20 ng of template DNA, 10 pmol of primers A1125 and A1208 (see Table 2), 2 mM $MgSO_4$, 200 µM of each dNTP, 1 × PCR buffer and 1 U Platinum *Taq* Hi-Fidelity (Invitrogen Ltd., UK) in a total reaction volume of 25 µl. The

PCR reaction was performed on a DNA Engine Tetrad2 Peltier Thermal Cycler (Bio-Rad) using a touchdown-PCR profile: initial denaturation for 2 mins at 95 °C, followed by 35 cycles of 30 sec denaturation (94 °C), 30 sec annealing, and 1 min 20 sec extension (68 °C), and a final extension at 68 °C for 10 mins at the end of the 35 cycles. The annealing temperature was 68 °C for the first cycle, decreasing by 1 °C per cycle until 58 °C was reached, then continuing at 58 °C in the annealing step of the remaining cycles. PCR products were then checked by electrophoresis on 1% agarose TBE gels, and visualized after ethidium bromide staining on an UV transilluminator. PCR amplicons were purified using AMPure magnetic beads (Agencourt®, Beckman Coulter) according to manufacturer's instruction.

All purified PCR products were bi-directionally sequenced with the ABI Prism® BigDye™ Terminator Cycle Sequencing Kit version 3.1 (Applied Biosystems). The sequencing reactions were performed in a final volume of 10 µl containing 3.2 pmol of primer, 0.25 µl Bigdye terminator premixture, 1.875 µl of 5 × sequencing buffer. The reaction conditions contained initial denaturation at 95 °C for 2 min, 25 cycles with 95 °C for 10 sec, 55 °C for 15 sec, 60 °C for 2 min 30 sec. Sequencing products were purified using Cleanseq magnetic beads (Agencourt®, Beckman Coulter) according to the manufacturer's protocol. All sequencing reactions were analyzed on an ABI 3130 DNA capillary sequencer (Applied Biosystems, Foster, CA).

Additionally, mitochondrial D-loop sequences were also retrieved for Iberian, American Criollo, and other common North American breeds from GenBank (Benson et

al. 2009). Sequences were preliminarily aligned using CLUSTALW version 2.0.12 (Larkin et al. 2007). Alignments were adjusted by hand in CodonCode (version 3.7.1, Codon Code Corporation). FASTA files were converted to Fluxus NETWORK input using Fluxus' DNA Aligner (Fluxus 2010a). In NETWORK version 4.5.16 (Fluxus 2012), mitochondrial haplotypes were consolidated using star contraction (Forster et al. 2001). Haplotype networks were constructed using the median-joining method (Bandelt, Forster, and Röhl 1999) followed by maximum parsimony post-processing (Polzin and Daneshmand 2003). Figures were generated in NETWORK PUBLISHER (Fluxus 2010b).

Mitochondrial D-loop sequences were also analyzed using PhyML version 3.0 (Guindon and Gascuel 2003). A general time reversible with invariant sites model was used to model sequence evolution. An initial tree was estimated using BioNeighbor-Joining, and Nearest Neighbor Interchange and Subtree Pruning Regrafting were used for searching tree topologies.

Birth date selection analysis

Birth dates were expressed as a continuous variable by subtracting 1950 from the birth year and expressing month and day as fractions of a year. In PLINK (Purcell et al. 2007; Purcell 2009), SNPs which mapped to chromosome unknown, had call rates less than 90%, or minor allele frequencies less than 0.01 were removed. SMARTPCA was run on all 5,139 samples, again using regression on the previous two SNPs to correct for linkage disequilibrium. With EMMAX, we calculated a Balding-Nichols matrix (Balding and Nichols 1995) as it takes into account population structure. Furthermore, we fit the first

three principal components as covariates in the analysis of birth date using EMMAX. Manhattan plots were created in R (R Development Core Team 2011), with R source code from (Turner 2011) which was altered to allow 31 chromosomes on the x-axis. Genomic regions 100 Kbp on either side of significant SNPs were visualized on the UCSC Genome Browser (Kent et al. 2002; Genome Bioinformatics Group of UC Santa Cruz 2012), using UMD3.1 as the reference sequence assembly. Gene ontology and pathways were retrieved from human and cattle databases on the NCBI Entrez Gene website (Maglott et al. 2007).

Breed type GWAS

Breed type information was obtained from (Felius 1995). A Balding-Nichols matrix and the first two principal components were used to correct for kinship and population structure within the data set. Associations with breed type were estimated using EMMAX. Gene information for SNP associated with breed type was retrieved as described for the birth date selection analysis.

Acknowledgments

We gratefully acknowledge the provision of samples and genotypes from breed associations, cattle breeders, semen distributors and the Bovine HapMap Project.

Figures

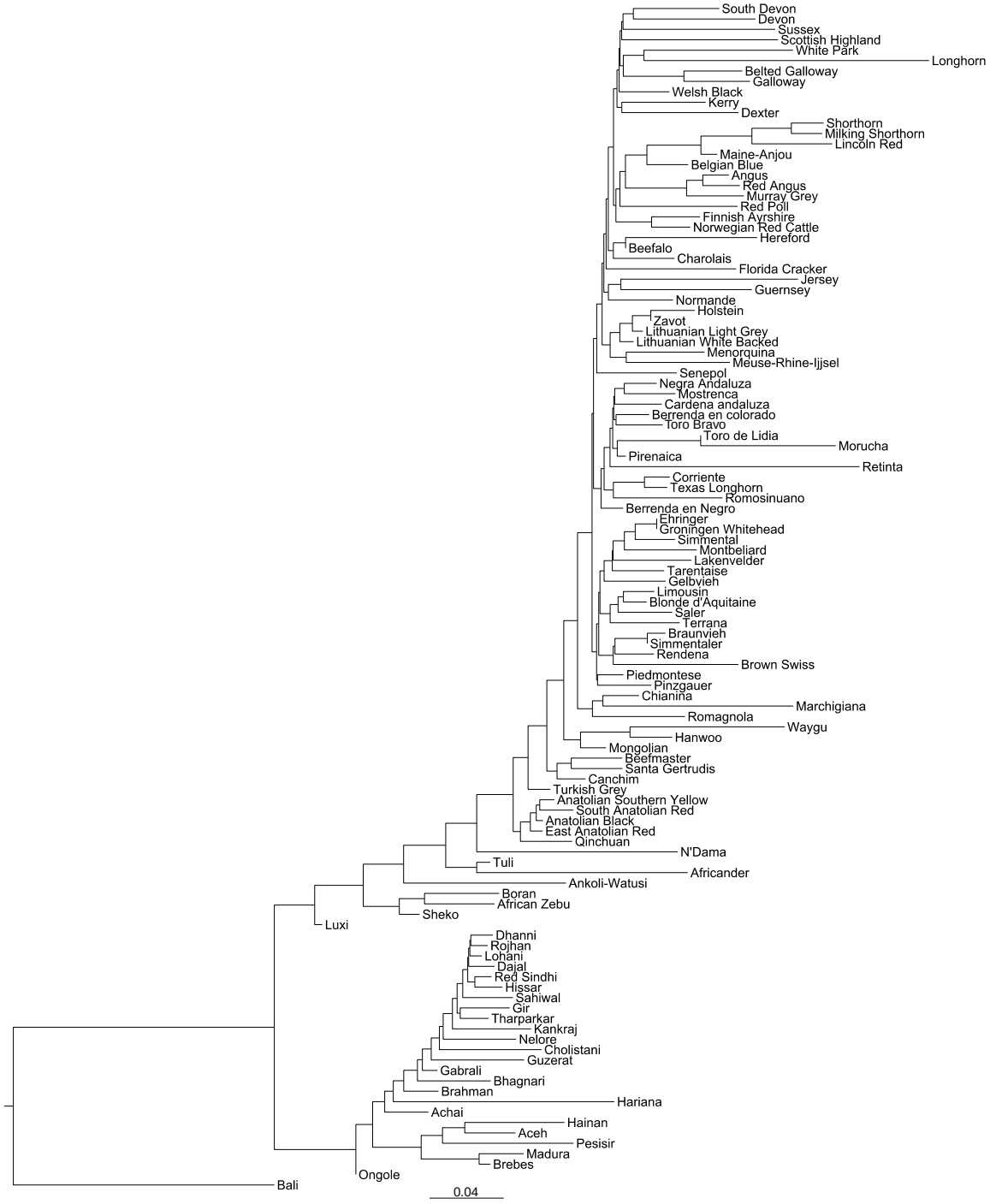

Figure 4.1. Neighbor joining tree calculated from F_{ST} distances.

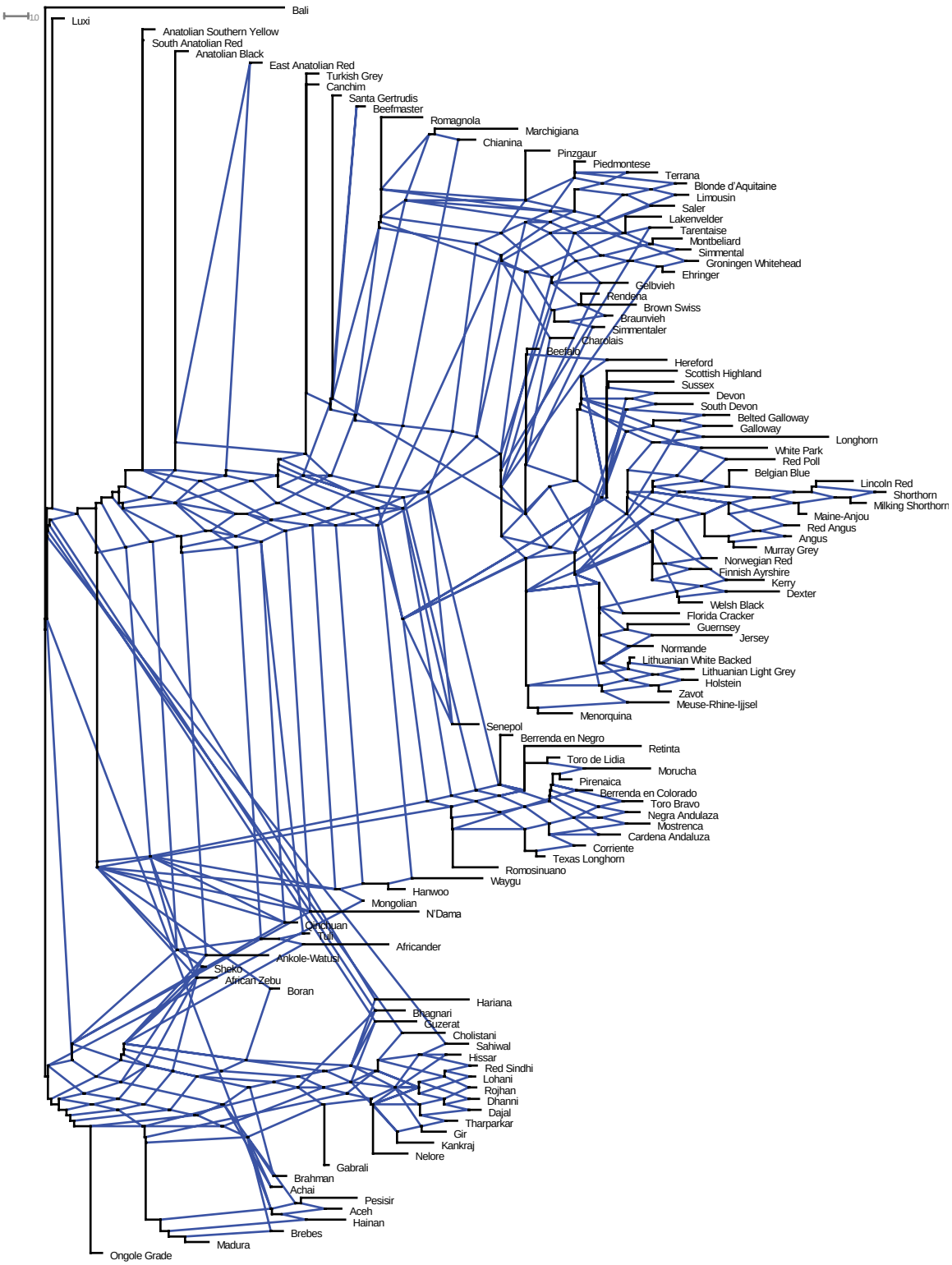

Figure 4.2. A cluster network drawn as a rectangular phylogram.
Splits were calculated using NeighborNet from F_{ST} distances.

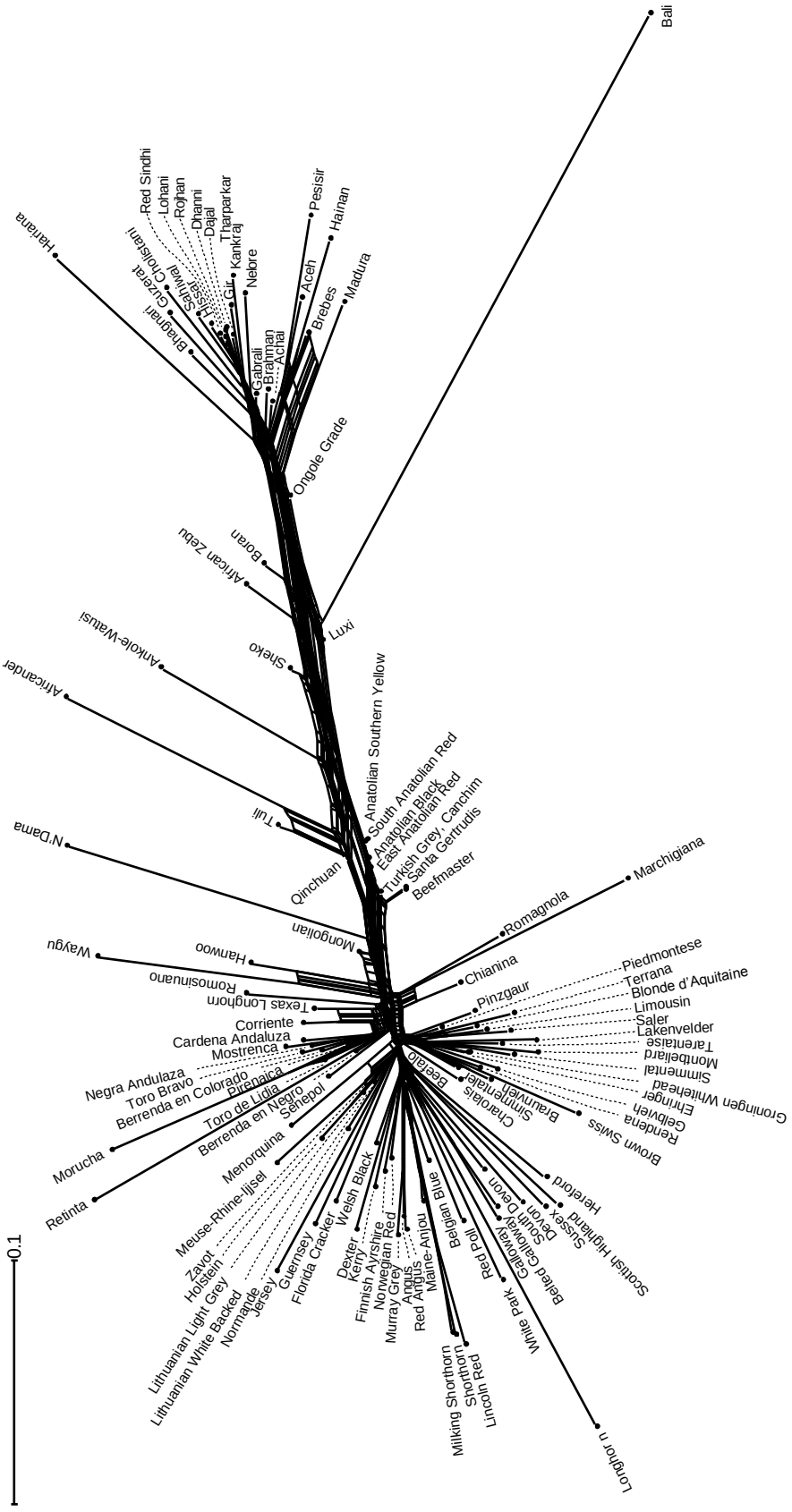

Figure 4.3. Equal angle NeighborNet network calculated from F_{ST} values for 114 cattle breeds.

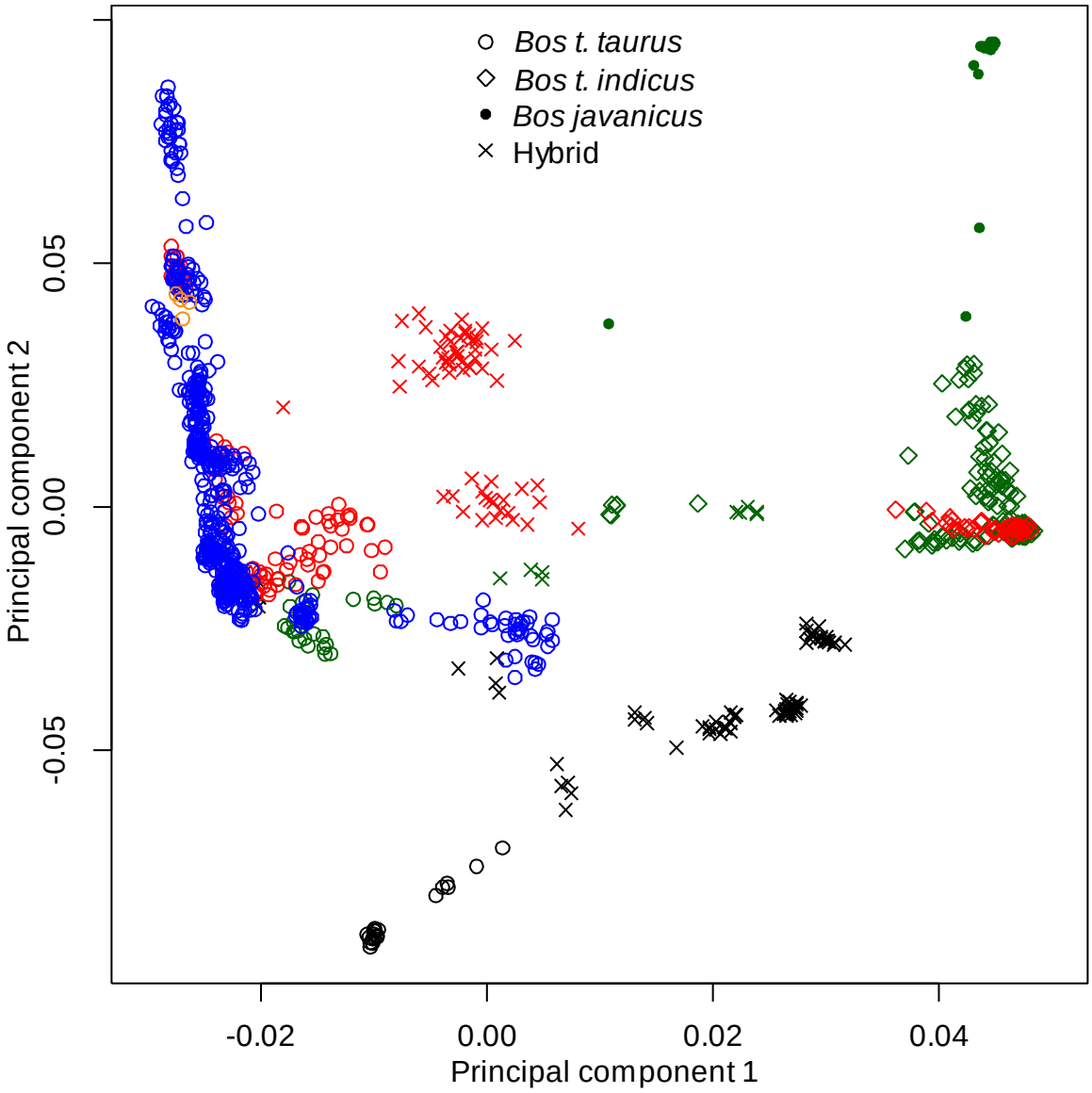

Figure 4.4. Principal component analysis of 1,143 animals genotyped for 47,282 autosomal SNPs.
Principal component 1 separates *Bos t. taurus* from *Bos t. indicus* and *Bos t. javanicus* animals. Principal component 2 separates *Bos t. javanicus* and Shorthorn breeds from N'Dama. In addition to the hybrid breeds, 4 of the Angole grade animals, several of the Anatolian animals, and 3 of the Bali (*Bos javanicus*) animals appear to be hybrids. Samples in green are from Asia, black are from Africa, blue are from Europe, orange are from Australia, and red are from the Americas.

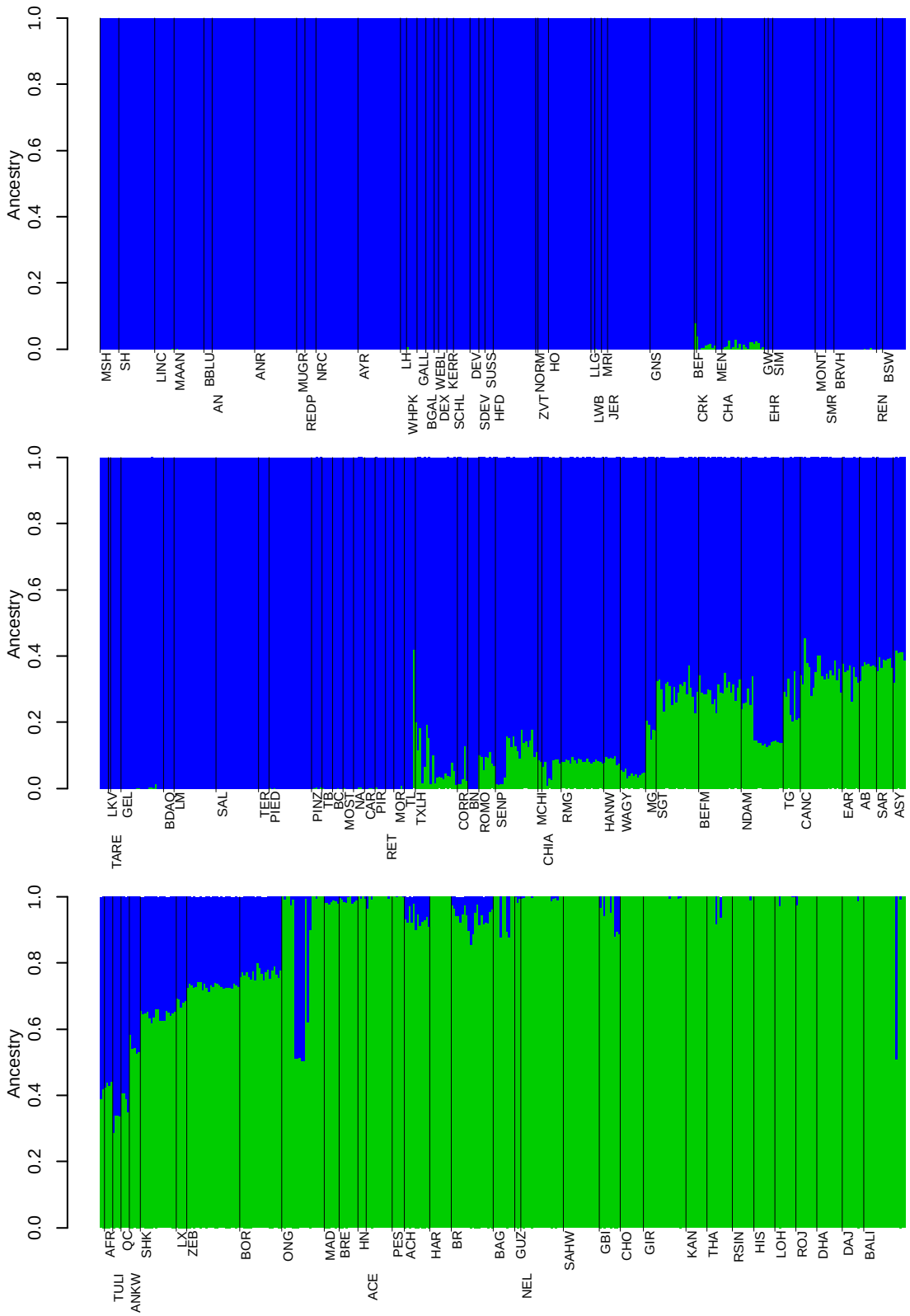

Figure 4.5. Plot of ancestry fractions for 1,143 animals with K = 2.
Breed key is in Table 4.1.

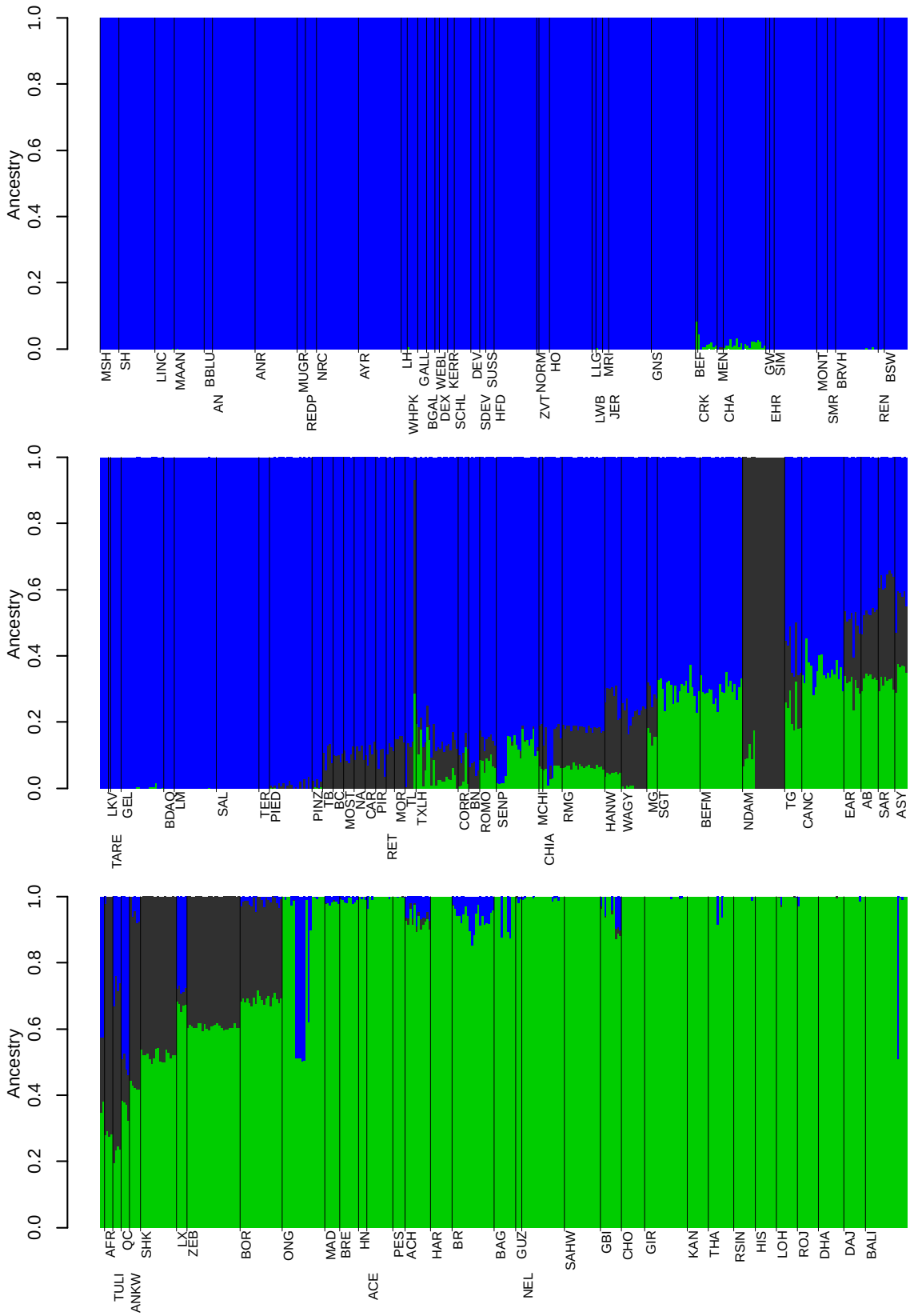

Figure 4.6. Plot of ancestry fractions for 1,143 animals with K = 3.

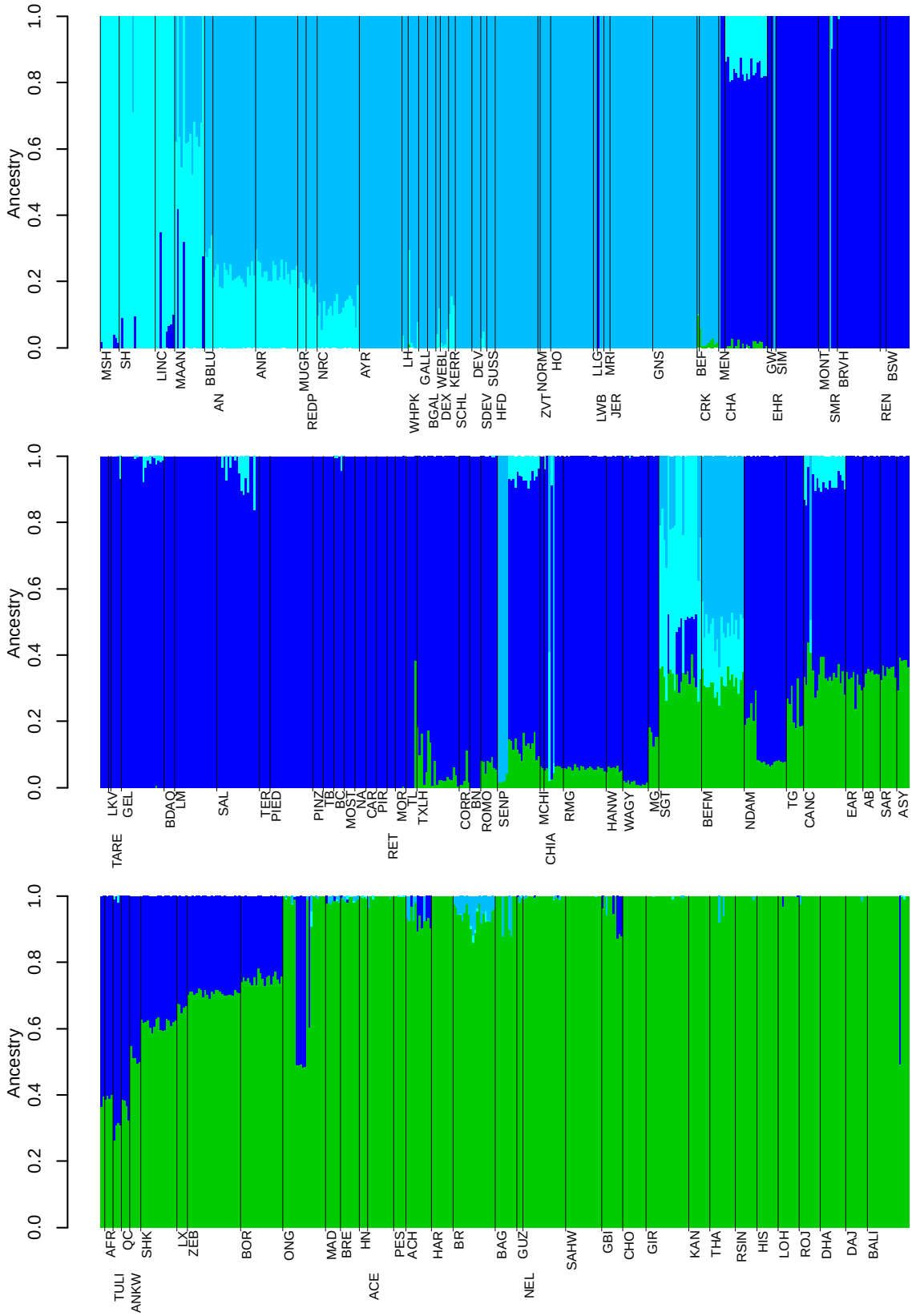

Figure 4.7. Plot of ancestry fractions for 1,143 animals with K = 4.

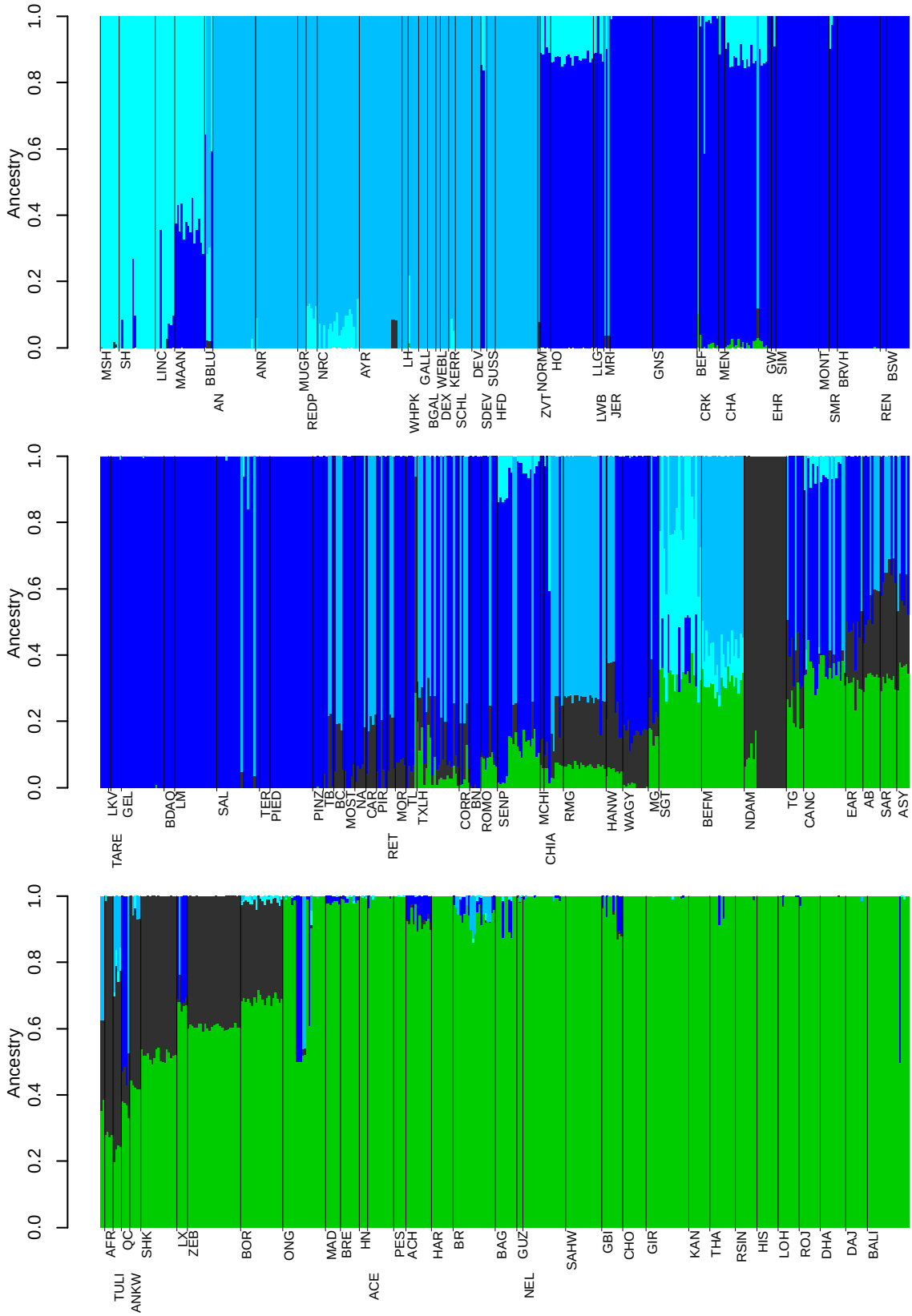

Figure 4.8. Plot of ancestry fractions for 1,143 animals with K = 5.

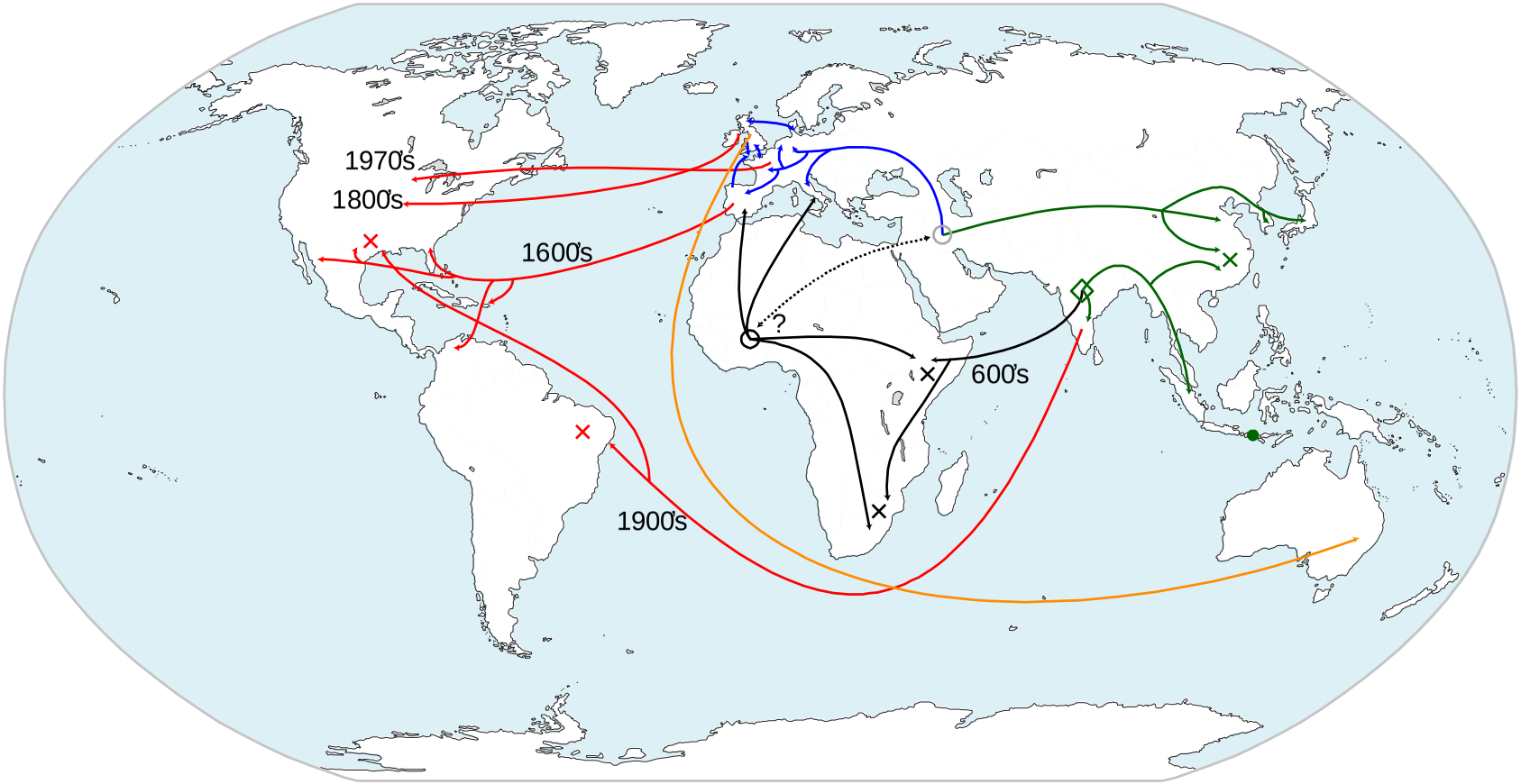

Figure 4.9. Map of worldwide exportations of domesticated cattle.
Circles represent ancient *Bos t. taurus* population centers. The diamond marks the region of *Bos t. indicus* domestication. Cross marks indicate locations of admixture between *Bos t. taurus* and *Bos t. indicaus*.

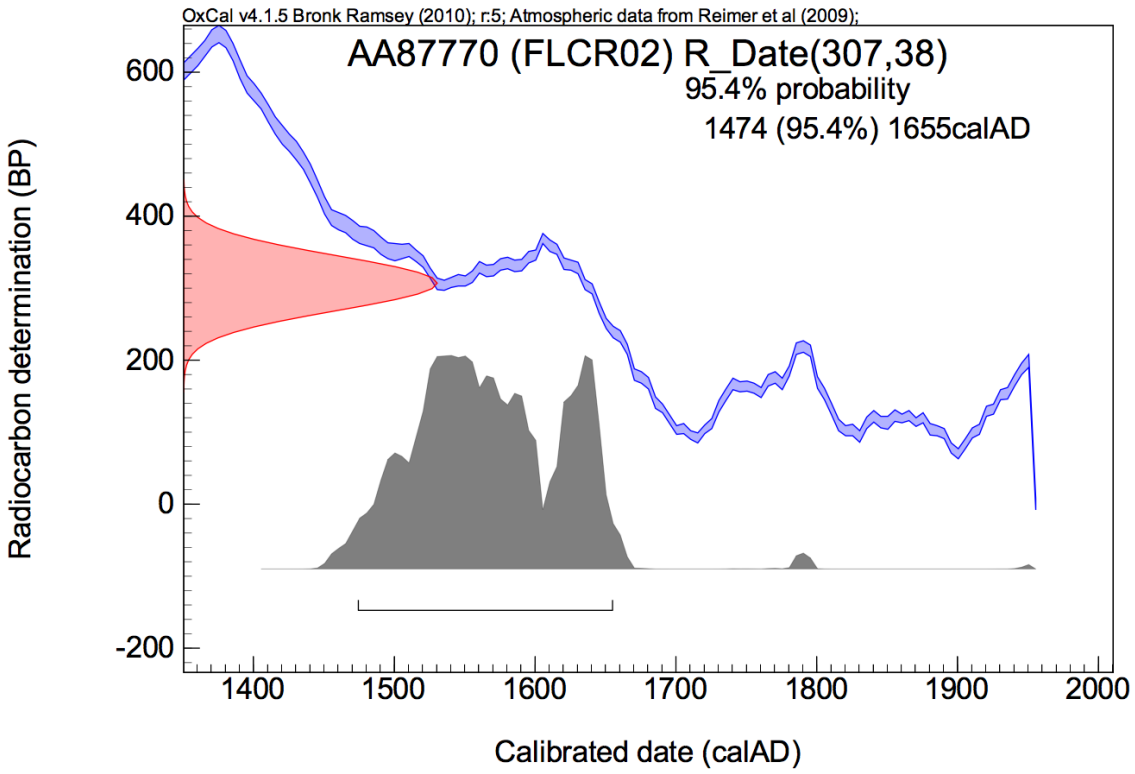

Figure 4.10. Radiocarbon date calibration for one of the ancient bovine teeth found in a well in St. Augustine, Florida.
The mean calibrated date for this sample is 1564 AD, with a 95.4% confidence interval of 1474 to 1655.

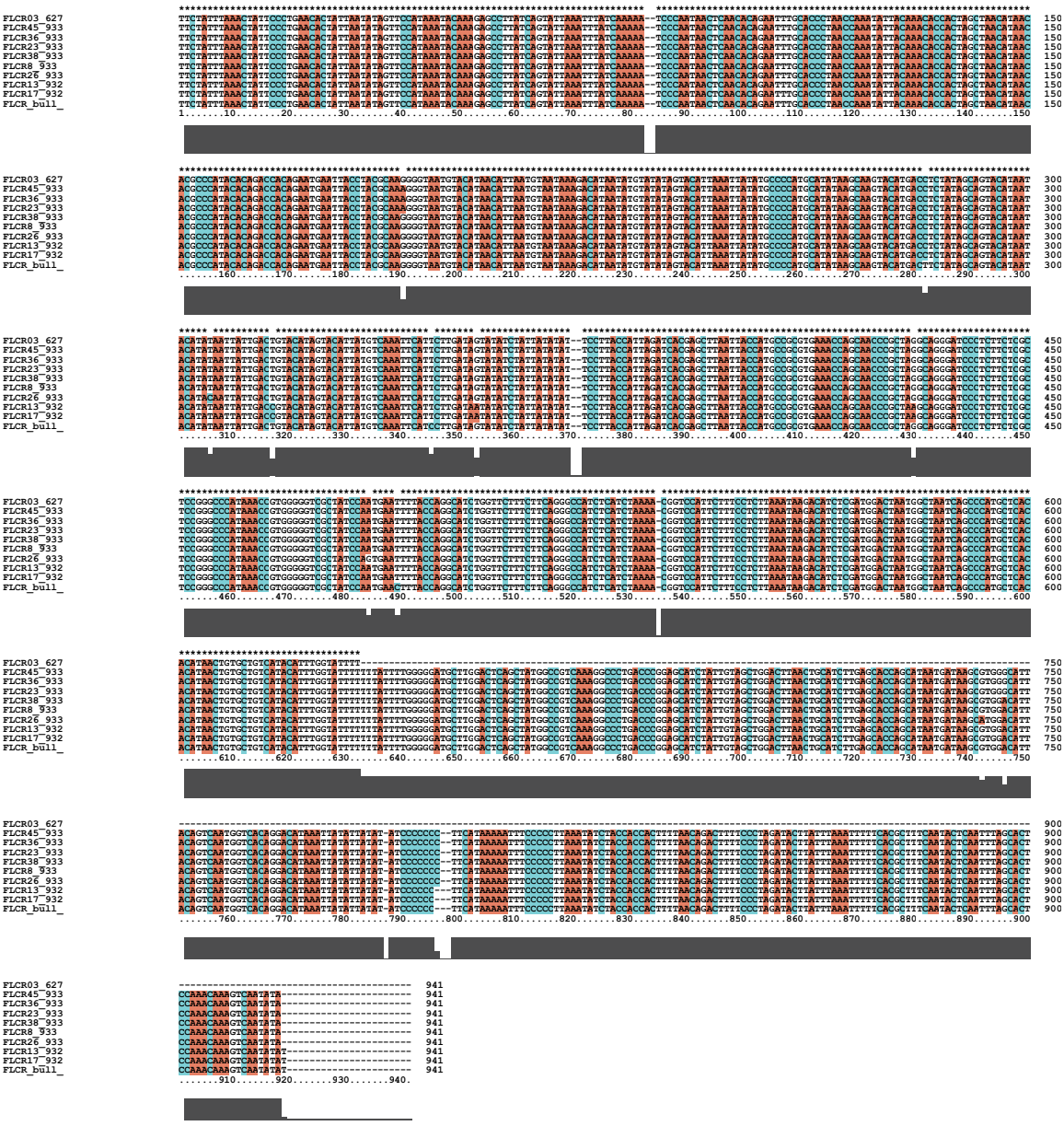

Figure 4.11. Multiple sequence alignment of mitochondrial D-loop sequences from the ancient tooth sample and 9 modern Florida Cracker samples.
Ancient tooth haplotype is denoted as FLCR03_627.

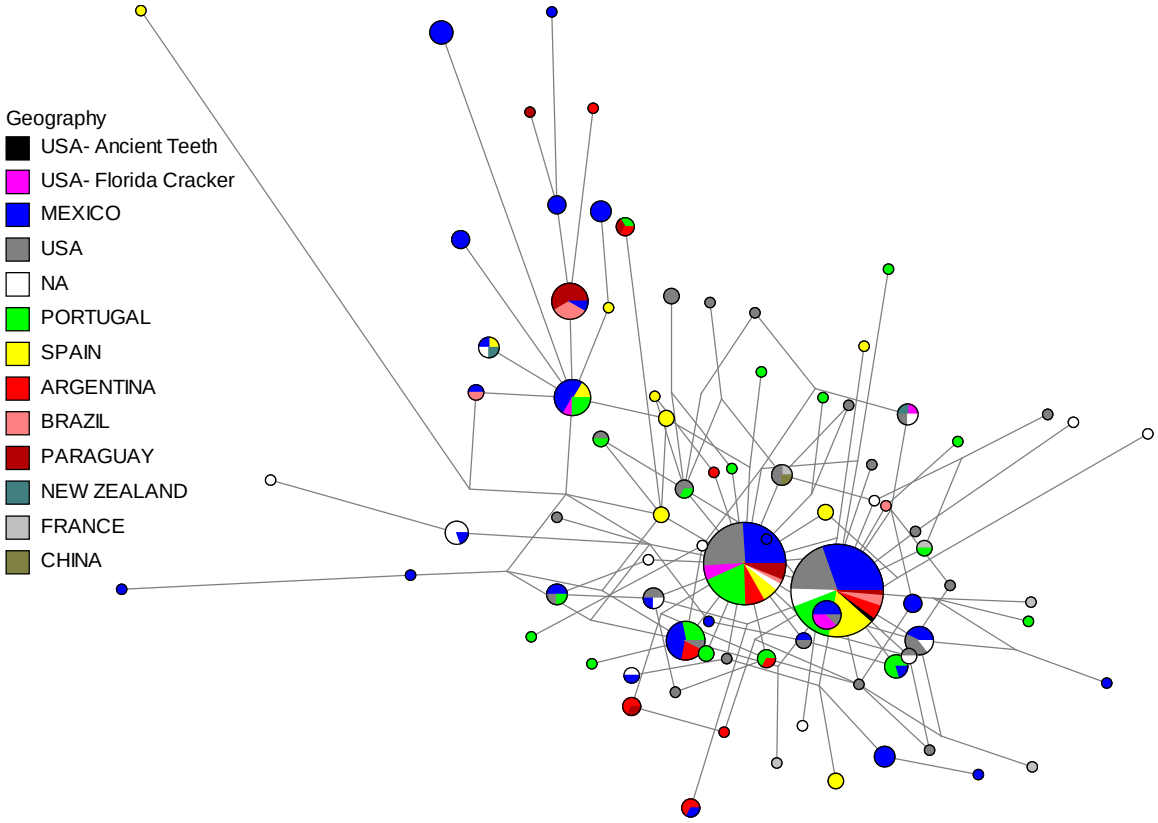

Figure 4.12. Haplotype network of 334 mitochondrial D-loop sequences with nodes colored by geographic origin of sample.

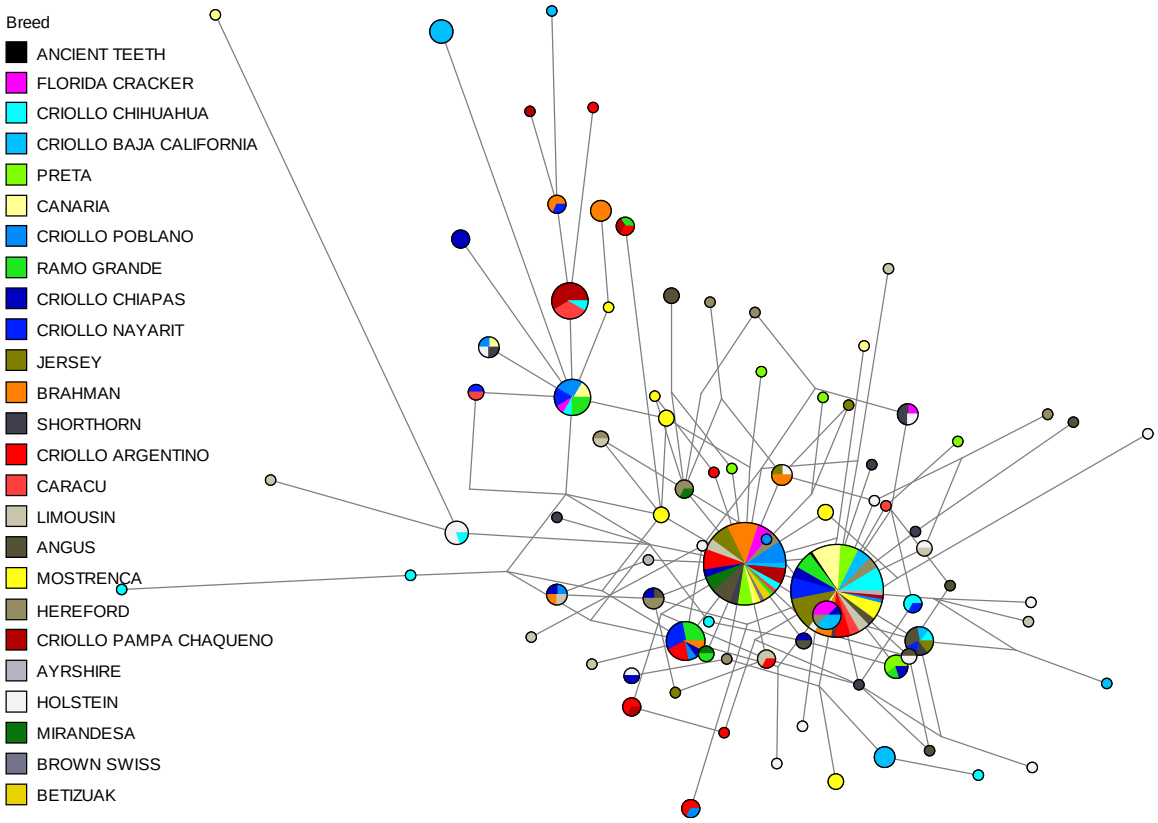

Figure 4.13. Haplotype network of 334 mitochondrial D-loop sequences with nodes colored by breed.

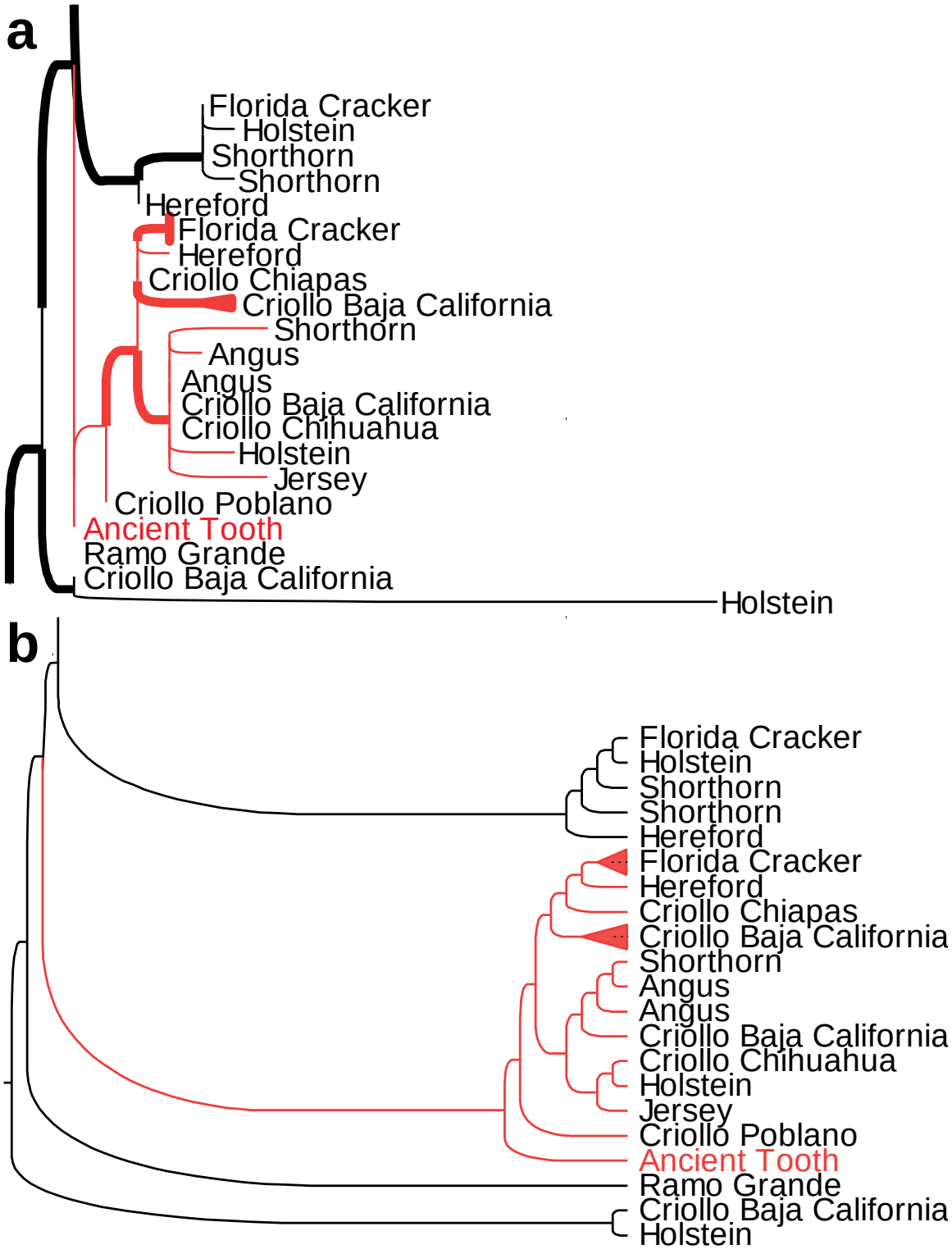

Figure 4.14. Portion of mitochondrial phylogeny generated by PhyML.
a. Phylogram with branch lengths proportional to distance. Branch widths are proportional to support. b. Cladogram where branches are not drawn to scale.

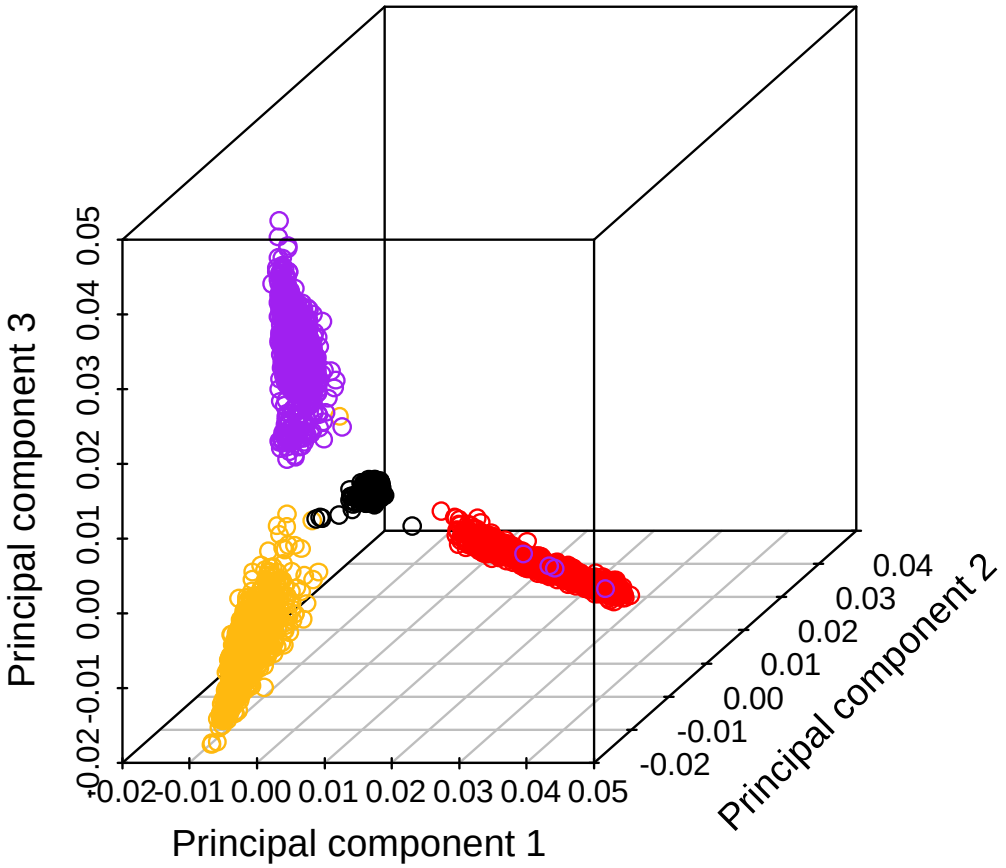

Figure 4.15. Principal component analysis of 5,169 samples used in signature of selection analysis with birth date as the dependent variable.
Points are color coded according to breed, red for Hereford, black for Holstein, yellow for Limousin, and purple for Simmental.

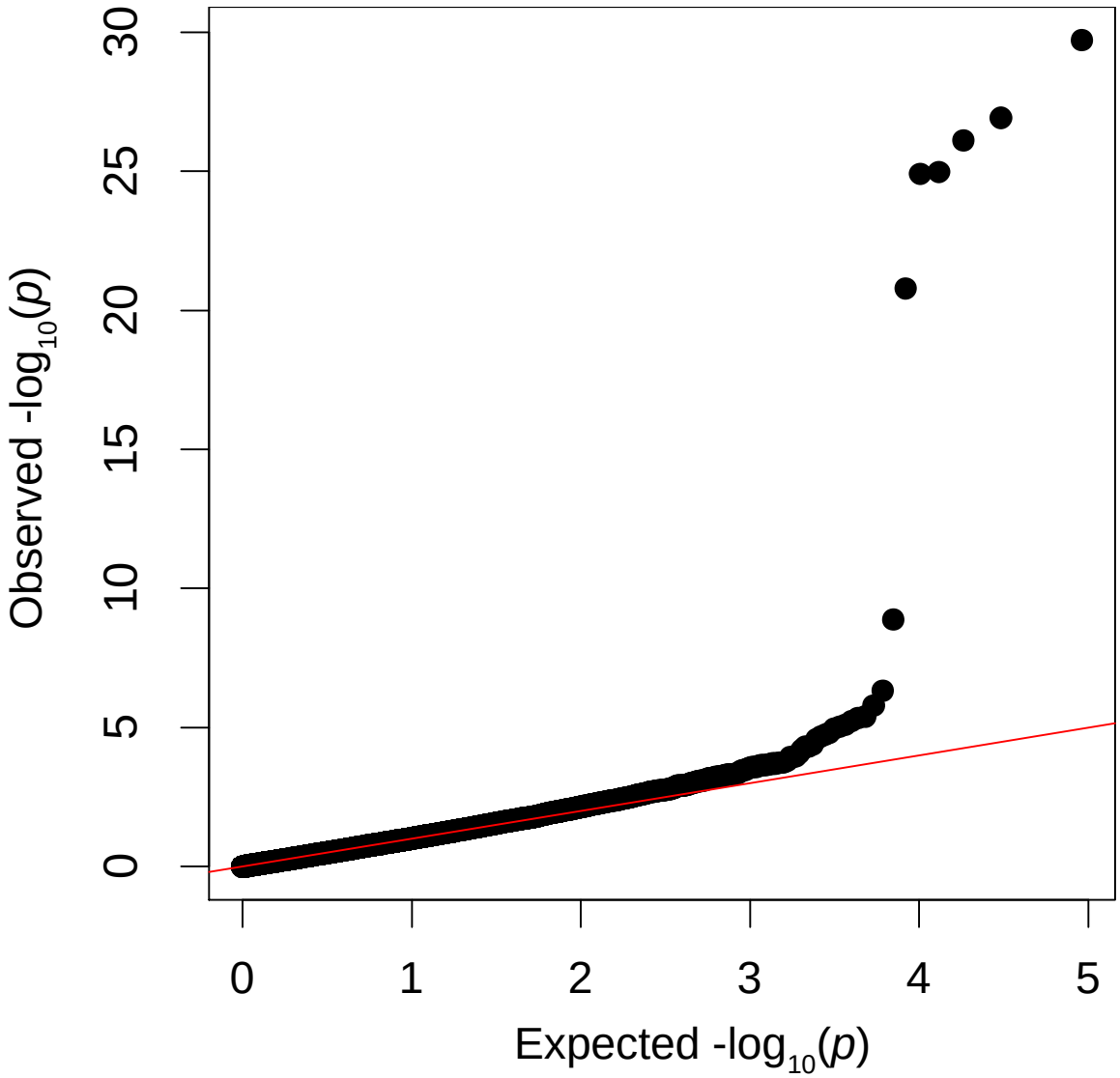

Figure 4.16. Q-Q Plot of *p*-values from analysis of birthdate for the combined data set of Hereford, Holstein, Limousin, and Simmental samples analyzed with EMMAX.

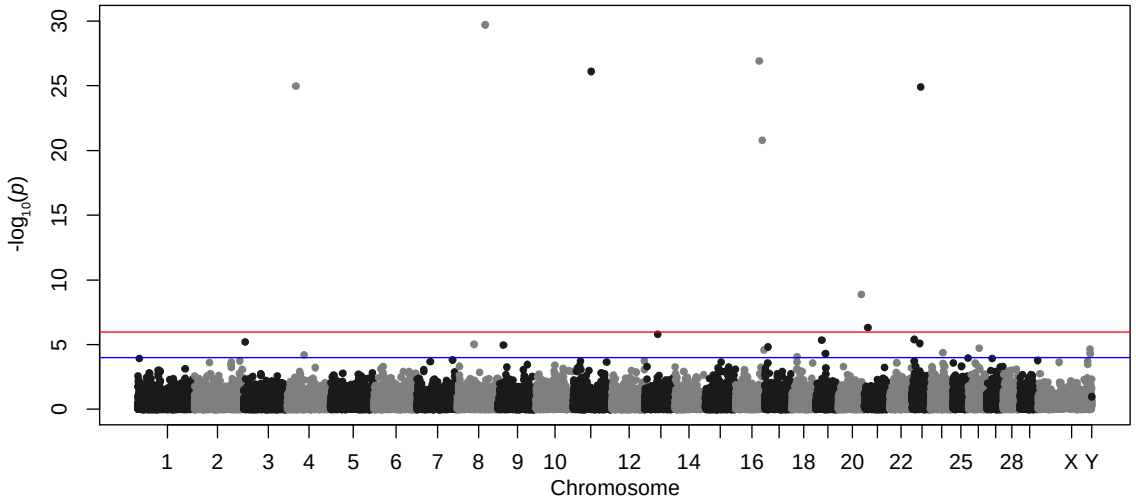

Figure 4.17. Manhattan plot of $-\log_{10}(p)$-values) from analysis of combined data set of Hereford, Holstein, Limousin, and Simmental samples using EMMAX.
Red line marks the Bonferroni corrected genome-wide significance cutoff of $-\log_{10}(p) > 5.96$, and the blue line marks the genome-wide suggestive cutoff of $-\log_{10}(p) > 4$.

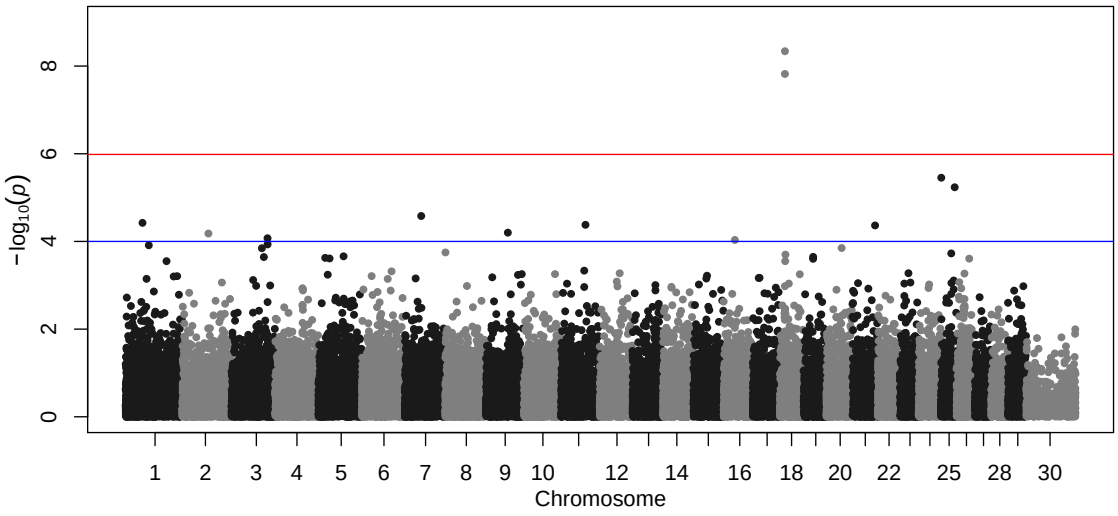

Figure 4.18. Manhattan plot of $-\log_{10}(p)$-values) from analysis of breed type.
Breeds used predominantly for beef were coded as 0, dairy as 2, and all others, including dual-purpose breeds, were coded as a 1. Red line marks the Bonferroni corrected genome-wide significance cutoff of $-\log_{10}(p) > 5.98$, and the blue line marks the genome-wide suggestive cutoff of $-\log_{10}(p) > 4$.

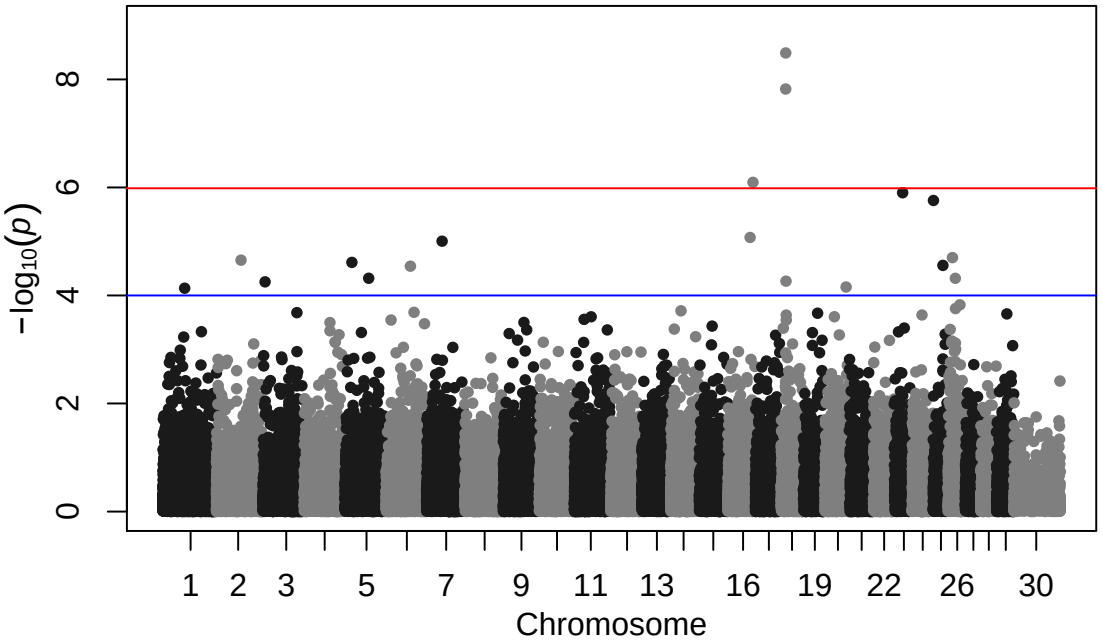

Figure 4.19. Manhattan plot of –log10(*p*-values) from analysis of breed type for animals with greater than 50% *Bos t. taurus* ancestry.
Breeds used predominantly for beef were coded as 0, dairy as 2, and all others, including dual-purpose breeds, were coded as a 1. Red line marks the Bonferroni corrected genome-wide significance cutoff of –log10(p) > 5.98, and the blue line marks the genome-wide suggestive cutoff of –log10(p) > 4.

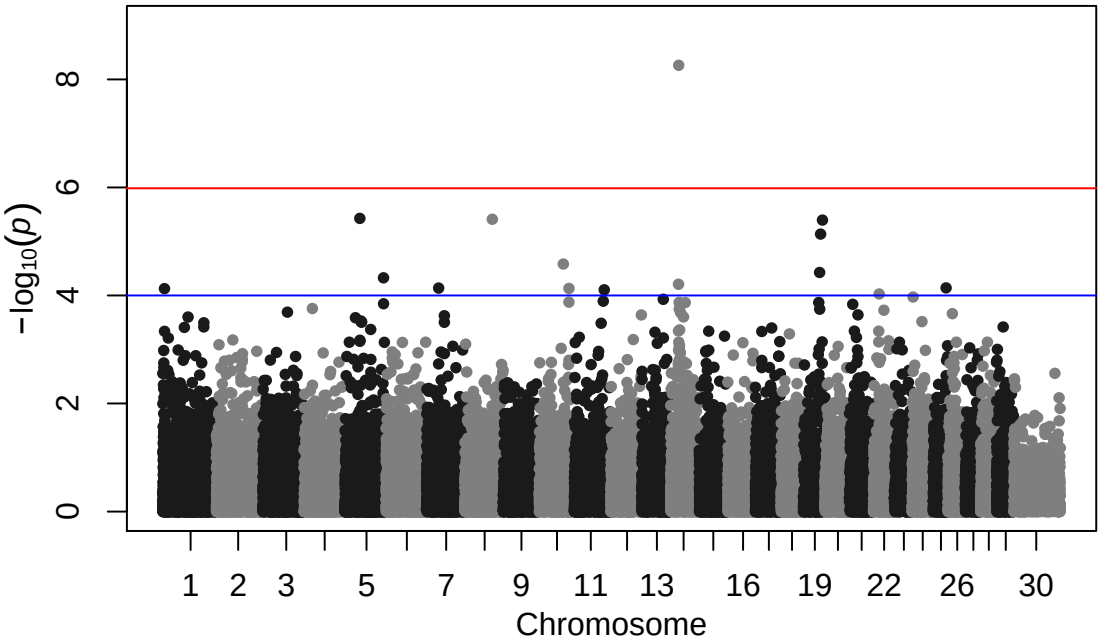

Figure 4.20. Manhattan plot of –log10(*p*-values) from analysis of breed type for animals with greater than 50% *Bos t. indicus* ancestry.
Breeds used predominantly for beef were coded as 0, dairy as 2, and all others, including dual-purpose breeds, were coded as a 1. Red line marks the Bonferroni corrected genome-wide significance cutoff of –log10(p) > 5.98, and the blue line marks the genome-wide suggestive cutoff of –log10(p) > 4.

Tables

Table 4.1. List of breeds, breed codes, numbers of samples, (sub)species, country, continent, type and type code for the 114 breeds used in the analysis.
Breed type was retrieved from (Felius 1995) and the breed type in bold is the primary use.

breed	breed code	Number of Samples	Number from Decker et al. 2009	(sub) species	country	continent	type	type code
Brahman	BR	20		*Bos indicus*	United States	Americas	beef	2
Nelore	NEL	20	5	*Bos indicus*	Brazil	Americas	beef	2
Aceh	ACE	12		*Bos indicus*	Indonesia	Asia	work/ beef	2
Achai	ACH	12		*Bos indicus*	Pakistan	Asia	NA	NA
Bhagnari	BAG	10		*Bos indicus*	Pakistan	Asia	work	1
Brebes	BRE	9		*Bos indicus*	Indonesia	Asia	NA	NA
Cholistani	CHO	11		*Bos indicus*	Pakistan	Asia	**dairy**/ beef	0
Dajal	DAJ	10		*Bos indicus*	Pakistan	Asia	dairy/ **work**	1
Dhanni	DHA	12		*Bos indicus*	Pakistan	Asia	dairy/ work/ beef	1
Gabrali	GBI	10		*Bos indicus*	Pakistan	Asia	dairy/ beef	1
Gir	GIR	20	9	*Bos indicus*	India	Asia	dairy/ work	1
Guzerat	GUZ	3	3	*Bos indicus*	India	Asia	dairy/ **beef**	2
Hariana	HAR	10		*Bos indicus*	India	Asia	dairy/ **work**	1
Hissar	HIS	10		*Bos indicus*	Pakistan	Asia	dairy/ work	1
Hainan	HN	4		*Bos indicus*	China	Asia	work/ beef	2
Kankraj	KAN	10		*Bos indicus*	India	Asia	dairy/ work	1

Lohani	LOH	10		*Bos indicus*	Pakistan	Asia	dairy/ **work**/ beef	1
Madura	MAD	7		*Bos indicus*	Indonesia	Asia	work/ racing/ fighting/ beef	2
Ongole Grade	ONG	20		*Bos indicus*	India	Asia	dairy/ work	1
Pesisir	PES	6		*Bos indicus*	Indonesia	Asia	beef	2
Rojhan	ROJ	10		*Bos indicus*	Pakistan	Asia	dairy/ **work**/ beef	1
Red Sindhi	RSIN	10		*Bos indicus*	Pakistan	Asia	dairy/ work/ beef	1
Sahiwal	SAHW	17	10	*Bos indicus*	Pakistan	Asia	dairy/ work/ beef	1
Tharparkar	THA	12		*Bos indicus*	Pakistan	Asia	dairy/ work	1
Bali	BALI	20		*Bos javanicus*	Indonesia	Asia	work/ beef	NA
N'Dama	NDAM	20	4	*Bos taurus*	Ivory Coast	Africa	dairy/ work/ beef	1
Red Angus	ANR	20	6	*Bos taurus*	United States	Americas	beef	2
Brown Swiss	BSW	15	7	*Bos taurus*	United States	Americas	**dairy**/ beef	0
Corriente	CORR	5	5	*Bos taurus*	Mexico	Americas	rodeo/ beef	2
Florida Cracker	CRK	9		*Bos taurus*	United States	Americas	beef	2
Romosinuano	ROMO	8	8	*Bos taurus*	Columbia	Americas	**dairy**/ beef	0
Senepol	SENP	20		*Bos taurus*	Virgin Islands	Americas	beef	2
Texas Longhorn	TXLH	20	8	*Bos taurus*	United States	Americas	beef	2
Hanwoo	HANW	8		*Bos taurus*	Korea	Asia	work/ beef	2

Mongolian	MG	5		*Bos taurus*	Mongolia	Asia	dairy/ work/ beef	1
Wagyu	WAGY	12	6	*Bos taurus*	Japan	Asia	fighting/ beef	2
Murray Grey	MUGR	4	4	*Bos taurus*	Australia	Australia	beef	2
Anatolian Black	AB	8		*Bos taurus*	Turkey	Europe	dairy/ work/ beef	1
Angus	AN	20	1	*Bos taurus*	Scotland	Europe	beef	2
Anatolian Southern Yellow	ASY	8		*Bos taurus*	Turkey	Europe	NA	NA
Finnish Ayrshire	AYR	20	2	*Bos taurus*	Scotland/ Finland	Europe	**dairy**/ beef	0
Belgian Blue	BBLU	4	4	*Bos taurus*	Belgium	Europe	**dairy**/ beef	0
Berrenda en Colorado	BC	5		*Bos taurus*	Spain	Europe	work/ beef	2
Blonde d'Aquitaine	BDAQ	5	5	*Bos taurus*	France	Europe	beef	2
Belted Galloway	BGAL	4	4	*Bos taurus*	Great Britian	Europe	beef	2
Berrenda en Negro	BN	5		*Bos taurus*	Spain	Europe	work/ **beef**	2
Braunvieh	BRVH	20		*Bos taurus*	Switzerland	Europe	**dairy**/ beef	0
Cardena Andaluza	CAR	5		*Bos taurus*	Spain	Europe	work/ **beef**	2
Charolais	CHA	20		*Bos taurus*	France	Europe	beef	2
Chianina	CHIA	9	7	*Bos taurus*	Italy	Europe	beef	2
Devon	DEV	4	4	*Bos taurus*	England	Europe	beef	2
Dexter	DEX	4	4	*Bos taurus*	Ireland	Europe	hobby/ dairy/ beef	1
East Anatolian Red	EAR	8		*Bos taurus*	Turkey	Europe	dairy/ beef	1
Ehringer	EHR	2		*Bos taurus*	Switzerland	Europe	cow-to-cow fighting/ dairy/ beef	0

Galloway	GALL	4	4	*Bos taurus*	Scotland	Europe	beef	2
Gelbvieh	GEL	20	5	*Bos taurus*	Germany	Europe	dairy/ **beef**	2
Guernsey	GNS	21	10	*Bos taurus*	Guernsey Island	Europe	**dairy**/ beef	0
Groningen Whitehead	GW	2		*Bos taurus*	Netherlands	Europe	**dairy**/ beef	0
Hereford	HFD	20	1	*Bos taurus*	Wales	Europe	beef	2
Holstein	HO	20		*Bos taurus*	Netherlands	Europe	dairy	0
Jersey	JER	20	7	*Bos taurus*	Jersey Island	Europe	**dairy**/ beef	0
Kerry	KERR	3	3	*Bos taurus*	Ireland	Europe	**dairy**/ beef	0
Longhorn	LH	3	3	*Bos taurus*	England	Europe	beef	2
Lincoln Red	LINC	9	9	*Bos taurus*	England	Europe	beef	2
Lakenvelder	LKV	1		*Bos taurus*	Netherlands	Europe	hobby/ dairy/ beef	1
Lithuanian Light Grey	LLG	2		*Bos taurus*	Lithuania	Europe	NA	NA
Limousin	LM	20		*Bos taurus*	France	Europe	beef	2
Lithuanian White Backed	LWB	3		*Bos taurus*	Lithuania	Europe	NA	NA
Maine-Anjou	MAAN	14	5	*Bos taurus*	France	Europe	dairy/ **beef**	2
Marchigiana	MCHI	2	2	*Bos taurus*	Italy	Europe	beef	2
Menorquina	MEN	3		*Bos taurus*	Spain	Europe	**dairy**/ beef	0
Montbeliard	MONT	5	5	*Bos taurus*	France	Europe	**dairy**/ beef	0
Morucha	MOR	5		*Bos taurus*	Spain	Europe	beef	2
Mostrenca	MOST	5		*Bos taurus*	Spain	Europe	beef	2
Meuse-Rhine-Ijjsel	MRI	3		*Bos taurus*	Netherlands	Europe	**dairy**/ beef	0
Milking Shorthorn	MSH	9	1	*Bos taurus*	England	Europe	**dairy**/ beef	0
Negra Andaluza	NA	5		*Bos taurus*	Spain	Europe	work/ **beef**	2
Normande	NORM	1		*Bos taurus*	France	Europe	**dairy**/ beef	0

Norwegian Red	NRC	20	9	*Bos taurus*	Norway	Europe	**dairy**/ beef	0
Piedmontese	PIED	20	9	*Bos taurus*	Italy	Europe	dairy/ **beef**	2
Pinzgauer	PINZ	5	5	*Bos taurus*	Austria	Europe	dairy/ beef	1
Pirenaica	PIR	5		*Bos taurus*	Spain	Europe	dairy/ **beef**	2
Red Poll	REDP	5	5	*Bos taurus*	England	Europe	dairy/ beef	1
Rendena	REN	3		*Bos taurus*	Central Alps	Europe	**dairy**/ beef	0
Retinta	RET	4		*Bos taurus*	Spain	Europe	work/ **beef**	2
Romagnola	RMG	20	10	*Bos taurus*	Italy	Europe	beef	2
Salers	SAL	20	4	*Bos taurus*	France	Europe	dairy/ **beef**	2
South Anatolian Red	SAR	8		*Bos taurus*	Turkey	Europe	dairy/ work/ beef	1
Scottish Highland	SCHL	8	8	*Bos taurus*	Scotland	Europe	beef	2
South Devon	SDEV	3	3	*Bos taurus*	England	Europe	beef	2
Beef Shorthorn	SH	17	7	*Bos taurus*	England	Europe	beef	2
Simmental	SIM	20		*Bos taurus*	Switzerland	Europe	beef	2
Simmentaler	SMR	4		*Bos taurus*	Switzerland	Europe	dairy/ beef	1
Sussex	SUSS	4	4	*Bos taurus*	England	Europe	beef	2
Tarentaise	TARE	5	5	*Bos taurus*	Central France	Europe	**dairy**/ beef	0
Toro Bravo	TB	5		Bos taurus	Spain	Europe	beef/ fighting	2
Terrana	TER	5		*Bos taurus*	Spain	Europe	work/ beef	2
Turkish Grey	TG	8		*Bos taurus*	Turkey	Europe	dairy/ work/ beef	1
Toro de Lidia	TL	5		*Bos taurus*	Spain	Europe	fighting/ beef	2
Welsh Black	WEBL	2	2	*Bos taurus*	Wales	Europe	beef	2

White Park	WHPK	5	4	*Bos taurus*	Wales	Europe	hobby/ beef	2
Zavot	ZVT	5		*Bos taurus*	Turkey	Europe	dairy/ beef	1
Africander	AFR	4		Hybrid	South Africa	Africa	beef	2
Ankole-Watusi	ANKW	5		Hybrid	Ruanda	Africa	**dairy**/ beef	0
Boran	BOR	20		Hybrid	Ethiopia	Africa	**dairy**/ beef	0
Sheko	SHK	17		Hybrid	Ethiopia	Africa	beef	2
Tuli	Tuli	4		Hybrid	Botswana	Africa	beef	2
East African Shorthorn Zebu	ZEB	25		Hybrid	Kenya	Africa	NA	NA
Beefalo	BEF	1		Hybrid	United States	Americas	beef	2
Beefmaster	BEFM	20		Hybrid	United States	Americas	beef	2
Canchim	CANC	20		Hybrid	Brazil	Americas	beef	2
Santa Gertrudis	SGT	20		Hybrid	United States	Americas	beef	2
Luxi	LX	5		Hybrid	China	Asia	work/ beef	2
Qinchuan	QC	4		Hybrid	China	Asia	work/ beef	2

Table 4.2. Statistics for values of *K* from 1 to 26 in admixture analysis.
The cross-validation for *K*=24 did not complete, and the model with *K*=26 did not converge.

K	Log-likelihood	Cross-validation error
1	-53287029.32	0.64255
2	-47327863.47	0.53243
3	-46816462.23	0.52395
4	-46593631.78	0.52117
5	-46211036.26	0.51521
6	-46058392.79	0.51352
7	-45819581.14	0.51011
8	-45379764.38	0.50254
9	-45502881.19	0.50719
10	-44951163.18	0.49719
11	-44832728.08	0.49602
12	-44606313.36	0.49331
13	-44460359.93	0.49085
14	-44417063.93	0.49214
15	-44309111.63	0.49102
16	-44130433.61	0.48980
17	-44004428.99	0.48774
18	-44008394.93	0.48946
19	-43775241.16	0.48647
20	-43696255.70	0.48584
21	-43562305.66	0.48567
22	-43530663.81	0.48566
23	-43442397.70	0.48536
24	-43300588.02	Did not complete
25	-43244973.60	0.48432
26	Did not converge	Did not converge

Table 4.3. Sample information for the 337 mitochondrial D-loop sequences.

(Sub)Species	Breed	Country	GenBank GI	Analysis Identifier	GenBank Accession #
Bos t. taurus	Ancient teeth	United States		FLCR03_627_1	
Bos t. taurus	Angus	NA	443737		L27712.1
Bos t. taurus	Angus	NA	443738		
Bos t. taurus	Angus	United States	256041669		
Bos t. taurus	Angus	United States	256041670		
Bos t. taurus	Angus	United States	256041671		
Bos t. taurus	Angus	United States	256041672		
Bos t. taurus	Angus	United States	256041673		
Bos t. taurus	Angus	United States	256041674		
Bos t. taurus	Angus	United States	256041675		
Bos t. taurus	Angus	United States	256041676		
Bos t. taurus	Angus	United States	256041677		
Bos t. taurus	Angus	United States	256041678		
Bos t. taurus	Angus	United States	256041679		
Bos t. taurus	Angus	United States	256041680		
Bos t. taurus	Angus	United States	256041681		
Bos t. taurus	Angus	United States	256041682		
Bos t. taurus	Angus	United States	256041683		
Bos t. taurus	Ayrshire	NA	2655345		AF034440.1
Bos t. taurus	Ayrshire	NA	36143095		
Bos t. taurus	Betizuak	Spain	157778271		EU177833.1
Bos t. taurus	Betizuak	Spain	157778285		EU177834.1
Bos t. taurus	Brown Swiss	NA	2655343		AF034438.1
Bos t. taurus	Canaria	Spain	256041412		
Bos t. taurus	Canaria	Spain	256041413		
Bos t. taurus	Canaria	Spain	256041414		
Bos t. taurus	Canaria	Spain	256041415		
Bos t. taurus	Canaria	Spain	256041416		
Bos t. taurus	Canaria	Spain	256041417		
Bos t. taurus	Canaria	Spain	256041418		
Bos t. taurus	Canaria	Spain	256041419		
Bos t. taurus	Canaria	Spain	256041420		
Bos t. taurus	Canaria	Spain	256041421		
Bos t. taurus	Canaria	Spain	256041422		
Bos t. taurus	Canaria	Spain	256041423		
Bos t. taurus	Canaria	Spain	256041424		

Bos t. taurus	Canaria	Spain	256041425
Bos t. taurus	Caracu	Brazil	256041464
Bos t. taurus	Caracu	Brazil	256041465
Bos t. taurus	Caracu	Brazil	256041466
Bos t. taurus	Caracu	Brazil	256041467
Bos t. taurus	Caracu	Brazil	256041468
Bos t. taurus	Caracu	Brazil	256041469
Bos t. taurus	Caracu	Brazil	256041470
Bos t. taurus	Caracu	Brazil	256041471
Bos t. taurus	Caracu	Brazil	256041472
Bos t. taurus	Caracu	Brazil	256041473
Bos t. taurus	Criollo Argentino	Argentina	256041443
Bos t. taurus	Criollo Argentino	Argentina	256041444
Bos t. taurus	Criollo Argentino	Argentina	256041445
Bos t. taurus	Criollo Argentino	Argentina	256041446
Bos t. taurus	Criollo Argentino	Argentina	256041447
Bos t. taurus	Criollo Argentino	Argentina	256041448
Bos t. taurus	Criollo Argentino	Argentina	256041449
Bos t. taurus	Criollo Argentino	Argentina	256041450
Bos t. taurus	Criollo Argentino	Argentina	256041451
Bos t. taurus	Criollo Argentino	Argentina	256041452
Bos t. taurus	Criollo Argentino	Argentina	256041453
Bos t. taurus	Criollo Argentino	Argentina	256041454
Bos t. taurus	Criollo Argentino	Argentina	256041455
Bos t. taurus	Criollo Argentino	Argentina	256041456
Bos t. taurus	Criollo Argentino	Argentina	256041457
Bos t. taurus	Criollo Argentino	Argentina	256041458
Bos t. taurus	Criollo Argentino	Argentina	256041459

Bos t. taurus	Criollo Argentino	Argentina	256041460
Bos t. taurus	Criollo Argentino	Argentina	256041461
Bos t. taurus	Criollo Argentino	Argentina	256041462
Bos t. taurus	Criollo Argentino	Argentina	256041463
Bos t. taurus	Criollo Baja California	Mexico	256041474
Bos t. taurus	Criollo Baja California	Mexico	256041475
Bos t. taurus	Criollo Baja California	Mexico	256041476
Bos t. taurus	Criollo Baja California	Mexico	256041477
Bos t. taurus	Criollo Baja California	Mexico	256041478
Bos t. taurus	Criollo Baja California	Mexico	256041479
Bos t. taurus	Criollo Baja California	Mexico	256041480
Bos t. taurus	Criollo Baja California	Mexico	256041481
Bos t. taurus	Criollo Baja California	Mexico	256041482
Bos t. taurus	Criollo Baja California	Mexico	256041483
Bos t. taurus	Criollo Baja California	Mexico	256041484
Bos t. taurus	Criollo Baja California	Mexico	256041485
Bos t. taurus	Criollo Baja California	Mexico	256041486
Bos t. taurus	Criollo Baja California	Mexico	256041487
Bos t. taurus	Criollo Baja California	Mexico	256041488
Bos t. taurus	Criollo Baja California	Mexico	256041489
Bos t. taurus	Criollo Baja California	Mexico	256041490
Bos t. taurus	Criollo Baja California	Mexico	256041491
Bos t. taurus	Criollo Baja California	Mexico	256041492

Bos t. taurus	Criollo Baja California	Mexico	256041493
Bos t. taurus	Criollo Chiapas	Mexico	256041513
Bos t. taurus	Criollo Chiapas	Mexico	256041514
Bos t. taurus	Criollo Chiapas	Mexico	256041515
Bos t. taurus	Criollo Chiapas	Mexico	256041516
Bos t. taurus	Criollo Chiapas	Mexico	256041517
Bos t. taurus	Criollo Chiapas	Mexico	256041518
Bos t. taurus	Criollo Chiapas	Mexico	256041519
Bos t. taurus	Criollo Chiapas	Mexico	256041520
Bos t. taurus	Criollo Chiapas	Mexico	256041521
Bos t. taurus	Criollo Chiapas	Mexico	256041522
Bos t. taurus	Criollo Chiapas	Mexico	256041523
Bos t. taurus	Criollo Chiapas	Mexico	256041524
Bos t. taurus	Criollo Chiapas	Mexico	256041525
Bos t. taurus	Criollo Chiapas	Mexico	256041526
Bos t. taurus	Criollo Chiapas	Mexico	256041527
Bos t. taurus	Criollo Chihuahua	Mexico	256041494
Bos t. taurus	Criollo Chihuahua	Mexico	256041495
Bos t. taurus	Criollo Chihuahua	Mexico	256041496
Bos t. taurus	Criollo Chihuahua	Mexico	256041497
Bos t. taurus	Criollo Chihuahua	Mexico	256041498
Bos t. taurus	Criollo Chihuahua	Mexico	256041499
Bos t. taurus	Criollo Chihuahua	Mexico	256041500

	Criollo		
Bos t. taurus	Chihuahua Criollo	Mexico	256041501
Bos t. taurus	Chihuahua Criollo	Mexico	256041502
Bos t. taurus	Chihuahua Criollo	Mexico	256041503
Bos t. taurus	Chihuahua Criollo	Mexico	256041504
Bos t. taurus	Chihuahua Criollo	Mexico	256041505
Bos t. taurus	Chihuahua Criollo	Mexico	256041506
Bos t. taurus	Chihuahua Criollo	Mexico	256041507
Bos t. taurus	Chihuahua Criollo	Mexico	256041508
Bos t. taurus	Chihuahua Criollo	Mexico	256041509
Bos t. taurus	Chihuahua Criollo	Mexico	256041510
Bos t. taurus	Chihuahua Criollo	Mexico	256041511
Bos t. taurus	Chihuahua Criollo	Mexico	256041512
Bos t. taurus	Nayarit Criollo	Mexico	256041528
Bos t. taurus	Nayarit Criollo	Mexico	256041529
Bos t. taurus	Nayarit Criollo	Mexico	256041530
Bos t. taurus	Nayarit Criollo	Mexico	256041531
Bos t. taurus	Nayarit Criollo	Mexico	256041532
Bos t. taurus	Nayarit Criollo	Mexico	256041533
Bos t. taurus	Nayarit Criollo	Mexico	256041534
Bos t. taurus	Nayarit Criollo	Mexico	256041535
Bos t. taurus	Nayarit Criollo	Mexico	256041536
Bos t. taurus	Nayarit Criollo	Mexico	256041537
Bos t. taurus	Nayarit	Mexico	256041538

	Criollo		
Bos t. taurus	Nayarit	Mexico	256041539
	Criollo		
Bos t. taurus	Nayarit	Mexico	256041540
	Criollo		
Bos t. taurus	Nayarit	Mexico	256041541
	Criollo		
Bos t. taurus	Nayarit	Mexico	256041542
	Criollo		
Bos t. taurus	Nayarit	Mexico	256041543
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041560
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041561
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041562
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041563
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041565
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041566
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041567
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041568
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041569
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041570
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041571
	Criollo		
	Pampa		
Bos t. taurus	Chaqueno	Paraguay	256041572

	Criollo			
	Pampa			
Bos t. taurus	Chaqueno	Paraguay	256041573	
	Criollo			
	Pampa			
Bos t. taurus	Chaqueno	Paraguay	256041574	
	Criollo			
	Pampa			
Bos t. taurus	Chaqueno	Paraguay	256041575	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041544	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041545	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041546	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041547	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041549	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041550	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041551	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041552	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041553	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041554	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041555	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041556	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041557	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041558	
	Criollo			
Bos t. taurus	Poblano	Mexico	256041559	
	Florida			
Bos t. taurus	Cracker	United States		FLCR_bull__1
	Florida			
Bos t. taurus	Cracker	United States		FLCR13_932_1
	Florida			
Bos t. taurus	Cracker	United States		FLCR17_932_1
Bos t. taurus	Florida	United States		FLCR23_933_1

	Cracker Florida			
Bos t. taurus	Cracker Florida	United States		FLCR26_933_1
Bos t. taurus	Cracker Florida	United States		FLCR36_933_1
Bos t. taurus	Cracker Florida	United States		FLCR38_933_1
Bos t. taurus	Cracker Florida	United States		FLCR45_933_1
Bos t. taurus	Cracker	United States		FLCR8_933_1
Bos t. taurus	Hereford	NA	443747	
Bos t. taurus	Hereford	NA	443748	
Bos t. taurus	Hereford	United States	256041696	
Bos t. taurus	Hereford	United States	256041697	
Bos t. taurus	Hereford	United States	256041698	
Bos t. taurus	Hereford	United States	256041699	
Bos t. taurus	Hereford	United States	256041700	
Bos t. taurus	Hereford	United States	256041701	
Bos t. taurus	Hereford	United States	256041702	
Bos t. taurus	Hereford	United States	256041703	
Bos t. taurus	Hereford	United States	256041704	
Bos t. taurus	Hereford	United States	256041705	
Bos t. taurus	Hereford	United States	256041706	
Bos t. taurus	Hereford	United States	256041707	
Bos t. taurus	Hereford	United States	256041708	
Bos t. taurus	Hereford	United States	256041709	
Bos t. taurus	Holstein	France	45643620	
Bos t. taurus	Holstein	France	45643622	
Bos t. taurus	Holstein	France	45643623	
Bos t. taurus	Holstein	France	45643624	
Bos t. taurus	Holstein	France	45643625	
Bos t. taurus	Holstein	NA	23429502	
Bos t. taurus	Holstein	NA	23452308	
Bos t. taurus	Holstein	NA	23452309	
Bos t. taurus	Holstein	NA	256041426	
Bos t. taurus	Holstein	NA	33321712	
Bos t. taurus	Holstein	NA	33321713	
Bos t. taurus	Holstein	NA	33321714	
Bos t. taurus	Holstein	NA	46404117	
Bos t. taurus	Holstein	NA	62363164	
Bos t. taurus	Holstein	NA	62363165	
Bos t. taurus	Holstein	NA	62363169	

Bos t. taurus	Holstein	NA	62363170	
Bos t. taurus	Holstein	NA	85375977	
Bos t. taurus	Holstein	NA	85375978	
Bos t. taurus	Jersey	United States	256041731	
Bos t. taurus	Jersey	United States	256041732	
Bos t. taurus	Jersey	United States	256041733	
Bos t. taurus	Jersey	United States	256041734	
Bos t. taurus	Jersey	United States	256041735	
Bos t. taurus	Jersey	United States	256041736	
Bos t. taurus	Jersey	United States	256041737	
Bos t. taurus	Jersey	United States	256041738	
Bos t. taurus	Jersey	United States	256041739	
Bos t. taurus	Jersey	United States	256041740	
Bos t. taurus	Jersey	United States	256041741	
Bos t. taurus	Jersey	United States	256041742	
Bos t. taurus	Jersey	United States	256041743	
Bos t. taurus	Jersey	United States	256041744	
Bos t. taurus	Jersey	United States	256041745	
Bos t. taurus	Jersey	United States	256041746	
Bos t. taurus	Jersey	United States	256041747	
Bos t. taurus	Jersey	United States	256041748	
Bos t. taurus	Limousin	NA	56410908	AY676856.1
Bos t. taurus	Limousin	Portugal	256041624	
Bos t. taurus	Limousin	Portugal	256041625	
Bos t. taurus	Limousin	Portugal	256041626	
Bos t. taurus	Limousin	Portugal	256041627	
Bos t. taurus	Limousin	Portugal	256041628	
Bos t. taurus	Limousin	Portugal	256041629	
Bos t. taurus	Limousin	Portugal	256041630	
Bos t. taurus	Limousin	Portugal	256041631	
Bos t. taurus	Limousin	Portugal	256041632	
Bos t. taurus	Limousin	Portugal	256041633	
Bos t. taurus	Limousin	Portugal	256041634	
Bos t. taurus	Limousin	Portugal	256041635	
Bos t. taurus	Limousin	Portugal	256041637	
Bos t. taurus	Limousin	Portugal	256041638	
Bos t. taurus	Limousin	Portugal	256041639	
Bos t. taurus	Mirandesa	Portugal	256041358	
Bos t. taurus	Mirandesa	Portugal	256041359	
Bos t. taurus	Mirandesa	Portugal	256041360	
Bos t. taurus	Mirandesa	Portugal	256041361	
Bos t. taurus	Mirandesa	Portugal	256041362	

Bos t. taurus	Mirandesa	Portugal	256041363
Bos t. taurus	Mostrenca	Spain	256041396
Bos t. taurus	Mostrenca	Spain	256041397
Bos t. taurus	Mostrenca	Spain	256041398
Bos t. taurus	Mostrenca	Spain	256041399
Bos t. taurus	Mostrenca	Spain	256041400
Bos t. taurus	Mostrenca	Spain	256041401
Bos t. taurus	Mostrenca	Spain	256041402
Bos t. taurus	Mostrenca	Spain	256041403
Bos t. taurus	Mostrenca	Spain	256041404
Bos t. taurus	Mostrenca	Spain	256041405
Bos t. taurus	Mostrenca	Spain	256041406
Bos t. taurus	Mostrenca	Spain	256041407
Bos t. taurus	Mostrenca	Spain	256041408
Bos t. taurus	Mostrenca	Spain	256041409
Bos t. taurus	Mostrenca	Spain	256041410
Bos t. taurus	Mostrenca	Spain	256041411
Bos t. taurus	Preta	Portugal	256041364
Bos t. taurus	Preta	Portugal	256041365
Bos t. taurus	Preta	Portugal	256041366
Bos t. taurus	Preta	Portugal	256041367
Bos t. taurus	Preta	Portugal	256041368
Bos t. taurus	Preta	Portugal	256041369
Bos t. taurus	Preta	Portugal	256041370
Bos t. taurus	Preta	Portugal	256041371
Bos t. taurus	Preta	Portugal	256041372
Bos t. taurus	Preta	Portugal	256041373
Bos t. taurus	Preta	Portugal	256041374
Bos t. taurus	Preta	Portugal	256041375
Bos t. taurus	Preta	Portugal	256041376
Bos t. taurus	Preta	Portugal	256041377
Bos t. taurus	Preta	Portugal	256041378
Bos t. taurus	Preta	Portugal	256041379
Bos t. taurus	Ramo Grande	Portugal	256041380
Bos t. taurus	Ramo Grande	Portugal	256041381
Bos t. taurus	Ramo Grande	Portugal	256041382
Bos t. taurus	Ramo Grande	Portugal	256041383
Bos t. taurus	Ramo Grande	Portugal	256041384

Bos t. taurus	Ramo Grande	Portugal	256041385
Bos t. taurus	Ramo Grande	Portugal	256041386
Bos t. taurus	Ramo Grande	Portugal	256041387
Bos t. taurus	Ramo Grande	Portugal	256041388
Bos t. taurus	Ramo Grande	Portugal	256041389
Bos t. taurus	Ramo Grande	Portugal	256041390
Bos t. taurus	Ramo Grande	Portugal	256041391
Bos t. taurus	Ramo Grande	Portugal	256041392
Bos t. taurus	Ramo Grande	Portugal	256041393
Bos t. taurus	Ramo Grande	Portugal	256041394
Bos t. taurus	Ramo Grande	Portugal	256041395
Bos t. taurus	Shorthorn	New Zealand	20372310
Bos t. taurus	Shorthorn	New Zealand	20372314
Bos t. taurus	Shorthorn	United States	256041749
Bos t. taurus	Shorthorn	United States	256041750
Bos t. taurus	Shorthorn	United States	256041751
Bos t. taurus	Shorthorn	United States	256041752
Bos t. taurus	Shorthorn	United States	256041753
Bos t. taurus	Shorthorn	United States	256041754
Bos t. taurus	Shorthorn	United States	256041755
Bos t. taurus	Shorthorn	United States	256041756
Bos t. taurus	Shorthorn	United States	256041757
Bos t. indicus	Brahman	China	114205636
Bos t. indicus	Brahman	Mexico	256041645
Bos t. indicus	Brahman	Mexico	256041646
Bos t. indicus	Brahman	Mexico	256041647
Bos t. indicus	Brahman	Mexico	256041648
Bos t. indicus	Brahman	Mexico	256041649
Bos t. indicus	Brahman	Mexico	256041650
Bos t. indicus	Brahman	Mexico	256041651
Bos t. indicus	Brahman	Mexico	256041652
Bos t. indicus	Brahman	Mexico	256041653
Bos t. indicus	Brahman	Mexico	256041654

Bos t. indicus	Brahman	Mexico	256041655	
Bos t. indicus	Brahman	Mexico	256041656	
Bos t. indicus	Brahman	Mexico	256041657	
Bos t. indicus	Brahman	Mexico	256041658	
Bos t. indicus	Brahman	Mexico	256041659	
Bos t. indicus	Brahman	NA	27462350	
Bos t. indicus	Brahman	NA	27462351	
Bos t. indicus	Brahman	United States	256041640	
Bos t. indicus	Brahman	United States	256041641	
Bos t. indicus	Brahman	United States	256041642	
Bos t. indicus	Brahman	United States	256041643	
Bos t. indicus	Brahman	United States	256041644	
Bubalus bubalis			126742614	EF464392.1
Bubalus bubalis			126742662	EF464440.1
Bubalus bubalis			126742653	EF464431.1

Table 4.4. Coordinates of significant associations from the genome-wide analysis of birth date for the combined data set of Herefords, Holsteins, Limousins, and Simmentals.

Genes within 100 Kbp of the significant SNP were identified and biological processes and pathways were acquired from NCBI Gene information for cow and human.

Chromosome	UMD 3.1 Position	p-value	Candidate gene symbol	Gene Ontology Biological Process	Pathways
4	24,004,871	1.04E-25	*MEOX2*	angiogenesis; blood circulation; limb development; multicellular organismal development; palate development; skeletal muscle tissue development; somite specification	
8	77,546,886	1.92E-30	*RASEF*	protein transport; small GTPase mediated signal transduction	
11	49,473,033	7.70E-27	*ELMOD3*	phagocytosis	
16	65,669,824	1.20E-27	*LAMC1*, *LAMC2*		Amoebiasis; ECM-receptor interaction; Focal adhesion; Prion diseases; Small cell lung cancer; Toxoplasmosis; Inflammatory Response Pathway
16	74,158,269	1.61E-21	*KCNH1*	ion transport; potassium ion transport; regulation of transcription, DNA-dependent; transmembrane transport	
20	63,865,337	1.30E-09	*SNORD123*, *SEMA5A*, *AGRP2*	appetite stimulation?	
21	10,087,575	4.72E-07			
23	24,667,121	1.21E-25	*TRAM2*, *EFHC1*	protein transport; collagen biosynthetic process	

Table 4.5. Primers used in aDNA amplification and sequencing.

	Primer	Primer Sequence (5' --- 3')	Length[(c)]
Set_A1	BovCR-16351F	CAACCCCCAAAGCTGAAG	~96bp
	BovCR-16457R	TGGTTRGGGTACAAAGTCTGTG	
Set_B1	BovCR-16420F	CCATAAATGCAAAGAGCCTCAYCAG	~172bp
	BovCR-16642R	TGCATGGGGCATATAATTTAATGTA	
Set_A2	BovCR-16507F	AATGCATTACCCAAACRGGG	~184bp
	BovCR-16755R	ATTAAGCTCGTGATCTARTGG	
Set-B2	BovCR-16633F[(a)]	GCCCCATGCATATAAGCAAG	~132bp
	BovCR-16810R[(a)]	GCCTAGCGGGTTGCTGGTTTCACGC	
Set_A3	BovCR-16765F[(a)]	GAGCTTAAYTACCATGCCG	~125bp
	BovCR-16998R	CGAGATGTCTTATTTAAGAGGAAAGAATGG	
Set_B3	BovCR-16960F	CATCTGGTTCTTTCTTCAGGGCC	~110bp
	BovCR-80R[(a)]	CAAGCATCCCCCAAAATAAA	

Two pairs of PCR primers derived from hypervariable control region and 12S-rRNA region of the mitochondrial genome were used for one-step simplex PCRs.

	Primer	Primer Sequence (5'--- 3')	Length[(c)]
Frag1	BovCR_16738MF[(b)]	***CACGACGTTGTAAAACGAC***ATYGTACATAGYACATTATGTCAA	67bp
	BovCR_16810TR[(b)]	***TACGACTCACTATAGGGCGA***GCCTAGCGGGTTGCTGGTTTCACGC	
Frag2	Mamm_12SE	CTATAATCGATAAACCCCGATA	96bp
	Mamm_12SH	GCTACACCTTGACCTAAC	

(a): Primers (BovCR-16633F, BovCR-16810R, BovCR-16765F, BovCR-80R) were published in Shapiro et al., 2004.

(b): To obtain good quality sequences for short fragment from directly sequencing, M13 (CAC GAC GTT GTA AAA CGA C) and T7 (TAC GAC TCA CTA TAG GGC GA) primer sequences were tagged at the primers BovCR_16738F and BovCR_16810R, respectively.

(c): Length of PCR amplicon is primer-excluded.

Table 4.6. Primers for sequencing mitochondrial control region from modern DNA.

Primer	Primer Location	Primer Sequences (5'--- 3')
A149	Bov_CR_113R	GTCCAGCTACAATAGATGCTC
A1125*	Bov_CR_L15737F	CTGCAGTCTCACCATCAACC
A1127	Bov_CR_L16161F	AATTACCATGCCGCGTGA
A1208*	Bov_CR_498R	ACTGGGGTGTAGATGCTTGC
A1453	Bos_CR_298R	GCTAAATTGAGTATTGAAAGYGTG

*Primers A1125 and A1208 were used for PCR amplification of mitochondrial control region (with an amplicon of 1.1 Kbp) from extant cracker cattle DNA. Primers A149, A1125, A1127, and A1453 were used to sequence the control region from modern cracker cattle DNA.

BIBLIOGRAPHY

Ajmone-Marsan, Paolo, Jose Fernando Garcia, Johannes A. Lenstra, and The Globaldiv Consortium. 2010. "On the Origin of Cattle : How Aurochs Became Cattle and Colonized the World." *Evolutionary Anthropology* 19: 148-157. doi:10.1002/evan20267.

Akey, Joshua M. 2009. "Constructing genomic maps of positive selection in humans: where do we go from here?" *Genome research* 19 (5): 711-722. doi:10.1101/gr.086652.108. http://www.ncbi.nlm.nih.gov/pubmed/19411596.

Akey, Joshua M., Ge Zhang, Kun Zhang, Li Jin, and Mark D. Shriver. 2002. "Interrogating a high-density SNP map for signatures of natural selection." *Genome research* 12 (12): 1805-1814. doi:10.1101/gr.631202. http://www.ncbi.nlm.nih.gov/pubmed/12466284.

Alexander, David H., John Novembre, and Kenneth Lange. 2009. "Fast model-based estimation of ancestry in unrelated individuals." *Genome research* 19 (9): 1655-1664. doi:10.1101/gr.094052.109. http://www.ncbi.nlm.nih.gov/pubmed/19648217.

Anon. 2011. European Bioinformatics Institute QuickGO Database. http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0060021#term=stats.

Anon. 2012. Angus - FAQs. *American Angus Association*. http://www.angus.org/Pub/FAQs.aspx.

Anon. Wye Angus History. http://wyeangus.umd.edu/History.cfm.

Anon. European Bioinformatics Institute QuickGO Database. http://www.ebi.ac.uk/QuickGO/GTerm?id=GO:0016605.

Anon. Angus History. http://www.angus.org/pub/Anghist.aspx.

Aulchenko, Yurii S., Stephan Ripke, Aaron Isaacs, and Cornelia M. van Duijn. 2007. "GenABEL: an R library for genome-wide association analysis." *Bioinformatics* 23 (10): 1294-1296. doi:10.1093/bioinformatics/btm108. http://www.ncbi.nlm.nih.gov/pubmed/17384015.

Ayroles, Julien F., Kimberly A. Hughes, Kevin C. Rowe, Melissa M. Reedy, Sandra L. Rodriguez-Zas, Jenny M. Drnevich, Carla E. Cáceres, Ken N. Paige, Carla E. Caceres,

and Ken N. Paige. 2009. “A genomewide assessment of inbreeding depression: gene number, function, and mode of action.” *Conservation Biology* 23 (4): 920-930. doi:10.1111/j.1523-1739.2009.01186.x. http://www.ncbi.nlm.nih.gov/pubmed/19627320.

Bacanu, Silviu-Alin, Bernie Devlin, and Kathryn Roeder. 2000. “The power of genomic control.” *The American Journal of Human Genetics* 66 (6): 1933-1944. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1378064&tool=pmcentrez&rendertype=abstract.

Baker, Marvel L., Cecil T. Blunn, and Mogens Plum. 1951. “‘DWARFISM’ IN ABERDEEN-ANGUS CATTLE.” *The Journal of heredity* 42 (3): 141-144.

Balding, David J., and Richard A. Nichols. 1995. “A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity.” *Genetica* 96 (1-2): 3-12. http://www.ncbi.nlm.nih.gov/pubmed/7607457.

Bandelt, Hans-Jurgen, Peter Forster, and Arne Röhl. 1999. “Median-joining networks for inferring intraspecific phylogenies.” *Molecular biology and evolution* 16 (1) (January): 37-48. http://www.ncbi.nlm.nih.gov/pubmed/10331250.

Baurain, Denis, Henner Brinkmann, and Hervé Philippe. 2007. “Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors?” *Molecular biology and evolution* 24 (1) (January): 6-9. doi:10.1093/molbev/msl137. http://www.ncbi.nlm.nih.gov/pubmed/17012374.

Beja-Pereira, Albano, David Caramelli, Carles Lalueza-Fox, Cristiano Vernesi, Nuno Ferrand, Antonella Casoli, Felix Goyache, et al. 2006. “The origin of European cattle: evidence from modern and ancient DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (21): 8113-8118. doi:10.1073/pnas.0509210103. http://www.ncbi.nlm.nih.gov/pubmed/16690747.

Benjamini, Yoav, and Y Hochberg. 1995. “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *Journal of the Royal Statistical Society, Series B* 57: 289-300.

Benson, Dennis A., Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, and Eric W. Sayers. 2009. “GenBank.” *Nucleic acids research* 37 (Database issue) (January): D26-31. doi:10.1093/nar/gkn723. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686462&tool=pmcentrez&rendertype=abstract.

Bibi, Faysal, and Elisabeth S. Vrba. 2010. “Unraveling bovin phylogeny: accomplishments and challenges.” *BMC biology* 8: 50. doi:10.1186/1741-7007-8-50. http://www.ncbi.nlm.nih.gov/pubmed/20525112.

Blackburn, H. D., Y. Toishibekov, M. Toishibekov, C. S. Welsh, S. F. Spiller, M. Brown, and S. R. Paiva. 2011. “Genetic diversity of Ovis aries populations near domestication centers and in the New World.” *Genetica* 139 (9) (September): 1169-78. doi:10.1007/s10709-011-9619-4. http://www.ncbi.nlm.nih.gov/pubmed/22105875.

Bolker, Benjamin M., Mollie E. Brooks, Connie J. Clark, Shane W. Geange, John R. Poulsen, M. Henry H. Stevens, and Jada-Simone S. White. 2009. “Generalized linear mixed models: a practical guide for ecology and evolution.” *Trends in ecology & evolution* 24 (3) (March): 127-35. doi:10.1016/j.tree.2008.10.008. http://www.ncbi.nlm.nih.gov/pubmed/19185386.

Bollongino, Ruth, Joachim Burger, Adam Powell, Marjan Mashkour, Jean-Denis Vigne, and Mark G. Thomas. 2012. “Modern Taurine Cattle descended from small number of Near-Eastern founders.” *Molecular Biology and Evolution* (March 14). doi:10.1093/molbev/mss092. http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mss092.

Bradley, Daniel G., David E. MacHugh, Patrick Cunningham, and Ronan T. Loftus. 1996. “Mitochondrial diversity and the origins of African and European cattle.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (10) (May 14): 5131-5. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=39419&tool=pmcentrez&rendertype=abstract.

Bronk Ramsey, Christopher. 2009. “Bayesian analysis of radiocarbon dates.” *Radiocarbon* 51 (1): 337-360. http://cat.inist.fr/?aModele=afficheN&cpsidt=21561335.

Bureš, D., L. Bartoň, R. Zahrádková, V. Teslík, and M. Fiedlerová. 2008. “Calving difficulty as related to body weights and measurements of cows and calves in a herd of Gascon breed.” *Czech Journal of Animal Science* 53 (5): 187-194.

Bushnell, Amy. 1978. “The Menéndez Marquéz Cattle Barony at La Chua and the Determinants of Economic Expansion in Seventeenth-Century Florida.” *The Florida Historical Quarterly* 56 (4): 407-431. http://www.jstor.org/stable/30150328.

Carlson, Christopher S., Daryl J. Thomas, Michael A. Eberle, Johanna E. Swanson, Robert J. Livingston, Mark J. Rieder, and Deborah A. Nickerson. 2005. “Genomic regions exhibiting positive selection identified from dense genotype data.” *Genome*

research 15 (11): 1553-1565. doi:10.1101/gr.4326505. http://www.ncbi.nlm.nih.gov/pubmed/16251465.

Cavalli-Sforza, Luigi L., Paolo Menozzi, and Alberto Piazza. 1993. “Demic expansions and human evolution.” *Science (New York, N.Y.)* 259 (5095) (January 29): 639-46. http://www.ncbi.nlm.nih.gov/pubmed/8430313.

Cheng, Peizhou, Xiao Liu, Guofan Zhang, and Yuewen Deng. 2006. “Heat-shock protein70 gene expression in four hatchery Pacific Abalone Haliotis discus hannai Ino populations using for marker-assisted selection.” *Aquaculture Research* 37 (13): 1290-1296. doi:10.1111/j.1365-2109.2006.01559.x. http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2109.2006.01559.x.

Clark, Andrew G., Melissa J. Hubisz, Carlos D. Bustamante, Scott H. Williamson, and Rasmus Nielsen. 2005. “Ascertainment bias in studies of human genome-wide polymorphism.” *Genome Research* 15 (11): 1496-1502. http://www.ncbi.nlm.nih.gov/pubmed/16251459.

Cockrill, W. Ross. 1974. *The husbandry and health of the domestic buffalo*. Ed. W. Ross Cockrill. Rome: Food and Agriculture Organization of the United Nations.

Deagan, Kathleen A. 1985. Spanish-Indian Interaction in Sixteenth-Century Florida and Hispaniola. In *Cultures in Contact*, ed. William W Fitzhugh, 281-318. Washington, DC: Anthropological Society of Washington and Smithsonian Institution Press. http://www.wou.edu/~smithr/496 INDIAN AMERICA/READINGS/Posted/Deagan_1985.pdf.

Dean, Frank B., Seiyu Hosono, Linhua Fang, Xiaohong Wu, A. Fawad Faruqi, Patricia Bray-Ward, Zhenyu Sun, et al. 2002. “Comprehensive human genome amplification using multiple displacement amplification.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (8): 5261-5266. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=122757&tool=pmcentrez&rendertype=abstract.

Decker, Jared E., J. Chris Pires, Gavin C. Conant, Stephanie D. McKay, Michael P. Heaton, Kefei Chen, Alan Cooper, et al. 2009. “Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (44) (November 3): 18644-18649. doi:10.1073/pnas.0904691106. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2765454&tool=pmcentrez&rendertype=abstract.

Devlin, Bernie, Silviu-Alin Bacanu, and Kathryn Roeder. 2004. Genomic Control to the extreme. *Nature Genetics*. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15514657.

Devlin, Bernie, and Kathryn Roeder. 1999. "Genomic control for association studies." *Biometrics* 55 (4) (December): 997-1004. http://www.ncbi.nlm.nih.gov/pubmed/11315092.

Diamond, Jared. 2002. "Evolution, consequences and future of plant and animal domestication." *Nature* 418 (6898) (August 8): 700-7. doi:10.1038/nature01019. http://www.ncbi.nlm.nih.gov/pubmed/12167878.

Editorial. 2008. "Prepare for the deluge." *Nature biotechnology* 26 (10) (October): 1099. doi:10.1038/nbt1008-1099. http://www.ncbi.nlm.nih.gov/pubmed/18846079.

Eggen, André. 2012. "The development and application of genomic selection as a new breeding paradigm." *Animal Frontiers* 2 (1) (January 18): 10-15. doi:10.2527/af.2011-0027. http://animalfrontiers.fass.org/cgi/doi/10.2527/af.2011-0027.

Elsik, Christine G., Ross L. Tellam, Kim C. Worley, Richard A. Gibbs, Donna M. Muzny, George M. Weinstock, David L. Adelson, et al. 2009. "The genome sequence of taurine cattle: a window to ruminant biology and evolution." *Science (New York, N.Y.)* 324 (5926) (April 24): 522-8. doi:10.1126/science.1169588. http://www.ncbi.nlm.nih.gov/pubmed/19390049.

Falconer, Douglas S., and Trudy F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Essex, UK.: Longmans Green, Harlow.

Felius, Marleen. 1995. *Cattle breeds - an Encyclopedia*. Doetinchem, Netherlands: Misset.

Felsenstein, Joseph. 1985. "Phylogenies and the Comparative Method." *American Naturalist* 125: 1-15.

———. 1989. "PHYLIP - Phylogeny Inference Package (Version 3.2)." *Cladistics* 5 (1): 164-166. http://www.citeulike.org/group/894/article/2344765.

Fernando, Rohan L., and Dorian J. Garrick. 2012. GenSel - User manual for a portfolio of genomic selection related analyses. http://taurus.ansci.iastate.edu/.

Fisher, Ronald Aylmer. 1918. “The correlation between relatives on the supposition of Mendelian inheritance.” *Transactions of the Royal Society of Edinburgh* 52: 399-433. http://digital.library.adelaide.edu.au/dspace/handle/2440/15097.

Fluxus. 2010a. DNA Alignment Software. *Fluxus Technology*. http://www.fluxus-technology.com/align.htm.

———. 2010b. Network | Network Publisher add-on. *Fluxus Technology*. http://www.fluxus-technology.com/nwpub.htm.

———. 2012. Free Phylogenetic Network Software. *Fluxus Technology*. http://www.fluxus-technology.com/sharenet.htm.

Forster, Peter, Antonio Torroni, Colin Renfrew, and Arne Röhl. 2001. “Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution.” *Molecular biology and evolution* 18 (10) (October): 1864-81. http://www.ncbi.nlm.nih.gov/pubmed/11557793.

Foss, Scott E., and Donald R. Prothero. 2007. Introduction. In *The Evolution of Artiodactyls*, ed. Scott E Foss and Donald R Prothero, 1-3. Baltimore, Maryland: The Johns Hopkins University Press.

Garrick, Dorian J., Jeremy F. Taylor, and Rohan L. Fernando. 2009. “Deregressing estimated breeding values and weighting information for genomic regression analyses.” *Genetics, selection, evolution : GSE* 41 (January): 55. doi:10.1186/1297-9686-41-55. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2817680&tool=pmcentrez&rendertype=abstract.

Gatesy, John, Deborah Yelon, Rob DeSalle, and Elisabeth S. Vrba. 1992. “Phylogeny of the Bovidae (Artiodactyla, Mammalia), based on mitochondrial ribosomal DNA sequences.” *Molecular biology and evolution* 9 (3): 433-446. http://www.ncbi.nlm.nih.gov/pubmed/1584013.

Gautier, Mathieu, and Michel Naves. 2011. “Footprints of selection in the ancestral admixture of a New World Creole cattle breed.” *Molecular Ecology* 20 (15) (August 20): 3128-3143. doi:10.1111/j.1365-294X.2011.05163.x. http://www.ncbi.nlm.nih.gov/pubmed/21689193.

Genome Bioinformatics Group of UC Santa Cruz. 2012. Cow (Bos taurus) Genome Browser Gateway. http://genome.ucsc.edu/cgi-bin/hgGateway.

Gibbs, Richard A., Jeremy F. Taylor, Curtis P. Van Tassell, William Barendse, Kellye A. Eversole, Clare A. Gill, Ronnie D. Green, et al. 2009. “Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds.” *Science* 324 (5926): 528-532. doi:10.1126/science.1167936. http://www.ncbi.nlm.nih.gov/pubmed/19390050.

Goddard, Michael E., and Ben J. Hayes. 2009. “Mapping genes for complex traits in domestic animals and their use in breeding programmes.” *Nature Reviews Genetics* 10 (6) (June): 381-91. doi:10.1038/nrg2575. http://www.ncbi.nlm.nih.gov/pubmed/19448663.

Goloboff, Pablo A., James S. Farris, and Kevin C. Nixon. 2008. “TNT, a free program for phylogenetic analysis.” *Cladistics* 24 (5): 13.

Gotherstrom, Anders, Cecilia Anderung, Linda Hellborg, Rengert Elburg, Colin Smith, Daniel G. Bradley, and Hans Ellegren. 2005. “Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe.” *Proceedings of the Royal Society B-Biological Sciences* 272 (1579): 2345-2350.

Guindon, Stéphane, and Olivier Gascuel. 2003. “A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.” *Systematic Biology* 52 (5): 696-704. http://sysbio.oxfordjournals.org/cgi/doi/10.1080/10635150390235520.

Habier, David, Rohan L. Fernando, Kadir Kizilkaya, and Dorian J. Garrick. 2011. “Extension of the bayesian alphabet for genomic selection.” *BMC bioinformatics* 12 (January): 186. doi:10.1186/1471-2105-12-186. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3144464&tool=pmcentrez&rendertype=abstract.

Hancock, Angela M., Gorka Alkorta-Aranburu, David B. Witonsky, and Anna Di Rienzo. 2010. “Adaptations to new environments in humans: the role of subtle allele frequency shifts.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365 (1552): 2459-2468. http://www.ncbi.nlm.nih.gov/pubmed/20643735.

Hancock, Angela M., David B. Witonsky, Edvard Ehler, Gorka Alkorta-Aranburu, Cynthia Beall, Amha Gebremedhin, Rem Sukernik, et al. 2010. “Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (Supplement_2): 8924-8930. http://www.ncbi.nlm.nih.gov/pubmed/20445095.

Hancock, Angela M., David B. Witonsky, Adam S. Gordon, Gidon Eshel, Jonathan K. Pritchard, Graham Coop, and Anna Di Rienzo. 2008. “Adaptations to Climate in

Candidate Genes for Common Metabolic Disorders." Ed. Dmitri A Petrov. *PLoS Genetics* 4 (2): 13. http://www.ncbi.nlm.nih.gov/pubmed/18282109.

Hanotte, Olivier, Daniel G. Bradley, Joel Ws Ochieng, Yasmin Verjee, Emmeline W. Hill, and J. Edward O. Rege. 2002. "African pastoralism: genetic imprints of origins and migrations." *Science* 296 (5566): 336-339. doi:10.1126/science.1069878. http://www.ncbi.nlm.nih.gov/pubmed/11951043.

Hassanin, Alexandre, and Emmanuel J. P. Douzery. 2003. "Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae." *Systematic biology* 52 (2): 206-228. doi:10.1080/1063515039019272. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12746147.

Hayes, Ben J., Phillip J. Bowman, Amanda C. Chamberlain, Klara Verbyla, and Michael E. Goddard. 2009. "Accuracy of genomic breeding values in multi-breed dairy cattle populations." *Genetics, selection, evolution : GSE* 41 (January): 51. doi:10.1186/1297-9686-41-51. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2791750&tool=pmcentrez&rendertype=abstract.

Hayes, Ben J., Phillip J. Bowman, Amanda J. Chamberlain, and Michael E. Goddard. 2009. "Invited review: Genomic selection in dairy cattle: progress and challenges." *Journal of dairy science* 92 (2) (February): 433-43. doi:10.3168/jds.2008-1646. http://www.ncbi.nlm.nih.gov/pubmed/19164653.

Hayes, Ben J., Amanda J. Chamberlain, S. Maceachern, K. Savin, H. McPartlan, I. MacLeod, L. Sethuraman, and Michael E. Goddard. 2009. "A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle." *Animal Genetics* 40 (2) (April): 176-184. doi:10.1111/j.1365-2052.2008.01815.x. http://www.ncbi.nlm.nih.gov/pubmed/19067671.

Hayes, Ben J., S. Lien, H. Nilsen, H. G. Olsen, P. Berg, Sean MacEachern, S. Potter, and T. H. Meuwissen. 2008. "The origin of selection signatures on bovine chromosome 6." *Animal genetics* 39 (2): 105-111. doi:10.1111/j.1365-2052.2007.01683.x. http://www.ncbi.nlm.nih.gov/pubmed/18241301.

Hayes, Ben J., Jennie Pryce, Amanda J. Chamberlain, Phillip J. Bowman, and Michael E. Goddard. 2010. "Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits." *PLoS genetics* 6 (9) (September): e1001139. doi:10.1371/journal.pgen.1001139.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2944788&tool=pmcentrez&rendertype=abstract.

Heath, Tracy A., Derrick J. Zwickl, Junhyong Kim, and David M. Hillis. 2008. “Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees.” *Systematic biology* 57 (1): 160-166. doi:10.1080/10635150701884640. http://www.ncbi.nlm.nih.gov/pubmed/18300029.

Heaton, Michael P., Gregory P. Harhay, Gary L. Bennett, Roger T. Stone, W. Michael Grosse, Eduardo Casas, John W. Keele, Timothy P. L. Smith, Carol G. Chitko-McKown, and William W. Laegreid. 2002. “Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle.” *Mammalian Genome* 13 (5): 272-281. http://www.ncbi.nlm.nih.gov/pubmed/12016516.

Henderson, Charles Roy. 1963. Selection index and expected genetic advance. In *Statistical genetics and plant breeding*, ed. W D Hanson and H F Robinson, 141-163. National Academy of Science.

Hill, William G. 2010. “Understanding and using quantitative genetic variation.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365 (1537) (January 12): 73-85. doi:10.1098/rstb.2009.0203. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2842708&tool=pmcentrez&rendertype=abstract.

Hill, William G, Michael E. Goddard, and Peter M. Visscher. 2008. “Data and theory point to mainly additive genetic variance for complex traits.” *PLoS genetics* 4 (2) (February): e1000008. doi:10.1371/journal.pgen.1000008. http://www.ncbi.nlm.nih.gov/pubmed/18454194.

Holsinger, Kent E., and Bruce S. Weir. 2009. “Genetics in geographically structured populations: defining, estimating and interpreting F(ST).” *Nature Reviews Genetics* 10 (9): 639-650. doi:10.1038/nrg2611. http://www.ncbi.nlm.nih.gov/pubmed/19687804.

Huang, Da Wei, Brad T. Sherman, and Richard A. Lempicki. 2009a. “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.” *Nature Protocols* 4 (1): 44-57. doi:10.1038/nprot.2008.211. http://www.ncbi.nlm.nih.gov/pubmed/19131956.

———. 2009b. “Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.” *Nucleic Acids Research* 37 (1): 1-13. doi:10.1093/nar/gkn923. http://www.ncbi.nlm.nih.gov/pubmed/19033363.

Huson, Daniel H., and David Bryant. 2006. “Application of phylogenetic networks in evolutionary studies.” *Molecular biology and evolution* 23 (2): 254-267. doi:msj030 [pii] 10.1093/molbev/msj030. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16221896.

Iwamoto, Kazuya, Junko Ueda, Yoko Nakano, Miki Bundo, Wataru Ukai, Eri Hashimoto, Toshikazu Saito, and Tadafumi Kato. 2007. “Evaluation of whole genome amplification methods using postmortem brain samples.” *Journal of Neuroscience Methods* 165 (1): 104-110. http://www.ncbi.nlm.nih.gov/pubmed/17604841.

Jakobsson, Mattias, Sonja W. Scholz, Paul Scheet, J. Raphael Gibbs, Jenna M. VanLiere, Hon-Chung Fung, Zachary A. Szpiech, et al. 2008. “Genotype, haplotype and copy-number variation in worldwide human populations.” *Nature* 451 (7181) (February 21): 998-1003. doi:10.1038/nature06742. http://www.ncbi.nlm.nih.gov/pubmed/18288195.

Kang, Hyun Min, Jae Hoon Sul, Susan K. Service, Noah A. Zaitlen, Sit-Yee Kong, Nelson B. Freimer, Chiara Sabatti, and Eleazar Eskin. 2010. “Variance component model to account for sample structure in genome-wide association studies.” *Nature genetics* 42 (4) (April): 348-54. doi:10.1038/ng.548. http://www.ncbi.nlm.nih.gov/pubmed/20208533.

Kang, Hyun Min, Noah A. Zaitlen, Claire M. Wade, Andrew Kirby, David Heckerman, Mark J. Daly, and Eleazar Eskin. 2008. “Efficient Control of Population Structure in Model Organism Association Mapping.” *Genetics* 178 (3): 1709-1723. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2278096&tool=pmcentrez&rendertype=abstract.

Kelley, Joanna L., Jennifer Madeoy, John C. Calhoun, Willie Swanson, and Joshua M. Akey. 2006. “Genomic signatures of positive selection in humans and the limits of outlier approaches.” *Genome research* 16 (8) (August): 980-9. doi:10.1101/gr.5157306. http://www.ncbi.nlm.nih.gov/pubmed/16825663.

Kenny, Eimear E., Nicholas J. Timpson, Martin Sikora, Muh-Ching Yee, Andres Moreno-Estrada, Celeste Eng, Scott Huntsman, et al. 2012. “Melanesian Blond Hair Is Caused by an Amino Acid Change in TYRP1.” *Science* 336 (6081) (May 3): 554-554. doi:10.1126/science.1217849. http://www.sciencemag.org/cgi/doi/10.1126/science.1217849.

Kent, W. James, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. 2002. “The Human Genome Browser at

UCSC." *Genome Research* 12 (6) (May 16): 996-1006. doi:10.1101/gr.229102. http://www.genome.org/cgi/doi/10.1101/gr.229102.

Kijas, James W., David Townley, Brian P. Dalrymple, Michael P. Heaton, Jillian F. Maddox, Annette McGrath, Peter Wilson, et al. 2009. "A genome wide survey of SNP variation reveals the genetic structure of sheep breeds." *PloS one* 4 (3) (January): e4668. doi:10.1371/journal.pone.0004668. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2652362&tool=pmcentrez&rendertype=abstract.

Killian, Gary J., David A. Chapman, and Lee A. Rogowski. 1993. "Fertility-associated proteins in Holstein bull seminal plasma." *Biology of Reproduction* 49 (6): 1202-1207. doi:10.1095/biolreprod49.6.1202. http://www.ncbi.nlm.nih.gov/pubmed/8286602.

Kimura, Motoo. 1983. *The neutral theory of molecular evolution*. Cambridge Cambridgeshire ; New York: Cambridge University Press. http://www.loc.gov/catdir/description/cam041/82022225.html.

Kristensen, Torsten Nygaard, Kamilla S. Pedersen, Cornelis J. Vermeulen, and Volker Loeschcke. 2010. "Research on inbreeding in the 'omic' era." *Trends in ecology & evolution* 25 (1) (January): 44-52. doi:10.1016/j.tree.2009.06.014. http://www.ncbi.nlm.nih.gov/pubmed/19733933.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, et al. 2007. "Clustal W and Clustal X version 2.0." *Bioinformatics (Oxford, England)* 23 (21) (November 1): 2947-8. doi:10.1093/bioinformatics/btm404. http://www.ncbi.nlm.nih.gov/pubmed/17846036.

Li, Jun Z., Devin M. Absher, Hua Tang, Audrey M. Southwick, Amanda M. Casto, Sohini Ramachandran, Howard M. Cann, et al. 2008. "Worldwide human relationships inferred from genome-wide patterns of variation." *Science (New York, N.Y.)* 319 (5866) (February 22): 1100-4. doi:10.1126/science.1153717. http://www.ncbi.nlm.nih.gov/pubmed/18292342.

Liron, J. P., C. M. Bravi, P. M. Mirol, P. Peral-Garcia, and G. Giovambattista. 2006. "African matrilineages in American Creole cattle: evidence of two independent continental sources." *Animal Genetics* 37 (4) (August): 379-382. doi:10.1111/j.1365-2052.2006.01452.x. http://www.ncbi.nlm.nih.gov/pubmed/16879351.

Loftus, Ronan T., David E. MacHugh, Daniel G. Bradley, Paul M. Sharp, and Patrick Cunningham. 1994. "Evidence for two independent domestications of cattle."

Proceedings of the National Academy of Sciences of the United States of America 91 (7): 2757-2761. http://www.ncbi.nlm.nih.gov/pubmed/8146187.

Lynch, Michael. 2010. “Rate, molecular spectrum, and consequences of human mutation.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (3): 961-968. http://www.ncbi.nlm.nih.gov/pubmed/20080596.

MacEachern, Sean, John McEwan, Alan McCulloch, Andrew Mather, Keith Savin, and Michael E. Goddard. 2009. “Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle?” *BMC genomics* 10 (January): 179. doi:10.1186/1471-2164-10-179. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2681479&tool=pmcentrez&rendertype=abstract.

MacHugh, David E., Mark D. Shriver, Ronan T. Loftus, Patrick Cunningham, and Daniel G. Bradley. 1997. “Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus).” *Genetics* 146 (3): 1071-1086. http://www.genetics.org/cgi/content/abstract/146/3/1071.

Maglott, Donna, Jim Ostell, Kim D. Pruitt, and Tatiana Tatusova. 2007. “Entrez Gene: gene-centered information at NCBI.” *Nucleic Acids Research* 35 (Database issue): D54-D58. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=539985&tool=pmcentrez&rendertype=abstract.

Maher, Brendan. 2008. “Personal genomes: The case of the missing heritability.” *Nature* 456 (7218) (November 6): 18-21. doi:10.1038/456018a. http://www.ncbi.nlm.nih.gov/pubmed/18987709.

Mannen, Hideyuki, M Kohno, Y Nagata, S Tsuji, Daniel G. Bradley, J S Yeo, D Nyamsamba, et al. 2004. “Independent mitochondrial origin and historical genetic differentiation in North Eastern Asian cattle.” *Molecular phylogenetics and evolution* 32 (2) (August): 539-44. doi:10.1016/j.ympev.2004.01.010. http://www.ncbi.nlm.nih.gov/pubmed/15223036.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. “Finding the missing heritability of complex diseases.” *Nature* 461 (7265) (October 8): 747-53. doi:10.1038/nature08494. http://www.ncbi.nlm.nih.gov/pubmed/19812666.

Marcot, Jonathan D. 2007. Molecular Phylogeny of Terrestrial Artiodactyls: Conflicts and Resolution. In *The Evolution of Artiodactyls*, ed. Scott E Foss and Donald R Prothero, 4-18. Baltimore, Maryland: The Johns Hopkins University Press.

Matukumalli, Lakshmi K., Cynthia T. Lawley, Robert D. Schnabel, Jeremy F. Taylor, Mark F. Allan, Michael P. Heaton, Jeffrey R. O'Connell, et al. 2009. "Development and characterization of a high density SNP genotyping assay for cattle." *PLoS One* 4 (4): e5350. doi:10.1371/journal.pone.0005350. http://www.ncbi.nlm.nih.gov/pubmed/19390634.

McClure, Matthew C., Holly R. Ramey, Megan M. Rolf, Stephanie D. McKay, Jared E. Decker, Richard H. Chapple, JaeWoo Kim, et al. 2012. "Genome-wide association analysis for quantitative trait loci influencing Warner-Bratzler shear force in five taurine cattle breeds." *Animal Genetics* (February 27): n/a-n/a. doi:10.1111/j.1365-2052.2012.02323.x. http://doi.wiley.com/10.1111/j.1365-2052.2012.02323.x.

McCue, Molly E., Danika L. Bannasch, Jessica L. Petersen, Jessica Gurr, Ernie Bailey, Matthew M. Binns, Ottmar Distl, et al. 2012. "A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies." Ed. Michel Georges. *PLoS Genetics* 8 (1) (January 12): e1002451. doi:10.1371/journal.pgen.1002451. http://dx.plos.org/10.1371/journal.pgen.1002451.

McKay, Stephanie D, Robert D. Schnabel, Brenda M Murdoch, Lakshmi K. Matukumalli, Jan Aerts, Wouter Coppieters, Denny Crews, et al. 2007. "Whole genome linkage disequilibrium maps in cattle." *BMC genetics* 8 (1) (January 25): 74. doi:10.1186/1471-2156-8-74. http://www.biomedcentral.com/1471-2156/8/74.

McVean, Gil. 2009. "A genealogical interpretation of principal components analysis." *PLoS genetics* 5 (10) (October): e1000686. doi:10.1371/journal.pgen.1000686. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2757795&tool=pmcentrez&rendertype=abstract.

Meuwissen, T. H. E., Ben J. Hayes, and Michael E. Goddard. 2001. "Prediction of total genetic value using genome-wide dense marker maps." *Genetics* 157 (4): 1819-1829. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461589&tool=pmcentrez&rendertype=abstract.

Mirol, P. M., G. Giovambattista, J. P. Lirón, and F. N. Dulout. 2003. "African and European mitochondrial haplotypes in South American Creole cattle." *Heredity* 91 (3): 248-254. http://www.ncbi.nlm.nih.gov/pubmed/12939625.

Moritz, Craig. 1995. “Uses of molecular phylogenies for conservation.” *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 349: 113-118.

Myles, Sean, Adam R. Boyko, Christopher L. Owens, Patrick J. Brown, Fabrizio Grassi, Mallikarjuna K. Aradhya, Bernard Prins, et al. 2011. “Genetic structure and domestication history of the grape.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (9) (March 1): 3530-5. doi:10.1073/pnas.1009363108. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3048109&tool=pmcentrez&rendertype=abstract.

Neale, Benjamin M., Yan Kou, Li Liu, Avi Ma’ayan, Kaitlin E. Samocha, Aniko Sabo, Chiao-Feng Lin, et al. 2012. “Patterns and rates of exonic de novo mutations in autism spectrum disorders.” *Nature* (April 4). doi:10.1038/nature11011. http://www.nature.com/doifinder/10.1038/nature11011.

Nugent, R. A., D. R. Notter, and W. E. Beal. 1991. “Body measurements of newborn calves and relationship of calf shape to sire breeding values for birth weight and calving ease.” *Journal of Animal Science* 69 (6): 2413-2421. http://www.ncbi.nlm.nih.gov/pubmed/1885359.

O’Roak, Brian J., Laura Vives, Santhosh Girirajan, Emre Karakoc, Niklas Krumm, Bradley P. Coe, Roie Levy, et al. 2012. “Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations.” *Nature* (April 4). doi:10.1038/nature10989. http://www.nature.com/doifinder/10.1038/nature10989.

Patterson, Nick J., Alkes L. Price, and David Reich. 2006. “Population structure and eigenanalysis.” *PLoS genetics* 2 (12) (December): e190. doi:10.1371/journal.pgen.0020190. http://www.ncbi.nlm.nih.gov/pubmed/17194218.

Pedersen, K. S., Torsten Nygaard Kristensen, and Volker Loeschcke. 2005. “Effects of inbreeding and rate of inbreeding in Drosophila melanogaster- Hsp70 expression and fitness.” *Journal of Evolutionary Biology* 18 (4) (July): 756-762. doi:10.1111/j.1420-9101.2005.00884.x. http://www.ncbi.nlm.nih.gov/pubmed/16033546.

Pellecchia, Marco, Riccardo Negrini, Licia Colli, Massimiliano Patrini, Elisabetta Milanesi, Alessandro Achilli, Giorgio Bertorelle, et al. 2007. “The mystery of Etruscan origins: novel clues from Bos taurus mitochondrial DNA.” *Proceedings. Biological sciences /*

The Royal Society 274 (1614) (May 7): 1175-9. doi:10.1098/rspb.2006.0258. http://www.ncbi.nlm.nih.gov/pubmed/17301019.

Poduri, Annapurna, Gilad D. Evrony, Xuyu Cai, Princess Christina Elhosary, Rameen Beroukhim, Maria K. Lehtinen, L. Benjamin Hills, et al. 2012. "Somatic Activation of AKT3 Causes Hemispheric Developmental Brain Malformations." *Neuron* 74 (1) (April 12): 41-8. doi:10.1016/j.neuron.2012.03.010. http://www.ncbi.nlm.nih.gov/pubmed/22500628.

Pollack, Andrew. 2011. "A Genome Deluge." *The New York Times*, December 1. http://www.nytimes.com/2011/12/01/business/dna-sequencing-caught-in-deluge-of-data.html?_r=1.

Polzin, Tobias, and Siavash Vahdati Daneshmand. 2003. "On Steiner trees and minimum spanning trees in hypergraphs." *Operations Research Letters* 31 (1): 12-20. doi:10.1016/S0167-6377(02)00185-2. http://linkinghub.elsevier.com/retrieve/pii/S0167637702001852.

Porter, Valerie. 1991. *Cattle: A handbook to the breeds of the world*. London: Christopher Helm (Publishers) Ltd.

Powell, Joseph E., Peter M. Visscher, and Michael E. Goddard. 2010. "Reconciling the analysis of IBD and IBS in complex trait studies." *Nature Reviews Genetics* 11 (11): 800-805. http://www.ncbi.nlm.nih.gov/pubmed/20877324.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. "Principal components analysis corrects for stratification in genome-wide association studies." *Nature genetics* 38 (8) (August): 904-9. doi:10.1038/ng1847. http://www.ncbi.nlm.nih.gov/pubmed/16862161.

Price, Alkes L., Noah A. Zaitlen, David Reich, and Nick J. Patterson. 2010. "New approaches to population stratification in genome-wide association studies." *Nature reviews. Genetics* 11 (7) (July): 459-63. doi:10.1038/nrg2813. http://www.ncbi.nlm.nih.gov/pubmed/20548291.

Pritchard, Jonathan K., Joseph K. Pickrell, and Graham Coop. 2010. "The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation." *Current Biology* 20 (4): R208-R215. http://www.ncbi.nlm.nih.gov/pubmed/20178769.

Pritchard, Jonathan K., M. Stephens, N. A. Rosenberg, and P. Donnelly. 2000. "Association mapping in structured populations." *American journal of human genetics* 67 (1) (July): 170-81. doi:10.1086/302959.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1287075&tool=pmcentrez&rendertype=abstract.

Pritchard, Jonathan K., and Anna Di Rienzo. 2010. “Adaptation - not by sweeps alone.” *Nature Reviews Genetics* 11 (10): 665-667. http://www.nature.com/doifinder/10.1038/nrg2880.

Purcell, Shaun M. 2009. PLINK. http://pngu.mgh.harvard.edu/purcell/plink/.

Purcell, Shaun M., Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: a tool set for whole-genome association and population-based linkage analyses.” *American journal of human genetics* 81 (3) (September): 559-75. doi:10.1086/519795. http://pngu.mgh.harvard.edu/purcell/plink/.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. “A genome-wide scan for signatures of recent selection in Holstein cattle.” *Animal genetics* 41 (4) (August): 377-89. doi:10.1111/j.1365-2052.2009.02016.x. http://www.ncbi.nlm.nih.gov/pubmed/20096028.

Qanbari, Saber, Daniel Gianola, Ben J. Hayes, Flavio Schenkel, Steve Miller, Stephen Moore, Georg Thaller, and Henner Simianer. 2011. “Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle.” *BMC genomics* 12 (January): 318. doi:10.1186/1471-2164-12-318. http://www.ncbi.nlm.nih.gov/pubmed/21679429.

Quaas, Richard L., R. D. Anderson, and A. R. Gilmour. 1984. Use of mixed models for prediction and for estimation of (co)variance components. In *BLUP school handbook*. Animal Genetics and Breeding Unit, University of New England, NSW, Australia.

Quaas, Richard L., and Emil J. Pollak. 1980. “Mixed Model Methodology for Farm and Ranch Beef Cattle Testing Programs.” *Journal of Animal Science* 51: 1277-1287.

R Development Core Team, R. 2011. R: A Language and Environment for Statistical Computing. Ed. R Development Core Team. *R Foundation for Statistical Computing*. R Foundation for Statistical Computing. http://www.r-project.org.

Reimer, P. J., M. G. L. Baillie, E. Bard, A. Bayliss, J. W. Beck, P. G. Blackwell, C. Bronk Ramsey, et al. 2009. “IntCal09 and Marine09 Radiocarbon Age Calibration Curves, 0-50,000 Years cal BP.” *Radiocarbon* 51 (4): 1111-1150. doi:10.2458/rc.v51i4.3569. http://digitalcommons.arizona.edu/objectviewer?o=http://radiocarbon.library.arizona.edu/Volume51/Number4/49691745-6a68-4e2c-a26f-08f0a16c1a53.

Reynolds, John, Bruce S. Weir, and C. Clark Cockerham. 1983. “Estimation of the coancestry coefficient: basis for a short-term genetic distance.” *Genetics* 105 (3): 767-779. http://www.ncbi.nlm.nih.gov/pubmed/17246175.

Rokas, Antonis, and Sean B. Carroll. 2005. “More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy.” *Molecular Biology and Evolution* 22 (5): 1337-1344. doi:10.1093/molbev/msi121. http://www.ncbi.nlm.nih.gov/pubmed/15746014.

Rosenblum, Erica Bree, and John Novembre. 2007. “Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard.” *The Journal of heredity* 98 (4): 331-336. http://www.ncbi.nlm.nih.gov/pubmed/17611259.

Saatchi, Mahdi, Matthew C. McClure, Stephanie D. McKay, Megan M. Rolf, Jaewoo Kim, Jared E. Decker, Tasia M. Taxis, et al. 2011. “Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation.” *Genetics, selection, evolution : GSE* 43 (1) (November 28): 40. doi:10.1186/1297-9686-43-40. http://www.ncbi.nlm.nih.gov/pubmed/22122853.

Sabeti, Pardis C., Patrick Varilly, Ben Fry, Jason Lohmueller, Elizabeth Hostetter, Chris Cotsapas, Xiaohui Xie, et al. 2007. “Genome-wide detection and characterization of positive selection in human populations.” *Nature* 449 (7164) (October 18): 913-8. doi:10.1038/nature06250. http://www.ncbi.nlm.nih.gov/pubmed/17943131.

Sambrook, Joseph, Ed F. Fritsch, and T. Maniatis. 1989. *Molecular cloning : a laboratory manual*. *Cold Spring Harbor Laboratory Press*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sanders, Stephan J., Michael T. Murtha, Abha R. Gupta, John D. Murdoch, Melanie J. Raubeson, A. Jeremy Willsey, A. Gulhan Ercan-Sencicek, et al. 2012. “De novo mutations revealed by whole-exome sequencing are strongly associated with autism.” *Nature* (April 4). doi:10.1038/nature10945. http://www.nature.com/doifinder/10.1038/nature10945.

Scheet, Paul, and Matthew Stephens. 2006. “A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.” *American journal of human genetics* 78 (4) (April): 629-44. doi:10.1086/502802. http://www.ncbi.nlm.nih.gov/pubmed/16532393.

Shapiro, Beth, Alexei J. Drummond, Andrew Rambaut, Michael C. Wilson, Paul E. Matheus, Andrei V. Sher, Oliver G. Pybus, et al. 2004. “Rise and fall of the Beringian steppe bison.” *Science (New York, N.Y.)* 306 (5701) (November 26): 1561-5. doi:10.1126/science.1101074. http://www.ncbi.nlm.nih.gov/pubmed/15567864.

Shriver, Mark D., Giulia C. Kennedy, Esteban J. Parra, Heather A. Lawson, Vibhor Sonpar, Jing Huang, Joshua M. Akey, and Keith W. Jones. 2004. "The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs." *Human Genomics* 1 (4): 274-286. http://www.ncbi.nlm.nih.gov/pubmed/15588487.

Sorensen, Jesper Givskov, Torsten Nygaard Kristensen, and Volker Loeschcke. 2003. "The evolutionary and ecological role of heat shock proteins." *Ecology Letters* 6 (11): 1025-1037. doi:10.1046/j.1461-0248.2003.00528.x. http://www.blackwell-synergy.com/links/doi/10.1046/j.1461-0248.2003.00528.x.

Stahl, Eli A., Daniel Wegmann, Gosia Trynka, Javier Gutierrez-Achury, Ron Do, Benjamin F. Voight, Peter Kraft, et al. 2012. "Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis." *Nature Genetics* (March 25): 1-9. doi:10.1038/ng.2232. http://www.nature.com/doifinder/10.1038/ng.2232.

Stavrinides, John, Honour C. McCann, and David S. Guttman. 2008. "Host-pathogen interplay and the evolution of bacterial effectors." *Cellular Microbiology* 10 (2): 285-292. http://www.ncbi.nlm.nih.gov/pubmed/18034865.

Steemers, Frank J., Weihua Chang, Grace Lee, David L. Barker, Richard Shen, and Kevin L. Gunderson. 2006. "Whole-genome genotyping with the single-base extension assay." *Nature methods* 3 (1) (January): 31-3. doi:10.1038/nmeth842. http://www.ncbi.nlm.nih.gov/pubmed/16369550.

Storey, John D., and Robert Tibshirani. 2003. "Statistical significance for genomewide studies." *Proceedings of the National Academy of Sciences of the United States of America* 100 (16): 9440-9445. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=170937&tool=pmcentrez&rendertype=abstract.

Swenson, Nathan G. 2009. "Phylogenetic resolution and quantifying the phylogenetic diversity and dispersion of communities." *PloS one* 4 (2) (January): e4390. doi:10.1371/journal.pone.0004390. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2633039&tool=pmcentrez&rendertype=abstract.

Van Tassell, Curtis P., Timothy P. L. Smith, Lakshmi K. Matukumalli, Jeremy F. Taylor, Robert D. Schnabel, Cynthia Taylor Lawley, Christian D. Haudenschild, Stephen S. Moore, Wesley C. Warren, and Tad S. Sonstegard. 2008. "SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries." *Nature methods* 5 (3) (March): 247-52. doi:10.1038/nmeth.1185. http://www.ncbi.nlm.nih.gov/pubmed/18297082.

Thornton, Timothy, and Mary Sara McPeek. 2010. “ROADTRIPS: Case-Control Association Testing with Partially or Completely Unknown Population and Pedigree Structure.” *The American Journal of Human Genetics* 86 (2): 172-184. http://www.ncbi.nlm.nih.gov/pubmed/20137780.

Turner, Stephen. 2011. Annotated Manhattan plots and QQ plots for GWAS using R, Revisited. Getting Genetics Done. http://gettinggeneticsdone.blogspot.com/2011/04/annotated-manhattan-plots-and-qq-plots.html.

VanRaden, Paul M. 2008. “Efficient methods to compute genomic predictions.” *Journal of dairy science* 91 (11) (November): 4414-23. doi:10.3168/jds.2007-0980. http://www.ncbi.nlm.nih.gov/pubmed/18946147.

VanRaden, Paul M., George Wiggans, Curtis P. Van Tassell, Tad S. Sonstegard, and Leigh Walton. 2008. Genomic prediction. http://aipl.arsusda.gov/reference/genomic prediction.html.

Visscher, Peter M., Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, and Nicholas G. Martin. 2006. “Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings.” *PLoS genetics* 2 (3) (March): e41. doi:10.1371/journal.pgen.0020041. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1413498&tool=pmcentrez&rendertype=abstract.

Voight, Benjamin F., Sridhar Kudaravalli, Xiaoquan Wen, and Jonathan K. Pritchard. 2006. “A map of recent positive selection in the human genome.” *PLoS biology* 4 (3) (March): e72. doi:10.1371/journal.pbio.0040072. http://www.ncbi.nlm.nih.gov/pubmed/16494531.

Vonholdt, Bridgett M., John P. Pollinger, Kirk E. Lohmueller, Eunjung Han, Heidi G. Parker, Pascale Quignon, Jeremiah D. Degenhardt, et al. 2010. “Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication.” *Nature* 464 (7290) (April 8): 898-902. doi:10.1038/nature08837. http://www.ncbi.nlm.nih.gov/pubmed/20237475.

Walker, Angela M., and R. Michael Roberts. 2009. “Characterization of the bovine type I IFN locus: rearrangements, expansions, and novel subfamilies.” *BMC Genomics* 10 (1471-2164 (Electronic) LA - eng PT - Journal Article PT - Research Support, N.I.H., Extramural SB - IM): 187. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2680415&tool=pmcentrez&rendertype=abstract.

Wall, E., S. Brotherstone, J.F. Kearney, J.A. Woolliams, and M.P. Coffey. 2005. “Impact of Nonadditive Genetic Effects in the Estimation of Breeding Values for Fertility and Correlated Traits.” *Journal of Dairy Science* 88 (1) (January): 376-385. doi:10.3168/jds.S0022-0302(05)72697-7. http://www.ncbi.nlm.nih.gov/pubmed/15591402.

Wang, Eric T., Greg Kodama, Pierre Baldi, and Robert K. Moyzis. 2006. “Global landscape of recent inferred Darwinian selection for Homo sapiens.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (1) (January 3): 135-40. doi:10.1073/pnas.0509691102. http://www.ncbi.nlm.nih.gov/pubmed/16371466.

Weir, Bruce S., Lon R. Cardon, Amy D. Anderson, Dahlia M. Nielsen, and William G. Hill. 2005. “Measures of human population structure show heterogeneity among genomic regions.” *Genome Research* 15 (11) (November 1): 1468-1476. doi:10.1101/gr.4398405. http://www.ncbi.nlm.nih.gov/pubmed/16251456.

Wiens, John J. 1998. “Does Adding Characters with Missing Data Increase of Decrease Phylogenetic Accuracy?” *Systematic biology* 47 (4): 625-640. http://www.ncbi.nlm.nih.gov/pubmed/12066307.

———. 2003. “Missing data, incomplete taxa, and phylogenetic accuracy.” *Systematic biology* 52 (4): 528-538. doi:10.1080/10635150390218330. http://sysbio.oxfordjournals.org/cgi/doi/10.1080/10635150390218330.

Yang, Jian, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, et al. 2010. “Common SNPs explain a large proportion of the heritability for human height.” *Nature genetics* 42 (7) (July): 565-9. doi:10.1038/ng.608. http://www.ncbi.nlm.nih.gov/pubmed/20562875.

Yu, Jianming, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, et al. 2006. “A unified mixed-model method for association mapping that accounts for multiple levels of relatedness.” *Nature Genetics* 38 (2): 203-208. http://www.ncbi.nlm.nih.gov/pubmed/16380716.

Zimin, Aleksey V., Arthur L. Delcher, Liliana Florea, David R. Kelley, Michael C. Schatz, Daniela Puiu, Finnian Hanrahan, et al. 2009. “A whole-genome assembly of the domestic cow, Bos taurus.” *Genome biology* 10 (4) (January): R42. doi:10.1186/gb-2009-10-4-r42. http://www.ncbi.nlm.nih.gov/pubmed/19393038.

Zuccon, A., and D. Zuccon. 2008. “MrEnt v.3. Program distributed by the authors.” http://www.mrent.org/.

VITA

Jared Egan Decker was born in 1982 in Durango, Colorado. Jared has known since he was 6 years old that he wanted to be a scientist; he just didn't know if he wanted to study dinosaurs, birds, or airplanes. In 1990, his family moved to a small farm in La Plata, New Mexico, and in 1995 they moved to a larger farm. It was Jared's competitive nature and love of cattle that sparked his interest in genetics. Jared wanted to win at the fair, so he was always searching the next great bull to breed to his Hereford heifers. In 2001, the summer after graduating from Farmington High School, Jared completed his goal and won grand champion heifer with a heifer he bred and raised.

In May of 2007, Jared graduated with Highest Honors from the College of Agriculture at New Mexico State University. During his time at New Mexico State, Jared worked in a nutritional toxicology laboratory and a physiological genetics laboratory. As part of his studies he also completed an Honors Thesis on scrotal growth in beef bulls. Most importantly, he was a summer undergraduate research intern at the University of Missouri working in Dr. Jeremy F. Taylor's animal genomics laboratory.

In August of 2007, Jared rejoined the Taylor group as a Ph.D. student. During his time in the Taylor lab, Jared has presented at international meetings, published in top tier journals, and received numerous awards.

In 2008, Jared married his lovely wife Mary. In addition to his step-daughter, Jared is the father of two boys born in 2009 and 2010.