PLOS GENETICS

RESEARCH ARTICLE

Powerful detection of polygenic selection and evidence of environmental adaptation in US beef cattle

Troy N. Rowan[1,2,3,4], **Harly J. Durbin**[1,2], **Christopher M. Seabury**[5], **Robert D. Schnabel**[1,2,6], **Jared E. Decker**[1,2,6]*

1 Division of Animal Sciences, University of Missouri, Columbia, Missouri, United States of America, **2** Genetics Area Program, University of Missouri, Columbia, Missouri, United States of America, **3** Department of Animal Science, University of Tennessee, Knoxville, Tennessee, United States of America, **4** College of Veterinary Medicine, Large Animal Clinical Science, University of Tennessee, Knoxville, Tennessee, United States of America, **5** Department of Veterinary Pathobiology, Texas A&M University, College Station, Texas, United States of America, **6** Institute for Data Science and Informatics, University of Missouri, Columbia, Missouri, United States of America

* deckerje@missouri.edu

Citation: Rowan TN, Durbin HJ, Seabury CM, Schnabel RD, Decker JE (2021) Powerful detection of polygenic selection and evidence of environmental adaptation in US beef cattle. PLoS Genet 17(7): e1009652. https://doi.org/10.1371/journal.pgen.1009652

Editor: Alex Buerkle, University of Wyoming, UNITED STATES

Received: September 22, 2020

Accepted: June 9, 2021

Published: July 22, 2021

Data Availability Statement: The raw data underlying the results presented in the study are available from the Red Angus Association of America (https://redangus.org/contact-us/, Ryan Boldt, Director of Breed Improvement, ryan@redangus.org, (940) 387-3502 ext. 12), American Simmental Association (https://simmental.org/site/index.php/contact#asadirectors, Jackie Atkins, Director Science and Education Operations, jatkins@simmgene.com, (406)587-4531), and American Gelbvieh Association (https://gelbvieh.

Abstract

Selection on complex traits can rapidly drive evolution, especially in stressful environments. This polygenic selection does not leave intense sweep signatures on the genome, rather many loci experience small allele frequency shifts, resulting in large cumulative phenotypic changes. Directional selection and local adaptation are changing populations; but, identifying loci underlying polygenic or environmental selection has been difficult. We use genomic data on tens of thousands of cattle from three populations, distributed over time and landscapes, in linear mixed models with novel dependent variables to map signatures of selection on complex traits and local adaptation. We identify 207 genomic loci associated with an animal's birth date, representing ongoing selection for monogenic and polygenic traits. Additionally, hundreds of additional loci are associated with continuous and discrete environments, providing evidence for historical local adaptation. These candidate loci highlight the nervous system's possible role in local adaptation. While advanced technologies have increased the rate of directional selection in cattle, it has likely been at the expense of historically generated local adaptation, which is especially problematic in changing climates. When applied to large, diverse cattle datasets, these selection mapping methods provide an insight into how selection on complex traits continually shapes the genome. Further, understanding the genomic loci implicated in adaptation may help us breed more adapted and efficient cattle, and begin to understand the basis for mammalian adaptation, especially in changing climates. These selection mapping approaches help clarify selective forces and loci in evolutionary, model, and agricultural contexts.

org/about/contact, Megan Slater, Executive Director, megans@gelbvieh.org, (303)465-2333, Ext 485) under a Data Use Agreement, but are not publicly available. Derived data (analytical results) are however available as supplementary files or as a Zenodo repository (doi:10.5281/zenodo.4455543) associated with this publication.

Funding: This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2016-68004-24827, no. MO-HAAS0027, and NSRP-8 from the USDA National Institute of Food and Agriculture awarded to JED. https://reeis.usda.gov/web/crisprojectpages/1008909-identifying-local-adaptation-and-creating-region-specific-genomic-predictions-in-beef-cattle.html Funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Interest in mapping the impacts of selection and local adaptation on the genome is increasing due to the novel stressors presented by climate change. Until now, approaches have largely focused on mapping "sweeps" on large-effect loci. Highly powered datasets that are both temporally and geographically distributed have not existed. Recently, large numbers of beef cattle have been genotyped across the United States, including influential individuals with cryopreserved semen. This has created multiple powerful datasets distributed over time and landscapes. Here, we map the recent effects of selection and local adaptation in three cattle populations. The results provide insight into the biology of mammalian adaptation and generate useful tools for selecting and breeding better-adapted cattle for a changing environment.

Introduction

As climate changes, organisms either migrate, rapidly adapt, or perish. The genes and alleles that underlie adaptation have been difficult to identify, except for a handful of large-effect variants that underwent selective sweeps [1]. It is becoming increasingly apparent that for adaptation, hard sweeps are likely to be the exception, rather than the rule [2]. Polygenic selection on complex traits can cause a significant change in the mean phenotype while producing only subtle changes in allele frequencies throughout the genome [3]. Additionally, we expect that polygenic selection is the major selective force both during and after domestication in agricultural species. Many selection mapping methods rely on allele frequency differences between diverged or artificially defined populations (e.g. F_{ST}, FLK, XP-CLR) [4–6], making the detection of selection within a largely panmictic population difficult. Others rely on detecting the disruption of normal LD patterns (iHS, EHH, ROH, etc.) [7–9]. In cattle, these methods have successfully identified genomic regions under selection that control Mendelian and simple traits like coat color, the absence of horns, or large-effect genes involved in domestication [10–15]. Further, in many cases these models are unable to derive additional power from massive increases in sample size [16]. Millions of North American *Bos taurus* beef cattle have been exposed to strong artificial and environmental selection for more than 50 years (~10 generations) [17], making them a powerful model for studying the impacts selection has on genomes over short time periods and across diverse environments.

Though the first cattle single nucleotide polymorphism (SNP) genotyping assay was developed just over a decade ago [18], numerous influential males who have been deceased for 30 to 40 years have been genotyped from cryopreserved semen (**S1 Fig, Table A in S1 Text**). These bulls add a temporally-stratified, multi-generational component to the datasets of thousands of contemporary animals genotyped from the numerically largest US beef breeds. Furthermore, the large number of animals genotyped from the most recent generations provide remarkable power for detecting small allele frequency changes due to ongoing selection. Under directional selection, alleles will be at significantly different frequencies in more recent generations compared with distant ones (**Fig 1C**). This creates a statistical association between allele frequencies at a selected locus and an individual's generation number. With multiple generations sampled and genotyped, we can disentangle small shifts in allele frequency due to directional selection from the stochastic small changes caused by drift (**Fig 1A**) using Generation Proxy Selection Mapping (GPSM) [17,19]. The GPSM method searches for allele frequency changes by identifying allelic associations with an individual's generation (or some proxy), while accounting for confounding population and family structure with a genomic

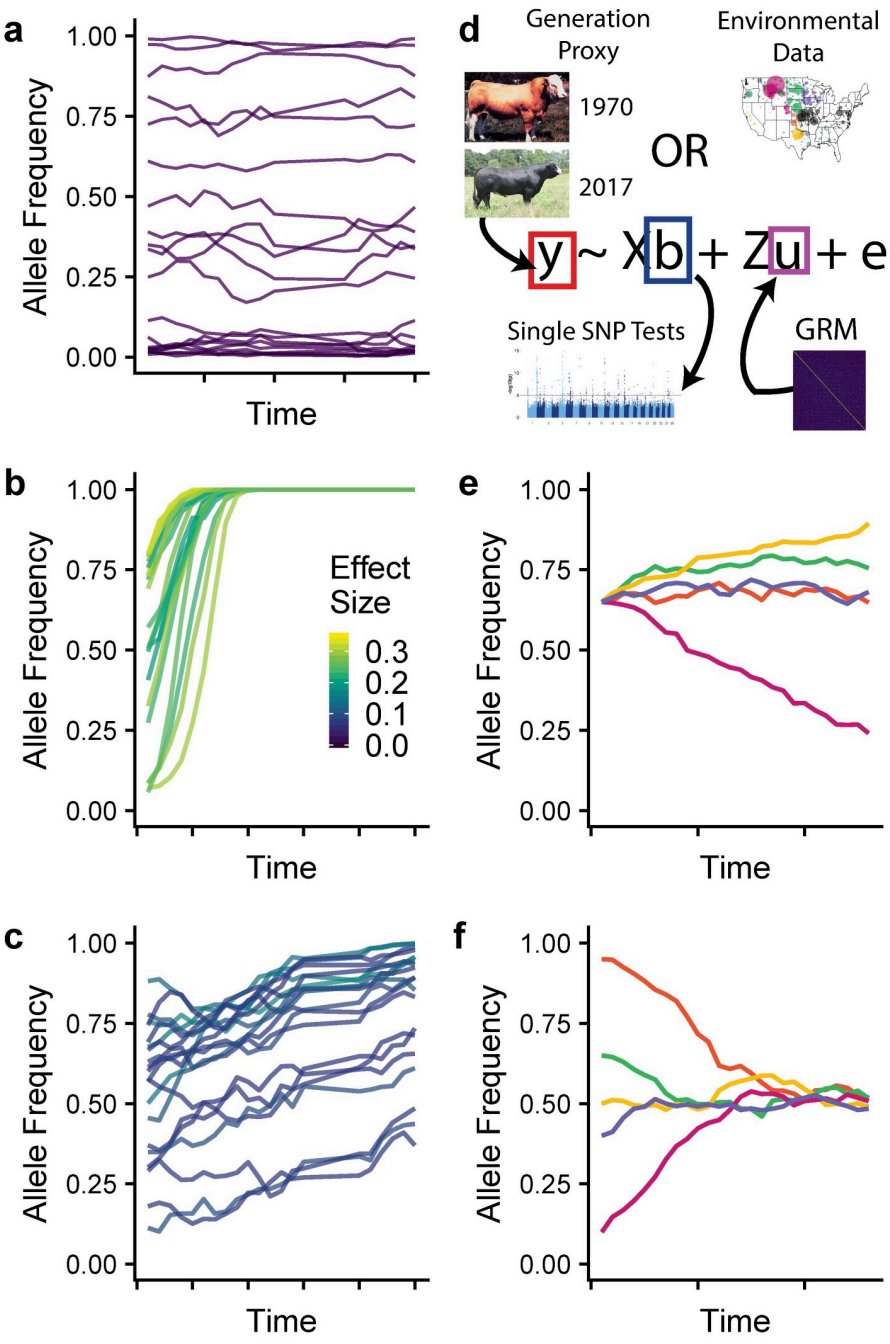

Fig 1. Simulated allele frequency trajectories and model overview. (a-c) Allele frequency trajectories for 20 SNPs colored by relative effect sizes from stochastic selection simulations. (a) Effect size = 0, representing stochastic changes in allele frequency due to genetic drift. (b) Large-effect alleles rapidly becoming fixed in the population representing selective sweeps. (c) Moderate-to-small effect size SNPs changing in frequency slowly over time, representing polygenic selection. (d) An overview of the linear mixed model approach used for Generation Proxy Selection Mapping and environmental GWAS. (e-f) A single SNP under ecoregion-specific selection. Different colors represent the trajectory of a given SNP in one of five different ecoregions. Ecoregion-specific selection can lead to allele frequencies that (e) diverge from or (f) converge to the population mean. Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).

https://doi.org/10.1371/journal.pgen.1009652.g001

relationship matrix (GRM) (**Fig 1D**) [20]. When pedigrees are missing, are missing large amounts of data, or have complex, overlapping generations, a proxy for generation can be used such as variety release date or birth date. Cattle producers are selecting on various combinations of growth, maternal, and carcass traits, but the genomic changes that result from this selection are not well-understood. Numerous genome-wide association studies have been undertaken on individual traits, but these say nothing about the underlying genomic changes that populations are experiencing due to selection. The GPSM method identifies the allele frequency changes underlying selection in a trait-agnostic manner, allowing us to observe the impacts of selection decisions in real time and understand how strong selection alters the genome over very short timescales.

From domestication to the present, humans have used phenotypic selection to change cattle. Since the 1980s and since the 2010s, genetic and genomic predictions, respectively, have been available to U.S. beef producers. However, even with these advanced tools, most beef producers still rely, at least partially, on phenotypic selection [21]. Prior to the 1980s, cattle were selected via phenotypic selection and there was very little movement of animals (i.e. gene flow) between regions. This strong, artificial phenotypic selection allows unintended selection on naturally-occurring abiotic and biotic stressor traits, akin to natural selection. Further, phenotypic selection could act on loci with genotype-by-environment effects (which BLUP breeding value-based selection would not), thus creating local adaptation. As an example, phenotypic selection could select for animals with better innate immune systems as they will grow faster and look more vigorous.

Though domesticated, beef cattle are exposed to a broad spectrum of unique environments and local selection pressures, as compared to other more intensely managed livestock populations. This suggests that local adaptation and genotype-by-environment interactions play important roles in the expression of complex traits. Local adaptation and genotype-by-environment (GxE) interactions are known to exist in closely related cattle populations [22,23]. Previous work has identified the presence of extensive GxE in beef cattle populations [24–28], but limited work exploring the genomic basis of local adaptation has occurred [29]. Further, we anticipate that the increased use of artificial insemination, responsible for dramatic increases in production efficiency, may be eroding environmentally adaptive allele frequency differences in populations. Understanding genetic interactions with the environment, and their presence in cattle populations will become increasingly important in the face of changing climates.

To identify genomic regions potentially contributing to local adaptation, we used continuous environmental variables as quantitative phenotypes or discrete ecoregions as case-control phenotypes in a linear mixed model framework. We refer to these approaches as "environmental genome-wide association studies", or envGWAS. Using a genomic relationship matrix in a LMM allows us to control the high levels of relatedness between spatially close individuals, and more confidently identify real environmentally-associated alleles. This method builds on the theory of the Bayenv approach from Coop et al. (2010) [30,31] that uses allele frequency correlations along environmental gradients to identify potential local adaptation.

Herein, we use two methods (**Fig 1**), the first for detecting complex polygenic selection (Generation Proxy Selection Mapping, GPSM), and the second for identifying local adaptation (environmental Genome-Wide Association Studies, envGWAS). When applied to three US beef cattle populations, each with ~15,000 genotyped individuals, we identified numerous genomic regions harboring directional or environmentally associated mutations. Further, using a meta-analysis approach, we identified loci with ecoregion-specific allele frequency changes (**Fig 1E and 1F**), largely due to the erosion of local adaptation caused by gene flow among ecoregions from the use of artificial insemination sires. This study is the first step in assisting beef cattle producers to identify locally adapted individuals, which will reduce the

industry's environmental footprint by increasing efficiency and resilience to stressors. Further, this repurposing of commercially-generated genomic data provides us unprecedented power to gain insight into the biology of polygenic selection and adaptation in mammalian species.

Results

Simulations

To identify the robustness of GPSM to distinguish selection from drift, we performed two major sets of simulations, stochastic and gene drop. First, we performed a set of stochastic simulations to demonstrate how selection in the context of different effective population sizes, selection intensities, generational sampling, and genomic architectures produce GPSM signal.

Stochastic simulations under multiple selection intensities, time periods, and trait architectures showed consistently that GPSM is able to map polygenic selection (**S1 Table**). Across all simulated scenarios and architectures, we identified an average of 38.5 selected loci (min 5.2, max 64.1) that reached genome-wide significance ($q < 0.1$) in GPSM. Depending on the genomic architecture of the simulated trait under selection, this represented between 0.57% and 32.54% of possible true positives (median = 9.94%). In most cases, we observed that significant hits were not the largest effect simulated QTL, but the loci that underwent the greatest allele frequency shifts over the course of genotype sampling. Usually, the largest effect QTL were fixed in the population during burn-in simulations, making their detection by GPSM impossible. Simulations suggested that GPSM effectively distinguished allele frequency changes due to selection from those associated with drift. Across all 36 scenarios of random selection, with drift acting as the only force by which alleles could change in frequency, we detected an average of 0.15 GPSM false positive SNPs (q-value < 0.1) per replicate. These rare false positives were not driven by changes in any single component of the simulations. The average number of false positive SNPs detected per replicate ranged from 0.0 to 0.45 across all scenarios, accounting for, at most, 1 false positive GPSM SNP per 100,000 tests.

Genotype sampling also significantly impacted the number of true and false positives detected by GPSM across scenarios. When genotypes were evenly sampled across generations, GPSM detected, on average, 18.05 more true positives (paired t-test $p < 2 \times 10^{-16}$) and 0.90 less false positives (paired t-test $p < 2 \times 10^{-16}$) compared with heavier sampling of recently born individuals. That said, across all scenarios, the uneven sampling scheme that more closely resembled our real datasets still detected over 20 selected loci on average across all simulations (min 1.9, max 36.4)

The proportion variation explained (PVE, **S1 Table**) of birth date across simulation scenarios reflects the number of generations and the number of crosses in the simulations across both selection and random scenarios ($p < 2e\text{-}16$). For random scenarios, the number of segregating QTL ($p = 9.94e\text{-}06$) was also associated with PVE. Importantly, for selection scenarios, the proportion of true positives ($p = 2.87e\text{-}07$) and QTL distribution ($p = 2.91e\text{-}13$) were also significantly associated with the PVE.

Gene dropping simulations using the Red Angus pedigree generated an average of 0.4 (sd = 0.52) significant GPSM loci ($q < 0.1$) per 200K markers tested, equating to 1–2 in our real genotype data (**Table A in S1 Text**). Thus, pedigree structure is not responsible for the significant SNPs that we detected in real datasets.

Detecting ongoing polygenic selection with Generation Proxy Selection Mapping (GPSM)

We used continuous birth date and high-density SNP genotypes for large samples of animals from three large US beef cattle populations; Red Angus (RAN; n = 15,295), Simmental (SIM;

n = 15,350), and Gelbvieh (GEL; n = 12,031) to map loci responding to selection (**Table B in S1 Text, S1 Fig**). The LMM estimated that the proportion of variance in individuals' birth dates explained by the additive genetic effects of SNPs was large [Proportion of Variance Explained (PVE) = 0.520, 0.588, and 0.459 in RAN, SIM, and GEL, respectively], indicating that the demographic histories of the populations and sampling strategy across the breeds were similar. We removed the link between generation proxy and genotype by randomly permuting the animals' birth date and on reanalysis of the permuted data we observed PVE to decrease to zero (**Table C in S1 Text**).

The GPSM analyses for these three populations identified 268, 548, and 763 statistically significant SNPs (q-value < 0.1), representing at least 52, 85, and 92 genomic loci associated with birth date in RAN, SIM, and GEL, respectively (**Fig 2A–2F, Tables 1 and S2**). Additionally, we found that despite birth date being a significantly skewed dependent variable (RAN skewness = 4.48, SIM skewness = 2.66, GEL skewness = 3.64), the GPSM p-values were well calibrated (**Fig 2G–2I**). Despite the tendency for genome-wide association studies (GWAS) to be biased in its detection of moderate frequency variants [32], we identify significant associations across the minor allele frequency range in our GPSM simulations and analyses (**Fig 2J–2L**). This suggests GPSM can differentiate drift from selection across the allele frequency spectrum. Rapid shifts in allele frequency create highly significant GPSM signals. For example, *rs1762920* on chromosome 28 has undergone large, recent changes in allele frequency in all three populations (**Fig 2G**), which in turn creates highly significant q-values (2.810×10^{-27}, 2.323×10^{-150}, 2.787×10^{-265} in RAN, SIM, and GEL, respectively). The allele frequency changes observed for this locus are extremely large compared to other significant regions, most of which have only small to moderate changes in allele frequency over the last ~10 generations. When we regressed allele frequency (0, 0.5, or 1.0 representing *AA*, *AB*, and *BB* genotypes per individual) on birth date, the average allele frequency changes per generation (ΔAF) for significant GPSM associations were 0.017, 0.024, and 0.022 for RAN, SIM, and GEL, respectively (**Table D in S1 Text**). In the analyses of each dataset, GPSM identified significant SNPs with $\Delta AF < 1.1\times10^{-4}$. The generally small allele frequency changes detected by GPSM are consistent with the magnitude of allele frequency changes expected for selection on traits with polygenic architectures [3].

We found that significant GPSM loci go largely undetected when using other selection mapping methods. In these datasets, the genomic regions identified by both TreeSelect and allele frequency trajectory (wfABC) methods for detecting selection were almost entirely different from those identified by GPSM. Further, the estimated selection coefficients (s) from wfABC [33] were lowly correlated with GPSM effect size estimates using both the full ($r = 0.002$) and relationship-pruned ($r = 0.003$) datasets. The slight increase in correlation is likely due to a violation of wfABC's assumption of random mating, because the correlation between estimated s between the full and relationship-pruned datasets was only 0.432. When we restricted our comparison to significant GPSM SNPs, the correlation between wfABC-estimated s and GPSM betas was higher ($r = 0.0675$), but still quite low. Interestingly the correlation between estimated selection coefficients and significant GPSM p-values was higher than with the GPSM effect size estimate ($r = 0.126$).

TreeSelect and GPSM test statistics were also lowly correlated. A three population TreeSelect analysis that treated each breed as a distinct population to detect between-breed selection showed low correlations between test statistics ($r = 0.009$, 0.002, 0.002 in Red Angus, Simmental, and Gelbvieh, respectively). The between-breed allele frequency differences detected by TreeSelect did not overlap at all with GPSM signatures. When used to detect selection within a breed using discrete groupings based on animal birth date, test statistic correlations were also low ($r = 0.007$, 0.018, and 0.005 for youngest, middle, and oldest groups of animals in Red

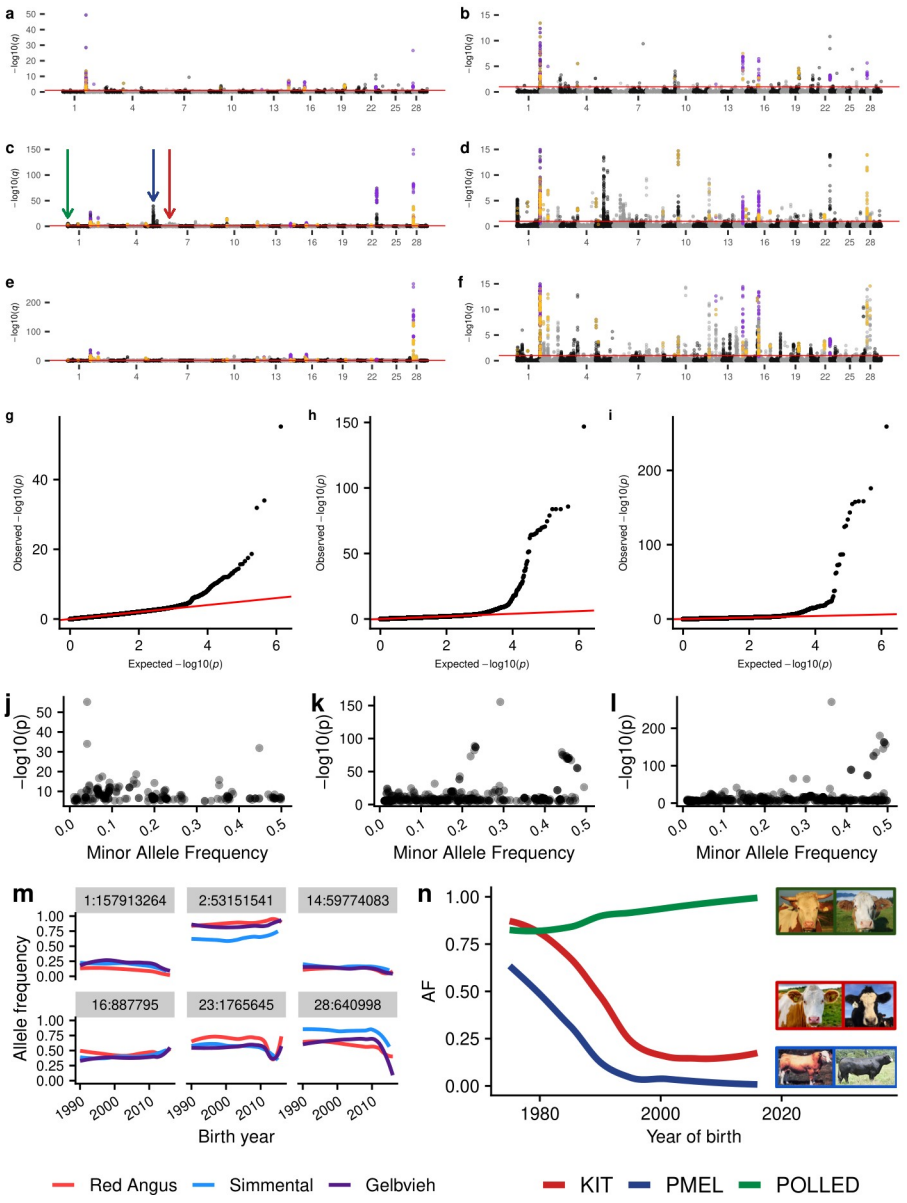

Fig 2. Generation Proxy Selection Mapping identifies signals of polygenic selection in three major U.S. cattle populations. Full and truncated ($-\log_{10}(q) < 15$) Manhattan plots for GPSM analysis of Red Angus (a & b), Simmental (c & d), and Gelbvieh (e & f). Purple points indicate SNPs significant in all three population-specific GPSM analyses and orange points indicate SNPs significant in two. Red lines in a-f indicate a significance cutoff of $q < 0.1$. Quantile-quantile plots of $-\log_{10}(p)$ values in GPSM analysis of (g) Red Angus, (h) Simmental, and (i) Gelbvieh populations. Minor allele frequency plotted versus -log10(p) values for significant SNPs in (j) Red Angus, (k) Simmental, and (l) Gelbvieh populations. (m) Smoothed allele frequency histories for the six most significant loci identified as being under selection in all three datasets. (n) Allele frequency histories for three known Mendelian loci that control differences in visual appearance between introduced European and modern US Simmental cattle. Arrows of the same color are used to distinguish the genomic locations of these loci in (c).

https://doi.org/10.1371/journal.pgen.1009652.g002

Angus). Despite generally low correlations between test statistics, TreeSelect with arbitrary population groupings detected some of the same loci that were identified by GPSM. In Red Angus, the within-breed TreeSelect analysis identified selection at six loci that were also

Table 1. Number of significant birth date-associated SNPs in each population at various significance thresholds.

Population	n individuals	q-value < 0.1	q-value < 0.05	Bonferroni (0.05/nSNPs)	Suggestive ($p < 10^{-5}$)
Red Angus	15,295	268	219	93	212
Simmental	15,350	548	466	201	397
Gelbvieh	12,031	763	634	267	517

https://doi.org/10.1371/journal.pgen.1009652.t001

significant in GPSM, including the two most significant loci on chromosome 23 near *KHDRBS2* and on chromosome 28 near *RHOU* (S2 Fig).

We performed a genomic restricted maximum likelihood (REML) analysis to identify how much of the variation in birth date was explained by various classes of GPSM SNPs. We built three GRMs using different SNP sets: One set with GPSM genome-wide significant SNPs ($q < 0.1$), the second with an equivalent number of the next most suggestive GPSM SNPs outside of significant loci (> 1 Mb from a $q < 0.1$ significant SNP), and the third with an equivalent number of random, moderate minor allele frequency (MAF > 0.15) SNPs not in the first two variant classes, intended to represent loci randomly drifting in the population. For each population, we observed that nearly all of the variation in birth date was explained by the significant and suggestive GRMs. While genome-wide significant loci explain the majority of genetic variance associated with birth date, an equivalent number of suggestive, but not significant SNPs resulted in only slightly smaller PVEs (Table 2). We suspect that these SNPs are undergoing directional allele frequency changes too small to detect at genome-wide significance, even in this highly-powered dataset. Since GPSM continues to gain power with additional samples, we suspect that future sample size increases will detect more of these signatures of polygenic selection at a genome-wide significance level. Regardless of the number of SNPs used in the drift GRM, the variance associated with drift was consistently minimal (Table 2).

As proof-of-concept, GPSM identified known targets of selection. In Simmental, we identified significant associations at three known Mendelian loci that explain the major differences in appearance between early imported European Simmental and modern US Simmental (Fig 2H). These loci: *POLLED* (absence of horns [34]), *ERBB3/PMEL* (European Simmental cream color [35]), and *KIT* (piebald coat coloration [36]) have not appreciably changed in allele frequency since 1995, making their GPSM signature significant, but less so than other loci actively changing in frequency.

In addition to these three known Mendelian loci, we detected numerous novel targets of selection within and across populations. While the majority of the genomic regions detected as being under selection were population-specific (79.8%, 79.8%, and 77.2% of the significant regions in RAN, SIM, and GEL, respectively), we identified seven loci that are under selection in all three populations, and fifteen more under selection in two (**Table E in** S1 Text). While GPSM is able to detect Mendelian selection, the overwhelming majority of signatures

Table 2. Variation in birth date explained by three classes of SNPs. The PVE estimates (standard error in parentheses) from a genomic restricted maximum likelihood (GREML) variance component analysis of birth date using three GRMs created from: 1) genome-wide significant SNPs ($q < 0.1$), 2) an equivalent number of the next most significant SNPs outside of genome-wide significant associated regions, and 3) an equivalent number of randomly sampled SNPs from genomic regions that did not harbor genome-wide significant associations.

Population	Genome-wide significant SNPs*	Suggestive significant SNPs*	Other SNPs*	Total
Red Angus	0.170 (0.021)	0.160 (0.016)	0.030 (0.004)	0.360 (0.020)
Simmental	0.223 (0.017)	0.199 (0.014)	0.046 (0.004)	0.468 (0.015)
Gelbvieh	0.232 (0.018)	0.175 (0.012)	0.021 (0.003)	0.428 (0.016)

*Contained 268, 548, and 763 SNPs for Red Angus, Simmental, and Gelbvieh, respectively

https://doi.org/10.1371/journal.pgen.1009652.t002

identified represent selection on complex, quantitative traits. Of the regions identified in multiple populations, many possessed positional candidate genes with production-related functions in cattle (*DACH1*-Growth [37,38], *LRP12*-Growth [39], *MYBPH*-Muscle Growth [40], *RHOU*-Carcass Weight [41], *BIRC5*-Feed Intake [42]). However, GPSM did not identify any of the well-established large-effect growth loci (e.g., chromosome 14 locus containing *PLAG1*, chromosome 6 locus containing *LCORL*). Growth phenotypes (e.g., birth, weaning, and yearling weights) are known to be under strong selection in all three populations [43], but antagonistic pleiotropic effects such as increased calving difficulty prevent directional selection from changing frequencies at these large-effect loci. Many of the selection signatures identified in at least two of the populations have no known functions or phenotype associations in cattle, highlighting the ability of GPSM to identify novel, important loci under polygenic selection, agnostic of phenotype.

Biological processes and pathways enriched in genes located proximal to GPSM SNP associations point to selection on drivers of production efficiency and on population-specific characteristics (**S3 Table**). In each population, we identified numerous biological processes involved in cell cycle control, which are directly involved in determining muscle growth rate [44], as being under selection. In Red Angus and Gelbvieh we identified multiple cancer pathways as being under selection. This likely represents further evidence of selection on cell cycle regulation and growth rather than on any cancer related phenotypes [45]. Red Angus cattle are known to be highly fertile with exceptional maternal characteristics [46]. We identified the "ovarian steroidogenesis" pathway as being under selection, a known contributor to cow fertility [47]. We also identify numerous other processes involved in the production and metabolism of hormones. Hormone metabolism is a central regulator of growth in cattle [48], but could also represent selection for increased female fertility in Red Angus. Further, Tissue Set Enrichment Analyses (TSEA) of Red Angus GPSM candidate genes showed suggestive expression differences ($p < 0.1$) in multiple human reproductive tissues (**S4 and S5 Tables**). Enrichments in these tissues did not exist in TSEA of Simmental or Gelbvieh GPSM gene sets, suggesting explicit within-population selection on fertility. Gelbvieh cattle are known for their rapid growth rate and carcass yield. Selection on these phenotypes likely drives the identification of the six biological processes identified which relate to muscle development and function in the Gelbvieh GPSM gene set. Consequently, this gene set is significantly enriched for expression in human skeletal muscle (**S4 and S5 Tables**), an enrichment unique to Gelbvieh. A complete list of genomic regions under population-specific selection and their associated candidate genes is in **S2 Table**.

Detecting environmental associations using envGWAS

Using an equivalent form of model to GPSM, but with continuous environmental variables (30 year normals for temperature, precipitation, and elevation) or statistically-derived discrete ecoregions as the dependent variable (rather than birth date in GPSM) allows us to identify environmentally-associated loci that have been subjected to artificial and, perhaps in this context more importantly, natural selection [49]. We refer to this method as environmental GWAS (envGWAS). envGWAS extends the theory of the Bayenv approach of Coop et al. (2010) which searches for allele frequency correlations along environmental gradients to identify potentially adaptive loci [30]. Similar approaches have been applied to plant datasets [50,51], but we extend envGWAS to panmictic, biobank-sized mammalian populations. Unlike many genome-environment association analyses which only used linear models [51,52], our large dataset and the use of multivariate models provides power to identify association while importantly controlling for geographic dependence between samples using a

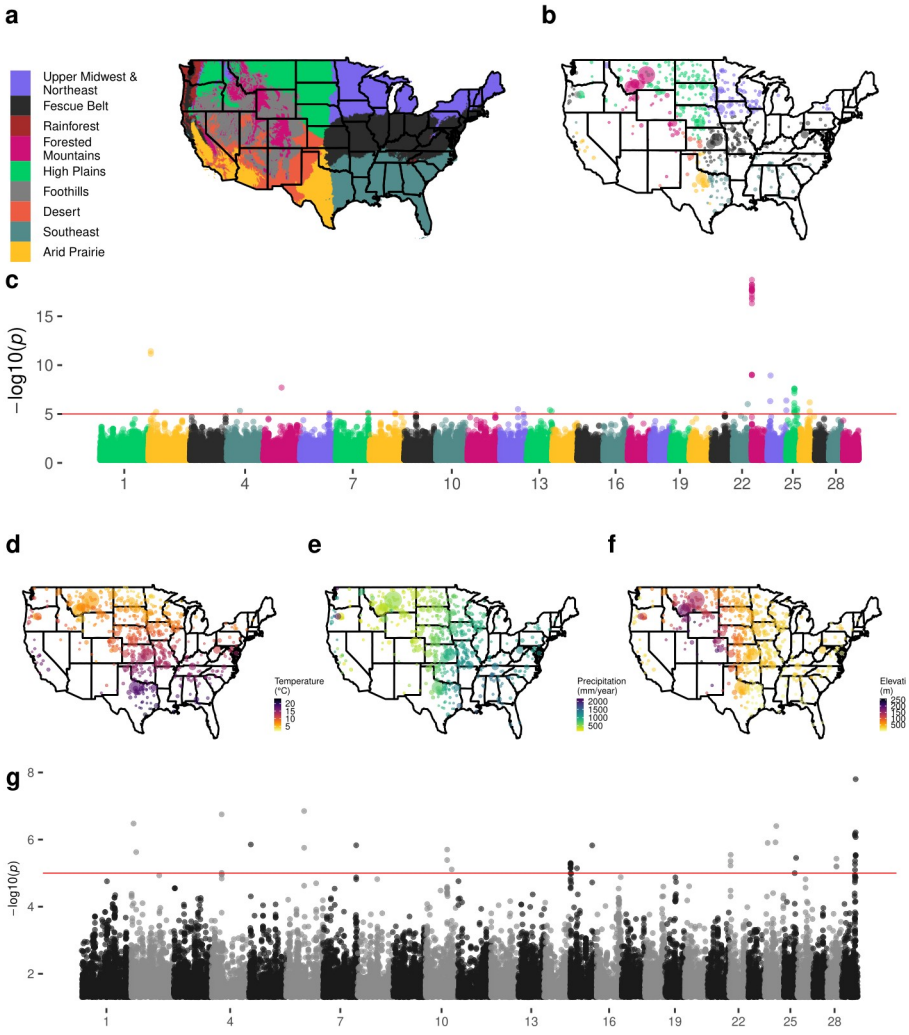

Fig 3. Manhattan plots for discrete and continuous envGWAS in Red Angus cattle. (a) Nine continental US ecoregions defined by *K*-means clustering of 30-year normal temperatures, precipitations, and elevations. (b) Locations of sampled Red Angus animals colored by breeder's ecoregion and sized by the number of animals at that location. (c) Multivariate discrete envGWAS (case-control for six regions with > 600 animals). Locations of sampled Red Angus animals colored by (d) 30-year normal temperature, (e) 30-year normal precipitation, and (f) elevation. (g) Multivariate continuous envGWAS with temperature, precipitation, and elevation as dependent variables. For all Manhattan plots the red line indicates the empirically-derived p-value significance threshold from permutation analysis ($p < 1 \times 10^{-5}$). Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).

https://doi.org/10.1371/journal.pgen.1009652.g003

genomic relationship matrix (S3 and S7 **Figs**). We used *K*-means clustering with 30-year normal values for temperature, precipitation, and elevation to partition the United States into 9 discrete ecoregions (Fig 3A). These ecoregions are largely consistent with those represented in previously-published maps from the environmetrics and atmospheric science literature [53], and reflect well-known differences in cattle production environments. The resulting ecoregions capture not only combinations of climate and environmental variables, but associated differences in forage type, local pathogens, and ecoregion-wide management differences to which animals are exposed. Thus, using these ecoregions as case-control phenotypes in envGWAS allowed us to detect more complex environmental associations. The three studied

populations are not universally present in all ecoregions (**Figs 3B, S4B and S5B, Table F in S1 Text**). Since the development of these US populations in the late 1960s and early 1970s, registered seedstock animals from these populations have a small footprint in desert regions with extreme temperatures and low rainfall.

Although environmental variables and ecoregions are not inherited, the estimated PVE measures the extent to which genome-wide genotypes change in frequency across the environments in which the animals were born and lived. The proportion of variance explained by SNPs ranged from 0.586 to 0.691 for temperature, 0.526 to 0.677 for precipitation, and 0.585 to 0.644 for elevation (**Table G in S1 Text**). In Red Angus, PVE for ecoregion membership ranged from 0.463 for the Arid Prairie to 0.673 for the Fescue Belt (**Table H in S1 Text**). We observe similar environmental PVE in both Simmental and Gelbvieh datasets. These measures suggest that genetic associations exist along both continuous environmental gradients and within discrete ecoregions. Permutation tests that shuffled environmental dependent variables, removing the relationship between the environment and the animal's genotype, resulted in all PVEs being reduced to ~ 0, strongly suggesting that the detected associations between genotype and environment were not spurious. An additional permutation test that permuted animals' zip codes, such that all animals from a given zip code were assigned the same "new" zip code from a potentially different ecoregion provided similar results, indicating that bias due to sampling at certain zip codes was not producing envGWAS signals. From 10 rounds of permutation, there were no SNP associations with p-values $< 1 \times 10^{-5}$. Consequently, we used this empirically-derived p-value threshold to determine SNP significance in all of the envGWAS analyses. Gene drop simulations suggested that a small portion of the identified associations are likely due to pedigree structure or founder effects (average of 1.36 false positive envGWAS loci per 200,000 tests). However, in this data, the pedigree structure reflects selection decisions of farmers and ranchers that are not beyond the influence of performance differences relative to environmental differences.

Discrete ecoregion envGWAS

In Red Angus, we identified 54 variants defining 18 genomic loci significantly associated with membership of an ecoregion in the discrete multivariate envGWAS analysis (**Fig 3C**). Despite locus-specific signal, principal component analysis (PCA) does not suggest that ecoregion-driven population structure exists in any of the populations (**S6 Fig**). Of these loci, only two overlapped with loci identified in the continuous envGWAS analyses, suggesting that using alternative definitions of environment in envGWAS may detect different sources of adaptation. Of the 18 significant loci, 17 were within or near (< 100 kb) positional candidate genes (**Table I in S1 Text, S6 Table**), many of which have potentially adaptive functions. For example, envGWAS identified SNPs immediately (22.13 kb) upstream of *CUX1* (Cut Like Homeobox 1) gene on chromosome 25. *CUX1* controls hair coat phenotypes in mice [54], and alleles within *CUX1* can be used to differentiate between breeds of goats raised for meat versus those raised for fiber [55]. The role of *CUX1* in hair coat phenotypes makes it a strong adaptive candidate in environments where animals are exposed to heat, cold, or toxic ergot alkaloids from fescue stress [56].

In Simmental, we identified 11 loci tagged by 39 variants significantly associated with membership of an ecoregion in the multivariate envGWAS analysis (**S4 Fig**). In Gelbvieh, 66 variants identified 33 local adaptation loci (**S5 Fig**). In the analyses of all three datasets, we identified a common local adaptation signature on chromosome 23 (peak SNP *rs1023574*). Multivariate analyses in all three populations identified alleles at this SNP to be significantly associated with one or more ecoregions ($q = 1.24 \times 10^{-13}$, 3.15×10^{-12}, 4.82×10^{-5} in RAN, SIM, and GEL, respectively). In all three datasets, we identified *rs1023574* as a univariate

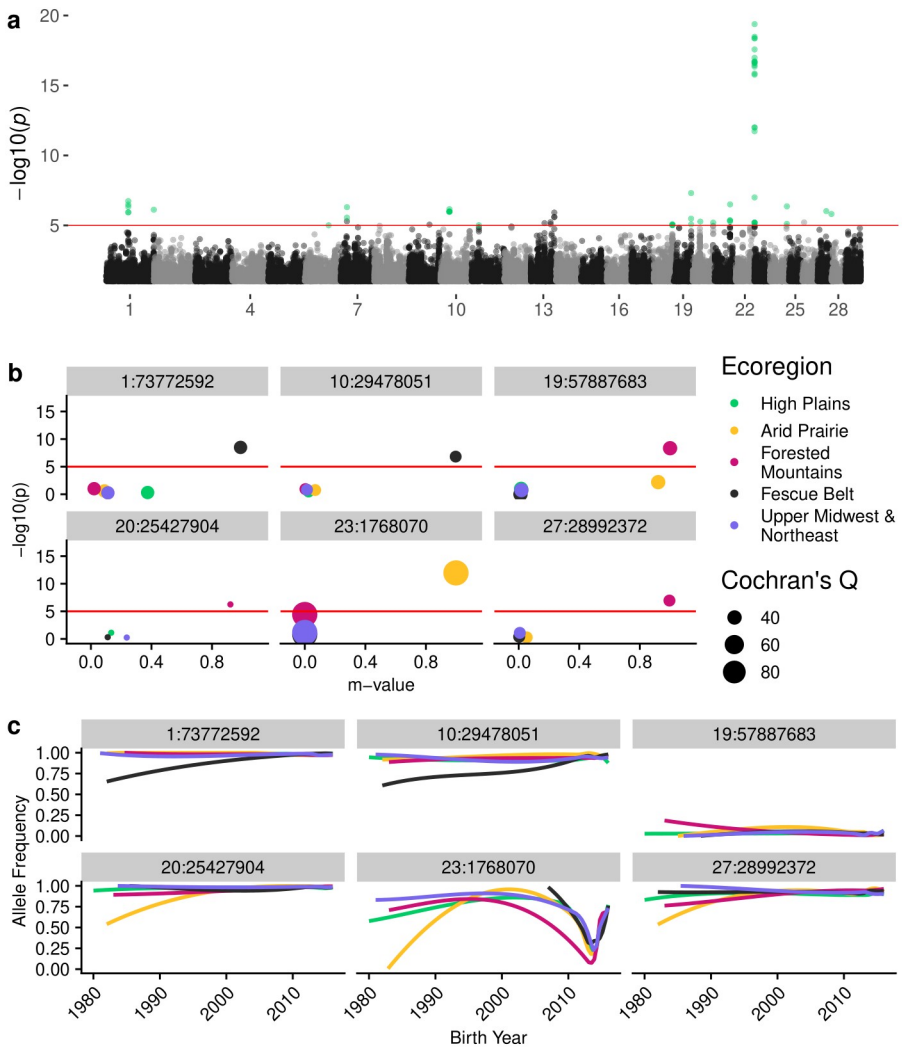

Fig 4. Meta-analysis of within-ecoregion GPSM for Red Angus cattle. (a) Manhattan plot of per-variant Cochran's Q p-values. Points colored green had significant Cochran's Q ($p < 1 \times 10^{-5}$) and were significant in at least one within-region GPSM analysis ($p < 1 \times 10^{-5}$). (b) Ecoregion effect plots for lead SNPs from six loci from (a). Points are colored by ecoregion and are sized based on Cochran's Q value. (c) Ecoregion-specific allele frequency histories for SNPs from (b), colored by ecoregion.

https://doi.org/10.1371/journal.pgen.1009652.g004

envGWAS association with membership of the Forested Mountains ecoregion. However, the most significant univariate association in Red Angus was with the Arid Prairie region which was excluded from both the Simmental and Gelbvieh analyses due to low within-region sample size. In the multivariate analysis for Red Angus, the associated locus spanned 18 SNPs from (1,708,914 to 1,780,836 bp) and contained the pseudogene *LOC782044*. The nearest annotated gene, *KHDRBS2* (KH RNA Binding Domain Containing, Signal Transduction Associated 2) has previously been identified by other adaptation studies in cattle, sheep, and pigs [57–59]. This variant was not significantly associated with any continuous environmental variable in Red Angus. However, *rs1023574* was significantly associated with temperature, elevation, and humidity variables in Simmental. The *KHDRBS2* locus was preferentially introgressed between *Bos taurus* and domestic yak [60]. Further, this locus shows an abnormal allele frequency trajectory (**Fig 4C**), indicating that it may be a target of balancing selection.

Continuous environmental variable envGWAS

Using continuous temperature, precipitation, and elevation data as quantitative dependent variables in a multivariate envGWAS analysis of Red Angus animals, we identified 46 significantly associated SNPs (**Fig 3G**). These SNPs tag 17 loci, many of which are within 100 kb of positional candidate genes. Univariate envGWAS identified 23, 17, and 10 variants associated with temperature, precipitation, and elevation, respectively (**S7 Fig**). The most significant multivariate association in Red Angus is located on chromosome 29 within *BBS1* (Bardet-Biedl syndrome 1), which is involved in energy homeostasis [61]. *BBS1* mutant knock-in mice show irregularities in photoreceptors and olfactory sensory cilia [62] functions that are likely important to an individual's ability to sense its local environment. This region was not significantly associated in any of the univariate analyses of environmental variables, and was not identified in any of the discrete ecoregion envGWAS. Of the other positional candidate genes identified in this Red Angus analysis, 9 have previously been implicated in adaptive functions in humans or cattle (**Table J in S1 Text**). For example, SNPs near *GRIA4* were implicated in body temperature maintenance in cold stressed Siberian cattle [63]. Significant SNPs and their corresponding candidate genes for all three datasets are reported in **S6 Table**.

While we found minimal candidate gene overlap between populations, we identified multiple shared biological pathways and processes (**S7 Table**) derived from lists of envGWAS positional candidate genes. Pathways in common between populations were driven by largely different gene sets. Across all populations, we identified the "axon guidance" pathway, and numerous other gene ontology (GO) terms related to axon development and guidance as enriched with environmentally-associated loci. Ai et al. (2015) suggested that axon development and migration in the central nervous system is essential for the maintenance of homeostatic temperatures by modulating heat loss or production [64]. In addition to axonal development, a host of other neural signaling pathways were identified in multiple populations. A genome-wide association study for gene-by-environment interactions with production traits in Simmental cattle by Braz et al. (2020) identified a similar set of enriched pathways [28]. These common neural signaling pathways identified by envGWAS are regulators of stress response, temperature homeostasis, and vasoconstriction [65]. We identified other shared pathways involved in the control of vasodilation and vasoconstriction (relaxin signaling, renin secretion, and insulin secretion). Vasodilation and vasoconstriction are essential to physiological temperature control in cattle and other species [66]. The ability to mount a physiological response to temperature stress has a direct impact on cattle performance, making vasodilation a prime candidate for environment-specific selection. Further, vasodilation and vasoconstriction likely also represent adaptation to hypoxic, high elevation environments. Pathways and processes identified by envGWAS signals are reported in **S7 Table.**

To further explore the biology underlying adaptive signatures, we performed Tissue Set Enrichment Analysis of our envGWAS candidate gene lists. These analyses using expression data from humans and distantly related worms (*C. elegans*) both identified brain and nerve tissues as the lone tissues where envGWAS candidate genes show significantly enriched expression (**S8**–**S11 Tables**). Tissue-specific expression in the brain further supports our observed enrichment of local adaptation pathways involved in neural signaling and development.

Identifying loci undergoing region-specific selection with GPSM ecoregion meta-analysis

envGWAS detects allelic associations with continuous and discrete environmental variables, but does not address whether selection is towards increased local adaptation, or whether local adaptation is being eroded by the exchange of germplasm between ecoregions via artificial

insemination. We used the spatiotemporal stratification of genotyped animals to identify loci undergoing ecoregion-specific selection. We performed GPSM within each sufficiently genotyped ecoregion and identified variants with high effect size heterogeneity (Cochran's Q statistic) between ecoregions. Variants with significant heterogeneity across regions that were also significant in at least one within-region GPSM analysis implied ecoregion-specific allele frequency change. These changes could have been due either to selection for local adaptation (**Fig 1E**), or locally different allele frequencies moving towards the population mean (**Fig 1F**). We identified 59, 38, and 46 significant SNPs in Red Angus, Simmental, and Gelbvieh, respectively undergoing ecoregion-specific selection. These represent 15, 21, and 26 genomic loci ($>$ 1 Mb to nearest next significant SNP) (**Fig 4A**). In most cases, these variants have an effect (posterior probability of an effect: m-value $>$ 0.9) in only one or two ecoregions (**Fig 4B**). Further, nearly all represent the decay of ecoregion-specific allele frequencies towards the population mean (**Fig 4C**) as opposed to on-going directional selection for ecoregion specific beneficial adaptations (**S10**–**S12 Figs**).

Adaptive alleles at these loci are being driven in frequency towards the population mean allele frequency (**Fig 4C**), which is typically a low minor allele frequency.

Discussion

We leveraged large commercially-generated genomic datasets from three major US beef cattle populations to map polygenic selection and environmental adaptation using GWAS applications [67]. Using temporally-stratified genotype data we detected very small selection-driven changes in allele frequency throughout the genome. This is consistent with expectations of polygenic selection acting on a large number of variants with individual small effects. Which phenotypes are being selected and driving the allele frequency changes at particular loci is not definitively known. GPSM is a heuristic model, and as a result the SNP effects are not immediately intuitive to interpret in a population genetic context. That said, it allows us to identify the genomic loci responding to selection, and particularly subtle changes due to polygenic selection. GPSM is agnostic to the selected phenotypes, and identifies important loci changing in frequency due to selection without the need to measure potentially difficult or expensive phenotypes. Our GPSM model is of course subject to false-positives and other short comings of genome-wide association models. However, in simulations, GPSM effectively differentiates between selection and drift while accounting for confounding effects such as uneven generation sampling, population structure, relatedness, and inbreeding. Population branch statistics require arbitrary definitions of subpopulations in panmictic populations, and allele frequency trajectories assume random mating, which is violated in our data. The ability of GPSM to account for relatedness and inbreeding likely accounts for the disagreement between GPSM and these other methods. Simulations suggest that GPSM has greater power to detect selection when genotyped individuals are uniformly sampled over time. When genotypes originate from individuals in only the most recent generations the power to detect "old" selection is lessened, and GPSM signatures are enriched for very recent, ongoing selection. As a result, we expect that applying a GPSM-like approach to experimental evolution studies would generate even clearer associations than observed in this study. With the availability of large samples our analytical framework can help solve the long-standing population genetics problem of identifying the loci subjected to polygenic selection.

Future studies exploring the effects of selection from the context of complex trait networks could explain how hundreds or thousands of selected genes act together to shape genomic diversity under directional selection. Candidate genes identified by GPSM suggest selection on pathways and processes involved in production efficiency (growth, digestion, muscle

development, and fertility). In addition to a small number of loci, for which function is known, we identify hundreds of novel signatures of ongoing selection.

The envGWAS identified 174, 125, and 130 SNPs associated with both continuous or discrete environmental factors in Red Angus, Simmental, and Gelbvieh, respectively. Using genes found near these environmentally-associated SNPs, we identified a consistent enrichment of pathways and tissues involved in neural development and signaling. These envGWAS associations emphasize the role that the nervous system likely plays in recognizing and responding to environmental stress in mammals, which will be valuable as society and agriculture cope with climate change. In addition to neural pathways, we observe significantly enriched expression of envGWAS genes in the brain tissues of humans, mice, and worms. Other pathways associated with environmental adaptation reveal the importance of mechanisms involved in regulating vasoconstriction and vasodilation, both of which are essential for responses to heat, cold, altitude, and toxic fescue stressors in cattle.

The statistical power and wide geographical distribution of the cattle comprising these data highlights that the utilized approaches can be leveraged to understand the genomic basis of adaptation in many other studies and species. While for GSPM there is a clear connection between the model and directional selection, for envGWAS the model is only identifying associations between the environment and SNP genotypes. These associations could be caused by a multitude of factors, one of which is local adaptation. The small allele frequency differences identified by envGWAS are consistent with a polygenic model of local adaptation, likely driven by small changes in gene expression [68]. Further, envGWAS identifies candidate genes (e.g. *KHDRBS2*) and pathways previously implicated as domestication-related [60]. This suggests that these genes may be under natural or balancing selection to cope with environmental stress, and not specifically part of the domestication process. Further, because different genes in the same pathways were detected in the analyses of the different populations, we hypothesize that these pathways influence local adaptation in many mammals and should be studied in other ecological systems. This knowledge will become increasingly valuable as species attempt to adjust to a changing climate.

The balance between artificial and natural selection in domesticated beef cattle is quite precarious. If natural selection is for a stressor which occurs nation-wide and is positively correlated with production traits, natural selection can effectively act in concert with artificial selection. The identification of genes involved in the immune system by the GPSM analysis reflects this interplay between artificial and natural selection; in other words, natural selection could be acting in the background in populations under artificial selection. However, if natural selection is acting at a local, ecoregion scale, then natural selection and artificial selection via additive genetic breeding values are likely to be at odds. Artificial insemination in cattle has allowed the ubiquitous use of males which have been found to be superior when progeny performance has been averaged across US environments. Due to limited gene flow and phenotypic selection which would act on loci with genotype-by-environment effects, local adaptation likely occurred prior to the 1980s. Our results suggest that environmental associations are present in cattle populations, but that the widespread use of artificial insemination resulting in gene flow and selection on breeding values, has caused US cattle populations to lose ecoregion-specific adaptive variants [69]. We identified 16, 21, and 30 loci undergoing ecoregion-specific selection in Red Angus, Simmental, and Gelbvieh, respectively. In almost every case, selection and use of artificial insemination has driven allele frequencies within an ecoregion back towards the population mean allele frequency (**Figs 1F and 4C**). In three independent datasets, we identified a single shared environmentally-associated locus near the gene *KHDRBS2*. This locus has been identified as introgressed in yak, and exhibits an irregular allele frequency trajectory which suggests that it may be subject to balancing selection [70]. Though

we identified only a single common envGWAS locus, we observed significant overlap in the pathways regulated by candidate genes within the associated loci. This reveals that adaptive networks are complex and that adaptation can be influenced by selection on functional variants within combinations of genes from these networks. As we work to breed more environmentally-adapted cattle, there will be a need for selection tools that incorporate genotype-by-environment interactions to ensure that cattle become increasingly locally adapted.

We further demonstrate that large commercially-generated genomic datasets from domesticated populations can be leveraged to detect polygenic selection [17] and local adaptation signatures [50]. The identification of adaptive loci can assist in selecting and breeding better adapted cattle for a changing climate. Further, both our statistical approaches and biological findings can serve as a blueprint for studying complex selection and adaptation in other agricultural or wild species. Our results suggest that neural signaling and development are essential components of mammalian adaptation, meriting further functional genomic study. Finally, we observe that local adaptation is declining in cattle populations, which will need to be preserved to sustainably produce protein in changing climates.

Materials and methods

RMarkdown files, snakemake files, and R scripts used in the manuscript are available at https://github.com/troyrowan/gpsm_envgwas.

Genotype data

SNP assays for three populations of genotyped *Bos taurus* beef cattle ranging in density from ~25K SNPs to ~770K SNPs were imputed to a common set of 830K SNPs using the large multi-breed imputation reference panel described by Rowan et al. 2019 [71]. Genomic coordinates for each SNP were from the ARS-UCD1.2 reference genome [72]. Genotype filtering for quality control was performed in PLINK (v1.9) [73], reference-based phasing was performed with Eagle (v2.4) [74], and imputation with Minimac3 (v2.0.1) [75]. Following imputation, all three datasets contained 836,118 autosomal SNP variants. All downstream analyses used only variants with minor allele frequencies > 0.01. Upon filtering, we performed a principal component analysis for each population in PLINK. This was to assess if there were discrete subpopulations within the populations and if there were patterns of structure related to ecoregions.

Generation Proxy Selection Mapping (GPSM)

To identify alleles that had changed in frequency over time, we fit a univariate genome-wide linear mixed model (LMM) using GEMMA (Version 0.98.1) [76]. Here, we used the model:

$$\begin{aligned} y &= Xg + Zu + e \\ u &\sim N(0, G\sigma_a{}^2) \\ e &\sim N(0, I\sigma_e{}^2) \end{aligned} \qquad \text{Eq1}$$

where **y** is an individual's generation proxy, in our case birth date, and **X** was an incidence matrix that related SNPs to birth dates within each individual and **g** was the estimated effect size for each SNP. An animal's age as of April 5, 2017 was used as the generation proxy in GPSM. We control for confounding population structure, relatedness, and inbreeding with a polygenic term **u** that uses a standardized genomic relationship matrix (GRM) **G** [20] and we estimated $\sigma_a{}^2$ *and* $\sigma_e{}^2$ using restricted maximum likelihood estimation (REML). Here, continuous age served as a proxy for generation number from the beginning of the pedigree. Other than the tested SNP effects, no fixed effects other than the overall mean were included in the

model. We tested each SNP for an association with continuous age. To control for multiple-testing, we converted p-values to FDR corrected q-values [77] and used a significance threshold of $q < 0.1$ to classify significant SNPs. We performed additional negative-control analyses in each dataset by permuting the date of birth associated with each animal's genotypes to ensure that the detected GPSM signals were likely to be true positives. Permutation was performed ten times for each population. To visualize the allele frequency history of loci undergoing the strongest selection, we fit a loess and simple linear regressions for date of birth and allele frequencies scored as 0, 0.5 or 1.0 within each individual using R [78]. Results were visualized using ggplot2 [79].

Simulations and gene drops

All simulations were performed in AlphaSimR [80]. Stochastic simulations were performed in 10 replicate sets using 10 sets of founder haplotypes as starting points. We generated founder haplotypes using the AlphaSimR wrapper around MaCS [81]. Using an approximation of the demographic history of cattle, we simulated 10 chromosomes with 20,000 segregating sites each for 2,000 founder individuals (1,000 males and 1,000 females). This resulted in a starting effective population size (N_e) of approximately 100, similar to estimates of U.S. beef cattle populations [17]. To test other N_e, we simulated populations with effective population sizes of 50 and 250. Based on the chosen genomic architecture, 1000, 500, or 200 purely additive QTL were randomly assigned to segregating sites. Effect sizes for simulated QTL were drawn from either a normal (mean = 0, variance = 1) or a gamma (shape = 0.42) distribution [82]. Prior to the two divergent selection regimes, we performed five generations of burn-in selection to establish LD in our populations.

After burn-in (generation 0), we performed selection of parents for the next generation in two parallel manners: randomly or using truncation selection on true breeding value. In each scenario, we held the effective population size by selecting appropriate numbers of males and females to be parents each generation. Selection intensity was altered by increasing or decreasing the number of crosses performed (1000, 2000, 4000, 8000). We also varied the number of generations of selection post-burn-in (20, 10, and 5 generations). For each scenario, we extracted 10,000 total simulated individuals for analysis in GPSM. To test the effects of uneven generation sampling that we see in real data, we performed two different strategies for sampling simulated genomes. In one case, we sample an equal number of individuals each generation. In the other, we sample more animals from the most recent generations. The number of sampled individuals is based on a negative exponential distribution that approximates the of ages observed in our real datasets (**S1 Fig**). Sampled individuals were chosen at random, and were not more or less likely to become parents in the next generation. In addition to sampling genotypes each generation, we calculated the allele frequency of simulated QTL each generation to track observed allele frequency changes over the course of selection. This process was performed in replicates of 10 for each scenario, allowing us to calculate descriptive statistics and compare GPSM's performance across scenarios.

We simulated haplotypes in MaCS for our 5,223 founder individuals. Founder haplotypes spanned 10 chromosomes, each with 20,000 segregating sites for a total of 200,000 SNPs. These founder haplotypes were then randomly dropped through the Red Angus pedigree, restricted to ancestors of genotyped individuals, in a Mendelian fashion, with recombinations occurring at a rate of one crossover per Mb.

After sampling genotypes, we created a standardized genomic relationship matrix (GRM) in GEMMA (v0.98.1) with all SNPs that had a MAF > 0.01. Using GEMMA, we fit the individuals' true generation number as the dependent variable in a genome-wide linear mixed model.

Outputs from GPSM were read, manipulated, and plotted in R using multiple tidyverse packages [83].

Finally, we fit regression models to assess the impact of various parameters on the PVE. For PVE, two regression models were analyzed. In the first, PVE from the random mating simulations was used as the dependent variable with proportion false positives, number of generations, number of crosses, number of QTL, QTL distribution, and number of segregating sites as explanatory variables. In the second, PVE from selection on true breeding value (TBV) was the dependent variable with proportion true positives, number of generations, number of crosses, number of QTL, QTL distribution, and number of segregating sites as dependent variables.

Birth date variance component analysis

To estimate the amount of variation in birth date explained by GPSM significant SNPs, we performed multi-GRM GREML analyses for birth date in GCTA (v1.92.4) [84]. We built separate GRMs using genome-wide significant markers and all remaining makers outside of significant GPSM loci (> 1 Mb from significant GPSM SNPs to control for markers physically linked to significant GPSM SNPs). To further partition the variance in birth date explained by subsets of SNPs, we performed a GREML analysis using three GRMs created with genome-wide significant ($q < 0.1$) SNPs, an equal number of the next most significant SNPs, and an equal number of randomly selected markers not present in the first two classes with minor allele frequencies > 0.15, to match the allele frequencies of significant SNPs. These three GRM were each constructed using 268, 548, and 763 SNPs for Red Angus, Simmental, and Gelbvieh, respectively.

Allele frequency time series methods for detecting selection

We used the software wfABC (Wright-Fisher Approximate Bayesian Computation) [33] to generate estimates of the selection coefficient at each locus. We partitioned individuals into generations based on the maximum generation number estimated from the Red Angus pedigree using the optiSel R package [85]. Since wfABC assumes random-mating populations of unrelated individuals, we also performed relationship pruning in PLINK, removing individuals with estimated relationship coefficients > 0.0625. The wfABC software estimated selection coefficients for each imputed SNP with MAF > 0.01, using an effective population size of 150 [17].

TreeSelect

Using TreeSelect [86], we artificially subdivided our genotyped populations to compare statistics from a population branch statistic (PBS) method with GPSM effect sizes and p-values. TreeSelect tests for allele frequency differences between discretely labeled, but closely related populations. We ran TreeSelect two ways: First, where each genotyped population (Red Angus, Simmental, and Gelbvieh) was its own branch, and then in three separate analyses where we subdivided each population into three equally sized groups based on an individual's birth date. These three age classes consisted of the oldest ⅓ of individuals, the middle ⅓ of the age distribution, and the youngest ⅓ of individuals. We compared TreeSelect chi-squared values with GPSM betas to quantify overall relationships between the statistics.

Environmental data

Thirty-year normals (1981–2010) for mean temperature ((average daily high (˚C) + average daily low (˚C)/2), precipitation (mm/year), and elevation (m above sea level) for each 4 km^2 of

the continental US were extracted from the PRISM Climate Dataset [87], and used as continuous dependent variables in envGWAS analysis. Optimal *K*-means clustering of these three variables grouped each 4 km^2 of the continental US into 9 distinct ecoregions. Using the reported breeder zip code for each individual, we linked continuous environmental variables to animals and partitioned them into discrete environmental cohorts for downstream analysis. For ecoregion assignments, latitude and longitude were rounded to the nearest 0.1 degrees. As a result, some zip codes were assigned to multiple ecoregions. Animals from these zip codes were excluded from the discrete region envGWAS but remained in analyses that used continuous measures as dependent variables.

Environmental Genome-wide Association Studies (envGWAS)

To identify loci segregating at different frequencies within discrete ecoregions or along continuous climate gradients, we used longitudinal environmental data for the zip codes attached to our study individuals as dependent variables in univariate and multivariate genome-wide LMMs implemented in GEMMA (Version 0.98.1). We fit three univariate envGWAS models that used 30-year normal temperature, precipitation, and elevation data as dependent variables. These used an identical model to **Eq 1**, but used environmental values as the dependent variable (**y**) instead of birth date. We also fit a combined multivariate model using all three environmental variables to increase power. To identify loci associated with entire climates as opposed to only continuous variables, we fit univariate and multivariate case-control envGWAS analyses using an individual's region assignment described in the "Environmental Data" section as binary phenotypes. Proportion of variation explained (PVE), phenotypic correlations, and genetic correlations were estimated for continuous environmental variables and discrete environmental regions using GEMMA's implementation of REML.

To ensure that envGWAS signals were not driven by spurious associations, we performed two separate permutation analyses. In the first, we randomly permuted the environmental variables and regions associated with an individual prior to performing each envGWAS analysis, detaching the relationship between an individual's genotype and their environment. In the second, to ensure that envGWAS signals were not driven by the over-sampling of individuals at particular zip codes, we permuted the environmental variables associated with each zip code prior to envGWAS analysis. These two types of permutation analyses were performed for each dataset and for each type of univariate and multivariate envGWAS analysis. We determined significance using a permutation-derived p-value cutoff ($p < 1 \times 10^{-5}$) [88].

GPSM meta-analyses

To identify variants undergoing ecoregion-specific allele frequency changes, we performed GPSM analyses within each region with more than 600 individuals. The SNP significance testing effects and standard errors from each of the within-region GPSM analyses were combined into a single meta-analysis for each population using METASOFT (v2.0.1) [89]. We identified loci with high heterogeneity in allele effect size, suggesting region-specific selection. An m-value indicating the posterior-probability of a locus having an effect in a particular ecoregion was calculated for each of these loci [90].

Gene set and tissue set enrichment analysis

Using the NCBI annotations for the ARS-UCD1.2 *Bos taurus* reference assembly, we located proximal candidate genes near significant SNPs from each of our analyses. We generated two candidate gene lists each from significant GPSM and envGWAS SNPs. Lists contained all annotated genes within 10 kb from significant SNPs. We consolidated significant SNPs from

all envGWAS analyses to generate a single candidate gene list for each breed. Using these candidate gene lists, we performed gene ontology (GO) and KEGG pathway enrichment analysis using Clue GO (v2.5.5) [91] implemented in Cytoscape (v3.7.2) [92]. We identified pathways and GO terms where at least two members of our candidate gene list comprised at least 1.5% of the term's total genes. We applied a Benjamini-Hochberg multiple-testing correction to reported p-values and GO terms with FDR corrected p-values < 0.1 were considered significant.

Using the above gene sets, we performed three separate Tissue Set Enrichment Analyses (TSEA) using existing databases of human, mouse, and worm gene expression data. We searched for enriched gene expression with data from the Human Protein Atlas [93] and Mouse ENCODE [94] using the Tissue Enrich tool (v1.0.7) [95]. Additionally, we performed another Tissue Set Enrichment Analysis using GTEx data [96] and a targeted Brain Tissue Set Enrichment Analysis in the pSI R package (v1.1) [97]. Finally, we used Ortholist2 [98] to identify *C. elegans* genes orthologous with members of our envGWAS and GPSM gene lists. We then queried these lists in WormBase's Tissue Enrichment Analysis tool [99,100] to identify specific tissues and neurons with enriched expression in *C. elegans*. We used each tool's respective multiple-testing correction to determine significance. We deemed an enrichment in a tissue "suggestive" when its p-value was < 0.1.

Summary data

Summary data from GPSM and envGWAS analyses are publicly available as a Zenodo repository [101].

Supporting information

S1 Text. Supplementary Information for Powerful detection of polygenic selection and environmental adaptation in US beef cattle. This file includes supplementary text, Tables A-J, and SI References.
(DOCX)

S1 Fig. Distributions of continuous birth date in sampled Red Angus, Simmental, and Gelbvieh populations. (a) Birth date histograms for complete datasets. (b) Histograms of animal birth dates born before 2000. (c) Q-Q plots of residual error from a GREML analysis of birth date in each population. Points represent individuals, and are colored by the number of years since the animal's birth date.
(TIF)

S2 Fig. TreeSelect results for Red Angus dataset. a) Single SNP -$\log_{10}$(p-values) for Red Angus branch of across-breed TreeSelect analysis. TreeSelect Manhattan plots for b) oldest ⅓, c) middle ⅓, and d) youngest ⅓ branches in within-breed analysis for the Red Angus population. Red line indicates significance at $p < 1 \times 10^{-5}$. Green points are SNPs that were significant ($q < 0.1$) in GPSM analysis of Red Angus dataset.
(TIF)

S3 Fig. Manhattan plots of discrete envGWAS in Red Angus cattle. Q-Q plots for envGWAS p-values of (a) a linear model for Forested Mountains ecoregion membership, (b) a linear mixed model for Forested Mountain ecoregion membership, and (c) a multivariate linear mixed model of ecoregion membership. Univariate discrete envGWAS for (d) Forested Mountain linear model, (e) Forested Mountains linear mixed model, (f) Southeast, (g) Fescue Belt, (h) Arid Prairie, (i) High Plains, and (j) Upper Midwest & Northeast ecoregions. In all Manhattan plots the red line indicates an empirically-derived p-value significance threshold from

permutation testing ($p < 1\times10–5$). Note the drastically inflated p-values from the linear model in (a). Further, note that associated loci are not consistent between linear model and linear mixed model, highlighting the need to control for geographic dependency with a genomic relationship matrix.
(TIF)

S4 Fig. Manhattan plots of discrete envGWAS in Simmental cattle. (a) Nine ecoregions of the continental United States defined by *K*-means clustering of 30-year normal temperature, precipitation, and elevation. (b) Locations of Simmental animals colored by breeder's ecoregion and sized by number of animals at that location. (c) Multivariate envGWAS (case-control for regions with > 600 animals). Univariate discrete envGWAS for (d) Desert, (e) Southeast, (f) Fescue Belt, (g) Forested Mountains, (h) High Plains, and (i) Upper Midwest & Northeast ecoregions. In all Manhattan plots the red line indicates an empirically-derived p-value significance threshold from permutation testing ($p < 1\times10^{-5}$). Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).
(TIF)

S5 Fig. Manhattan plots of discrete envGWAS in Gelbvieh cattle. (a) Nine ecoregions of the continental United States defined by *K*-means clustering of 30-year normal temperature, precipitation, and elevation. (b) Locations of Gelbvieh animals colored by breeder's ecoregion and sized by number of animals at that location. (c) Multivariate envGWAS (case-control for regions with > 600 animals). Univariate discrete envGWAS for (d) Desert, (e) Southeast, (f) Fescue Belt, (g) Forested Mountains, (h) High Plains, and (i) Upper Midwest & Northeast ecoregions. In all Manhattan plots the red line indicates an empirically-derived p-value significance threshold from permutation testing ($p < 1\times10^{-5}$). Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).
(TIF)

S6 Fig. Plots of first eight principal components from PCA analysis. Plots for Red Angus (A-D), Simmental (E-H), and Gelbvieh (I-L). Points indicate individuals, colored by their assigned ecoregion.
(TIF)

S7 Fig. Continuous environmental variable envGWAS in Red Angus cattle. Q-Q plots for envGWAS p-values of (a) a linear model for temperature, (b) a linear mixed model for temperature, and (c) a multivariate linear mixed model of temperature, precipitation, and elevation. Geographic distributions colored by (d) temperature, (e) precipitation, (f) elevation. Manhattan plots for univariate envGWAS analysis of (g) temperature, (h) precipitation, (i) elevation. Red lines indicate permutation-derived p-value cutoff of 1×10^{-5}. Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).
(TIF)

S8 Fig. Continuous environmental variable envGWAS in Simmental cattle. (a) Multivariate envGWAS of temperature, precipitation, and elevation for Simmental cattle. Geographic distributions colored by (b) temperature, (d) precipitation, (f) elevation. Manhattan plots for univariate envGWAS analysis of (c) temperature, (e) precipitation, (g) elevation. Red lines indicate permutation-derived p-value cutoff of 1×10^{-5}. Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps

(version 3.1, https://cran.r-project.org/web/packages/maps/).
(TIF)

S9 Fig. Continuous environmental variable envGWAS in Gelbvieh cattle. (a) Multivariate envGWAS of temperature, precipitation, and elevation for Gelbvieh cattle. Geographic distributions colored by (b) temperature, (d) precipitation, (f) elevation. Manhattan plots for univariate envGWAS analysis of (c) temperature, (e) precipitation, (g) elevation. Red lines indicate permutation-derived p-value cutoff of 1×10^{-5}. Maps were plotted using public domain data from the US Department of Commerce, Census Bureau via the R package maps (version 3.1, https://cran.r-project.org/web/packages/maps/).
(TIF)

S10 Fig. PM-plots and region-specific allele frequency trajectories for meta-analysis SNPs of interest in the Red Angus population ecoregions with > 1,000 genotyped animals. (a) PM-plots for lead SNPs of significant within-region GPSM meta-analysis (Cochran's Q p-value $> 1\times10^{-5}$ and significant in at least one region-specific GPSM analysis $p < 1\times10^{-5}$). Each box represents the lead SNP, colored by ecoregion, and sized by Cochran's Q value (for heterogeneity). (b) Region-specific allele frequency trajectories for lead SNPs since 1980, generated by fitting smoothed loess regression of allele frequency on birth date. Trajectories are colored by ecoregion.
(TIF)

S11 Fig. PM-plots and region-specific allele frequency trajectories for meta-analysis SNPs of interest in the Simmental population ecoregions with > 1,000 genotyped animals. (a) PM-plots for lead SNPs of significant within-region GPSM meta-analysis (Cochran's Q p-value $> 1\times10^{-5}$ and significant in at least one region-specific GPSM analysis $p < 1\times10^{-5}$). Each box represents the lead SNP, colored by ecoregion, and sized by Cochran's Q value (for heterogeneity). (b) Region-specific allele frequency trajectories for lead SNPs since 1980, generated by fitting smoothed loess regression of allele frequency on birth date. Trajectories are colored by ecoregion.
(TIF)

S12 Fig. PM-plots and region-specific allele frequency trajectories for meta-analysis SNPs of interest in the Gelbvieh population ecoregions with > 1,000 genotyped animals. (a) PM-plots for lead SNPs of significant within-region GPSM meta-analysis (Cochran's Q p-value $> 1\times10^{-5}$ and significant in at least one region-specific GPSM analysis $p < 1\times10^{-5}$). Each box represents the lead SNP, colored by ecoregion, and sized by Cochran's Q value (for heterogeneity). (b) Region-specific allele frequency trajectories for lead SNPs since 1980, generated by fitting smoothed loess regression of birth of allele frequency on birth date. Trajectories are colored by ecoregion.
(TIF)

S1 Table. GPSM stochastic simulation results. Descriptions of 36 selection scenarios, and the corresponding true and false positive rates for GPSM detecting simulated QTL under selection (simulated QTL GPSM q-value < 0.1). Each scenario's true and false positive statistics were calculated based on 10 replicates starting with different founder populations and selected randomly (false positives) or based on true breeding value (true positives). For each scenario, we report the number simulated QTL, the number of total crosses performed using 50 males and 500 females, the distribution from which QTL effects were drawn from, and the number of generations of selection performed. The mean PVE was reported for both selection on true breeding value (TBV) and random mating. We report results when 10,000 simulated

individuals were randomly chosen to be genotyped (evenly each generation) and when more recent animals were genotyped more frequently (uneven sampling).
(XLSX)

S2 Table. Summary statistics and candidate genes from significant GPSM SNPs for Red Angus, Simmental, and Gelbvieh populations. Variants are significant if GPSM q-value < 0.1. Genomic locations are reported based on coordinates from ARS-1.2 genome assembly. Candidate genes were assigned to a SNP if within 10 kb of a significant SNP.
(XLSX)

S3 Table. Gene enrichment analysis of GPSM candidate genes in Red Angus, Simmental, and Gelbvieh populations. Candidate genes were annotated genes < 10 kb to significant GPSM SNPs ($q < 0.1$). Significant (FDR-corrected p-values < 0.1) KEGG pathways and GO biological processes are reported for each breed.
(XLSX)

S4 Table. TissueEnrich analysis using GPSM gene sets from Red Angus, Simmental, and Gelbvieh populations. TSEA from TissueEnrich software using Human Protein Atlas gene expression data. Enrichment analysis carried out for candidate genes within 10 kb of significant envGWAS SNPs. For each test, we report the number of tissue specific genes, their average fold change, and the FDR-corrected $\log_{10}$ p-value for tissue enriched expression.
(XLSX)

S5 Table. Tissue enrichment analysis results from GPSM gene sets in Red Angus, Simmental, and Gelbvieh populations using the pSI R package and human GTEx expression data. Enrichment significance values for four specificity index thresholds (pSI) of 25 human tissues types. Each combination of stringency for enrichment (pSI) and tissue reports a p-value for Fisher's Exact Test and a Benjamini Hochberg corrected p-value reported in parentheses. Tissue-gene-set combinations that are significant (Benjamini-Hochberg p-value < 0.1) are highlighted in red, those that are suggestive (raw p-value < 0.1) are highlighted in green.
(XLSX)

S6 Table. envGWAS significant SNPs and candidate genes. SNPs were significant when $p < 1\times10^{-5}$. Candidate genes are genes within < 10 kb of significant envGWAS SNPs. We report significant SNPs from all univariate and multivariate analyses for both continuous environmental variables and discrete environments in Red Angus, Simmental, and Gelbvieh populations.
(XLSX)

S7 Table. Gene enrichment analysis of envGWAS candidate genes in Red Angus, Simmental, and Gelbvieh populations. Candidate genes were annotated genes < 10 kb to significant envGWAS SNPs (p-value $< 1\times10^{-5}$). A single gene list for each breed was generated using significant SNPs from all combinations of univariate/multivariate, continuous/discrete envGWAS. Significant (FDR-corrected p-values < 0.1) KEGG pathways and GO biological processes are reported, along with the associated genes.
(XLSX)

S8 Table. TissueEnrich analysis using envGWAS gene sets from Red Angus, Simmental, and Gelbvieh populations. TSEA from TissueEnrich software using Human Protein Atlas gene expression data. Enrichment analysis carried out for candidate genes within 10 kb of significant envGWAS SNPs. For each test, we report the number of tissue specific genes, their

average fold change, and the FDR-corrected $\log_{10}$ p-value for tissue enriched expression.
(XLSX)

S9 Table. Tissue enrichment analysis results from envGWAS gene sets in Red Angus, Simmental, and Gelbvieh populations using the pSI R package and human GTEx expression data. Enrichment significance values for four specificity index thresholds (pSI) of 25 human tissues types. Each combination of stringency for enrichment (pSI) and tissue reports a p-value for Fisher's Exact Test and a Benjamini Hochberg corrected p-value reported in parentheses. Tissue-gene-set combinations that are significant (Benjamini-Hochberg p-value < 0.1) are highlighted in red, those that are suggestive (raw p-value < 0.1) are highlighted in green.
(XLSX)

S10 Table. Brain region and cell-type enrichment analysis results from envGWAS gene sets in Red Angus, Simmental, and Gelbvieh populations using the pSI R package with expression data from the Allen Brain Atlas. Enrichment significance values for four specificity index thresholds (pSI) of six brain regions and 35 brain cell types. Each combination of stringency for enrichment (pSI) and brain region/cell type reports a p-value for Fisher's Exact Test and a Benjamini Hochberg corrected p-value reported in parentheses. Combinations that are significant (Benjamini-Hochberg p-value < 0.1) are highlighted in red, those that are suggestive (raw p-value < 0.1) are highlighted in green.
(XLSX)

S11 Table. Cell-type specific expression of homologous *C. elegans* genes derived from envGWAS candidate gene lists of Red Angus, Simmental, and Gelbvieh populations. *C. elegans* gene homologs were generated from Ortholist2, requiring that genes be present in at least three data sources to be included in enrichment analysis. For each breed's gene list, we include a list of worm tissues with significant enrichment of listed genes (q-value < 0.1).
(XLSX)

Acknowledgments

We appreciate comments from Wes Warren, Jeremy Taylor, and William Lamberson while writing this manuscript and advice from Iain Mathieson on allele trajectory analyses. We valued input on simulations from members of the AlphaGenes group at the Roslin Institute, especially John Hickey, Gregor Gorjanc, Chris Gaynor, Martin Johnsson, Owen Powell, and Ivan Pocrnic. We also appreciate farmers and ranchers for collecting this data and for data being shared from the Red Angus Association of America, American Simmental Association, and American Gelbvieh Association.

Author Contributions

Conceptualization: Troy N. Rowan, Jared E. Decker.

Data curation: Robert D. Schnabel.

Formal analysis: Troy N. Rowan.

Funding acquisition: Jared E. Decker.

Investigation: Troy N. Rowan, Harly J. Durbin, Christopher M. Seabury, Robert D. Schnabel, Jared E. Decker.

Methodology: Troy N. Rowan, Jared E. Decker.

Project administration: Jared E. Decker.

Supervision: Jared E. Decker.

Validation: Troy N. Rowan.

Visualization: Troy N. Rowan.

Writing – original draft: Troy N. Rowan, Jared E. Decker.

Writing – review & editing: Troy N. Rowan, Harly J. Durbin, Christopher M. Seabury, Jared E. Decker.

References

1. Gutiérrez-Gil B, Arranz JJ, Wiener P. An interpretive review of selective sweep studies in Bos taurus cattle populations: identification of unique and shared selection signals across breeds. Front Genet. 2015; 6: 167. https://doi.org/10.3389/fgene.2015.00167 PMID: 26029239
2. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011; 331: 920–924. https://doi.org/10.1126/science.1198878 PMID: 21330547
3. Höllinger I, Pennings PS, Hermisson J. Polygenic adaptation: From sweeps to subtle frequency shifts. PLoS Genet. 2019; 15: e1008035. https://doi.org/10.1371/journal.pgen.1008035 PMID: 30893299
4. Wright S. THE GENETICAL STRUCTURE OF POPULATIONS. Ann Eugen. 1949; 15: 323–354.
5. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, et al. Detecting selection in population trees: the Lewontin and Krakauer test extended. Genetics. 2010; 186: 241–262. https://doi.org/10.1534/genetics.104.117275 PMID: 20855576
6. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. Genome Res. 2010; 20: 393–402. https://doi.org/10.1101/gr.100545.109 PMID: 20086244
7. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007; 449: 913–918. https://doi.org/10.1038/nature06250 PMID: 17943131
8. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002; 419: 832–837. https://doi.org/10.1038/nature01140 PMID: 12397357
9. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006; 4: e72. https://doi.org/10.1371/journal.pbio.0040072 PMID: 16494531
10. Ramey HR, Decker JE, McKay SD, Rolf MM, Schnabel RD, Taylor JF. Detection of selective sweeps in cattle using genome-wide SNP data. BMC Genomics. 2013; 14: 382. https://doi.org/10.1186/1471-2164-14-382 PMID: 23758707
11. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 2014; 10: e1004148. https://doi.org/10.1371/journal.pgen.1004148 PMID: 24586189
12. Rothammer S, Seichter D, Förster M, Medugorac I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics. 2013; 14: 908. https://doi.org/10.1186/1471-2164-14-908 PMID: 24359457
13. Utsunomiya YT, Pérez O'Brien AM, Sonstegard TS, Van Tassell CP, do Carmo AS, Mészáros G, et al. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. PLoS One. 2013; 8: e64280. https://doi.org/10.1371/journal.pone.0064280 PMID: 23696874
14. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol. 2015; 47: 49. https://doi.org/10.1186/s12711-015-0127-3 PMID: 26089079
15. Boitard S, Boussaha M, Capitan A, Rocha D, Servin B. Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. Genetics. 2016; 203: 433–450. https://doi.org/10.1534/genetics.115.181594 PMID: 27017625
16. Ma Y, Ding X, Qanbari S, Weigend S, Zhang Q, Simianer H. Properties of different selection signature statistics and a new strategy for combining them. Heredity. 2015; 115: 426–436. https://doi.org/10.1038/hdy.2015.42 PMID: 25990878

17. Decker JE, Vasco DA, McKay SD, McClure MC, Rolf MM, Kim J, et al. A novel analytical method, Birth Date Selection Mapping, detects response of the Angus (Bos taurus) genome to selection on complex traits. BMC Genomics. 2012; 13: 606. https://doi.org/10.1186/1471-2164-13-606 PMID: 23140540

18. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One. 2009; 4: e5350. https://doi.org/10.1371/journal.pone.0005350 PMID: 19390634

19. Walsh B, Lynch M. Evolution and selection of quantitative traits. 2018. Available: https://books.google.ca/books?hl=en&lr=&id=L2liDwAAQBAJ&oi=fnd&pg=PP1&ots=y9dWVmdg1F&sig=pOREAZIAXXiV3gcMJ2WO-qKSEkc

20. VanRaden PM. Efficient methods to compute genomic predictions. J Dairy Sci. 2008; 91: 4414–4423. https://doi.org/10.3168/jds.2007-0980 PMID: 18946147

21. Rutherford B. U.S. beef herd is mostly black but changing slightly. 29 Jan 2014 [cited 18 Jan 2021]. Available: https://www.beefmagazine.com/cattle-genetics/us-beef-herd-mostly-black-changing-slightly

22. Burns WC, Koger M, Butts WT, Pahnish OF, Blackwell RL. Genotype by Environment Interaction in Hereford Cattle: II. Birth and Weaning Traits. J Anim Sci. 1979; 49: 403–409.

23. Hohenboken W, Jenkins T, Pollak J, Bullock D, Radakovich S. Genetic improvement of beef cattle adaptation in America. Proceedings of the Beef Improvement Federation's 37th annual research symposium and annual meeting. 2005. pp. 115–120.

24. Bradford HL, Fragomeni BO, Bertrand JK, Lourenco DAL, Misztal I. Genetic evaluations for growth heat tolerance in Angus cattle. J Anim Sci. 2016; 94: 4143–4150. https://doi.org/10.2527/jas.2016-0707 PMID: 27898850

25. Fennewald DJ, Weaber RL, Lamberson WR. Genotype by environment interaction for stayability of Red Angus in the United States. Journal of Animal Science. 2018. pp. 422–429. https://doi.org/10.1093/jas/skx080 PMID: 29385483

26. Carvalheiro R, Costilla R, Neves HHR, Albuquerque LG, Moore S, Hayes BJ. Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. Genet Sel Evol. 2019; 51: 29. https://doi.org/10.1186/s12711-019-0470-x PMID: 31221081

27. Smith JL, Wilson ML, Nilson SM, Rowan TN, Oldeschulte DL, Schnabel RD, et al. Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle. BMC Genomics. 2019; 20: 926. https://doi.org/10.1186/s12864-019-6231-y PMID: 31801456

28. Braz CU, Rowan TN, Schnabel RD, Decker JE. Extensive genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. https://doi.org/10.1101/2020.01.09.900902

29. Blackburn HD, Krehbiel B, Ericsson SA, Wilson C, Caetano AR, Paiva SR. A fine structure genetic analysis evaluating ecoregional adaptability of a Bos taurus breed (Hereford). PLoS One. 2017; 12: e0176474. https://doi.org/10.1371/journal.pone.0176474 PMID: 28459870

30. Coop G, Witonsky D, Di Rienzo A, Pritchard JK. Using environmental correlations to identify loci underlying local adaptation. Genetics. 2010; 185: 1411–1423. https://doi.org/10.1534/genetics.110.114819 PMID: 20516501

31. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. Genetics. 2013; 195: 205–220. https://doi.org/10.1534/genetics.113.152462 PMID: 23821598

32. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet. 2014; 15: 335–346. https://doi.org/10.1038/nrg3706 PMID: 24739678

33. Foll M, Shim H, Jensen JD. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. Mol Ecol Resour. 2015; 15: 87–98. https://doi.org/10.1111/1755-0998.12280 PMID: 24834845

34. Wiedemar N, Tetens J, Jagannathan V, Menoud A, Neuenschwander S, Bruggmann R, et al. Independent polled mutations leading to complex gene expression differences in cattle. PLoS One. 2014; 9: e93435. https://doi.org/10.1371/journal.pone.0093435 PMID: 24671182

35. Mészáros G, Petautschnig E, Schwarzenbacher H, Sölkner J. Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. Anim Genet. 2015; 46: 65–68. https://doi.org/10.1111/age.12249 PMID: 25515556

36. Fontanesi L, Tazzoli M, Russo V, Beever J. Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. Anim Genet. 2010; 41: 295–303. https://doi.org/10.1111/j.1365-2052.2009.02007.x PMID: 19968642

37. Serão NV, González-Peña D, Beever JE, Faulkner DB, Southey BR, Rodriguez-Zas SL. Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. BMC Genet. 2013; 14: 94. https://doi.org/10.1186/1471-2156-14-94 PMID: 24066663

38. Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, et al. Genome-wide association study of growth in crossbred beef cattle. J Anim Sci. 2010; 88: 837–848. https://doi.org/10.2527/jas.2009-2257 PMID: 19966163

39. Seabury CM, Oldeschulte DL, Saatchi M, Beever JE, Decker JE, Halley YA, et al. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. BMC Genomics. 2017; 18: 386. https://doi.org/10.1186/s12864-017-3754-y PMID: 28521758

40. Chelh I, Picard B, Hocquette J-F, Cassar-Malek I. Myostatin inactivation induces a similar muscle molecular signature in double-muscled cattle as in mice. Animal. 2011; 5: 278–286. https://doi.org/10.1017/S1751731110001862 PMID: 22440772

41. Espigolan R, Baldi F, Boligon AA, Souza FRP, Fernandes Júnior GA, Gordo DGM, et al. Associations between single nucleotide polymorphisms and carcass traits in Nellore cattle using high-density panels. Genet Mol Res. 2015; 14: 11133–11144. https://doi.org/10.4238/2015.September.22.7 PMID: 26400344

42. Chen Y, Gondro C, Quinn K, Herd RM, Parnell PF, Vanselow B. Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. Anim Genet. 2011; 42: 475–490. https://doi.org/10.1111/j.1365-2052.2011.02182.x PMID: 21906099

43. Kuehn LA, Thallman RM. Across-Breed EPD Tables For The Year 2016 Adjusted To Breed Differences For Birth Year Of 2014. 2016 [cited 9 Feb 2020]. Available: https://digitalcommons.unl.edu/hruskareports/380/

44. Guo B, Greenwood PL, Cafe LM, Zhou G, Zhang W, Dalrymple BP. Transcriptome analysis of cattle muscle identifies potential markers for skeletal muscle growth rate and major cell types. BMC Genomics. 2015; 16: 177. https://doi.org/10.1186/s12864-015-1403-x PMID: 25887672

45. Rolf MM, Taylor JF, Schnabel RD, McKay SD, McClure MC, Northcutt SL, et al. Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet. 2012; 43: 367–374. https://doi.org/10.1111/j.1365-2052.2011.02273.x PMID: 22497295

46. Breeds—Red Angus. In: The Cattle Site [Internet]. [cited 28 Feb 2020]. Available: https://www.thecattlesite.com/breeds/beef/99/red-angus/

47. Gareis NC, Huber E, Hein GJ, Rodríguez FM, Salvetti NR, Angeli E, et al. Impaired insulin signaling pathways affect ovarian steroidogenesis in cows with COD. Anim Reprod Sci. 2018; 192: 298–312. https://doi.org/10.1016/j.anireprosci.2018.03.031 PMID: 29622349

48. Davis SL, Hossner KL, Ohlson DL. Endocrine Regulation of Growth in Ruminants. In: Roche JF, O'Callaghan D, editors. Manipulation of Growth in Farm Animals: A Seminar in the CEC Programme of Coordination of Research on Beef Production, held in Brussels December 13–14, 1982. Dordrecht: Springer Netherlands; 1984. pp. 151–178.

49. Hill WG. Applications of population genetics to animal breeding, from wright, fisher and lush to genomic prediction. Genetics. 2014; 196: 1–16. https://doi.org/10.1534/genetics.112.147850 PMID: 24395822

50. Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic signature of adaptation to climate in Medicago truncatula. Genetics. 2014; 196: 1263–1275. https://doi.org/10.1534/genetics.113.159319 PMID: 24443444

51. Li J, Chen G-B, Rasheed A, Li D, Sonder K, Zavala Espinosa C, et al. Identifying loci with breeding potential across temperate and tropical adaptation via EigenGWAS and EnvGWAS. Mol Ecol. 2019; 28: 3544–3560. https://doi.org/10.1111/mec.15169 PMID: 31287919

52. Lasky JR, Upadhyaya HD, Ramu P, Deshpande S, Hash CT, Bonnette J, et al. Genome-environment associations in sorghum landraces predict adaptive traits. Sci Adv. 2015; 1: e1400218. https://doi.org/10.1126/sciadv.1400218 PMID: 26601206

53. Sathiaraj D, Huang X, Chen J. Predicting climate types for the Continental United States using unsupervised clustering techniques. Environmetrics. 2019. p. e2524. https://doi.org/10.1002/env.2524

54. Sansregret L, Nepveu A. The multiple roles of CUX1: insights from mouse models and cell-based assays. Gene. 2008; 412: 84–94. https://doi.org/10.1016/j.gene.2008.01.017 PMID: 18313863

55. Bertolini F, Servin B, Talenti A, Rochat E, Kim ES, Oget C, et al. Signatures of selection and environmental adaptation across the goat genome post-domestication. Genet Sel Evol. 2018; 50: 57. https://doi.org/10.1186/s12711-018-0421-y PMID: 30449276

56. Aiken GE, Klotz JL, Looper ML, Tabler SF, Schrick FN. Disrupted hair follicle activity in cattle grazing endophyte-infected tall fescue in the summer insulates core body temperatures1. The Professional Animal Scientist. 2011; 27: 336–343.

57. León CD, De León C, Manrique C, Martínez R, Rocha JF. Research Article Genomic association study for adaptability traits in four Colombian cattle breeds. Genetics and Molecular Research. 2019. https://doi.org/10.4238/gmr18373

58. Guo J, Tao H, Li P, Li L, Zhong T, Wang L, et al. Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. Sci Rep. 2018; 8: 10405. https://doi.org/10.1038/s41598-018-28719-w PMID: 29991772

59. Gurgul A, Jasielczuk I, Ropka-Molik K, Semik-Gurgul E, Pawlina-Tyszko K, Szmatoła T, et al. A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. BMC Genet. 2018; 19: 95. https://doi.org/10.1186/s12863-018-0681-0 PMID: 30348079

60. Medugorac I, Graf A, Grohs C, Rothammer S, Zagdsuren Y, Gladyr E, et al. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nat Genet. 2017; 49: 470–475. https://doi.org/10.1038/ng.3775 PMID: 28135247

61. Guo D-F, Cui H, Zhang Q, Morgan DA, Thedens DR, Nishimura D, et al. The BBSome Controls Energy Homeostasis by Mediating the Transport of the Leptin Receptor to the Plasma Membrane. PLoS Genet. 2016; 12: e1005890. https://doi.org/10.1371/journal.pgen.1005890 PMID: 26926121

62. Davis RE, Swiderski RE, Rahmouni K, Nishimura DY, Mullins RF, Agassandian K, et al. A knockin mouse model of the Bardet–Biedl syndrome 1 M390R mutation has cilia defects, ventriculomegaly, retinopathy, and obesity. Proc Natl Acad Sci U S A. 2007; 104: 19422–19427. https://doi.org/10.1073/pnas.0708571104 PMID: 18032602

63. Igoshin AV, Yurchenko AA, Belonogova NM, Petrovsky DV, Aitnazarov RB, Soloshenko VA, et al. Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations. BMC Genet. 2019; 20: 26. https://doi.org/10.1186/s12863-019-0725-0 PMID: 30885142

64. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet. 2015; 47: 217–225. https://doi.org/10.1038/ng.3199 PMID: 25621459

65. Morrison SF. Central control of body temperature. F1000Res. 2016;5. https://doi.org/10.12688/f1000research.7958.1 PMID: 27239289

66. Garner JB, Douglas ML, Williams SRO, Wales WJ, Marett LC, Nguyen TTT, et al. Genomic Selection Improves Heat Tolerance in Dairy Cattle. Sci Rep. 2016; 6: 34114. https://doi.org/10.1038/srep34114 PMID: 27682591

67. Decker JE. Agricultural Genomics: Commercial Applications Bring Increased Basic Research Power. PLoS Genet. 2015; 11: e1005621. https://doi.org/10.1371/journal.pgen.1005621 PMID: 26539986

68. Fraser HB. Gene expression drives local adaptation in humans. Genome Res. 2013; 23: 1089–1096. https://doi.org/10.1101/gr.152710.112 PMID: 23539138

69. Lenormand T. Gene flow and the limits to natural selection. Trends Ecol Evol. 2002; 17: 183–189.

70. Castric V, Bechsgaard J, Schierup MH, Vekemans X. Repeated adaptive introgression at a gene under multiallelic balancing selection. PLoS Genet. 2008; 4: e1000168. https://doi.org/10.1371/journal.pgen.1000168 PMID: 18769722

71. Rowan TN, Hoff JL, Crum TE, Taylor JF, Schnabel RD, Decker JE. A multi-breed reference panel and additional rare variants maximize imputation accuracy in cattle. Genet Sel Evol. 2019; 51: 77. https://doi.org/10.1186/s12711-019-0519-x PMID: 31878893

72. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. Gigascience. 2020; 9. https://doi.org/10.1093/gigascience/giaa021 PMID: 32191811

73. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. https://doi.org/10.1086/519795 PMID: 17701901

74. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016; 48: 1443. https://doi.org/10.1038/ng.3679 PMID: 27694958

75. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016; 48: 1284–1287. https://doi.org/10.1038/ng.3656 PMID: 27571263

76. Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014; 11: 407–409. https://doi.org/10.1038/nmeth.2848 PMID: 24531419

77. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100: 9440–9445. https://doi.org/10.1073/pnas.1530509100 PMID: 12883005

78. R Core Team R, Others. R: A language and environment for statistical computing. R foundation for statistical computing Vienna, Austria; 2013.

79. Wickham H. ggplot2. Wiley Interdisciplinary Reviews: Computational. 2011. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.147

80. Faux A-M, Gorjanc G, Gaynor RC, Battagin M, Edwards SM, Wilson DL, et al. AlphaSim: Software for Breeding Program Simulation. Plant Genome. 2016;9. https://doi.org/10.3835/plantgenome2016.02.0013 PMID: 27902803

81. Chen GK, Marjoram P, Wall JD. Fast and flexible simulation of DNA sequence data. Genome Res. 2009; 19: 136–142. https://doi.org/10.1101/gr.083634.108 PMID: 19029539

82. Hayes B, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol. 2001; 33: 209–229. https://doi.org/10.1186/1297-9686-33-3-209 PMID: 11403745

83. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. JOSS. 2019; 4: 1686.

84. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88: 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011 PMID: 21167468

85. Wellmann R. Optimum contribution selection for animal breeding and conservation: the R package optiSel. BMC Bioinformatics. 2019; 20: 25. https://doi.org/10.1186/s12859-018-2450-5 PMID: 30642239

86. Bhatia G, Patterson N, Pasaniuc B, Zaitlen N, Genovese G, Pollack S, et al. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. Am J Hum Genet. 2011; 89: 368–381. https://doi.org/10.1016/j.ajhg.2011.07.025 PMID: 21907010

87. PRISM Climate Group. PRISM 30-year Normal Climate Data. Available: http://prism.oregonstate.edu

88. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. Genet Epidemiol. 2008; 32: 227–234. https://doi.org/10.1002/gepi.20297 PMID: 18300295

89. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011; 88: 586–598. https://doi.org/10.1016/j.ajhg.2011.04.014 PMID: 21565292

90. Han B, Eskin E. Interpreting meta-analyses of genome-wide association studies. PLoS Genet. 2012; 8: e1002555. https://doi.org/10.1371/journal.pgen.1002555 PMID: 22396665

91. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25: 1091–1093. https://doi.org/10.1093/bioinformatics/btp101 PMID: 19237447

92. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13: 2498–2504. https://doi.org/10.1101/gr.1239303 PMID: 14597658

93. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347: 1260419. https://doi.org/10.1126/science.1260419 PMID: 25613900

94. Mouse ENCODE Consortium, Stamatoyannopoulos JA, Snyder M, Hardison R, Ren B, Gingeras T, et al. An encyclopedia of mouse DNA elements (Mouse ENCODE). Genome Biol. 2012; 13: 418. https://doi.org/10.1186/gb-2012-13-8-418 PMID: 22889292

95. Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis. Bioinformatics. 2019; 35: 1966–1967. https://doi.org/10.1093/bioinformatics/bty890 PMID: 30346488

96. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 2015; 348: 648–660. https://doi.org/10.1126/science.1262110 PMID: 25954001

97. Xu X, Wells AB, O'Brien DR, Nehorai A, Dougherty JD. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. J Neurosci. 2014; 34: 1420–1431. https://doi.org/10.1523/JNEUROSCI.4488-13.2014 PMID: 24453331

98. Kim W, Underwood RS, Greenwald I, Shaye DD. OrthoList 2: A New Comparative Genomic Analysis of Human and Caenorhabditis elegans Genes. Genetics. 2018; 210: 445–461. https://doi.org/10.1534/genetics.118.301307 PMID: 30120140

99. Angeles-Albores D, N Lee RY, Chan J, Sternberg PW. Tissue enrichment analysis for C. elegans genomics. BMC Bioinformatics. 2016; 17: 366. https://doi.org/10.1186/s12859-016-1229-9 PMID: 27618863

100. Angeles-Albores D, Lee RYN, Chan J, Sternberg PW. Two new functions in the WormBase Enrichment Suite. microPublication Biology: 2018. https://doi.org/10.17912/W25Q2N PMID: 32550381

101. Rowan TN, Durbin HJ, Seabury CM, Schnabel RD, Decker JE. Data From: Powerful detection of polygenic selection and environmental adaptation in US beef cattle. Zenodo; 2021. https://doi.org/10.5281/zenodo.4455543