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

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

Abstract
Selection is a basic principle of evolution. Many approaches and studies have identified DNA mutations rapidly selected to high frequencies, which leave pronounced signatures on the surrounding sequence (selective sweeps). However, for many important complex, quantitative traits, selection does not leave these intense signatures. Instead, hundreds or thousands of loci experience small changes in allele frequencies, a process called polygenic selection. We know that directional selection and local adaptation are actively changing genomes; however we have been unable to identify the genomic loci responding to this polygenic, complex selection. Here we show that the use of novel dependent variables in linear mixed models (which allow us to account for population structure, relatedness, inbreeding) identify complex, polygenic selection and local adaptation. Thousands of loci are responding to artificial directional selection and hundreds of loci have evidence of local adaptation. While advanced reproduction and genomic technologies are increasing the rate of directional selection, local adaptation is being lost. In a changing climate, the loss of local adaptation may be especially problematic. These selection mapping approaches can be used in evolutionary, model, and agriculture contexts.

In a changing climate, organisms are forced to adapt to environmental pressures rapidly or perish. The genomic changes that underlie adaptation are not well known and have proven difficult to map, aside from a handful of large-effect selective sweeps [1]. It is becoming increasingly apparent that in adaptation, hard sweeps are likely the exception, not the rule [2,3]. Selection on complex traits causes significant changes in the mean phenotype while producing only subtle changes in allele frequencies across the genome [2,4–6]. Most methods require discrete grouping of genetically similar individuals, making the identification of selection within a panmictic population difficult. American *Bos taurus* beef cattle have been under well-documented strong artificial and environmental selection over the last ~50 years, making them an intriguing model to study how selection acts on the genome over short periods of time and across diverse environments. Though domesticated, beef cattle are exposed to a host of unique environments and local selection pressures compared to other more intensely managed livestock populations. This suggests that local adaptation and genotype-by-environment interactions may play an important role in beef cattle. Here, we present two methods for detecting complex polygenic selection and local adaptation using genome-wide linear mixed models (LMM) in conjunction with novel dependent variables. These models identify ongoing polygenic selection on complex traits and local adaptation within populations while explicitly controlling for family structure. When applied to three US beef cattle populations, we identify multiple locally adaptive genomic regions that provide insight into the biology of adaptation in mammalian species, all while assisting producers in identifying and selecting locally adapted individuals to further reduce the industry's environmental footprint.

Detecting ongoing selection with Generation Proxy Selection Mapping (GPSM)

Though the first cattle genotyping assay was developed just over a decade ago [7], influential individuals from deep within the pedigree have been genotyped, providing a temporal distribution of samples at least ten generations deep for most well-established breeds. Under directional selection, alleles will be at significantly different frequencies in more recent generations compared to distant ones. This creates a statistical association between allele frequencies at a selected locus and an individual's generation number. Historical selective sweeps on simple traits in many species, including cattle, has been successfully studied, but recent and ongoing selection for complex traits is more difficult to characterize. Most methods used to identify selection rely on allele frequency differences between diverged populations (e.g. F_{ST}, FLK, XP-CLR) [8–10], or on the disruption of normal LD patterns (iHS, EHH, etc.) [11–13]. In cattle, these methods have successfully identified genomic regions under selection that control a handful of Mendelian and simple traits like coat color, the absence of horns, or large-effect genes involved in domestication [14–17 1,18,1914–17]. 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. To identify selected alleles changing in frequency over time, we fit an individual's generation number or a proxy, such as its birth year, as the dependent variable in a genome-wide linear mixed model (LMM) [20]. Using a linear mixed model approach explicitly controls for confounding family and population structure with a genomic relationship matrix [21]. Our method, Generation Proxy Selection Mapping (GPSM), is unique in its ability to detect subtle allele frequency changes caused by recent and ongoing polygenic selection within homogeneous populations (like breeds or smaller related groups). When applied to simulations and real cattle populations, GPSM is able to effectively distinguish selection from genetic drift. In identifying alleles actively changing in frequency, we can study selection on complex traits as an active process as opposed to retrospective analyses of completed or near-complete sweeps.

We performed simulations under a variety of genetic architectures, selection intensities, effective population sizes, and timescales to test (1) if GPSM can distinguish between selection and drift, (2) what characterizes the allele frequency shifts that produce signal in GPSM, (3) the scenarios that produce these detectable shifts, (4) and whether selection on complex phenotypes in cattle could produce them. Simulations and their results are summarized in (**Supplementary File 1**). For each simulation we performed selection on a single additive trait in three parallel regimes; on true breeding value, on phenotype, and random-mating. Across all simulations, we rarely observed false-positive GPSM signal in our randomly-selected population, suggesting that drift is not producing detectable changes in frequency, even in populations with small effective sizes (N_e). The most significant hits in GPSM were typically not the largest effect simulated QTL, but the variants that are actively undergoing changes in frequency. Since many large effect variants become fixed in early generations, intermediate-effect variants are enriched in GPSM signals. We detected signatures of selection under all genetic architectures that were not purely infinitesimal (where each SNP has some small effect). GPSM signals persists when we reduce the number of generations of selection simulated from 20 to 10 to 5. Our simulations also demonstrated the benefits of a linear mixed model framework. Without using a genomic relationship matrix to control population and family structure, we observe massively inflated p-values and a large number of false-positive signals. Finally, in situations where small numbers of early-generation individuals are genotyped, we are able to detect selection with a similar effectiveness as when sampling evenly across generations. Simulations randomly chose which individuals from each generation were genotyped. When applying GPSM to real data, early generation animals are almost always influential sires for the modern breed. This means that they are likely much more genetically similar to subsequent generation's animals than randomly sampled individuals, making the observed allele frequency changes over time appear smaller than they actually were in the entire population.

We applied GPSM to detect genomic regions under selection in three common breeds of United States beef cattle; Red Angus, Simmental, and Gelbvieh. Strong, population-altering selection on both Mendelian and quantitative traits has created quantifiable genetic differences over the last ~10 generations (50 years) in these populations. In addition to observable changes in visual appearance, clear genetic trends for complex traits involved in growth, carcass quality, and maternal performance exist in each breed [22]. When performing GPSM, we estimate that the proportion of additive genetic variance of an individual's birth year is large [Proportion Variance Explained (PVE) = 0.520, 0.588, 0.459 in Red Angus, Simmental, and Gelbvieh, respectively], meaning that we can predict an animal's generation (or relevant proxy) with ~70% accuracy if we know its genotype [23]. Obviously, birth year is not a heritable trait, but these PVE estimates indicate significant associations between genotype and birth year across the genome. The large amount of genetic variance explained by birth year persists even when restricting the analysis to individuals born in the last 10 years (~2 generations) or 20 years (~4 generations) (**Table 1**), suggesting that GPSM has the power to detect subtle allele frequency changes over extremely short time periods. We removed the link between generation proxy and genotype by randomly permuting the birth years assigned to each animal and observed PVEs drop to effectively zero (**Table 1**).

Table 1. The proportion of variation in birth year explained (PVE) by markers in GPSM analysis. PVE calculated for each breed dataset in full, and subsets of individuals born within the last 20 or 10 years. The standard errors of PVE estimates are reported in parentheses.

Breed	Full Dataset PVE (se)	20-year PVE (se)	10-year PVE (se)	Shuffled PVE (se)
Red Angus	0.520 (0.013)	0.358 (0.013)	0.406 (0.014)	9.82×10^{-6} (0.002)
Simmental	0.588 (0.009)	0.551 (0.010)	0.401 (0.014)	9.70×10^{-4} (0.003)
Gelbvieh	0.459 (0.015)	0.454 (0.015)	0.361 (0.014)	5.05×10^{-4} (0.004)

Changes at a handful of Mendelian loci explain the major differences in appearance between European Simmental introduced in the late 1960s and the modern American Simmental population (**Figure 1**). Since their initial import to America in 1968, breeders of Simmental cattle have selected against horns and preferred solid black or red hair coats to spotted cream coloring. We use these traits and their known Mendelian alleles as positive controls for our GPSM method. GPSM identifies significant associations between birth year and allele frequency near the *POLLED* (absence of horns [24]), *ERBB3/PMEL* (European Simmental cream color [25]), and *KIT* (piebald coat coloration [26]) loci. No appreciable changes in frequency have occurred for these alleles since 1995, yet, a signal of selection remains despite only 200 individuals born prior to 1995 being included in the analysis. While these Mendelian loci account for basic differences in appearance between imported European and modern American Simmental, they account for a small proportion of total GPSM signatures (3 of 85 total regions) and explain only a small fraction of the "phenotypic variance" in birth year.

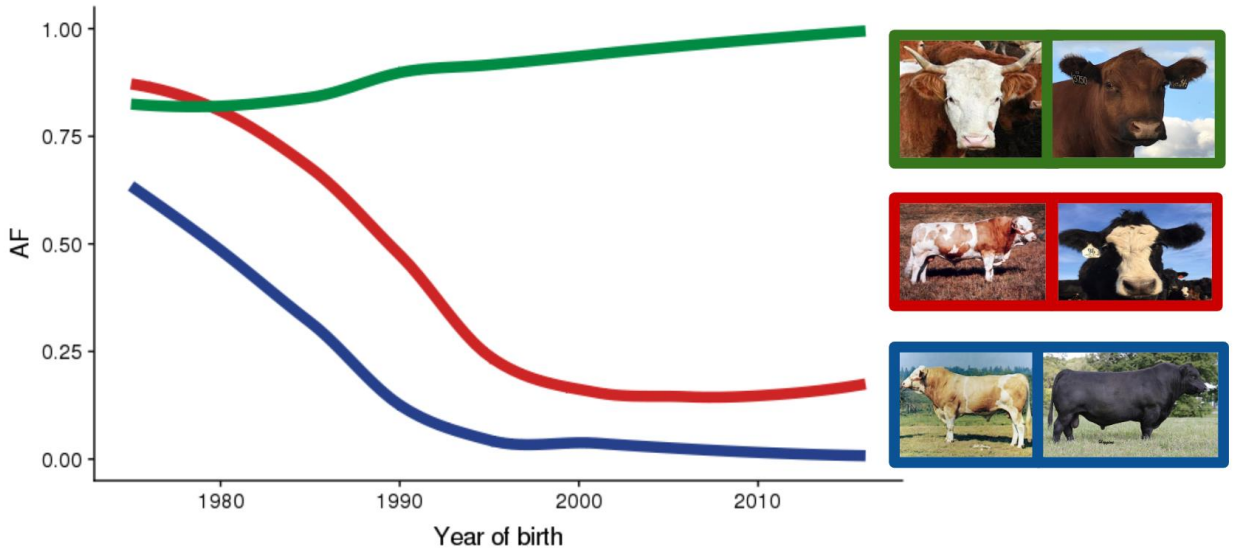

Figure 1. Allele frequency trajectories for three Mendelian loci known to be under selection in American Simmental cattle populations. Line colors correspond with a single locus; *POLLED* is green, *KIT* is red, and *PMEL/ERBB3* is blue.

Shared genomic regions are under selection across populations

Alleles actively undergoing strong selection create an even more powerful signal than those under selection in earlier generations. For example, a variant 56,341 bp upstream of *RHOU* (ras homolog family member U) on BTA28 (GPSM q-value = 2.323×10^{-150}) has increased in frequency from 0.154 to 0.544 in animals born in 2012 compared to 2015 (**Figure 2G**). This rapid shift in frequency demonstrates the ability of strong selection pressure to change allele frequencies over short periods of time. We also observe this variant, under strong selection in Red Angus and Gelbvieh populations (q-values = 2.810×10^{-27} and 2.787×10^{-265} respectively) (**Figure 2G**). *RHOU* is a GTP-binding protein that is involved in cell-shape morphology and cytoskeletal rearrangement [27]. In humans *RHOU* has been shown to play a role in epidermal growth factor signaling [28]. While no previous studies in cattle have pointed to *RHOU* as a large-effect QTL for any economically-relevant traits, its role in cell shape and organization could be involved in increased growth rate. This GPSM locus also resides ~166 kb from a cluster of olfactory receptors. In Simmental and Gelbvieh datasets, significant GPSM markers in this locus extend into this cluster of olfactory genes. It is unclear if this is a continuation of the peak closer to *RHOU*, or separate, but significant selection on these olfactory genes. The allele frequency changes observed for this GPSM peak are extreme compared to other significant regions, most of which are undergoing only small to moderate shifts in frequency over the last ~10 generations.

The largest genomic region detected by GPSM lies at the end of BTA1 (157.5 Mb - 158.5 Mb). The lead SNP in this peak (*rs1755753*) lies within the long non-coding RNA *LOC112448253*. The selected region is immediately upstream of *PRDM9*, a modulator of recombination in most mammalian species, including cattle [29–31]. While no variants within PRDM9 reach genome-wide significance, a variant ~10.7 kb from the TSS does appear to be responding to selection.

We identified six other common genomic regions under strong selection in all three populations, encompassing 106 statistically significant markers (**Table 2**). Another 90 SNPs overlap in at least two

populations, corresponding to 15 additional genomic regions under selection (**Table 3**). These shared GPSM signals suggest that not only are there similar selection pressures, but common genomic architectures under selection in these three breeds. All three breeds have been selected for increased growth and efficiency traits over the last 50 years [22], and GPSM identifies two genomic regions that have been previously associated with feed efficiency and growth traits. The common peak on BTA12 (lead SNP *rs1389713*) is ~182 kb upstream of the *DACH1* (Dachshund homolog 1), a transcription factor associated with post-weaning gain, various indicators of feed efficiency [32,33], and backfat thickness [34] in cattle. The shared GPSM peak on BTA14 resides near a known large-effect QTL for post-weaning gain [35].

In addition to selection on loci that appear directly involved in growth and efficiency, we identify multiple selection targets likely involved in aspects of immune function. A shared significant peak on BTA2 (lead SNP *rs1080110*) resides within the *ARHGAP15* gene (Rho GTPase Activating Protein 15) that is essential for Trypanosomiasis resistance in African cattle populations [36–38]. Though Trypanosomiasis and other tsetse fly-transmitted diseases are restricted to Africa, genetic tolerance to similar immune disturbances may account for the positive selection observed in these three American cattle populations. Variants within *ADORA1* (Adenosine A1 Receptor) are also detected as being under selection by GPSM. In cattle, *ADORA1* plays a role in the activation of polymorphonuclear neutrophilic leukocytes, which are important for peripartial immune responses in cattle [39], and likely play roles in other immune functions. *ADORA1* and other purinergic receptors play an important role in bone metabolism [40] and likely growth in cattle. This signature within *ADORA1* also spans a potentially regulatory region for *MYBPH*, an important gene in muscle formation and development, and another potential target of selection. In this case and others, we identify multiple logical candidate genes within regions undergoing selection.

We observe 22 genomic regions that are changing in frequency in at least two of our datasets, evidence of shared networks and genetic architectures under selection. The lead SNPs for each shared peak and positional candidate genes are reported in **Table 3**. A GO enrichment analysis of 46 genes residing in or near the 29 shared GPSM signatures identified by at least two datasets identified multiple biological processes undergoing selection. The most significant pathways involved G-protein coupled signaling (Benjamini–Hochberg adjusted p-value = 0.038) and purinergic receptor signaling pathways (Benjamini–Hochberg adjusted p-value = 0.034, associated genes *ADORA1* and *P2RY8*). Purinergic receptors have been identified as important drivers of immune responses in cattle [39]. Selection on immune genes and pathways is likely driven by the increased production efficiency of healthy calves [41,42]. Shared selection on *SLC2A5* and *BIRC5* point towards biological pathways involved in sensing carbohydrate, hexose, and monosaccharide stimuli (Benjamini-Hochberg adjusted p-values = 0.01). We also expect that an enhanced metabolic response to carbohydrates would result in increased animal efficiency. Finally, genes involved in the regulation of arterial blood pressure (Benjamini–Hochberg adjusted p-value = 0.045: *ADORA1*, *SLC2A5*) made up the lone other significant gene class under selection across all three breeds.

Table 2. Overlapping GPSM variants in all three breeds. Lead SNPs from significant regions identified in GPSM analyses of Red Angus, Simmental, and Gelbvieh cattle populations. Candidate gene is the annotated gene closest to lowest p-value SNP in peak if < 200 kb away. Associations are from cattle literature unless otherwise reported.

CHR	POS	Nearest Candidate Gene(s) (Distance)	Known Candidate Gene Associations	References
1	157,913,264	LOC112448253 (within)	lncRNA, misplaced BoLA SNPs?, PRDM9 association?	
2	53,151,541	*ARHGAP15* (within)	Immune functions, Trypanosomiasis resistance	[36–38]
12	46,730,506	*DACH1* (182.1 kb)	Feed efficiency/growth	[32]
14	59,774,083	*LRP12* (261.5 kb), LOC112449532 (113.6 kb)	(*LRP12*) Feed efficiency/growth	[35,43]
16	955,146	*ADORA1* (within), *MYBPH* (2.0 kb)	Fertility/immune (*ADORA1*)/muscle growth (*MYBPH*)	[39]
23	1,768,070	LOC782044 (25.7 kb)		
28	640,998	*RHOU* (56.3 kb), Olfactory gene cluster (166.5kb)	Bone development, Innate immune system	[44]

Table 3. Overlapping GPSM variants in two of three breeds. Lead SNPs from significant regions identified at least two of our three cattle populations. Candidate gene is the annotated gene closest to lowest p-value SNP in peak if < 200 kb away. Associations are from cattle literature unless otherwise reported.

CHR	POS	Nearest Candidate Gene (Distance)	Known Candidate Gene Associations	References	Breeds
1	3,144,864	URB1 (within)	Embryonic lethal (pigs)	[45]	SIM/GEL
1	73,642,609	> 200 kb to a gene			SIM/GEL
3	119,429,700	CSF2RA (within)	TB Resistance	[46]	RAN/SIM
5	5,266,489	ENSBTAG00000050164 (within)	lncRNA		SIM/GEL
8	113,265,058	> 200 kb to a gene			SIM/GEL
9	78,704,840	PWWP2B (23.6 kb)	Bovine fetus muscle expression	[47]	RAN/GEL

9	104,228,150	PDCD2 (111 kb)			SIM/GEL
10	940,793	MCC (within)			SIM/GEL
12	518,595	PCDH20 (37.7 kb)			SIM/GEL
15	77,106,772	DDB2 (within)	Calving Ease	[48]	RAN/GEL
15	84,680,379	LOC617614	Olfactory Receptor		SIM/GEL
16	78,129,493	> 200 kb to a gene			SIM/GEL
19	53,947,810	BIRC5 (within)	RFI (DE), Fertility and Reproduction	[49]	RAN/GEL
22	24,625,216	Between CNTN6 and CNTN4	Known Milk Yield QTL	[50]	SIM/GEL
28	20,762,645	> 200 kb to a gene			SIM/GEL

Table 4. GPSM Gene Ontology Terms. Significant GO terms (FDR-corrected p-values < 0.1) for genes < 10kb of significant GPSM SNPs (q < 0.1) in Red Angus, Simmental, or Gelbvieh datasets.

https://docs.google.com/spreadsheets/d/1HP79hSM_PGObrN3ZIMc-vtLKaIKoKFgGU8w5HUlB2Sk/edit?usp=sharing

Generation Proxy Selection Mapping identifies polygenic selection

In total, GPSM identifies 268, 548, and 763 statistically significant SNPs (q-value < 0.1) associated with year-of-birth in Red Angus, Simmental, and Gelbvieh cattle populations, respectively (**Figure 2**). These SNPs encompass at least 52, 85, and 92 total genomic loci. The median per-generation allele frequency change (ΔAF) for significant SNPs are modest, but significantly larger than the upper limits of the per generation drift variance (**Table 4**). If we assume a biallelic SNP with a starting allele frequency of 0.5 in a population where $N_e = 100$, the maximum expected per-generation change in allele frequency would be 0.0012. While we identify some SNPs undergoing massive shifts in allele frequency, most significant GPSM SNPs are undergoing modest allele frequency changes. These small allele frequency shifts detected by GPSM are consistent with the magnitude of shifts expected for traits with polygenic trait architectures [6,51]. As a result, we suspect that many of these loci would go largely undetected when using other selection mapping methods.

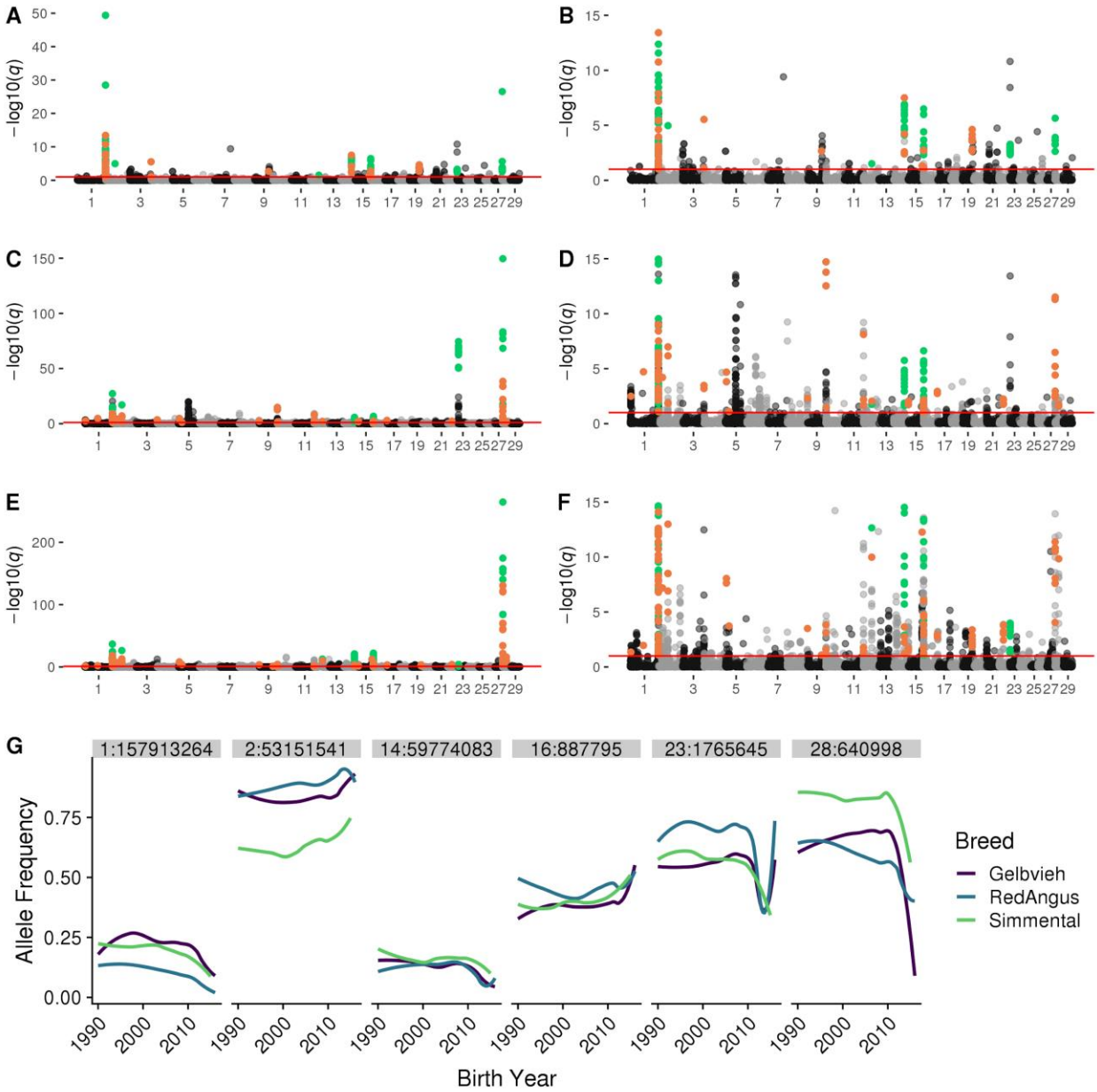

Figure 2. Generation Proxy Selection Mapping (GPSM) analysis of three major US cattle breeds. Manhattan plots of GPSM p-values for (A) Red Angus, (C) Simmental, and (E) Gelbvieh cattle populations. Corresponding truncated (-log10(*q*) < 15) GPSM Manhattan plots for (B) Red Angus, (D) Simmental, and (F) Gelbvieh. Significant SNPs ($q < 0.1$ or $p < 1x10^{-5}$) identified in all three datasets are colored green and SNPs identified in two of three analyses are colored orange. (G) Allele frequency trajectories over time for lead SNPs of six common GPSM peaks in three cattle populations.

Table 4. **Summary statistics of allele frequency change (ΔAF) per generation for significant GPSM SNPs.** ΔAF is the slope of the regression of allele frequency on birth year.

Breed	N SNPs (GPSM q < 0.1)	Mean ΔAF per generation (sd)	Median ΔAF per generation	Min ΔAF per generation	Max ΔAF per generation
Red Angus	268	0.018 (0.011)	0.017	4.97×10^{-5}	0.076
Simmental	548	0.024 (0.017)	0.022	6.79×10^{-5}	0.093
Gelbvieh	762	0.033 (0.028)	0.024	1.01×10^{-4}	0.223

To identify the amount of variation in birth year explained by significant signals, we performed a REML analysis using two GRMs; one created using genome-wide significant GPSM markers (q < 0.1), and another with the remaining non-significant markers. The PVEs for each dataset are reported in **Table 5**. This suggests that while the relatively low proportion of genome-wide significant SNPs explain an outsized portion of the variation in birth year, there still exists a significant amount of signal in the remaining suggestive SNPs.

Table 5. Genetic variation in birth year explained by significant GPSM SNPs from GREML analysis.

Breed	Significant SNP PVE (se)	Non-significant PVE (se)	Total 2-GRM PVE (se)	Single GRM PVE (se)
Red Angus	0.168 (0.234)	0.437 (0.016)	0.604 (0.014)	0.558 (0.011)
Simmental	0.232 (0.020)	0.418 (0.014)	0.650 (0.012)	0.595 (0.010)
Gelbvieh	0.282 (0.023)	0.279 (0.015)	0.561 (0.017)	0.528 (0.014)

We performed a similar analysis with GRMs by further dividing the non-significant SNPs into "suggestive" and "other" classes (**Table 6**). To eliminate signal associated with genome-wide significant LD peaks, we removed variants within 1 Mb of significant markers for "suggestive" and "other" variant classes. We then built a suggestive-variant GRM with the 2,000 variants with the smallest remaining p-values, and an other-variant GRM with 2,000 random remaining variants with p > 0.5, but still segregating at a moderate frequency (MAF > 0.15) in each dataset. The other-variant GRM is intended to only include SNPs under drift. The resulting variance component estimates show that nearly all of the observed variation in birth year is accounted for by significant and suggestive SNP sets and likely is not being driven by stochastic changes in allele frequency due to drift. While our genome-wide significant variants account for a relatively small fraction of the genome, suggestive variants are evenly spread throughout the genome, consistent with our expectations when traits are under polygenic selection. Variants contributing to variance in birth year, but that don't reach genome-wide significance are likely undergoing quite small directional shifts in allele frequency, congruent with polygenic selection, that will be detectable as datasets increase in size.

Table 6. Genetic variation in birth year explained by three classes of GPSM significant SNPs.

Breed	Genome-wide Significant SNP PVE (se)	Suggestive significant SNP PVE (se)	Other SNP PVE (se)	Total PVE (se)
Red Angus	0.107 (0.017)	0.331 (0.014)	0.025 (0.003)	0.464 (0.015)
Simmental	0.166 (0.016)	0.339 (0.013)	0.028 (0.003)	0.533 (0.012)
Gelbvieh	0.162 (0.017)	0.297 (0.014)	0.000 (0.003)	0.465 (0.015)

While a clear function or nearby candidate gene existed for many of the observed GPSM hits, many of our lead SNPs were near, but not within genes (123, 242, 301 significant markers outside of, but within 100 kb of annotated gene in Red Angus, Simmental, and Gelbvieh, respectively). We assigned the closest gene to a significant marker as the "candidate gene." This generalization ignores the potential for a marker being in stronger LD with a different, more distant gene. With limited information describing the regulatory elements in the bovine genome [52], we were also unable to ascribe possible *trans* regulatory functions to these variants. Further, our 835K genotypes are unable to provide the base-pair resolution needed to further explore the actual effects of variants undergoing selection (frameshifts, loss-of-function, etc.).

Directional selection for production traits acts on similar genomic networks between breeds

Despite the multiple shared GPSM signatures, the majority of GPSM signatures detected are breed-specific (78.8%, 78.8%, and 77.2% of total significant regions in RAN, SIM, and GEL, respectively). While some of these signatures may be due to a slightly different selection emphasis placed on one breed, but not the others, it is more likely the result of selection on different genes or networks that control a similar suite of traits. Beef cattle have been historically selected for increased growth rates and reproductive efficiency, both of which are particularly complex traits, likely controlled by thousands of variants of varying effect sizes that act in both additive and maternal fashions [53,54]. We expect that the many significant GPSM regions will be tagging allele frequency changes driven by selection. That said, we observe relatively few of the regions identified by GPSM overlap with major growth QTL identified using identical Gelbvieh and Simmental datasets [55,56]. Most of the major growth-related QTL still segregating in these breeds are effectively unchanged over the course of our sampled datasets. Genes identified in previous growth GWAS analyses with these datasets such as *PLAG1* and *LCORL* increase the economically-relevant phenotypes weaning and yearling weights, but have antagonistic effect by increasing birth weights and calving difficulties [57]. As a result, even if variation exists at known, large-effect QTL, the allele frequency changes over the temporal scope of the data may be moderate across the population.

Candidate genes identified by GPSM in Simmental point towards selection on pathways involved in digestion and metabolic functions (Arachidonic acid metabolism, Gastric acid secretion, Salivary secretion, Carbohydrate digestion and absorption) that are likely due to increases in production efficiency [58–60]. In each breed, we identify numerous biological processes involved in cell cycle control under selection, a group of processes directly involved in muscle growth rate in cattle [61]. Additionally, in Red Angus and Gelbvieh we identify multiple cancer pathways under selection. This likely represents further evidence of selection on cell cycle regulation and growth rather than on actual cancer pathways.

While multi-population candidate genes account for some of these shared cell-cycle terms, (e.g. *PRDM9*, *RHOU*), significant pathways and GO terms are driven largely by breed-specific candidate genes.

In Simmental cattle we detect 23 GPSM candidate genes involved in olfactory transduction. While not related directly to growth traits, olfactory receptors have identified as selection targets in many mammalian species, including cattle [18,62,63]. This rapidly-evolving class of genes is also associated with growth and carcass traits, suggesting a wide range of functions that are not limited to detecting smell [35,64]. Other dataset-specific gene classes under selection highlight known within-breed characteristics. For example, Red Angus cattle are known to be a highly fertile breed with exceptional maternal characteristics [65]. We identify the "ovarian steroidogenesis" pathway under selection, a potential driver of hormone balance and female fertility [66,67]. We also identify numerous other processes involved in the production and metabolism of hormones. Hormone metabolism is a central regulator of growth in cattle [68–71], but could also represent selection for increased female fertility in the breed. Gelbvieh cattle are known for their quick growth rate and carcass quality. We identify 6 different biological processes related to muscle development and function from Gelbvieh GPSM gene sets.

Identifying genomic regions contributing to local adaptation

Beef cattle are one of the few livestock species in the United States whose production environments remain largely unaltered by humans. Cattle are distributed across almost every possible environment in the continental United States [72], and virtually none are raised in controlled confinement like pigs or chickens. Local adaptation and genotype-by-environment (GxE) interactions exist in closely related cattle populations [73,74]. The most common solution for breeding heat adapted cattle to date has been through crossbreeding *Bos taurus* and *Bos indicus* cattle [75]. While this has led to cattle better equipped to handle hot and humid environments, it has come at the cost of decreased meat quality [76,77]. Further, *Bos indicus* crossbreeding is unable to improve adaptation to a wide variety of other environmental stressors such as cold temperatures, high elevations, and a host of other biotic and abiotic factors. Ideally, we could exploit within breed variation to produce pure *Bos taurus* individuals suited to the full range of environments in the US without sacrificing performance for growth, maternal, or carcass traits. Previous work has identified the presence of extensive GxE in beef cattle populations [78–80], but limited work exploring the genomic basis of local adaptation has occurred [81].

Environmental selection has driven the development of locally adapted breeds and populations of cattle around the world. Well-adapted animals better express their genetic potential in their local environment, and thus are more likely to be selected as parents than their poorly-adapted counterparts. This leads to changes in allele frequency at loci that modulate adaptation in this local population compared to the full population. In the past 30 years, the use of artificial insemination has increased in U.S. beef cattle populations. This technology has allowed elite germplasm to be propagated ubiquitously across environments and has led to dramatic increases in production efficiency. However, as a consequence this has made breeds more homogeneous. We are interested in identifying whether detectable allele frequency differences exist between environments, and how human-imposed selection and mating has changed the magnitude of these differences.

Ecoregions and envGWAS

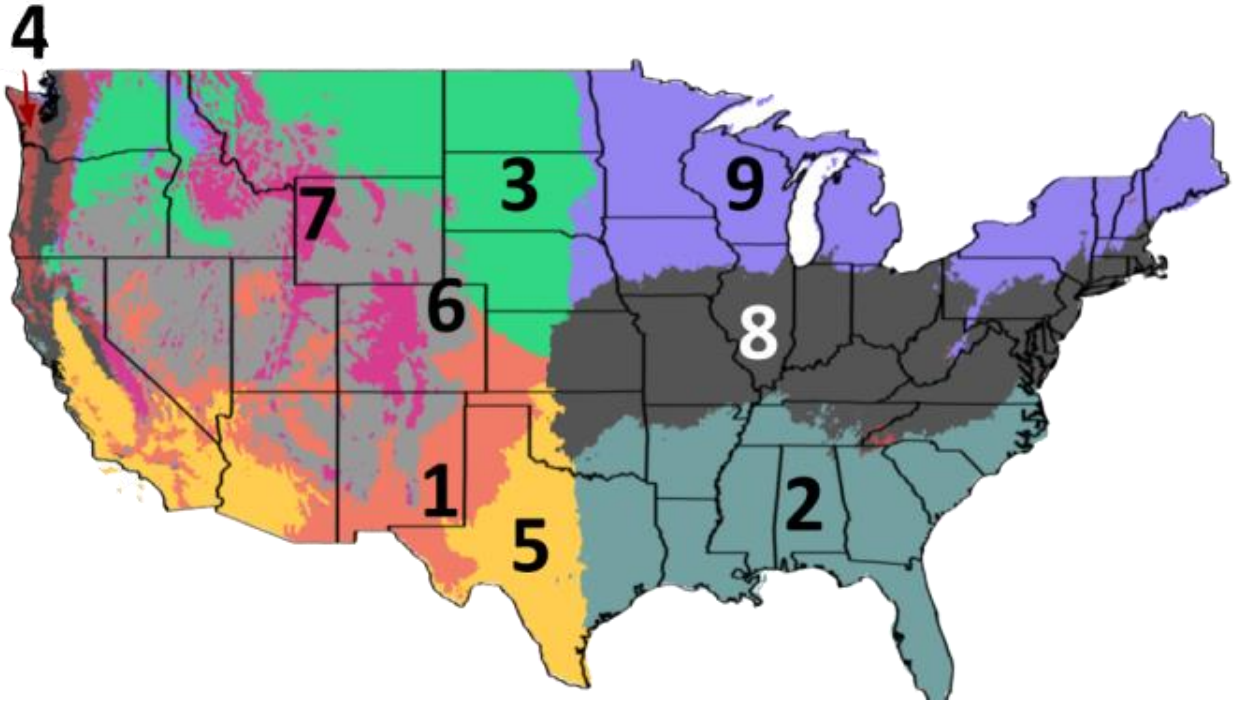

Figure 3. Nine discrete ecoregions of Continental United States. Each 1 km^2 was assigned to one of nine ecoregions based on a k-means clustering analysis of 30-year mean temperature, precipitation, and elevation from the PRISM Climate Dataset.

We used k-means clustering with 30-year normal values for temperature, precipitation, and elevation [82] to divide the United States into 9 clusters (**Figure 3**). These ecoregion assignments largely agree with previously-published maps in the environmetrics and atmospheric science literature [83,84], and reflect well-known differences in cattle production environments. For instance, the "Fescue Belt ', the region of the US where tall fescue (*Festuca arundinacea*) is the most common forage, is almost perfectly defined using only these three environmental variables. In using clustering to define ecoregions, we expect that we are capturing not only combinations of environmental variables, but the associated differences in forage type, local pathogens, and ecoregion-wide management differences that animals are exposed to.

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 true signatures of local adaptation. This method builds on the theory of the Bayenv approach from Coop et al. (2010) [85,86] that uses allele frequency correlations along environmental gradients to identify potential local adaptation. Using genome-wide LMMs allowed us to extend the Bayenv method to extremely large datasets (eventually hundreds of thousands of animals with millions of SNPs). A GWAS model in conjunction with a GRM controls for cryptic family and population structure that frequently obscures these studies.

Before performing any genomic analyses, we observe that the three analyzed breeds are not universally present in all ecoregions (**Figure 4 & Table 7**). This uneven regional distribution suggests that since the introduction of these breeds in the late 1960s and early 1970s, registered seedstock animals have a small footprint in desert regions with extreme temperatures and minimal rainfall. While seedstock operations exist at a wide range of elevations and temperatures, they are mostly absent from dry environments with low forage-availability (Desert ecoregion). Our assumption is that similar trends, though less extreme, would produce statistical signal in finer scale environments.

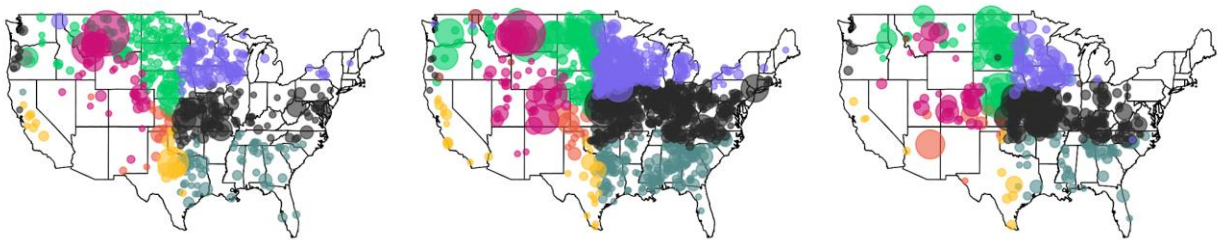

Figure 4. Geographic distributions of (A) Red Angus, (B) Simmental, and (C) Gelbvieh across the Continental United States. Each point is a zip code, sized by number of genotyped animals at that zip code and colored by the ecoregion that zip code resides in.

Table 7. Counts of analyzed individuals in each region for each dataset after filtering and region assignment based on individual's breeder zip code.

Region	Red Angus	Simmental	Gelbvieh
Desert	367[1]	321[1]	408[1]
Southeast	615[1]	1,073[1,2]	422[1]
High Plains	2,796[1,2]	3,645[1,2]	4,022[1,2]
Rainforest	0	62	1
Arid Prairie	1,206[1,2]	181	87
Foothills	136	0	0
Forested Mountains	4,525[1,2]	2,589[1,2]	704[1,2]
Fescue Belt	3,011[1,2]	4,393[1,2]	4,482[1,2]
Upper Midwest & Northeast	1,513[1,2]	2,524[1,2]	1,072[1,2]
Total	**14,169**	**14,788**	**11,198**

[1] included in multivariate discrete envGWAS analysis
[2] included in "large region" multivariate envGWAS analysis

Although environmental variables and regions are not inherited, the estimated Proportion Variation Explained (PVE) provides a measure of the genome-wide variation associated with environment and the degree to which cattle families are distributed across the United States. In Red Angus, the proportion of variation explained by SNPs (PVE) was 0.597, 0.526, and 0.594 for temperature, precipitation, and elevation respectively. PVEs for ecoregion identity ranged from 0.463 for the Arid Prairie to 0.673 for the Fescue Belt. These measures suggest that genetic similarity exists along both continuous environmental gradients and within discrete ecoregions. Genetic and phenotypic correlations for all three datasets are reported in **Tables 8 and 9**. Despite this genetic signal existing along environmental gradients and within ecoregions, regional relationships does not create observable population structure in the first 10 principal components (**Supplementary File 2**).

Table 8. Univariate REML estimates of PVE for continuous environmental variables in genotyped Red Angus, Simmental, and Gelbvieh cattle populations. Standard errors for PVE estimates are reported in parentheses.

Variable	Red Angus PVE (se)	Simmental PVE (se)	Gelbvieh PVE (se)
Temperature	0.597 (0.010)	0.586 (0.011)	0.691 (0.010)
Precipitation	0.526 (0.011)	0.602 (0.011)	0.677 (0.010)
Elevation	0.594 (0.010)	0.585 (0.011)	0.644 (0.011)

Table 9. Univariate estimates of PVE for discrete ecoregion assignment in genotyped Red Angus, Simmental, and Gelbvieh cattle populations. Standard errors for PVE estimates are reported in parentheses.

Variable	Red Angus PVE (se)	Simmental PVE (se)	Gelbvieh PVE (se)
Desert	0.646 (0.010)	0.517 (0.012)	0.726 (0.010)
Southeast (SE)	0.408 (0.010)	0.547 (0.013)	0.478 (0.013)
High Plains (HP)	0.641 (0.011)	0.588 (0.010)	0.694 (0.010)
Arid Prairie (AP)	0.463 (0.011)	0.566 (0.014)	NA
Forested Mountains (FM)	0.575 (0.011)	0.594 (0.010)	0.615 (0.013)
Fescue Belt (FB)	0.673 (0.010)	0.545 (0.011)	0.649 (0.011)
Upper Midwest & Northeast (UMWNE)	0.548 (0.012)	0.509 (0.012)	0.609 (0.013)

envGWAS detects allele frequency differences between discrete ecoregions

We hypothesized that using discrete ecoregions as the dependent variables genome-wide linear mixed models (discrete envGWAS) would identify different genomic regions than the same analysis using continuous environmental variables (continuous envGWAS). Using nine regions defined with k-means clustering of environmental data, we grouped individuals into environmental cohorts and fit univariate and multivariate case-control envGWAS models to identify allele frequency differences between environments that may be the result of local adaptation. The use of discrete ecoregion assignments as phenotypes allowed us to identify associations with entire environments as opposed to individual environmental variables. In addition to their shared climates, locations within these regions are more likely to share common forage sources, parasites, diseases, and management practices, providing a more complete representation of shared environment. We performed univariate discrete envGWAS in Red Angus, Simmental, and Gelbvieh datasets for regions with > 250 individuals. We performed two separate multivariate analyses, one that included all regions with > 250 individuals, and another more restrictive analysis that only used regions with > 700 individuals.

In Red Angus, we identified 54 variants significantly associated (significance threshold $p < 1x10^{-5}$ determined from permutation) with membership of an ecoregion in the discrete multivariate envGWAS analysis, encompassing 18 total genomic regions (**Figure 5A**). Of these regions, only two overlapped with regions identified in continuous envGWAS, suggesting that using different definitions of environment with envGWAS may detect different sources of adaptation. Of the 18 significant regions, 17 were within or near (< 100 kb) a candidate gene. Known functions of these genes are diverse. 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 [87]. Alleles within *CUX1* can be used to differentiate between breeds of goats raised for meat versus those raised for fiber [88]. The role of CUX1 in hair coat phenotypes makes it a prime adaptive candidate in environments where animals are under heat, cold, or toxic fescue stress. Other candidate genes identified by envGWAS have been previously identified as targets of selection between breeds of cattle (*MAGI2*, *CENPP*), or in other species (*DIRC1* in humans, *GORASP2* in fish, *ADRB1* in dogs and lizards). Adaptive signatures shared between cattle and other species may point to shared biological processes that drive environmental adaptation. We expect that novel adaptive genes and pathways identified in these analyses could be of value to studies of adaptation in other species.

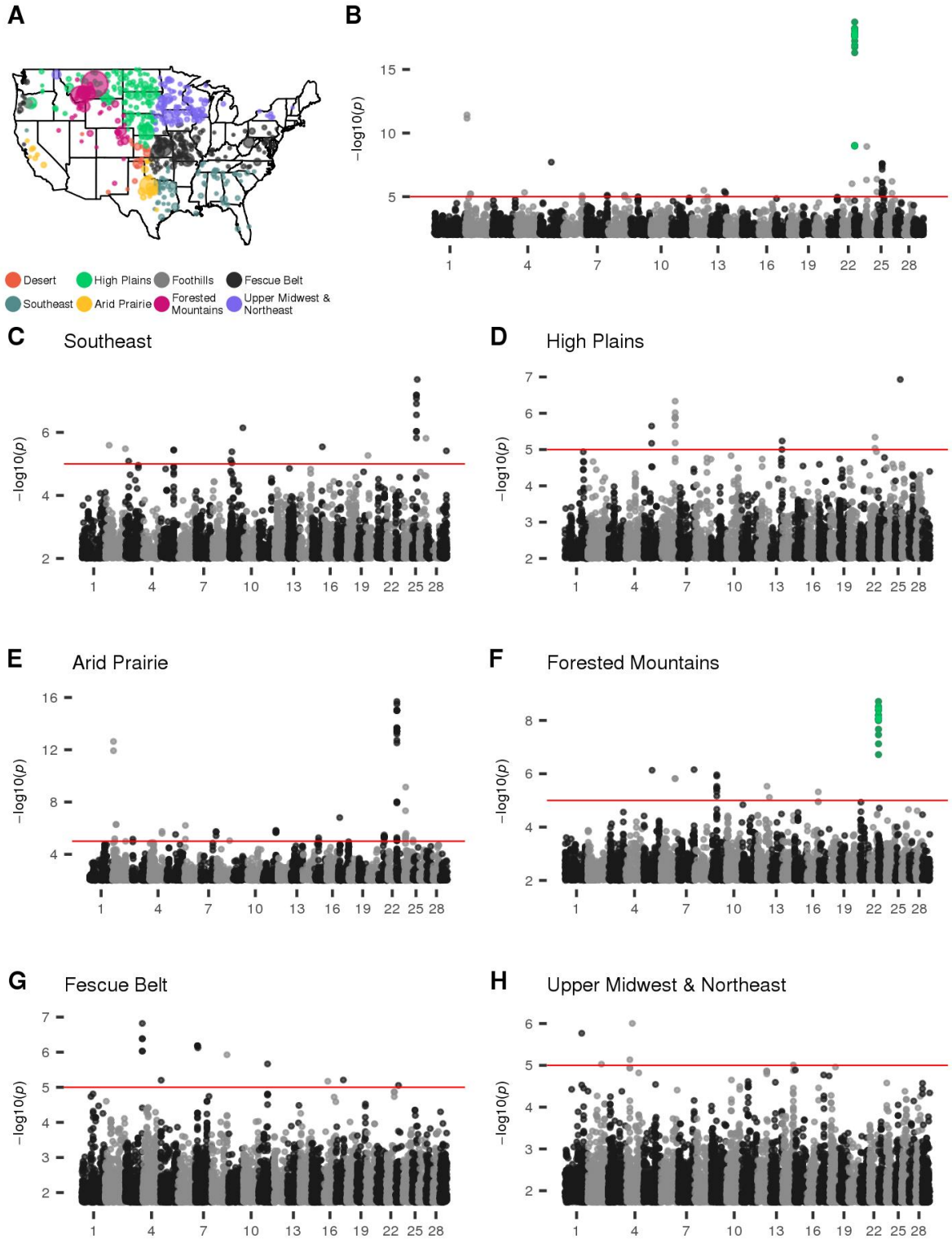

Figure 5. Manhattan plots of discrete envGWAS in Red Angus cattle. Locations of animals (A). Multivariate envGWAS (B), and univariate discrete envGWAS analysis for sufficiently-sized ecoregions (C-H). Green points were identified in the corresponding analysis in Simmental or Gelbvieh datasets.

From the Red Angus multivariate envGWAS, we identified four adaptive candidate genes directly involved in immune function (*RASGEF1B*, *SPN*, *ZMYND8*, *LOC100298064*/*HAVCR1*). GPSM analyses identified variants within or proximal to immune genes under ongoing selection across all three cattle populations (**Table 10**). These envGWAS regions suggest that genetic adaptations conferring resistance or tolerance to local pathogens and immune stressors may be as important as adaptations to more explicit environmental stressors (i.e. heat tolerance).

Table 10. Discrete multivariate envGWAS candidate genes for Red Angus. Lead SNP in envGWAS peak is reported along with nearest plausible candidate genes (provided < 250 kb from lead SNP).

CHR	POS	Nearest Candidate Gene(s) (Distance)	Univariate Continuous Association	Univariate Ecoregion Association	Candidate Gene Adaptive Associations	Reference	Breed(s)
2	7,571,508	DIRC1 (110.3 kb), COL5A2 (212.7 kb)	Multivariate, Temperature	FM	Sweep region (human), blood pressure (human)	[89,90]	RAN
2	25,519,381	GORASP2 (4.47 kb)	N/A	AP	Adaptive signature (fish)	[91]	RAN
4	43,928,337	MAGI2 (within)	N/A	UMWNE	Selection signature, imprinted (cattle)	[92]	RAN
5	59,498,938	LOC788524 (OR9K2) (1.42 kb) [Olfactory cluster]	N/A	FM	Olfactory receptor cluster, selection signature (bison)	[63]	RAN
6	96,217,506	RASGEF1B (within)	N/A	HP	Immune function, adaptation signature (human), Response to viral infections.	[93,94]	RAN
8	84,146,584	CENPP (within)	N/A	N/A	African cattle CNV, hypoxia (human)	[95,96]	RAN
12	58,438,020	N/A	N/A	N/A			RAN
13	75,784,164	ZMYND8 (within)			Immune function, DNA	[97,98]	RAN

					damage repair		
13	81,716,869	PFDN4 (71.97 kb)	N/A	HP	Dermatitis (humans)	[99]	RAN
22	48,873,180	DUSP7 (40.83 kb)			Heat stress/sperm motility (cattle), Stress response	[100,101]	RAN
23	1,768,070			AP, FM			RAN
24	10,628,280	CDH7 (151.7 kb)	MV	AP	Developmental processes	[102]	RAN
24	62,228,300	LOC100298064 (HAVCR1) (15.2kb)		AP	Immune function	[103]	RAN
25	26,511,450	SPN (CD43) (within)	N/A	SE	MHC-Class I, Immune response (cattle)	[104]	RAN
25	35,085,041	CUX1 (67.8 kb)	Multivariate	N/A	Hair phenotype (goats, mice)	[87,88]	
26	34,432,722	ADRB1 (81.5 kb)	Temperature		Selection (sporting dogs), climate adaptation (lizards)	[105,106]	RAN

In all three datasets, we identify a common local adaptation signature on chromosome 23 (peak SNP at 1,768,070 bp). Multivariate analyses in all three breeds identify alleles at this SNP to be significantly associated with a region or regions ($q = 1.24x10^{-13}$, $3.15x10^{-12}$, $4.82x10^{-5}$ in RAN, SIM, and GEL, respectively). In all three datasets, we identify this same SNP as a univariate envGWAS association with membership of the Forested Mountains ecoregion. Its most significant univariate association in the Red Angus dataset was with the Arid Prairie region which is excluded from both Simmental and Gelbvieh analyses due to low within-region sample size. In the Red Angus multivariate analysis, this peak spanned 18 SNPs from (1,708,914 bp to 1,780,836 bp) and contained the pseudogene *LOC782044*. This pseudogene, an annotated exon of *ALKBH8* (alkylated DNA repair protein alkB homolog 8-like), has no previously reported function or associations in cattle. The nearest annotated gene is *KHDRBS2* (KH RNA Binding Domain Containing, Signal Transduction Associated 2), which has been identified by other adaptation studies in cattle, sheep, and pigs [107–109]. Interestingly, this variant was not significantly associated with a single environmental variable in any analysis, suggesting that its adaptive benefits are unique to a particular combination of ecoregions or environmental variables. The use of discretely defined environments increases our power to detect alleles associated with a particular combination of

environmental variables. This lead SNP was also significantly associated with birth year in all three GPSM analyses. Selection for this variant in the Forested Mountains ecoregion, our most highly populated region in all three breeds, may be creating this population-wide signal.

envGWAS detects variants segregating along continuous environmental gradients

Using continuous temperature, precipitation, and elevation as quantitative phenotypes in a multivariate envGWAS analysis of Red Angus animals, we identify 21 significant environmental associations ($q < 0.1$) and an additional 25 loci with p-values below our empirically-derived cutoff ($p < 1 \times 10^{-5}$) (**Figure 6**). These loci tag 17 genomic regions, many of which have clear candidate genes nearby (< 100 kb). The most significant genomic region identified is located on chromosome 29 is located within *BBS1* (Bardet-Biedl syndrome 1), a gene directly involved in energy homeostasis in mice and humans [110,111]. *BBS1* mutant knock-in mice show irregularities in photoreceptors and olfactory sensory cilia [112]. Many of these environmental variable-associated genes have been previously identified as being adaptive or under selection in either human or cattle populations (**Table 11**). From the multivariate envGWAS in Red Angus, 9 of the 16 positional candidate genes have been identified by previous studies of adaptation or selection in humans or cattle (*DIRC1*, *ABCB1*, *TBC1D1*, *AP5M1*, *GRIA4*, *LRRC4C*, *RBMS3*, *GADL1*, *PLA2G12B*) [89,113,114]. For example, *GRIA4* has been identified as a target of selection in cold-adapted Yakutian cattle populations [115]. Of the other 7 candidate genes, three have been previously associated with adaptive phenotypes; body size (*E2F7*, *BBS1*) and circadian rhythm (*ADCYAP1*).

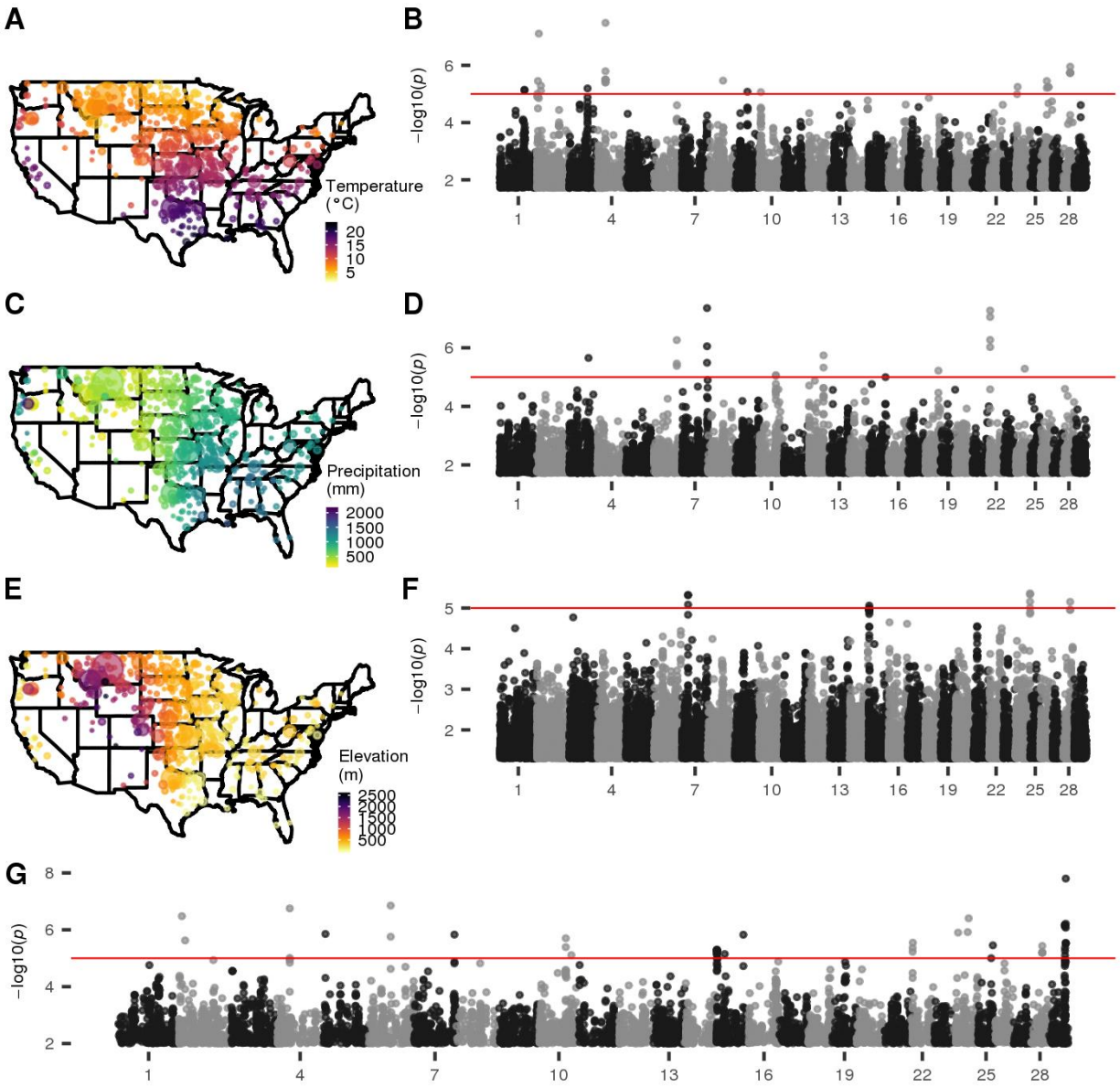

Figure 6. Continuous environmental variable envGWAS in Red Angus cattle. Geographic distributions colored by temperature (A), precipitation (C) , and elevation (E). Manhattan plots for univariate envGWAS analyses of temperature (B), precipitation (D), elevation (F), and a multivariate envGWAS of temperature, precipitation, and elevation (G). Red lines indicate permutation-derived p-value cutoff of $1x10^{-5}$.

Table 11. Candidate genes identified in multivariate envGWAS analyses using continuous environmental attributes as dependent variables. Chromosome and genomic positions are for lead SNP in peak. Closest gene is identified as candidate (if < 250 kb from lead SNP).

Chr	Position of lead SNP	Nearest Annotated Gene (Distance)	Univariate association (if $p < 1x10^{-5}$)	Adaptation-related associations	References	Breed(s)
2	7,571,508	DIRC1 (110.3 kb), COL5A2 (212.7 kb)	Temperature	Sweep region (human), blood pressure (human)	[89,90]	RAN
2	16,014,918	No genes < 200 kb	Temperature			RAN
4	32,945,494	ABCB1 (within)	Temperature	Drug resistance, human adaptation, cattle health traits	[116,117]	RAN
5	6,551,121	E2F7 (68.592 kb)		Body weight, bone density (human)	[118,119]	RAN
6	57,392,860	TBC1D1 (within)		Selection (cattle, chickens), body size (chickens, mice), immune traits (lymphocytes, etc.), obesity (humans)	[18,120–124]	RAN
7	106,527,874	EFNA5 (within)	Precipitation			RAN
10	69,692,497	AP5M1 (within)		Selection (humans)	[125]	RAN
10	84,307,302	DPF3 (within)	Precipitation			RAN
15	2,232,970	GRIA4 (within)	Elevation	Cold tolerance (cattle)	[115]	RAN
15	71,402,156	LRRC4C (within)		Altitude adaptation (humans)	[126]	RAN
22	4,815,225	RBMS3 (156.7 kb)	Precipitation	Tropical adaptation (humans), Sweep region	[89,127,128]	RAN

				(humans), Pleiotropic QTL (cattle)		
22	5,297,061	GADL1 (within)		Associated with climate variables in Mediterranian cattle, blood metabolites, human adaptation	[129–131]	RAN
24	10,628,280	CDH7 (150.58 kb)				RAN
24	35,718,635	ADCYAP1 (14.3 kb)		Circadian rhythm (birds), chronotype (humans)	[132–134]	RAN
25	35,085,041	CUX1 (67.8 kb)		Hair phenotypes (goats, mice)	[87,88]	RAN
25	39,464,325	LOC101904513	Precipitation	ncRNA		RAN
28	28,979,090	PLA2G12B (16.0 kb)	Temperature and Elevation	Local adaptation (humans)	[114,135]	RAN
29	44,555,972	BBS1 (within)		Obesity, energy homeostasis (human)	[110–112]	RAN
1	26,621,249	ROBO1 (within)		Neuron development (dogs, cattle, pigs),selection signature (cattle), sporting dogs, temperature acclimation (pigs)	[19,105,136,137]	SIM
1	134,777,473	CEP63 (within)		Height (human), thermal adaptation (fish)	[138,139]	SIM
3	2,966,062	UCK2 (18.65 kb)		Osmoregulation (fish), disease response (cattle)	[140,141]	SIM
8	64,286,414	ENSBTAG0000005		lncRNA		SIM

		4262				
10	16,366,200	KIF23 (24.83 kb)		Hepatic function (cattle)	[142,143]	SIM
17	51,685,973	LOC100847522 (88.07 kb)		ncRNA		SIM
20	54,365,387		Precipitation, Elevation			SIM
23	1,768,070	LOC782044	Precipitation, Elevation			SIM
26	34,125,126	NRAP (within)		Meat traits (cattle), under selection in Eastern Finncattle, neuron development	[144–146]	SIM
28	640,998	RHOU (56.34 kb)	Elevation			SIM
29	36,766,544	LOC112444895 (137.67 kb)	Precipitation, Elevation	ncRNA		SIM
2	116,117,760	SPHKAP (within)	Elevation	Insulin secretion, kidney disease susceptibility (human)	[147,148]	GEL
3	65,627,833	ADGRL4 (39.34 kb)	Temperature			GEL
4	35,457,854	SEMA3D (within)		Calving ease, neuron/axon guidance	[149–151]	GEL
4	95,515,907	LOC785077 (30.81 kb)	Precipitation			GEL
11	81,911,781	FAM49A (26.09 kb)	Temperature			GEL
14	73,663,377	CALB1 (within)	Elevation	Selection signature (pigs)	[152]	GEL
23	1,760,296	LOC782044 (17.22 kb)				GEL

25	663,051	LOC531296 (within)/MSLN	Temperature	Feed intake (cattle)	[49]	GEL

While there was minimal candidate gene overlap between envGWAS signatures in the other breeds, we do identify similar overarching themes, and similarities in the pathways and biological processes under selection. In Simmental, we identify two genes, *ROBO1* (Roundabout homolog 1) and *NRAP* (Nebulin Related Anchoring Protein) involved in neuron development and guidance [136,146]. Variants in and near the *ROBO1* have been identified as being differentially selected between breeds of cattle [19] and dogs [105]. Variants of *NRAP* were identified as being potentially adaptive in selection scans of Eastern Finncattle, a cold-adapted breed of cattle from Finland [145]. A genomic region within *SEMA3D* (Semaphorin 3D), another gene involved in axonal guidance, was significantly associated with continuous environmental variables in the Gelbvieh dataset [151,153]. Other candidate genes associated with continuous environments are reported with their potentially adaptive functions in (**Table 11**).

envGWAS identifies adaptive pathways and processes

Local adaptation is likely highly complex and controlled by many areas of the genome. Though we detected minimal overlap in candidate genes across datasets, we identified multiple conserved biological processes and pathways that appear to play roles in local adaptation across populations.

Though there was minimal candidate gene overlap between continuous and discrete envGWAS (19 genes in common in Red Angus), we identified many shared pathways and gene ontologies. In most cases where we observed GO or pathway overlap, statistical significance was greater for the discrete envGWAS analysis, simply due to the difference in number of provided genes (168 vs 38). Most of the shared terms and pathways were driven by the genes *BAD*, *EFNA5*, *LRRC4C*, *MARK2*, *PLCB3*, and *PRKG2* detected in both analyses. In these overlapping terms, additional genes from the discrete zone envGWAS further supplemented the identified terms.

Table 12. Shared KEGG pathways identified by envGWAS. Enrichment analyses were performed within breed using

Pathway	Num. Genes	% genes	Term p-value (FDR Corrected)	Gene Names	Dataset
Axon guidance	6	3.409	0.028	[EFNA5, EFNB2, FYN, LRRC4C, SEMA6A, UNC5C]	RAN
Cholinergic synapse	4	3.540	0.050	[FYN, LOC529425, PLCB3, PRKACB]	RAN
Dopaminergic synapse	4	2.985	0.065	[GRIA4, LOC529425, PLCB3, PRKACB]	RAN
Glutamatergic synapse	4	3.540	0.050	[GRIA4, LOC529425, PLCB3, PRKACB]	RAN
Platelet	5	4.167	0.027	[FYN, LYN, PLCB3, PRKACB,	RAN

activation				PRKG2]	
Relaxin signaling pathway	4	2.963	0.066	[LOC529425, PLCB3, PRKACB, VEGFB]	RAN
Retrograde endocannabinoid signaling	5	3.226	0.040	[GRIA4, LOC529425, NDUFA12, PLCB3, PRKACB]	RAN
Cholinergic synapse	3	2.655	0.028	[CACNA1A, KCNQ5, PRKACA]	SIM
Dopaminergic synapse	3	2.239	0.031	[CACNA1A, MAPK10, PRKACA]	SIM
Glutamatergic synapse	2	1.770	0.066	[CACNA1A, PRKACA]	SIM
Platelet activation	2	1.667	0.068	[COL1A1, PRKACA]	SIM
Relaxin signaling pathway	3	2.222	0.029	[COL1A1, MAPK10, PRKACA]	SIM
Retrograde endocannabinoid signaling	3	1.935	0.033	[CACNA1A, MAPK10, PRKACA]	SIM
Dopaminergic synapse	2	1.667	0.068	[CACNA1A, PRKACA]	SIM
Axon guidance	3	1.705	0.073	[PLXNA4, SEMA3D, SEMA3E]	SIM
Cholinergic synapse	3	2.655	0.037	[GNG13, ITPR2, KCNJ12]	GEL
GABAergic synapse	2	2.247	0.082	[ABAT, GNG13]	GEL
Glutamatergic synapse	2	1.770	0.104	[GNG13, ITPR2]	GEL
Platelet activation	2	1.667	0.110	[ITPR2, PPP1R12A]	GEL
Serotonergic synapse	2	1.667	0.110	[GNG13, ITPR2]	GEL

Across all three breeds, we consistently identified the “axon guidance” pathway, and numerous GO terms relating to axon development and guidance under region-specific selection. Ai et al. (2015) [137] 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 [154]. The direction and organization of axons is an essential component of the olfactory system which is frequently implicated in environmental adaptation through the recognition of local environmental cues. [89,155]. Other pathways identified across all three datasets include “cholinergic synapse”, “glutamatergic synapse”, and “platelet activation”. Cholinergic signaling drives cutaneous vascular responses to heat stress in humans [156–158]. Additionally, cholinergic receptors act as the major neural driver of sweating in humans [159].

Glutamatergic synapses are involved in neural vasoconstriction [160,161]. "Retrograde endocannabinoid signaling", "dopaminergic synapse", "GABAergic synapse", and "serotonergic synapse" pathways are also significantly enriched by envGWAS candidate genes. Nearly all of these important neural signaling pathways are also enriched in a series of gene-by-environment interaction GWAS for birth weight, weaning weight, and yearling weight in Simmental cattle [56]. Taken together, these results suggest that pathways involved in neuron development and neurotransmission are essential components of local adaptation in cattle. Temperature homeostasis is largely controlled by the central nervous system [162], making environment-specific selection on these pathways an efficient way for populations to adapt.

Other pathways identified by envGWAS appear to play important roles in vasodilation and vasoconstriction. Relaxin signalling was identified in both Red Angus and Simmental populations as a locally adaptive pathway under selection. Relaxin, initially identified as a pregnancy-related hormone, is an important modulator of vasodilation [163]. In Simmental this pathway association originates from local adaptation signatures near the genes *COL1A1*, *MAPK10*, and *PRKACA* identified both in our discrete multivariate envGWAS and in the Desert ecoregion univariate analysis. Relaxin signalling was also identified in Red Angus, but with four entirely different genes (*LOC529425*, *PLCB3*, *PRKACB*, *VEGFB*). While all four of the Red Angus genes were identified in multivariate analyses, we identify two of them in the Desert ecoregion univariate envGWAS. Vasodilation is an essential component of physiological temperature adaptation in cattle and other species [164–166]. The ability to mount a physiological response to heat stress has a direct impact on cattle performance. Heat stressed cattle have decreased feed intake, slower growth rates, and decreased fertility [167]. The Renin secretion pathway, which is also directly involved in vasoconstriction was identified in Red Angus (*ADRB1*, *PLCB3*, *PRKACB*, *PRKG2*), and has also been previously implicated in physiological responses to heat stress in cattle [168]. In each breed we identify multiple biological processes related to the regulation of insulin secretion. Insulin secretion is elevated in heat stressed cattle [169] and pigs [170], suggesting that it plays a role in metabolism and thermoregulation. Insulin secretion in response to the presence of glucose may also be related to different diets and forage availability along these continuous environmental gradients [171].

Other pathways identified by envGWAS candidate genes from Simmental point towards the immune system's role in local adaptation. "Th1 and Th2 cell differentiation" and "Th17 cell differentiation" were significant in a KEGG pathway analysis. The development of these cell types is essential for adaptive immune responses [172,173]. These signals were driven in part by region-specific allele frequency differences in or near *MAPK10* an innate immunity gene identified in human studies as an adaptive targets of selection [174]. We also identify multiple cardiac-related pathways from Simmental envGWAS genes ("Dilated cardiomyopathy", "Arrhythmogenic right ventricular cardiomyopathy", "Hypertrophic cardiomyopathy") driven by a pair of related genes *SGCA* and *SGCD*. Like other circulatory-related pathways, the alleles driving this signal were identified in both multivariate and Desert ecoregion envGWAS. We expect that these cardiac-related pathways, like renin secretion and relaxin signaling affect the efficiency of circulation and improved temperature homeostasis when exposed to heat or cold stress. A complete list of pathways and biological pathways identified by candidate genes from all envGWAS analyses are reported in **Supplementary File 2.**

Supplementary File 2. Significantly enriched (Benjamini-Hochberg FDR < 0.1) pathways and biological processes identified based on candidate genes < 10kb from genome-wide significant envGWAS SNPs.

https://docs.google.com/spreadsheets/d/1HP79hSM_PGObrN3ZIMc-vtLKaIKoKFgGU8w5HUlB2Sk/edit?usp=sharing

We identify minimal overlap between genomic regions identified by envGWAS across our three different datasets, but an enrichment analysis of pathways and biological processes identified considerable overlap across breeds. We expect that environment-specific selection in each breed is acting on largely different gene sets that are members of the same networks. This is especially encouraging as we work to breed genetically adapted cattle for a changing climate. We expect that considerable amounts of standing variation involved in these genomic networks exists in most major breeds of beef cattle. We expect that selection for well-adapted individuals is possible, and that the additional information from adaptive variants, genes, and pathways identified by envGWAS will make it even more effective.

envGWAS Summary

While numerous methods have been developed to identify local adaptation in populations, most rely on highly diverged populations inhabiting different environments [1,175]. The underlying theory of envGWAS is described by Coop et al. (2010) [85], but our implementation in a genome-wide LMM framework allows the theory of Bayenv to be extended to larger datasets with tens of thousands of individuals and hundreds of thousands to millions of SNPs. As opposed to comparing largely diverged populations, we used the entire spectrum of environmental variation in a polygenic framework to identify associations between genotypes and continuous and discrete environments. Using a genomic relationship matrix controls for false-positive signatures driven by familial relationships in a particular area. In a largely panmictic population, we still had adequate statistical power to detect multiple genomic regions and pathways that are likely involved in the genetic control of adaptation (**Table 12**).

Table 12. Counts of variants significantly associated with continuous environmental measures in univariate and multivariate continuous or discrete envGWAS analysis. Variants were identified as significant when $p < 1x10^{-5}$ (Empirically-derived significance threshold). Number of SNPs with $q < 0.1$ are reported in parentheses.

	Red Angus	**Simmental**	**Gelbvieh**
Temperature (univariate)	23 (2)	7 (0)	14 (0)
Precipitation (univariate)	17 (7)	20 (0)	6 (0)
Elevation (univariate)	10 (0)	36 (27)	12 (3)
Multivariate climate variable	**46 (21)**	**40 (36)**	**19 (1)**
Desert (univariate)	47 (30)	40 (19)	29 (4)
Southeast (univariate)	20 (8)	14 (2)	87 (87)
High Plains (univariate)	12 (1)	12 (0)	8 (0)
Arid Prairie (univariate)	62 (62)	48 (33)	NA
Forested Mountains (univariate)	29 (25)	36 (30)	37 (18)
Fescue Belt (univariate)	13 (8)	15 (0)	7 (0)

Upper Midwest/NE (univariate)	5 (0)	12 (0)	33 (19)
Multivariate zone	**176 (176)**	**39 (20)**	**66 (60)**

Previous work in cattle has identified loci that are involved in genotype-by-environment (GxE) interactions. GxE variants have different effects in different environments [55,56,176,177]. These approaches are different, but complementary to our envGWAS analysis. Instead of identifying interactions, envGWAS identifies direct associations between attributes of an individual's location and its genotype. If a favorable GxE variant is segregating in a region or along an environmental gradient, we may also identify it using envGWAS. It is also important to note that discussion of results from our genome scans have focused primarily on proximal candidate genes. This identification of candidate genes ignores the possibility significant variants may be acting in *trans* to change the expression of adaptive genes. Ongoing work to map functional elements of the bovine genome [52,178,179] coupled with sequence-level imputed genotypes will allow us to more accurately identify the important genes, pathways, and networks that contribute to local adaptation.

Ongoing work is incorporating informative adaptation SNPs into region-specific genomic predictions via genomic feature selection [180,181]. We anticipate that these predictions will produce significant rerankings of individuals across environments. These tools will empower producers to make environment-informed selection decisions to maximize efficiency.

Region-specific GPSM meta-analysis identifies variants actively undergoing region-specific selection

envGWAS detects allelic associations with continuous and discrete environment, but does not address whether local adaptation is ongoing, or being eroded by the increased exchange of germplasm between ecoregions. We used the spatiotemporal stratification of genotyped animals to identify genomic regions undergoing region-specific selection. This approach combines our ecoregions and GPSM in a meta-GWAS framework. This approach draws on the GxE interpretation of metaGWAS analyses from Kang et al. (2014) [182], where a heterogeneity of p-values at a locus between "treatments", or in our case ecoregions, is indicative of a treatment-specific association. This analysis identified variants undergoing region-specific changes in allele frequency. If selection on locally adaptive variants is ongoing, we would expect this analysis to identify variants actively diverging from the population mean in a particular ecoregion. The converse, an erosion of allele frequency differences back to the population mean, would suggest that selection for locally adaptive variants is not strong enough to overcome the influx of alleles from the outside population via artificial insemination.

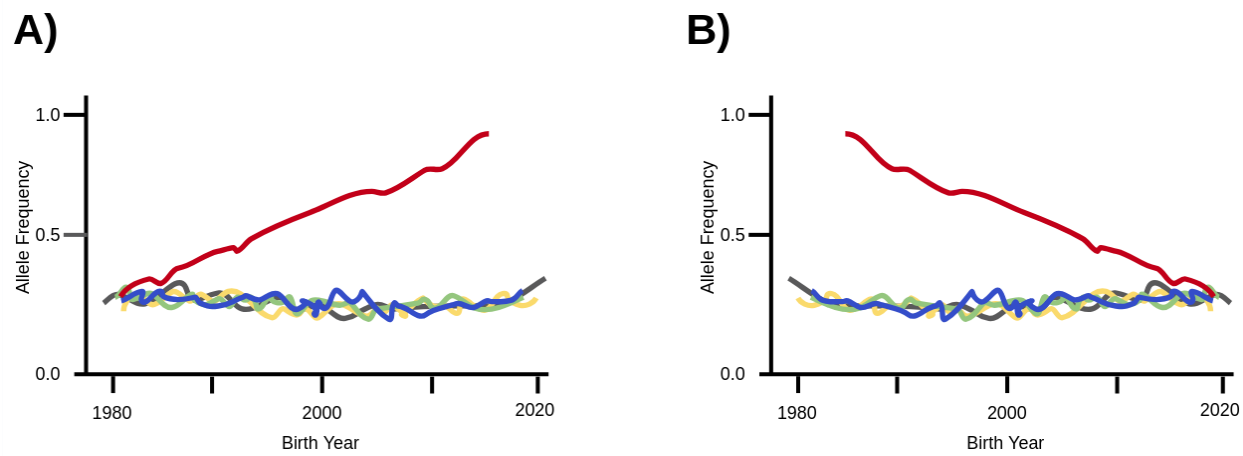

Figure 7. Allele frequency trajectory schematic for two possible types of ecoregion-specific selection on a locus. Different colored lines represent the allele frequency change of a locus in different ecoregions. In each case the locus is under ecoregion-specific selection in only the red environment. A) Ongoing local adaptation. B) Dissipation of local adaptation.

We performed a meta-analysis in METASOFT [183] by incorporating individual GPSM analyses for each ecoregion with more than 1,000 sampled individuals. For each marker tested in the meta-analysis, a Cochran's Q value is assigned based on the heterogeneity of effects between ecoregion. Variants with high Q values are significantly associated with birth year and are undergoing different shifts in allele frequency in one (or a few) ecoregions. The test also generates within-region GPSM p-values, an across-study p-value optimized for detecting associations in contexts of increased heterogeneity [183], and within-region m-values, the posterior probability that an effect exists within that region (**Figure 7B**) [184].

Variants with significant Cochran's Q statistics ($p < 1x10^{-5}$) that were also significant in at least one within-region GPSM analysis ($p < 1x10^{-5}$) imply region-specific allele frequency change. These changes can be due either to selection for local adaptation, or locally different allele frequencies moving towards the population mean. We identified 59, 38, and 46 significant SNPs in Red Angus, Simmental, and Gelbvieh datasets, respectively representing 15, 21, and 26 genomic regions (> 1 Mb to nearest other significant SNP) (**Figure 8A**). In all cases where we detect region-specific selection at a locus, it represents changes that make the ecoregion's allele frequency appear more similar to the general population's (**Figure 8C**).

The major driver of this is artificial insemination, and the ability to use germplasm from genetically superior animals from outside of one's home region. The increase in artificial insemination in the Red Angus breed has increased nearly tenfold since 1980 (28,896 to 284,788) (based on annual domestic semen sales reported by National Association of Animal Breeders).

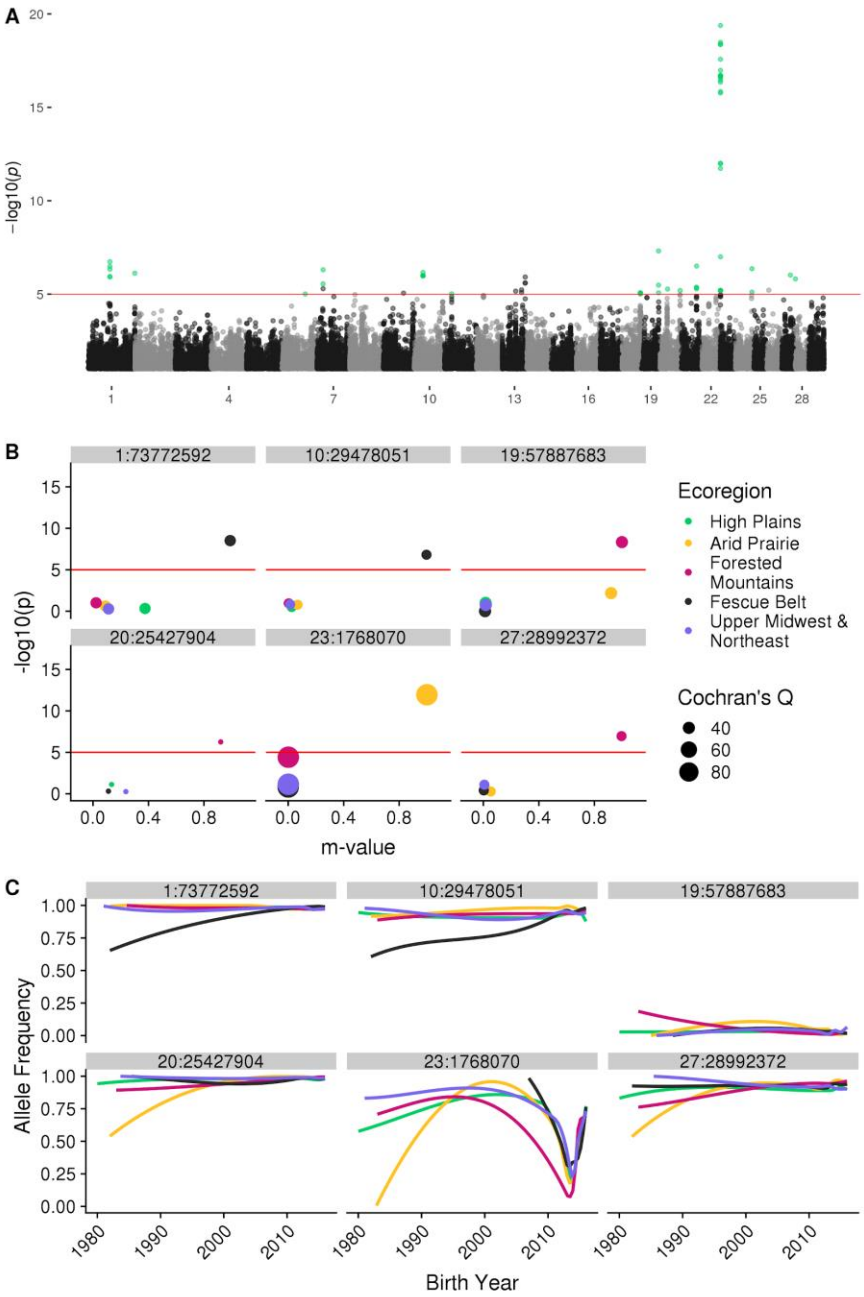

Figure 8. Meta-analysis of within-ecoregion GPSM for Red Angus cattle. A) Manhattan plot of per-variant Cochran's Q p-values. Points colored green had significant Cochran's Q ($p < 1x10^{-5}$) and were significant in at least one within-region GPSM analysis ($p < 1x10^{-5}$). B) Ecoregion effect plots for lead SNPs in six peaks from (A). Points are colored by ecoregion and sized based on Cochran's Q value. C) Region-specific allele frequency trajectories for SNPs from (B), colored by ecoregion.

Despite the apparent ongoing decay of local adaptation signatures, this meta-analysis on ecoregion-specific GPSM does identify multiple intriguing adaptive candidate genes. A peak on BTA 1 at ~73.8 Mb (lead snp rs254372) lies within the *OPA1* (OPA1 mitochondrial dynamin like GTPase) gene, and is under selection in the Fescue Belt ecoregion. *OPA1* has been implicated in circadian rhythm control in mice

[185,186], and is a known regulator of metabolic and cardiac adaptations [187,188] through its interactions with mitochondria. We also identify variants located within *ADAMTS16* (ADAM Metallopeptidase With Thrombospondin Type 1 Motif 16), a gene that has been identified in human adaptation studies, and that regulates blood pressure in mice [189,190]. Other genomic regions identified in the meta-analysis reside in candidate genes that do not have a clear adaptive function. This analysis is almost exclusively identifying the degradation of regional allele frequency differences. While loci identified in this analysis may be locally adaptive, they may also merely be artifacts of sparse sampling of AI sires in early generations.

CONCLUSIONS:

Using large, spatiotemporally distributed datasets from three United States beef cattle populations allowed us to begin mapping ongoing polygenic selection and possible locally adaptive variants. Using a proxy for generation number as the dependent variable in a genome-wide linear mixed model, we map widespread polygenic selection across the genome. This selection includes shared signatures of ongoing selection as well as significant amounts of breed specific selection. Our envGWAS analysis identifies subtle allele frequency differences along continuous environmental gradients and within discrete statistically-derived ecoregions. While envGWAS identifies largely different significant markers and candidate genes between datasets, numerous shared pathways related to adaptation exist. Generally, these pathways are related to neural development, neural signaling, and vasoconstriction, all known to affect temperature and energy homeostasis. While genetic differences exist between environments, it appears that local adaptation is being actively eroded in cattle populations through the use of artificial insemination. The use of artificial insemination has made massive improvements to the beef cattle population, but the effects of this migration overpower local adaptation. This further underlines the need for environment-aware predictions to exploit GxE and local adaptation.

METHODS:

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 square of the continental United States were extracted from the PRISM Climate Dataset [82], and used as continuous dependent variables in envGWAS analysis. Optimal k-means clustering of these three variables grouped each 8km square the continental US into 9 distinct ecoregions. Using 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.

Genotype Data:

SNP assays for three populations of genotyped *Bos taurus* beef cattle ranging in density from ~25,000 SNPs to ~770,000 SNPs were imputed to a common set of 830K SNPs using the large multi-breed imputation reference panel [196]. Map positions originated from the ARS-UCD1.2 reference genome (Rosen et al. 2020). Genotypes underwent quality control in PLINK (v1.9) [197], reference-based phasing with Eagle (v2.4) [198], and imputation with Minimac3 (v2.0.1) [199]. Following imputation, all three datasets contained 836,118 autosomal variants. All downstream analyses used only variants with minor allele frequencies > 0.01.

Exploratory Analyses:

We performed a principal component analysis using SMARTPCA from the EIGENSOFT suite [200]. (**Supplementary file ##**)

Generation Proxy Selection Mapping (GPSM):

To identify alleles changing in frequency over time we fit a univariate genome-wide linear mixed model (LMM) using GEMMA (Version 0.98.1) [201]. Here, we used the model:

EQUATION 1:

$$y = Xg + Zu + e$$
$$g \sim N(0, G\sigma_a{}^2)$$
$$e \sim N(0, \sigma^2{}_e I)$$

Where **y** is an individual's age on the date of analysis. Here, continuous age served as a proxy for generation number from the beginning of the pedigree. No fixed effects were used in our model, but a vector of polygenic random effects, **u,** was used to control for relationships and population structure within each dataset. This was estimated using a standardized genomic relationship matrix, **G**, [21] and estimates of **λ** (the ratio of variance components) and **τ** (the residual variance). **X** was an incidence matrix that related SNPs to individuals and **g** was the estimated effect size for each marker. Using this model, we tested each SNP for an association with continuous birth year. We converted p-values to FDR corrected q-values and used a significance threshold cutoff of q < 0.1. We performed multiple additional negative-control analyses in each dataset by shuffling the age associated with each genotype to ensure that observed GPSM signals were real. This was performed ten times for each breed dataset. To visualize the change in allele frequency of loci undergoing the strongest selection, we divided additive genotypes (coded as 0, 1, or 2) of significant variants by 2 (to reflect allele frequencies as 0, 0.5, or 1), then fit a loess and simple linear regressions for age and allele frequency in R [202]. These results and others were visualized with ggplot2 [203]. For regions with sufficient sample sizes (> 700 individuals), we fit within-region GPSM models to identify alleles changing in frequency only in particular environments. Significant loci were compared across regions, and to the population-scale GPSM results from each dataset.

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 (v0.98) [201]. We fit three univariate envGWAS models that used 30-year normal temperature, precipitation, and elevation as dependent variables. These used an identical model to **EQUATION 1**, but used environmental values as the dependent variable (**y**) instead of birth year. We also fit a combined multivariate model using all three environmental variables to account for shared genetic associations between temperature, elevation, and precipitation. 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 variation explained (PVE), phenotypic correlations, and genetic

correlations were estimated for continuous environmental variables and discrete environmental regions in two separate multivariate REML analyses in MTG2 (Lee and van der Werf 2016).

To ensure that envGWAS signals were not driven by spurious associations, we performed two types of permuted analyses. In the first, we shuffled the environmental variables and regions associated with an individual prior to performing each envGWAS analysis, completely detaching the link 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 shuffled the environmental variables associated with each zip code prior to envGWAS analysis. These two types of shuffled analysis were performed in each dataset and for each type of univariate and multivariate envGWAS analysis.

GPSM meta-analyses:

To identify variants undergoing frequency changes within particular ecoregions, we performed GPSM analyses within each region with more than 700 individuals (**Table 7**). The outputs from within-region GPSM analyses were combined into a single meta-analysis for each breed using METASOFT [183]. We report p-values for both the overall significance, which should be approximately equivalent to the initial GPSM analysis, as well as within-region p-values. We identified loci with high heterogeneity in p-values, indicative of possible region-specific selection. An m-value was calculated for each of these loci. In this context, m-values indicated the probability that a variant is changing in frequency more significantly in one region compared to others [184].

Pathway and biological process enrichment analysis:

Using the NCBI annotations for the ARS-UCD1.2 *Bos taurus* reference assembly (Rosen et al. 2020), we located proximal candidate genes near significant SNPs from each of our analyses. We identified candidate gene lists for all genes located +/- 10 kb from significant envGWAS SNPs. We consolidated significant SNPs from all envGWAS analyses into a single candidate gene list for each breed. Using these candidate gene lists, we performed gene ontology (GO) enrichment analysis using ClueGO (v2.5.5) [205] implemented in Cytoscape (v3.7.2). [206]. Using a *Bos taurus* gene list, we identified GO terms where at least two members of our candidate gene list made up at least 1.5% of the term's total genes. We applied a Benjamini-Hochberg multiple-testing correction to reported p-values. GO terms with FDR-corrected p-values < 0.1 were considered significant.

Birth year variance component analysis

To estimate the amount of variation in birth year explained by SNPs significant in GPSM, we performed multi-GRM GREML analyses for birth year in GCTA (v1.92.4) [207]. We built separate GRMs using genome-wide significant markers and all remaining makers outside of GPSM peaks (> 1Mb from significant GPSM markers). To further partition the variance in birth year explained by subsets of markers, we performed a GREML analysis using created GRMs with genome-wide significant ($p < 1x10^{-5}$), the next 2,000 lowest p-values, and 2,000 random unassociated ($p > 0.5$) markers.

1. Gutiérrez-Gil, B., Arranz, J. J. & 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.* **6**, 167 (2015).
2. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–15 (2010).
3. Hernandez, R. D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
4. Beissinger, T. M. *et al.* A genome-wide scan for evidence of selection in a maize population under long-term artificial selection for ear number. *Genetics* **196**, 829–840 (2014).
5. Berg, J. J. *et al.* Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* **8**, (2019).
6. Höllinger, I., Pennings, P. S. & Hermisson, J. Polygenic adaptation: From sweeps to subtle frequency shifts. *PLoS Genet.* **15**, e1008035 (2019).
7. Matukumalli, L. K. *et al.* Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* **4**, e5350 (2009).
8. Wright, S. THE GENETICAL STRUCTURE OF POPULATIONS. *Ann. Eugen.* **15**, 323–354 (1949).
9. Bonhomme, M. *et al.* Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics* **186**, 241–262 (2010).
10. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
11. Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
12. Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
13. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
14. Ramey, H. R. *et al.* Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics* **14**, 382 (2013).
15. Qanbari, S. *et al.* Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* **10**, e1004148 (2014).
16. Rothammer, S., Seichter, D., Förster, M. & Medugorac, I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genomics* **14**, 908 (2013).
17. Utsunomiya, Y. T. *et al.* Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS One* **8**, e64280 (2013).
18. Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D. P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* **47**, 49 (2015).
19. Boitard, S., Boussaha, M., Capitan, A., Rocha, D. & Servin, B. Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics* **203**, 433–450 (2016).
20. Decker, J. E. *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* **13**, 606 (2012).
21. VanRaden, P. M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **91**, 4414–4423

(2008).
22. Kuehn, L. A. & Thallman, R. M. Across-Breed EPD Tables For The Year 2016 Adjusted To Breed Differences For Birth Year Of 2014. (2016).
23. Lynch, M. & Walsh, B. Genetics and analysis of quantitative traits. *invemar.org.co*.
24. Wiedemar, N. *et al.* Independent polled mutations leading to complex gene expression differences in cattle. *PLoS One* **9**, e93435 (2014).
25. Mészáros, G., Petautschnig, E., Schwarzenbacher, H. & Sölkner, J. Genomic regions influencing coat color saturation and facial markings in Fleckvieh cattle. *Anim. Genet.* **46**, 65–68 (2015).
26. 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.* **41**, 295–303 (2010).
27. Canovas Nunes, S. *et al.* The small GTPase RhoU lays downstream of JAK/STAT signaling and mediates cell migration in multiple myeloma. *Blood Cancer J.* **8**, 20 (2018).
28. Zhang, J.-S., Koenig, A., Young, C. & Billadeau, D. D. GRB2 couples RhoU to epidermal growth factor receptor signaling and cell migration. *Mol. Biol. Cell* **22**, 2119–2130 (2011).
29. Berg, I. L. *et al.* PRDM9 variation strongly influences recombination hot-spot activity and meiotic instability in humans. *Nat. Genet.* **42**, 859–863 (2010).
30. Baudat, F. *et al.* PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836–840 (2010).
31. Ma, L. *et al.* Cattle Sex-Specific Recombination and Genetic Control from a Large Pedigree Analysis. *PLoS Genet.* **11**, e1005387 (2015).
32. Serão, N. V. *et al.* Single nucleotide polymorphisms and haplotypes associated with feed efficiency in beef cattle. *BMC Genet.* **14**, 94 (2013).
33. Snelling, W. M. *et al.* Genome-wide association study of growth in crossbred beef cattle. *J. Anim. Sci.* **88**, 837–848 (2010).
34. Mateescu, R. G., Garrick, D. J. & Reecy, J. M. Network Analysis Reveals Putative Genes Affecting Meat Quality in Angus Cattle. *Front. Genet.* **8**, 171 (2017).
35. Seabury, C. M. *et al.* Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* **18**, 386 (2017).
36. Noyes, H. *et al.* Genetic and expression analysis of cattle identifies candidate genes in pathways responding to Trypanosoma congolense infection. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9304–9309 (2011).
37. Smetko, A. *et al.* Trypanosomosis: potential driver of selection in African cattle. *Front. Genet.* **6**, 137 (2015).
38. Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I. & Goyache, F. African Cattle do not Carry Unique Mutations on the Exon 9 of the ARHGAP15 Gene. *Anim. Biotechnol.* **27**, 9–12 (2016).
39. Seo, J., Osorio, J. S. & Loor, J. J. Purinergic signaling gene network expression in bovine polymorphonuclear neutrophils during the peripartal period. *J. Dairy Sci.* **96**, 7675–7683 (2013).
40. Mediero, A. & Cronstein, B. N. Adenosine and bone metabolism. *Trends Endocrinol. Metab.* **24**, 290–300 (2013).
41. Weber, K. L. *et al.* Identification of Gene Networks for Residual Feed Intake in Angus Cattle Using Genomic Prediction and RNA-seq. *PLoS One* **11**, e0152274 (2016).

42. Alexandre, P. A. *et al.* Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics* **16**, 1073 (2015).
43. Müller, M.-P. *et al.* Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *Journal of Dairy Science* vol. 100 1987–2006 (2017).
44. Espigolan, R. *et al.* Associations between single nucleotide polymorphisms and carcass traits in Nellore cattle using high-density panels. *Genet. Mol. Res.* **14**, 11133–11144 (2015).
45. Derks, M. F. L. *et al.* Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* **15**, e1008055 (2019).
46. Meade, K. G. *et al.* Antigen stimulation of peripheral blood mononuclear cells from Mycobacterium bovis infected cattle yields evidence for a novel gene expression program. *BMC Genomics* **9**, 447 (2008).
47. Cassar-Malek, I., Boby, C., Picard, B., Reverter, A. & Hudson, N. J. Molecular regulation of high muscle mass in developing Blonde d'Aquitaine cattle foetuses. *Biol. Open* **6**, 1483–1492 (2017).
48. Höglund, J. K., Guldbrandtsen, B., Lund, M. S. & Sahana, G. Analyzes of genome-wide association follow-up study for calving traits in dairy cattle. *BMC Genet.* **13**, 71 (2012).
49. Chen, Y. *et al.* Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Anim. Genet.* **42**, 475–490 (2011).
50. Marete, A. G. *et al.* A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations. *Front. Genet.* **9**, 522 (2018).
51. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).
52. Andersson, L. *et al.* Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* **16**, 57 (2015).
53. Deese, R. E. & Koger, M. Maternal Effects on Preweaning Growth Rate in Cattle. *J. Anim. Sci.* **26**, 250–253 (1967).
54. Meyer, K. Variance components due to direct and maternal effects for growth traits of Australian beef cattle. *Livestock Production Science* vol. 31 179–204 (1992).
55. Smith, J. L. *et al.* Genome-wide association and genotype by environment interactions for growth traits in U.S. Gelbvieh cattle. *BMC Genomics* **20**, 926 (2019).
56. Braz, C. U., Rowan, T. N., Schnabel, R. D. & Decker, J. E. Extensive genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. doi:10.1101/2020.01.09.900902.
57. Saatchi, M., Schnabel, R. D., Taylor, J. F. & Garrick, D. J. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* **15**, 442 (2014).
58. Diez-Gonzalez, F., Callaway, T. R., Kizoulis, M. G. & Russell, J. B. Grain feeding and the dissemination of acid-resistant Escherichia coli from cattle. *Science* **281**, 1666–1668 (1998).
59. Bailey, C. B. Saliva secretion and its relation to feeding in cattle: 3.* The rate of secretion of mixed saliva in the cow during eating, with an estimate of the magnitude of the total daily secretion of mixed saliva. *Br. J. Nutr.* **15**, 443–451 (1961).
60. Nozière, P., Ortigues-Marty, I., Loncke, C. & Sauvant, D. Carbohydrate quantitative digestion and absorption in ruminants: from feed starch and fibre to nutrients available for tissues. *Animal* **4**, 1057–1074 (2010).

61. Guo, B. *et al.* Transcriptome analysis of cattle muscle identifies potential markers for skeletal muscle growth rate and major cell types. *BMC Genomics* **16**, 177 (2015).
62. Li, M. *et al.* Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **45**, 1431–1438 (2013).
63. Gautier, M. *et al.* Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Mol. Biol. Evol.* **33**, 2801–2814 (2016).
64. Magalhães, A. F. B. *et al.* Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. *PLoS One* **11**, e0157845 (2016).
65. Breeds - Red Angus. *The Cattle Site* https://www.thecattlesite.com/breeds/beef/99/red-angus/.
66. Spicer, L. J. & Echternkamp, S. E. Ovarian follicular growth, function and turnover in cattle: a review. *J. Anim. Sci.* **62**, 428–451 (1986).
67. Gareis, N. C. *et al.* Impaired insulin signaling pathways affect ovarian steroidogenesis in cows with COD. *Anim. Reprod. Sci.* **192**, 298–312 (2018).
68. Davis, S. L., Hossner, K. L. & Ohlson, D. L. Endocrine Regulation of Growth in Ruminants. in *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* (eds. Roche, J. F. & O'Callaghan, D.) 151–178 (Springer Netherlands, 1984).
69. Brockman, R. P. & Laarveld, B. Hormonal regulation of metabolism in ruminants; a review. *Livestock Production Science* **14**, 313–334 (1986).
70. Hocquette, J. F., Ortigues-Marty, I., Pethick, D., Herpin, P. & Fernandez, X. Nutritional and hormonal regulation of energy metabolism in skeletal muscles of meat-producing animals. *Livestock Production Science* **56**, 115–143 (1998).
71. Hocquette, J. F. Endocrine and metabolic regulation of muscle growth and body composition in cattle. *Animal* **4**, 1797–1809 (2010).
72. Drouillard, J. S. Current situation and future trends for beef production in the United States of America — A review. *Asian-Australasian Journal of Animal Sciences* vol. 31 1007–1016 (2018).
73. Burns, W. C., Koger, M., Butts, W. T., Pahnish, O. F. & Blackwell, R. L. Genotype by Environment Interaction in Hereford Cattle: II. Birth and Weaning Traits. *J. Anim. Sci.* **49**, 403–409 (1979).
74. Hohenboken, W., Jenkins, T., Pollak, J., Bullock, D. & Radakovich, S. Genetic improvement of beef cattle adaptation in America. in *Proceedings of the Beef Improvement Federation's 37th annual research symposium and annual meeting* vol. 37 115–120 (2005).
75. Hammond, A. C. *et al.* Heat tolerance in two tropically adapted Bos taurus breeds, Senepol and Romosinuano, compared with Brahman, Angus, and Hereford cattle in Florida. *J. Anim. Sci.* **74**, 295–303 (1996).
76. Crouse, J. D., Cundiff, L. V., Koch, R. M., Koohmaraie, M. & Seideman, S. C. Comparisons of Bos Indicus and Bos Taurus Inheritance for Carcass Beef Characteristics and Meat Palatability. *J. Anim. Sci.* **67**, 2661–2668 (1989).
77. Rodrigues, R. T. de S. *et al.* Differences in Beef Quality between Angus (Bos taurus taurus) and Nellore (Bos taurus indicus) Cattle through a Proteomic and Phosphoproteomic Approach. *PLoS One* **12**, e0170294 (2017).
78. Bradford, H. L., Fragomeni, B. O., Bertrand, J. K., Lourenco, D. A. L. & Misztal, I. Genetic evaluations for growth heat tolerance in Angus cattle. *J. Anim. Sci.* **94**, 4143–4150 (2016).

79. Fennewald, D. J., Weaber, R. L. & Lamberson, W. R. Genotype by environment interaction for stayability of Red Angus in the United States. *Journal of Animal Science* vol. 96 422–429 (2018).
80. Carvalheiro, R. *et al.* Unraveling genetic sensitivity of beef cattle to environmental variation under tropical conditions. *Genet. Sel. Evol.* **51**, 29 (2019).
81. Blackburn, H. D. *et al.* A fine structure genetic analysis evaluating ecoregional adaptability of a Bos taurus breed (Hereford). *PLoS One* **12**, e0176474 (2017).
82. PRISM Climate Group. PRISM 30-year Normal Climate Data.
83. Fovell, R. G. & Fovell, M.-Y. C. Climate Zones of the Conterminous United States Defined Using Cluster Analysis. *J. Clim.* **6**, 2103–2135 (1993).
84. Sathiaraj, D., Huang, X. & Chen, J. Predicting climate types for the Continental United States using unsupervised clustering techniques. *Environmetrics* vol. 30 e2524 (2019).
85. Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J. K. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411–1423 (2010).
86. Günther, T. & Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220 (2013).
87. Sansregret, L. & Nepveu, A. The multiple roles of CUX1: insights from mouse models and cell-based assays. *Gene* **412**, 84–94 (2008).
88. Bertolini, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* **50**, 57 (2018).
89. Williamson, S. H. *et al.* Localizing recent adaptive evolution in the human genome. *PLoS Genet.* **3**, e90 (2007).
90. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
91. Wang, G. *et al.* Gene expression responses of threespine stickleback to salinity: implications for salt-sensitive hypertension. *Front. Genet.* **5**, 312 (2014).
92. Barbaux, S. *et al.* A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics* **7**, 1079–1090 (2012).
93. Andrade, W. A. *et al.* Early endosome localization and activity of RasGEF1b, a toll-like receptor-inducible Ras guanine-nucleotide exchange factor. *Genes Immun.* **11**, 447–457 (2010).
94. Lopez, M. *et al.* Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. *Curr. Biol.* **29**, 2926–2935.e4 (2019).
95. Wang, M. D., Dzama, K., Hefer, C. A. & Muchadeyi, F. C. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC Genomics* **16**, 894 (2015).
96. Alkorta-Aranburu, G. *et al.* The genetic architecture of adaptations to high altitude in Ethiopia. *PLoS Genet.* **8**, e1003110 (2012).
97. Delgado-Benito, V. *et al.* The Chromatin Reader ZMYND8 Regulates Igh Enhancers to Promote Immunoglobulin Class Switch Recombination. *Mol. Cell* **72**, 636–649.e8 (2018).
98. Gong, F. *et al.* Screen identifies bromodomain protein ZMYND8 in chromatin recognition of transcription-associated DNA damage that promotes homologous recombination. *Genes Dev.* **29**, 197–211 (2015).
99. Hirota, T. *et al.* Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. *Nat. Genet.* **44**, 1222–1226 (2012).

100. Srikanth, K., Kwon, A., Lee, E. & Chung, H. Characterization of genes and pathways that respond to heat stress in Holstein calves through transcriptome analysis. *Cell Stress Chaperones* **22**, 29–42 (2017).
101. Rahman, M. B. *et al.* Altered chromatin condensation of heat-stressed spermatozoa perturbs the dynamics of DNA methylation reprogramming in the paternal genome after in vitro fertilisation in cattle. *Reprod. Fertil. Dev.* **26**, 1107–1116 (2014).
102. Aramaki, M. *et al.* Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome. *Birth Defects Res. A Clin. Mol. Teratol.* **79**, 50–57 (2007).
103. Nakajima, T. *et al.* Evidence for natural selection in the HAVCR1 gene: high degree of amino-acid variability in the mucin domain of human HAVCR1 protein. *Genes Immun.* **6**, 398–406 (2005).
104. McLoughlin, K. E. *et al.* RNA-seq Transcriptional Profiling of Peripheral Blood Leukocytes from Cattle Infected with Mycobacterium bovis. *Front. Immunol.* **5**, 396 (2014).
105. Kim, J. *et al.* Genetic selection of athletic success in sport-hunting dogs. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7212–E7221 (2018).
106. Rodríguez, A. *et al.* Genomic and phenotypic signatures of climate adaptation in an Anolis lizard. *Ecol. Evol.* **7**, 6390–6403 (2017).
107. León, C. D., De León, C., Manrique, C., Martínez, R. & Rocha, J. F. Research Article Genomic association study for adaptability traits in four Colombian cattle breeds. *Genetics and Molecular Research* vol. 18 (2019).
108. Guo, J. *et al.* Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci. Rep.* **8**, 10405 (2018).
109. Gurgul, A. *et al.* A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. *BMC Genet.* **19**, 95 (2018).
110. Guo, D.-F. *et al.* The BBSome Controls Energy Homeostasis by Mediating the Transport of the Leptin Receptor to the Plasma Membrane. *PLoS Genet.* **12**, e1005890 (2016).
111. Rouabhi, M., Guo, D. F. & Rahmouni, K. Bardet-Biedl Syndrome 1 Gene in the Ventromedial Hypothalamus is Required for Energy Homeostasis. *The FASEB Journal* **30**, 750.4–750.4 (2016).
112. Davis, R. E. *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.* **104**, 19422–19427 (2007).
113. Hussin, J., Nadeau, P., Lefebvre, J.-F. & Labuda, D. Haplotype allelic classes for detecting ongoing positive selection. *BMC Bioinformatics* **11**, 65 (2010).
114. Pemberton, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.* **91**, 275–292 (2012).
115. Igoshin, A. V. *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.* **20**, 26 (2019).
116. Wang, H. *et al.* Comparative and evolutionary pharmacogenetics of ABCB1: complex signatures of positive selection on coding and regulatory regions. *Pharmacogenet. Genomics* **17**, 667–678 (2007).
117. Wang, Z. *et al.* Signatures of recent positive selection at the ATP-binding cassette drug transporter superfamily gene loci. *Hum. Mol. Genet.* **16**, 1367–1380 (2007).
118. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum.*

Genet. **104**, 65–75 (2019).
119. Tachmazidou, I. *et al.* Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *Am. J. Hum. Genet.* **100**, 865–884 (2017).
120. Cardoso, D. F. *et al.* Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genet. Sel. Evol.* **50**, 22 (2018).
121. Rubin, C.-J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591 (2010).
122. Stone, S. *et al.* TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition. *Hum. Mol. Genet.* **15**, 2709–2720 (2006).
123. Chadt, A. *et al.* Tbc1d1 mutation in lean mouse strain confers leanness and protects from diet-induced obesity. *Nat. Genet.* **40**, 1354–1359 (2008).
124. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).
125. Duforet-Frebourg, N., Bazin, E. & Blum, M. G. B. Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Mol. Biol. Evol.* **31**, 2483–2495 (2014).
126. Wang, B. *et al.* On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS One* **6**, e17002 (2011).
127. Amorim, C. E. G., Daub, J. T., Salzano, F. M., Foll, M. & Excoffier, L. Detection of convergent genome-wide signals of adaptation to tropical forests in humans. *PLoS One* **10**, e0121557 (2015).
128. Granka, J. M. *et al.* Limited evidence for classic selective sweeps in African populations. *Genetics* **192**, 1049–1064 (2012).
129. Flori, L. *et al.* A genomic map of climate adaptation in Mediterranean cattle breeds. *Mol. Ecol.* **28**, 1009–1029 (2019).
130. Shin, S.-Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat. Genet.* **46**, 543–550 (2014).
131. Key, F. M., Fu, Q., Romagné, F., Lachmann, M. & Andrés, A. M. Human adaptation and population differentiation in the light of ancient genomes. *Nat. Commun.* **7**, 10775 (2016).
132. Mueller, J. C., Pulido, F. & Kempenaers, B. Identification of a gene associated with avian migratory behaviour. *Proc. Biol. Sci.* **278**, 2848–2856 (2011).
133. Bazzi, G. *et al.* Adcyap1 polymorphism covaries with breeding latitude in a Nearctic migratory songbird, the Wilson's warbler (Cardellina pusilla). *Ecol. Evol.* **6**, 3226–3239 (2016).
134. Jones, S. E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).
135. Sjöstrand, A. E., Sjödin, P. & Jakobsson, M. Private haplotypes can reveal local adaptation. *BMC Genet.* **15**, 61 (2014).
136. Dickinson, R. E. & Duncan, W. C. The SLIT-ROBO pathway: a regulator of cell function with implications for the reproductive system. *Reproduction* **139**, 697–704 (2010).
137. Ai, H. *et al.* Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
138. Wu, D.-D. & Zhang, Y.-P. Positive selection drives population differentiation in the skeletal genes in modern humans. *Hum. Mol. Genet.* **19**, 2341–2346 (2010).
139. Chen, Z. Physiological, transcriptomic and genomic mechanisms of thermal adaptation in

Oncorhynchus mykiss. (University of British Columbia, 2017). doi:10.14288/1.0340726.
140. do Prado, F. D. *et al.* Parallel evolution and adaptation to environmental factors in a marine flatfish: Implications for fisheries and aquaculture management of the turbot (Scophthalmus maximus). *Evol. Appl.* **11**, 1322–1341 (2018).
141. Twomey, A. J. *et al.* Genome-wide association study of endo-parasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. *Genet. Sel. Evol.* **51**, 15 (2019).
142. Li, R. W. & Li, C. Butyrate induces profound changes in gene expression related to multiple signal pathways in bovine kidney epithelial cells. *BMC Genomics* **7**, 234 (2006).
143. Ringseis, R., Zeitz, J. O., Weber, A., Koch, C. & Eder, K. Hepatic transcript profiling in early-lactation dairy cows fed rumen-protected niacin during the transition from late pregnancy to lactation. *J. Dairy Sci.* **102**, 365–376 (2019).
144. Williams, J. L. *et al.* Discovery, characterization and validation of single nucleotide polymorphisms within 206 bovine genes that may be considered as candidate genes for beef production and quality. *Anim. Genet.* **40**, 486–491 (2009).
145. Weldenegodguad, M. *et al.* Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions. *Front. Genet.* **9**, 728 (2018).
146. Lei, N., Mellem, J. E., Brockie, P. J., Madsen, D. M. & Maricq, A. V. NRAP-1 Is a Presynaptically Released NMDA Receptor Auxiliary Protein that Modifies Synaptic Strength. *Neuron* **96**, 1303–1316.e6 (2017).
147. Wang, Y., Harashima, S.-I., Liu, Y., Usui, R. & Inagaki, N. Sphingosine kinase 1-interacting protein is a novel regulator of glucose-stimulated insulin secretion. *Sci. Rep.* **7**, 779 (2017).
148. Cañadas-Garre, M. *et al.* Genetic Susceptibility to Chronic Kidney Disease - Some More Pieces for the Heritability Puzzle. *Front. Genet.* **10**, 453 (2019).
149. Purfield, D. C., Bradley, D. G., Evans, R. D., Kearney, F. J. & Berry, D. P. Genome-wide association study for calving performance using high-density genotypes in dairy and beef cattle. *Genet. Sel. Evol.* **47**, 47 (2015).
150. Ang, C. E. *et al.* The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife* **8**, (2019).
151. Lu, W.-C. *et al.* The protocadherin alpha cluster is required for axon extension and myelination in the developing central nervous system. *Neural Regeneration Res.* **13**, 427–433 (2018).
152. Yang, R. *et al.* Genome-wide analysis of structural variants reveals genetic differences in Chinese pigs. *PLoS One* **12**, e0186721 (2017).
153. Fujii, T. *et al.* Possible association of the semaphorin 3D gene (SEMA3D) with schizophrenia. *J. Psychiatr. Res.* **45**, 47–53 (2011).
154. Boulant, J. A. & Dean, J. B. Temperature receptors in the central nervous system. *Annu. Rev. Physiol.* **48**, 639–654 (1986).
155. Mombaerts, P. Axonal wiring in the mouse olfactory system. *Annu. Rev. Cell Dev. Biol.* **22**, 713–737 (2006).
156. Kellogg, D. L., Jr *et al.* Cutaneous active vasodilation in humans is mediated by cholinergic nerve cotransmission. *Circ. Res.* **77**, 1222–1228 (1995).
157. Kellogg, D. L., Jr *et al.* Cholinergic mechanisms of cutaneous active vasodilation during heat stress in cystic fibrosis. *J. Appl. Physiol.* **103**, 963–968 (2007).

158. Joyner, M. J. & Dietz, N. M. Sympathetic vasodilation in human muscle. *Acta Physiol. Scand.* **177**, 329–336 (2003).
159. Smith, C. J. & Johnson, J. M. Responses to hyperthermia. Optimizing heat dissipation by convection and evaporation: Neural control of skin blood flow and sweating in humans. *Auton. Neurosci.* **196**, 25–36 (2016).
160. Takemoto, Y. Amino acids that centrally influence blood pressure and regional blood flow in conscious rats. *J. Amino Acids* **2012**, 831759 (2012).
161. Meng, W., Tobin, J. R. & Busija, D. W. Glutamate-induced cerebral vasodilation is mediated by nitric oxide through N-methyl-D-aspartate receptors. *Stroke* **26**, 857–62; discussion 863 (1995).
162. Morrison, S. F. Central control of body temperature. *F1000Res.* **5**, (2016).
163. Conrad, K. P. Unveiling the vasodilatory actions and mechanisms of relaxin. *Hypertension* **56**, 2–9 (2010).
164. Daanen, H. A. M. & Van Marken Lichtenbelt, W. D. Human whole body cold adaptation. *Temperature (Austin)* **3**, 104–118 (2016).
165. Choshniak, I., McEwan-Jenkinson, D., Blatchford, D. R. & Peaker, M. Blood flow and catecholamine concentration in bovine and caprine skin during thermal sweating. *Comp. Biochem. Physiol. C* **71C**, 37–42 (1982).
166. Garner, J. B. *et al.* Genomic Selection Improves Heat Tolerance in Dairy Cattle. *Sci. Rep.* **6**, 34114 (2016).
167. Rhoads, M. L. *et al.* Effects of heat stress and plane of nutrition on lactating Holstein cows: I. Production, metabolism, and aspects of circulating somatotropin. *J. Dairy Sci.* **92**, 1986–1997 (2009).
168. El-Nouty, F. D., Elbanna, I. M., Davis, T. P. & Johnson, H. D. Aldosterone and ADH response to heat and dehydration in cattle. *J. Appl. Physiol.* **48**, 249–255 (1980).
169. Itoh, F., Obara, Y., Rose, M. T., Fuse, H. & Hashimoto, H. Insulin and Glucagon Secretion in Lactating Cows During Heat Exposure1.
170. Sanz Fernandez, M. V. *et al.* Heat stress increases insulin sensitivity in pigs. *Physiol Rep* **3**, (2015).
171. Keogh, K., Kenny, D. A., Kelly, A. K. & Waters, S. M. Insulin secretion and signaling in response to dietary restriction and subsequent re-alimentation in cattle. *Physiol. Genomics* **47**, 344–354 (2015).
172. Tesmer, L. A., Lundy, S. K., Sarkar, S. & Fox, D. A. Th17 cells in human disease. *Immunol. Rev.* **223**, 87–113 (2008).
173. Romagnani, S. T-cell subsets (Th1 versus Th2). *Ann. Allergy Asthma Immunol.* **85**, 9–18; quiz 18, 21 (2000).
174. Deschamps, M. *et al.* Genomic Signatures of Selective Pressures and Introgression from Archaic Hominins at Human Innate Immunity Genes. *Am. J. Hum. Genet.* **98**, 5–21 (2016).
175. Fan, S., Hansen, M. E. B., Lo, Y. & Tishkoff, S. A. Going global by adapting local: A review of recent human adaptation. *Science* **354**, 54–59 (2016).
176. Streit, M. *et al.* Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle. *G3* **3**, 1085–1093 (2013).
177. Rauw, W. M. & Gomez-Raya, L. Genotype by environment interaction and breeding for robustness in livestock. *Front. Genet.* **6**, 310 (2015).
178. Foissac, S. *et al.* Multi-species annotation of transcriptome and chromatin structure in

domesticated animals. *BMC Biol.* **17**, 108 (2019).
179. Giuffra, E., Tuggle, C. K. & FAANG Consortium. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu Rev Anim Biosci* **7**, 65–88 (2019).
180. Bermingham, M. L. *et al.* Application of high-dimensional feature selection: evaluation for genomic prediction in man. *Sci. Rep.* **5**, 10312 (2015).
181. MacLeod, I. M. *et al.* Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144 (2016).
182. Kang, E. Y. *et al.* Meta-analysis identifies gene-by-environment interactions as demonstrated in a study of 4,965 mice. *PLoS Genet.* **10**, e1004022 (2014).
183. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
184. Han, B. & Eskin, E. Interpreting meta-analyses of genome-wide association studies. *PLoS Genet.* **8**, e1002555 (2012).
185. Perganta, G. *et al.* Non-image-forming light driven functions are preserved in a mouse model of autosomal dominant optic atrophy. *PLoS One* **8**, e56350 (2013).
186. Davies, V. J. *et al.* Opa1 deficiency in a mouse model of autosomal dominant optic atrophy impairs mitochondrial morphology, optic nerve structure and visual function. *Hum. Mol. Genet.* **16**, 1307–1318 (2007).
187. Patten, D. A. *et al.* OPA1-dependent cristae modulation is essential for cellular adaptation to metabolic demand. *EMBO J.* **33**, 2676–2691 (2014).
188. Piquereau, J. *et al.* Down-regulation of OPA1 alters mouse mitochondrial morphology, PTP function, and cardiac adaptation to pressure overload. *Cardiovasc. Res.* **94**, 408–417 (2012).
189. Dong, K. *et al.* Genomic scan reveals loci under altitude adaptation in Tibetan and Dahe pigs. *PLoS One* **9**, e110520 (2014).
190. Gopalakrishnan, K. *et al.* Targeted disruption of Adamts16 gene in a rat genetic model of hypertension. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 20555–20559 (2012).
191. Alavi, M. V. *et al.* A splice site mutation in the murine Opa1 gene features pathology of autosomal dominant optic atrophy. *Brain* **130**, 1029–1042 (2007).
192. Atemnkeng, V. A. *et al.* Deoxyhypusine hydroxylase from Plasmodium vivax, the neglected human malaria parasite: molecular cloning, expression and specific inhibition by the 5-LOX inhibitor zileuton. *PLoS One* **8**, e58318 (2013).
193. Gronostajski, R. M. Roles of the NFI/CTF gene family in transcription and development. *Gene* **249**, 31–45 (2000).
194. Tizioto, P. C. *et al.* Gene expression differences in Longissimus muscle of Nelore steers genetically divergent for residual feed intake. *Scientific Reports* vol. 6 (2016).
195. Jain, I. H. *et al.* Hypoxia as a therapy for mitochondrial disease. *Science* **352**, 54–61 (2016).
196. Rowan, T. N. *et al.* A Multi-Breed Reference Panel and Additional Rare Variation Maximizes Imputation Accuracy in Cattle. *bioRxiv* 517144 (2019) doi:10.1101/517144.
197. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
198. Loh, P.-R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443 (2016).

199. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
200. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
201. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
202. R Core Team, R. & Others. R: A language and environment for statistical computing. (2013).
203. Wickham, H. ggplot2. *Wiley Interdisciplinary Reviews: Computational* (2011).
204. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
205. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
206. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
207. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).