Using singleton densities to detect recent selection in *Bos taurus*

Matthew Hartfield[1,2], Nina Aagaard Poulsen[3], Bernt Guldbrandtsen[4,5,6], Thomas Bataillon[1]

1) Bioinformatics Research Centre, Aarhus University, DK–8000 Aarhus, Denmark.

2) Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, UK.

3) Department of Food Science, Aarhus University, Agro Food Park 48, 8200 Aarhus N, Denmark.

4) Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK–8830 Tjele, Denmark.

5) Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Tierwissenschaften, Tierzucht, Endenicher Allee 15, D-53115 Bonn, Germany.

6) Department of Veterinary Sciences, Copenhagen University, DK–1870 Frederiksberg C, Denmark.

Corresponding author: Matthew Hartfield (m.hartfield@ed.ac.uk).

Key words: selection, genomics, *Bos taurus*, milk protein, milk fat, stature.

Abstract: Many quantitative traits are subject to polygenic selection, where several genomic regions undergo small, simultaneous changes in allele frequency that collectively alter a phenotype. The widespread availability of genome data, along with novel statistical techniques, has made it easier to detect these changes. We apply one such method, the 'Singleton Density Score', to the Holstein breed of *Bos taurus* to detect recent selection (arising up to around 740 years ago). We identify several genes as candidates for targets of recent selection, including some relating to cell regulation, catabolic processes, neural-cell adhesion and immunity. We do not find strong evidence that three traits that are important to humans – milk protein content, milk fat content, and stature – have been subject to directional selection. Simulations demonstrate that since *B. taurus* recently experienced a population bottleneck, singletons are depleted so the power of SDS methods are reduced. These results inform on which genes underlie recent genetic change in *B. taurus*, while providing information on how polygenic selection can be best investigated in future studies.

Impact statement: Many traits of ecological or economic importance (including height, disease propensity, climatic adaptation) are ‘polygenic’. That is, they are affected by a large number of genetic variants, with each one only making a small contribution to a trait, but collectively influence variation. As selection acts on all of these variants simultaneously, it only changes the frequency of each one by a small amount, making it hard to detect such selection from genome data. This situation has changed in recent years, with the proliferation of whole–genome data from many individuals, along with the development of methods to detect the subtle effects of polygenic selection. Here, we use data from 102 genomes from domesticated cattle (*Bos taurus*) that has experienced intense artificial selection since domestication, and test whether we can detect signatures of recent selection (arising up to 740 years ago). Domesticated species are appealing for this kind of study, as they are subject to extensive genome sequencing studies, and genetic variants can be related to traits under selection. We carried out our analysis in two parts. We first performed a genome–wide scan to find individual genetic regions that show signatures of recent selection. We identify some relating to cell regulation, catabolic processes, neural-cell adhesion and immunity. In the second part, we then analysed genetic regions associated with three key traits: milk protein content, milk fat content, and stature. We tested whether these regions collectively showed a signature of selection, but did not find a significant result in either case. Simulations suggest that the domestication history of cattle affected the power of these methods. We end with a discussion on how to best detect polygenic selection in future studies.

Introduction

Determining which genomic regions have been subject to selection is a major research goal in evolutionary genetics. Traditional methods have focused on detecting strong selection affecting individual genes (Nielsen, 2005; Vitti *et al.*, 2013; Stephan, 2019). An alternative process is 'polygenic selection', where many loci contribute to genetic variation in a trait, so selection acting on it is expected to generate small and simultaneous allele frequency changes at multiple loci (Pritchard & Di Rienzo, 2010; Pritchard *et al.*, 2010). Many polygenic models have been formulated to account for both the response to phenotypic selection, and the maintenance of genetic variance in quantitative traits [reviewed by Sella & Barton (2019); Barghi *et al.* (2020)]. Among them is Fisher's infinitesimal model, which is important for its historical role in uniting population and quantitative genetics, and its recent renaissance in the context of genome–wide association studies (Fisher, 1918; Barton & Keightley, 2002; Barton *et al.*, 2017; Charlesworth & Edwards, 2018; Visscher & Goddard, 2019). However, whereas it has been possible to identify which genetic regions contribute to trait variation, it has historically been hard to infer which alleles have been involved in the polygenic selection response. Extensive theoretical studies of how alleles at multiple loci act when a population adapts to a new optimum generally find that 'large–effect' alleles, which strongly affect a trait, are the first to spread and fix while 'small–effect' alleles take much longer to reach high frequencies (de Vladar & Barton, 2014; Wollstein & Stephan, 2014; Jain & Stephan, 2015, 2017a, 2017b; Stetter *et al.*, 2018; Thornton, 2019; Hayward & Sella, 2019). Furthermore, if epistasis exists between variants, many selected alleles do not reach fixation as they eventually become deleterious (de Vladar & Barton, 2014; Jain & Stephan, 2017b).

82 The spread of large–effect alleles may also be impeded if a faster adaptive response
83 can be otherwise realised through changes at many small–effect alleles (Lande,
84 1983; Chevin & Hospital, 2008; Pavlidis *et al.*, 2012; Chevin, 2019). Alternatively, if
85 the optimum shift is sufficiently big, then large-effect mutations that first go to fixation
86 can subsequently be replaced by small–effect variants over longer timescales (on the
87 order of the population size; Hayward and Sella (2019)). Overall, only a small
88 proportion of loci affected by polygenic selection are expected to fix sufficiently
89 quickly to leave selection signatures in genomic data (Pavlidis *et al.*, 2012; Thornton,
90 2019).

91 Due to this difficulty, earlier methods for detecting polygenic selection focused
92 on cases where selection favours distinct phenotypes in different populations, so trait
93 differentiation amongst populations will be greater than expected under neutral drift.
94 Tests for this form of selection relied on comparing Q_{st} and F_{st} statistics, which
95 respectively measured mean genetic differentiation at the trait itself and a set of
96 neutral loci (Whitlock, 2008; Le Corre & Kremer, 2012; Savolainen *et al.*, 2013). Yet
97 these methods do not determine which genomic regions are subject to selection. This
98 situation has now changed with the increased number of genome–wide association
99 study (GWAS) data that link genotypes and phenotypes, as exemplified by the
100 development of large cohort studies [e.g., the UK Biobank; Bycroft *et al.* (2018)]. The
101 release of these data spurred a series of studies and new methods designed
102 specifically to detect polygenic selection. These methods usually involve determining,
103 which SNPs affecting a phenotype show correlated changes in frequency (Berg &
104 Coop, 2014; Racimo *et al.*, 2018; Sanjak *et al.*, 2018; Josephs *et al.*, 2019; Berg *et*
105 *al.*, 2019a, 2019b; Uricchio *et al.*, 2019; Edge & Coop, 2019; Kreiner *et al.*, 2020;

106 Wieters *et al.*, 2021; Gramlich *et al.*, 2021); which sets of alleles are associated with
107 certain environmental or climatic variations (Coop *et al.*, 2010; Turchin *et al.*, 2012;
108 Robinson *et al.*, 2015; Yeaman *et al.*, 2016; Exposito-Alonso *et al.*, 2018; Zan &
109 Carlborg, 2018; Exposito-Alonso *et al.*, 2019; MacLachlan *et al.*, 2021; Ehrlich *et al.*,
110 2021; Fuhrmann *et al.*, 2021; Rowan *et al.*, 2021); or determining which SNPs or
111 genetic regions explain a large fraction of phenotypic variance and trait heritability
112 (Zhou *et al.*, 2013; Yang *et al.*, 2015; Gazal *et al.*, 2017; Zeng *et al.*, 2018; Schoech
113 *et al.*, 2019; Exposito-Alonso *et al.*, 2020; Duntsch *et al.*, 2020; Zeng *et al.*, 2021).
114 Some of these approaches use overlapping methods.

115 Detecting recent polygenic selection is much harder, as long periods of time
116 (number of generations on the order of the population size; Hayward and Sella, 2019;
117 Thornton, 2019) may be needed to cause detectable frequency changes in alleles
118 with small effect sizes. Over shorter timescales, these frequency changes are
119 expected to be more modest and harder to detect (Stephan, 2016; Jain & Stephan,
120 2017a). A recent breakthrough in detecting these subtle changes was the
121 development of the 'Singleton Density Score' (SDS), a statistic tailored to detect
122 recent and coordinated allele frequency changes over many SNPs (Field *et al.*,
123 2016). Recent selection at a locus favouring one variant will lead to a reduction in the
124 number of singletons (i.e., variants that are only observed once) around it. The SDS
125 detects regions that exhibit a reduction in the density of singletons, to determine
126 candidate regions that have been subject to recent selection. Using this approach,
127 Field et al. (2016) found correlations between SDS scores at SNPs and their
128 associated GWAS effect sizes for several polygenic traits in the modern UK human
129 population, including increased height, infant head circumference and fasting insulin.

130 Their findings suggested that these traits have been subject to recent selection
131 during the most recent 75 or so generations (about 2,000 years). However, these
132 (and other) results that detect selection for increased height may instead reflect
133 previously unaccounted–for population structure (Novembre and Barton, 2018;
134 Barton et al., 2019; Sohail et al., 2019; Berg et al., 2019; Uricchio et al., 2019; Edge
135 and Coop, 2019).

136 The SDS method is ideally suited to organisms where large amount of whole-
137 genome data are available, along with QTL or GWAS information that link genotypes
138 to phenotypes, Domesticated species are attractive systems for studying recent
139 selection, as selected phenotypes are often already known and these species are
140 subject to large–scale sequencing studies. Investigating the genetic architecture
141 underlying rapid selection in these species is also important to determine how they
142 respond to agricultural practices, and uncover selection targets that can be used to
143 improve breeding programs (Georges *et al.*, 2018). Domestic cattle *Bos taurus* has
144 been subject to intensive genomics analyses to improve artificial selection for traits
145 that are important for human use, including milk protein yield, milk fat content, and
146 stature (Hayes *et al.*, 2009; Meuwissen *et al.*, 2013; Wray *et al.*, 2019). These traits
147 are influenced in part by an individual's genome, with significant heritability estimates
148 being recorded, some as high as 80% (Soyeurt *et al.*, 2007; Haile-Mariam *et al.*,
149 2013; Buitenhuis *et al.*, 2016). Previous selection scans on *B. taurus* reported
150 individual regions that were likely to be subject to recent selection, some of which
151 were close to genetic regions for stature and milk protein content (Lemay *et al.*, 2009;
152 MacEachern *et al.*, 2009; Qanbari *et al.*, 2010; Boitard & Rocha, 2013; Qanbari *et al.*,
153 2014; Zhao *et al.*, 2015; Boitard *et al.*, 2016a; Bouwman *et al.*, 2018). However,

154 stature and milk protein content are polygenic traits, with several genetic regions and
155 QTLs associated with each (Lemay *et al.*, 2009; Boitard *et al.*, 2016a; Bouwman *et*
156 *al.*, 2018; van den Berg *et al.*, 2020). While recent methods have been developed to
157 detect polygenic environmental adaptation (Rowan *et al.,* 2021), there has yet to be a
158 formal test of whether these intrinsic traits show evidence of polygenic selection.
159 Here, we applied the SDS method to whole–autosome sequencing data from
160 102 *B. taurus* Holstein individuals. We first determined genetic regions that have
161 been subject to recent directional selection, and subsequently tested whether
162 evidence exists for recent selection acting on a set of QTLs underlying either milk
163 protein content, milk fat content, or stature in this breed.
164

165 ## ***Results***

166 ### *Methods outline*

167 We filtered the data to retain only bi–allelic SNPs that had a sensible level of
168 coverage and did not lie in putatively over–assembled regions (i.e., duplicated
169 sections that caused many reads to assemble at a specific genetic location). Over–
170 assembled regions appear as highly heterozygous with elevated coverage, and can
171 exhibit false signatures of recent selection. We also obtained a set of singletons and
172 filtered them to retain high-quality variants where both alleles were equally well
173 covered to remove potentially erroneous calls. We polarised test SNPs using
174 outgroup sequences and applied the SDS test of Field et al. (2016) to detect recent
175 selection, with increased SDS values reflecting selection favouring derived SNPs
176 over ancestral variants. We standardised SDS scores with those of a similar
177 frequency, so they are normally distributed [similar normalisation was also carried out

178 by Field et al. (2016)]. These values are denoted sSDS for 'standardised SDS'.
179 Further details are available in the *Methods* in the Supplementary Text.
180

181 ## *Estimating timescale of selection*

182 We first determined the timescale over which we expect to detect selection in
183 *B. taurus* using the SDS method. SDS measures the changes in singleton numbers
184 around putatively selected SNPs, relative to background numbers in the absence of
185 selection. As singletons arise on the tips of the underlying gene trees, the average tip
186 length in the genealogy of sequenced samples determines the timescale over which
187 the SDS detects a signal (Field et al., 2016). As more haploid genomes are included
188 in the study, the time to first coalescence between two samples decreases, reducing
189 the tip lengths and therefore shortening the timescale over which SDS detects
190 selection (Field *et al.*, 2016). We hence simulate tip-ages over a range of sample
191 sizes to investigate how this timescale changes accordingly.
192 To calculate the mean tip age, we simulated gene genealogies under two
193 scenarios. We first simulated the Holstein population demography inferred by Boitard
194 et al. (2016b), which suggested that this population experienced a sudden decline in
195 effective population size (N_e) since domestication, but with a present–day N_e (~793)
196 that is much larger than that inferred from pedigree data [~49; Sørensen et al.
197 (2005)] or from temporal variation in SNP frequencies (~48; Jiménez–Mena et al.
198 2016). Hence, we also simulated genealogies under a second model that used the
199 Boitard et al. (2016b) demographic model, but with the present–day N_e set to 49.
200 These scenarios will be referred to as the 'High N_0' and 'Low N_0' models,
201 respectively.

202 Figure 1 shows simulation results. Depending on the assumed present–day
203 N_e, the tip length in our sample of 204 alleles (i.e., assuming two per diploid
204 individual) goes back either 65 or 148 generations. Assuming 5 years per generation
205 (Boitard *et al.*, 2016b), this timescale corresponds to between 325 and 740 years
206 ago. Since *B. taurus* domestication started around 10,000 years ago (Zeder, 2008)
207 the sample size used in this study will only capture selection acting in the very recent
208 past that is more relevant for breed formation, rather than selection during *B. taurus*
209 domestication. Sample sizes and tip-ages are linearly related on a log-log scale,
210 meaning that an increase in sample size will greatly decrease the timescale over
211 which SDS detects selection. For example, with 500 haplotypes then SDS will detect
212 selection acting no more than 50 generations ago, depending on the underlying
213 demographic model.
214 We will focus on detecting selection signatures assuming the high N_0 model.
215 Results using the low N_0 model to calibrate scores were broadly similar. They are
216 outlined in the Supplementary Text; we will highlight when differences arise.
217

218 *Genome–wide sSDS*

219 Figure 2 plots sSDS values (at SNPs with minor allele frequency greater than
220 5%) across all autosomes, excluding chromosome 25 (due to an insufficient number
221 of singletons needed to obtain SDS scores after filtering). Many SNPs have elevated
222 sSDS scores (158 SNPs at $FDR < 0.05$; 306 for the low N_0 model). Several regions
223 contain SNPs with significantly high sSDS values (Bonferroni–corrected nominal $P <$
224 0.05; actual $P < \sim 2.7 \times 10^{-8}$). To further investigate potential selection targets, we
225 looked for genes that either overlapped significant SNPs or lay 10kb up– or

226 downstream of them. Linkage disequilibrium (LD), as measured by r^2, decays to
227 around 0.2 over 50kb in Danish Holstein breeds (Buitenhuis *et al.*, 2016), so genes
228 within 10kb should be in LD with regions harbouring high sSDS scores. Table 1 lists
229 these genes, with more targets present under the low N_0 model. Most of these genes
230 are of unknown function (as listed on UniProt); the list also includes an snRNA.
231 *PPM1L* is involved with cellular regulation and the activation of stress–activated
232 protein kinases. *TDO2* is involved in tryptophan–related catabolic processes, while
233 *NTM* is implicated in neural cell adhesion. SNPs with significantly elevated scores
234 are also found on chromosome 23 near the MHC region, which may reflect over–
235 dominant selection. All Bonferroni–significant SNPs were removed from subsequent
236 tests of recent polygenic selection to prevent directional selection from skewing the
237 underlying sSDS distributions. Figure S1 shows results for the low N_0 model.
238

239 *Testing for polygenic selection acting on milk protein and stature*

240 If polygenic selection were acting on specific traits, we expect a positive
241 correlation between the effect size of variant underpinning it, and selection acting on
242 it as measured by sSDS. We collated sSDS scores of SNPs that lie close to QTLs
243 reported for either milk fat percentage, milk protein percentage (van den Berg *et al.*,
244 2020), or those that lie close to stature QTLs (Bouwman *et al.*, 2018). The latter were
245 inferred from a meta–analysis of GWAS studies conducted in seven Holstein
246 populations, but not every QTL had an effect size reported in each population. We
247 hence investigated two overlapping consensus QTL sets, where an effect size was
248 either reported in at least 6 of 7 populations (yielding 42 QTLs with sSDS scores
249 associated with them), or where effect sizes were reported in at least 5 of 7

250 populations (58 QTLs had sSDS scores). We re-polarised SDS scores so that a
251 positive score reflected a trait-increasing effect; we denote these values 'tSDS'
252 following Field *et al.* (2016). We then determine if there was a positive correlation
253 between the absolute $\log_{10}$-value of the QTL *P*-value (a proxy for the effect size) and
254 tSDS.

255 Figure 3 shows the relationship between QTL *P*-values and tSDS for SNPs
256 that lie close to QTLs. Although positive trends are observed, they all exhibit non-
257 significant correlations (milk fat percentage Spearman ρ = 0.0990, P = 0.603; milk
258 protein percentage Spearman ρ = 0.0354, P = 0.758; stature from 6 breeds
259 Spearman ρ = -0.0739, P = 0.642; stature from 5 breeds Spearman ρ = -0.00966, P
260 = 0.943). Relationships remain non-significant after removing an outlier point for the
261 milk traits whose QTL has an extremely low *P*-value (Figure S2), and also under the
262 low N_0 model (Figure S3; see figure legends for regression *P*-values).

263 sSDS (and tSDS) can become correlated along the genome if focal SNPs are
264 in LD with one another, which was not accounted for in the preceding analyses. To
265 determine whether LD could have affected these correlations, we randomly
266 subsampled sSDS scores from SNPs that shared the same chromosome and bin of
267 derived-allele frequency as the SNPs used in the above analyses, and re-polarised
268 them to transform them into tSDS values. We then determined the Spearman's ρ
269 associated with these permuted values to determine whether that for the true data
270 was significantly elevated (see Methods for details). In all cases, the observed value
271 was not significantly higher than for permuted values (see Figure S4 for histograms
272 and exact *P*–values, which all exceed 0.05). We therefore conclude that these QTL
273 datasets do not harbour SNPs with significantly different tSDS scores compared to

274 the rest of the genome.

275

276 Discussion

277 *Summary of results*

278 We analysed an extensive *B. taurus* genomic dataset to identify signatures of
279 recent selection in the Holstein breed, and to determine whether the data contained a
280 signal of polygenic selection acting on milk proteins and QTLs underlying phenotypic
281 variation in stature. Given the sample size and the demographic history of Holsteins,
282 the SDS method can detect very recent selection events arising no more than
283 approximately 740 years ago (Figure 1). A whole–genome scan for sSDS scores
284 identified several targets of recent directional selection that overlap or lie close to
285 protein–coding genes (Figure 2; Table 1). The genes whose functions are known are
286 involved in protein regulation, catabolic processes, and neural-cell adhesion.
287 Significant values were also observed near the MHC region. We subsequently
288 investigated whether either milk protein genes or SNPs near stature QTLs
289 collectively showed evidence of polygenic selection. We did so by testing whether
290 there is a relationship between the QTL effect size, as measured by its *P*-value, and
291 tSDS values to SNPs near them. However, no relationship was observed, even after
292 performing a permutation test (Figures 3, S2-S4). Hence, while sSDS could reveal
293 specific instances of recent selection, tests based on collective scores of variants
294 associated with known selected traits yielded no signal of polygenic selection.

295

296 *Potential reasons for a lack of polygenic selection signal*

297 *Impact of Holstein demographic history*

298 While the SDS method detected individual candidate genes for very recent
299 selection, we were unable to find strong evidence for polygenic selection acting on
300 three traits that were subject to artificial selection since domestication. This result is *a*
301 *priori* surprising, given that these traits have been subject to recent intense artificial
302 selection. Recent studies generally find non-zero heritability estimates for them,
303 indicating that there should be the potential for genetic variants underpinning them to
304 change in response to artificial selection (Soyeurt *et al.*, 2007; Haile-Mariam *et al.*,
305 2013; Buitenhuis *et al.*, 2016). In addition, the ratio of the mutation and recombination
306 rates in cattle is just over three (Boitard *et al.*, 2016b; Harland *et al.*, 2018), indicating
307 that several informative SNPs exist per haplotypes that should improve the power of
308 the SDS method [in contrast, this ratio is approximately equal to one in humans
309 (Field *et al.*, 2016)].

310 One potential reason for this lack of signal is due to the population history of
311 *Bos taurus*. The effective population size of many *B. taurus* breeds appears to have
312 undergone a decline since domestication (Sørensen et al., 2005; Boitard et al.,
313 2016b), which likely reflects successive bottlenecks due to domestication, breed
314 formation and intense recent selection. Population size reductions are known to
315 reduce the number of low-frequency variants and increases the prevalence of
316 intermediate-frequency variants (Harpending *et al.*, 1998), which can affect the power
317 of the SDS method. To understand if the history of *B. taurus* affects the detection of
318 recent selection in Holstein cattle using SDS, we ran coalescent simulations to
319 determine its ability to detect ongoing selection, given realistic Holstein population
320 history and genetic parameters (see Methods for details). We simulated a partial

321 sweep occurring in the middle of a 10Mb region, either assuming a mutation rate in
322 line with what has been inferred for Holstein, or one 10-fold higher to replicate
323 diversity expected in a genetic region with an elevated mutation rate.

324 For the standard mutation rate, no SDS scores were produced for any
325 simulations. After inspecting the simulation results, we see that there is a large skew
326 in the distribution of singleton numbers per individual with a large number of
327 individuals (over 20 on average) that do not carry singletons at the end of
328 simulations, preventing the calculation of a local SDS score (Figure 4). This fraction
329 remained the same irrespective of whether the simulated region was neutral or
330 subject to selection; the main effect of a sweep was to reduce the mean number of
331 singletons per individual, which is the signal measured by SDS (Field *et al.*, 2016).
332 This reduction in overall singleton numbers is consistent with the known effects of
333 population size contraction on reducing tip lengths (Harpending *et al.*, 1998).

334 With a 10-fold higher mutation rate, there were fewer cases where no
335 individual harboured singletons (Figure 4). Accordingly, SDS scores could be
336 calculated for 65 and 66 out of 100 simulations for the neutral and selective cases
337 respectively. In these cases, sSDS values were significantly higher in the selected
338 case than for the neutral case (Figure 5; two-sided Wilcox Test $P = 1.1\text{x}10^{-5}$).
339 However, note that sSDS values is less than one for the selected case, which does
340 not exceed the FDR threshold in our study (for the high N_0 case, the smallest sSDS
341 value with FDR < 0.05 is 4.46).

342 Although singleton numbers differ between the two cases, a reduction in
343 power could also be caused by a more general reduction in diversity due to the small
344 recent effective population sizes of cattle. To investigate this effect, we estimated the

345 fixed *Ne* that would yield the same number of segregating sites in simulations using
346 the standard mutation rate, based on Watterson's estimator (Watterson, 1975;
347 Hudson, 1990; see Methods for details). In both cases where selection is present or
348 absent, *Ne* estimates lie at around 25,000, which is that inferred at approximately
349 halfway between the onset of domestication and the present day. (Boitard *et al.*,
350 2016b; Figure S5). Given that estimates are similar irrespective of whether a sweep
351 was present or not, the reduced population size caused by domestication could have
352 also affected power due to limiting genetic variation and thus the potential to detect
353 subtle sweep signatures associated with polygenic selection.
354 Overall, these simulations are consistent with population size reductions in *B.*
355 *taurus* both reducing the overall genetic diversity and the number of singletons, which
356 limits its ability to detect partial sweeps. SDS is more likely to detect signals in
357 regions of elevated mutation rate, suggesting there will likely be an ascertainment
358 bias in where signals are detected in the genome. The reduction in singletons also
359 reduces the power to investigate SDS values in telomeric regions. SDS values are
360 calculated using the distance up- and downstream from a SNP to the nearest
361 singleton, and are undefined if a certain number of samples do not harbour
362 singletons in either direction (Field *et al.*, 2016). SDS values are hence less likely to
363 be defined in telomeric regions, as it is generally less feasible to observe singletons
364 up until the end of the chromosome. This problem is exacerbated if there are few
365 singletons overall.
366
367 *Other potential reasons for a lack of signal*
368 Another potential reason for a lack of signal is that the selection response on

369 these traits may have been driven by large-effect variants that have already fixed in
370 the population, with a smaller contribution from small-effect mutations. Theoretical
371 models have shown that more major–effect QTLs are likely to fix if the population lies
372 further from a fitness optimum (Lande, 1983; Jain & Stephan, 2017b; Thornton,
373 2019). Domesticated species, which experience strong and sustained directional
374 artificial selection, especially in recent generations, could thereby fix more adaptive
375 mutation via sweep–like processes compared to populations evolving in more stable
376 environments (Lande, 1983; Jain & Stephan, 2017a). Furthermore, once a population
377 has adapted to a new environment (the domestication phenotype in this case), then
378 any remaining major–effect mutations are likely to be superseded by variants with
379 weaker effects, which are harder to detect (Hayward & Sella, 2019). The response to
380 polygenic selection will be further weakened in smaller populations (John & Stephan,
381 2020), which could be a factor given the reduced effective population sizes of *B.*
382 *taurus* (Sørensen *et al.*, 2005; Boitard *et al.*, 2016b). There is some evidence of this
383 explanation; selective sweeps signatures are associated with stature QTLs
384 (Bouwman *et al.*, 2018), and the study of van den Berg et al. (2020) was more likely
385 to identify milk QTLs that had a moderate to high minor allele frequency, suggesting
386 reduced power to detect low-frequency variants that are potential contributors to
387 polygenic selection. Conversely, the stature meta-analysis by Bouwman et al. (2018)
388 found significant SNPs that explained up to 13.8% of the variance in stature, which is
389 similar to that explained by significant SNPs for human height (16%), which is a
390 classic trait for polygenic selection studies. Hence, there may be sufficient polygenic
391 SNPs present to test for polygenic selection, but the power will still be reduced due to
392 the demographic history of Holstein cattle.

393 Potential solutions to increase power include increasing sample sizes; using
394 alternative methods; or analysing different kinds of genome data to detect polygenic
395 selection. Applying SDS to a larger sample size would increase the power to detect
396 selection acting in the recent past [Figure 1; see also Field et al. (2016)], but overall
397 power will still be limited by the tip-length of neutral genealogies. Recent
398 developments in methodology involve directly inferring trees from genome data, and
399 using these to identify subtle sweep signatures associated with trait variants (Edge &
400 Coop, 2019; Speidel *et al.*, 2019; Stern *et al.*, 2021). These methods have greater
401 power to detect weakly-selected mutations that may be segregating for longer than
402 the tip-length of the population.
403 Another approach would be to look beyond sequence data and focus on gene
404 networks [reviewed by Fagny & Austerlitz (2021)]. The recently–proposed 'omnigenic'
405 model (Boyle *et al.*, 2017; Liu *et al.*, 2019) posits that variation in quantitative traits is
406 principally affected by a plethora of 'peripheral' genes that indirectly affect them,
407 rather than a limited set of 'core' genes that directly modify a trait. These numerous
408 peripheral genes may exert their influence via regulatory effects (e.g., gene
409 expression changes), but are also expected to be highly pleiotropic. Fully testing the
410 omnigenic model will require larger datasets and novel experimental designs (Wray
411 *et al.*, 2018). A recent example is from an experiment with *Drosophila melanogaster*,
412 where gene knockouts that do not pass a GWAS significance threshold for pupal
413 length still significantly affect it (Zhang *et al.*, 2021). There is also nascent evidence
414 that gene regulation may underlie directional polygenic selection. Boitard et al.
415 (2016a) found that some adaptive signatures of *B. taurus* are located in intergenic
416 regions; regulatory changes were also proposed to guide polygenic selection in

Arabidopsis (He *et al.*, 2016). Analyses of gene–sets associated with infection responses or immunity also found evidence for polygenic selection in humans and primates (Daub *et al.*, 2013, 2017; Svardal *et al.*, 2017). Immunity gene–sets might be exceptional cases, as they are more likely to contain genes subject to very strong selection (Castellano *et al.*, 2019). Further investigations using regulatory information and a broader range of gene–sets could be a promising approach to determine the impact of polygenic selection.

Materials and Methods

Full methods are available in the Supplementary Text.

Acknowledgements. We would like to thank Simon Boitard for sharing his results on *B. taurus* demographic inference; Simon Boitard, Jon Slate and two anonymous reviewers for providing feedback on the manuscript. MH is supported by a NERC Independent Research Fellowship (NE/R015686/1). NAP, BG and TB are funded by a synergistic research grant from the Faculty of Science and Technology, Aarhus University, Denmark. MH and TB also acknowledge financial support from the European Research Council under the European Union's Seventh Framework Program (FP7/20072013, ERC Grant 311341). The authors declare no conflicts of interest.

Author contributions. All authors contributed to the study design. NAP and BG provided data. MH performed the analyses and wrote the manuscript, with feedback from NAP, BG and TB.

Data archiving. Raw SDS scores and polarisation information has been deposited on Dryad (https://doi.org/10.5061/dryad.547d7wm8q). Data analysis and simulation scripts are available on GitHub (https://github.com/MattHartfield/CattleSDS).

References

Barghi, N., Hermisson, J. & Schlötterer, C. (2020) Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.*, **21**, 769–781.

Barton, N.H., Etheridge, A.M. & Véber, A. (2017) The infinitesimal model: Definition, derivation, and implications. *Theor. Popul. Biol.*, **118**, 50–73.

Barton, N.H., Hermisson, J. & Nordborg, M. (2019) Why structure matters. *eLife*, **8**, e45380.

Barton, N.H. & Keightley, P.D. (2002) Understanding quantitative genetic variation. *Nat. Rev. Genet.*, **3**, 11–21.

Berg, I. van den, Xiang, R., Jenko, J., Pausch, H., Boussaha, M., Schrooten, C., *et al.* (2020) Meta-analysis for milk fat and protein percentage using imputed sequence variant genotypes in 94,321 cattle from eight cattle breeds. *Genet. Sel. Evol.*, **52**, 37.

Berg, J.J. & Coop, G. (2014) A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.*, **10**, e1004412.

Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., *et al.* (2019a) Reduced signal for polygenic adaptation of height in UK Biobank. *eLife*, **8**, e39725.

Berg, J.J., Zhang, X. & Coop, G. (2019b) Polygenic Adaptation has Impacted Multiple Anthropometric Traits. *bioRxiv*, 167551.

Boitard, S., Boussaha, M., Capitan, A., Rocha, D. & Servin, B. (2016a) Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics*, **203**, 433–450.

Boitard, S. & Rocha, D. (2013) Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. *Anim. Genet.*, **44**, 579–583.

470 Boitard, S., Rodríguez, W., Jay, F., Mona, S. & Austerlitz, F. (2016b) Inferring
471 Population Size History from Large Samples of Genome-Wide Molecular Data - An
472 Approximate Bayesian Computation Approach. *PLoS Genet.*, **12**, e1005877.
473 Bouwman, A.C., Daetwyler, H.D., Chamberlain, A.J., Ponce, C.H., Sargolzaei, M.,
474 Schenkel, F.S., *et al.* (2018) Meta-analysis of genome-wide association studies for
475 cattle stature identifies common genes that regulate body size in mammals. *Nat.*
476 *Genet.*, **50**, 362–367.
477 Boyle, E.A., Li, Y.I. & Pritchard, J.K. (2017) An Expanded View of Complex Traits:
478 From Polygenic to Omnigenic. *Cell*, **169**, 1177–1186.
479 Buitenhuis, B., Poulsen, N.A., Gebreyesus, G. & Larsen, L.B. (2016) Estimation of
480 genetic parameters and detection of chromosomal regions affecting the major milk
481 proteins and their post translational modifications in Danish Holstein and Danish
482 Jersey cattle. *BMC Genet.*, **17**, 114.
483 Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., *et al.* (2018)
484 The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**,
485 203–209.
486 Castellano, D., Uricchio, L.H., Munch, K. & Enard, D. (2019) Viruses rule over
487 adaptation in conserved human proteins. *bioRxiv*, 555060.
488 Charlesworth, B. & Edwards, A.W.F. (2018) A century of variance. *Significance*, **15**,
489 20–25.
490 Chevin, L.-M. (2019) Selective Sweep at a QTL in a Randomly Fluctuating
491 Environment. *Genetics*, **213**, 987–1005.
492 Chevin, L.-M. & Hospital, F. (2008) Selective Sweep at a Quantitative Trait Locus in
493 the Presence of Background Genetic Variation. *Genetics*, **180**, 1645–1660.

Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, **185**, 1411–1423.

Daub, J.T., Hofer, T., Cutivet, E., Dupanloup, I., Quintana-Murci, L., Robinson-Rechavi, M., *et al.* (2013) Evidence for Polygenic Adaptation to Pathogens in the Human Genome. *Mol. Biol. Evol.*, **30**, 1544–1558.

Daub, J.T., Moretti, S., Davydov, I.I., Excoffier, L. & Robinson-Rechavi, M. (2017) Detection of Pathways Affected by Positive Selection in Primate Lineages Ancestral to Humans. *Mol. Biol. Evol.*, **34**, 1391–1402.

Duntsch, L., Tomotani, B.M., Villemereuil, P. de, Brekke, P., Lee, K.D., Ewen, J.G., *et al.* (2020) Polygenic basis for adaptive morphological variation in a threatened Aotearoa l New Zealand bird, the hihi (*Notiomystis cincta*). *Proc. R. Soc. B.*, **287**, 20200948.

Edge, M.D. & Coop, G. (2019) Reconstructing the History of Polygenic Scores Using Coalescent Trees. *Genetics*, **211**, 235–262.

Ehrlich, M.A., Wagner, D.N., Oleksiak, M.F. & Crawford, D.L. (2021) Polygenic Selection within a Single Generation Leads to Subtle Divergence among Ecological Niches. *Genome Biol. Evol.*, **13**, evaa257.

Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H.A., Bossdorf, O., Nielsen, R. & Weigel, D. (2019) Natural selection on the *Arabidopsis thaliana* genome in present and future climates. *Nature*, **573**, 126–129.

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H.A. & Weigel, D. (2018) Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nat. Ecol. Evol.*, **2**, 352–358.

Exposito-Alonso, M., Wilton, P. & Nielsen, R. (2020) Non-additive polygenic models

518 improve predictions of fitness traits in three eukaryote model species. *bioRxiv*,
519 2020.07.14.194407.
520 Fagny, M. & Austerlitz, F. (2021) Polygenic Adaptation: Integrating Population
521 Genetics and Gene Regulatory Networks. *Trends Genet.*, **37**, 631–638.
522 Field, Y., Boyle, E.A., Telis, N., Gao, Z., Gaulton, K.J., Golan, D., *et al.* (2016)
523 Detection of human adaptation during the past 2000 years. *Science*, **354**, 760–764.
524 Fisher, R.A. (1918) The correlation between relatives on the supposition of
525 Mendelian inheritance. *Trans. R. Soc. Edinb.*, **52**, 399–433.
526 Fuhrmann, N., Prakash, C. & Kaiser, T.S. (2021) Polygenic adaptation from standing
527 genetic variation allows rapid ecotype formation. *bioRxiv*, 2021.04.16.440113.
528 Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P.-R., Palamara, P.F., Liu, X., *et al.*
529 (2017) Linkage disequilibrium–dependent architecture of human complex traits
530 shows action of negative selection. *Nat. Genet.*, **49**, 1421–1427.
531 Georges, M., Charlier, C. & Hayes, B. (2018) Harnessing genomic information for
532 livestock improvement. *Nat. Rev. Genet.*
533 Gramlich, S., Liu, X., Favre, A., Alex Buerkle, C. & Karrenberg, S. (2021) A polygenic
534 architecture with conditionally neutral effects underlies ecological differentiation in
535 *Silene*. *bioRxiv*, 2021.07.06.451304.
536 Haile-Mariam, M., Nieuwhof, G.J., Beard, K.T., Konstatinov, K.V. & Hayes, B.J.
537 (2013) Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle
538 from genomic and pedigree data and implications for genomic evaluations:
539 Implication of genomic heritability for genomic evaluation. *J. Anim. Breed. Genet.*,
540 **130**, 20–31.
541 Harland, C., Coppieters, W., Mullaart, E., Charlier, C. & Georges, M. (2018) *Rate of*

542 *de novo mutation in dairy cattle and potential impact of reproductive technologies*.
543 Proc. World Congr. Genet. Appl. to Livest. Prod.
544 Harpending, H.C., Batzer, M.A., Gurven, M., Jorde, L.B., Rogers, A.R. & Sherry, S.T.
545 (1998) Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA*, **95**, 1961–
546 1967.
547 Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E. (2009) Invited review:
548 Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.*, **92**, 433–
549 443.
550 Hayward, L.K. & Sella, G. (2019) Polygenic adaptation after a sudden change in
551 environment. *bioRxiv*, 792952.
552 He, F., Arce, A.L., Schmitz, G., Koornneef, M., Novikova, P., Beyer, A., *et al.* (2016)
553 The Footprint of Polygenic Adaptation on Stress-Responsive Cis-Regulatory
554 Divergence in the *Arabidopsis* Genus. *Mol. Biol. Evol.*, **33**, 2088–2101.
555 Hudson, R.R. (1990) Gene Genealogies and the Coalescent Process. In *Oxford*
556 *Surveys in Evolutionary Biology* (ed. by Futuyma, D.J. & Antonovics, J.). Oxford Univ.
557 Press, Oxford, pp. 1–42.
558 Jain, K. & Stephan, W. (2015) Response of Polygenic Traits Under Stabilizing
559 Selection and Mutation When Loci Have Unequal Effects. *G3*, **5**, 1065–1074.
560 Jain, K. & Stephan, W. (2017a) Modes of Rapid Polygenic Adaptation. *Mol. Biol.*
561 *Evol.*, **34**, 3169–3175.
562 Jain, K. & Stephan, W. (2017b) Rapid Adaptation of a Polygenic Trait After a Sudden
563 Environmental Shift. *Genetics*, **206**, 389–406.
564 Jiménez-Mena, B., Tataru, P., Brøndum, R.F., Sahana, G., Guldbrandtsen, B. &
565 Bataillon, T. (2016) One size fits all? Direct evidence for the heterogeneity of genetic

566 drift throughout the genome. *Biol. Lett.*, **12**.

567 John, S. & Stephan, W. (2020) Important role of genetic drift in rapid polygenic

568 adaptation. *Ecol. Evol.*, **10**, 1278–1287.

569 Josephs, E.B., Berg, J.J., Ross-Ibarra, J. & Coop, G. (2019) Detecting Adaptive

570 Differentiation in Structured Populations with Genomic Data and Common Gardens.

571 *Genetics*, **211**, 989–1004.

572 Kreiner, J.M., Tranel, P.J., Weigel, D., Stinchcombe, J.R. & Wright, S.I. (2020) The

573 genetic architecture and genomic context of glyphosate resistance in *Amaranthus*

574 *tuberculatus*. *bioRxiv*, 2020.08.19.257972.

575 Lande, R. (1983) The response to selection on major and minor mutations affecting a

576 metrical trait. *Heredity*, **50**, 47–65.

577 Le Corre, V. & Kremer, A. (2012) The genetic differentiation at quantitative trait loci

578 under local adaptation. *Mol. Ecol.*, **21**, 1548–1566.

579 Lemay, D.G., Lynn, D.J., Martin, W.F., Neville, M.C., Casey, T.M., Rincon, G., *et al.*

580 (2009) The bovine lactation genome: insights into the evolution of mammalian milk.

581 *Genome Biol.*, **10**, R43.

582 Liu, X., Li, Y.I. & Pritchard, J.K. (2019) Trans Effects on Gene Expression Can Drive

583 Omnigenic Inheritance. *Cell*, **177**, 1022-1034.e6.

584 MacEachern, S., Hayes, B., McEwan, J. & Goddard, M. (2009) An examination of

585 positive selection and changing effective population size in Angus and Holstein cattle

586 populations (*Bos taurus*) using a high density SNP genotyping platform and the

587 contribution of ancient polymorphism to genomic diversity in Domestic cattle. *BMC*

588 *Genom.*, **10**, 181.

589 MacLachlan, I.R., McDonald, T.K., Lind, B.M., Rieseberg, L.H., Yeaman, S. & Aitken,

590 S.N. (2021) Genome-wide shifts in climate-related variation underpin responses to
591 selective breeding in a widespread conifer. *Proc. Natl. Acad. Sci. USA*, **118**,
592 e2016900118.
593 Meuwissen, T., Hayes, B. & Goddard, M. (2013) Accelerating Improvement of
594 Livestock with Genomic Selection. *Annu. Rev. Anim. Biosci.*, **1**, 221–237.
595 Nielsen, R. (2005) Molecular Signals of Natural Selection. *Annu. Rev. Genet.*, **39**,
596 197–218.
597 Novembre, J. & Barton, N.H. (2018) Tread Lightly Interpreting Polygenic Tests of
598 Selection. *Genetics*, **208**, 1351–1355.
599 Pavlidis, P., Metzler, D. & Stephan, W. (2012) Selective Sweeps in Multilocus Models
600 of Quantitative Traits. *Genetics*, **192**, 225–239.
601 Pritchard, J.K. & Di Rienzo, A. (2010) Adaptation - not by sweeps alone. *Nat. Rev.*
602 *Genet.*, **11**, 665–667.
603 Pritchard, J.K., Pickrell, J.K. & Coop, G. (2010) The Genetics of Human Adaptation:
604 Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr. Biol.*, **20**, R208–R215.
605 Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T.M., Fries, R., *et al.* (2014)
606 Classic Selective Sweeps Revealed by Massive Sequencing in Cattle. *PLoS Genet.*,
607 **10**, e1004148.
608 Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R., *et al.*
609 (2010) A genome-wide scan for signatures of recent selection in Holstein cattle.
610 *Anim. Genet.*
611 Racimo, F., Berg, J.J. & Pickrell, J.K. (2018) Detecting Polygenic Adaptation in
612 Admixture Graphs. *Genetics*, **208**, 1565–1584.
613 Robinson, M.R., Hemani, G., Medina-Gomez, C., Mezzavilla, M., Esko, T.,

614 Shakhbazov, K., *et al.* (2015) Population genetic differentiation of height and body
615 mass index across Europe. *Nat. Genet.*, **47**, 1357.
616 Rowan, T.N., Durbin, H.J., Seabury, C.M., Schnabel, R.D. & Decker, J.E. (2021)
617 Powerful detection of polygenic selection and evidence of environmental adaptation
618 in US beef cattle. *PLoS Genet.*, **17**, e1009652.
619 Sanjak, J.S., Sidorenko, J., Robinson, M.R., Thornton, K.R. & Visscher, P.M. (2018)
620 Evidence of directional and stabilizing selection in contemporary humans. *Proc. Natl.*
621 *Acad. Sci. USA*, **115**, 151–156.
622 Savolainen, O., Lascoux, M. & Merila, J. (2013) Ecological genomics of local
623 adaptation. *Nat. Rev. Genet.*, **14**, 807–820.
624 Schoech, A.P., Jordan, D.M., Loh, P.-R., Gazal, S., O'Connor, L.J., Balick, D.J., *et al.*
625 (2019) Quantification of frequency-dependent genetic architectures in 25 UK Biobank
626 traits reveals action of negative selection. *Nat. Commun.*, **10**, 790.
627 Sella, G. & Barton, N.H. (2019) Thinking About the Evolution of Complex Traits in the
628 Era of Genome-Wide Association Studies. *Annu. Rev. Genom. Hum. Genet.*, **20**,
629 461–493.
630 Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., *et al.*
631 (2019) Polygenic adaptation on height is overestimated due to uncorrected
632 stratification in genome-wide association studies. *eLife*, **8**, e39702.
633 Sørensen, A.C., Sørensen, M.K. & Berg, P. (2005) Inbreeding in Danish Dairy Cattle
634 Breeds. *J. Dairy Sci.*, **88**, 1865–1872.
635 Soyeurt, H., Gillon, A., Vanderick, S., Mayeres, P., Bertozzi, C. & Gengler, N. (2007)
636 Estimation of Heritability and Genetic Correlations for the Major Fatty Acids in Bovine
637 Milk. *J. Dairy Sci.*, **90**, 4435–4442.

638 Speidel, L., Forest, M., Shi, S. & Myers, S.R. (2019) A method for genome-wide
639 genealogy estimation for thousands of samples. *Nat Genet.*, **51**, 1321–1329.
640 Stephan, W. (2016) Signatures of positive selection: from selective sweeps at
641 individual loci to subtle allele frequency changes in polygenic adaptation. *Mol. Ecol.*,
642 **25**, 79–88.
643 Stephan, W. (2019) Selective Sweeps. *Genetics*, **211**, 5–13.
644 Stern, A.J., Speidel, L., Zaitlen, N.A. & Nielsen, R. (2021) Disentangling selection on
645 genetically correlated polygenic traits via whole-genome genealogies. *Am. J. Hum.*
646 *Genet.*, **108**, 219–239.
647 Stetter, M.G., Thornton, K. & Ross-Ibarra, J. (2018) Genetic architecture and
648 selective sweeps after polygenic adaptation to distant trait optima. *PLoS Genet.*, **14**,
649 e1007794.
650 Svardal, H., Jasinska, A.J., Apetrei, C., Coppola, G., Huang, Y., Schmitt, C.A., *et al.*
651 (2017) Ancient hybridization and strong adaptation to viruses across African vervet
652 monkey populations. *Nat. Genet.*, **49**, 1705–1713.
653 Thornton, K.R. (2019) Polygenic Adaptation to an Environmental Shift: Temporal
654 Dynamics of Variation Under Gaussian Stabilizing Selection and Additive Effects on a
655 Single Trait. *Genetics*, **213**, 1513–1530.
656 Turchin, M.C., Chiang, C.W., Palmer, C.D., Sankararaman, S., Reich, D., Genetic
657 Investigation of ANthropometric Traits (GIANT) Consortium, *et al.* (2012) Evidence of
658 widespread selection on standing variation in Europe at height-associated SNPs. *Nat*
659 *Genet.*, **44**, 1015–1019.
660 Uricchio, L.H., Kitano, H.C., Gusev, A. & Zaitlen, N.A. (2019) An evolutionary
661 compass for detecting signals of polygenic selection and mutational bias. *Evol. Lett.*,

3, 69–79.

Visscher, P.M. & Goddard, M.E. (2019) From R.A. Fisher's 1918 Paper to GWAS a Century Later. *Genetics*, **211**, 1125–1130.

Vitti, J.J., Grossman, S.R. & Sabeti, P.C. (2013) Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet.*, **47**, 97–120.

Vladar, H.P. de & Barton, N. (2014) Stability and Response of Polygenic Traits to Stabilizing Selection and Mutation. *Genetics*, **197**, 749–767.

Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, **7**, 256–276.

Whitlock, M.C. (2008) Evolutionary inference from F_{ST}. *Mol. Ecol.*, **17**, 1885–1896.

Wieters, B., Steige, K.A., He, F., Koch, E.M., Ramos-Onsins, S.E., Gu, H., *et al.* (2021) Polygenic adaptation of rosette growth in Arabidopsis thaliana. *PLoS Genet.*, **17**, e1008748.

Wollstein, A. & Stephan, W. (2014) Adaptive Fixation in Two-Locus Models of Stabilizing Selection and Genetic Drift. *Genetics*, **198**, 685–697.

Wray, N.R., Kemper, K.E., Hayes, B.J., Goddard, M.E. & Visscher, P.M. (2019) Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans. *Genetics*, **211**, 1131–1141.

Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J. & Visscher, P.M. (2018) Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell*, **173**, 1573–1580.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., *et al.* (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat. Genet.*, **47**, 1114–1120.

Yeaman, S., Hodgins, K.A., Lotterhos, K.E., Suren, H., Nadeau, S., Degner, J.C., *et al.* (2016) Convergent local adaptation to climate in distantly related conifers. *Science*, **353**, 1431.

Zan, Y. & Carlborg, Ö. (2018) A Polygenic Genetic Architecture of Flowering Time in the Worldwide Arabidopsis thaliana Population. *Mol. Biol. Evol.*, **36**, 141–154.

Zeder, M.A. (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl. Acad. Sci. USA*, **105**, 11597–11604.

Zeng, J., Vlaming, R. de, Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., *et al.* (2018) Signatures of negative selection in the genetic architecture of human complex traits. *Nat. Genet.*, **50**, 746–753.

Zeng, J., Xue, A., Jiang, L., Lloyd-Jones, L.R., Wu, Y., Wang, H., *et al.* (2021) Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nat. Commun.*, **12**, 1164.

Zhang, W., Reeves, G.R. & Tautz, D. (2021) Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association. *Curr. Biol.*, **31**, 1092-1098.e6.

Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D.P. (2015) Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.*, **47**, 49.

Zhou, X., Carbonetto, P. & Stephens, M. (2013) Polygenic Modeling with Bayesian Sparse Linear Mixed Models. *PLoS Genet.*, **9**, e1003264.

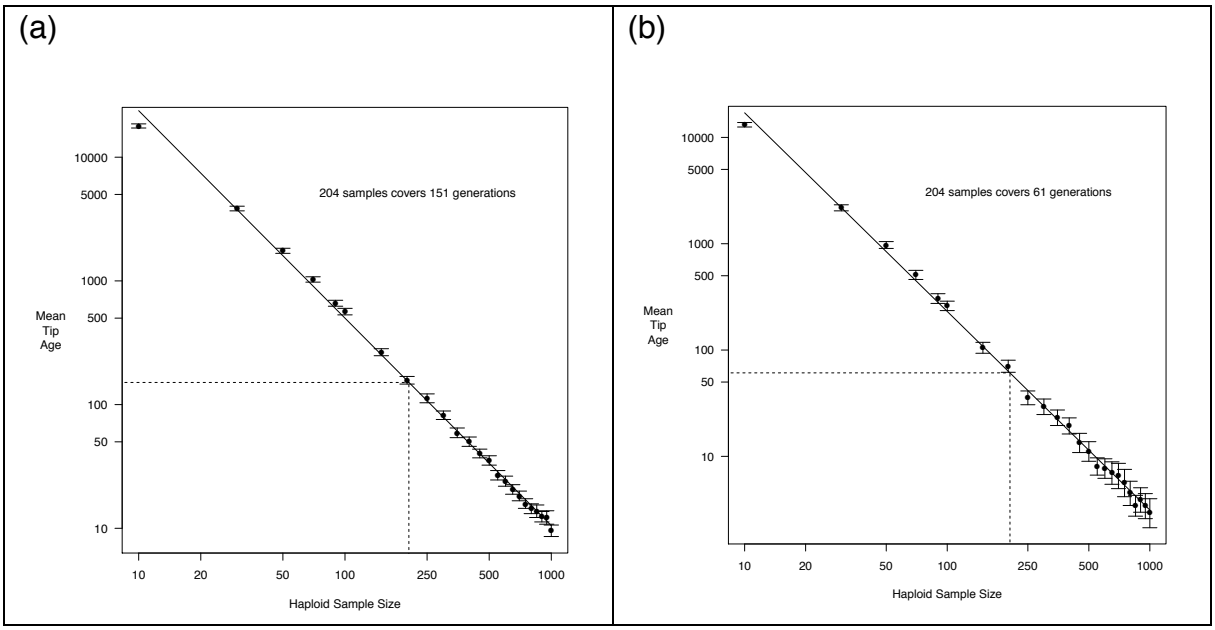


Figure 1: Simulated mean tip age for *B. taurus*, as a function of the number of haploid samples. Simulations assumed either (a) demography as inferred by Boitard et al. (2016b) (the 'High N_0' model), or (b) the same but with a smaller present–day N_e of 49 (the 'Low N_0' model). Points are the mean values; bars show 95% confidence intervals. The solid line is the best linear fit to the log of both values; dotted lines show the predicted tip age for 204 alleles.

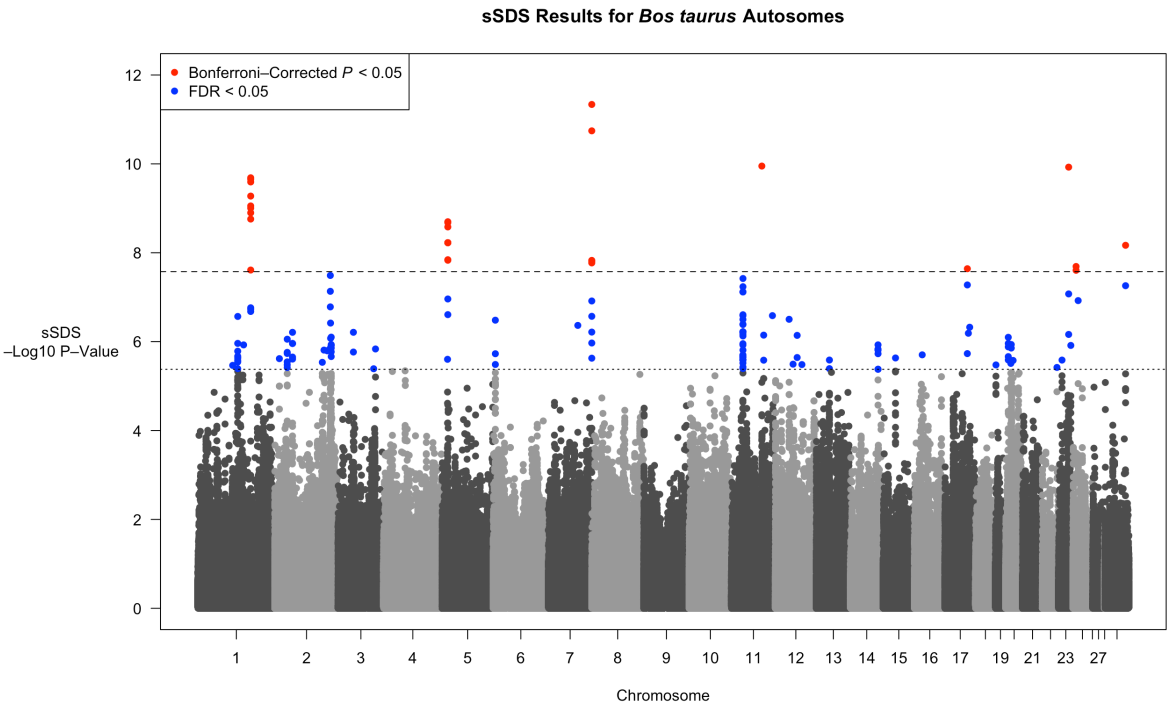
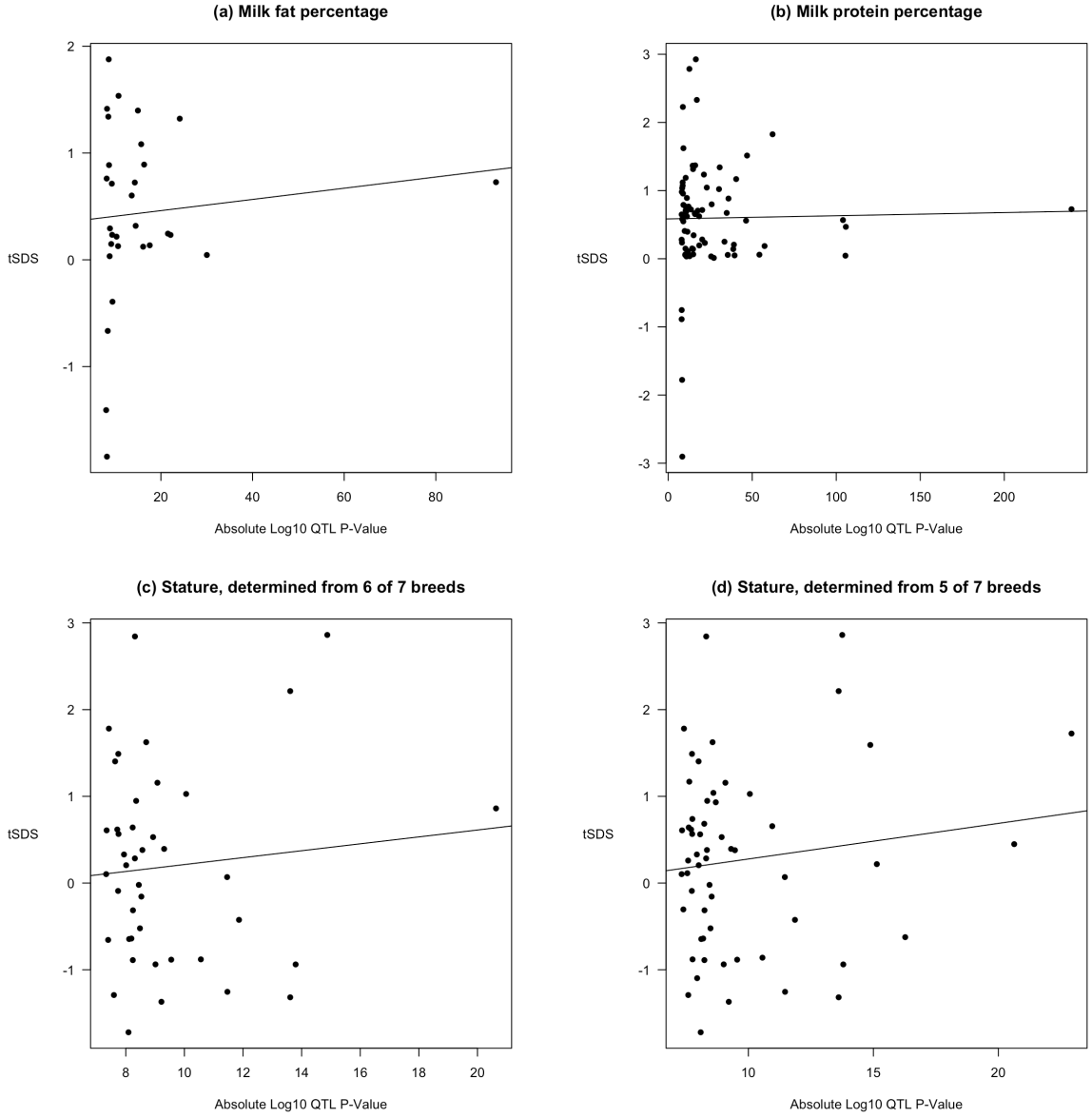


Figure 2: *P*–values of sSDS scores across *B. taurus* autosomes, as plotted on a negative Log_{10} scale, as a function of the chromosome. Alternating black and grey points show (non–significant) values from different chromosomes. Blue points are SNPs with *FDR* < 0.05, with the cut-off denoted by a horizontal dotted line. Red points are SNPs with Bonferroni–corrected *P*–value < 0.05 (actual *P*–value < ~2.7 x 10^{-8}), with the cut-off denoted by a horizontal dashed line. Figure S1 shows results for the Low N_0 model.



723

724 Figure 3: Relationship between tSDS scores near milk or stature QTLs, as noted in

725 the subheadings, and the absolute log *P*-value of QTLs. Lines show a linear model

726 regression fit. Figure S3 shows results assuming a low N_0 model.

727

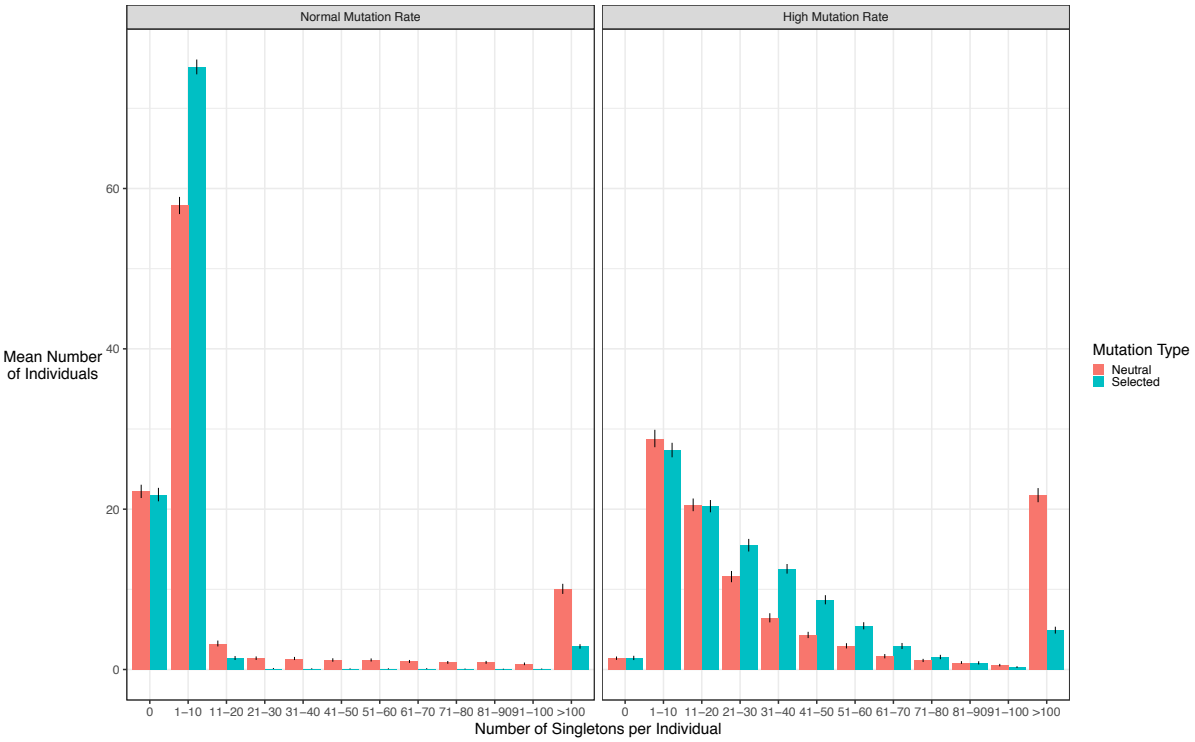


Figure 4: Mean distribution of singleton numbers per individual for each simulation, wither assuming a standard mutation rate (left) or a 10-fold higher mutation rate (right). Bars represent 95% confidence intervals.

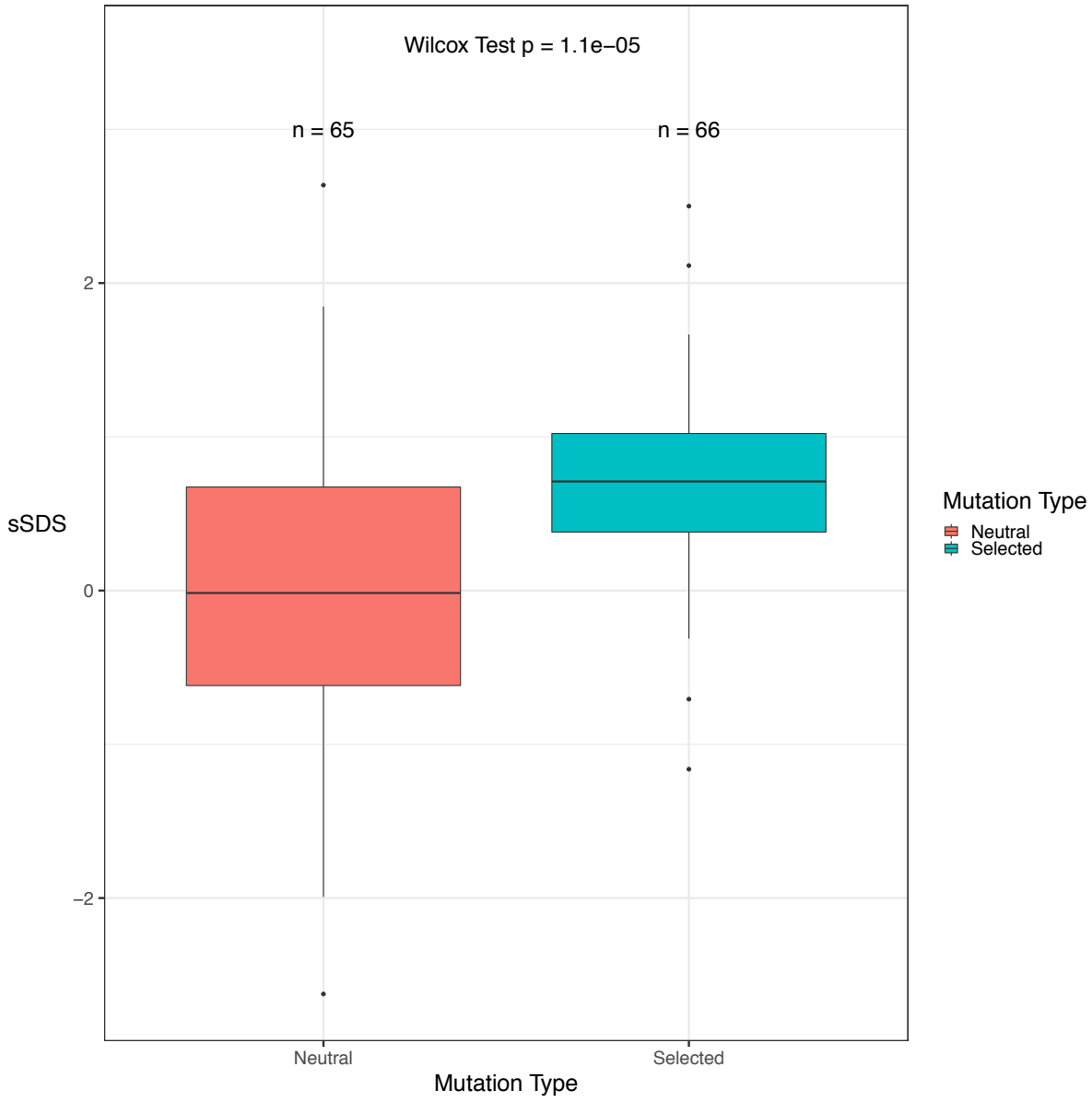


Figure 5: Distribution of simulated sSDS scores assuming a high mutation rate, for the neutral and selected cases. Numbers above each box plot denotes how many simulations produced SDS scores and were included in the plot.

Chromosome	Gene Name	Start Position	End Position	Gene Biotype	High, Low N_0
1	PPM1L	106405113	106727070	Protein Coding	High, Low
5	TMCC3	24306913	24595494	Protein Coding	High, Low
5	CEP83	24070404	24345243	Protein Coding	High, Low
17	U6	43381106	43381209	snRNA	Low
17	CTSO	43364999	43381605	Protein Coding	Low
17	TDO2	43386894	43403747	Protein Coding	High, Low
23	OR12D2H	29291787	29292713	Protein Coding	High, Low
23	OR12D2E	29305933	29309785	Protein Coding	High, Low
24	GAREM1	24694637	24927333	Protein Coding	High, Low
29	NTM	34576918	34994005	Protein Coding	High, Low

Table 1: Genes that overlap or lie close to Bonferroni–significant sSDS regions. The 'High, Low N_0' column specifies which genes are close to significant SNPs for each N_0 model.