ABSTRACT

The Finnish Ayrshire (FAY) belongs to the Nordic Red breeds and is characterized by high milk yield, high milk components, good fertility, and functional conformation. The FAY breeding program is based on genomic selection. Despite the benefits of selection on breeding values, autozygosity in the genome may increase due to selection, and increased autozygosity may cause inbreeding depression in selected traits. However, there is lack of studies concerning selection signatures in the FAY after genomic selection introduction.

The aim of this study was to identify signatures of selection in FAY after the introduction of genomic selection. Genomic data included 45,834 SNPs. The genotyped animals were divided into 2 groups: animals born before genomic selection introduction (6,108 cows) and animals born after genomic selection introduction (47,361 cows). We identified the selection signatures using 3 complementary methods: 2 based on identification of selection signatures from runs of homozygosity (ROH) islands and one based on the decay of site-specific extended haplotype between populations at SNP sites (Rsb).

In total, we identified 34 ROH islands on chromosomes 1, 3, 6, 8, 12–15, 17, 19, 22, and 26 in FAY animals born before genomic selection (between 1980 and 2011) and 30 ROH islands on chromosomes 1–3, 13–17, 22, and 25–26 in FAY animals born after genomic selection introduction (between 2015 and 2020). We additionally detected 22 ∆ROH islands on chromosomes 2–3, 11, 13, 14, 16, 18, 20, and 25–26. Finally, a total of 31 Rsb regions on chromosomes 2, 3, 14, 18, 20, and 25 were identified.

Based on the results, genomic selection has favored certain alleles and haplotypes on genomic regions related to traits relevant in the FAY breeding program: milk production, fertility, growth, beef production traits, and feed efficiency. Several genes related to these traits, e.g., PLA2G4A, ME CR, CHUK, COX15, RICTOR, SHISA9, and SEMA4G overlapped or partially overlapped the observed selection signature regions. The association of genotypes within these regions and their effects on traits relevant in the FAY breeding program should be studied and genetic regions undergoing selection monitored in the FAY population.

Keywords: cattle, ROH island, diversity, genomic selection, extended homozygosity

INTRODUCTION

The Finnish Ayrshire (FAY) belongs to the Nordic Red breeds (RDC) (Finnish Ayrshire, Swedish Red, and Danish Red) (NAV, 2023B). FAY is characterized by high milk yield, high milk components, good fertility, and functional conformation (Faba, 2023; ProAgria, 2023). The 2 most common dairy breeds are Holstein (58.5% in 2022) and Ayrshire (39.5% in 2022), and the national milk recording in Finland includes approximately 70,000 FAY cows (Faba, 2023; ProAgria, 2023). The FAY breeding program is based on genomic selection (GS) (Faba, 2023). Breeding values are estimated jointly for the RDC in Finland, Sweden, and Denmark (NAV, 2023B), and breeding animals are selected based on the Nordic Total Merit Index (NTM) (NAV, 2023A). Production traits (milk, protein, and fat) have the highest weights in the NTM (NAV, 2023A).

GS has been applied in dairy cattle breeding for over a decade. To date, genomic breeding values are generally estimated for all genotyped RDC females and artificial insemination bulls (NAV, 2023A). Despite benefits of selection on breeding values, both traditional BLUP-based selection and GS cause increased autozygosity in the genome (Forutan et al., 2018). As a consequence of autozygosity, inbreeding depression in fertility and production traits in FAY has been detected (Martikainen et al., 2018; Martikainen et al., 2020).

Runs of homozygosity (ROH) are autozygous stretches of genome, resulting from both inbreeding and selection (Broman and Weber, 1999; Kim et al., 2013; Forutan et al., 2018). Selection on traditional BLUP-EBV is expected to increase the frequency of
long ROHs and ROH length, while GS may increase the frequency of small and medium length ROHs (Forutan et al., 2018). In addition, recombination splits ROHs resulting from inbreeding and selection into shorter segments (Kirin et al., 2010; Howrigan et al., 2011). However, Zhang et al. (2015) noted that ROHs emerging from inbreeding are randomly distributed along the genome. Consequently, a high ROH frequency at certain genomic regions in a population are considered to result from selection and are called ROH islands (Nething-agel et al., 2010) — population-specific indicators of selection pressure (Purfield et al., 2017; Gorssen et al., 2019). Additionally, by comparing ROH frequencies between populations, ΔROH islands may be identified which implies ongoing intense selection or novel selection signatures. By studying ROH islands and ΔROH islands, selection signatures have been identified e.g., in cattle (Gorssen et al., 2021; Liu et al., 2021; Lozada-Soto et al., 2022), sheep (Gorssen et al., 2021; Cesarani et al., 2022), pigs (Gorssen et al., 2019; Schiavo et al., 2020; Gorssen et al., 2021), and goats (Cortellari et al., 2021; Gorssen et al., 2021).

In addition to the increased local frequency of ROHs, positive selection may change allele frequencies so that homozygosity is extended to exceptionally long distances (Sabeti et al., 2002). Extended homozygosity and strong selection sweeps that have occurred in one of the 2 populations can be identified from genomic data using the Rsb approach (Tang et al., 2007; Weigand and Leese, 2018; Klassmann and Gautier, 2022). The Rsb approach is based on a standardized log-ratio of integrated site-specific extended haplotype homozygosity between pairs of populations (Klassmann and Gautier, 2022). The Rsb approach is robust to departures from neutrality that are due to demographic parameters other than selection (Tang et al., 2007) and does not require phased or polarized data (the ancestral versus derived state of each allele known) or recombination maps (Tang et al., 2007; Weigand and Leese 2018; Klassmann and Gautier, 2022). The Rsb approach has commonly been applied to detect signatures of selection in indigenous or local breeds and in crossbred or composite cattle breeds (Goszczynski et al., 2018; Alshawi et al., 2019; Mekonnen et al., 2019; van der Nest et al., 2021; Seo et al., 2022). However, only a few studies have applied information on temporal groups within the same cattle breed with the Rsb approach (e.g., Mastrangelo et al., 2020; Seo et al., 2022).

Increased homozygosity due to GS are expected near QTL affecting desired traits (Sonesson et al., 2010). Favorable haplotypes are selected to advance e.g., production, health, and fertility traits, but both desired and undesired haplotypes may simultaneously be ‘hitchhiked’ near by the selected regions (Ma et al., 2019; Xu et al., 2020). Signatures of selection and the hitchhiking effect may be identified by studying genomic regions with increased homozygosity. However, a lack of studies exists concerning selection signatures in FAY after GS introduction.

In this study, we aim to detect signatures of selection in the Finnish Ayrshire population. Using complementary methods — ROH island and ΔROH islands detection and the Rsb method — we will identify genomic regions where selection has caused increased levels of homozygosity during the introduction of GS.

**MATERIALS AND METHODS**

No animal experiments were carried out in this study. Genotype data were obtained from the Nordic Cattle Genetic Evaluation (NAV). The data included genotypes of 53,469 FAY cows with 46,914 imputed autosomal SNP. Animals were considered FAY if they were registered as RDC with the country code FIN (Finland). Information on birth years of the genotyped animals were obtained from the Finnish Animal Breeding Association (Faba).

Genotyped animals were divided into 2 groups: group **BGS**: animals born before the introduction of GS (between 1980 and 2011, 6,108 cows) and group **AGS**: animals born after GS introduction (between 2015 and 2020, 47,361 cows). Most of the animals were genotyped with Illumina BovineLD v.1 Beadchip (Boichard et al., 2012). Before analysis, genotypes from low-density chips were imputed to 50K density using the FImpute software (Sargolzaei et al., 2014). Because genotypes were based on the UMD3.1 assembly, the positions of the SNPs were updated according to the ARS-UCD1.2 assembly (NCBI, 2023). This resulted in the exclusion of 1,080 SNP variants that were either located on the sex chromosome or the mitochondrial chromosome in the ARS-UCD1.2 assembly, lacked chromosomal location, were missing from the ARS-UCD1.2 assembly, or shared a position in the assembly. In total, 45,834 SNP variants remained for further analyses.

**ROH identification**

ROHs were identified using PLINK version 1.90b6.20 (Chang et al., 2015; Purcell and Chang, 2021). Based on the average heterozygosity (32% in both groups), the minimum number of SNPs in a ROH was set to 61 and the sliding window length to 61 SNPs following Meyermans et al. (2020). Threshold t is defined as the hit rate of homozygous sliding windows that must contain the given SNP to define the SNP as part of a ROH. Threshold t was calculated as

---

**Journal of Dairy Science Vol. TBC No. TBC, TBC**
\[ t = \frac{N_{\text{out}} + 1}{L}, \]

where \( N_{\text{out}} \) is the desired number of outer SNPs excluded from a final ROH at both ends of a segment and \( L \) is the sliding window size (Meyermans et al., 2020). Threshold \( t \) was rounded down to 2 decimals. Thus, the sliding window size of 61 and \( N_{\text{out}} = 4 \) resulted in threshold \( t = 0.08 \). Additionally, to control false positive ROH (Meyermans et al., 2020), the minimum ROH length was set to 500 kb. The minimum density of SNPs on a ROH was set to 1 SNP/65 kb, and the maximum gap size between adjacent SNPs in a ROH was limited to 500 kb to obtain sufficient genome coverage. Finally, as imputed genotypes were used, one heterozygote was allowed per sliding window to account for genotyping and imputation errors. The data were not pruned for minor allele frequency or deviations from the Hardy-Weinberg equilibrium (HWE). Individual genomic inbreeding (\( F_{\text{ROH}} \)) was estimated as the number of SNPs in ROHs divided by the total number of SNPs (Pryce et al., 2014).

**ROH islands and ∆ROH islands.** Selection signature detection was based on ROH islands and \( \Delta \)ROH islands, where ROH frequency has increased at certain genomic regions. ROH islands (representing selection signatures detected within a population) were identified i) in the BGS group and ii) in the AGS group based on SNP-wise H-scores (the proportion of SNPs belonging to ROH) (Purfield et al., 2017). The \( \Delta \)ROH islands (representing selection signatures in one population relative to the other) were identified based on SNP-wise \( \Delta \)H-scores, i.e., the difference between SNP-wise H-scores of the BGS and AGS groups.

The SNPs in the top 1% of H-score and \( \Delta \)H-score distributions, with \( p \)-value < 0.01, formed the bases of the ROH islands and \( \Delta \)ROH islands. The SNPs belonging to the top 1% SNP set and separated by a maximum of 2 SNPs were considered a part of the same ROH island or \( \Delta \)ROH island. A ROH island or \( \Delta \)ROH island had to include a minimum of 2 SNPs.

**Rsb approach**

To complement the \( \Delta \)ROH island approach, we identified selective sweeps due to strong positive selection also with the Rsb method using the “rehh” package in R (Klassmann et al., 2021; Klassmann and Gautier, 2022). Before analysis, SNPs significantly deviating from the HWE \( (P < 1 \times 10^{-6}) \) in either group were excluded (648 markers in total). A method by Klassmann and Gautier (2022) was applied to obtain a ratio between the groups (Rsb), briefly, as follows. First, using a SNP (\( t \)) as a reference, a SNP (\( t \)), where a site-specific extended haplotype homozygosity score (EHHS) decreased below a chosen cut-off limit, was chosen. Then, the number of homozygous individuals for a region that extended from SNP \( s \) to SNP \( t \) was calculated \( I_{s,t} \). The EHHS between SNPs \( s \) and \( t \) (EHHS\(_{s,t} \)) was determined independently in both directions starting from SNP \( s \). The EHHS\(_{s,t} \) was calculated for both groups by dividing the \( I_{s,t} \) by the number of homozygous individuals for SNP \( s \) (\( I_{s,s} \)):

\[
\text{EHHS}_{s,t} = \frac{I_{s,t}}{I_{s,s}}.
\]

This was followed by integration over the physical distance and normalization of EHHS\(_{s,t} \) to its value at marker \( t = s \), producing integrated normalized EHHS scores

\[
\text{inES}(s) = \int \frac{1}{ \text{EHHS}_{s,t}} \text{EHHS}_{s,t} \text{dt}.
\]

Because unphased data were used, the cut-off limit for EHHS was increased from the default 0.05 to 0.1 (Klassmann et al., 2021). The higher cut-off limit was recommended by Klassmann and Gautier (2022) to reduce statistical noise, i.e., to prevent high EHHS scores due to one or a few extremely long, shared haplotypes. In addition, EHHS was only considered for those SNP s with at least 5 homozygous individuals (sample-wise), and the integration was stopped if less than 2 individuals remained homozygous at SNP \( t \) (Klassmann et al., 2021). The unstandardized log-ratios of inES scores between populations were obtained by

\[
\text{unRsbs} = \ln \left( \frac{\text{inES}_{\text{BGS}}(s)}{\text{inES}_{\text{AGS}}(s)} \right),
\]

where inES\(_{\text{BGS}} \) and inES\(_{\text{AGS}} \) refer to groups BGS and AGS, respectively.

Finally, ln(Rsb) \((s)\) scores were obtained by standardization of the unRsbs \((s)\) ratio by the median;

\[
\ln(Rsb(s)) = \frac{\text{unRsbs} - \text{median(\text{unRsbs})}}{\text{sd(\text{unRsbs})}},
\]

where median(Rsb) and sd(Rsb) are the median and standard deviation of unRsbs computed over all analyzed markers.

The ln(Rsb) \((s)\) scores were assumed to be normally distributed under neutrality (Gautier and Naves 2011). Hence, the desired significance threshold of the ln(Rsb
(s) scores was calculated as the log_{10}-transformed reciprocal of the p-value, where p-value is associated with the neutral hypothesis (no selection) (Gautier and Naves, 2011; Weigand and Leese, 2018). To control the false-discovery rate, q-values associated with the p-values were obtained from the “qvalue” package in R (Storey et al., 2023). The SNPs with ln(Rsb (s)) ≤ −3.88 (SNPs with very high EHHS within the AGS group compared with the BGS group) had very low false-discovery rates (q < 0.01) and can be considered to be under positive selection. Consecutive significant SNPs were joined to form regions under positive selection (hereafter Rsb regions). Because imputed genotypes were used, only Rsb regions with more than one significant SNP were considered.

**RESULTS**

**ROH and ΔROH islands**

We identified a total of 34 ROH islands on chromosomes 1, 3, 6, 8, 12–15, 17, 19, 22, and 26 in the BGS group (Figure 1). The number of ROH islands within these chromosomes varied from one to 9. After concatenating the ROH islands, the total number of SNPs within the ROH islands was 1,401, with an average of 41.2 SNPs, varying from 2 to 192. H-scores of SNPs within the ROH islands varied from 0.110 to 0.170 (the average H-score was 0.055 across all 45,834 SNPs). ROH island length varied from 0.03 Mb to 10.6 Mb, with a mean of 2.06 Mb (Supplemental Table S1; Sarviaho, 2024). All SNPs within each ROH island were less than 500 Mb apart.

Similarly, in the AGS group, we identified a total 30 ROH islands on chromosomes 1–3, 13–17, 22, and 25–26 (Figure 2). The number of ROH islands within these chromosomes varied from one to 7. After concatenating the ROH islands, the total number of SNPs within the ROH islands was 1,396, with an average of 46.5 SNPs, varying from 2 to 203. H-scores of SNPs within the ROH islands varied from 0.110 to 0.170 (the average H-score was 0.055 across all 45,834 SNPs). ROH island length varied from 0.03 Mb to 10.6 Mb, with a mean of 2.06 Mb (Supplemental Table S1; Sarviaho, 2024). All SNPs within each ROH island were less than 500 Mb apart.

Out of 159 significant SNPs, 60 were single SNPs without surrounding significant SNPs, and were discarded. A total of 99 SNPs formed 31 Rsb regions on chromosomes 2, 3, 14, 18, 20, and 25 (Supplemental Table S5, Supplemental Figure S1; Sarviaho, 2024). All SNPs within these regions were less than 500 kb apart from each other. Rsb region length ranged from 0.02 to 0.44 Mb. The total area covered by the observed Rsb regions (2.89 Mb) and their mean length (0.09 Mb) were relatively small compared with the coverage (61.4 Mb) and mean length (2.79 Mb) of the ΔROH islands. Out of 31 observed Rsb regions, 24 overlapped 7 ΔROH islands completely or partially on chromosomes 2, 3, 14, 18, and 25 (Table 1). One of these ΔROH islands was a novel ΔROH island (ΔROH_21).

**DISCUSSION**

A limited number of methods are available for detecting selection signatures between 2 or more populations using unphased SNP genotype data. The variety is even narrower with methods that do not require the polarization of alleles or recombination maps. In our study, we applied 3 complementary methods to detect signatures of selection using unphased and unpolarized genotype data in the FAY population: one based on ROH islands and one based on the standardized log-ratios of integrated site-specific extended haplotype homozygosity between pairs of populations (Rsb).

**Pruning the data**

As UMD3 bovine genome assembly has been used in designing the Illumina BovineLD v.1 Beadchip (Boichard et al., 2012), we updated the SNP positions according to the ARS-UCD 1.2 assembly. This resulted in the exclusion of 1,080 SNPs out of 46,914. Because of this,
it is possible that certain genotypes have been imputed erroneously. Moreover, some ROHs have possibly been spliced or slightly elongated due to updated positions of the SNPs. Nevertheless, updating the SNP positions should have minor effects on the final results of this study because i) we allowed one heterozygote within the sliding window in ROH detection, ii) we allowed 2 nonsignificant SNPs between SNPs within ROH island and ΔROH islands, iii) we did not consider regions with only one SNP as a ROH island or ΔROH island, or a Rsb region, and iv) it is unlikely that long homozygous regions contained single heterozygous SNPs between many homozygous SNPs. Due to the latter, it is unlikely that ROHs were considerably elongated due to updated SNP positions, i.e., a homozygous SNP added between 2 ROHs would unlikely result in one false ROH segment. However, it is possible that updating the SNP positions causes bias in the Rsb approach in relation to imputed genotypes. More specifically, if many individuals have SNPs imputed in such positions that the adapted frequency of alleles deviates notably from the true frequency, it is possible that the integration of EHHS scores is prematurely stopped. Despite this, as genotype imputation accuracy from low density to medium is relatively high (Boichard et al., 2012), imputation errors should have negligible effects on the Rsb regions detected in our study.

We did not prune the data used in the ROH island or ΔROH island studies for deviations from HWE. In addition to being genotyping errors, deviations from HWE can also be due to selection (Cox and Kraft, 2006). Because we were interested in selection, we allowed one heterozygote in a sliding window in ROH detection to account for genotyping errors. On the contrary, we pruned the data used with the Rsb approach for significant deviations ($P < 1 \times 10^{-6}$) from HWE due to the requirement of the HWE when unphased data were used (Klassmann and Gautier, 2022). This excluded 648 SNPs from the genotype data and resulted in 45,186 markers left for further analyses.

We did not prune the genotype data for MAF either in the ROH island or ΔROH island studies or with the Rsb approach because: i) pruning for MAF could lead to fewer detected ROHs (Meyermans et al., 2020), ii) notable changes in the frequency of SNPs with MAF < 0.05 were unlikely, and iii) pruning for MAF could cause considerable loss of information when the Rsb approach was used (Klassmann et al., 2021). Finally, Meyermans et al. (2020) do not recommend pruning genotype data for linkage disequilibrium (LD) in ROH studies if medium-density genotypes are used. Because of this, we avoided false-ROHs arising from LD by setting the minimum ROH length i) in SNPs and ii) in kbs.

Figure 1. Distribution of H-scores along the genome in the BGS group. SNPs within ROH islands are highlighted in green. The blue horizontal line indicates the cut-off threshold (0.116, corresponding to a $p$-value of < 0.01) of the H-scores.
**ROH identification**

The amount of heterozygosity in the 2 groups varied little. Average heterozygosity in the BGS group was 32.7%, which, in terms of recommended parameters in PLINK, would mean a sliding window of 56 SNPs and a minimum length of ROHs 56 SNPs (Meyermans et al., 2020). In contrast, average heterozygosity in the AGS group was 32.8%, and the optimal sliding window size and the minimum ROH length was 61 SNPs. Despite this, we assumed similar average heterozygosity, even if sample size was increased, in both groups, and used equal sliding window size and a minimum ROH length of 61 SNPs for both groups.

Although ROHs resulting from linkage disequilibrium and low recombination rates in cattle are less than 1 Mb long (Ferenčaković et al., 2012), short ROHs may also result from LD that is due to selection (Sabeti et al., 2002). In addition, ROHs, whether emerged from selection or inbreeding, are spliced into shorter segments by recombination (Kirin et al., 2010; Howrigan et al., 2011). We therefore additionally limited the minimum ROH length to 500 kb. This may have led to some ROHs less than 500 kb or 61 SNPs in length to go undetected and some homozygous segments 0.5–1 Mb in length, arising from low recombination rates, considered as ROHs. However, we expect that the parameters limiting ROH length had negligible effects on the final results of this study.

**Identification of ROH- and ΔROH islands and Rsb approach**

A high frequency of ROHs at certain genomic regions in a population form ‘ROH islands’ (Nothnagel et al., 2010), which are considered selection signatures (Purffield et al., 2017; Gorssen et al., 2019). Even though, signatures of selection have been detected in various cattle breeds, ROH-patterns vary between breeds (Szmatała et al., 2016; Mastrangelo et al., 2018) which complicates the results comparison between studies. Moreover, there is no consensus in the definition of ROH islands between previous studies.

The appropriate significance threshold for considering an SNP as part of a ROH island depends on the population inbreeding level: the higher the inbreeding level, the less SNPs surpass a strict significance threshold (Gorssen et al., 2021). In previous studies, the threshold for the top significant SNPs has generally varied between 0.1% (Moscarelli et al., 2020; Gorssen et al., 2021) and 1% (Szmatała et al., 2016; Szmatała et al., 2019; Cesarani et al., 2022). In our study, the mean $F_{ROH}$ was moderate — 0.055 (SD 0.025) in the BGS group and 0.057 (SD 0.022) in the AGS group.

![Figure 2. Distribution of H-scores along the genome in the AGS group. SNPs within ROH islands are highlighted in green. The blue horizontal line indicates the cut-off threshold (0.120, corresponding to a p-value of < 0.01) of the H-scores.](image-url)
fore, 1% was used as the threshold for significance. For continuity of the ROH islands, we allowed the significant SNPs to be separated by a maximum of 2 nonsignificant SNPs to be called in a ROH island and considered the in-between SNPs to be significant. Despite this, the maximum inter-marker distances between consecutive SNPs within ROH islands and ∆ROH islands were less than 0.5 Mb, which falls within the extent of LD in cattle (MacKay et al., 2007). Nevertheless, it is possible that selection pressure is still shared by certain distinct ROH islands or ∆ROH islands. Finally, as bovine autosomes are acrocentric, with a centromere near the chromosome start, ROH islands emerging due to the centromeric location of ROHs are expected to locate within the few first megabases of a chromosome (Nandolo et al., 2018).

Due to the small number of studies on ∆ROH islands, the requisites for calling a ∆ROH island were

Table 1. Selection signatures that emerged after the introduction of genomic selection. Positions are based on ARS-UCD1.2 assembly

<table>
<thead>
<tr>
<th>∆ROH island</th>
<th>BTA</th>
<th>Start (bp)</th>
<th>End (bp)</th>
<th>Overlaps ROH_AGS</th>
<th>Overlaps ROH_BGS</th>
<th>Overlaps Rsb</th>
</tr>
</thead>
<tbody>
<tr>
<td>∆ROH_2</td>
<td>2</td>
<td>67,703,328</td>
<td>76,276,484</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_3*</td>
<td>2</td>
<td>119,815,236</td>
<td>124,592,874</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_4</td>
<td>3</td>
<td>66,511,334</td>
<td>68,294,393</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_5</td>
<td>3</td>
<td>68,508,313</td>
<td>71,230,722</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_7</td>
<td>3</td>
<td>82,579,661</td>
<td>88,821,997</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_9*</td>
<td>13</td>
<td>40,115,588</td>
<td>40,200,790</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_10</td>
<td>14</td>
<td>36,225,604</td>
<td>45,459,160</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_11*</td>
<td>14</td>
<td>45,772,293</td>
<td>46,102,133</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_16*</td>
<td>16</td>
<td>68,000,619</td>
<td>68,691,616</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_17*</td>
<td>16</td>
<td>69,398,109</td>
<td>70,587,239</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_18*</td>
<td>16</td>
<td>70,836,094</td>
<td>71,066,302</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>∆ROH_19</td>
<td>18</td>
<td>44,323,486</td>
<td>52,399,424</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_21*</td>
<td>25</td>
<td>5,760,199</td>
<td>13,934,351</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>∆ROH_22*</td>
<td>26</td>
<td>19,531,798</td>
<td>22,144,981</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

1Novel ∆ROH island.
the same as for the ROH islands. However, it is likely that not all observed ΔROH islands are regions truly under positive selection due to the limited sample size of animals born before 2012.

Similar to ROH island studies, there is no consensus of the appropriate thresholds to use for detecting the regions under selection with the Rsb approach. Commonly, selection is inferred using a standard normal distribution (Weigand and Leese, 2018), and the threshold of significance for SNPs under selection has generally varied between ln(Rsb) ≤ ± 2 and ln(Rsb) ≤ ± 4, corresponding to p-values associated with the neutral hypothesis (no selection) between 0.01 and 0.0001, respectively (e.g., Goszczyński et al., 2018; Mastrangelo et al., 2020). To control the false-discovery rate, we calculated q-values associated with the p-values, and only the SNPs with very low false-discovery rates (q < 0.01) were considered positive selection.

**Observed selection signatures**

FAY belong to the Nordic Red breeds that are selected based on the NTM (NAV, 2023A). Production traits (milk, protein, and fat) have the highest weights in the NTM (NAV, 2023A). As expected, we observed ΔROH islands overlapping genes related to milk yield and composition. For example, MECR, CHUK, and COX15, candidate genes for milk fatty acid traits (Ibeagha-Awemu et al., 2016; Palombo et al., 2018; Shi et al., 2019), overlap the novel selection signatures ΔROH_3 and ΔROH_22. Milk fatty acids are correlated with milk fat and milk protein contents (Mele et al., 2009). Additionally, PLA2G4A, a gene associated with milk yield in Holstein (Lu et al., 2022), is located in the novel selection signature region ΔROH_16. Furthermore, QTL regions related to milk fat yield in the RDC breed (Iso-Touru et al., 2016; Fang et al., 2018) overlap the novel selection signature ΔROH_21. Selection signatures have also been identified in these regions in Holstein (Szmatała et al., 2016; Mastrangelo et al., 2018; Szmatała et al., 2019; Lozada-Soto et al., 2022). Furthermore, SEMA4G was located in the novel selection signature region ΔROH_22. Changes in the mRNA expression of SEMA4G, in relation to the amount of dietary energy, have been reported in nonlactating Holstein cows (Moisá et al., 2017). Selection signatures overlapping these genes have also been identified in Holstein and Jersey (Szmatała et al., 2016; Mastrangelo et al., 2018; Szmatała et al., 2019; Lozada-Soto et al., 2022).

Fertility has the second highest weight in the NTM (NAV, 2023A). RICTOR, a regulator for highly expressed genes in fertile cows (Matsuyama et al., 2022) is within ΔROH_20. Selection signatures overlapping RICTOR have also been identified in Holstein (Szmatała et al., 2016; Mastrangelo et al., 2018; Szmatała et al., 2019; Lozada-Soto et al., 2022). In addition, a QTL region for fertility (Fang et al., 2018) overlaps the above-mentioned selection signature ΔROH_21. Although FAY is most importantly a dairy breed, a growth index is also included in the NTM (NAV, 2023A). ANGPTL3, a gene associated with growth and meat quality traits in cattle (Chen et al., 2015), is located in region ΔROH_7. This ΔROH island was located within ROH islands observed in the BGS and AGS groups. In addition, SHISA9, a candidate gene for maturity rate in beef cattle (Duan et al., 2021) is located in the novel ΔROH_21. Additionally, SLC44A5, associated with birth weight in Holstein (Sugimoto et al., 2012) overlaps ΔROH_5, and HSPB6, influencing the intramuscular fat and meat and carcass traits in dual-purpose cattle (Picard et al., 2023), overlaps ΔROH_19.

Interestingly, the previously mentioned RICTOR, which overlaps ΔROH_20, is a candidate gene also for metabolic efficiency in dairy cattle (Waerp et al., 2019), a trait included in the ‘Saved feed’ index in NTM (NAV 2023A). The same region (ΔROH_20 overlapping RICTOR also overlaps OSMR, a gene associated with feed efficiency in beef cattle (Abo-Ismail et al., 2018). Selection signatures covering RICTOR and OSMR have previously been identified in Holstein (Szmatała et al., 2016; Mastrangelo et al., 2018; Szmatała et al., 2019; Lozada-Soto et al., 2022). Furthermore, SEMA4G was located in the novel selection signature region ΔROH_22. Changes in the mRNA expression of SEMA4G, in relation to the amount of dietary energy, have been reported in nonlactating Holstein cows (Moisá et al., 2017). Selection signatures overlapping these genes have also been identified in Holstein and Jersey (Szmatała et al., 2016; Mastrangelo et al., 2018; Szmatała et al., 2019; Lozada-Soto et al., 2022).

**CONCLUSIONS**

In conclusion, we detected selection signatures that were related to milk production, fertility, growth, and feed efficiency in FAY. The results of this study indicate that, after genomic selection has been introduced, selection has favored genomic regions related to traits relevant in the FAY breeding program. Because homozygosity within these genomic regions has increased, the effect of genotypes in a homozygous state within the selected regions should be studied. In addition, we recommend monitoring the level of homozygosity within these regions that are undergoing selection in the FAY population.

**ACKNOWLEDGMENTS**

The authors wish to thank Naudanjalostussäätiö for funding this research. We are grateful for NAV (Nordic Cattle Genetic Evaluation, Aarhus, Denmark) and Faba (The Finnish Animal Breeding Association, Hollola, Finland) for providing the data.
Funding information Naudanjalostussäätiö has funded this research.

Data availability The data supporting the findings of this study are available from the Nordic Cattle Genetic Evaluation (NAG), Aarhus, Denmark and the Finnish Animal Breeding Association (Faba), Hollola, Finland. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the corresponding author upon reasonable request and with permission from NAV and Faba.

Conflicts of interest The authors have no conflict of interest to declare.

REFERENCES


Sarviaho et al.: Signatures selection in the Finnish Ayrshire


Sarviaho et al.: Signatures selection in the Finnish Ayrshire


ORCIDS

Katri Sarviaho @ https://orcid.org/0000-0002-8810-1177
Pekka Uimari @ https://orcid.org/0000-0002-5349-3932
Katja Martikainen @ https://orcid.org/0000-0002-0693-8687