Genome-wide identification of runs of homozygosity islands and associated genes in local dairy cattle breeds


1Dipartimento Scienze Agrarie, Alimentari e Forestali, Università degli Studi di Palermo, 90128 Palermo, Italy; 2Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, 40127 Bologna, Italy

(Received 25 September 2017; Accepted 23 February 2018)

Runs of homozygosity (ROH) are widely used as predictors of whole-genome inbreeding levels in cattle. They identify regions that have an unfavorable effect on a phenotype when homozygous, but also identify the genes associated with traits of economic interest present in these regions. Here, the distribution of ROH islands and enriched genes within these regions in four dairy cattle breeds were investigated. Cinisara (71), Modicana (72), Reggiana (168) and Italian Holstein (96) individuals were genotyped using the 50K v2 Illumina BeadChip. The genomic regions most commonly associated with ROHs were identified by selecting the top 1% of the single nucleotide polymorphisms (SNPs) most commonly observed in the ROH of each breed. In total, 11 genomic regions were identified in Cinisara and Italian Holstein, and eight in Modicana and Reggiana, indicating an increased ROH frequency level. Generally, ROH islands differed between breeds. The most homozygous region (>45% of individuals with ROH) was found in Modicana on chromosome 6 within a quantitative trait locus affecting milk fat and protein concentrations. We identified between 126 and 347 genes within ROH islands, which are involved in multiple signaling and signal transduction pathways in a wide variety of biological processes. The gene ontology enrichment provided information on possible molecular functions, biological processes and cellular components under selection related to milk production, reproduction, immune response and resistance/susceptibility to infection and diseases. Thus, scanning the genome for ROH could be an alternative strategy to detect genomic regions and genes related to important economic traits.

Keywords: runs of homozygosity islands, genomic regions, candidate genes, local dairy cattle, bovine beadchip 50K

Implications

The genomic regions subjected to selection tend to generate runs of homozygosity (ROH) islands or hotspots. The aim of this work was to identify the differences between breeds and use the location of ROH islands to identify genes potentially involved in economically important traits. We identified several genes within ROH involved in a wide variety of biological processes, such as milk yield and composition, reproduction, immune response, resistance/susceptibility to infectious and diseases. These results showed that scanning the genome for ROH could be an alternative strategy to detect genomic regions and genes related with important economically traits.

Introduction

The development of single nucleotide polymorphism (SNP) arrays to scan the genome allow us to distinguish non-autozygotic segments that are identical by state from autozygotic and identical by descent segments (Peripolli et al., 2016). A potential alternative method, called ROH, has been used in livestock for the identification of homozygous genomic regions (Purfield et al., 2012; Ferenčaković et al., 2013a). Runs of homozygosity are contiguous lengths of homozygous genotypes that are present in an individual because the parents transmitted identical by descending haplotypes to their offspring (Gibson et al., 2006). Runs of homozygosity has been widely used as predictors of whole-genome inbreeding levels (Zhang et al., 2015a; Mastrangelo et al., 2016). Moreover, ROH have been used in livestock genomic studies, confirming the correlation between shared ROH and genomic regions putatively under selection (Kim et al., 2013; Gaspa et al., 2014; Metzger et al., 2015; Szmatora et al., 2016; Kukučková et al., 2017; Purfield et al., 2017). In fact, the genomic regions subjected to selection frequently show signatures, such as reduced nucleotide diversity, and tend to generate ROH islands or hotspots, which have high levels of homozygosity around a selected
locus compared with the rest of the genome (Szmata et al., 2016; Purfield et al., 2017). Runs of homozygosity islands are not randomly distributed across the genome and are shared among individuals within a breed (Zhang et al., 2015b).

A large number of cattle breeds are defined by marked phenotypic differences and, therefore, constitute valuable models to study genome evolution in response to processes such as selection and domestication. Thus, in livestock species, ROH may contribute to the detection of genomic regions that could explain phenotypic differences among breeds that affect traits of economic importance. We previously described ROH structures in three local cattle breeds (Reggiana, Cinisara and Modicana) and in Holstein cattle (Mastrangelo et al., 2016). The aim of this work was to further study the distribution of ROH islands across the genome of these four cattle breeds, which may provide insights into the mechanisms underlying their genomic differences. In addition, it aimed to characterize ROH islands and identify enriched genes that could potentially explain the effects of these homozygous regions on economically important traits.

Material and methods

Samples, genotyping and data filtering
A total of 407 animals (Cinisara = 71, Modicana = 72, Reggiana = 168 and Italian Holstein = 96) were used for the analyses. All of the individuals were genotyped using the Illumina BovineSNP50 v2 BeadChip assay (Illumina Inc., San Diego, CA, USA). Single nucleotide polymorphisms were filtered to exclude loci assigned to unmapped contigs, and only those SNPs located on autosomes were considered. Quality control included call frequency \(\geq 0.95\), minor allele frequency (MAF) \(\geq 0.01\), and Hardy–Weinberg Equilibrium with a \(P > 0.001\). SNPs that did not satisfy these quality criteria were excluded. Single nucleotide polymorphisms were mapped using the Bos taurus UMD 3.1.1 genome assembly.

Genetic relationship between individuals
The genetic relationship among individuals was estimated by principal components analysis (PCA) of genetic distances. This analysis was based on the identity by state (IBS) matrices of genetic distances between individuals. Principal components analysis of the genetic distance (\(D\)) matrix was performed using the multidimensional scaling option in PLINK v.1.07 (Purcell et al., 2007). The graphical representation was depicted using the statistical R software (http://www.R-project.org/).

Runs of homozygosity detection
Runs of homozygosity were estimated, for each individual, using a sliding window approach of 50 SNPs in PLINK v.1.07 (Purcell et al., 2007). The minimum length that constituted the ROH was set to 4 Mb. The density of the SNP panel used to generate data for ROH identification is an important factor that strongly affects autozygosity estimates. The 50K panel overestimates the number of small segments (Purfield et al., 2012; Ferenčaković et al., 2013b). The following criteria were used to define the ROH: (i) one missing SNP was allowed in the ROH and up to one possible heterozygous genotype, (ii) the minimum number of consecutive SNPs that constituted a ROH was set to 30, (iii) minimum density of 1 SNP every 100 kb, and (iv) maximum gap between consecutive SNPs of 1 Mb.

Identification of genomic regions and genes within runs of homozygosity
To identify the genomic regions of high homozygosity, the amount of times that each SNP appeared in the ROH was considered and normalized by dividing it by the number of animals included in the analysis. These values were plotted against the position of the SNP along the chromosome. The genomic regions were defined according to Szmata et al. (2016). Adjacent SNPs having a proportion of ROH occurrences over the adopted threshold formed ROH islands. Mean linkage disequilibrium (LD) was estimated using HAPLOVIEW v. 4.2 (Barrett et al., 2005) for all pairwise combinations of SNPs within each ROH island. Genomic coordinates for all identified ROH islands were also used for the annotation of genes that were fully or partially contained within each selected region using the UCSC Genome Browser (http://genome.ucsc.edu/). The genes were further analyzed with the Panther Classification System (Mi et al., 2013) to identify significant (\(P \leq 0.05\)) gene ontology (GO) terms. Finally, to investigate the biological function of each annotated gene contained in ROH islands, an accurate literature search was also conducted.

Results

A PCA was used to visualize and explore the genetic relationships among breeds. The PCA (Figure 1) showed that breeds formed non-overlapping clusters and were clearly separated populations. After data quality and genetic relationships analyses, no outliers were detected.
A total of 44,875 SNPs in Cinisara, 42,687 SNPs in Modicana, 35,270 SNPs in Reggiana, and 41,569 SNPs in Italian Holstein cattle breeds were retained after quality control for ROH detection. The top 1% of SNPs observed in the ROH was selected, and adjacent SNPs over this threshold were merged into genomic regions corresponding to ROH islands (Szmatola et al., 2016). In ROH islands detected here, each SNP showed a percentage of occurrence $> 10\%$ (Figure 2). This approach resulted in the identification of 11 ROH islands in Cinisara and Italian Holstein, and eight in Modicana and Reggiana (Table 1). Two overlapping ROH islands were observed between breed pairs. Modicana and Reggiana breeds showed a common genomic region on Bos taurus autosome (BTA) 6 (6:38,689,886 to 39,346,170 bp) and Cinisara and Italian Holstein breeds on BTA10 (10:56,464,919 to 56,792,715 bp). The genomic distribution of ROH islands was clearly non-uniform among breeds and across autosomes (Table 1). The longest ROH island was observed in Italian Holstein on BTA10 (12.42 Mb), while the shortest one was observed in Reggiana on BTA3 (0.03 Mb). BTA6 in Modicana breed had the ROH with the highest peak (Figure 2) which consisted of 38 SNPs with an occurrence in ROH $> 45\%$ and a length of 2.05 Mb.

The mean $r^2$ value, a standard descriptive LD parameter, was estimated for all pairwise combinations of SNPs within each ROH island (Supplementary Material Table S1). In Cinisara breed, the majority of SNPs within ROH islands showed low level of LD ($< 0.080$), and $r^2$ ranged from 0.024 to 0.290. The other breeds showed intermediate levels of LD within ROH islands (from 0.006 to 0.280). The highest LD level was found in the ROH island on BTA6 in Reggiana breed (0.387).

Within all of the ROH islands here reported, we identified from 126 to 347 genes (347 Italian Holstein, 250 Modicana, 190 Cinisara and 126 Reggiana). A list of genes found in the ROH islands of each breed underwent a GO enrichment analysis. Multiple categories were statistically significant ($P \leq 0.05$). The genes within ROH islands encompass a wide spectrum of molecular function, biological process, and cellular components. A PANTHER gene list analysis revealed a high percentage of genes involved in catalytic activity (GO:0003824), cellular processes (GO:0009987), cell part (GO:0044464), metabolic processes (GO:0008152), binding (GO:0005488) as well as biological regulations (GO:0065007) and response to stimulus (GO:0050896) in all of the ROH islands of the analyzed breeds (Table 2). Supplementary Material Table S2 provides the chromosome position, number of SNPs and number of genes per genomic region, gene symbol and full name for all of the annotated genes in each breed.

**Discussion**

We analyzed animals from four Italian cattle breeds with different inbreeding background and selection histories. Mastrangelo et al. (2016), in a previous study on evaluation on ROH in these breeds, reported the highest value of...
inbreeding (F) based on ROH ($F_{ROH} = 0.055$) for Modicana, whereas Reggiana showed the lowest one ($F_{ROH} = 0.035$). The individuals of Italian Holstein and Reggiana showed high number of short ROH segments. Modicana and Cinisara showed similar results between them with the total length of ROH characterized by the presence of large segments due to a recent inbreeding. In this study, we reported the distribution of ROH islands across the genome of these cattle breeds to provide insights into the mechanisms underlying genomic differences among them.

**Genomic regions with high frequency in runs of homozygosity**

In our study, we did not perform LD pruning, but, owing to the minimum 4 Mb size of ROH segments, we tried to avoid small autozygous segments caused by LD. Indeed, a strong LD, typically extending up to ~200 kb, is common throughout the bovine genome (Mastrangelo et al., 2014), and short ROH are very prevalent. To exclude these short and very common ROH, the minimum length for ROH was set to $>4$ Mb.

The top 1% of SNPs with the highest number of occurrence was chosen as an indication of a possible ROH island in the genome. The same threshold was reported in studies on cattle (Szmatola et al., 2016) and sheep (Purfield et al., 2017). Gaspa et al. (2014) and Sölkner et al. (2014) used top regions with percentage of SNP in ROH $>40\%$ within breed, whereas Mészáros et al. (2015) applied a threshold of 10%. Recently, a common ROH proportion higher than 7.5% was chosen as an indicator of potential autozygosity islands in the genome.

### Table 1

<table>
<thead>
<tr>
<th>Breed</th>
<th>BTA</th>
<th>Number of SNPs</th>
<th>Number of genes</th>
<th>Start bp</th>
<th>End bp</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cinisara</td>
<td>8</td>
<td>5</td>
<td>0</td>
<td>18 112 643</td>
<td>18 420 652</td>
<td>308 010</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>56 464 919</td>
<td>56 792 715</td>
<td>327 797</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>30 530 185</td>
<td>30 878 341</td>
<td>348 157</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>23</td>
<td>22</td>
<td>43 922 935</td>
<td>45 552 538</td>
<td>1 629 604</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>84</td>
<td>16</td>
<td>60 163</td>
<td>6 423 288</td>
<td>6 363 126</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>7</td>
<td>6</td>
<td>10 870 036</td>
<td>11 251 946</td>
<td>381 911</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>5</td>
<td>2</td>
<td>13 517 193</td>
<td>13 793 884</td>
<td>276 692</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>69</td>
<td>49</td>
<td>27 655 543</td>
<td>32 996 400</td>
<td>5 340 858</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>88</td>
<td>20</td>
<td>34 157 181</td>
<td>39 007 759</td>
<td>4 850 579</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>9</td>
<td>1</td>
<td>39 700 262</td>
<td>40 191 764</td>
<td>491 503</td>
</tr>
<tr>
<td>Residuated</td>
<td>28</td>
<td>136</td>
<td>60</td>
<td>40 782 045</td>
<td>46 224 056</td>
<td>5 441 652</td>
</tr>
<tr>
<td>Italian Holstein</td>
<td>3</td>
<td>35</td>
<td>27</td>
<td>91 930 742</td>
<td>93 497 168</td>
<td>1 566 427</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>99 527 745</td>
<td>99 569 438</td>
<td>41 694</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>49 145 480</td>
<td>49 715 020</td>
<td>569 541</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>83</td>
<td>71</td>
<td>34 907 534</td>
<td>40 294 545</td>
<td>5 387 012</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>214</td>
<td>76</td>
<td>49 889 790</td>
<td>62 309 052</td>
<td>12 419 263</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>47</td>
<td>17</td>
<td>63 095 461</td>
<td>67 118 053</td>
<td>4 022 593</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>61</td>
<td>100</td>
<td>51 880 463</td>
<td>56 190 025</td>
<td>4 309 563</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>33</td>
<td>11</td>
<td>24 266 877</td>
<td>26 460 587</td>
<td>2 193 711</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>28</td>
<td>16</td>
<td>29 545 545</td>
<td>31 848 979</td>
<td>2 303 435</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>27</td>
<td>8</td>
<td>34 817 221</td>
<td>36 570 529</td>
<td>1 753 309</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>26</td>
<td>19</td>
<td>19 727 292</td>
<td>21 226 405</td>
<td>1 499 114</td>
</tr>
<tr>
<td>Modicana</td>
<td>1</td>
<td>39</td>
<td>25</td>
<td>130 168 696</td>
<td>132 182 348</td>
<td>2 013 653</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>46</td>
<td>4</td>
<td>35 763 942</td>
<td>37 877 098</td>
<td>2 113 157</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>113</td>
<td>20</td>
<td>51 406 099</td>
<td>57 744 446</td>
<td>6 338 348</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>84</td>
<td>136</td>
<td>27 542 987</td>
<td>33 508 142</td>
<td>5 965 156</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>51</td>
<td>18</td>
<td>78 776 781</td>
<td>82 786 530</td>
<td>4 009 750</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>112</td>
<td>25</td>
<td>34 324 052</td>
<td>41 343 408</td>
<td>7 019 357</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>53</td>
<td>4</td>
<td>29 767 566</td>
<td>32 749 041</td>
<td>2 981 476</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>10</td>
<td>1</td>
<td>40 422 559</td>
<td>40 921 256</td>
<td>498 698</td>
</tr>
<tr>
<td>Reggiana</td>
<td>1</td>
<td>25</td>
<td>14</td>
<td>150 141 293</td>
<td>151 550 746</td>
<td>1 409 454</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>15</td>
<td>3</td>
<td>151 736 540</td>
<td>152 412 336</td>
<td>675 997</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>71 141 852</td>
<td>71 167 977</td>
<td>26 126</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>71</td>
<td>28</td>
<td>73 035 441</td>
<td>79 378 528</td>
<td>6 343 088</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>38 689 886</td>
<td>39 346 170</td>
<td>656 285</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>97</td>
<td>31</td>
<td>56 941 968</td>
<td>61 788 328</td>
<td>4 846 361</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>23</td>
<td>8</td>
<td>9 078 964</td>
<td>10 441 474</td>
<td>1 362 511</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>74</td>
<td>28</td>
<td>15 819 913</td>
<td>23 142 122</td>
<td>7 322 210</td>
</tr>
</tbody>
</table>
cattle (Kukučková et al., 2017). Therefore, we have employed a stricter criteria compared with the last two works mentioned above.

The ROH peaks were distributed and shared among individuals, and it was clear that they were signs of common ROH islands within breeds. Some of these genomic regions overlapped with ROH islands found in other studies (Table 3). The ROH islands reported on BTA4 and BTA5 in Modicana overlapped with ROH islands reported in Pinzgau (Kukučková et al., 2017) and Simmental (Szamatola et al., 2016). Several studies (Sölkner et al., 2014; Mészáros et al., 2015; Szamatola et al., 2016; Kukučková et al., 2017) showed ROH islands located on BTA6. These regions overlapped with the ones obtained in our study for Modicana (34.32 to 41.34 Mb) and Reggiana (38.69 to 39.35 Mb). Sölkner et al. (2014) studying Taurine and Indicine cattle breeds, identified a region in BTA16 (43.80 to 44.97 Mb) visible only in Taurine. This overlapped with a region obtained in our study in

Candidate genes within runs of homozygosity

<table>
<thead>
<tr>
<th>Breeds</th>
<th>Molecular function (GO) terms enriched (P &lt; 0.05) based on runs of homozygosity islands and number of involved genes (n) for each cattle breed</th>
<th>Biological process (GO) terms enriched (P &lt; 0.05) based on runs of homozygosity islands and number of involved genes (n) for each cattle breed</th>
<th>Cellular component (GO) terms enriched (P &lt; 0.05) based on runs of homozygosity islands and number of involved genes (n) for each cattle breed</th>
</tr>
</thead>
</table>
| **Cinisara** | Binding (GO:0005488) n = 32  
Receptor activity (GO:0004872) n = 7  
Structural molecule activity (GO:0005198) n = 2  
Signal transducer activity (GO:0004871) n = 6  
Catalytic activity (GO:0003824) n = 33  
Transporter activity (GO:0005215) n = 11 | Cellular component organization (GO:0071840) n = 18  
Cellular process (GO:0009987) n = 57  
Localization (GO:0051179) n = 20  
Reproduction (GO:0000003) n = 5  
Biological regulation (GO:00065007) n = 36  
Response to stimulus (GO:0050896) n = 20  
Developmental process (GO:0032502) n = 15  
Immune System process (GO:0002376) n = 1  
Multicellular organismal process (GO:0032501) n = 15  
Biological adhesion (GO:0022610) n = 1  
Locomotion (GO:0040011) n = 1  
Metabolic process (GO:0006152) n = 50  
Growth (GO:0040007) n = 1 | Membrane (GO:0016020) n = 12  
Macromolecular complex (GO:0032991) n17  
Cell part (GO:0044464) n = 52  
Organelle (GO:0043226) n = 32  
Extracellular region (GO:0005576) n = 12  
Synapse (GO:0045202) n = 1 |
| **Modicana** | Binding (GO:0005488) n = 37  
Receptor activity (GO:0004872) n = 20  
Structural molecule activity (GO:0005198) n = 10  
Signal transducer activity (GO:0004871) n = 17  
Catalytic activity (GO:0003824) n = 30  
Transporter activity (GO:0005215) n = 11 | Cellular component organization (GO:0071840) n = 19  
Cellular process (GO:0009987) n = 82  
Localization (GO:0051179) n = 10  
Biological regulation (GO:00065007) n = 42  
Response to stimulus (GO:0050896) n = 40  
Developmental process (GO:0032502) n = 29  
Locomotion (GO:0040011) n = 11  
Biological adhesion (GO:0022610) n = 3  
Metabolic process (GO:0006152) n = 45  
Growth (GO:0040007) n = 5  
Immune System process (GO:0002376) n = 7 | Cell junction (GO:0030054) n = 2  
Membrane (GO:0016020) n = 9  
Macromolecular complex (GO:0032991) n = 29  
Extracellular matrix (GO:0031012) n = 2  
Cell part (GO:0044464) n = 57  
Organelle (GO:0043226) n = 35  
Extracellular region (GO:0005576) n = 8 |
| **Reggiana** | Binding (GO:0005488) n = 28  
Receptor activity (GO:0004872) n = 1  
Structural molecule activity (GO:0005198) n = 3  
Signal transducer activity (GO:0004871) n = 6  
Catalytic activity (GO:0003824) n = 30  
Transporter activity (GO:0005215) n = 11 | Cellular component organization (GO:0071840) n = 6  
Cellular process (GO:0009987) n = 44  
Localization (GO:0051179) n = 11  
Biological regulation (GO:00065007) n = 14  
Response to stimulus (GO:0050896) n = 11  
Developmental process (GO:0032502) n = 11  
Multicellular organismal process (GO:0032501) n = 2  
Biological adhesion (GO:0022610) n = 2  
Metabolic process (GO:0006152) n = 31  
Growth (GO:0040007) n = 5  
Immune System process (GO:0002376) n = 7 | Synapse (GO:0045202) n = 1  
Membrane (GO:0016020) n = 4  
Macromolecular complex (GO:0032991) n = 11  
Cell part (GO:0044464) n = 35  
Organelle (GO:0043226) n = 22 |
| **Italian**  | Translation regulator activity (GO:0045182) n = 1  
Binding (GO:0005488) n = 74  
Receptor activity (GO:0004872) n = 12  
Structural molecule activity (GO:0005198) n = 10  
Signal transducer activity (GO:0004871) n = 8  
Catalytic activity (GO:0003824) n = 86  
Transporter activity (GO:0005215) n = 19 | Cellular component organization (GO:0071840) n = 31  
Cellular process (GO:0009987) n = 138  
Localization (GO:0051179) n = 37  
Reproduction (GO:0000003) n = 8  
Biological regulation (GO:00065007) n = 63  
Response to stimulus (GO:0050896) n = 35  
Developmental process (GO:0032502) n = 33  
Multicellular organismal process (GO:0032501) n = 20  
Biological adhesion (GO:0022610) n = 7  
Locomotion (GO:0040011) n = 7  
Metabolic process (GO:0006152) n = 112  
Immune System process (GO:0002376) n = 10 | Synapse (GO:0045202) n = 2  
Cell junction (GO:0030054) n = 3  
Membrane (GO:0016020) n = 17  
Macromolecular complex (GO:0032991) n = 39  
Cell part (GO:0044464) n = 121  
Organelle (GO:0043226) n = 64  
Extracellular region (GO:0005576) n = 8 |
Comparison among overlapped runs of homozygosity (ROH) islands here detected and those reported in previous studies

<table>
<thead>
<tr>
<th>Breed</th>
<th>BTA Position (Mb)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinzgau</td>
<td>4 52.42 to 65.05</td>
<td>Kukučková et al. (2017)</td>
</tr>
<tr>
<td>Modicana</td>
<td>4 51.41 to 57.74</td>
<td>This study</td>
</tr>
<tr>
<td>Simmental</td>
<td>5 78.71 to 80.94</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Modicana</td>
<td>5 78.78 to 82.79</td>
<td>This study</td>
</tr>
<tr>
<td>Pinzgau</td>
<td>6 35.46 to 42.31</td>
<td>Kukučková et al. (2017)</td>
</tr>
<tr>
<td>Tyrol Grey</td>
<td>6 36.28 to 41.12</td>
<td>Mészáros et al. (2015)</td>
</tr>
<tr>
<td>Modicana</td>
<td>6 34.32 to 41.34</td>
<td>This study</td>
</tr>
<tr>
<td>Simmental</td>
<td>6 38.34 to 40.10</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Taurine</td>
<td>6 38.27 to 39.45</td>
<td>Sölckner et al. (2014)</td>
</tr>
<tr>
<td>Reggiana</td>
<td>6 38.69 to 39.35</td>
<td>This study</td>
</tr>
<tr>
<td>Taurine</td>
<td>16 43.80 to 44.97</td>
<td>Sölckner et al. (2014)</td>
</tr>
<tr>
<td>Red Polish</td>
<td>16 43.52 to 46.19</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Simmental</td>
<td>16 42.89 to 46.77</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Limousin</td>
<td>16 43.37 to 46.07</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Cinisara</td>
<td>16 43.92 to 45.55</td>
<td>This study</td>
</tr>
<tr>
<td>Holstein</td>
<td>20 28.33 to 32.29</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Italian Holstein</td>
<td>20 29.54 to 31.85</td>
<td>This study</td>
</tr>
<tr>
<td>Holstein</td>
<td>20 34.47 to 35.48</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Italian Holstein</td>
<td>20 34.82 to 36.57</td>
<td>This study</td>
</tr>
<tr>
<td>Holstein</td>
<td>26 21.15 to 23.00</td>
<td>Gaspa et al. (2014)</td>
</tr>
<tr>
<td>Italian Holstein</td>
<td>26 19.73 to 21.23</td>
<td>This study</td>
</tr>
<tr>
<td>Simmental</td>
<td>28 39.77 to 40.57</td>
<td>Szmatola et al. (2016)</td>
</tr>
<tr>
<td>Cinisara</td>
<td>28 39.70 to 40.19</td>
<td>This study</td>
</tr>
</tbody>
</table>

Some ROH islands actually overlapped with regions having low recombination rates (Table 1) as reported in previous studies in sheep (Purfield et al., 2017). Moreover, a previous study on cattle (Purfield et al., 2012) reported a correlation between extensive LD and high incidence of ROH. The majority of SNPs within ROH islands showed similar LD levels as those computed for the entire chromosome (Supplementary Material Table S2), with the exception of two ROH islands (on BTA16 in Cinisara and on BTA 6 in Reggiana). Therefore, their existence was not easy explained on the basis of just LD (Nothnagel et al., 2009).

Identification of candidate genes within runs of homozygosity

We found that some SNPs occurred in regions of poor gene content. Some of the identified ROH islands, such as on BTA10 in the Cinisara breed, contained only one annotated gene (WDR72) or uncharacterized genes (i.e. LOC107132862). This may reflect selection acting on uncharacterized regulatory regions or simply the fixation of non-coding DNA by genetic drift due to the absence of any selection (Qanbari et al., 2011). An enrichment of genes involved in several GO-terms was observed in the four cattle breeds. We have not discussed in detail all of the genomic regions associated with ROH islands. Instead, we focused on selected genes in highly GO-enriched terms that, on the basis of the literature, showed associations with several specific traits related to livestock. Therefore, the functions of candidate genes within ROH islands play important roles in cattle and other livestock species are summarized for each breed.

In Cinisara, the ROH islands were identified on BTA8, 10, 13, 16, 23 and 28. A total of 40 genes were identified as being related to catalytic activity (GO:0003824), with genes implicated in immune response and immune regulation (PIK3CD and SPSB1, respectively) (Ramey et al., 2013). A high number of genes (n = 57) were identified as being related to cellular process (GO:0009987). Among these, some candidate genes mapped on BTA16, such as PEX14, which is related to dairy production, KIF1B, which is under...
strong selection in dairy Holstein cattle (Flori et al., 2009), and RERE, which is implicated in embryonic growth and reproductive development (Ramey et al., 2013). Moreover, 52 identified genes were also related to cell part (GO:0044464) in which we highlighted the ADK gene on BTA28, which is involved in a physiological state (Ramey et al., 2013). Other candidate genes within the ROH islands on BTA28 were NRG3 and PPHY1, which are related with bovine mammary gland development and milk production, respectively (Ogorevc et al., 2009).

The ROH islands in the Modicana breed were identified on BTA1, 4, 5, 6, and 8. A total of 37, 82 and 69 genes were identified as being related to binding (GO:0005488), cellular process (GO:0009987) and cell part (GO:0044464), respectively. Several enriched GO-terms contained genes related with milk production, such as the LALBA gene, a major whey protein that showed a significant association with the milk protein profile (Huang et al., 2012). On BTA6, the most homogenous region (>45% of individuals having the ROH island) was found (6:37 019 972 to 39 069 719 bp) and it contained an intriguing element. A quantitative trait locus (QTL) on this chromosome affecting milk fat and protein concentrations has been reported (Zhang et al., 1998). The QTL, containing six genes (ABCG2, PKD2, SPP1, MEPE, IBSP and LAP3), was identified within one ROH island in our study. In this chromosomal region, several genes associated with milk production traits are annotated, such as FAM13A1, a gene involved in mammary gland function and milk production, respectively (Ogorevc et al., 2009). The ROH islands in the Modicana breed were identified on BTA1, 4, 5, 6, and 8. A total of 37, 82 and 69 genes were identified as being related to binding (GO:0005488), cellular process (GO:0009987) and cell part (GO:0044464), respectively. Several enriched GO-terms contained genes related with milk production, such as the LALBA gene, a major whey protein that showed a significant association with the milk protein profile (Huang et al., 2012). On BTA6, the most homogenous region (>45% of individuals having the ROH island) was found (6:37 019 972 to 39 069 719 bp) and it contained an intriguing element. A quantitative trait locus (QTL) on this chromosome affecting milk fat and protein concentrations has been reported (Zhang et al., 1998). The QTL, containing six genes (ABCG2, PKD2, SPP1, MEPE, IBSP and LAP3), was identified within one ROH island in our study. In this chromosomal region, several genes associated with milk production traits are annotated, such as FAM13A1, a gene involved in mammary gland function and milk production, respectively (Ogorevc et al., 2009).

As reported above, several enriched GO-terms were related to milk production, reproduction, immune response, and resistance/susceptibility to infections and diseases. This indicated that the analyzed individuals may have experienced selective pressure on their genomes for these specific traits. Some genomic regions may be fixed in individuals within a population as a result of artificial or natural selection for reasons such as adaptability or productivity. Cinisara and Modicana are two breeds that have excellent abilities to adapt to harsh environments, high resistance levels to infections and diseases, good maternal aptitudes, and high-quality milk production. Genes that are involved in these traits were detected in our study using the ROH approach and were consistent with the phenotypic characteristics of these two breeds. Recently, a study on local sheep breeds (Mastrangelo et al., 2017) revealed the presence of ROH islands in genomic regions that harbor candidate genes for selection in response to environmental stress and which underlie local adaptation. The presence of many immune system-related genes in the identified ROH islands could reflect selection (natural or artificial) for disease resistance. Reggiana and Italian Holstein are two breeds reared and selected for milk production, and in accordance with this phenotypic trait, our results emphasized the presence of dairy-related genes within the ROH islands. Currently, in
dairy cattle, such as Holstein, the systemic decline of fertility is being observed, in agreement with the several genes implicated in affecting the reproductive traits highlighted in this work. Kim et al. (2013) found that several genomic regions within ROH were associated with economically important traits, including milk, fat and protein yields. Therefore, the annotated genes that mapped to these ROH islands were perceived as exposed to selection.

Conclusion
In this work, we examined the distributions of ROH islands across the genomes of four cattle breeds with similar production aptitudes but different selection histories. We confirmed that the ROH islands were clearly non-uniform among breeds, subject to selective pressure. Our results contributed to understanding how selection can shape the distribution of ROH islands and suggested that ROH islands can be used to identify genes potentially involved in economically important traits. Further research must be performed to compare selection signatures and ROH islands, and to incorporate the use of ROH island distributions across the genome to limit the number of false positives identified and to modify current procedures.

Acknowledgments
This research was financed by PON02_00451_3133441, CUP: B61C1200076005 funded by MIUR.

Declaration of interest
None.

Ethics statement
None.

Software and data repository resources
None.

Supplementary material
To view supplementary material for this article, please visit https://doi.org/10.1017/S1751731118000629

References
Candidate genes within runs of homozygosity


Perrelli S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575.


