

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Analysis of Runs of Homozygosity and their Relationship with Inbreeding in Five Cattle Breeds Farmed in Italy

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1686984> since 2019-02-08T11:31:11Z

Published version:

DOI:10.1111/age.12259

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

This is the author's final version of the contribution published as:

Marras, G. , Gaspa, G. , Sorbolini, S. , Dimauro, C. , Ajmone - Marsan, P. ,
Valentini, A. , Williams, J. L. and Macciotta, N. P. ,
**Analysis of runs of homozygosity and their relationship with inbreeding in five cattle
breeds farmed in Italy,**
ANIMAL GENETICS, 46: 110-121, 2015,
doi: 10.1111/age.12259

The publisher's version is available at:

<https://onlinelibrary.wiley.com/doi/full/10.1111/age.12259>

When citing, please refer to the published version.

Link to this full text:

<http://hdl.handle.net/2318/1686984>

This full text was downloaded from iris-AperTO: <https://iris.unito.it/>

1 **Analysis of Runs of Homozygosity and their Relationship with Inbreeding in**
2 **Five Cattle Breeds Farmed in Italy.**

3

4 Gabriele Marras*, Giustino Gaspa*¹, Silvia Sorbolini*, Corrado Dimauro*, Paolo Ajmone
5 Marsan‡, Alessio Valentini†, John L. Williams¶, Nicolò Pietro Paolo Macciotta*

6

7 * Dipartimento di Agraria–Sezione Scienze Zootecniche, Università di Sassari,Sassari07100,Italy

8 ‡Istituto di Zootecnica, Università Cattolica del Sacro Cuore, Piacenza, 29100,Italy.

9 † Dipartimento per l’innovazione dei sistemi biologici agroalimentari e forestali DIBAF, Università
10 della Tuscia, Viterbo 01000, Italy.

11 ¶ Parco Tecnologico Padano, Lodi 26900, Italy.

12

13

14 ¹Correspondingauthor: Giustino Gaspa, Dipartimento di Agraria–Sezione Scienze Zootecniche,
15 Università di Sassari,07100,Italy, Email: gigaspa@uniss.it +39079229309

16

17 **Summary**

18 Increased inbreeding is an inevitable consequence of selection in livestock populations. The
19 analysis of high density Single Nucleotide Polymorphisms (SNP) facilitates the identification of
20 long and uninterrupted “Runs of Homozygosity” (ROH) that can be used to identify chromosomal
21 regions that are identical by descent. In this work the distribution of ROH of different length in five
22 Italian cattle breeds is described. A total of 4095 bulls from 5 cattle breeds were genotyped at 54K
23 SNP loci: 2093 Italian Holstein, 749 Italian Brown, 364 Piedmontese, 410 Marchigiana and 479
24 Italian Simmental. ROH were identified and used to estimate molecular inbreeding coefficients
25 (F_{ROH}), which were compared with inbreeding coefficients estimated from pedigree information
26 (F_{PED}) and using the genomic relationship matrix (F_{GRM}). The average number of ROH per animal
27 ranged from 54 ± 7.2 in Piedmontese to 94.6 ± 11.6 in Italian Brown. The highest number of short
28 ROH (related to ancient consanguinity) were found in Piedmontese, followed by Simmental. The
29 Italian Brown and Holstein had a higher proportion of longer ROH distributed across the whole
30 genome, revealing recent inbreeding. The F_{PED} were moderately correlated with $F_{ROH>1\text{ Mb}}$ (0.662,
31 0.700 and 0.669 in BRW, HOL and SIM) but poor correlated with F_{GRM} (0.134, 0.128 and 0.448 for
32 BRW, HOL and SIM respectively). The inclusion of $ROH>8\text{ Mb}$ in the inbreeding calculation
33 improved the correlation of F_{ROH} with F_{PED} and F_{GRM} . ROH are a direct measure of autozygosity at
34 the DNA level and can overcome approximations and errors resulting from incomplete pedigree
35 data. In populations with high LD and recent inbreeding (e.g. HOL and BRW) a medium density
36 marker panel, such as the one used here, may provide a good estimate of inbreeding. However, in
37 populations with low LD and ancient inbreeding, marker density would have to be increased to
38 identify short ROH that are IBD more precisely.

39 **Keywords:** Runs of Homozygosity, Autozygosity, *Bos taurus*, molecular inbreeding

40

41

42 **Introduction**

43 The inbreeding coefficient (kinship or F) is defined as the probability that in a locus sampled
44 randomly in a population a pair of alleles is identical by descent (IBD) with respect to a base
45 population where all alleles are independent (Wright 1922). Inbreeding is a consequence of mating
46 among closely related individuals, and the resulting detrimental effects on the performance and
47 fitness of the progeny have been widely documented both in natural and domesticated animal
48 populations (Keller & Waller 2002, Charlesworth & Willis 2009; Bjelland *et al.* 2013).
49 Inbreeding is unavoidable in populations under selection, as only a sub-set of individuals is used for
50 breeding. The consequences of inbreeding are the loss of genetic variation, accumulation of
51 recessive lethal genetic mutations, worsening of performance in production traits and fertility,
52 which will impact on the profitability and sustainability of farms (Miglior *et al.* 1995; González-
53 Recio *et al.* 2007; Mc Parland *et al.* 2007; Bjelland *et al.* 2013). The inbreeding level in a
54 population could be limited using specific mating strategies, but this may also constrain the genetic
55 gain in traits under selection (Meuwissen *et al.* 1997). A key requirement for the implementation of
56 mating strategies to restrict inbreeding is accurate pedigree recording, however, a reliable pedigree
57 is not always available, especially for local breeds.

58 Inbreeding coefficients have usually been calculated from the pedigree, and the probability
59 that a pair of alleles are IBD is estimated from statistical expectations. However, the recent
60 availability of the genome sequence and information on a large numbers of SNP loci has opened
61 new opportunities to use genomic information in animal breeding, including the development of
62 genomic selection programmes (Meuwissen *et al.* 2001). Using high density single nucleotide
63 polymorphism (SNP) information the level of inbreeding can be estimated for livestock populations
64 even when no pedigree available. The probability of an allele at a locus being IBD can be estimated

65 by direct inference from the alleles inherited by an individual, which can be done for hundreds of
66 thousands of loci spanning the genome. Theoretically, using this approach the true inbreeding value
67 of an individual can be calculated using the whole genome sequence.

68 Genomic inbreeding coefficients can be calculated using two approaches: the first by
69 examining IBS information marker-by-marker, using a genomic relationship matrix (GRM,
70 VanRaden *et al.* 2008; Yang *et al.* 2010;); the second, using runs of homozygosity (ROH), which
71 was proposed first for humans by McQuillan *et al.*(2008) and later by Ferencakovic *et al.* (2011) for
72 cattle.

73 ROH are DNA segments that harbour uninterrupted stretches of loci which are homozygous
74 in the individual but that are polymorphic in the population. Broman and Weber (1999) were the
75 first to suggest that ROH are likely to be autozygous, and proposed a statistical method based on
76 LOD score to evaluate the proportion of autozygosity using data for ~8000 short tandem-repeat
77 polymorphisms (STRP) on 134 subjects in human outbreed populations. However, the presence of a
78 long stretch of homozygous loci in an individual does not necessarily imply that the region was
79 inherited from a common ancestor without recombination. Subsequently, such extended regions of
80 homozygosity (ROH) have been used to provide an estimate of autozygosity at genome-wide level
81 (McQuillan *et al.*2008; Kirin *et al.* 2010; Keller *et al.* 2011). Although ROH can arise for several
82 reasons, the primary cause of ROH is believed to be the inbreeding (Broman & Weber 1999;
83 Gibson *et al.* 2006).

84 Both microsatellite loci and dense SNP loci have been proposed for estimating inbreeding
85 (Toro *et al.* 2002; Slate *et al.* 2004; Carothens *et al.* 2006; VanRaden *et al.* 2011). The ROH
86 approach uses a genome-wide average of the proportion of sequence in ROH segments. Thus, for an
87 individual the inbreeding coefficient is the proportion of genome that is IBD, which can be
88 estimated from ROH (Keller *et al.* 2011). Genome based inbreeding metrics are particularly useful

89 when pedigrees are either incomplete or not particularly deep, as they provide an improved estimate
90 of the inbreeding coefficients. In cattle, ROH have been used to analyze population history
91 following the recent selection (Purfield *et al.* 2012), to estimate inbreeding coefficients
92 (Ferencakovic *et al.* 2011; Ferencakovic *et al.* 2013a; Ferencakovic *et al.* 2013b); to study the
93 detrimental effect of inbreeding on traits affecting farm profitability (Bjelland *et al.* 2013) and to
94 control the increase of inbreeding in genome assisted breeding schemes (Pryce *et al.* 2012).
95 Measures inbreeding using ROH are particularly appealing, as the number of generations of
96 inbreeding and history of recent selection events can be inferred from the extent and frequency of
97 ROH regions. As recombination will break long chromosome segments, it is expected that long
98 autozygous segments in an individual genome would be found when there is a recent common
99 ancestor and shorter segments would be found when the common ancestor is more distant (Broman
100 & Weber 1999). Hence, the longer the homozygous segments the more recent the inbreeding.

101 Even when deep, reliable pedigrees are available, inbreeding coefficients do not allow
102 ancient relatedness to be estimated (Ferencakovic *et al.* 2011). In exploring the inbreeding history
103 of a population, the founders are usually considered not to be inbred. Inbreeding coefficients based
104 on pedigrees ignore stochastic differences in the proportion of the genome inherited IBD relative to
105 the statistical expectation. Hence, molecular and pedigree derived inbreeding estimates tend to
106 differ. In addition, not all the ROH segments are autozygous, and short ROH may occur by chance
107 and are not IBD. Studies that compared ROH identified using SNP markers of different densities
108 found that segments shorter than 4 Mb are less likely to be IBD (Purfield *et al.* 2012; Ferencakovic
109 *et al.* 2013b). Moreover, inbreeding estimates differ depending on the length of the ROH segments
110 used to calculate the coefficients, in contrast, GRM methods give a more uniform estimate of
111 inbreeding.

112

113 The objective of this work was to compare molecular estimates of inbreeding calculated
114 from ROH (F_{ROH}) belonging to different length classes in five cattle breeds farmed in Italy and
115 characterized by different breeding goals and selection histories. The F_{ROH} coefficients were then
116 compared with the pair-wise IBS estimates using the GRM approach (F_{GRM}) and classical pedigree
117 based coefficients (F_{PED}).

118 **Materials and Methods**

119 **Data**

120 The animals used in this study consisted of 4,095 bulls from five Italian cattle breeds, specifically
121 selected to include dairy, beef and dual-purpose cattle populations. A total of 2,093 and 749 bulls
122 were selected from the Italian Holstein (HOL) and Italian Brown (BRW) dairy breeds, respectively.
123 Three hundred and sixty-four Piedmontese (PMT) and 410 Marchigiana (MAR) bulls represented
124 beef breeds. A sample of 479 dual-purpose Italian Simmental bulls (SIM) were also included in the
125 dataset. The bulls were selected using pedigree information to avoid, as far as possible, closely
126 related animals in order to maximize genetic variability. All bulls were genotyped with the Illumina
127 Bovine SNP50 bead-chip v.1 containing 54,001 SNP loci, in the framework of two big national
128 projects (Ajmone-Marsan *et al.* 2010). Markers that did not map to any chromosome, that belonged
129 to the X chromosome or had within-breed call rate $\leq 97.5\%$, were removed from dataset. SNP that
130 were monomorphic in all breeds were also removed, as were those that had an overall minor allele
131 frequency ≤ 0.01 . SNP markers that deviated from Hardy-Weinberg equilibrium were left in the
132 data set. Missing values were not considered in the analysis. After editing, 44,325 SNP were
133 retained for the study. No pruning was performed based on LD, but a minimum ROH length was set
134 (see criteria used for ROH detection) to exclude short, common ROH deriving from LD following

135 Ferencakovic *et al.* (2013b). Thus all markers that passed the quality control were used to calculate
136 ROH.

137 **ROH detection and distribution**

138 ROH were detected using a SAS 9.2 script (SAS Institute, Cary NC) designed to find stretches with
139 a specified number of contiguous homozygous SNP. The following constraints were applied to limit
140 the number of spurious ROH detected: *i*) the minimum number of SNP included in a ROH was
141 fixed at 15; *ii*) the minimum length of a ROH was set at 1 Mb; *iii*) the maximum distance between
142 adjacent SNP was 1 Mb; *iv*) neither heterozygous nor missing genotypes were allowed in a ROH; *v*)
143 sliding windows were not used to detect ROH in order to avoid the introduction of artificial ROH
144 that were shorter than the window (Ferencakovic *et al.* 2013b). The expected length of DNA
145 segments IBD was derived by Fisher (1954) and follows an exponential distribution with mean
146 equal to $1/2g$ (Morgan), where g is the number of generations since the common ancestor. There is
147 an approximate correlation between ROH length and the number of generations that separates an
148 individual from the common ancestor that contributed the IBD fragment: fragments of 1 Mb and 16
149 Mb are ~ 50 and ~ 6 generations from a common ancestor, respectively (see Howrigan *et al.* 2011 for
150 more details). In the present study the minimum length of ROH was set at 1 Mb on the basis of the
151 theoretical relationship between distribution of IBD fragments and the number of generations since
152 a common ancestor. ROH were placed into five classes using the nomenclature of Kirin *et al.*
153 (2010) and Ferencakovic *et al.* (2013a): ROH < 2 Mb, 2-4Mb, 4-8Mb, 8-16Mb and 16<Mb
154 identified as ROH_{1-2Mb}; ROH_{2-4Mb}, ROH_{4-8Mb}, ROH_{8-16Mb}, ROH_{>16Mb} respectively. For each
155 individual in each of the five breeds, and for each ROH length category, the total number of ROH
156 detected (N_{ROH}), the average length of ROH (L_{ROH} , in Mb) and the sum of all ROH segments by

157 animals (S_{ROH} , in Mb) were calculated. The distribution of S_{ROH} within breed were assessed using
158 box plots.

159 **Pedigree and genomic inbreeding analyses**

160 Three types of inbreeding coefficients were calculated: the F_{PED} , F_{GRM} and F_{ROH} . F_{PED}
161 estimates for HOL, BRW and SIM were provided by Italian national breed associations and were
162 calculated using >10 generation of pedigree for HOL and 7 and 4 for BRW and SIM respectively.
163 Pedigrees of PMT and MAR were not available. F_{GRM} estimates were calculated following Van
164 Raden *et al.* (2008). The GRM matrix were calculated as $\mathbf{ZZ}' / \sum_{j=1}^{n_{SNP}} 2p_j(1 - p_j)$ where p_j was the
165 allelic frequency at j -th locus, \mathbf{Z} containing the values of $(1-2p)$ for heterozygous genotypes, $(0-2p)$
166 and $(2-2p)$ for opposite homozygous genotypes, respectively (where p was set at 0.5 in the base
167 population). The diagonal of this matrix was used to assess the genomic inbreeding. F_{ROH}
168 inbreeding coefficients were calculated following McQuillan *et al.* (2008), who defined the
169 proportion of the autosomal genome in ROH as $F_{ROH}^i = S_{ROH}^i / L_{GEN}$, where S_{ROH}^i is the sum across
170 the genome of ROH >1 Mb (including centromeric regions) for the i -the bulls and where L_{GEN} is the
171 autosomal genome length covered by SNP, in the present case this corresponded to ~2,556 Mb.
172 Four different F_{ROH} coefficients were calculated for ROH of different minimum length for each
173 animal ($F_{ROH>1Mb}$, $F_{ROH>4Mb}$, $F_{ROH>8Mb}$ and $F_{ROH>16Mb}$). Within breed one-way ANOVA was used
174 evaluate differences between the coefficients (F_{PED} , F_{GRM} and F_{ROH} of different minimum length).
175 Correlation analysis and a linear regression between F_{PED} (when available) and genomic inbreeding
176 coefficients (F_{GRM} or F_{ROH} with different minimum length) was performed in order to test the
177 similarity among the different estimates of inbreeding.

178

179 To evaluate the effects of the breed and the chromosome on the $F_{ROH>1Mb}$, the following linear
 180 mixed model was used: $y_{ijk} = BREED_i + CHR_j + (BREED \times CHR)_{ij} + BULL_{k(i)} + e_{ijk}$ where,
 181 y_{ijk} was F_{ROH} measured for the k -th individual ($n=4,095$) of the i -th BREED ($n=5$), located in the j -
 182 th autosome (CHR, $n=29$). BREED and CHR were treated as fixed effect, random effect of
 183 $BULL_{k(i)}$ nested within each breed were also fitted. The random effect of $BULL_{k(i)}$ and residual
 184 random term was assumed to be IID distributed $\sim N(0, \mathbf{I}\sigma_{BULL}^2)$ $e \sim N(0, \mathbf{I}\sigma_e^2)$ respectively; the
 185 TukeyHSD test was used to correct for multiple comparisons in order to set the significance level.

186

187 **Effective population size estimation.**

188 The effective population size (N_e) for the five cattle breeds was also estimated. Exploiting the
 189 known relationship between LD (measured by expected r^2) and the inter-marker genetic distance c
 190 between two loci with N_e , the later was calculated for each breed according to Sved (1971)

191
$$N_e = \frac{[(E(r^2) - 1/n)]^{-1} - 1}{4c}$$
, where n is twice the number of animals.

192 All the pair-wise comparisons of r^2 within an interval of 10,000 kb were calculated using
 193 Haploview software (Barret *et al.*, 2005). r^2 was progressively averaged within bins of 100 kb (from
 194 100 kb to 10,000 kb), separately for each chromosome, then, the means for chromosomal averages
 195 for each bin were calculated and plotted against the inter-marker distance (or number of generations
 196 g). 1,000 Kb was considered to be approximately equivalent at 0.1 Morgan (M) and N_e were
 197 calculated per breed for $1/2c$ generations ago, for each c inter-marker distance. In the present study
 198 N_e of 5 generations ago was used to track more recent inbreeding ($1/2g = 0.1$ M, or ~ 10 Mb,
 199 corresponding the longer ROH), whilst N_e of 50 generations ago (ROH of ~ 1 Mb) was used to
 200 capture ancient relatedness, which is deeper than pedigree records.

201

202 **Results**

203 **ROH detection and distributions**

204 The number of ROH per animals varied both within breeds and length classes. The total
205 N_{ROH} ranged from 54 ± 7.2 (PMT) to 94.6 ± 11.6 (BRW). In general, dairy and dual purpose breeds
206 showed the highest N_{ROH} per animal, whereas the lowest values were found for the beef breeds. The
207 number of ROH in the class ROH_{1-2Mb} were similar across breed, with the lowest and highest N_{ROH}
208 in PMT (43.5) and SIM (66.6), respectively. The relative frequency of N_{ROH} by length class is
209 shown in Figure 1. ROH of 1-2Mb were the most frequent in all breeds, representing 71% and 81%
210 of ROH in SIM and PMT, respectively. BRW and HOL bulls had fewer ROH in this class
211 compared to SIM, PMT and MAR. Whereas ROH >4 Mb were at 2% in SIM and 4% in BRW, and
212 all animals of these breeds had at least one ROH >16 Mb. The average number of SNP falling into a
213 ROH was consistent among breeds and ROH length category, ranging from 22.6 (ROH_{1-2Mb}) to
214 476.3 ($ROH_{>16Mb}$) SNP. The longest ROH measured 2,231 SNP and was found in MAR.

215 The average ROH length was longer in BRW and HOL bulls ($L_{ROH} \sim 3.9$ and 3.6 Mb,
216 respectively) compared with PMT and SIM ($L_{ROH} \sim 1.9$ vs 2.2 Mb respectively). MAR had
217 intermediate values, but with several outliers. S_{ROH} varied among dairy, beef or dual purpose breeds
218 (Figure 2). In dairy breeds, the average S_{ROH} was 371 Mb for BRW and 297 Mb for HOL. The
219 largest proportion of the genome within ROH was found in SIM and MAR which had ~ 210 Mb in
220 ROH, followed by PMT with 106 Mb. MAR had on average 7% of the genome estimated as
221 autozygous, however some individuals had up to 30% of their genome in ROH.

222 **Pedigree and genomic inbreeding analysis**

223 Pedigree based inbreeding coefficients were available for BRW, HOL and SIM bulls but not for
224 MAR and PMT. F_{PED} were calculated using all available pedigree information (the minimum
225 number of generations was 4 for SIM), whereas the F_{GRM} and F_{ROH} were calculated for all the five
226 breeds. The average inbreeding coefficients estimated using the different approaches and their
227 distributions are presented in Table 2 and Figure S1, respectively. The highest F_{PED} was observed
228 for HOL, followed by BRW and SIM bulls. The F_{GRM} estimates were significantly higher (p-
229 values < 0.001) than F_{PED} estimates for BRW, HOL and SIM. The proportion of genome that was
230 found to be autozygous from $F_{ROH>1Mb}$ significantly exceed both F_{PED} and F_{GRM} (p-values < 0.001),
231 however, F_{ROH} decreased as the minimum length of the ROH increased, and did not differ from
232 F_{PED} for HOL for $F_{ROH>8Mb}$, or in SIM and BRW for $F_{ROH>16Mb}$. Both the genomic based inbreeding
233 coefficients (F_{GRM} and F_{ROH}) and F_{PED} differed significantly among breeds, and the highest values
234 for each was found in BRW and the lowest for SIM.

235

236 The Pearson correlations between the F_{GRM} and F_{PED} were quite low for BRW (0.134, p-
237 value < 0.001) and HOL (0.128, p-value < 0.001), and moderate in SIM (0.448, p-values < 0.001). The
238 correlations between either F_{PED} or F_{GRM} and F_{ROH} (with ROH of different minimum length) are
239 shown in Table 3. The highest correlations between F_{GRM} and F_{PED} were found for the two dairy
240 breeds when all the ROH > 1 Mb were considered, whereas for the SIM a stronger correlation was
241 observed when ROH longer than 8 Mb were used to estimate F_{ROH} . The correlations between F_{GRM}
242 and F_{ROH} were generally lower than those between F_{ROH} and F_{PED} for all the ROH classes
243 considered. F_{GRM} were moderately associated with F_{ROH} for beef (up to 0.715 in MAR) or dual
244 purpose breeds, but poorly correlated for dairy breeds. The highest correlations between F_{GRM} and
245 F_{ROH} were observed for $F_{ROH>8Mb}$ across all breeds, with exception of the BRW. Regression of F_{ROH}

246 (for different minimum ROH length) on F_{PED} suggested that F coefficients estimated using ROH >8
247 Mb would best fit with pedigree estimates (Figure 3).

248

249 Least square means of F_{ROH} for the different chromosomes were estimated using a linear
250 mixed model ($R^2=0.50$). The effect of chromosome and breed on F_{ROH} were significant (p-values
251 <0.001). F_{ROH} estimates by chromosome and breed are shown in Figure 4, separately for dairy and
252 dual purpose (Figure 4a) and beef breeds (Figure 4b). A significant interaction between
253 chromosome and breed was also observed (p-values <0.001). The highest chromosome effect for
254 F_{ROH} was found on BTA 5 (from 6% of chromosomal length in ROH for PMT to 25.5% in BRW)
255 and BTA 6 (from 5.8% in PMT up to 24.3% in BRW). These two chromosomes harbour a higher
256 percentage of homozygous segments than other chromosomes, particularly in the BRW, MAR and
257 SIM breeds, although other chromosomes do contain regions with a high percentage of ROH .

258 **Discussion**

259 Inbreeding has traditionally been estimated from pedigree data. However, with the availability of
260 high density single nucleotide polymorphism markers, several authors have explored the used of
261 autozygosity identified through ROH to estimate inbreeding. These studies, initially in human and
262 evolutionary genetics investigated the origins of inbreeding at molecular level, estimated the effects
263 of inbreeding on fitness traits, or attempted to map complex disease traits (Gibson *et al.* 2006;
264 McQuillan *et al.* 2008; Lencz *et al.* 2007; Keller *et al.*, 2012). In animal breeding, ROH have been
265 recently used to estimate F coefficients that are more accurate than those obtained using pedigree
266 information, to better quantify inbreeding depression in different animal populations. In the present
267 paper, an analysis of the distribution of autozygous segments in 5 Italian cattle populations is

268 presented, and F statistics obtained using genomic data compared with those obtained using
269 pedigree based (or GRM based) approaches.

270 **ROH detection and distribution**

271 In the present study, longer ROH were found far less frequently than shorter ones in all the breeds
272 analysed (Figure 1). This was also observed in studies of ROH in human and cattle populations
273 (Kirin *et al.* 2010; Ferencakovic *et al.* 2013a). In the present study, dairy and dual purpose breeds
274 had a higher number of ROH compared with beef breed. The results for the BRW and SIM
275 presented here, were in close agreement with a study of ROH in the Austrian Brown Swiss and
276 Simmental using the 54K SNP panel (Ferencakovic *et al.* 2013a). This suggests similarity in the N_e
277 between the populations of the same breed raised in the two countries.

278

279 In the present work, the number of SNP falling into a ROH was consistent among different breeds,
280 although the number of consecutive homozygous SNP was affected by genotyping errors and
281 missing data. Several approaches have been used to manage missing SNP calls: Ferencakovic *et al.*
282 (2013b) suggested allowing some heterozygous SNP within a ROH, especially when handling high
283 density SNP data, while Howrigan *et al.* (2011) argued that incorporating heterozygous SNP would
284 increase the power of detecting truly autozygous segments, which was shown, at least in an *in silico*
285 comparative study where high and low genotyping error rates were simulated. In the present work
286 only homozygous SNP were allowed in ROH. When 1 or 2 heterozygous genotypes were allowed,
287 the number of longest ROH increased dramatically. However, as SNP density was relatively low
288 and HD SNP data was not available to make comparisons heterozygous genotypes were not allowed
289 in the ROH.

290

291

292 In this study, L_{ROH} showed lower variation across breeds than the S_{ROH} , which suggests that L_{ROH} is
293 not a good descriptor of ROH, most likely because of the skewed distribution. In the literature S_{ROH}
294 has often been used in place of L_{ROH} (Kirin *et al.* 2010; Purfield *et al.* 2012; Silio *et al.* 2013),
295 because S_{ROH} gives a better separation among breeds and the length of autozygous segments can be
296 used to interpret population history (McQuillan *et al.* 2008). The average S_{ROH} in BRW was similar
297 to values reported in other studies of Brown Swiss (Ferencakovic *et al.* 2013a), Holstein and PMT
298 (Purfield *et al.* 2012). In this study, beef breeds had a lower S_{ROH} than dairy breeds. This is most
299 likely because dairy breeds have been under more intensive selection, and may be related to recent
300 increase in consanguineous matings resulting from the small number of high genetic merit sires
301 used for artificial insemination. This would be expected to increase inbreeding levels over time,
302 which is what is observed, e.g. in Italian and other European and US Holstein populations (Biffani
303 2002; Miglior *et al.* 2005; Mc Parland *et al.* 2007).

304 The average S_{ROH} found for the SIM was in agreement with values reported for Austrian Simmental
305 (Ferencakovic *et al.* 2013a). The high genetic flow known to occur between these two populations
306 would partially explain these results. Furthermore, the lower S_{ROH} for SIM compared with the BRW
307 and HOL is consistent with differences in breeding strategies: whereas HOL uses a relatively small
308 number of sires, the Simmental breeding programme uses large number of young unproven bulls
309 (~50%) in the commercial population. In addition, a restricted number of semen doses (no more
310 than 7 000 per bull) are stored and used; this limits the number of offspring per bulls, and indirectly
311 controls the level of inbreeding (ANAPRI, Udine 2010; Gaspa *et al.* 2013). PMT was different
312 from other breeds in showing the lowest average S_{ROH} for all the length classes. Unfortunately, no
313 pedigree information was available for this breed, nor for MAR.

314 In general, ROH are found when an individual inherits the same DNA segments from both parents,
315 i.e. the ROH that are truly IBD, and occurs when the parents themselves received the chromosomal
316 fragments from a relatively recent common ancestor. Another possibility explanation for ROH is a
317 lack of recombination in a specific region (LD Hotspots), which means that ancestral chromosomal
318 regions persist in the population. This latter situation is generally associated with shorter ROH. In
319 case of MAR or PMT the origin of ROH could not be assessed in absence of detailed pedigree
320 information and other indicators should be used to explain S_{ROH} in MAR and PMT, for example LD
321 patterns or N_e estimates.

322 **Pedigree and genomic inbreeding analysis**

323 Three different estimates of F were calculated for each bull: Pedigree (F_{PED}) and the other two
324 based on SNP-by-SNP measurement (F_{GRM}) or on the proportion of genome covered by ROH of
325 different minimum size (F_{ROH}). Inbreeding coefficients estimated from the pedigree were lower than
326 those estimated from GRM (with the exception of BRW) or calculated using ROH >1Mb. The
327 difference in average F_{PED} depends on the number of generations in the pedigree. The difference
328 between F_{ROH} and F_{PED} may be explained by considering that ROH capture both ancient and recent
329 relatedness, represented by shorter and longer IBD fragments respectively. Whereas F_{PED} estimates
330 provide only inbreeding estimates based on the recorded pedigree, which may only extend back a
331 few generations. When longer ROH ($F_{ROH>16Mb}$), which arise from recent inbreeding events, are
332 considered, the results for F_{PED} and F_{ROH} are closer. A similar increase in estimated inbreeding for
333 genome based calculations vs pedigree based coefficients has been reported previously for both
334 humans and cattle populations (McQuillan *et al.* 2008; VanRaden *et al.* 2011; Ferencakovic *et al.*
335 2013a; Ferencakovic *et al.* 2013a and b). The F_{ROH} of the shortest ROH (<4 Mb) should be treated
336 with caution as they may not be related to autozygosity, leading to an overestimation of inbreeding

337 coefficients (Ferencakovic *et al.* 2013b). ROH identified in 9 cattle breeds using 54K SNP data
338 were not confirmed when the 800 K SNP was used, indicating that many ROH shorter than 5Mb
339 detected with the 54K SNP are not reliable (Purfield *et al.* 2012). When shorter ROH are included in
340 the calculation of F_{ROH} , the inbreeding coefficients have been shown to be systematically
341 overestimated using a lower density of SNP (Ferencakovic *et al.* 2013b).

342 In the data presented here, both pedigree and molecular based inbreeding coefficients were
343 characterized (Figure 5). The variation of F, transformed by the \log_{10} of the variance, according to
344 Keller *et al.* (2011), shows that the pattern of ROH varies among the F calculated by the three
345 different methods. F_{GRM} shows the highest variance, followed by F_{ROH} and F_{PED} . F_{PED} shows the
346 lowest variation in breeds for which pedigree data is available. This is because F_{PED} does not
347 capture the variation due to recombination, that other two coefficients do by accounting for
348 mendelian sampling and recombination. The variance of $F_{ROH>16MB}$ overlaps the variance of F_{PED} ,
349 probably because $F_{ROH>16MB}$ track very long and hence recent IBD segments. The second
350 consideration is related to the relationship between F_{ROH} and effective population size (see Figure 5,
351 and Table 4). Keller *et al.* (2011) found that the variance of F decreases as the N_e increases, using
352 both simulated and real data. In present study 5 generations were simulated using a pair-wise r^2
353 calculation (Figure S2). In the genotype data this relationship is confirmed for all breeds, with the
354 exception of MAR. In general, the larger the population size, the lower the variance of F, thus the
355 possibility to detect inbreeding from genomic data is reduced. In the breeds examined the lowest
356 and highest N_e (estimated 5 generations ago) were observed in BRW and PMT, respectively (Table
357 4). Effective population size and the level of inbreeding found in the present work could be
358 explained by the recent history of selection for these two breeds. The BRW has been under strong
359 selection for dairy production over the last 80 years from the original Alpine Brown Swiss triple
360 purpose breed to become the Italian Brown dairy breed. This selection process made extensive use

361 of American Brown sires and artificial insemination with a restricted number of sires that brought
362 about a reduction in the effective population size. The same was true for HOL, with an even more
363 restricted number of sires.

364

365 The comparison of F_{PED} and F_{ROH} revealed a poor correlation when F_{GRM} was compared with F_{ROH}
366 This poor correlation between F_{GRM} and F_{ROH} was lower than correlation between F_{PED} and F_{ROH} ,
367 which was also reported by Keller *et al.* (2011). All the F estimates are intrinsically incorrect as
368 determining the actual level of autozygosity IBD is inaccurate. In the present study a small number
369 of generations were available to estimate F_{PED} which introduces errors, as historic inbreeding is not
370 identified. The differences observed in the correlation among different F estimates may also be due
371 to the different N_e of the 5 populations, effect of sampling, missing pedigree information, and the
372 methods used to identify ROH. It was assumed that animals sampled were representative of the 5
373 breeds, however, this may not have been the case for PMT and MAR as both are local breeds which
374 may have a stratified population structure, which would only be identified from a larger sample
375 size. The relationship between ROH and inbreeding has been studied in human populations with
376 different demographic histories and depth of pedigree information (McQuillan *et al.* 2008; Kirin *et*
377 *al.* 2010), the conclusions drawn from these studies were in agreement with those made here. High
378 correlations between the ROH based estimates and pedigree based inbreeding coefficients were
379 reported for 18 human populations for which reliable pedigree information was available, when
380 ROH greater than 5Mb were considered. A correlation of about 0.75 has been reported between
381 ROH and pedigree based inbreeding coefficients for cattle using medium and high density SNP
382 panels (Purfield *et al.* 2012). This is higher than the correlation found in the present paper (Table 3).
383 The correlation between genomic and pedigrees inbreeding values for both HOL and BRW did not
384 change substantially whether ROH shorter than 4Mb are included or not. This is because for these

385 two breeds the percentage of homozygous segments in the shortest classes (74%,BRW and 78%
386 HOL was lower than for SIM (91%).

387

388 ROH may also be used to identify regions of the genome under selection with possibility to
389 map genes affecting traits of interest for animal breeders. Several study have used ROH to map
390 deleterious variants of complex disease in human (see Ku *et al.* 2011 for a review). The ROH
391 clustering in specific genomic regions provide a signal to search for genes under selection. Peaks in
392 ROH were seen on chromosomes 5 and 6, which may be evidence of signatures of selection,
393 resulting from intense selection for dairy traits. These chromosomes harbour genes that affect milk
394 production traits including **ABCG2**, **FAM31A1**, **OPN** (Cohen *et al.* 2004; Cohen-Zinder *et al.*
395 2005; Schnabel *et al.* 2005) on chromosome 5 and the casein cluster on BTA6 (Blott *et al.* 2003)
396 (Figure S3). Several SNP fall into the same ROH on chromosome 6, for 100% of BRW, ~65% of
397 SIM and ~60% MAR individuals. A ROH in PMT at the beginning of the chromosome 2 was found
398 in ~90% of animals.The *myostatin* (**MSTN**) locus which controls the double muscling phenotype
399 (6,532,697 and 6,539,265 bp) is found in this region and a specific **MSTN** variant (p.Cys313Tyr) is
400 reported to be fixed in the PMT breed (Casas *et al.*, 1999).

401

402

403 **Conclusions**

404 This work analysed relationships between runs of homozygosity at SNP loci across the genome and
405 the genetic diversity of the five breeds, and examined the relationship between the ROH and the
406 level of inbreeding. ROH were identified in five Italian cattle breeds genotyped using 54K SNP.
407 The distribution of ROH found was in agreement with previous studies carried out in populations of

408 the same breeds farmed in other countries. In particular, in this work, dairy breeds were found to
409 have longer and more frequent ROH compared with beef and dual purpose cattle, demonstrating
410 that inbreeding is more recent in the dairy breeds. Relationships between pedigree based
411 inbreeding coefficients derived from ROH showed that they could be used to reveal recent selection
412 history of these breeds. One of the advantages of using ROH base inbreeding over pedigree-based
413 estimates is the ability explore historic inbreeding events which is not possible as the depth of
414 pedigree information is usually limited. ROH therefore, facilitate the estimation of inbreeding
415 where pedigree data is limited. However, the regression between the traditional pedigree based
416 inbreeding coefficient and the ROH using different sets of ROH according to the length (from >1
417 Mb to >16 Mb) indicates that including ROH < 4 Mb in the inbreeding calculation produces an
418 overestimation of inbreeding when 54 K SNP are used. ROH are a direct measure of autozygosity at
419 the DNA level and can overcome approximations and errors resulting from incomplete pedigree
420 data. However, distinguishing ROH that are identical by state from those that are truly identical by
421 descent is approximate using low and medium marker density. In populations with high LD and
422 recent inbreeding (e.g. HOL and BRW) the 54K SNP may provide a good estimate of inbreeding,
423 but in populations with low LD and ancient inbreeding, higher marker density would be needed to
424 identify short ROH that are IBD. Ideally whole genome sequence data would be used. ROH can
425 also be used to investigate signatures of selection left on the genome by artificial selection: the
426 homozygous stretches of the genome can help to dissect the genetic architecture of quantitative
427 traits and to map genes of interest for animal breeders. In summary, ROH are a powerful tool for
428 estimating inbreeding coefficients, evaluating its effect on traits of economic importance and
429 controlling its level in breeding programs.

430

431

432 **Acknowledgments**

433 Research was funded by the Italian Ministry of Agriculture (grants SELMOL and INNOVAGEN)
434 and Fondazione CARIPLLO (grant PROZOO). Authors wish to acknowledge Italian Holstein
435 (ANAFI), Italian Brown (ANARB), Italian Piedmontese (ANABORAPI) and Italian Beef cattle of
436 Central Italy (ANABIC) associations for providing data.

437

438 **Conflict of interest.**

439 The authors declare no conflict of interest in the publication of this information.

440 **References**

- 441 Ajmone-Marsan P., Nicolazzi E., Negrini R., Macciotta N., Fontanesi L., Russo V., Bagnato A., Santus E.,
442 Vicario D., van Kaam J.B.C.H.M., Albera A., Filippini F., Marchitelli C., Mancini G., Nardone A.,
443 Valentini A. (2010). Integrating population genomics and genomic selection. *Interbull Bulletin* **41**,
444 39-42.
- 445 ANAPRI (Udine I. (2010) L'indice di selezione. In:
446 http://www.anapri.eu/index.php?option=com_content&view=article&id=68&Itemid=97.
- 447 Barrett J. C., Fry B., Maller J. & Daly M. J. (2005). Haploview: analysis and visualization of LD and
448 haplotype maps. *Bioinformatics* **21**, 263-265
- 449 Biffani S. S.A.a.C.F. (2002) Inbreeding depression for production, reproduction and functional traits in
450 Italian Holstein cattle. In: *Proc. World Congr. Genet. Applied to Livest. Prod.*, , pp. 183-6,
451 Montpellier, France.
- 452 Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in
453 Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding.
454 *Journal of Dairy Sci* **96**, 4697 - 706.
- 455 Blott S., Kim J.-J., Moisis S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B.,
456 Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F. &
457 Coppieters W. (2003) Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-
458 Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is
459 Associated With a Major Effect on Milk Yield and Composition. *Genetics* **163**, 253-66.
- 460 Broman K.W. & Weber J.L. (1999) Long homozygous chromosomal segments in reference families from the
461 Centre d'Etude du Polymorphisme Humain. *American Journal of Human Genetics* **65**, 1493-500.
- 462 Carothers A.D., Rudan I., Kolcic I., Polasek O., Hayward C., Wright A.F., Campbell H., Teague P., Hastie
463 N.D. & Weber J.L. (2006) Estimating human inbreeding coefficients: Comparison of genealogical
464 and marker heterozygosity approaches. *Annals of Human Genetics* **70**, 666-76.

- 465 Casas E., Keele J.W., Fahrenkrug S.C., Smith T.P., Cundiff L.V. & Stone R.T. (1999) Quantitative analysis
466 of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating
467 an inactive myostatin allele. *Journal of Animal Science* **77**, 1686-92.
- 468 Charlesworth D. & Willis J.H. (2009) FUNDAMENTAL CONCEPTS IN GENETICS The genetics of
469 inbreeding depression. *Nature Reviews Genetics* **10**, 783-96.
- 470 Cohen-Zinder M., Seroussi E., Larkin D.M., Looor J.J., Everts-van der Wind A., Lee J.H., Drackley J.K.,
471 Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I. & Ron M. (2005) Identification of a
472 missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6
473 affecting milk yield and composition in Holstein cattle. *Genome Research* **15**, 936-44.
- 474 Cohen M., Reichenstein M., Everts-van der Wind A., Heon-Lee J., Shani M., Lewin H.A., Weller J.I., Ron
475 M. & Seroussi E. (2004) Cloning and characterization of FAM13A1--a gene near a milk protein
476 QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. *Genomics*
477 **84**, 374-83.
- 478 Ferencakovic M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-
479 wide autozygosity in the Austrian fleckvieh cattle. *Agric Conspec Sci* **76**, 325 - 8.
- 480 Ferencakovic M., Hamzic E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Solkner J. (2013a)
481 Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected
482 cattle populations. *Journal of Animal Breeding and Genetics* **130**, 286-93.
- 483 Ferencakovic M., Solkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information:
484 effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.
- 485 Fisher R.A. (1954) A fuller theory of junctions in inbreeding. *Heredity* **8**, 187-97.
- 486 Gaspa G., Pintus M.A., Nicolazzi E.L., Vicario D., Valentini A., Dimauro C. & Macciotta N.P. (2013) Use
487 of principal component approach to predict direct genomic breeding values for beef traits in Italian
488 Simmental cattle. *Journal of Animal Science* **91**, 29-37.
- 489 Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations.
490 *Human Molecular Genetics* **15**, 789 - 95.
- 491 González-Recio O., López de Maturana E. & Gutiérrez J.P. (2007) Inbreeding Depression on Female
492 Fertility and Calving Ease in Spanish Dairy Cattle. *Journal of Dairy Science* **90**, 5744-52.
- 493 Howrigan D., Simonson M. & Keller M. (2011) Detecting autozygosity through runs of homozygosity: a
494 comparison of three autozygosity detection algorithms. *Bmc Genomics* **12**, 460.
- 495 Keller L.F. & Waller D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**,
496 230-41.
- 497 Keller M.C., Visscher P.M. & Goddard M.E. (2011) Quantification of Inbreeding Due to Distant Ancestors
498 and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics* **189**, 237-U920.
- 499 Keller M.C., Simonson M.A., Ripke S., Neale B.M., Gejman P.V., Howrigan D.P., Lee S.H., Lencz T.,
500 Levinson D.F., Sullivan P.F. & Wide S.P.G. (2012) Runs of Homozygosity Implicate Autozygosity
501 as a Schizophrenia Risk Factor. *Plos Genetics* **8**, 425-35.
- 502 Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of
503 homozygosity record population history and consanguinity. *Plos One* **5**, e13996.
- 504 Ku C.S., Naidoo N., Teo S.M. & Pawitan Y. (2011) Regions of homozygosity and their impact on complex
505 diseases and traits. *Human Genetics* **129**, 1-15.
- 506 Lencz T., Lambert C., DeRosse P., Burdick K., Morgan T., Kane J., Kucherlapati R. & Malhotra A. (2007)
507 Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci*
508 *USA* **104**, 1994,2 - 7.
- 509 Mc Parland S., Kearney J.F., Rath M. & Berry D.P. (2007) Inbreeding effects on milk production, calving
510 performance, fertility, and conformation in Irish Holstein-Friesians. *Journal of Dairy Science* **90**,
511 4411-9.
- 512 McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-
513 Narancic N., Janicijevic B., Polasek O., Tenesa A., MacLeod A.K., Farrington S.M., Rudan P.,
514 Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H. & Wilson J.F.
515 (2008) Runs of Homozygosity in European Populations. *American Journal of Human Genetics* **83**,
516 658-.

517 Meuwissen T. (1997) Maximizing the response of selection with a predetermined rate of inbreeding. *Journal*
518 *of Animal Science* **75**, 934 - 40.

519 Meuwissen T., Hayes B. & Goddard M. (2001) Prediction of total genetic value using genome-wide dense
520 marker maps. *Genetics* **157**, 1819 - 29.

521 Miglior F., Burnside E.B. & Kennedy B.W. (1995) Production Traits of Holstein Cattle - Estimation of
522 Nonadditive Genetic Variance-Components and Inbreeding Depression. *Journal of Dairy Science*
523 **78**, 1174-80.

524 Miglior F., Muir B.L. & Van Doormaal B.J. (2005) Selection Indices in Holstein Cattle of Various
525 Countries. *Journal of Dairy Science* **88**, 1255-63.

526 Pryce J.E., Hayes B.J. & Goddard M.E. (2012) Novel strategies to minimize progeny inbreeding while
527 maximizing genetic gain using genomic information. *Journal of Dairy Science* **95**, 377-88.

528 Purfield D., Berry D., McParland S. & Bradley D. (2012) Runs of homozygosity and population history in
529 cattle. *BMC Genetics* **13**, 70.

530 SAS (2012) Sas User Guide. Inc, Cary NC.

531 Schnabel R.D., Kim J.J., Ashwell M.S., Sonstegard T.S., Van Tassell C.P., Connor E.E. & Taylor J.F. (2005)
532 Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin
533 gene. *Proc Natl Acad Sci U S A* **102**, 6896-901.

534 Silio L., Rodriguez M., Fernandez A., Barragan C., Benitez R., Ovilo C. & Fernandez A. (2013) Measuring
535 inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal*
536 *of Animal Breeding and Genetics* **130**, 349 - 60.

537 Slate J., David P., Dodds K.G., Veenvliet B.A., Glass B.C., Broad T.E. & McEwan J.C. (2004)
538 Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity:
539 theoretical expectations and empirical data. *Heredity* **93**, 255-65.

540 Sved J.A. (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations.
541 *Theoretical Population Biology* **2**, 125-41.

542 Toro M., Barragan C., Ovilo C., Rodriganez J., Rodriguez C. & Silio L. (2002) Estimation of coancestry in
543 Iberian pigs using molecular markers. *Conservation Genetics* **3**, 309-20.

544 VanRaden P.M. (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* **91**,
545 4414-23.

546 VanRaden P.M., Olson K.M., Wiggans G.R., Cole J.B. & Tooker M.E. (2011) Genomic inbreeding and
547 relationships among Holsteins, Jerseys, and Brown Swiss. *Journal of Dairy Science* **94**, 5673-82.

548 Wright S. (1922) Coefficients of Inbreeding and Relationship. *The American Naturalist* **56**, 330-8.

549 Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C.,
550 Martin N.G., Montgomery G.W., Goddard M.E. & Visscher P.M. (2010) Common SNPs explain a
551 large proportion of the heritability for human height. *Nature Genetics* **42**, 565-9.

552

553

554 **Table 1.** Average N_{ROH} per animals detected for each ROH length class in five Italian cattle breeds,
 555 with the *standard deviation* given in brackets.

ROH class ¹	BRW	HOL	MAR	PMT	SIM
ROH _{1-2Mb}	50.0(7.2)	46.5 (6.8)	45.5(7.1)	43.5(6.2)	66.6(9.2)
ROH _{2-4Mb}	20.3(5.1)	17.0(4.2)	14.2(4.6)	7.3(2.8)	19.2(5.2)
ROH _{4-8Mb}	12.7(4.5)	9.7(3.5)	7.2(3.2)	2.3(1.6)	5.9(2.8)
ROH _{8-16Mb}	8.1(3.3)	5.9(2.7)	3.0(2.2)	1.9(1.5)	2.3(1.5)
ROH _{16Mb}	3.8(2.2)	3.0(1.8)	1.6(3.4)	1.4 (0.9)	2.0(1.5)
Tot ROH	94.6(11.6)	81.7(9.7)	71.4(11.1)	54.0(7.2)	94.3(12.2)

557 ¹ Number of observations used to calculate N_{ROH} varied from 706 to 749 (mean=738.8) in Brown (BRW), 1,839-2,093
 558 (mean=2,039) in Holstein (HOL), 237-401 (mean=366) in Marchigiana (MAR), 88-364 (mean=257.8) in
 559 Piedmontese(PMT) and from 176-479 (mean=398.2) in Simmental (SIM).

560

561 **Table 2.** Estimated mean (*min-max*) of pedigree based inbreeding coefficients (F_{PED}), GRM based
562 inbreeding coefficient (F_{GRM}) and ROH based inbreeding coefficients (F_{ROH}) for five Italian Cattle
563 breeds. F_{ROH} greater than a specific length class $F_{\text{ROH}>\text{class}}$ (>1 Mb, >4 Mb, >8 Mb, and >16 Mb)
564 were reported.

Coefficient	BRW	HOL	MAR	PMT	SIM
F_{PED}^1	0.026 ^F (.000-.141)	0.044 ^D (.000-.179)	- -	- -	0.008 ^E (.000-.078)
F_{GRM}	0.196 ^A (.000-.690)	0.101 ^B (.000-.611)	0.043 ^B (.000-.373)	0.024 ^B (.000-.146)	0.074 ^B (.000-.201)
$F_{\text{ROH}>1\text{Mb}}$	0.145 ^B (.039-.256)	0.116 ^A (.038-.277)	0.084 ^A (.014-.364)	0.041 ^A (.023-0.170)	0.083 ^A (.038-.188)
$F_{\text{ROH}>4\text{Mb}}$	0.097 ^C (.002-.203)	0.073 ^C (.006-.233)	0.046 ^B (.001-.341)	0.011 ^C (.000-.124)	0.028 ^C (.000-.142)
$F_{\text{ROH}>8\text{Mb}}$	0.068 ^D (.000-.173)	0.051 ^D (.000-.197)	0.031 ^D (.000-.312)	0.007 ^D (.000-.124)	0.015 ^D (.000-.118)
$F_{\text{ROH}>16\text{Mb}}$	0.034 ^E (.000-.134)	0.026 ^E (.000-.167)	0.017 ^E (.000-.282)	0.004 ^D (.000-.078)	0.007 ^E (.000-.097)
s.e.	0.001	<0.001	0.002	<0.001	0.001

565 * Brown (BRW), Holstein (HOL), Marchigiana (MAR), Piedmontese (PMT) and Simmental (SIM)

566 **Estimated mean with those that differ significantly within each breed indicated by a different letter (p-values <0.001)

567 ¹ F_{PED} were not available for MAR and PMT breed

568
569

570 **Table 3.** Correlation between pedigree based inbreeding coefficients (F_{PED}) and ROH based
571 inbreeding coefficients, $r(F_{PED}, F_{ROH})$, corresponding to the minimum size of the ROH used
572 ($F_{ROH>1Mb}$, $>4Mb$, $>8Mb$, $>16Mb$). Correlation between inbreeding coefficients based on genomic
573 relationship matrix (F_{GRM}) and ROH based inbreeding coefficients, $r(F_{GRM}, F_{ROH})$
574

Correlation	BRW	HOL	MAR	PMT	SIM
$r(F_{PED}, F_{ROH})$					
$F_{ROH>1Mb}$	0.662 ^{***}	0.700 ^{***}	-	-	0.669 ^{***}
$F_{ROH>4Mb}$	0.661 ^{***}	0.696 ^{***}	-	-	0.747 ^{***}
$F_{ROH>8 Mb}$	0.654 ^{***}	0.651 ^{***}	-	-	0.765 ^{***}
$F_{ROH>16 Mb}$	0.588 ^{***}	0.561 ^{***}	-	-	0.712 ^{***}
$r(F_{GRM}, F_{ROH})$					
$F_{ROH>1Mb}$	0.271 ^{***}	0.332 ^{***}	0.661 ^{***}	0.420 ^{***}	0.433 ^{***}
$F_{ROH>4Mb}$	0.249 ^{***}	0.350 ^{***}	0.691 ^{***}	0.447 ^{***}	0.547 ^{***}
$F_{ROH>8 Mb}$	0.199 ^{***}	0.389 ^{***}	0.714 ^{***}	0.452 ^{***}	0.576 ^{***}
$F_{ROH>16 Mb}$	0.079 [*]	0.390 ^{**}	0.712 ^{**}	0.416 ^{**}	0.553 ^{**}

575 Brown (BRW), Holstein (HOL), Marchigiana (MAR), Piedmontese (PMT) and Simmental (SIM).

576 *** p-values <0.001; ** p-values <0.01; * p-values <0.05.

577

578

579 **Table 4.** Average LD at 1 Mb and 10 Mb, with the effective population size estimated from LD
 580 values.

Breed	$r^2 \pm sd$		N_e	
	10 Mb	1Mb	5 generation	50 generation
BRW	0.029±0.040	0.098±0.119	90.7	237.6
HOL	0.026±0.035	0.082±0.102	98.7	284.3
MAR	0.013±0.018	0.041±0.056	315.2	662.8
PMT	0.011±0.015	0.022±0.031	435.4	1539.6
SIM	0.012±0.015	0.034±0.048	363.5	821.1

581

582 **Figure Captions**

583 **Figure 1.** Frequency distribution of the number of ROH in different length classes and for each
584 breed.

585 **Figure 2.** Box Plots of within breed average Sum of Length of ROH (S_{ROH} , in Mb) calculated
586 across all the animals. (HOL=Holstein; BRW=Brown; PMT=Piedmontese; MAR=Marchigiana;
587 SIM=Simmental)

588 **Figure 3.**Plot of the regression of F_{ROH} of different minimum length (>1Mb, >4Mb, >8Mb and
589 >16Mb) on F_{PED} , for two Italian dairy breeds (HOL, BRW) and a dual purpose breed (SIM).

590 **Figure 4.** Least squared mean estimates of F_{ROH} along *Bos Taurus* Autosomes (BTA) in five breeds
591 derived using linear mixed model. a) Dairy and Dual Purpose Breeds (BRW=Brown,
592 HOL=Holstein and SIM=Simmental) b) Beef Breeds (MAR=Marchigiana, PMT= Piemontese).

593 **Figure 5.** Variance of F for different inbreeding estimates (F_{PED} , F_{GRM} and F_{ROH} of different
594 minimum length). Effective population size 5 generation ago is also reported for BRW=Brown,
595 HOL=Holstein, MAR=Marchigiana, PMT=Piedmontese and SIM=Simmental.

596

597 **Supporting information**

598 **Figure S1**

599 Histogram of the distribution of different inbreeding coefficients in the five breeds. For breeds with
600 pedigree (BRW, HOL, SIM) F_{PED} , F_{GRM} , $F_{ROH>1MB}$ $F_{ROH>16MB}$ are reported and for MAR and PMT
601 F_{GRM} , $F_{ROH>1MB}$ $F_{ROH>8MB}$ $F_{ROH>16MB}$ are reported.

602

603 **Figure S2**

604 a) LD estimation (r^2) between pair-wise markers within an interval of 10000 kb in 5 Italian cattle
605 breeds (BRW= Brown, HOL=Holstein, MAR= Marchigiana, PMT=Piedmontese, SIM=
606 Simmental). b) Effective population size estimates (N_e) from 5 to 250 generation ago in the 5
607 Italian cattle breeds.

608

609 **Figure S3**

610 Occurrence of each SNP in a ROH measured by the percentage of animals for which a particular
611 SNP falls into a ROH vs the position (Mb) along the chromosomes BTA2, BTA5, BTA6, BTA18.

612 Position of genes known to affect dairy and beef traits are indicated by a black triangle.

613