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

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

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

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

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

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

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

¶ Parco Tecnologico Padano, Lodi 26900, Italy.

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

Summary

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

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

Introduction

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

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

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

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

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

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

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

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

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

Materials and Methods

Data

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

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

ROH detectionand distribution

ROH were detected using a SAS 9.2 script (SAS Institute, Cary NC) designed to find stretches with a specified number of contiguous homozygous SNP. The following constrains were applied to limit the number of spurious ROH detected: *i*) the minimum number of SNP included in a ROH was fixed at 15; *ii*) the minimum length of a ROH was set at 1 Mb; *iii*) the maximum distance between adjacent SNP was 1 Mb; *iv*) neither heterozygous nor missing genotypes were allowed in a ROH; *v*) sliding windows were not used to detect ROH in order to avoid the introduction of artificial ROH that were shorter than the window (Ferencakovic *et al*. 2013b). The expected length of DNA segments IBD was derived by Fisher (1954) and follows an exponential distribution with mean equal to $1/2g$ (Morgan), where g is the number of generations since the common ancestor. There is an approximate correlation between ROH length and the number of generations that separates an individual from the common ancestor that contributed the IBD fragment: fragments of 1 Mb and 16 Mb are ~50 and ~6 generations from a common ancestor, respectively (see Howrigan *et al.* 2011 for more details). In the present study the minimum length of ROH was set at 1 Mb on the basis of the theoretical relationship between distribution of IBD fragments and the number of generations since a common ancestor. ROH were placed into five classes using the nomenclature of Kirin *et al.* (2010) and Ferencakovic *et al.* (2013a): ROH < 2 Mb, 2-4Mb, 4-8Mb, 8-16Mb and 16<Mb identified as $ROH_{1\text{-}2Mb}$; $ROH_{2\text{-}4Mb}$, $ROH_{4\text{-}8Mb}$, $ROH_{8\text{-}16Mb}$, $ROH_{>16Mb}$ respectively. For each individual in each of the five breeds, and for each ROH length category, the total number of ROH 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 $\mathbf{ZZ}' / \sum_{j=1}^{nSNP} 2p_j(1 - p_j)$ where p_j was the
165 allelic frequency at *j*-th locus, **Z** containing the values of (1-2*p*) for heterozygous genotypes, (0-2*p*)
166 and (2-2*p*) 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 $F^i_{ROH} = S^i_{ROH} / L_{GEN}$,where S^i_{ROH} 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 $F_{ROH>16\ Mb}$). 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

To evaluate the effects of the breed and the chromosome on the $F_{ROH>1Mb}$, the following linear mixed model was used: $y_{ijk} = \text{BREED}_i + \text{CHR}_j + (\text{BREED} \times \text{CHR})_{ij} + \text{BULL}_{k(i)} + \text{e}_{ijk}$ where, 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*-th autosome (CHR, n=29). BREED and CHR were treated as fixed effect, random effect of $\text{BULL}_{k(i)}$ nested within each breed were also fitted. The random effect of $\text{BULL}_{k(i)}$ and residual random term was assumed to be IID distributed $\sim N(0, \mathbf{I}\sigma^2_{\text{BULL}})$ $e \sim N(0, \mathbf{I}\sigma^2_e)$ respectively; the TukeyHSD test was used to correct for multiple comparisons in order to set the significance level.

Effective population size estimation.

The effective population size (N_e) for the five cattle breeds was also estimated. Exploiting the known relationship between LD (measured by expected r^2) and the inter-marker genetic distance *c* between two loci with N_e, the later was calculated for each breed according to Sved (1971) $N_e = \frac{[(E(r^2) - 1/n)]^{-1} - 1}{4c}$, where *n* is twice the number of animals.

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

Results

ROH detection and distributions

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

The average ROH length was longer in BRW and HOL bulls (L_{ROH}~3.9 and 3.6 Mb, respectively) compared with PMT and SIM (L_{ROH}~1.9 vs2.2 Mb respectively). MAR had intermediate values, but with several outliers. S_{ROH} varied among dairy, beef or dual purpose breeds (Figure 2). In dairy breeds, the average S_{ROH} was 371 Mb for BRW and 297 Mb for HOL. The largest proportion of the genome within ROH was found in SIM and MAR which had ~210 Mb in ROH, followed by PMT with 106 Mb. MAR had on average 7% of the genome estimated as 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

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

ROH detection and distribution

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

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

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

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

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

Pedigree and genomic inbreeding analysis

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

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

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

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

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

Conclusions

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

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

Acknowledgments

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

Conflict of interest.

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

References

Ajmone-Marsan P., Nicolazzi E., Negrini R., Macciotta N., Fontanesi L., Russo V., Bagnato A., Santus E., Vicario D., van Kaam J.B.C.H.M., Albera A., Filippini F., Marchitelli C., Mancini G., Nardone A., Valentini A. (2010). Integrating population genomics and genomic selection. *Interbull Bulletin* **41,** 39-42.

ANAPRI (Udine I. (2010) L'indice di selezione. In: *http://www.anapri.eu/index.php?option=com_content&view=article&id=68&Itemid=97*.

Barrett J. C., Fry B., Maller J. & Daly M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265

Biffani S. S.A.a.C.F. (2002) Inbreeding depression for production, reproduction and functional traits in Italian Holstein cattle. In: *Proc. World Congr. Genet. Applied to Livest. Prod.*, , pp. 183–6, Montpellier, France.

Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journalof Dairy Sci* **96**, 4697 - 706.

Blott S., Kim J.-J., Moisio S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Vilkki J., Georges M., Farnir F. & Coppieters W. (2003) Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. *Genetics* **163**, 253-66.

Broman K.W. & Weber J.L. (1999) Long homozygous chromosomal segments in reference families from the Centre d'Etude du Polymorphisme Humain. *American Journal of Human Genetics* **65**, 1493-500.

Carothers A.D., Rudan I., Kolcic I., Polasek O., Hayward C., Wright A.F., Campbell H., Teague P., Hastie N.D. & Weber J.L. (2006) Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. *Annals of Human Genetics* **70**, 666-76.

Casas E., Keele J.W., Fahrenkrug S.C., Smith T.P., Cundiff L.V. & Stone R.T. (1999) Quantitative analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds segregating an inactive myostatin allele. *Journal of Animal Science* **77**, 1686-92.

Charlesworth D. & Willis J.H. (2009) FUNDAMENTAL CONCEPTS IN GENETICS The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783-96.

Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der Wind A., Lee J.H., Drackley J.K., Band M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I. & Ron M. (2005) Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Research* **15**, 936-44.

Cohen M., Reichenstein M., Everts-van der Wind A., Heon-Lee J., Shani M., Lewin H.A., Weller J.I., Ron M. & Seroussi E. (2004) Cloning and characterization of FAM13A1--a gene near a milk protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holsteins. *Genomics* **84**, 374-83.

Ferencakovic M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agric Conspec Sci* **76**, 325 - 8.

Ferencakovic M., Hamzic E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Solkner J. (2013a) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding and Genetics* **130**, 286-93.

Ferencakovic M., Solkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.

Fisher R.A. (1954) A fuller theory of junctions in inbreeding. *Heredity* **8**, 187-97.

Gaspa G., Pintus M.A., Nicolazzi E.L., Vicario D., Valentini A., Dimauro C. & Macciotta N.P. (2013) Use of principal component approach to predict direct genomic breeding values for beef traits in Italian Simmental cattle. *Journalof Animal Science* **91**, 29-37.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789 - 95.

González-Recio O., López de Maturana E. & Gutiérrez J.P. (2007) Inbreeding Depression on Female Fertility and Calving Ease in Spanish Dairy Cattle. *Journal of Dairy Science* **90**, 5744-52.

Howrigan D., Simonson M. & Keller M. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *Bmc Genomics* **12**, 460.

Keller L.F. & Waller D.M. (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**, 230-41.

Keller M.C., Visscher P.M. & Goddard M.E. (2011) Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics* **189**, 237-U920.

Keller M.C., Simonson M.A., Ripke S., Neale B.M., Gejman P.V., Howrigan D.P., Lee S.H., Lencz T., Levinson D.F., Sullivan P.F. & Wide S.P.G. (2012) Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. *Plos Genetics* **8**, 425-35.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of homozygosity record population history and consanguinity. *Plos One* **5**, e13996.

Ku C.S., Naidoo N., Teo S.M. & Pawitan Y. (2011) Regions of homozygosity and their impact on complex diseases and traits. *Human Genetics* **129**, 1-15.

Lencz T., Lambert C., DeRosse P., Burdick K., Morgan T., Kane J., Kucherlapati R. & Malhotra A. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci USA* **104**, 1994,2 - 7.

Mc Parland S., Kearney J.F., Rath M. & Berry D.P. (2007) Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *Journal of Dairy Science* **90**, 4411-9.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., MacLeod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H. & Wilson J.F. (2008) Runs of Homozygosity in European Populations. *American Journal of Human Genetics* **83**, 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
$ROH_{1\text{-}2Mb}$	50.0*(7.2)*	46.5 *(6.8)*	45.5*(7.1)*	43.5*(6.2)*	66.6*(9.2)*
$ROH_{2\text{-}4Mb}$	20.3*(5.1)*	17.0*(4.2)*	14.2*(4.6)*	7.3*(2.8)*	19.2*(5.2)*
$ROH_{4\text{-}8Mb}$	12.7*(4.5)*	9.7*(3.5)*	7.2*(3.*2)	2.3*(1.6)*	5.9*(2.8)*
$ROH_{8\text{-}16Mb}$	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)*

556

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 $F_{ROH->class}$ (>1 Mb, >4 Mb**,** >8 Mb, and >16 Mb)
564 were reported.

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

	$r^2 \pm sd$		N_e	
Breed	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

Figure Captions

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

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

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

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

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

Supporting information

Figure S1

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

Figure S2

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

Figure S3

Occurrence of each SNP in a ROH measured by the percentage of animals for which a particular SNP falls into a ROH *vs* the position (Mb) along the chromosomes BTA2, BTA5, BTA6, BTA18. Position of genes known to affect dairy and beef traits are indicated by a black triangle.