Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds

S. Mastrangelo [1], M. Tolone [1], R. Di Gerlando [1], L. Fontanesi [2], M.T. Sardina [1] and B. Portolano [1]

[1]*Dipartimento Scienze Agrarie e Forestali, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy*

[2]*Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Fanin 46, Bologna, 40127, Italy*

Corresponding author: Salvatore Mastrangelo. Email: salvatore.mastrangelo@unipa.it

Short title: Runs of homozygosity in three local cattle breeds

Abstract

In the local breeds with small population size, one of the most important problems is the increase of inbreeding coefficient (*F*). High levels of inbreeding lead to reduced genetic diversity and inbreeding depression. The availability of high-density single nucleotide polymorphism (SNP) arrays has facilitated the quantification of *F* by genomic markers in farm animals. Runs of homozygosity (ROH) are contiguous lengths of homozygous genotypes and represent an estimate of the degree of autozygosity at genome-wide level. The current study aims to quantify the genomic *F* derived from ROH (F_{ROH}) in three local dairy cattle breeds. F_{ROH} values were

compared with *F* estimated from the genomic relationship matrix (F_{GRM}), based on the difference between observed versus expected number of homozygous genotypes (F_{HOM}) and the genomic homozygosity of individual *i* ($F_{MOL\ i}$). The molecular coancestry coefficient ($f_{MOL\ ij}$) between individuals *i* and *j* was also estimated. Individuals of Cinisara (71), Modicana (72) and Reggiana (168) were genotyped with the 50K v2 Illumina BeadChip. Genotypes from 96 animals of Italian Holstein cattle breed were also included in the analysis. We used a definition of ROH as tracts of homozygous genotypes that were >4 Mb. Among breeds, 3661 ROH were identified. Modicana showed the highest mean number of ROH per individual and the highest value of F_{ROH}, whereas Reggiana showed the lowest ones. Differences among breeds existed for the ROH lengths. The individuals of Italian Holstein showed high number of short ROH segments, related to ancient consanguinity. Similar results showed the Reggiana with some extreme animals with segments covering 400 Mb and more of genome. Modicana and Cinisara showed similar results between them with the total length of ROH characterized by the presence of large segments. High correlation was found between F_{HOM} and F_{ROH} ranged from 0.83 in Reggiana to 0.95 in Cinisara and Modicana. The correlations among F_{ROH} and other estimated *F* coefficients were generally lower ranged from 0.45 ($F_{MOL\ i,}$ - F_{ROH}) in Cinisara to 0.17 (F_{GRM} - F_{ROH}) in Modicana. On the basis of our results, recent inbreeding was observed in local breeds, considering that 16 Mb segments are expected to present inbreeding up to 3 generations ago. Our results showed the necessity of implementing conservation programs to control the rise of inbreeding and coancestry in the three Italian local dairy cattle breeds.

Keywords: Genomic inbreeding; local cattle breeds; Runs of Homozygosity

Implications

In the local breeds with small population size, one of the most important problems is the increase of inbreeding that leads to different negative effects as a reduction in phenotypic values. The current study aims to quantify the genomic inbreeding derived from ROH (F_{ROH}) in three Italian local dairy cattle breeds. According to ROH results, recent inbreeding was well detected in the investigated local dairy cattle breeds. Our results showed the necessity of implementing conservation programs to preserve the local breeds in order to avoid further loss of genetic distinctiveness. Therefore, determining the occurrence of IBD segments in potential parents, thereby measuring their relatedness and coancestry, can be used to minimize the occurrence of long ROH in the offspring.

Introduction

Animal genetic resources must be preserved because of their contribution to human livelihood, now and in the future (Toro *et al.,* 2011). Most local livestock breeds are the result of a particular adaptation to production systems environmentally conditioned, and in many cases no other breed could survive in the same habitat if the local breed goes extinct. In addition, such local populations might harbor specific genetic variants that are worth retaining and that might be used to recover the loss of genetic diversity that occurs in mainstream breeds because of very intensive selection on production traits (Fernández *et al.,* 2011). Apart from that, these populations represent local culture, history, and tradition and are often linked to traditional products of farm animals (milk, meat, eggs etc).

.

Typically, local breeds are small populations and their size put them at risk of extinction. Consequently, the genetic diversity stored in each of them should be treated with great care and management strategies that insure the viability, and maintenance of the population should be implemented. Selection programs in local breeds with small population size are limited by the low number of animals (families) and by the need to control inbreeding (Fontanesi *et al.,* 2015) which represent one of the most important problems. The individual inbreeding coefficient (*F*) is defined as the proportion of an individual's genome that is autozygous, that has homozygous identical by descent (IBD) status, or equivalently the probability of a randomly sampled locus in the genome to be autozygous (Ferenčaković *et al.,* 2013a). The increase of *F* leads to different negative effects as reduction in phenotypic values for some traits, reduction of genetic variance, higher frequency of homozygous genotypes with the reduction of individual performance (inbreeding depression), and lower population viability (Ouborg *et al.,* 2010). Therefore, to avoid inbreeding depression, an accurate and sensitive estimation of *F* is very important, especially in local breeds/populations. Traditionally, *F* is estimated on the basis of pedigree information but in most cases this is unavailable or inaccurate. Moreover, the probabilistic approach of pedigree analysis does not take into account the stochastic nature of recombination (McQuillan *et al.,* 2008). Recently, with the availability of high-density single nucleotide polymorphism (SNP) arrays, *F* can be estimated accurately in absence of pedigree information (Allendorf *et al.,* 2010). There are two categories of genomic inbreeding measures based on genome-wide SNPs. The first category is based on marker-by-marker estimates such as the diagonal elements of the genomic relationship matrix (GRM) (VanRaden *et al.,* 2011), the canonical estimate based on excess SNP homozygosity in PLINK (Purcell *et al.,* 2007) and

101 molecular coancestry estimates (Caballero and Toro, 2002). The second one is
102 based on Runs of Homozygosity (ROH) detection. ROH are contiguous lengths of
103 homozygous genotypes that are present in an individual due to parents transmitting
104 identical haplotypes to their offspring (Gibson *et al.,* 2006). Nowadays, *F* estimated
105 from ROH (F_{ROH}) is considered to be the most powerful method of detecting
106 inbreeding effects among several alternative estimates of inbreeding (Keller *et al*.,
107 2011). F_{ROH} provided a good measure of individual genome-wide autozygosity and
108 allows to distinguish between recent and ancient inbreeding (McQuillan *et al.,* 2008).
109 Because recombination events interrupt long chromosome segments, long ROH (~
110 10 Mb) arise as result of recent inbreeding (up to five generation ago), while shorter
111 ROH (~ 1 Mb) can indicate more distant ancestral effect (up to 50 generation ago)
112 such as breed founder effects (Howrigan *et al*., 2011). Therefore, estimate of *F* using
113 ROH is particularly appealing as the number of generations of inbreeding and the
114 history of recent selection events can be inferred from the extend and frequency of
115 ROH regions (Purfield *et al*., 2012). Although ROH from high-throughput genotyping
116 analyses have been studied extensively in humans, these estimates are rare in
117 cattle, particular in local breeds, and in other livestock species (Purfield *et al*., 2012;
118 Ferenčaković *et al*., 2013a; Silió *et al.,* 2013; Pertoldi *et al*., 2014).
119 The current study aims to quantify the genomic inbreeding derived from ROH in three
120 economically important Italian local dairy cattle breeds, Cinisara, Modicana, and
121 Reggiana, characterized by the same breeding goals but different selection histories.
122 Moreover, genotypes from Italian Holstein, the most important dairy cattle breed
123 reared in Italy, were also included in these analyses in order to compare results
124 among breeds.
125

Material and methods

Breeds, genotypes and quality control

A total of 407 individuals were used for the analyses. DNA samples belonged to four different cattle breeds: Cinisara (71), Modicana (72), Reggiana (168), and Italian Holstein (96). For these breeds pedigree data were not available. Sampling was carried out in several farms and individuals were selected on the basis of information supplied by the farmers to avoid, as much as possible, closely related animals. The Cinisara, Modicana, and Reggiana are three economically important local breeds with small population size (number of reared animals <4000). Cinisara and Modicana are two cattle breeds well adapted to the harshness of Sicilian marginal mountain areas and their economic importance lies on the traditional production systems of two typical 'pasta filata' cheeses: Caciocavallo Palermitano and Ragusano P.D.O. (Protected Designation of Origin), respectively. Recently, Mastrangelo *et al*. (2014) reported the effective population size values estimated from rate of *F* per year (19 and 12) and from rate of coancestry (*f*) (4 and 8 individuals) in Cinisara and Modicana cattle breeds, respectively. Reggiana is a local cattle breed reared in the province of Reggio Emilia in Northern Italy specialized for the production of a niche brand of Parmigiano-Reggiano P.D.O. cheese.

All animals were genotyped for 54 609 SNPs using Bovine SNP50K v2 BeadChip (Illumina Inc., San Diego, CA). Data quality control was performed separately for each breed. We excluded all SNPs not assigned to a chromosome (BTA) or assigned to chromosomes X and Y. Markers were filtered according to quality criteria that included: i) call frequency (≥0.95), ii) minor allele frequency (MAF≥0.01), and iii) Hardy-Weinberg equilibrium (HWE; *P*-value=0.001). SNPs that did not satisfy these quality criteria were excluded. Moreover, considering that high linkage disequilibrium

(LD) can lead to detection of ROH that are not truly identical by descent, LD pruning was also performed before the ROH call to increase power, as suggested by Purcell *et al*. (2007) and applied by several authors (Howrigan *et al*., 2011; Bjelland *et al*., 2013). Therefore, unlinked SNPs were selected using *-indep* option of PLINK with the following parameters: 50 SNPs per window, a shift of 5 SNPs between windows, and r^2 threshold of 0.5. A total of 38 937 SNPs in Cinisara, 32 179 SNPs in Modicana, 29 483 SNPs in Reggiana, and 27 586 SNPs in Italian Holstein cattle breeds were retained after quality control and were used to estimate F_{ROH}. The main difference for the number of SNPs used for each breed, in particular the highest number of SNPs used for Cinisara, was due to different values of LD among breeds. In fact, Cinisara showed the lowest value of LD and, therefore, the lowest number of excluded SNPs.

Run of homozygosity calling option

F_{ROH} were calculated as the proportion of genome in runs of homozygosity over the overall length of the genome covered by the involved SNPs (2 541 174 kb) using the PLINK whole-genome association analysis toolset (Purcell *et al.,* 2007). The following criteria were used to define the ROH: i) the minimum number of SNPs included in the ROH was fixed to 40; ii) the minimum length that constituted the ROH was set to 4 Mb; iii) two missing SNPs were allowed in the ROH; iv) minimum density of 1 SNP every 100 kb; v) maximum gap between consecutive SNPs of 1 Mb. Moreover, the number of allowed heterozygous SNPs was set to different values: from one to three. Mean F_{ROH} values obtained allowing different numbers of heterozygous SNPs were compared within the same breed using paired *t*-tests. The mean number of ROH per individual per breed (MN_{ROH}), the average length of ROH (L_{ROH}), and the sum of all ROH segments per animal (S_{ROH}) were estimated. The

distribution of S_{ROH} within breed was assessed using box plots. Additionally, chromosomal (BTA) F_{ROH} (F_{ROHBTA}) values were also estimated for each breed, as $F_{ROHBTA} = L_{ROHBTA} / L_{BTA}$ (Silió *et al.,* 2013), in which L_{ROHBTA} is the total length of an individual's ROH in each BTA and L_{BTA} is the length of each chromosome covered by the involved SNPs (Supplementary Table S1). ROH were classified into three classes (4-8 Mb, 8-16 Mb, and >16 Mb) using the same nomenclature reported by other authors (Ferenčaković *et al.,* 2013a; Marras *et al*., 2014) except for two classes (<2 and 2-4 Mb) which were not considered in our study. The number and percentage of ROH within each ROH length category for breed were also determined.

Genomic inbreeding analyses

Alternative estimates of inbreeding and coancestry coefficients were also calculated. In particular: (1) the values of the diagonal elements of the genomic relationship matrix (GRM) (F_{GRM}) proposed by VanRaden *et al.* (2011); (2) the genomic inbreeding coefficient based on the difference between observed versus expected number of homozygous genotypes (F_{HOM}) using PLINK (Purcell *et al.,* 2007); (3) the molecular coancestry coefficient ($f_{MOL\ ij}$) between individuals *i* and *j* (Caballero and Toro, 2002); (4) the molecular inbreeding coefficient ($F_{MOL\ i}$) of individual *i*, calculated as $F_{MOL\ i} = 2\ f_{MOL\ ii} - 1$ ($f_{MOL\ ii}$ is the molecular self-coancestry). Spearman's rank correlation among different genomic inbreeding measures was calculated.

Effective population size

The effective population sizes (N_e) were calculated as $N_e = (1/4c)*(1/r^2\text{-}1)$ (Sved, 1971) where r^2 (the squared correlation coefficient of allele frequencies at pair of loci)

is the value of LD and c is the genetic distance in Morgans between SNPs. Physical distances between SNP pairs were converted to genetic distances with the assumption of 1 cM ~ 1 Mb. Each genetic distance *c* corresponds to a value of *t* generation in the past, and this value was calculated as $t=1/(2c)$,assuming a linear population growth (Hayes *et al*., 2003). All pairwise combinations of SNPs were estimated using LD plot function in Haploview v 4.2 software (Barrett *et al*., 2005). For this analysis, markers were filtered according to quality criteria reported above, except for LD pruning; in fact N_e estimates could be biased if calculated from LD pruned SNPs. A total of 44 875 SNPs in Cinisara, 42 687 SNPs in Modicana, 35 720 SNPs in Reggiana, and 41 596 SNPs in Italian Holstein cattle breeds were used. For each chromosome, pairwise r^2 was calculated for SNPs between 0 and 50 Mb apart. To visualize the LD pattern per chromosome, r^2 values were stacked and plotted as a function of inter-marker distance categories.

Results and Discussion

The main aim of this study was to analyze estimates of inbreeding derived from ROH in three important Italian local cattle breeds. Moreover, genotypes from Italian Holstein were also included in these analyses in order to compare results among breeds.

We used a definition of ROH as tracts of homozygous genotypes that were >4 Mb in length identified with a minimum number of 40 SNPs. In fact, the density of SNP panel used to generate the data for ROH identification is an important factor that strongly affects autozygosity estimates. Ferenčaković *et al.* (2013b) showed that the 50K panel revealed an abundance of small segments and overestimated the numbers of segments 1-4 Mb long, suggesting that it is not sensitive enough for the

226 precise determination of small segments. We estimated mean $F_{ROH>4Mb}$ values
227 allowing one, two and three heterozygous SNPs and paired *t*-tests were conducted
228 within each cattle breed. In fact, considering that genotyping errors in SNP chip data
229 do occur, it seems reasonable to allow some heterozygous calls, especially for long
230 segments that are more frequent in cattle populations (Ferenčaković *et al.,* 2013b)
231 than in human species (Kirin *et al.,* 2010). The results showed different values
232 depending on whether one, two and three heterozygous genotypes were allowed
233 (Table 1). The differences between F_{ROH} estimated using one and two heterozygous
234 SNPs were very small in all breeds and did not have important effects on estimates
235 of inbreeding levels, with the highest value of 0.003 units in Italian Holstein and
236 Modicana (Table 1). The highest different values of F_{ROH} were observed when one
237 and three heterozygous SNPs were compared, with the highest value of 0.007 units
238 for the same above mentioned breeds. Ferenčaković *et al.* (2013b) suggested that
239 for long ROH (which can have more than 5000-6000 SNPs) some heterozygous calls
240 must be allowed, especially with high-density chip, but at the same time, the number
241 of allowable heterozygous calls should be limited. In fact, the same authors showed
242 that allowing certain minimum numbers of heterozygous SNPs leads to inaccurate
243 ROH calls, in particular at the ends of ROH. Marras *et al*. (2014), in a study of ROH
244 using medium-density chip, reported that when heterozygous SNPs were allowed,
245 the number of longer ROH increased dramatically, and preferred not use them in the
246 ROH. Therefore, considering that in our study medium-density SNP data were used,
247 and that the longest segment was below 2000 SNPs, only one heterozygous SNP
248 was allowed in the ROH in order to avoid underestimation of long ROH.
249 We analyzed animals from four Italian cattle breeds with different inbreeding
250 background and selection histories. In particular, Cinisara and Modicana are two

ancient Sicilian breeds that are not subject to breeding programs (Mastrangelo *et al*., 2014), whereas Reggiana is characterized by limited selection program. For this breed, only few studies have been carried out so far with the aim to identify associations with production traits that might be useful to refine selection and conservation programs (Fontanesi *et al.,* 2015). Holstein dairy cattle has dominated the milk production industry over decades. Intense and accurate artificial selection practiced over many years has resulted in high rates of genetic gain; however the high rates of gain have been accompanied by large increase of inbreeding (Rodriguez-Ramilo *et al.,* 2015).

A total of 3661 ROH were identified among the four breeds. All individuals of Italian Holstein displayed at least two ROH, whereas in the local breeds there were individuals that did not showed ROH >4 Mb. In all breeds, except for Reggiana, the number of ROH per chromosome was greater in BTA1 and BTA2, and tended to decrease with chromosome length. The maximum size of ROH was 112.65 Mb and was found on BTA 8 in Cinisara breed. Kim *et al*. (2013) showed similar results in Holstein cow with the maximum size of ROH of 87.13 Mb on BTA 8. Modicana and Italian Holstein breeds showed the maximum size of ROH on BTA 9 (89.61 Mb and 70.11 Mb, respectively), whereas the Reggiana breed on BTA 4 (102.18 Mb). Modicana breed showed the highest MN_{ROH} per individual and the highest value of $F_{ROH>4Mb}$ (11.03 and 0.055, respectively), whereas Reggiana breed showed the lowest ones (7.15 and 0.035, respectively) (Table 2). L_{ROH} values indicated low variation among the four breeds showing that this value is not a good descriptor of ROH as reported by other authors (Marras *et al*., 2014). The comparison of ROH is not straightforward since different studies used different criteria in particular for the minimum length of ROH and the minimum number of SNPs involved in ROH.

Furthermore, the number of SNPs, density of the SNP chip, and selection criteria for SNPs used to determine the genomic inbreeding can have a huge effect on these values (Bjelland *et al*., 2013). Ferenčaković *et al*. (2013a) found higher number of ROH in four analyzed cattle breeds probably because of the shorter length considered to define the ROH (>1 Mb). Similar results of F_{ROH}>4 Mb were reported by Ferenčaković *et al.* (2013b) using a 50K panel for Pinzgauer (0.037) and Tyrol Grey (0.042) local cattle breeds, and by Marras *et al*. (2014) in Marchigiana (0.046) beef cattle breed. Differences among breeds existed also for the ROH length. Figure 1 showed the total number of ROH and the total lengths of genome in ROH for each individual of the four breeds. Considerable differences among animals and breeds have been found. The individuals of Italian Holstein breed showed high number of short ROH segments. Similar results were showed for Reggiana breed with some extreme animals with segments covering 400 Mb and more of genome, and with a number of ROH per individual greater than 25. The Sicilian breeds showed analogous results between them with the total length of ROH characterized by the presence of large segments. S_{ROH} varied among breeds (Figure 2). The highest average S_{ROH} was 132 Mb in Cinisara, whereas the lowest one was 90 Mb in Reggiana. Considering the median values, the highest one was found in Italian Holstein, whereas the lowest one was found in Reggiana. The average reported S_{ROH} values were lower than the ones reported in other studies (Purfield *et al*., 2012; Ferenčaković *et al.,* 2013a). The three most homozygous animals present in our dataset were from Cinisara (676.9 Mb), Modicana (681.2 Mb), and Reggiana (725.2 Mb) with almost a quarter of their genome classified as ROH. In all breeds, most ROH segment coverage was in the shorter length categories (4-8 Mb), in particular Modicana (51%) and Italian Holstein (50%) (Table 3). In fact, as reported in studies

of ROH in human (Kirin *et al.,* 2010) and cattle populations (Ferenčaković *et al.,* 2013a; Marras *et al*., 2014) longer ROH were found less frequently than shorter ones. The expected length of autozygous segments that are identical by descent follows an exponential distribution with mean equal to ½ *g* Morgans, where *g* is the number of generations since the common ancestor (Howrigan *et al.,* 2011). Therefore, considering that 16 Mb segments are expected to present inbreeding up to 3 generations ago, recent inbreeding was observed in the studied local breeds due to the higher frequencies of ROH in this length category (Table 3), whereas the short ROH segments observed in Italian Holstein (4 Mb) was related to more ancient inbreeding, occurring 12.5 generation ago (about 75 years ago). However, the findings suggest that the local breeds experienced both recent and ancient inbreeding events, since that some animals lacked such long ROH, whereas other showed long segments. The results also indicated that these breeds have not recently been extensively crossed with other ones otherwise the long ROH would have broken down.

One of the main advantages of genomic coefficients is the availability of chromosomal inbreeding coefficients. F_{ROHBTA} estimates were reported in Figure 3. In general, for each breed, the F_{ROHBTA} values followed the same pattern as those computed for the whole genome. Higher F_{ROHBTA} values were found on BTA28 (for Cinisara), BTA16 (Modicana), BTA26 (Italian Holstein), and BTA23 (Reggiana) whereas for all breeds the lowest one was found in BTA5. In a previous study on Italian Holstein, Gaspa *et al*. (2014) identified an interesting region of ~ 2 Mb on BTA26 that harbors some genes involved in the metabolism of mammary gland. Similar values were reported by Marras *et al.* (2014) in Italian Simmental and local Marchigiana cattle breeds.

In the absence of pedigree information, the origin of ROH could also be explained using other indicators, as LD and N_e. In fact, another explanation for ROH is the lack of recombination in a specific region. Pairwise r^2 values were averaged over all autosomes and plotted as a function of genomic distance between markers (Figure 4). The highest level of r^2 was found in Italian Holstein whereas the lowest one in Cinisara. The extent of LD was used to estimate current and past N_e that is an important parameter for the assessment of genetic diversity and helps to explain how population evolved (Tenesa *et al*., 2007). In the four breeds, the highest N_e (estimated five generation ago) was observed in Cinisara (94.58) whereas the lowest one was observed in Modicana (59.84) (Table 4). For Sicilian breeds, the N_e estimates based on LD were substantially higher than those reported in a previous study (Mastrangelo *et al*., 2014) calculated from the rates of *F* and *f*. Different estimates for N_e were also reported in Iberian pigs with complete and accurate pedigree records, where N_e calculated from the rates of molecular *F* and *f* were 17 and 10, respectively (Saura *et al*., 2013), whereas N_e estimate using information from LD and recombination rate was 36 (Saura *et al*., 2014). Therefore, the discrepancies were due to the different used methods. In fact, as for the pedigree-based methods, the different molecular methods may give divergent results depending on the sampling strategy or the parameters used to compute N_e (Leroy *et al*., 2013). These methods differ also in terms of time scale investigated and the amount of available information. The rates of *F* and *f* only give estimates of N_e based on limited time period, and taking into account the year of birth of individuals (that in local breeds as Cinisara and Modicana may be incorrect) may result in biased estimates. LD-based method uses more information, leads to an accurate estimates (Waples and Do, 2010; Waples and England, 2011; Saura *et al*., 2015), with the possibility of

investigating the change of N_e over time, as LD between loci at a specific recombination distance reflects the ancestral N_e 1/2c generations ago (Hayes *et al.*, 2003), if the population grows linearly over time. However, it should be underlined that some parameters, as density and frequency of SNP pairs and distribution of MAF, affect the estimations of LD (Ober *et al*., 2013) and then of N_e. Moreover, the methods used to convert physical distances between SNP pairs to genetic distance may result in different estimated N_e values (García-Gámez *et al.*, 2012). Estimate of N_e obtained in this study for Italian Holstein was closed to those previously published for other Holstein population (Rodríguez-Ramilo *et al*., 2015). In general, the breed with the highest average inbreeding coefficient had the lowest N_e, as in Modicana breed. Moreover, LD and N_e were influenced by the recent history of selection. In fact, the strong selection for milk production and artificial insemination in Holstein and the highest inbreeding in Modicana have led to a reduction in the N_e.

In Table 5 the average inbreeding and coancestry molecular coefficients estimated using different approaches were reported. Cinisara presented the highest values for all F coefficients (F_{GRM}, F_{HOM}, and $F_{MOL\ i}$); Modicana showed the lowest values for F_{GRM} and F_{HOM} and the highest value for $f_{MOL\ ij}$ (Table 5). Italian Holstein breed showed the lowest values of $f_{MOL\ ij}$ and $F_{MOL\ i}$. Estimates of inbreeding coefficients depend on the used methods. In fact, F coefficients estimated using allele frequencies (F_{HOM} and F_{GRM}) showed considerable variation among breeds respect to F_{ROH} and $F_{MOL\ i}$. In all breeds, $f_{MOL\ ij}$ and $F_{MOL\ i}$ values were much higher than the other coefficients because these two methods (that are obtained on a SNP-by-SNP basis) do not discriminate alleles that are IBD or IBS (Rodríguez-Ramilo *et al*., 2015). However, these estimates computed from SNP array data were strongly correlated with genealogical estimates, represent a useful alternative to genealogical

376 information for measuring and maintaining genetic diversity and are very accurate in
377 predicting genealogical coancestry (Gómez-Romano *et al*., 2013; Saura *et al*., 2013).
378 Spearman's rank correlation between F_{ROH} and the other genomic inbreeding
379 estimated measures was calculated (Table 6). High correlation was found between
380 F_{HOM} and F_{ROH} ranged from 0.83 in Reggiana to 0.95 in Cinisara and Modicana. The
381 correlations among F_{ROH} and other inbreeding estimates (F_{GRM}, F_{HOM}, and $F_{MOL\ i}$) were
382 generally lower ranged from 0.45 ($F_{MOL\ i}$ - F_{ROH}) in Cinisara to 0.17 (F_{GRM} - F_{ROH}) in
383 Modicana (Table 6). High correlation between F_{HOM} and F_{ROH} (0.84) was also
384 reported by Zhang *et al*. (2014) in a study on pig in which ROH > 5 Mb after LD-
385 pruning were detected, whereas really different values (0.06, 0.35, and 0.61) were
386 reported by Zhang *et al*. (2015) in three cattle breeds. Ferenčaković *et al*. (2013a)
387 reported high correlation between F_{HOM} and F_{ROH} based on short segments (ROH > 1
388 and > 2 Mb). The poor correlation reported in our study between F_{GRM} and F_{ROH} was
389 in according to other studies (Marras *et al*., 2014; Zhang *et al*., 2015). Zavarez *et al*.
390 (2015) in a study on autozygosity using high-density SNPs, showed that the
391 correlation between F_{GRM} and F_{ROH} decreased from 0.74 per ROH > 0.5 Mb to 0.41
392 per ROH > 16 Mb, probably due to the properties of the G matrix which is based on
393 individual loci, whereas F_{ROH} is based on chromosomal segments. A higher
394 correlation between $F_{MOL\ i}$ and F_{ROH} were reported by Gómez-Romano *et al*. (2014) in
395 Austrian Brown Swiss cattle (0.76) and Rodríguez-Ramilo *et al.* (2015) in Spanish
396 Holstein breed (0.88). However, while the alternative used estimates of inbreeding
397 and coancestry coefficients could not distinguish between recent and ancient
398 inbreeding, F_{ROH} provided the direct estimated level of autozygosity in the current
399 populations and allowed us to detect recent inbreeding (up to three generations ago)
400 in the local cattle breeds, in particular for Cinisara and Modicana ones. In fact, in the

Sicilian farming system, natural mating is the common practice for local breeds, and the exchange of animal among flocks is quite unusual, with an increase of inbreeding within the population due to uncontrolled mating of related individuals (Mastrangelo *et al*., 2012). As pedigree data were unavailable for animals in this study, comparison of genomic and pedigree inbreeding coefficients was not possible. However, the strong correlation between the pedigree inbreeding coefficient and the sum of ROH reported by several authors (Purfield *et al*., 2012; Ferenčaković *et al*.*,* 2013b) suggests that in absence of animal's pedigree data, the extent of a genome under ROH may be used to infer aspects of recent population history even from relatively few samples. It should be underlined that the occurrence of ROH in an individual may be the result of inbreeding events but they may also be present in outbreed populations as result of other phenomena. In fact, an increased frequency of common extended haplotypes can also be a consequence of selection pressure on genomic regions involved in functional roles (Gaspa *et al.,* 2014), but as reported above, Sicilian cattle breeds are not subject to selection programs, therefore the presence of ROH in these two breeds was only due to inbreeding effect. Moreover, recent studies showed that the genomic estimates of inbreeding can be used to calculate the effects of inbreeding on performance and fitness traits. Pryce *et al*. (2014), in a study on the identification of genomic regions associated with inbreeding depression in Holstein cattle breed, showed that long ROH (>60 SNPs or 3.5 Mb), as those identified in our breeds, were associated with a reduction in milk yield, independently of the proportion of the genome that was homozygous. Therefore, our results showed the necessity of implementing conservation programs to preserve the local breeds in order to avoid further loss of genetic distinctiveness.

Selection and mating strategies have been proposed in the past for controlling inbreeding and coancestry. The best know strategy to achieve these goals is optimizing the contributions of the parents to minimize global coancestry in their offspring (Fernandez *et al*., 2003). Recently, measures of coancestry based on IBD segments (de Cara *et al*., 2013) and on shared segments of the genome (Bosse *et al*., 2015) have been proposed as good balance between maintaining diversity and fitness, with a higher fitness than managing with molecular coancestry and higher diversity than managing with genealogical coancestry. Therefore, determining the occurrence of IBD segments in potential parents, thereby measuring their relatedness and coancestry, can be used to minimize the occurrence of long ROH in the offspring. The availability of genome-wide genotyping platforms allows us now to study populations from a more detailed perspective, providing information on the genetic status and on their evolution across time.

Conclusion

This study has reported for the first time the genome-wide inbreeding estimate using runs of homozygosity in three Italian local cattle breeds. The obtained results highlight differences in detection and in distribution of ROH among breeds. In particular, Cinisara and Modicana breeds showed long ROH segments and the presence of inbreeding due to recent consanguineous mating. Therefore, our results showed the necessity of implementing conservation programs with the aim to control the level of inbreeding. The control of coancestry would restrict inbreeding depression, the probability of losing beneficial rare alleles, and therefore the risk of extinction for these local cattle breeds, and may be crucial for implementing genetic improvement programs. Breeders should be aware of this situation, and breeding

systems should be designed to foster and maintain genetic variation in these populations. Avoiding mating among relatives, together with other actions (e.g. sires/dams ratio, balanced progeny sizes) are strategies to control the increase of inbreeding.

Acknowledgements

This research was financed by PON02_00451_3133441, CUP: B61C1200076005 funded by MIUR. The authors would like to thank two anonymous referees for valuable comments which helped to improve the manuscript.

References

Allendorf FW, Hohenlohe PA and Luikart G 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11, 697-709.

Barrett JC, Fry B, Maller J and Daly MJ 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263-265.

Bjelland DW, Weigel KA, Vukasinovic N and Nkrumah JD 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. Journal of Dairy Science 96, 4697-4706.

Bosse M, Megens HJ, Madsen O, Crooijmans RP, Ryder OA, Austerlitz F, Groenen MAM and de Cara MAR 2015. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Research 25, 970-981.

Caballero A and Toro MA 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conservation Genetics 3,289-299.

de Cara MÁR, Villanueva B, Toro, MA and Fernández J 2013. Using genomic tools to maintain diversity and fitness in conservation programmes. Molecular Ecology 22, 6091-6099.

Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I and Sölkner J 2013a. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics 130, 286-293.

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

Fernández J, Toro MA and Caballero A 2003. Fixed contributions designs vs. minimization of global coancestry to control inbreeding in small populations. Genetics 165, 885-894.

Fernández J, Meuwissen THE, Toro MA and Mäki-Tanila A 2011. Management of genetic diversity in small farm animal populations. Animal 5, 1684-1698.

Fontanesi L, Scotti E, Samorè AB, Bagnato A and Russo V 2015. Association of 20 candidate gene markers with milk production and composition traits in sires of Reggiana breed, a local dairy cattle population. Livestock Science 176, 14-21.

García-Gámez E, Sahana G, Gutiérrez-Gil B and Arranz JJ 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genetics 13, 43.

Gaspa G, Marras G, Sorbolini S, Ajmone Marsan P, Williams JL, Valentini A, Dimauro C and Macciotta NPP 2014. Genome-Wide Homozygosity in Italian Holstein Cattle using HD Panel. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 17-22 August 2014, Vancouver, Canada.

Gibson J, Morton N and Collins A 2006. Extended tracts of homozygosity in outbred human populations. Human Molecular Genetics 15, 789-795.

Gómez-Romano F, Villanueva B, de Cara MAR and Fernández J 2013. Maintaining genetic diversity using molecular coancestry: the effect of marker density and effective population size. Genetics Selection Evolution 45, 38.

Hayes BJ, Visscher PM, McPartlan HC and Goddard ME 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research 13, 635-643.

505 Howrigan DP, Simonson MA and Keller MC 2011. Detecting autozygosity through runs of
506 homozygosity: a comparison of three autozygosity detection algorithms. BMC
507 Genomics 12, 460.
508 Keller MC, Visscher PM and Goddard ME 2011. Quantification of inbreeding due to distant
509 ancestors and its detection using dense single nucleotide polymorphism data. Genetics
510 189, 237-249.
511 Kim ES, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y and
512 Sonstegard TS 2013. Effect of artificial selection on runs of homozygosity in U.S.
513 Holstein cattle. PLoS One 8, e80813.
514 Kirin M, McQuillan R, Franklin C, Campbell H, McKeigue P and Wilson J 2010. Genomic
515 runs of homozygosity record population history and consanguinity. PLoS One 5,
516 e13996.
517 Leroy G, Mary-Huard T, Verrier E, Danvy S, Charvolin E and Danchin-Burge C 2013.
518 Methods to estimate effective population size using pedigree data: Examples in dog,
519 sheep, cattle and horse. Genetics Selection Evolution 45, 1-10.
520 Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsam P, Valentini A, Williams JL and
521 Macciotta NPP 2014. Analysis of runs of homozygosity and their relationship with
522 inbreeding in five cattle breeds farmed in Italy. Animal Genetics 46, 110-121.
523 Mastrangelo S, Sardina MT, Riggio V and Portolano B 2012. Study of polymorphisms in the
524 promoter region of ovine *β*-lactoglobulin gene and phylogenetic analysis among the
525 Valle del Belice breed and other sheep breeds considered as ancestors. Molecular
526 Biology Reports 39, 745-751.
527 Mastrangelo S, Saura M, Tolone M, Salces-Ortiz J, Di Gerlando R, Bertolini F, Fontanesi L,
528 Sardina MT, Serrano M and Portolano B 2014. The genome-wide structure of two
529 economically important indigenous Sicilian cattle breeds. Journal of Animal Science 92,
530 4833-4842.
531 McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L,
532 Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM,

533 Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H
534 and Wilson JF 2008. Runs of homozygosity in European populations. The American
535 Journal of Human Genetics 83,359-372.
536 Ober U, Malinowski A, Schlather M and Simianer H 2013. The expected linkag disequilibrium
537 in finite populations revisited. ArXiv preprint 1304.4856.
538 Ouborg NJ, Pertoldi C, Loeschcke V, Bijlsma R and Hedrick PW 2010. Conservation
539 genetics in transition to conservation genomics. Trends in Genetics 26, 177-187.
540 Pertoldi C, Purfield DC, Berg P, Jensen TH, Bach OS, Vingborg R and Kristensen TN 2014.
541 Genetic characterization of a herd of the endangered Danish Jutland cattle. Journal of
542 Animal Science 92, 2372-2376.
543 Pryce JE, Haile-Mariam M, Goddard ME and Hayes BJ 2014. Identification of genomic
544 regions associated with inbreeding depression in Holstein and Jersey dairy cattle.
545 Genetics Selection Evolution 46, 71.
546 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De
547 Bakker PI, Daly MJ and Sham PC 2007. PLINK: A tool set for whole-genome
548 association and population-based linkage analyses. The American Journal of Human
549 Genetics 81, 559-575.
550 Purfield DC, Berry DP, McParland S and Bradley DG 2012. Runs of homozygosity and
551 population history in cattle. BMC Genetics 13, 70.
552 Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D and Villanueva B 2015.
553 Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in
554 the Spanish Holstein Population. PloS one 10, 4.
555 Saura M, Fernández A, Rodríguez MC, Toro, MA, Barragán C, Fernández AI and Villanueva
556 B 2013. Genome-wide estimates of coancestry and inbreeding depression in an closed
557 herd of ancient Iberian pigs. PloS one 8, e78314.
558 Saura M, Woolliams JA, Tenesa A, Fernández A and Villanueva B 2014. Estimation of
559 ancient and recent effective population size from linkage disequilibrium in a closed herd

of Iberian pigs. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 17-22 August 2014, Vancouver, Canada.

Saura M, Tenesa A, Woolliams JA, Fernández A and Villanueva B 2015. Evaluation of the linkage-disequilibrium method for the estimation of effective population size when generations overlap: an empirical case. BMC Genomics16, 922.

Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C and Fernández AI 2013. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. Journal of Animal Breeding and Genetics 130, 349-360.

Sved JA 1971. Linkage disequilibrium of chromosome segments. Theoretical Population Biology 141, 125-141.

Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME and Visscher PM 2007. Recent human effective population size estimated from linkage disequilibrium. Genome Research 17, 520-526.

Toro MA, Meuwissen THE, Fernández J, Shaat I and Mäki-Tanila A 2011. Assessing the genetic diversity in small farm animal populations. Animal 5, 1669-1683.

VanRaden PM, Olson KM, Wiggans GR, Cole JB and Tooker ME 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. Journal of Dairy Science 94, 5673-5682.

Waples RS and Do C 2010. Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. Evolutionary Applications 3, 244-262.

Waples R and England PR 2011. Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration. Genetics 189, 633-644.

Zavarez LB, Utsunomiya YT, Carmo AS, Neves HH, Carvalheiro R, Ferenčaković M, O'Brien P, Curik I, Cole JB, Van Tassell CP, da Silva MVGB, Sonstegard TS, Sölkner J and Garcia JF 2015. Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Frontiers in Genetics 6, 5.

Zhang Q, Calus MP, Guldbrandtsen B, Lund MS and Sahana G 2015. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genetics 16, 88.

Zhang Y, Young JM, Wang C, Sun X, Wolc A and Dekkers JCM 2014. Inbreeding by Pedigree and Genomic Markers in Selection Lines of Pigs. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, 17-22 August 2014, Vancouver, Canada.

Table 1 *Comparison of inbreeding derived from Runs of Homozygosity (F_{ROH}) values obtained by allowing different numbers of heterozygous (het) SNPs*

Breed	$F_{ROH>4Mb}$		
	1 het SNP	2 het SNPs	3 het SNPs
Cinisara	0.052[a]	0.054[b]	0.056[c]
Modicana	0.055[a]	0.058[b]	0.062[c]
Italian Holstein	0.042[a]	0.045[b]	0.049[c]
Reggiana	0.035[a]	0.036[b]	0.039[c]

[a, b, c] Different letters indicate statistical significance within the same breed (P<0.001, paired t-test).

599 **Table 2** *Descriptive statistics for Runs of Homozygosity (ROH) for each cattle breed*

Breed	MN_{ROH}	$F_{ROH>4Mb}$	L_{ROH}	SNPs
Cinisara	9.38	0.052±0.064	13.57	49-1771
	(0 – 34)	(0.000 - 0.266)	(4 – 112.65)	
Modicana	11.03	0.055±0.053	12.31	45-1010
	(0 – 40)	(0.000 - 0.268)	(4 – 89.61)	
Italian Holstein	10.42	0.042±0.023	10.16	48-716
	(2 – 22)	(0.006 - 0.163)	(4 – 70.11)	
Reggiana	7.15	0.035±0.040	11.78	44-1135
	(0 – 47)	(0.000 - 0.285)	(4 – 102.18)	

600 MN_{ROH} = mean number of ROH per individual with minimum and maximum value in brackets; $F_{ROH>4Mb}$
601 = mean ROH-based inbreeding coefficient with standard deviation and minimum and maximum value
602 in brackets; L_{ROH} = average length of ROH in Mb with minimum and maximum value in brackets; SNPs
603 = minimum and maximum number of single nucleotide polymorphisms (SNPs) involved in ROH.
604

605 **Table 3** *Descriptive statistics of the number and the frequency distribution of Runs of*

606 *Homozygosity (ROH) in different ROH length categories (Mb) for each cattle breed*

	ROH length categories (Mb)					
	4-8		8-16		>16	
	n ROH	Freq	n ROH	Freq	n ROH	Freq
Cinisara	294	0.44	207	0.31	165	0.25
Modicana	403	0.51	217	0.27	173	0.22
Italian Holstein	504	0.50	371	0.37	125	0.13
Reggiana	531	0.44	426	0.35	245	0.21

607 n ROH = number of Runs of Homozygosity; Freq = relative frequency of Runs of Homozygosity on
608 different ROH length categories.
609

Table 4 *Effective population size (Ne) estimated from Linkage Disequilibrium values for each cattle breed*

	Effective population size	
Breed	50 generations ago	5 generations ago
Cinisara	657.42	94.58
Modicana	341.70	59.84
Italian Holstein	320.25	69.61
Reggiana	519.21	87.20

614 **Table 5** *Estimated mean of genomic inbreeding and coancestry coefficients for each*

615 *cattle breed*

Breed	F_{GRM}	F_{HOM}	$F_{MOL\ i}$	$f_{MOL\ ij}$
Cinisara	0.098	0.025	0.669	0.662
Modicana	0.036	-0.015	0.664	0.670
Italian Holstein	0.042	-0.014	0.653	0.658
Reggiana	0.074	-0.009	0.659	0.661

616 F_{GRM} = inbreeding coefficient based on genomic relationship matrix; F_{HOM} = inbreeding coefficient
617 based on the difference between observed versus expected number of homozygous genotypes; $F_{MOL\ i}$
618 = molecular inbreeding coefficient of individual *i*; $f_{MOL\ ij}$ = molecular coancestry coefficient between
619 individuals *i* and *j.*

Table 6 *Correlation between F_{ROH} and other genomic inbreeding coefficients for each cattle breed*

Correlation	Cinisara	Modicana	Italian Holstein	Reggiana
F_{HOM} - F_{ROH}	0.95***	0.95***	0.89***	0.83***
F_{GRM} - F_{ROH}	0.42***	0.17	0.18	0.26**
$F_{MOL\ i}$ - F_{ROH}	0.45***	0.27*	0.31*	0.44***

F_{HOM} = inbreeding coefficient based on the difference between observed versus expected number of homozygous genotypes; F_{ROH} = inbreeding coefficient based on the Runs of Homozygosity; F_{GRM} = inbreeding coefficient based on genomic relationship matrix; $F_{MOL\ i}$ = molecular inbreeding coefficient of individual *i*. **P*-values < 0.05; ***P*-values < 0.01; ****P*-values < 0.001.

Figure captions

Figure 1 Relationship between the total number of Runs of Homozygosity (ROH)>4 Mb (on y axes) and length in kb of genome in such ROH (on x axes) for individuals in Cinisara, Modicana, Reggiana, and Holstein cattle breeds. Each dot represents an individual.

Figure 2 Box plots of within-breed average and median sum of all ROH segments per individual. CIN, Cinisara; MOD, Modicana; HOL, Holstein; REG, Reggiana.

Figure 3 Distribution of inbreeding coefficient estimates for each chromosome ($F_{ROH_{BTA}}$) calculated as the proportion of BTA in ROH over the length of the BTA covered by the involved SNPs. CIN, Cinisara; MOD, Modicana; HOL, Holstein; REG, Reggiana.

Figure 4 Linkage disequilibrium across the genome as a function of genomic distance (Mb).

Supplementary material

Supplementary Table S1 Length of each chromosome (L_{BTA}) covered by SNPs

Figure 1

Click here to download Figure Figure 1.tif

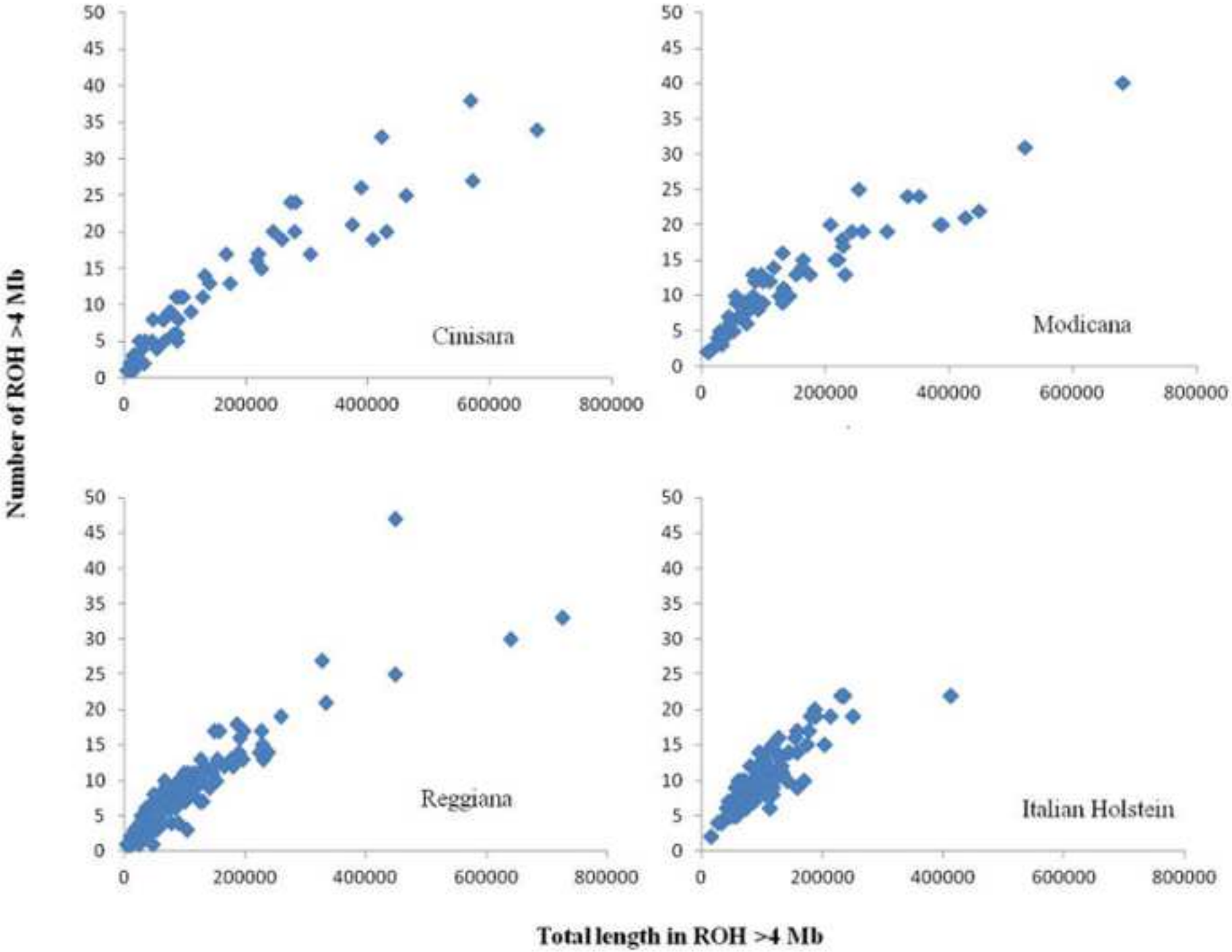

Figure 2

Click here to download Figure Figure 2.tiff

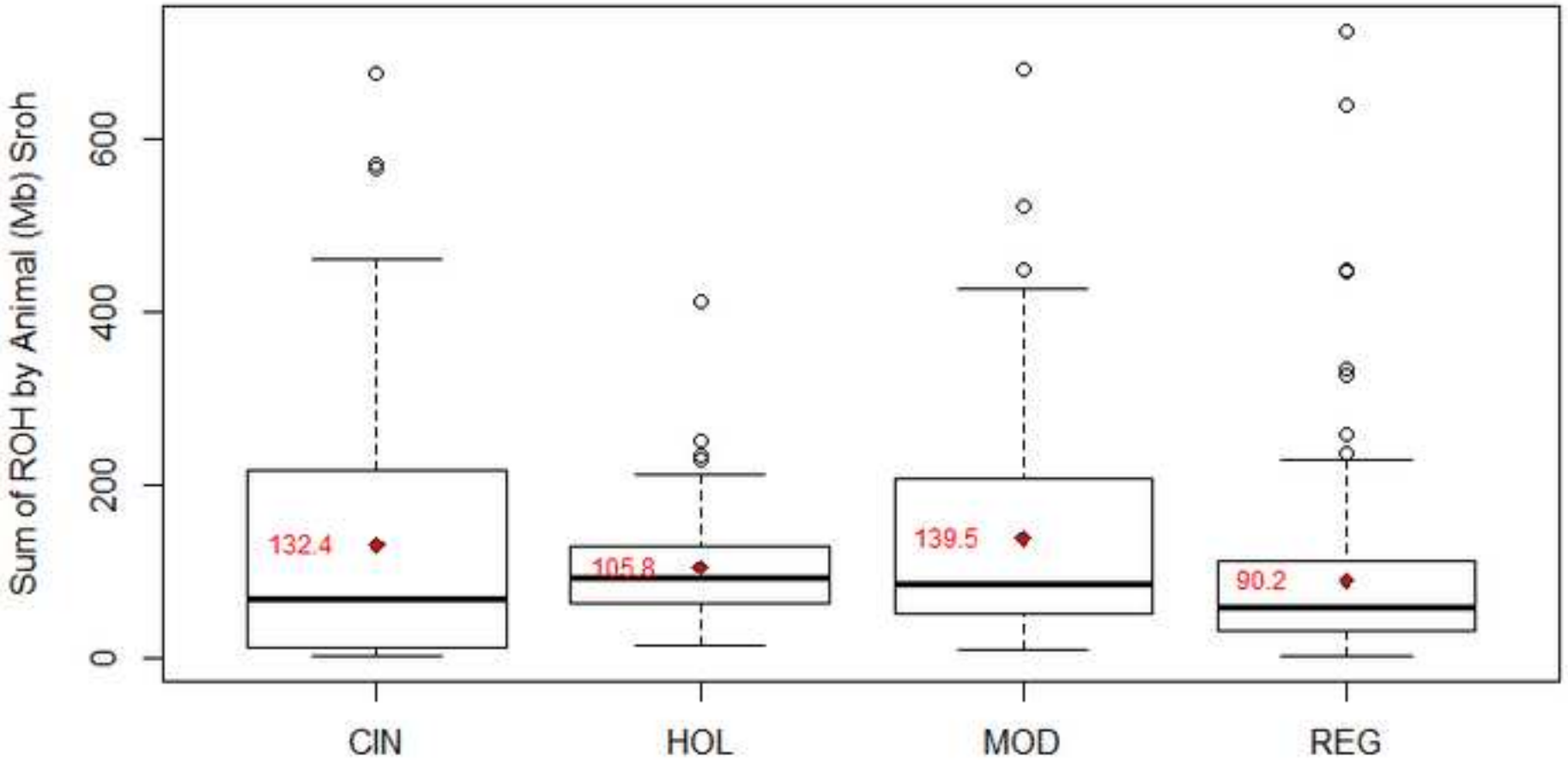

Figure 3

Click here to download Figure Figure 3.tif

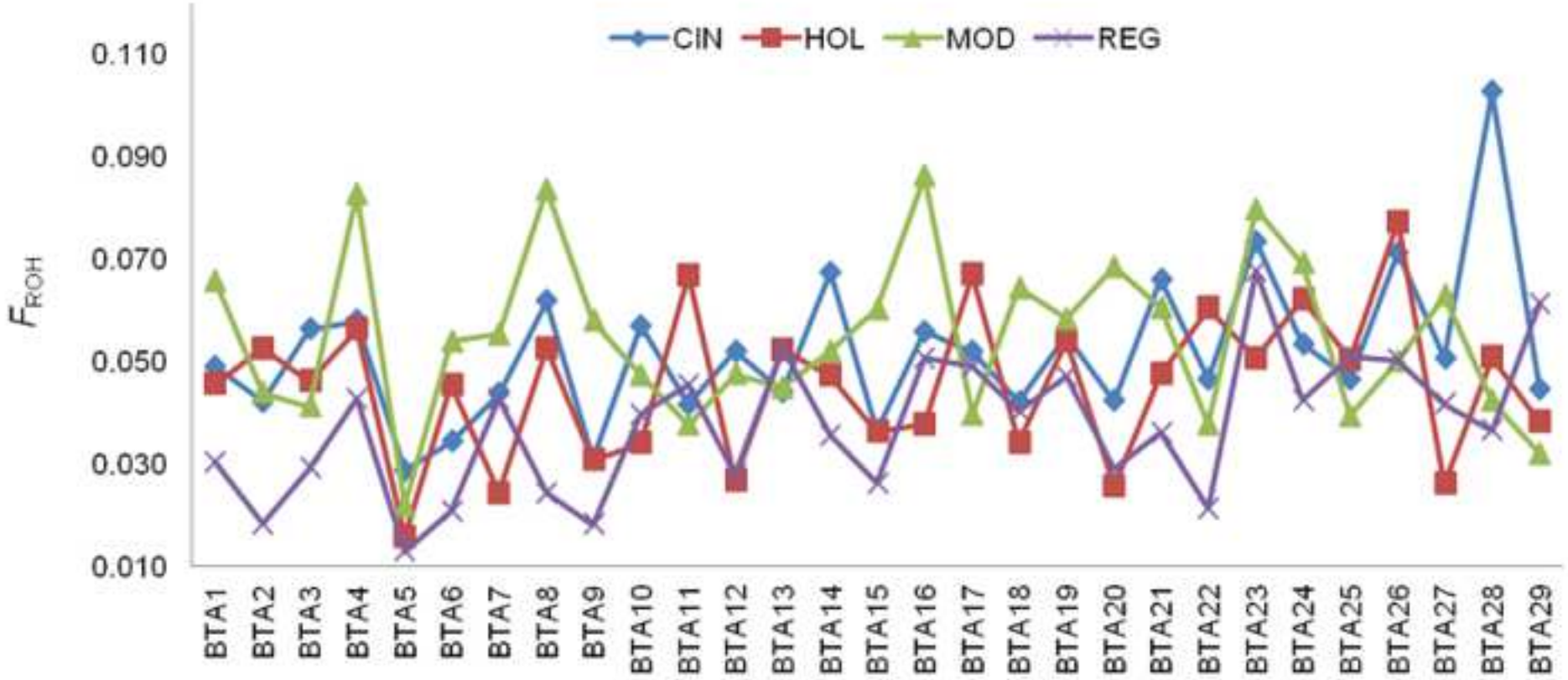

Figure 4

Click here to download Figure Figure 4.tif