



UNIVERSITÀ DI PARMA

ARCHIVIO DELLA RICERCA

University of Parma Research Repository

Genetic Diversity in the Italian Holstein Dairy Cattle Based on Pedigree and SNP Data Prior and After Genomic Selection

This is the peer reviewed version of the following article:

Original

Genetic Diversity in the Italian Holstein Dairy Cattle Based on Pedigree and SNP Data Prior and After Genomic Selection / Ablondi, Michela; Sabbioni, Alberto; Stocco, Giorgia; Cipolat-Gotet, Claudio; Dadousis, Christos; Kaam, Jan-Thijs van; Finocchiaro, Raffaella; Summer, Andrea. - In: FRONTIERS IN VETERINARY SCIENCE. - ISSN 2297-1769. - 8:(2022). [10.3389/fvets.2021.773985]

Availability:

This version is available at: 11381/2910489 since: 2022-03-17T09:01:55Z

Publisher:

Published

DOI:10.3389/fvets.2021.773985

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

1 Genetic diversity in the Italian Holstein dairy cattle based on pedigree 2 and SNP data prior and after genomic selection

3 Michela Ablondi¹, Alberto Sabbioni¹, Giorgia Stocco¹, Claudio Cipolat-Gotet^{1*}, Christos
4 Dadousis¹, Jan-Thijs van Kaam², Raffaella Finocchiaro², Andrea Summer¹

5 ¹Dipartimento di Scienze Medico-Veterinarie, University of Parma, Via del Taglio 10, 43126, Parma,
6 Italy

7 ²Associazione Nazionale Allevatori della Razza Frisona Bruna e Jersey Italiana, Cremona, Italy

8 * Correspondence:

9 Corresponding Author Claudio Cipolat-Gotet

10 claudio.cipolatgotet@unipr.it

11 **Keywords: Inbreeding, Runs of Homozygosity, Effective population size, Genomics, Dairy**
12 **industry**

13 Abstract

14 Genetic diversity has become an urgent matter not only in small local breeds but also in more
15 specialized ones. While the maximization of genetic gain has reached astonishing levels thanks to the
16 genomic era, it is counterbalanced by the potential loss of variability. Thus, in this study, we aimed to
17 investigate the genetic diversity using data from Italian Holstein cows born between 2002 and 2020.
18 We estimated variation in inbreeding, effective population size and generation interval and compare
19 those aspects prior to and after the introduction of genomic selection in the breed. A total of 84,443
20 SNPs and 74,485 cows were kept after the quality control. Pedigree depth based on complete
21 generation equivalent was equal to 10.67. The average F_{ped} was 0.07 while the average F_{ROH} was more
22 than double, being equal to 0.17. The pattern of the effective population size based on pedigree and
23 genotype data was similar although different in scale, with a constant decrease till last five generations
24 ago. The overall inbreeding rate (ΔF) per year was equal to +0.27% and +0.44% for F_{ped} and F_{ROH}
25 throughout the studied period, which corresponds to a $\sim +1.35\%$ and $\sim +2.2\%$ per generation,
26 respectively. A significant increase in the ΔF was found since the introduction of genomic selection in
27 the breed. This study in the Italian Holstein cattle showed the importance of controlling the loss of
28 genetic diversity to ensure the long-term sustainability of this breed, as well as to guarantee future
29 market demands.

30 1 Introduction

31 The Holstein dairy cattle have been bred in a pioneering effort to increase milk yield over the last
32 century and, more recently, emphasis on functional traits has increased. This breed is counted in over
33 150 countries, and it is the most common dairy cattle breed worldwide (1). Despite its abundance,
34 concerns on the decreasing level of genetic diversity of the Holstein populations have been recently
35 raised in several countries (2–4). These current alarms stem from the intense selection practiced in the
36 last century which has caused loss of genetic variability. The nineteenth century harbored the adoption
37 of selection theories, such as the selection index theory (5) and advanced statistical methods, such as
38 the best linear unbiased prediction (BLUP) (6), promoting a remarkable improvement in dairy cattle
39 breeds. In addition, the implementation of artificial insemination (AI) over 80 years ago, boosted the

40 impact of elite bulls worldwide, thus allowing an even superior genetic gain per generation (7). In the
41 last twenty years, the advances in high-throughput genotyping procedures allowed the development of
42 single nucleotide polymorphism (SNP) chips at reasonable price, thereby determining the application
43 of genomic selection (GS) based on SNP arrays (8) in numerous Holstein breeding programs among
44 different countries (4,9–11). A recent study evaluated the impact of the GS implementation in the US
45 Holstein and showed an outstanding increase in the annual genetic gain rate, ranging from 50% to
46 100% for yield traits and from 200% to 300% for fitness traits (12). The increase in the reliability of
47 genomic breeding values (GEBV) over traditional estimated breeding values (EBV) for young bulls is
48 remarkable, reaching up to a 20% increase for some traits (11). Since young bulls can now be selected
49 as parents based on their GEBV at a very early stage, generation intervals have been shortened
50 significantly (2,4,13,14). However, the maximization of genetic gain is counterbalanced by a reduction
51 of genetic diversity. It has been shown that the annual inbreeding rate has increased in several dairy
52 cattle populations after the implementation of GS (2,4,13). While the overall number of sires of bulls
53 has increased since the introduction of GS, the number of popular bulls (siring half of the young bulls
54 entering artificial insemination) has remained fairly stable (15).

55 Managing genetic diversity determines the long-term sustainability of the livestock production sector.
56 The impact of climate change on livestock, market demand fluctuations together with the increase in
57 human population, urgently require a sufficient reservoir of genetic diversity (16). It is therefore
58 crucial to evaluate genetic diversity in both commercial and local breeds to preserve biodiversity. The
59 most traditional approach to assess genetic diversity relies on pedigree information, where inbreeding
60 can be estimated as the probability of an individual to have two identical alleles by descent (17).
61 Analyses at pedigree level are particularly effective to evaluate the state of genetic diversity in small
62 and under-development populations with limited financial resources, thus causing the unavailability of
63 more advanced technologies such as genomic data (18–21). However, those estimates highly depend
64 on quality and depth of the pedigree information, and they rely on the assumption that no relationship
65 exists among founder animals; hence, they generally underestimate actual inbreeding coefficients
66 (3,22). The advent of genomics allowed researchers to acquire more precise insights on genetic
67 diversity by using genotype data (23). One of the most well-established methods to detect within-breed
68 loss of genotypic diversity is the runs of homozygosity (ROH) detection (24). Runs of homozygosity
69 are long consecutive homozygous segments distributed across the genome, which arise from identical-
70 by-descendent haplotype (25,26). Hence, ROHs have been commonly used to estimate genomic
71 inbreeding (F_{ROH}) in several species such as cattle (13,27–29), horses (30–33), pigs (34), sheep (35,36)
72 and goat (37,38). In contrast to pedigree-based estimates, F_{ROH} can capture the variation due to
73 Mendelian sampling and linkage during gamete formation (23).

74 In the Italian context, the most reared cattle breed is the Italian Holstein, counting more than 1,000,000
75 alive animals and about 9,500 breeders, with an average of 10,386 kg of milk produced per
76 lactation/cow in 2020 (39). However, fewer studies have evaluated the level of genetic diversity in this
77 breed. In 2015, inbreeding was evaluated based on pedigree and 50K SNP data in 2,093 Italian Holstein
78 bulls (40), while in 2018 50K SNP information from 96 animals were analyzed to identify ROHs (41).
79 A recent study that aimed to evaluate the presence of genomic divergence in Italian Holstein cows bred
80 in different production chains, calculated the inbreeding based on ROH in 1,000 Italian Holstein cows
81 (42). Nevertheless, to the best of the authors' knowledge, the genetic variability in the Italian Holstein
82 has not been fully explored, especially using a large database on the female side and with historical
83 data. Thus, in this study, we aimed to investigate the genetic diversity using data from Italian Holstein
84 cows born between 2002 and 2020. The specific aims of the study were: i) calculate the inbreeding
85 based on pedigree and genotype data, ii) evaluate changes in the effective population size throughout
86 generations and iii) test the effect of genomic selection on genetic diversity and generation interval.

87 2 Material and Methods

88 Records used in this study were obtained from archived data from the Italian National Association of
89 Holstein, Brown Swiss and Jersey Breeders (ANAFIBJ) and as such, no approval was required for
90 animal experimental purposes from the Animal Care Committee unit of the University of Parma. The
91 consent for data use was obtained by ANAFIBJ.

92 2.1 Pedigree Data

93 Pedigree records for the genotyped animals were provided by ANAFIBJ. Pedigree information
94 consisted of 393,607 individuals born between 1898 till 2020 with 26,226 males and 367,381 females
95 over 24 generations pedigree depth. To evaluate the role of pedigree depth in the inbreeding estimates,
96 the complete generation equivalent (CGE) was calculated using the optiSel package (43) in R software
97 (44).

98 2.2 Genotype Data

99 A total of 95,497 genotyped Italian Holstein cows born between 2002 to 2020 were available for this
100 study. Cows were genotyped with a variety of SNP panels, ranging from low to high density panels.
101 The animals genotyped with low density panels were imputed to medium density (85K) using
102 PedImpute (45). To guarantee high accuracy during the imputation pipeline, females were retained
103 only when both sire and sire of the dam were i) genotyped and ii) used in the imputation pipeline.
104 Quality Control (QC) excluding poorly genotyped and faulty data was performed on the 29 autosomal
105 chromosomes by using PLINK v1.90 (46). The QC was based on the following criteria: call rate <95%,
106 parent-offspring SNP mismatch <0.01, minor allele (<0.01) and genotype (<0.001) frequencies and
107 extreme deviation from Hardy-Weinberg equilibrium ($P < 0.005$).

108 2.3 Pedigree and genotype-based inbreeding coefficients

109 Pedigree-based inbreeding coefficient (F_{ped}) was defined as the probability of an individual to present
110 two identical alleles by descent (47). The coefficients were computed by using the optiSel (43) package
111 in R (44). SNP-based inbreeding coefficient was derived by means of ROH assessments, a segment-
112 based approach. The ROH segments were detected by using the detectRUNS package (48) in R (44),
113 and defined as follows: (i) at least 15 SNPs in a run, (ii) a minimum length of a run equal to 1 Mb, (iii)
114 a maximum distance between consecutive SNPs in a window 500 kb, (iv) a lower density limit of 1
115 SNP per 100 kb and (v) by allowing for a maximum of one missing and one heterozygous SNP in a
116 run. The genomic inbreeding coefficient (F_{ROH}) was calculated as follow (49):

$$117 \quad F_{ROH} = \frac{\sum L_{ROH}}{L_{AUTO}}$$

118 with L_{ROH} being the sum of the length of ROHs per cow and L_{AUTO} the total length of the autosomal
119 genome covered by SNPs (in this study 2.48 Gbp). The correlation between F_{ped} and F_{ROH} was
120 calculated by means of Pearson's product-moment correlation (r). To compare the variability of F_{ped}
121 and F_{ROH} , we calculated the coefficient of variation of these two measurements as the ratio between
122 the standard deviation of inbreeding and its overall mean. Moreover, the $r_{F_{ped}-F_{ROH}}$ was estimated per
123 year. To evaluate the effect of CGE on the relationship between F_{ped} and F_{ROH} , the database was divided
124 in cows with $CGE \leq 10$ ($N = 21,028$) and with $CGE > 10$ ($N = 53,457$). Based on the hypothesis that
125 ROH length reflects the chronological time points when inbreeding happened, the genomic inbreeding
126 was expressed separately for six length ROH categories (1>ROH ≤ 2 , 2> ROH ≤ 4 , 4> ROH ≤ 8 , 8>

127 ROH ≤ 16 , $16 < \text{ROH} \leq 32$, $\text{ROH} > 32$ Mbp) by dividing the database in four birth year classes,
128 considering an average generation interval of 5 years as found based on pedigree data.

129 2.4 Effective Population size

130 Effective population size (N_e) of an actual population can be defined as the size of a hypothetical ideal
131 population resulting in the same amount of genetic drift as is present in the real population (47). In this
132 study, we estimated the historical and recent N_e both based on pedigree and SNP data. The optiSel
133 (43) package in R (44) was used to estimate the N_e based on pedigree data from 1960 to 2020. The
134 SNeP v.1.1 software was used to estimate the trends of historical N_e based on Linkage Disequilibrium
135 (LD) on SNP data (50) by using animals born in the latest year (677 cows in 2020). The N_e was
136 estimated for the latest 30 generations as the first introduction of the Dutch Friesian in Italy dates back
137 to 1870 followed by North American Holstein Friesians from 1923. Since different methods for N_e
138 estimation based on Linkage Disequilibrium (LD) are available in the literature, two analyses were
139 performed by: a) using default settings, except for the recombination rate, which was inferred following
140 Sved & Feldman method (51) and the mutation rate in cattle [$\alpha = 2.2$ (52)]; b) including also a
141 restriction on the maximum distance used to calculate LD. The latter parameterization together with
142 the Sved & Feldman's mutation rate modifier allowed an estimation of N_e for the most recent
143 generations (52,53).

144 2.5 The role of genomic selection on genetic diversity and generation interval

145 The rate of inbreeding (ΔF) was calculated as the inverse of the slope of the regression of $\ln(1 - \bar{x})$
146 on year of birth, where \bar{x} was equal to the average of the parameter each year (F_{ped} and F_{ROH}) (2). The
147 annual rate was multiplied by the generation interval (in the Italian Holstein being equal on average to
148 5 years as shown in this study) to obtain the rate per generation (ΔF_{gen}). To assess the effect of different
149 selection strategies on genetic diversity in the Italian Holstein breed, i.e., classical progeny testing
150 (PTS) vs. GS, we divided the database in two 5 years birth cohorts by taking all animals born before
151 (2006 to 2010) and after the introduction of GS (2015 to 2019). We then tested the equality of PTS and
152 GS means. Moreover, to evaluate the impact of the introduction of GS on genetic diversity, we used
153 the following linear model, using the R function `lm` (44):

$$154 \quad Y_i = \begin{cases} \alpha_1 + \beta_{\text{PTS}}x_i + \varepsilon_i, & 2006 \leq x_i \leq 2010 \\ \alpha_1 + (\beta_{\text{PTS}} + \delta)x_i + \varepsilon_i, & 2015 \leq x_i \leq 2019 \end{cases}$$

155 where Y_i is the variable of interest for each cow i (F_{ped} and F_{ROH}), x_i is the birth year of each cow i , and
156 β_{PTS} is the associated coefficient of regression if cow i was born in the PTS cohort or $\beta_{\text{PTS}} + \delta$ if cow
157 i was born in the GS cohort. The impact of GS on the inbreeding rate was measured with the δ
158 coefficient. Analysis of variance was used to test the significance of the δ value. The relative change
159 (RC) of the slopes of regression before and after GS was computed as $RC = \frac{\delta}{\beta_1}$ with β_1 being the slope
160 in the second evaluated period. The assessment of the RC allowed the evaluation of slope value with
161 time (13). The effect of the introduction of GS was also evaluated via the length of the generation
162 interval (GI). The GI, defined as the average age of parents when their offspring were born, was
163 calculated for all the four selection pathways (sire of bulls, dam of bulls, sire of cows, and dam of
164 cows) using the 393,607 individuals present in the pedigree. The generation interval was estimated
165 using the Pedig software (54).

166 3 Results

167 3.1 Pedigree and genotype-based inbreeding coefficients

168 A total of 84,443 SNPs and 74,485 cows were kept after QC. The average genotyping call rate was
169 0.99 and the average pedigree depth based on CGE was equal to 10.67 (SD = 1.12). The cows
170 descended from 3,058 sires and 59,377 dams. In total, over 50% of the cows was born between 2016
171 and 2020 (last five years), whereas 34,286 cows were born between 2002 and 2015. The mean F_{ped} was
172 equal to 0.07 (SD = 0.02) ranging between 0.01 and 0.32 (CV = 0.29). The F_{ROH} mean value was more
173 than doubled (0.17; SD = 0.03) with minimum and maximum of 0.05 and 0.50, respectively (CV =
174 0.20). A total of 839 cows showed an F_{ped} exceeding the mean + 3 SD ($F_{ped} \geq 0.13$), which were defined
175 as highly inbred females. In the case of the F_{ROH} , 507 cows presented an inbreeding higher than the
176 mean +3 SD ($F_{ROH} \geq 0.27$) (Figure 1). Approximately 23% of the highly inbred cows were in common
177 from the comparison between the two methodologies. Pearson correlation between F_{ped} and F_{ROH} was
178 equal to 0.68 (confidence interval: 0.676 – 0.683), P -value < 2.2e-16) as shown in Supplementary
179 Figure S1. The average CGE in the entire database was equal to 10.06, showing an increase throughout
180 the studied period from 2002 (average CGE = 7.5) till 2019 (average CGE = 11.90) (Figure a). The
181 correlation ranged from 0.44 in 2005 (N = 136) to 0.89 in 2003 (N = 16); however, those estimates
182 should be evaluated with caution due to the low number of available animals for those years. Since
183 2010, the number of born and genotyped cows per year has been above 1,000, and the $r_{F_{ped}-F_{ROH}}$ was
184 steadier and ranged between 0.58 and 0.69. The r was higher in cows with CGE > 10 compared to those
185 with CGE \leq 10 (0.68 and 0.59, respectively; Figure 2b).

186 The F_{ROH} estimates and average number of ROH per animal based upon the six ROH length classes (to
187 differentiate old and recent inbreeding) of each of the four birth year cohorts are presented in Table 1.
188 In the case of ROH length classes: 1-2 Mbp, 2-4 Mbp and 4-8 Mbp, all the cows, regardless of the birth
189 year cohort, exhibited some degree of inbreeding with a mean value ranging from 0.02 in the 2-4 Mbp
190 and 4-8 Mbp in the two oldest birth year cohorts (2002-2005 and 2006-2010) to 0.04 in the shortest
191 length class for all birth classes. From the 8-16 Mbp class and above, not all the cows exhibited ROH
192 of such lengths. Only 39% of the females exhibited ROH in the longest length class (>32 Mbp) with
193 most of them belonging to the latest two birth year cohorts (2011-2015 and 2016-2020) (data not
194 shown).

195 3.2 Effective Population size based on genotype and pedigree data

196 The two settings for the N_e estimation based on LD resulted comparable with a N_e reduction from
197 generation 30 till generation 13 (Figure 3a). The N_e was equal to 140 animals at generation 30 and to
198 96 at generation 13 based on both settings. The second parameterization (Figure 3b) allowed the
199 evaluation of the N_e trend at more recent generations and showed a sharp N_e increase (almost doubled)
200 within the last five generations, reaching to 120 in the most recent years.

201 The patterns of the N_e based on pedigree and SNP data were similar to each other, with differences in
202 scale (Supplementary Figure S2). The N_e based on pedigree data was calculated from 1960 - 1964
203 (generation n.12) till 2016 - 2020 (generation n.1) (Supplementary Figure S2). This was equal to 87
204 and 55 animals in the oldest and in earliest generation, respectively. A decrease in the N_e was found
205 from generation 12 till generation 6 (1991 - 1995), with a N_e of 43 in generation 6. In contrast, since
206 generation 5 (1996 - 2000) a general increase was observed. Likewise for the N_e based on SNP data,
207 a decrease was observed since generation 9 (1976 - 1980), followed by a plateau, with N_e equal to 89
208 animals from generation 8 till generation 6 (1981 - 1995). An increase in the N_e was observed with a
209 value of 120 animals in the latest generation.

210 3.3 The role of genomic selection on genetic diversity and generation interval

211 Generation intervals based on pedigree information were calculated from 1960 till 2018 for all four
 212 pathways of selection by using all animals in the pedigree file (Figure 4). For sire of bulls and cows,
 213 an increase in GI was observed till 1984, with a GI equal to 11.08 years for bulls and 8.74 years for
 214 cows, respectively. In contrast, a tendency to decrease was found from 1985 onwards for both
 215 pathways. A noticeable drop occurred both in the sire of bulls and sire of cows from 2011 till 2018,
 216 with the lowest GI in 2017 for the former (2.34 years) and in 2018 for the latter (3.6 years). The dam
 217 of bulls pathway decreased from 1962 (8.64 years) to 1992 (3.8 years) with some oscillations. The GI
 218 in this pathway remained approximately steady from 1993 till 2011, with a minimum of 3.68 years in
 219 2010 and a maximum of 4.24 years in 1996, whereas it declined from 2012 till 2018 (2.43 years). In
 220 the case of dam of cows' pathway, a decrease was visible from 1960 (8.22 years) till 1992 (4.03 years)
 221 with few fluctuations. From 1993 onwards the GI remained stable with an average value of 3.68 years,
 222 varied between 3.01-4.0 years.

223 The overall inbreeding rate per year was equal to +0.27% and +0.44% for F_{PED} and F_{ROH} throughout
 224 the studied period, which corresponds to a $\sim +1.35\%$ and $\sim +2.2\%$ ΔF_{gen} , respectively. The mean
 225 difference in F_{PED} and F_{ROH} based on a two-sample t-test between the two periods (PTS and GS) was
 226 significant ($P < 0.001$) for both inbreeding estimates (Figure 5). The average F_{PED} was equal to 0.05 in
 227 the PTS and 0.07 in the GS. The average F_{ROH} was equal to 0.14 and 0.17 in the PTS and GS,
 228 respectively. The overall inbreeding rate per year in the PTS was equal to 0.14% and 0.32% based on
 229 F_{PED} and F_{ROH} , whereas it increased up to 0.47% (based on F_{PED}) and 0.70% (based on F_{ROH}) in the
 230 GS. The RC in inbreeding comparing the two periods (PTS and GS) was equal to 2.36 and 1.19 from
 231 the F_{PED} and F_{ROH} (Table 2, Figure 5). The overall GI (calculated as the average among the four
 232 pathways) decreased in the GS by a factor of 1.8 compared to the PTS period (Figure 5). Since this
 233 latter reduction, also the ΔF_{gen} changed between GS and PTS, being equal to +0.75% (based on F_{PED})
 234 and +1.72% (based F_{ROH}) in the PTS and 1.41% (based on F_{PED}) and 2.1% (based F_{ROH}) in the GS
 235 period.

236 4 Discussion

237 4.1 Pedigree and genotype-based inbreeding coefficients

238 In this study, we evaluated the genetic diversity in the Italian Holstein breed by using pedigree and
 239 SNP data from cows genotyped throughout a period of 19 years. The primary goal was to estimate
 240 variation in inbreeding, effective population size and generation interval and to compare those aspects
 241 prior to and after the introduction of genomic selection (GS) in the breed. In line with the development
 242 and spread of the genotyping technology in the dairy sector (55), from 2010 the number of genotyped
 243 Italian Holstein cows per year rapidly increased, and it reached over 15,000 cows in 2016. This amount
 244 of information provided new opportunities on exploring genetic diversity at genomic level. The most
 245 traditional source of information to evaluate genetic diversity is pedigree information (47,56).
 246 However, the accuracy of genetic diversity estimated using pedigree data highly relies on the quality
 247 and depth of the recorded data (57). In this study, the pedigree depth showed high values with an
 248 average CGE of 10.67, which agrees with previous studies on the Dutch Holstein breed (CGE between
 249 ~ 10 and ~ 14) (58,59). Nevertheless, the CGE influenced the correlation between F_{ped} and F_{ROH} . Indeed,
 250 the correlation dropped to 0.59 when considering cows with a CGE lower than 10, and it reached a
 251 value of 0.68 for animals with CGE above 10 ($r = 0.68$). The reduction of the correlation values for
 252 animals with lower CGE was expected, as the genealogical approach strongly depends on pedigree
 253 completeness and quality. Nevertheless, the obtained correlations are in line with the values found in
 254 previous studies, regardless of the CGE assessment. In the Spanish Holstein population, a correlation
 255 of 0.57 was shown (14), and in four different cattle breeds (Brown Swiss, Fleckvieh, Norwegian Red

Genetic diversity in the Italian Holstein

256 and Tyrol Grey) correlations were reported to be between 0.50 to 0.72 (60). A unity correlation is not
257 expected between F_{ped} and F_{ROH} , since the former does not account for the Mendelian sampling
258 variation, as the latter does (26,61).

259 4.2 Effective Population size based on genotype and pedigree data

260 Prior to herdbook formation, the N_e was large for most cattle populations, i.e. in the order of tens of
261 thousands (7,62). However, the advent of closed policy in the breeding management from the 19th
262 century onwards, reduced drastically the gene flow among populations, causing loss of genetic
263 diversity. The major changes in breeding programs in the last century have exacerbated this reduction.
264 It has been shown that current N_e in most modern and specialized breeds is close to 100 animals (7).
265 Estimates of N_e found in this study based on pedigree and genotype data are in the range of those
266 formerly published in other Holstein populations. Pedigree-based N_e estimates ranged from 39 in the
267 US Holstein in 2000 (63) to 114 in the Canadian Holstein population between 2000 - 2007 (64). In this
268 study, the N_e based on pedigree in 2000 was equal to 47 animals, thus slightly higher compared to the
269 US Holstein population, yet comparable. In contrast, over two-time lower N_e between 2000 - 2007 (N_e
270 = 49) was found in our study compared to what reported in the Canadian Holstein population. In that
271 period, there was a reduction inbreeding rate per generation in Canadian Holstein (64) due to the
272 introduction of the average relationship values in the breeding program by the Canadian Dairy Network
273 in 2000. Since in the Italian Holstein this type of control was not implemented at that time, we suspect
274 this might be the reason behind those differences among the two Holstein populations in terms of N_e .
275 Genotype-based N_e estimates ranged from 69 in the Dutch Holstein (2) to roughly 135 in the New
276 Zealand Holstein cattle (65). In contrast to pedigree estimates, where the whole information available
277 on genealogical data is generally used, the genotype-based estimates are highly dependent of the
278 genotyped animals used for the estimation and on the SNP data used. For this reason, the comparison
279 with other studies might be less straightforward. In addition, several methods for the N_e estimations
280 are available based on genotype data (as an example: based on homozygous segments, marker-by-
281 marker homozygosity and similarity and based on linkage disequilibrium) leading to different results.
282 The N_e values found in the Dutch Holstein population, were calculated using several genotype-based
283 estimates (ranging between 69 to 100) (2), and most of them were lower than those found for the Italian
284 Holstein. A possible explanation of this difference might be due to the use of different sexes in the two
285 studies: in the Dutch study, proved bulls were genotyped whereas in this work only females were used.
286 Thus, due to differences in the intensity of selection, we expect to have higher N_e based on female
287 data, compared to proved sires. New Zealand and Australian Holstein cattle (65) showed analogous
288 estimates to the Italian Holstein in the latest generations in our study. The present study and the one by
289 de Roos et al., 2008 were based on LD-based N_e estimates, making the comparison more forthright.
290 Nevertheless, an akin N_e across countries is expected for highly specialized worldwide breeds, due to
291 the extensive exchange of genetic material and similar applied breeding strategies. The Holstein fully
292 represents this scenario as it has been bred in over 150 countries, and it currently dominates commercial
293 dairy production worldwide (1). The overall breeding goal is relatively uniform, with high selection
294 intensities and a small number of sires being used and propagating worldwide (66). In the Italian dairy
295 industry, over 80% of the milk is produced by this breed, which has been used over time for a multitude
296 of purposes, from drinkable milk to highly specialized consortia for PDO cheese production (39). This
297 latter aspect might have caused a peculiar genomic architecture in the Italian Holstein which should be
298 investigated further in comparison with other International Holstein populations. Nevertheless, in the
299 Italian Holstein several bulls from other states are used with roughly 49% of the bulls being local bulls
300 (Italian nationality), followed by United State (~23%) and Canadian (~7%) origins, highlighting that
301 what we found in the Italian Holstein might resemble the status quo of other countries as well.
302 Unfortunately, this implies that, despite the census of millions of animals, the Holstein breed is exposed

303 to the same genetic drift and accumulation of inbreeding as a population of roughly 100 individuals.
304 Surely, the use of a small number of sires propagated worldwide through artificial insemination and
305 other advanced reproductive technologies has boosted this issue (65). Analyses on genetic distance and
306 variability including all Holstein populations at international level are suggested as they provide
307 additional information on the genetic distance among populations which finally can serve to further
308 quantify the genetic pool available in this breed.

309 For the latest generation considered in this work, the N_e was equal to 55 and to 120 animals from
310 pedigree and genotype-based estimates, respectively. The Food and Agriculture Organization of the
311 United Nations (FAO) set as 50 animals the N_e critical value, from which the population is expected
312 to lose fitness and viability in the long term (67). Since the Italian Holstein population seems to be very
313 close to the critical value from the pedigree estimation, appropriate strategies to reduce the loss of
314 genetic diversity are needed to preserve the breed in the long term. The reduction in the N_e from 1960
315 till 1995 found in this study agrees with previous studies in the Canadian, Dutch, French, and US
316 Holstein populations (64,68). Interestingly, in the Canadian Holstein a similar N_e increase from 2000
317 was shown (64). This increase was even more evident in our study when looking at SNP N_e estimates,
318 which showed an increase of about 20 animals in the latest three generations. However, the latter result
319 seems counterintuitive if evaluated with the general increase in inbreeding over time found in the
320 Italian Holstein. A potential reason might be the more diverse offer in terms of selected candidates
321 thanks to GS (69), leading to a reduction in LD over greater recombinant distances. Thanks to GS, we
322 can predict the Mendelian sampling term component, potentially reducing the co-selection of sibs.
323 Thus, we expect a positive effect on population size, which was indeed shown in this study both based
324 on SNP and pedigree estimates. Nevertheless, we suspect that the more visible increase in N_e based on
325 SNP compared to pedigree is due to the distinct methodologies employed based on the source of data
326 available. In the case of N_e based on pedigree, the estimate depends on the quality and completeness
327 of the genealogical data, whereas, in the case of SNP N_e , the estimate relies on the relationship between
328 the LD variance and effective population size. Therefore, we believe that the latter estimate might be
329 more sensible on the detection N_e variation compared to pedigree data.

330 **4.3 Impact of genomics and future breeding strategies**

331 The average GI in the Italian Holstein has decreased since 1985, with two sharp declines from 1985-
332 1990 and from 2009 onwards. The initial reduction in the generation interval might be attributed to the
333 implementation and use of the BLUP evaluation (70). Nevertheless, from 1992 till GS, an almost
334 constant GI was registered which might be due to the general tendency of using proven bull sires. The
335 remarkable reduction in the GI since 2009 might be attributed to GS, as predicted in a previous study
336 (71). The severe drop in the GI found in this study is in line with what we know from the literature.
337 After 2009 in the Canadian Holstein a drop of 38% in the GI was registered as well as in the US
338 Holstein (37%) and in the Dutch Holstein (~35%) (2,4,12). The ΔF per generation also showed some
339 fluctuation throughout the studied period, and it increased from 0.14% to 0.47% based on pedigree,
340 and from 0.32% to 0.70% based on ROH comparing the PTS to the GS period. Similar rate of
341 inbreeding were found in a previous study where over 4,000 Holstein bulls were analyzed, highlighting
342 the effect of genomic selection both on the annual genetic gain increase and on the inbreeding rate (13).
343 The observed ΔF in the Italian Holstein is higher than the 1% per generation suggested by FAO
344 guidelines as critical value for the maintenance of genetic variability in the long term (72).

345 Right now, a web interface is available for breeders to evaluate and manage inbreeding within their
346 herds, where advice on mating strategies is provided by specialists from the breeding association.
347 However, the routinely use of this tool by breeders is still not fully known and likely limited. Thus,
348 activities to raise awareness on inbreeding control among breeders are highly advised to prevent any

Genetic diversity in the Italian Holstein

349 further loss of genetic variability. This latter aspect is extremely important if evaluated together with
350 the current and coming expansion of international demand for dairy products due to emerging
351 economies, the need for high quality milk protein in developing countries and world population
352 expansion (73). Since from this study we revealed threatening level of loss of genetic diversity in the
353 Italian Holstein, which might cause inbreeding depression, we highly advice to optimize selection
354 programs to ensure long-term productivity of this breed. So far, in the dairy sector, the management of
355 genetic diversity has been kept within individual herds where ultimately breeding decisions are made.
356 Unfortunately, this is a drawback for the implementation of one of the most effective methods to
357 manage genetic variability which is the optimal contribution selection (OCS) (74). Nevertheless,
358 thanks to the genomic era, the dairy genetic industry is slowly changing, moving toward scenarios
359 where tighter control of the population as a whole is becoming possible (i.e., vertically integrated
360 industries) (75). Thus, the implementation of genomic OCS is highly recommended to prevent any
361 further loss of genetic variability in this breed. Finally, the implementation of OCS in the Italian
362 Holstein breed should specifically use inbreeding estimates based on genotype data, as we have shown
363 they can provide a more precise evaluation of the available variability.

364 **5 Conclusion**

365 The presented study in the Italian Holstein cattle showed the urgent matter of controlling the loss of
366 genetic diversity in a highly specialized breed, and that this loss is not enclosed to small and local
367 breeds only. The implementation of genomic OCS is thus highly advised to prevent any further loss of
368 genetic variability. The preservation of genetic resources is key to ensure the long-term sustainability
369 of this breed which represents one of the most important players in the dairy chain as well as to
370 guarantee future market demands.
371

372 **5.1 Tables**373 **Table 1. Inbreeding based on ROH divided by six length classes per each of the four birth year**
374 **cohorts.**

Inbreeding coefficients based on ROH (F_{ROH})												
Length Class (Mbp)	2002-2005 (n. 248) ¹			2006-2010 (n. 3,883) ¹			2011-2015 (n. 30,156) ¹			2016-2020 (n. 40,198) ¹		
	Mean	SD	N. ²	Mean	SD	N. ²	Mean	SD	N. ²	Mean	SD	N. ²
1> ROH ≤2	0.04	0.01	69.4	0.04	0.01	68.7	0.04	0.01	69.8	0.04	0.01	70.4
2> ROH ≤4	0.02	0.01	20.1	0.02	0.01	20.9	0.03	0.01	22.8	0.03	0.01	24.0
4> ROH ≤8	0.03	0.01	11.4	0.03	0.01	12.6	0.03	0.01	14.1	0.04	0.01	15.6
8> ROH ≤16	0.03	0.01	6.10	0.04	0.01	6.98	0.04	0.01	7.88	0.04	0.01	9.07
16> ROH ≤32	0.02	0.01	2.12	0.03	0.02	2.67	0.03	0.02	2.96	0.03	0.02	3.38
>32	0.02	0.01	0.37	0.02	0.01	0.47	0.02	0.01	0.52	0.02	0.01	0.60
Total	0.15	0.03	109.5	0.17	0.03	112.5	0.18	0.03	118.1	0.19	0.03	123.2

375 ¹n.=number of animals per each birth year cohort, ²N.=average number of ROH per animal376 **Table 2. Parameters used to estimate the differences in the inbreeding trend based on pedigree**
377 **and genotype data between the progeny testing (PTS) and genomic selection (GS) periods.**

Parameter	F_{PED}	F_{ROH}
$b_{PTS} (\pm SE)^a$	0.14% (2.0×10^{-04})	0.32% (4.5×10^{-04})
$b_{GS} (\pm SE)^b$	0.47% (7.7×10^{-05})	0.70% (1.4×10^{-04})
δ^c	0.0033	0.0031
p-value of δ	<0.0001	<0.0001
RC ^d	2.36	1.19

378 ^a b_{PTS} is the slope in percentage of regression per F_{PED} and F_{ROH} for cows born between 2006-2010 (progeny testing
379 selection); ^b b_{GS} is the slope in percentage of regression per F_{PED} and F_{ROH} for cows born between 2015-2019 (genomic
380 selection); ^c δ is the difference between the slopes of regression of each inbreeding measurement depending on the two 5-
381 year birth class; ^dRC is the relative change equals to $\frac{\delta}{\beta_1}$ with β_1 the slope of the second evaluated period.

382 **5.2 Figures caption**

Genetic diversity in the Italian Holstein

383 Figure 1. Violin plots of the inbreeding coefficients in the Italian Holstein cows. On the left side the
384 inbreeding based on pedigree (F_{ped}) and on the right the inbreeding based on ROH data (F_{ROH}). Black
385 horizontal line within the boxplot represents the median. Extreme values (above and below the mean
386 ± 3 SD) are presented in magenta.

387 Figure 2. a) Pearson correlation between F_{ped} and F_{ROH} is shown in grey color and the complete
388 generation equivalent (CGE) from 2002 to 2020 in Italian Holstein cows is shown in light blue color,
389 2 b) Pearson correlation between F_{ped} and F_{ROH} dividing the sample in two subgroups based on CGE
390 ($CGE \leq 10$ on the left side, and $CGE > 10$ on the right side). The Pearson correlation is shown above
391 diagonal, the scatterplot below the diagonal and the density plots of inbreeding coefficients measured
392 by ROH (F_{ROH}) and pedigree data (F_{ped}) in the Italian Holstein dairy cows are shown on the
393 diagonal.

394 Figure 3. The effective population size based on SNP data calculated in the SNeP software from
395 generation 30th using on the left a) defaults settings, recombination rate according to Sved and
396 Feldman (1973) and occurrence of mutation at 2.2; on the right b) a restriction on maximum distance
397 to calculate linkage disequilibrium, allowing estimations in the recent generations.

398 Figure 4. Generation Interval (GI) in years from 1960 to 2018 for the four pathways of selection: sire
399 of bulls, sire of cows, dam of bulls and dam of cows is sh.

400 Figure 5. Inbreeding estimates from F_{PED} , F_{ROH} and the generation interval between pre (called
401 “Progeny test selection” period – from 2006-2010) and post the introduction of genomic selection
402 (GS) (called “Progeny and Genomic Selection – from 2015-2019) are shown.

403

404 6 Supplementary Material

405 Supplementary Figure S1. Pearson correlation (above diagonal), scatterplot (below diagonal) and
406 density (diagonal) of inbreeding coefficients measured by ROH (F_{ROH}) and pedigree data (F_{ped}) in the
407 Italian Holstein dairy cows.

408 Supplementary Figure S2. Effective population size based on pedigree and SNP data (using the
409 second method applied in the study for the Ne calculation) from 1960 till 2020.

410

411 7 Conflict of Interest

412 *The authors declare that the research was conducted in the absence of any commercial or financial*
413 *relationships that could be construed as a potential conflict of interest.*

414 8 Author Contributions

415 A.Su. C.C.G. and M.A. conceived the idea and formulated the objectives of this study. C.C.G., J-
416 T.K. and G.S. helped in data preparation. M.A. conducted the analysis and wrote the first draft of the
417 paper. A.Su. and A.Sa. supervised the project. C.D. and G.S. contributed on the data visualization.
418 A.Su. C.C.G., A.Sa, G.S., R.F., C.D. critically reviewed the text. All authors read and approved the
419 final manuscript.

420 **9 Reference styles**

- 421 1. Buchanan DS. “Breeds of Dairy Cattle (Major *Bos taurus* Breeds),” in *Reference Module in*
422 *Food Science* (Elsevier), 1–9. doi:10.1016/B978-0-08-100596-5.00624-7
- 423 2. Doekes HP, Veerkamp RF, Bijma P, Hiemstra SJ, Windig JJ. Trends in genome-wide and
424 region-specific genetic diversity in the Dutch-Flemish Holstein–Friesian breeding program
425 from 1986 to 2015. *Genet Sel Evol* (2018) **50**:15. doi:10.1186/s12711-018-0385-y
- 426 3. Forutan M, Ansari Mahyari S, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and
427 runs of homozygosity before and after genomic selection in North American Holstein cattle.
428 *BMC Genomics* (2018) **19**:98. doi:10.1186/s12864-018-4453-z
- 429 4. Makanjuola BO, Miglior F, Abdalla EA, Maltecca C, Schenkel FS, Baes CF. Effect of
430 genomic selection on rate of inbreeding and coancestry and effective population size of
431 Holstein and Jersey cattle populations. *J Dairy Sci* (2020) **103**:5183–5199.
432 doi:10.3168/jds.2019-18013
- 433 5. Hazel LN, Lush JL. The efficiency of three methods of selection. *J Hered* (1942) **33**:393–399.
434 doi:10.1093/oxfordjournals.jhered.a105102
- 435 6. Henderson CR. Best Linear Unbiased Estimation and Prediction under a Selection Model.
436 *Biometrics* (1975) **31**:423. doi:10.2307/2529430
- 437 7. Brotherstone S, Goddard M. Artificial selection and maintenance of genetic variance in the
438 global dairy cow population. *Philos Trans R Soc B Biol Sci* (2005) **360**:1479–1488.
439 doi:10.1098/rstb.2005.1668
- 440 8. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-
441 wide dense marker maps. *Genetics* (2001) **157**:1819–1829. doi:10.1093/genetics/157.4.1819
- 442 9. Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United
443 States: Past, present, future. *J Dairy Sci* (2011) **94**:3202–3211. doi:10.3168/jds.2010-3866
- 444 10. Lund MS, de Roos AP, de Vries AG, Druet T, Ducrocq V, Fritz S, Guillaume F,
445 Guldbrandsen B, Liu Z, Reents R, et al. A common reference population from four European
446 Holstein populations increases reliability of genomic predictions. *Genet Sel Evol* (2011) **43**:43.
447 doi:10.1186/1297-9686-43-43
- 448 11. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in
449 dairy cattle: Progress and challenges. *J Dairy Sci* (2009) **92**:433–443. doi:10.3168/jds.2008-
450 1646
- 451 12. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP.
452 Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle
453 as a result of genomic selection. *Proc Natl Acad Sci* (2016) **113**:E3995–E4004.
454 doi:10.1073/pnas.1519061113
- 455 13. Doublet AC, Croiseau P, Fritz S, Michenet A, Hozé C, Danchin-Burge C, Laloë D, Restoux G.
456 The impact of genomic selection on genetic diversity and genetic gain in three French dairy

Genetic diversity in the Italian Holstein

- 457 cattle breeds. *Genet Sel Evol* (2019) **51**:1–13. doi:10.1186/s12711-019-0495-1
- 458 14. Rodríguez-Ramilo ST, Fernández J, Toro MA, Hernández D, Villanueva B. Genome-Wide
459 estimates of coancestry, inbreeding and effective population size in the spanish holstein
460 population. *PLoS One* (2015) **10**:1–11. doi:10.1371/journal.pone.0124157
- 461 15. Miglior F, Beavers L. Genetic diversity and inbreeding: Before and after genomics.
462 Progressive Dairyman. (2014) Available at: [https://www.progressivedairy.com/topics/a-i-](https://www.progressivedairy.com/topics/a-i-breeding/genetic-diversity-and-inbreeding-before-and-after-genomics)
463 [breeding/genetic-diversity-and-inbreeding-before-and-after-genomics](https://www.progressivedairy.com/topics/a-i-breeding/genetic-diversity-and-inbreeding-before-and-after-genomics) [Accessed July 28,
464 2021]
- 465 16. Rojas-Downing MM, Nejadhashemi AP, Harrigan T, Woznicki SA. Climate change and
466 livestock: Impacts, adaptation, and mitigation. *Clim Risk Manag* (2017) **16**:145–163.
467 doi:10.1016/j.crm.2017.02.001
- 468 17. Wright S. Coefficients of Inbreeding and Relationship. *Am Nat* (1922) **56**:330–338.
469 doi:10.1086/279872
- 470 18. Mariani E, Summer A, Ablondi M, Sabbioni A. Genetic variability and management in nero di
471 parma swine breed to preserve local diversity. *Animals* (2020) **10**: doi:10.3390/ani10030538
- 472 19. Ablondi M, Vasini M, Beretti V, Superchi P, Sabbioni A. Exploring genetic diversity in an
473 Italian horse native breed to develop strategies for preservation and management. *J Anim*
474 *Breed Genet* (2018) **135**:450–459. doi:10.1111/jbg.12357
- 475 20. Fabbri MC, de Rezende MPG, Dadousis C, Biffani S, Negrini R, Carneiro PLS, Bozzi R.
476 Population structure and genetic diversity of Italian beef breeds as a tool for planning
477 conservation and selection strategies. *Animals* (2019) **9**:880. doi:10.3390/ani9110880
- 478 21. Cavani L, Silva RM de O, Carreño LOD, Ono RK, Bertipaglia TS, Farah MM, Millen DD,
479 Fonseca R da. Genetic diversity of Brazilian Brahman cattle by pedigree analysis. *Pesqui*
480 *Agropecuária Bras* (2018) **53**:74–79. doi:10.1590/s0100-204x2018000100008
- 481 22. Schurink A, Arts DJG, Ducro BJ. Genetic diversity in the Dutch harness horse population
482 using pedigree analysis. *Livest Sci* (2012) **143**:270–277. doi:10.1016/j.livsci.2011.10.005
- 483 23. Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: Inbreeding in the genomics era:
484 Inbreeding, inbreeding depression, and management of genomic variability. *J Dairy Sci* (2017)
485 **100**:6009–6024. doi:10.3168/jds.2017-12787
- 486 24. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity:
487 current knowledge and applications in livestock. *Anim Genet* (2017) **48**:255–271.
488 doi:10.1111/age.12526
- 489 25. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows
490 into population history and trait architecture. *Nat Rev Genet* (2018) **19**:220–234.
491 doi:10.1038/nrg.2017.109
- 492 26. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: A possible solution
493 to an old problem. *Livest Sci* (2014) **166**:26–34. doi:10.1016/j.livsci.2014.05.034

- 494 27. Makanjuola BO, Maltecca C, Miglior F, Schenkel FS, Baes CF. Effect of recent and ancient
495 inbreeding on production and fertility traits in Canadian Holsteins. *BMC Genomics* (2020)
496 **21**:1–15. doi:10.1186/s12864-020-07031-w
- 497 28. Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, Liu G, Da Y,
498 Sonstegard TS. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein
499 Cattle. *PLoS One* (2013) **8**:e80813. doi:10.1371/journal.pone.0080813
- 500 29. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population
501 history in cattle. *BMC Genet* (2012) **13**:70. doi:10.1186/1471-2156-13-70
- 502 30. Grilz-Seger G, Mesarič M, Cotman M, Neuditschko M, Druml T, Brem G. Runs of
503 Homozygosity and Population History of Three Horse Breeds With Small Population Size. *J*
504 *Equine Vet Sci* (2018) **71**:27–34. doi:10.1016/j.jevs.2018.09.004
- 505 31. Sadeghi R, Moradi-Shahrbabak M, Ashtiani SRM, Schlamp F, Cosgrove EJ, Antczak DF.
506 Genetic diversity of Persian arabian horses and their relationship to other native iranian horse
507 breeds. *J Hered* (2019) **110**:173–182. doi:10.1093/jhered/esy061
- 508 32. Mancin E, Ablondi M, Mantovani R, Pigozzi G, Sabbioni A, Sartori C. Genetic Variability in
509 the Italian Heavy Draught Horse from Pedigree Data and Genomic Information. *Animals*
510 (2020) **10**:1310. doi:10.3390/ani10081310
- 511 33. Ablondi M, Dadousis C, Vasini M, Eriksson S, Mikko S, Sabbioni A. Genetic diversity and
512 signatures of selection in a native italian horse breed based on SNP data. *Animals* (2020) **10**:1–
513 15. doi:10.3390/ani10061005
- 514 34. Schiavo G, Bovo S, Bertolini F, Tinarelli S, Dall'Olio S, Nanni Costa L, Gallo M, Fontanesi
515 L. Comparative evaluation of genomic inbreeding parameters in seven commercial and
516 autochthonous pig breeds. *animal* (2020) **14**:910–920. doi:10.1017/S175173111900332X
- 517 35. Addo S, Klingel S, Thaller G, Hinrichs D. Genetic diversity and the application of runs of
518 homozygosity-based methods for inbreeding estimation in German White-headed Mutton
519 sheep. *PLoS One* (2021) **16**:e0250608. doi:10.1371/journal.pone.0250608
- 520 36. Rodríguez-Ramilo ST, Elsen JM, Legarra A. Inbreeding and effective population size in
521 French dairy sheep: Comparison between genomic and pedigree estimates. *J Dairy Sci* (2019)
522 **102**:4227–4237. doi:10.3168/jds.2018-15405
- 523 37. Dadousis C, Cecchi F, Ablondi M, Fabbri MC, Stella A, Bozzi R. Keep Garfagnina alive. An
524 integrated study on patterns of homozygosity, genomic inbreeding, admixture and breed
525 traceability of the Italian Garfagnina goat breed. *PLoS One* (2021) **16**:e0232436.
526 doi:10.1371/journal.pone.0232436
- 527 38. Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, Drögemüller
528 C, Flury C. Genetic diversity analyses reveal first insights into breed-specific selection
529 signatures within Swiss goat breeds. *Anim Genet* (2016) **47**:727–739. doi:10.1111/age.12476
- 530 39. Allevatori AIA. Italy: milk recording activity - Official Statistics - Year 2020. Roma (2020).
531 Available at: http://bollettino.aia.it/Contenuti.aspx?CD_GruppoStampe=RS&CD_Specie=C4

Genetic diversity in the Italian Holstein

- 532 40. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams JL,
533 Macciotta NPP. Analysis of runs of homozygosity and their relationship with inbreeding in
534 five cattle breeds farmed in Italy. *Anim Genet* (2015) **46**:110–121. doi:10.1111/age.12259
- 535 41. Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L, Portolano B.
536 Genome-wide identification of runs of homozygosity islands and associated genes in local
537 dairy cattle breeds. *Animal* (2018) **12**:2480–2488. doi:10.1017/S1751731118000629
- 538 42. Ablondi M, Malacarne M, Cipolat-Gotet C, van Kaam J-T, Sabbioni A, Summer A. Genome-
539 wide scan reveals genetic divergence in Italian Holstein cows bred within PDO cheese
540 production chains. *Sci Rep* (2021) **11**:12601. doi:10.1038/s41598-021-92168-1
- 541 43. Wellmann R. Optimum contribution selection for animal breeding and conservation: the R
542 package optiSel. *BMC Bioinformatics* (2019) **20**:25. doi:10.1186/s12859-018-2450-5
- 543 44. R Development Core Team. R: A Language and Environment for Statistical Computing.
544 (2011)
- 545 45. Nicolazzi EL, Biffani S, Jansen G. Short communication: Imputing genotypes using
546 PedImpute fast algorithm combining pedigree and population information. *J Dairy Sci* (2013)
547 **96**:2649–2653. doi:10.3168/jds.2012-6062
- 548 46. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MARR, Bender D, Maller J, Sklar P,
549 de Bakker PIWW, Daly MJ, et al. PLINK: A tool set for whole-genome association and
550 population-based linkage analyses. *Am J Hum Genet* (2007) **81**:559–575. doi:10.1086/519795
- 551 47. Wright S. *Evolution and the genetics of populations. Volume 4: variability within and among*
552 *natural populations*. University of Chicago Press, Chicago, IL, USA. (1978).
- 553 48. Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: an R package to detect runs of
554 homozygosity and heterozygosity in diploid genomes. (2019) Available at: [https://cran.r-](https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references)
555 [project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references](https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references)
- 556 49. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L,
557 Smolej-Narancic N, Janicijevic B, Polasek O, Tenesa A, et al. Runs of Homozygosity in
558 European Populations. *Am J Hum Genet* (2008) **83**:359–372. doi:10.1016/j.ajhg.2008.08.007
- 559 50. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in
560 recent effective population size trajectories using genome-wide SNP data. *Front Genet* (2015)
561 **6**: doi:10.3389/fgene.2015.00109
- 562 51. Sved JA, Feldman MW. Correlation and probability methods for one and two loci. *Theor*
563 *Popul Biol* (1973) **4**:129–132. doi:10.1016/0040-5809(73)90008-7
- 564 52. Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population
565 size using linkage disequilibria with marker data. *J Anim Breed Genet* (2012) **129**:257–270.
566 doi:10.1111/j.1439-0388.2012.01003.x
- 567 53. Barbato M, Hailer F, Upadhyay M, Del Corvo M, Colli L, Negrini R, Kim E-S, Crooijmans
568 RPMA, Sonstegard T, Ajmone-Marsan P. Adaptive introgression from indicine cattle into

- 569 white cattle breeds from Central Italy. *Sci Rep* (2020) **10**:1279. doi:10.1038/s41598-020-
570 57880-4
- 571 54. Boichard D. PEDIG : A FORTRAN PACKAGE FOR PEDIGREE ANALYSIS SUITED FOR
572 LARGE POPULATIONS. in *7th World Congress on Genetics Applied to Livestock*
573 *Production, August 19-23, 2002, Montpellier, France*, 13–28.
- 574 55. Wiggans GR, Cole JB, Hubbard SM, Sonstegard TS. Genomic Selection in Dairy Cattle: The
575 USDA Experience. *Annu Rev Anim Biosci* (2017) **5**:309–327. doi:10.1146/annurev-animal-
576 021815-111422
- 577 56. Meuwissen T, Luo Z. Computing inbreeding coefficients in large populations. *Genet Sel Evol*
578 (1992) **24**:305. doi:10.1186/1297-9686-24-4-305
- 579 57. Oliehoek PA, Bijma P. Effects of pedigree errors on the efficiency of conservation decisions.
580 *Genet Sel Evol* (2009) **41**:9. doi:10.1186/1297-9686-41-9
- 581 58. Doekes HP, Veerkamp RF, Bijma P, de Jong G, Hiemstra SJ, Windig JJ. Inbreeding
582 depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genet*
583 *Sel Evol* (2019) **51**:54. doi:10.1186/s12711-019-0497-z
- 584 59. Doekes HP, Veerkamp RF, Bijma P, Hiemstra SJ, Windig J. Value of the Dutch Holstein
585 Friesian germplasm collection to increase genetic variability and improve genetic merit. *J*
586 *Dairy Sci* (2018) **101**:10022–10033. doi:10.3168/jds.2018-15217
- 587 60. Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, Sölkner J.
588 Estimates of autozygosity derived from runs of homozygosity: empirical evidence from
589 selected cattle populations. *J Anim Breed Genet* (2013) **130**:286–293. doi:10.1111/jbg.12012
- 590 61. Baes CF, Makanjuola BO, Miglior F, Marras G, Howard JT, Fleming A, Maltecca C.
591 Symposium review: The genomic architecture of inbreeding: How homozygosity affects
592 health and performance. *J Dairy Sci* (2019) **102**:2807–2817. doi:10.3168/jds.2018-15520
- 593 62. Hill WG. Understanding and using quantitative genetic variation. *Philos Trans R Soc B Biol*
594 *Sci* (2010) **365**:73–85. doi:10.1098/rstb.2009.0203
- 595 63. Weigel KA. Controlling Inbreeding in Modern Breeding Programs. *J Dairy Sci* (2001)
596 **84**:E177–E184. doi:10.3168/jds.S0022-0302(01)70213-5
- 597 64. Stachowicz K, Sargolzaei M, Miglior F, Schenkel FS. Rates of inbreeding and genetic
598 diversity in Canadian Holstein and Jersey cattle. *J Dairy Sci* (2011) **94**:5160–5175.
599 doi:10.3168/jds.2010-3308
- 600 65. de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage Disequilibrium and Persistence
601 of Phase in Holstein–Friesian, Jersey and Angus Cattle. *Genetics* (2008) **179**:1503–1512.
602 doi:10.1534/genetics.107.084301
- 603 66. Powell RL, Norman HD, Sanders AH. Progeny testing and selection intensity for holstein
604 bulls in different countries. *J Dairy Sci* (2003) **86**:3386–3393. doi:10.3168/jds.S0022-
605 0302(03)73942-3

Genetic diversity in the Italian Holstein

- 606 67. Sherf B. *The Second Report on the State of the World's Animal Genetic Resources for Food*
607 *and Agriculture*. FAO (2015). doi:10.4060/I4787E
- 608 68. Danchin-Burge C, Hiemstra SJ, Blackburn H. Ex situ conservation of Holstein-Friesian cattle:
609 Comparing the Dutch, French, and US germplasm collections. *J Dairy Sci* (2011) **94**:4100–
610 4108. doi:10.3168/jds.2010-3957
- 611 69. Daetwyler HD, Villanueva B, Bijma P, Woolliams JA. Inbreeding in genome-wide selection. *J*
612 *Anim Breed Genet* (2007) **124**:369–376. doi:10.1111/j.1439-0388.2007.00693.x
- 613 70. Gianola D, Rosa GJM. One hundred years of statistical developments in animal breeding.
614 *Annu Rev Anim Biosci* (2015) **3**:19–56. doi:10.1146/annurev-animal-022114-110733
- 615 71. Schaeffer LR. Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed*
616 *Genet* (2006) **123**:218–223. doi:10.1111/j.1439-0388.2006.00595.x
- 617 72. Scherf BD, Pilling D. The second report on the state of the world's animal genetic resources
618 for food and agriculture. Rome (2015). Available at: <http://www.fao.org/3/a-i4787e/index.html>
- 619 73. OECD-FAO Agricultural Outlook 2019-2028. OECD (2019). doi:10.1787/agr_outlook-2019-
620 en
- 621 74. Meuwissen TH, Sonesson AK. Maximizing the response of selection with a predefined rate of
622 inbreeding: overlapping generations. *J Anim Sci* (1998) **76**:2575. doi:10.2527/1998.76102575x
- 623 75. Maltecca C, Tiezzi F, Cole JB, Baes C. Symposium review: Exploiting homozygosity in the
624 era of genomics—Selection, inbreeding, and mating programs. *J Dairy Sci* (2020) **103**:5302–
625 5313. doi:10.3168/jds.2019-17846

626

627 **10 Data Availability Statement**

628 Data supporting this paper were obtained from ANAFIBJ. The genotype data are available only upon
629 agreement with ANAFIBJ.