1   **Towards deciphering the structure of long homozygous stretches in cattle genome**

2   **Michael Grigorievich Smaragdov**

3

4   Russian Research Institute of Farm Animal Genetics and Breeding - Branch

5   of the l.K. Ernst Federal Science Center for Animal Husbandry, St. Petersburg,

6   Pushkin, Russian Federation. E-mail: spbvniigen@mail.ru

7   Corresponding author: M.G. Smaragdov   e-mail: mik7252@yandex.ru

8

9

10

11

12

13

14

15

16

17

18

19

Abstract

Up-to-day there is no as universally accepted software tool and threshold parameters to identify runs of homozygosity (**ROH**). The relative position of POH segments in the cattle genome has not been studied extensively. Specific objective of this study was to evaluate the effect of allowed missing and heterozygous SNPs in ROH on their number, on the estimate of inbreeding level, and on structure of ROH segments in the cattle genome. In this study 371 Holsteinized cows from six herds were genotyped with BovineSNP50 array. To identify ROH, the consecutive and sliding runs were carried out with detectRUNS and Plink tools. Neither effect was shown for missing SNPs genotype calls. Allowing even one heterozygous SNP resulted in significant bias of ROH data. Furthermore, the sliding runs identified less ROH than consecutive runs. The mean coefficient of inbreeding across herds was 0.111 ± 0.003 and 0.104 ± 0.003 based on consecutive and sliding runs respectively. It was shown how, using the heterozygous SNPs in ROH, may be possible to derive a distribution of ROH segments in the cow genome. We suggested it was similar to normal distribution. Furthermore, frequency of ROH in the chromosomes did not depend on their length. Of 29 chromosomes, the most abundant with ROH were BTA 14, BTA 7, and BTA 18. The result of this study confirmed more accurately identification of ROH with consecutive runs, uneven their distribution in the cattle genome, significant bias of the data due to allowing heterozygous SNPs in ROH.

Background

Inbreeding in dairy cattle is an inevitable phenomenon of artificial selection. Traditionally coefficient of inbreeding is calculated based on ancestry [1]. With the advent of SNPs arrays become possible to investigate autozygosity at a previously unattainable level [2]. In fact, due

2

43   to runs of homozygosity (**ROH**) approach the analysis of the animal genome for extended

44   homozygous SNPs stretches is still ongoing. The primary cause of autozygosity in livestock

45   measured with ROH is assumed to be inbreeding [2] or consanguineous marriage in humans

46   [3]. For identifying of ROH, the softwares based on either identity by descent (IBD)

47   GERMLINE [4], or Hidden Markov Model (HMM) Beagle [5] and BCFtools [6] were

48   elaborated. An addition, the software based on scanning with SNPs window Plink [7],

49   overlapping sliding window SNP101[8], or both consecutive and sliding runs detectRUNS

50   [9] as well software based on other scripts [10,11], cgaTON [12] have been provided. There

51   is the commercial software SNP & Vriation Suite (Golden Helix SNP & Variation Suite). It

52   was shown that software based on HMM and IBD occurred to be inferior to other software

53   mentioned above [13].The major difficulty faced by scientists is the lack of consistent criteria

54   among studies regarding the threshold values in each parameters analyzed to define the ROH

55   [2]. The most crucial parameters to be used in software are the number of allowed

56   heterozygous or missing SNPs calls in ROH. There is inconsistency among thresholds to be

57   applied in published studies. Some authors disallowed any number of heterozygous SNPs in

58   ROH [14-17], other allowed one, two and more heterozygous SNPs depending on the length

59   of ROH segments [18-25]. Anyway, Ferencakovic et al. [22] suggested that allowing some

60   number of genotype errors in long ROH may minimize underestimation of those segments.

61   Though, Mastrangelo et al. [24] showed small increasing in inbreeding coefficient, if

62   heterozygous genotypes were allowed.

63   There are relatively a few studies assessing what a set of these parameters is optimal to detect

64   ROH, so as to better understand their effect on identified autozygosity. Ferencakovic et al.

65   [22] have shown that SNP array density and genotyping errors introduce patterns of bias in

66   the estimation of autozygosity. These authors observed that allowance of the heterozygous

67   SNPs in ROH may has led to merging of adjacent ROH segments that resulted in biased

68   estimates of the ROH number. Based on simulation data Howrigan et al. [13] recommended

69   not allowing any heterozygous SNPs existence in a called ROH. Summarizing, there is

70   currently no generally accepted opinion about the reasonable number of heterozygous SNPs

71   in ROH to avoid bias in the ROH data.

72   The Leningrad region is the highest average milk yield producing region in Russia, with

73   approximately 60,000 cows of Holsteinized Black and White cattle. Dutch, Danish, and

74   Swedish Black and White bulls and heifers were imported to Russia during the 1930s. The

75   Black and White breed was officially registered in Russia in 1959. To improve milk traits of

76   Black and White cattle, local farmers started to use imported from USA (since 1978) and The

77   Netherlands (since 2002) Holstein bulls and semen. Currently, the commercial Russian Black

78   and White cattle population can be considered as Holstein due to long-term crossing only

79   with Holstein bulls.

80   Main objectives of this study were: (i) to estimate the number of ROH segments in the cow

81   genome with consecutive and sliding runs, (ii) to assess their dependence on allowed number

82   of heterozygous and missing SNPs calls in ROH, (iii) to quantify the coefficient of

83   inbreeding of the cows in the herds studied, (iv) to compare the results obtained by these two

84   approaches (v) to elucidate the structure and distribution of ROH in cow genome.

Materials and Methods

Ethics statement

87   Th*is* study does not involve any endangered or protected species. Samples were derived from

88   the Committee on Agro-Industrial Complex of Leningrad region. The permission to carry out

89   the sampling at each farm was obtained directly from the owners. All the samples were

90   collected during routine veterinary checks and in accordance with local/national regulations

91   and ethical rules in force at the time of sampling. The study was approved by the Ethical &

92   Animal care committee of the RRIFAGB — Branch of the L. K. Ernst Federal Science

93   Centre for Animal Husbandry. Animals were handled with respect to the Russian Federal

94   Law No. 498-FZ on Responsible Treatment of Animals and on Amendments to Certain

95   Legislative Acts of the Russian Federation.

Animal resources and SNPs genotyping

97   Data and genotypes were obtained from Committee on Agro-Industrial Complex of

98   Leningrad region. Cows genotypes were available from six breeding herds located in

99   Leningrad region (Russia) born in range from 2010 to 2013. Animals for genotyping were

100   selected by farmers with regarding the pedigree structure of the herd. Number of animals

101   genotyped depends on number of sires used in herd at least one daughter per sire. In case of

102   multiple daughters per single sire, the sires of dams were different. Sampled animals included

103   8–15% of total number of milking cows in herds. Altogether, 371 cows were genotyped with

104   BovineSNP50 v.2.0 array (Illumina Ca. USA). In the first quality control step, SNPs with

105   quality score less than 0.7 were removed. In the second step, genotyping quality control was

106   done with PLINK 1.9 [7]. Only autosomal chromosomes were considered. In data set missing

107   rate per SNP was no more than 5% and probability of deviation from Hardy-Weinberg

108   equilibrium (HWE) was less than 1.0E-03. SNPs with MAF < 0.01 were removed resulted in

109   48,108 SNPs. Total genotyping rate was > 0.99.

Detection the runs of homozygosity

111   The ROH segments were identified using detectRUNS [9] implemented in R environment

112   [26], and Plink tool [7]. The parameters and threshold applied to define ROH with

113   detectRUNS for consecutive runs method were: (i) minimal number of SNPs needed to

114   define segment as a ROH: 15 and 20, (ii) number of missing calls allowed in a ROH segment:

115   0 - 4, (iii) number of heterozygous calls allowed in a ROH segment: 0 - 240, (iv) minimal

116   length of ROH segments: 250 Kb, (v) maximal gap between ROH segments: 1 Mb.

117   For sliding window method the parameters and thresholds were: (i) window size: 15 SNPs

118   and 20 SNPs, (ii) threshold: 0.05, (iii) minimal number of SNPs needed to define segment as

119   a ROH: 15 and 20, (iv) number of missing calls allowed in a ROH segment: 0 - 4, (v)

120   number of heterozygous calls allowed in a ROH segment: 0 - 16, (vi) minimal length of

121   ROH: 250 Kb, (vii) maximal gap between ROH segments: 1 Mb, (viii) minimal allowed

122   density of SNPs: 1 SNP per 1 Mb.

123   The parameters and threshold applied to define ROH with Plink were: (i) sliding window: 20

124   SNPs, (ii) the proportion of homozygous overlapping windows: 0.05, (iii) the minimum

125   number of SNPs in ROH: 20, (iv) density was one SNP per 50 Kb, (v) the number of missing

126   SNPs: zero, (vi) the number of heterozygous SNPs: zero.

127   Inbreeding coefficient was calculated as the sum of ROH length of an animal divided by the

128   total length of the autosomes covered by SNPs (2508.706681 Mb). A probability (quantile-

129   quantile (Q-Q) plot was performed in R environment [26].

130   Results

131   In this section the effect of the number of heterozygous and missing SNPs on ROH data was

132   evaluated by consecutive and sliding runs. Primarily, the effect of missing SNPs allowed in

133   ROH on the data was evaluated by consecutive and sliding runs. Both methods were not

134   show any effect on ROH data when from one to four missing SNPs calls were allowed in

135   ROH. Therefore, to evaluate the ROH findings further, this value has been fixed to zero.

136   Effect of heterozygous SNPs on ROH data

137   ROH identification based on consecutive runs

138   To detect the number of ROH segments in the cow genome, 15 SNPs (S1 Table) and 20

139   SNPs (Table 1) consecutive runs across cows genome were carried out. When ROH segments

140   were not interrupted with heterozygous SNPs the mean number of ROH was greater in 1.9

141   times at 15 SNPs runs. In fact, the average number of ROH across herds was 182.0 ± 2.8 at

142   15 SNPs runs, as compared with 94.4 ± 2.7 at 20 SNPs runs. To avoid an autozygous ROH

143   overestimation due to short ROH segments, the 20 SNPs runs were used further.

144   **Table 1. Estimated mean (± SE) ROH number in the herds based on 20 SNPs**

145   **consecutive runs**

Herd	1	2	3	4	5	6
0*						
The mean number of ROH	99.6 ± 6.5	90.5 ± 1.0	91.5 ± 1.4	106.4 ± 14.7	93.0 ± 1.3	91.4 ± 1.5
Maximum	360	112	148	757	111	125
Minimum	2	66	65	48	75	73
1						
The mean number of ROH	161.9 ± 10.2	145.4 ± 1.3	146.5 ± 1.6	155.8 ± 12.1	149.2 ± 1.5	148.8 ± 1.6
Maximum	565	179	195	692	175	174
Minimum	7	117	109	82	125	126
2						
The mean number of ROH	262.2 ± 13.1	243.1 ± 1.6	244.7 ± 1.8	253.7 ± 7.4	245.5 ± 1.6	248.2 ± 1.8
Maximum	761	277	281	564	270	283
Minimum	21	211	215	145	214	207
4						
The mean number of ROH	583.8 ± 13.8	573.6 ± 2.1	579.3 ± 2.5	582.7 ± 7.4	576.3 ± 2.7	580.3 ± 3.3
Maximum	1017	625	631	655	636	630
Minimum	116	528	536	383	533	525
8						
The mean number of	1161.4 ± 9.6	1183.3 ± 2.9	1194.5 ± 3.3	1182.2 ± 22.1	1182.5 ± 3.3	1192.0 ± 3.7

ROH						
Maximum	1227	1254	1271	1263	1233	1239
Minimum	818	1126	1125	224	1129	1132
16						
The mean number of ROH	848.4 ± 11.8	857.1 ± 2.1	862.5 ± 2.4	852.2 ± 16.8	852.6 ± 2.5	860.3 ± 2.9
Maximum	1297	914	898	1024	887	893
Minimum	531	810	805	132	805	804
20						
The mean number of ROH	696.1 ± 9.7	703.4 ± 1.7	707.7 ± 1.9	699.4 ± 13.8	699.9 ± 1.9	706.3 ± 2.3
Maximum	1069	753	738	840	732	735
Minimum	417	668	659	107	664	663
60						
The mean number of ROH	257.2 ± 3.4	259.6 ± 0.6	261.2 ± 0.7	258.9 ± 4.9	258.2 ± 0.7	260.5 ± 0.8
Maximum	390	275	272	312	260	272
Minimum	160	248	242	49	248	244
100						
The mean number of ROH	163.9 ± 2.1	165.5 ± 0.4	166.3 ± 0.5	164.6 ± 3.0	164.5 ± 0.5	165.8 ± 0.5
Maximum	247	174	176	195	173	173
Minimum	103	156	157	38	155	157
240						
The mean number of ROH	80.9 ± 0.9	81.9 ± 0.2	82.2 ± 0.2	81.3 ± 1.0	81.5 ± 0.3	82.1 ± 0.3
Maximum	114	87	86	93	87	86
Minimum	55	77	78	36	77	77

Note: * - The number of hetrozygous SNPs in ROH

Table 2. Estimated mean (± SE) across herds ROH number in length classes based on 20 SNPs consecutive run

Class (Mb)	0-2	2-4	4-8	6-16	>16
0*	3585 ± 277	1155 ± 9	662 ± 63	319 ± 32	113 ± 17
1	6792 ± 560	1391 ± 103	704 ± 63	344 ± 34	123 ± 17
2	12382 ± 1080	1840 ± 142	750 ± 69	362 ± 31	131 ± 18
4	31023 ± 2857	3552 ± 308	874 ± 76	395 ± 31	139 ± 18
8	59676 ± 5699	11991 ± 1084	1261 ± 122	463 ± 34	161 ± 17
16	16178 ± 1570	31742 ± 3010	4476 ± 387	612 ± 65	205 ± 19

8

20	5049 ± 486	29569 ± 2816	8089 ± 747.6	738 ± 76	229 ± 20
60	249 ± 26	490 ± 46	4618 ± 440	10076 ± 953	688 ± 2
240	62 ± 7	119 ± 14	478 ± 47	528 ± 51	3891 ± 364

150 Note: * - The number of allowed heterozygous SNPs in the ROH segments

151 The effect of allowed heterozygous SNPs on the number of ROH segments was evaluated

152 when their values varied from 0 to 240. Descriptive statistics of the data are given in Table 1.

153 The mean number of ROH varied across herds. Nevertheless, between herds mean values

154 were insignificant when heterozygous SNPs were disallowed. The large variation of the ROH

155 number among the cows in fourth herd was observed. This result was due to large number of

156 ROH at one cow (757 ROH segments). Exclusion this cow resulted in the mean ROH number

157 for forth herd 85.9 ± 2.1. As a result, this herd, using t-test estimated with Bonferroni

158 correction, became significantly differed from fifth herd ($P < 0.05$). Initially, the average

159 number of ROH increased by more than 10 times from 94.4 ± 2.7 up to 1182.7 ± 4.9 when

160 heterozygous SNPs in ROH were allowed from 1 to 8 (Table 1). Then, the mean value

161 decreased to 81.7 ± 0.2 when the number of heterozygous SNPs was increased to 240.

162 The length of ROH segments was classified into five categories (1-2 Mb, 2-4 Mb, 4-8 Mb, 8-

163 16 Mb, and >16 Mb) (Table 2). The most abundant in the number of ROH was 1-2 Mb class.

164 Significant increase in the number of ROH in this class occurred when 15 SNPs runs were

165 used (Table 2 vs. S2 Table). The largest proportion of ROH number had the same class up to

166 eight allowed heterozygous SNPs, and then its proportion begun noticeable to decline (Table

167 2). The same type of the distribution was faithful for 2-4 Mb class but with other values of

168 allowed heterozygous SNPs. Namely, the number of ROH began to increase when two

169 heterozygous SNPs were allowed, then the distribution peaked at 16 heterozygous SNPs and

170 gradually decreased further (Fig 1). This type of distribution was the same for 4-8 Mb class

171 but with a lower maximum and at more number of allowed heterozygous SNPs. For 8-16 Mb

172   class increasing of ROH number occurred when sixty heterozygous SNPs were allowed. The

173   mean length of ROH segments for each class is shown in S3 Table. These values did not

174   vary mach, except segments in >16 Mb class with 240 allowed heterozygous SNPs (S3

175   Table).

ROH identification based on sliding runs

177   As for consecutive runs, 15 SNP window (S4 Table) and 20 SNPs window (Table 3) for

178   sliding runs were used. Interestingly, the data for 15 SNPs runs identified both methods

179   nearly coincided (S1 Table *vs*. S4 Table) while for 20 SNPs runs the data differed

180   significantly in 1.9 times (P < 0.001) (Table 1 *vs*. Table 3). Therefore, for this approach as

181   well as consecutive runs, the shorter window was, the shorter ROH segments were identified.

182   **Table 3. Estimated mean (± SE) ROH number in the herds based on 20 SNPs sliding**

183   **runs**

Herd	1	2	3	4	5	6
0*						
The mean number of ROH	91 ± 6	82 ± 1	83 ± 1	92 ± 14	85 ± 1	83 ± 1
Maximum	336	194	138	731	106	111
Minimum	1	62	59	44	67	64
1						
The mean number of ROH	157 ± 9	142 ± 1	143 ± 1	146 ± 5	147 ± 1	145 ± 1
Maximum	513	174	175	335	175	169
Minimum	5	117	114	82	118	120
2						
The mean number of ROH	268 ± 10	255 ± 1	254 ± 2	256 ± 4	255 ± 2	257 ± 2
Maximum	631	284	287	294	285	288
Minimum	21	225	220	109	237	222
4						
The mean number of ROH	524 ± 8	535 ± 2	540 ± 2	533 ± 12	534 ± 2	542 ± 3
Maximum	576	578	586	608	579	590
Minimum	133	494	484	39	492	502
8						
The mean number	266	271	272	264	270	270

of ROH	± 8	± 1	± 2	± 6	± 2	± 2
Maximum	580	305	314	402	294	302
Minimum	50	246	238	36	241	238
12						
The mean number of ROH	43.3 ± 2.1	41,3 ± 0.2	41.3 ± 0.3	41.5 ± 0.6	40.4 ± 0.2	41.1 ± 0.3
Maximum	161	47	48	63	45	48
Minimum	36	37	37	36	36	37
16						
The mean number of ROH	36	36	36	36	36	36
Maximum	NA	NA	NA	NA	NA	NA
Minimum	NA	NA	NA	NA	NA	NA

184 Note: * - The number of allowed heterozygous SNPs in the ROH segments

185 To have a comparable result to consecutive runs, 20 SNPs window was used further.

186 Descriptive statistic for 20 SNPs sliding data is given in Table 3. Between herds mean

187 number of ROH was insignificant for disallowed heterozygous SNPs in ROH. But, similar to

188 consecutive runs after exclusion the most deviated cow (731 ROH segments) among fourth

189 herd, the mean number ROH became 77.6 ± 2.0 and fourth herd significantly differed from

190 third, fifth and sixth herds (P < 0.02) using t-test estimated with Bonferroni correction. The

191 mean value increased six times from 86.0 ± 2.6 to 534.7 ± 2.5 when four heterozygous SNPs

192 in ROH were allowed, then decreased to 36 when the number of heterozygous SNPs

193 increased to sixteen.

194 The length of ROH segments were classified into the same five categories (1-2 Mb, 2-4 Mb,

195 4-8 Mb, 8-16 Mb, and >16 Mb) as it was done for consecutive runs (Table 4). The most

196 abundant in the number of ROH was 1-2 Mb class as well. This fact indicates that with an

197 increase in the number of allowed heterozygous SNPs, initially the number of short 1-2 Mb

198 ROH had a bursting increase, and only then the number of longer ROH increased (Fig 2).

199 Hence, close location of numerous shorter than 1 Mb ROH segments were identified in cow

200 genome. Like this, starting from two allowed heterozygous SNPs, 4-8 Mb ROH segments

201 began to increase, and starting from four heterozygous SNPs, 8-16 Mb and >16 Mb ROH

202   segments likewise began to increase. Finally, only beyond eight heterozygous SNPs the

203   number of 8-16 Mb ROH segments began to decrease. No doubt, sliding approach led to

204   faster filling in of the cow genome with ROH segments compared to consecutive runs. It is

205   necessary to pay attention to a sizable increase in the average length of ROH segments in the

206   >16 Mb class when the number of allowed heterozygous SNPs exceeded 12 (S5 Table).

207   Table 4. **Estimated mean (± SE) across herds ROH number in length classes based on 20**
208   **SNPs sliding runs**

Classes (Mb)	0-2	2-4	4-8	6-16	>16
0*	3190 ± 241	1048 ± 83	643 ± 61	309 ± 32	111 ± 16
1	6488 ± 554	1419 ± 122	690 ± 66	358 ± 31	139 ± 15
2	12205 ± 1100	2444 ± 220	817 ± 88	385 ± 39	152 ± 14
4	22024 ± 2139	8342 ± 738	2169 ± 208	545 ± 54	185 ± 19
8	2744 ± 267	3333 ± 312	4421 ± 447	3795 ± 379	2476 ± 226
12	23 ± 3	21.7 ± 4.5	281 ± 24	181 ± 19	2071 ± 193
16			249 ± 23	124 ± 12	1866 ± 174

209   Note: *- The number of allowed heterozygous SNPs in ROH

210

ROH identification based on Plink tool

212   Plink software is widely used for ROH identification. Therefore, it was applied to estimate a

213   possible data bias due to different software tools being used. It turn out, that the mean number

214   of ROH obtained with Plink was 74.9 ± 1.9, which was significantly less of those calculated

215   with detectRUNS based on sliding runs ($P < 0.001$). The fact, that Plink identified less ROH

216   segments mainly in 1-2 Mb class than detectRUNS led to a larger mean ROH length in this

12

217 class (S6 Table *vs*. S3 Table and S5 Table). Thus, the data obtained for the shortest ROH

218 length class is highly dependent on the software and parameters used**.**

Comparative analysis of consecutive and sliding runs

220 Comparative analysis of the consecutive and sliding data led to the following conclusions. If

221 heterozygous SNPs were disallowed the consecutive runs identified slightly more ROH than

222 sliding runs (94.4 ± 2.7 *vs*. 86.0 ± 2.8, $P < 0.05$). The shorter SNPs window at sliding runs or

223 minimal number of SNPs needed to define segment at consecutive runs were, the more

224 abundant in the ROH number was 1-2 Mb class (Table 2 and Table 4 *vs*. S2 Table and S6

225 Table). If the number of allowed heterozygous SNPs in ROHs continued to increase, the

226 distribution of ROH segments peaked faster (on the number of heterozygous SNPs scale)

227 when the sliding runs were used (Fig 1 *vs*. Fig 2). Necessary to note that the distributions

228 based on consecutive runs had several ROH maximum (Fig 1) while those based on sliding

229 runs had only two ROH maximum (Fig 2). There are further points to be made here. The cow

230 genome had been filled in with ROH segments when sixteen heterozygous SNPs were

231 allowed at sliding runs while their number was increased up to 240 for consecutive runs (Fig

232 3). Further, as it is shown on Fig 3, the sliding method identified the less number of ROH

233 segments with smaller number of allowed heterozygous SNPs. Summarizing comparative

234 analysis of both methods, we can conclude the large differences between findings of these

235 methods. In the section below will be discussed how the ROH segments are distributed in

236 cow genome.

Evaluation of the distances between ROH segments

238 The number of allowed heterozygous SNPs might be converted into nucleotide distance. For

239 example, the distances between two non-interrupted by heterozygous SNPs ROH segments

240 (**niROH**) separated by one heterozygous SNP will be equal to the mean distance between

241   SNPs in the cow genome multiplied by two. If two heterozygous SNP were allowed then,

242   either the distances between three niROH would be the same as for one heterozygous SNP or

243   it would be double of mean length between SNPs in genome, plus yet one mean distance, if

244   two niROH were separated by two heterozygous SNPs. Anyway, if the number of allowed

245   heterozygous SNPs was being continued to increase, the mean distance between niROH also

246   was increasing. That is the number of allowed heterozygous SNPs may be used as a measure

247   of the distance between niROH. Summarizing the above mentioned reasoning, the upper limit

248   of distance between niROH segments will be $D = (n + 1)m$, where n is the number of allowed

249   heterozygous SNPs and m is the mean distance between SNPs in cow genome equal to 52.1

250   Kb (2508.706681 Mb (genome length covered by SNPs) / 48108 (number of genotyped

251   SNPs) = 52.1 Kb). Based on this formula it is possible to provide some estimates for each

252   length class about maximum and minimum distance between niROH and their length in ROH

253   clusters. For each class there was a threshold beyond of which being impossible to form a

254   ROH clusters from niROH that would not exceed the maximum length for this class. Then, a

255   longer length class will begin to form. But, SNPs unevenly located in the cow genome [26].

256   Owing to this phenomenon and different length of niROHs, the number of heterozygous

257   SNPs suitable for every class length limit would be hardly to predict if not based on the mean

258   distances between SNPs.

259   Calculation of the distances between ROH segments

260   Let us calculate upper and lower limits between niROH in the length classes. If one

261   heterozygous SNP was allowed then mean distance between two niROH segments would be

262   104 Kb and those niROH form an initial ROH clusters. If eight heterozygous SNPs were

263   allowed, then different number of niROH segments might formed the ROH clusters

264   belonging to the 1-2 Mb class. In this case, the maximum mean upper limit on distance

265 between two niROH will be 450 Kb. Otherwise, this distance should begin to decrease if the

266 number of niROH will increase. Given that maximum length 2 Mb could not be exceeded for

267 1-2 Mb class, then minimal mean lower limit on distances between niROH would be nine 133

268 Kb length niROH separated by eight 100 Kb heterozygous SNP segments. Based on Fig 1 it

269 can be concluded, with an increase in the number of heterozygous SNPs over 8, the number

270 of niROH clusters satisfying 1–2 Mb class began to decrease. Further increase in the number

271 of heterozygous SNPs led to an increase in the number of ROH in the next class. The same

272 calculation is valid for other length classes but each of them has its own maximal and

273 minimal number of niROH and distances between them.

274 Since the distances between SNPs in the cattle genome vary considerably [27], some SNPs

275 may be located extremely close to each other. It can be illustrated for 1-2 Mb class when

276 more than 60 heterozygous SNPs were allowed (Table 2). Indeed, for 60 allowed

277 heterozygous SNPs in ROH, summed mean distance between 60 niROH would be ≈ 3.5 Mb,

278 that exceeds the length 2 Mb for 1–2 Mb class. As it was mentioned above there were

279 noticeable differences in results of two methods. The sliding window method underestimated

280 the distances between ROH and, therefore, the position of the peaks were shifted and merged

281 relative to those for consecutive runs (Fig 1 *vs*. Fig 2).Therefore, for these methods the

282 number of ROH and distances between niROH considerably differed and obtained findings

283 were inconsistent. Visualization of ROH formation owing to niROH clusters is illustrated on

284 Fig S1 – S3, on which the BTA 28 was chosen having the smallest number of ROH

285 segments. When heterozygous SNPs were disallowed, only niROH were identified on BTA

286 28 (Fig S1). These niROH segments 1-18 Mb length scattered along BTA 28. When one

287 heterozygous SNP was allowed, additional niROH clasters forming new short ROH segments

288 can be seen on BTA 28 (Fig S2). Their length was mostly 1-2 Mb. As the number of

15

289   heterozygous SNPs increased to two, new niROH clusters forming ROH segments different

290   length can be seen on the chromosome 28 (Fig S3).

Distribution of ROH in the cow genome

292   In order to evaluate the chromosomes with largest number of ROH segments, calculation of

293   ROH proportion in them based on normalization to each chromosome length was performed.

294   For both runs the ranked chromosomes list is given in Table 5. Of 29 chromosomes, the

295   highest proportion of ROHs was calculated at BTA14, BTA16, and BTA7 not BTA1

296   **Table 5. Normalized proportion of ROH in the cow chromosomes**

Chromosome	14	16	7	26	8	13	1	22	17	4
Consecutive runs*	1.350	1.280	1.240	1.200	1.190	1.120	1.110	1.098	1.060	1.050
Sliding runs	1.388	1.295	1.218	1.170	1.169	1.161	1.086	1.011	1.050	1.030
Chromosome	20	6	19	21	3	24	10	12	11	5
Consecutive runs	1.024	1.018	1.016	0.964	0.955	0.929	0.928	0.926	0.920	0.911
Sliding runs	0.980	0.999	1.010	0.965	0.963	0.955	0.931	0.918	0.909	0.911
Chromosome	9	2	25	29	23	15	18	27	28	
Consecutive runs	0.908	0.900	0.876	0.844	0.814	0.782	0.754	0.743	0.679	
Sliding runs	0.902	0.898	0.920	0,865	0.851	0.800	0.743	0.774	0.680	

297   Note: * - The values of proportions were ranged from maximum to minimum only for
298   consecutive runs

299   (seventh position in the list), BTA2 (22th position in the list) and BTA3 (15th position in the

300   list). Thus, the number of ROH in chromosomes was not directly proportional to the

301   chromosome length. Spearman's correlation between consecutive and sliding runs data set

302   was r = 1.0 (P ≤ $2.0 \cdot 10^{-7}$).

303   As can be seen on Fig 1 and Fig 2, for both methods the distributions of ROH number from

304   allowed number of the heterozygous SNPs within them was similar to normal. To measure a

305   proximity to the normal distribution Q-Q testing was used. It was shown their similarity to

16

306   normal distribution but to varying degree for different length classes (S4 Fig –S13 Fig). It

307   should be borne in mind that for small samples a greater variation is expected. Obviously, the

308   distributions for classes 1-2 Mb, 2-4 Mb and 8-16 Mb had a positive skewness and light

309   tailed form, while for classes 8-16 Mb and >16 Mb the distributions were the most similar to

310   normal except class >16 Mb for sliding runs (S13 Fig).

Inbreeding

312   To estimate a level of inbreeding in the herds, the mean inbreeding coefficient across herds

313   was calculated (Table 6 and Table 7). If heterozygous SNPs were disallowed, then the mean

314   inbreeding coefficients amounted to 0.111 ± 0.003 and 0.104 ± 0.003 for consecutive and

315   sliding runs respectively, and difference between them was insignificant. The F_{ROH} estimated

316   with Plink was 0.105 ± 0.004 what consistent with that for sliding runs. The largest

317   variability of inbreeding was observed in fourth herd. This was mainly due to highly inbred

318   cow in this herd. Exclusion this cow led to F_{ROH} 0.096 ± 0.005 and 0.089 ± 0.005 for

319   consecutive and sliding runs. In this case, fourth herd significantly differed from herds 1 and

320   herd 5 ($P < 0.01$) and herd 2 ($P < 0.02$), given t-test estimated with Bonferroni correction. It

321   should be noted that in this herd insemination of the cows only from The Netherland bulls has

322   occurred, while for other herds the bulls semen from North America and Canada were used as

323   well. After excluding the highly inbreed cow from fourth herd, the average F_{ROH} in this herd

324   became less than for other herds. This result indicates the correct selection of the bulls, even

325   they were used from a single country. It was shown, the forth herd largely deviated from

326   other herds when it was measured by fixation index and Principal Component Analysis [32].

327   When one heterozygous SNP was allowed in ROH, then F_{ROH} were 0.145 ± 0.003 and $144 \pm$

328   0.003 based on consecutive and sliding runs. Thus, an allowing of even one heterozygous

329   SNP led to significant increase in F_{ROH} ($P \leq 0.001$). Therefore, to evaluate inbreeding in the

17

330   herds, the heterozygous SNPs should not be allowed in ROH due to sizable bias. Distribution

331   of F_{ROH} from the number of heterozygous SNPs is shown on Figure 4. Initially, a relatively

332   gradual increase of F_{ROH} occurred. Then, an increasing greatly accelerated. Finally,

333   increasing gradually slowed down again. In fact, this curve indicated the relative distances

334   between niROH on heterozygous SNPs scale. From Fig 4 it can be seen that 20% of niROH

335   were located at short nucleotide distances less 1 Mb, 80% of ROH clusters formed from

336   niROH were located at intermediate distances, and at least 20% of niROH clusters were

337   located at most longest distances. Thus, in the cow genome there is a complex hierarchical

338   structure of niROH clusters.

339   **Table 6. Estimated mean (± SE) inbreeding coefficient in the herds based on 20 SNPs**

340   **consecutive runs**

Herd	1	2	3	4	5	6
0*						
Inbreeding coefficient	0.117 ± 0.004	0.112 ± 0.002	0.105 ± 0.002	0.111 ± 0.016	0.116 ± 0.003	0.105 ± 0.003
Maximum	0.0006	0.060	0.068	0.047	0.075	0.071
Minimum	0.227	0.153	0.160	0.779	0.166	0.158
1						
Inbreeding coefficient	0.153 ± 0.006	0.143 ± 0.002	0.143 ± 0.002	0.147 ± 0.018	0.147 ± 0.002	0.137 ± 0.003
Maximum	0.003	0.091	0.091	0.072	0.104	0.102
Minimum	0.367	0.184	0.184	0.908	0.191	0.192
2						
Inbreeding coefficient	0.208 ± 0.008	0.195 ± 0.002	0.189 ± 0.002	0.198 ± 0.017	0.199 ± 0.002	0.190 ± 0.003
Maximum	0.010	0.139	0.148	0.102	0.158	0.155
Minimum	0.511	0.235	0.229	0.948	0.240	0.246
4						
Inbreeding coefficient	0.384 ± 0.011	0.372 ± 0.002	0.369 ± 0.002	0.377 ± 0.01	0.376 ± 0.002	0.369 ± 0.002
Maximum	0.053	0.317	0.333	0.229	0.349	0.333
Minimum	0.742	0.407	0.406	0.972	0.416	0.410
8						
Inbreeding coefficient	0.787 ± 0.008	0.788 ± 0.001	0.788 ± 0.001	0.791 ± 0.005	0.790 ± 0.001	0.788 ± 0.001
Maximum	0.396	0.763	0.764	0.667	0.773	0.772
Minimum	0.934	0.809	0.813	0.981	0.814	0.808
16						

Inbreeding coefficient	0.9399 ± 0.0009	0.9395 ± 0.0002	0.9391 ± 0.0002	0.9399 ± 0.001	0.9397 ± 0.0002	0.9391 ± 0.0002
Maximum	0.903	0.934	0.935	0.924	0.936	0.936
Minimum	0.964	0.944	0.944	0.985	0.943	0.943
20						
Inbreeding coefficient	0.9492 ± 0.0008	0.9488 ± 0.0002	0.9487 ± 0.0002	0.9495 ± 0.0009	0.9489 ± 0.0002	0.9488 ± 0.0002
Maximum	0.919	0.945	0.945	0.938	0.946	0.945726
Minimum	0.971	0.953	0.952	0.987	0.953	0.952828
100						
Inbreeding coefficient	0.9814 ± 0.0002	0.98124 ± 0.00009	0.98117 ± 0.00009	0.9814 ± 0.0002	0.98115 ± 0.00008	0.9814 ± 0.0001
Maximum	0.975	0.979	0.979	0.978	0.980	0.979
Minimum	0.986	0.983	0.982	0.990	0.982	0.983
240						
Inbreeding coefficient	0.9872 ± 0.0001	0.98641 ± 0.00007	0.98635 ± 0.00007	0.9864 ± 0.0001	0.98626 ± 0.00007	0.98642 ± 0.00007
Maximum	0.983	0.984	0.985	0.985	0.985	0.985
Minimum	0.989	0.988	0.987	0.990	0.988	0.988

Note: * - The number of allowed heterozygous SNPs in ROH

Table 7. Estimated mean (± SE) inbreeding coefficient in the herds based on 20 SNPs sliding runs

Herd	1	2	3	4	5	6
Zero heterozygous SNP in ROH						
Inbreeding coefficient	0.110 ± 0.004	0.105 ± 0.002	0.098 ± 0.002	0.103 ± 0.015	0.109 ± 0.003	0.098 ± 0.003
Minimum	0.0003	0.055	0.063	0.040	0.070	0.065
Maximum	0.204	0.146	0.153	0.739	0.158	0.151
One heterozygous SNP in ROH						
Inbreeding coefficient	0.154 ± 0.007	0.143 ± 0.002	0.137 ± 0.002	0.146 ± 0.018	0.148 ± 0.002	0.138 ± 0.003
Minimum	0.002	0.092	0.094	0.071	0.104	0.104
Maximum	0.392	0.186	0.184	0.938	0.192	0.194
Two heterozygous SNP in ROH						
Inbreeding coefficient	0.227 ± 0.01	00.213 ± 0.002	0.207 ± 0.002	0.217 ± 0.018	0.217 ± 0.002	0.208 ± 0.003
Minimum	0.010	0.157	0.167	0.108	0.173	0.173
Maximum	0.583	0.254	0.244	0.980	0.254	0.258
Four heterozygous SNP in ROH						
Inbreeding coefficient	0.474 ± 0.012	0.463 ± 0.002	0.461 ± 0.002	0.470 ± 0.012	0.467 ± 0.002	0.463 ± 0.002
Minimum	0.072	0.417	0.417	0.298	0.439	0.429
Maximum	0.859	0.503	0.504	0.989	0.502	0.506
Eight heterozygous SNP in ROH						
Inbreeding	0.924	0.929	0.928	0.930	0.930	0.929

coefficient	± 0.007	± 0.0005	± 0.0007	± 0.003	± 0.0006	± 0.0007
Minimum	0.562	0.918	0.915	0.826	0.921	0.919
Maximum	0.988	0.940	0.941	0.100	0.942	0.942
Twelve heterozygous SNP in ROH						
Inbreeding coefficient	0.988 ± 0.0005	0.989 ± 4.6·10⁻⁵	0.989 ± 6.15·10⁻⁵	0.989 ± 0.0001	0.989 ± 6.16·10⁻⁵	0.989 ± 5.58·10⁻⁵
Minimum	0.961	0.987	0.988	0.985	0.987	0.988
Maximum	0.9898	0.9897	0.9897	0.9898	0.9898	0.9896
Sixteen heterozygous SNP in ROH						
Inbreeding coefficient	0.989824	0.989824	0.989824	0.989824	0.989824	0.989824
Minimum	NA	NA	NA	NA	NA	NA
Maximum	NA	NA	NA	NA	NA	NA

344

Discussion

346 Over past decade the ROH approach has been widely used both on humans [3] and farm

347 animals [2]. Hallmark of ROH studies is a lot of software and thresholds criteria to be used in

348 them. The most widely applied software tools to identify ROH stretches are either sliding

349 window or consecutive runs. We preferred detectRUNS where both approach were

350 implemented [9]. The consecutive runs resulted in the average number of ROH 94.4 ± 2.7

351 while sliding runs 86.0 ± 2.6. This values for North American [8], Italian [16], European

352 Holstein [21], and Polish Holstein Black-and White variety [28] are 82.3 ± 9.8(SD), 81.7 ±

353 9.7(SD), 74.6 ± 2.3(SE), and 53.3 ± 7.3(SD) respectively. The first three values do not differ

354 significantly from our ones, while the value for Polish cattle considerably differs. It was

355 shown that allowance of even one heterozygous SNP in ROH significantly increases the

356 number of ROH by 56.9 and 60.7 points for consecutive and sliding runs respectively (Table

357 1 and Table 3). A limited number of studies investigated the effect of allowed heterozygous

358 SNPs on ROH data. Howrigan et al. [13] recommended disallowing any heterozygous SNPs

359 in ROH, while Ferencakovic et al. [22] suggested the number of heterozygous SNPs allowed

360 should be determined separately for each ROH length of interest and for each SNPs density.

20

361 The relative frequency of ROH number in different length classes for consecutive runs were

362 61.4% (1–2 Mb), 19.8% (2-4 Mb), 11.3% (4–8 Mb), 5.5% (8–16 Mb) and (longer than 16

363 Mb) 1.9%, while for sliding runs these values were 60%, 19.8%, 12.1%, 5.8%, 2.1%. Thus,

364 the largest number of ROH was identified in the shortest 1–2 Mb class. By Plink-running of

365 the cow genome was identified the following frequency of ROH in five categories 52% (0-2

366 Mb), 25% (2–4 Mb), 14% (4–8 Mb), 7% (8-16 Mb) and (longer than 16 Mb) 2.5% whose

367 distribution is slightly different from those identified with detectRUNS. For North American

368 Holstein animals these values were 43.5%, 23.9%, 17.7%, 10.5%, and 4.7% [8].

369 Corresponding values for Italian Holstein bulls were 56.9%, 20.8, 11.9, 7.2, and 3.7% [16]

370 and Polish Holstein 21%, 37.1%, 23.1%, 13.3, and 5.6% [28]. Thus, when we used

371 detectRUNS for genome scanning of our local Holsteinzied cows, we obtained an abundant

372 number of short ROH resulted from haplotypes reflecting the ancient relationship within

373 breeding animals. But, when we used Plink, the values were similar to those for American

374 and Italian data. It should be considered that authors of the article [28] used cgaTOH software

375 and their data considerably differ from other data. Whether this result was due to cgaTOH

376 software or/and breeding management requires furthers analysis. Estimation of the true

377 amount of short ROH is important, since 0.1–3 Mb ROH segments have the more number of

378 deleterious variants than segments greater of 3 Mb [20]. Furthermore, to evaluate genomic

379 estimated breeding value (GEBV), the short ROH is important for genomic ROH-based

380 relationship matrix (G_{ROH}) construction [29]. According to our estimates, the most number of

381 ROH was in 1–2 Mb class, which included niROHs shorter than 1 Mb, located at short

382 distances and forming 1-2 Mb length ROH clusters, thereby determining the structure of the

383 cattle genome (Fig S1-S3). Finally, the results of the animal ROH genome scanning can

384 substantially depend on not only from selection strategy, but software used. This fact should

385 be taken into account in comparative analysis of the ROH data.

386 Currently evaluated with detectRUNS the mean inbreeding coefficient for six herds was

387 0.111 ± 0.003 and 0.104 ± 0.003. It is equal to 0.105 ± 0.004 based on Plink runs. These

388 values do not differ significantly (t-test). But, F_{ROH} increased significantly (P < 0.001) from

389 0.111 ± 0.003 to 0.145 ± 0.003, when one heterozygous SNP was allowed in ROH (Table 6).

390 However, Mastrangelo et al. [24] showed that the differences between F_{ROH} estimated using

391 one, two, and three heterozygous SNPs were very small in Italian Holstein (0.042, 0.045,

392 0.049) when F_{ROH} was estimated for >4 Mb ROH. F_{ROH} for American Holstein was 0.12 in

393 2011 year and after applying genomic selection it increased until 0.15 in 2018 year [8]. For

394 European [21], Italian [16], and Polish Holstein [26] these values were 0.108 ± 0.006 (SE),

395 0.116 ± 0.001 (SE), and 0.118 ± 0.027 (SD) respectively. In the studies cited above, ROH

396 data were based on 50k arrays, thereby not shorter than 1 Mb ROH segments were identified.

397 Thus, short ROH were underestimated that might be led to an inaccurate estimate of

398 inbreeding occurred several tenth years ago. The accurate knowledge of inbreeding in the

399 herds is necessary for both calculation of inbreeding trend and evaluation of the selection

400 strategies. In short ROH the deleterious mutations may be concentrated [20]. To overcome

401 this problem a high density arrays or whole genome sequencing (WGS) should be used.

402 Comparing 50K and HD panels provide evidence that data from 50K panel lead to imprecise

403 determination of short ROH segments [22]. However, detection the ROH based on high-

404 density or 50k chip data was shown to be able to give the estimates of current inbreeding the

405 most similar to ROH values obtained from sequence data [30]. Bhali et al. [31] provided the

406 comprehensive WGS data for Branvich cattle. The medium – sized ROH (0.1–2 Mb) were

407 the most frequent class (50.46%) and contributed most (75%) of the total genomic

408 inbreeding, while short ROH 50 – 100 Kb occurred almost as frequent (49.17%) as medium-

409 sized ROH, they contributed only 19.52% to total genomic inbreeding. These findings

410 provide an accurate estimate of short ROH in the cattle genome and their input to total

411 inbreeding. The average F_{ROH} estimated from WGS data was 0.14 in Branvich cattle. This

412 value is less than WGS F_{ROH} in Holstein 0.18 [31]. Thus, the results inferred from WGS

413 provide a larger F_{ROH} than those from SNPs array (0.15 *vs*. 0.18). Thus, 50k panel can not

414 accurately capture inbreeding occurred several tenth years in the past.

415 It may be assumed that for a herd or population has to be an event horizon beyond, which it is

416 impossible to obtain only from ROH data valid information about their breeding history

417 events. Our local population with both reduced effective population size, ongoing admixture

418 and inbreeding throughout its breeding history, accompanied by recombination, should has

419 the greatest number of short ROH less than 1 Mb in herds. These short ROH can be

420 considered as a relict ROH segments formed due to some population events, such as

421 admixture, drift, bottleneck, and inbreeding occurred many decades ago. Previously, with

422 Principal Component Analysis the bottleneck in our local herds was not proved [32]. An

423 accurate interpretation of these ROH can be troublesome without knowledge of the herd

424 management history. Nevertheless, it is very important to know the true number of short

425 ROHs at the animals analyzed (see above mentioned WGS data). Thus, the event horizon

426 might be depended as from pedigree information, ROH length profile (used SNPs array or

427 WGS) as well as on algorithm-defined approach for ROHs identification. However, ROH

428 segments shorter than 1 Mb can be considered as being beyond the event horizon due to

429 significant shortening by recombination. An addition, short ROH characterized by strong LD

430 among markers are not always considered autozygous but, nevertheless, are due to mating to

431 distantly related animals [33]. Summarizing, it should be assume that inbreeding data based

432 on array can only be relatively correct for an ROH larger than 1 Mb (no more 50 generation

433 back).

434 **Distribution of ROH in the cow genome**

435   In several studies have noticed an uneven distribution of ROH in the chromosomes [e.g. 10,

436   14, 34]. Giving the number of ROH in the chromosomes, their proportion normalized on each

437   chromosome length was calculated (Table 5). Of 29 chromosomes, the most fill in with ROH

438   segments were BTA14, BTA16 and BTA7 for both runs approach used. Purfield at al. [17]

439   noticed that the largest extent of overlapping ROH positions among studied breeds had ROH

440   segments on BTA14 and BTA16. The genome regions with the highest frequency of ROH

441   occurrence were called "ROH islands" [35, 36]. The ROH islands on BTA14 and BTA16

442   were identified among Polish Holstein-Friesian animals [26]. In American Holstein ROH

443   distribution was more variable among the genomes of selected animals in comparison to a

444   more even ROH distribution at unselected animals [37]. Further, in this study more than 40

445   genetic regions under selection on BTA2, BTA7, BTA16 and BTA20 were identified.

446   Regions with high level of ROH for American and New Zealand Jersey cows and bulls were

447   detected on BTA3 and BTA7 [10]. One strongest ROH region common for Kholmogor and

448   Holstein breeds and one region common for Yaroslavl and Holstein breeds was found on

449   BTA14 and BTA16 [21]. An extremely non-uniform ROH patterns among bovine

450   populations of Angus, Brown Swiss, and Fleckvieh breeds were mainly located on BTA6,

451   BTA7, BTA16, and BTA21 [34]. On BTA7 was found the highest number of ROH islads

452   among all Neilore breed lineages [38]. Furthermore, on BTA14 in dairy Gyr breed (Bos

453   Indicus) has shown an enrichment of genes affecting traits of interest for dairy breeds [39].

454   Goszczynski et al. [40] analyzed ROH >16 Mb (three generations since the common

455   ancestor) in highly inbred Retinta bulls. Among other chromosomes the highest occurrence of

456   ROH was found on BTA7. Summarizing studies cited above, it can be suggested that BTA7

457   is outstanding regarding ROH islands occurrence in the cattle genome but there is no overall

458   direct relationship between frequency of ROH segments in chromosomes and identified ROH

459   islands there.

460 Having determined proportion of ROHs in the chromosomes, it is necessary to define a type

461 of distribution over which ROH segments located in genome. Considering the dependence of

462 the ROH number from the number of allowed heterozygous SNPs in them, promoted to

463 realize this task. For this, recalculation of the number of allowed heterozygous SNPs in ROH

464 into distances between niROH was performed. To further elucidate the structure of ROH

465 segments they were classified into five length classes. It turned out that in each class the

466 number of ROH segments, which were formed from niROH were distributed similar to the

467 normal distribution (Fig 1 and Fig 2). It was true for total set of ROH segments as well (Fig

468 3).This indicates a cluster structure of ROH in the cow genome. The ROH clusters form a

469 hierarchical structure from short to longer as can be visualized on Fig S1-S3. As was shown,

470 the number of short ROH was highly depended on the software and genotyping method used.

471 Moreover, we show that the consecutive runs more accurately identified ROH pattern in cow

472 genome. Nevertheless, it would like to point out that both methods coincide in estimate of

473 ROH distribution type being similar to normal. Taken together our findings it should be

474 admit about uneven distribution ROH segments in the cow genome as a result of different

475 population events occurred during it history.

Conclusions

476

477 Our analysis of the ROH data was able to estimate that consecutive runs the most accurately

478 identified ROH in the cow genome. It was shown that missing SNPs had not sizable effect

479 on the number of ROH, while an allowance of even one heterozygous SNP in ROH had

480 significant effect. Therefore, caution should be taken to allow any number of heterozygous

481 SNPs in ROH. The mean inbreeding coefficient for our local herds was 0.111 ± 0.003, which

482 was not differ from those for the Holstein breed over the world. It was suggested that ROH

483 segments had a tendency for clustering in the cow genome. Distribution of ROH clusters in

484 the cow genome was similar to normal. Moreover, the number of ROH segments in the

485 chromosomes did not depend on their length and the most abundant in ROH segments were

486 BTA14, BTA7, and BTA18.

Author contribution

References

489 **1**. Meuwissen T, Luo Z. Computing inbreeding coefficient in large populations. Genet Sel
490   Evol. 1992; 24(4): 305-313. doi.org/10.1186/1297-9686-24-4-305

491 **2**. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of
492   homozygosity: current knowledge and applications in livestock. Animal Genet. 2016; 48:
493   255–271. doi: 10.1111/age.12526.

494 **3**. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity:
495   windows into population history and trait architecture. Nat Rev Genet. 2018;19(4):220-
496   234. doi: 10.1038/nrg.2017.109

497 **4.** Gusev A, Lowe JK, Stoffel M, Daly MJ, Altshuler D, Breslow JL, Friedman JM, Peer I.
498   Whole population, genome-wide mapping of hidden relatedness. Genome Res. 2009; 19:
499   318–326. doi: 10.1101/gr.081398.108

500 **5**. Browning SR, Browning BL. High-resolution detection of identity by descent in unrelated
501   individuals. Am J Hum Genet. 2010; 86: 526–539. doi: 10.1016/j.ajhg.2010.02.021

502 **6**. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-smith C, Durbin R. Application note
503   BCFtools / ROH: a hidden Markov model approach for detecting autozygosity from next-
504   generation sequencing data. Bioinformatics. 2016; 32(11): 1749-1751. doi:
505   10.1093/bioinformatics/btw044

506 **7**. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association

507 and population-based linkage analyses. Am J Hum Genet. 2007; 81: 559–575. doi:

508 10.1086/519795

509 **8**. Forutan M, Mahyari SA, Baes C, Melzer N, Schenkel FS, Sargolzaei M. Inbreeding and

510 runs of homozygosity before and after genomic selection in North American Holstein

511 cattle. BMC Genomics. 2018; 19:98. DOI 10.1186/s12864-018-4453-z

512 **9**. Biscarini F, Paolo Cozzi P, Gaspa G, Marras G. detectRUNS: Detect Runs of

513 homozygosity and runs of heterozygosity in diploid genomes. R package version 0.9.6.

514 2020; Retrieved from https://CRAN.Rproject.org/package=detectRUNS

515 **10**. Howard JT, Maltecca C, Haile-Mariam M, Hayes BJ, Pryce JE. Characterizing

516 homozygosity across United States, New Zealand and Australian Jersey cow and bull

517 populations. BMC Genomics. 2015; 16: 187. DOI 10.1186/s12864-015-1352-4

518 **11**. Kim E-S, Sonstegard TS, Van Tassell CP, Wiggans G, Rothschild MF. The relationship

519 between runs of homozygosity and inbreeding in Jersey cattle under selection. PLoS One.

520 2015; 10: e0129967. doi: 10.1371/journal.pone.0129967

521 **12**. Zhang L, Orloff MS, Reber S, Li S, Zhao Y, Eng C. CGATOH: extended approach for

522 identifying tracts of homozygosity. PLoS One. 2013; 8: e57772.

523 doi.org/10.1371/journal.pone.0057772

524 **13**. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of

525 homozygosity: a comparison of three autozygosity detection algorithms. BMC

526 Genomics. 2011; 12: 460. http://www.biomedcentral.com/1471-2164/12/460

527   **14**. Ferencakovic M, Hamzic E, Gredler B, Curik I, Johann S. Runs of homozygosity reveal

528   genome-wide autozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus

529   Scientificus (ACS). 2011; 76(4): 325-329.

530   **15**. Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding

531   depression in Holstein cattle using whole-genome SNP markers and alternative measures

532   of genomic inbreeding. J Dairy Sci. 2013; 96(7): 4697–4706. doi: 10.3168/jds.2012-

533   6435.

534•   **16**. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams

535   JL, Macciotta NPP. Analysis of runs of homozygosity and their relationship with inbreeding

536   in five cattle breeds farmed in Italy. Animal Genet. 2014; 46: 110–121.

537   DOI: 10.1111/age.12259

538   **17**. Purfield DC, Berry D, McParland S, Bradley DG. Runs of homozygosity and population

539   history in cattle. BMC Genetics. 2012; 13: 70. doi: 10.1186/1471-2156-**13-70**

540   **18**. Karimi Z. Runs of homozygosity patterns in taurine and indicine cattle breeds. Doctoral

541   thesis, University of Natural Resources and Life Sciences, Vienna. 2013.

542   **19**. Zavarez LB, Utsunomiya YT, Carmo AS, Neves HHR, Carvalheiro R, Ferencalkovic M,

543   et al. Assessment of autozygosity in Nellore cows (Bos indicus) through high density

544   SNP genotypes. Frontiers in Genetics. 2015; 6: 5. doi: 10.3389/fgene.2015.00005

545   **20**. Zhang Q, Guldbrandtsen B, Bosse M, Lund MS, Sahana G. Runs of homozygosity and

546   distribution of functional variants in the cattle genome. BMC Genomics. 2015; 16: 542.

547   doi: 10.1186/s12864-015-1715-x.

548 **21**. Zinovieva NA, Dotsev AV, Sermyagin AA, Deniskova TE, Abdelmanova AS,

549   Kharzinova VR, et al. Selection signatures in two oldest Russian native cattle breeds

550   revealed using high density single nucleotide polymorphism analysis. PLoS ONE. 2020;

551   15(11): e0242200. doi: 10.1371/journal.pone.0242200

552 **22**. Ferenčaković M , Sölkner J, Curik I. Estimating autozygosity from high-throughput

553   information: effects of SNP density and genotyping errors. Genet Sel Evol. 2013; 45: 42.

554   doi: 10.1186/1297-9686-**45-42**

555 **23**. Ceballos FC, Hazelhurst S, Ramsay M. Assessing runs of Homozygosity: a comparison

556 of SNP Array and whole genome sequence low coverage data.   BMC Genomics. 2018; 19:

557   106. doi: 10.1186/s12864-018-4489-0

558 **24**. Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT, Portolano B.

559   Genomic inbreeding estimation in small populations: evaluation of runs of

560   homozygosityin three local dairy cattle breeds. Animal. 2016; 10(5): 746-754.

561   doi.org/10.1017/S1751731115002943

562 **25**. Addo S, Klingel S, Hinrichs D, Thaller G. Runs of homozygosity and NetView analyses

563   provide new insigt into the genome –wide diversity and admixture of three German

564   cattle breeds. PLoS ONE. 2019; 14(12): e0225847. doi: 10.1371/journal.pone.0225847

565 **26**. R Development Core Team R: a language and enviroment for statistical computing. R

566   foundation for statistical computing. Viena. http://www.Rproject.org. Accessed 1 Dec

567   2019

568 **27**. Matukumalli LK, Cynthia T, Lawley CT, Robert D,. Schnabel RD, Taylor JF, et al.

569   Development and characterization of a high density SNP genotyping assay for cattle.

570   PLoS ONE. 2009; 4(4): e5350. doi:10.1371/journal.pone.0005350

571 **28**. Szmatoła T, Gurgul A , Jasielczuk I, Zabek T, Ropka-Molik K, Litwinczuk Z, Bugno-

572   Poniewierska M. Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle

573   Breeds Representing Different Production Types. Animals. 2019; 9: 1024.

574   doi.org/10.3390/ani9121024.

575 **29**. Luan T, Yu X, Dolezal M, Bagnato A, Meuwissen THE. Genomic prediction based on

576   runs of homozygosity. Genet Sel Evol. 2014; 46: 64. doi: 10.1186/s12711-014-0064-6

577 **30**. Zhang Q, Calus MPL, Guldbrandtsen B, Lund MS, Sahana G. Estimation of inbreeding

578   using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds.

579   BMC Genetics.2015; 16: 88. doi: 10.1186/s12863-015-0227-7

580 **31**. Bhati M, Kadri NK, Crysnanto D, Pausch H. Assesing genomic diversity and signatures

581   of selection in Original Braunvich cattle using whole –genome sequencing data. BMC

582   Genomics. 2020; 21: 27. doi: 10.1186/s12864-020-6446-y

583 **32**. Smaragdov MG, Kudinov AA. Assessing the power of principal components and

584   wright's fixation index analyzes applied to reveal the genomewide genetic differences

585   between herds of Holstein cows. BMC Genetics. 2020; 21: 47. doi.org/10.1186/s12863-

586   020-00848-0

587 **33**. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, et al.

588   Whole genome linkage disequilibrium maps in cattle. BMC Genetics. 2007; 8: 74.

589   doi:10.1186/1471-2156-8-74

590 **34**. Sölkner J, Ferencakovic M, Karimi Z, Pérez O'Brien AM, Mészáros G, Eaglen S, Boison

591   SA, Curik I. Extremely non-uniform: patterns of runs of homozygosity in bovine

592   populations. 10th World Congr. Genet Appl Livest. Prod. 2014.

35. Nothnagel M, Lu TT, Kayser M, Krawczak M. Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. Hum Mol Genet. 2010; 19(15): 2927–2935. https://doi.org/10.1093/hmg/ddq198

36. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet. 2012; 91(2): 275–292. doi: 10.1016/j.ajhg.2012.06.014

37. Kim ES, Cole JB, Huson H, Wiggans GR, van Tassel CP, Crooker BA, et al. Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS ONE. 2013; 8: e80813. doi.org/10.1371/journal.pone.0080813

38. Peripolli E, Metzger J, Antuner de Lemos MV, Stafuzza NB, Kluska S, Olivieri BF, et al. Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits. BMC Genomics. 2018; 19: 680. doi: 10.1186/s12864-018-5060-8.

39. Peripolli E, Stafuzza NB, Munari DP, Lima ALF, Irgang R, Machado MA, et al. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (*Bos indicus*) dairy cattle. BMC Genomics. 2018; 19: 34. doi: 10.1186/s12864-017-4365-3.

40. Goszczynski D, Molina A, Tera E, Morales-Durand H, Ross P, et al. Runs of homozygosity in a selected cattle population with extremely inbred bulls: Descriptive and functional analyses revealed highly variable patterns. PLoS ONE. 2018; 13(7): e0200069. doi: 10.1371/journal.pone.0200069.

Supporting information

615 **S1 Table. Estimated mean (± SE) ROH number in the herds based on 15 SNPs**
616 **consecutive runs.**

617 (DOCX)

618 **S2 Table. Estimated mean (± SE) across herds ROH number in length classes based on**
619 **15 SNPs consecutive runs.**

620 (DOCX)

621 **S3 Table. The mean (± SE) across herds ROH length in the classes based on 20 SNPs**
622 **consecutive runs.**

623 (DOCX)

624 **S4 Table**. **Estimated mean (± SE) ROH number in the herds based on 15 SNPs sliding**
625 **runs.**

626 (DOCX)

627 **S5 Table**. **The mean across herds ROH length in the classes based on 20 SNPs sliding**
628 **runs**.

629 (DOCX)

630 **S6 Table**. **The mean (± SE) across herds ROH length in the classes based on Plink runs.**

631 (DOCX)

632 Figures

633 **S1 Fig. Distribution of ROH segments along BTA 28 based on consecutive runs.**

634 Each row represents a cow from first herd and each column represents a genotyped SNPs
635 position. Each ROH is a niROH. The heterozygous SNPs were disallowed.

636 (PDF)

637 **S2 Fig. Distribution of ROH segments along BTA 28 based on consecutive runs.**

638 Each row represents a cow from first herd and each column represents a genotyped SNPs
639 position. Each ROH consists of some niROH number. One heterozygous SNPs were allowed.

640 (PDF)

641 **S3 Fig. Distribution of ROH segments along BTA 28 based on consecutive runs.**

642 Each row represents a cow from first herd and each column represents a genotyped SNPs
643 position. Each ROH consists of some niROH number. Two heterozygous SNPs were
644 allowed.

32

645   (PDF)

646   **S4 Fig. Q-Q plot based on consecutive runs**.

647   Distribution of ROH segments from 1-2 Mb class.

648   (PDF)

649   **S5 Fig. Q-Q plot based on consecutive runs**.

650   Distribution of ROH segments from 2-4 Mb class.

651   (PDF)

652   **S6 Fig. Q-Q plot based on consecutive runs**.

653   Distribution of ROH segments from 4-8 Mb class.

654   (PDF)

655   **S7 Fig. Q-Q plot based on consecutive runs**.

656   Distribution of ROH segments from 8-16 Mb class.

657   (PDF)

658   **S8 Fig. Q-Q plot based on consecutive runs.**

659   Distribution of ROH segments from >16 Mb class.

660   (PDF)

661   **S9 Fig. Q-Q plot based on sliding runs**.

662   Distribution of ROH segments from 1-2 Mb class.

663   (PDF)

664   **S10 Fig. Q-Q plot based on sliding runs**.

665   Distribution of ROH segments from 2-4 Mb class.

666   (PDF)

667   **S11 Fig. Q-Q plot based on sliding runs**.

668   Distribution of ROH segments from 4-8 Mb class.

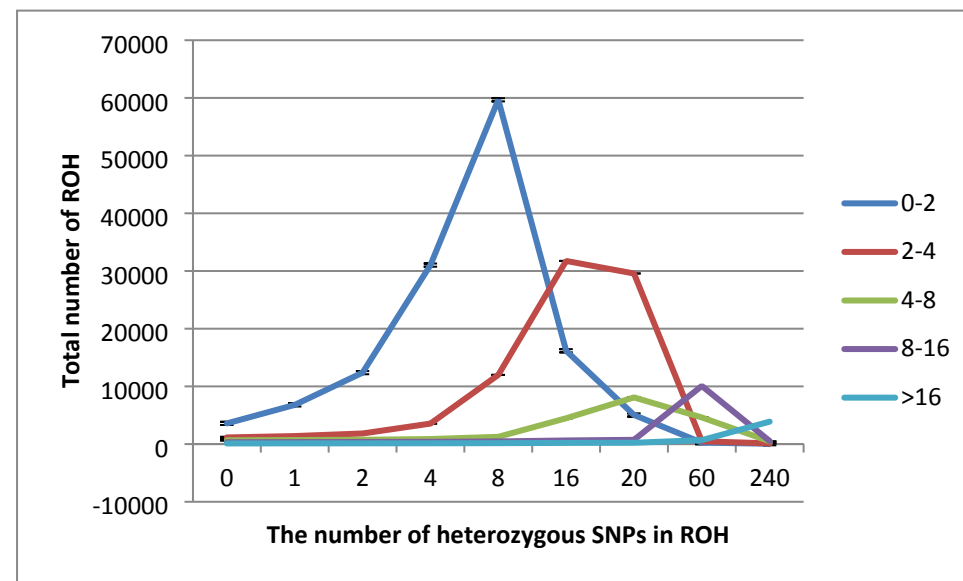669   (PDF)

670   **S12 Fig. Q-Q plot based on sliding runs**.

671   Distribution of ROH segments from 8-16 Mb class.
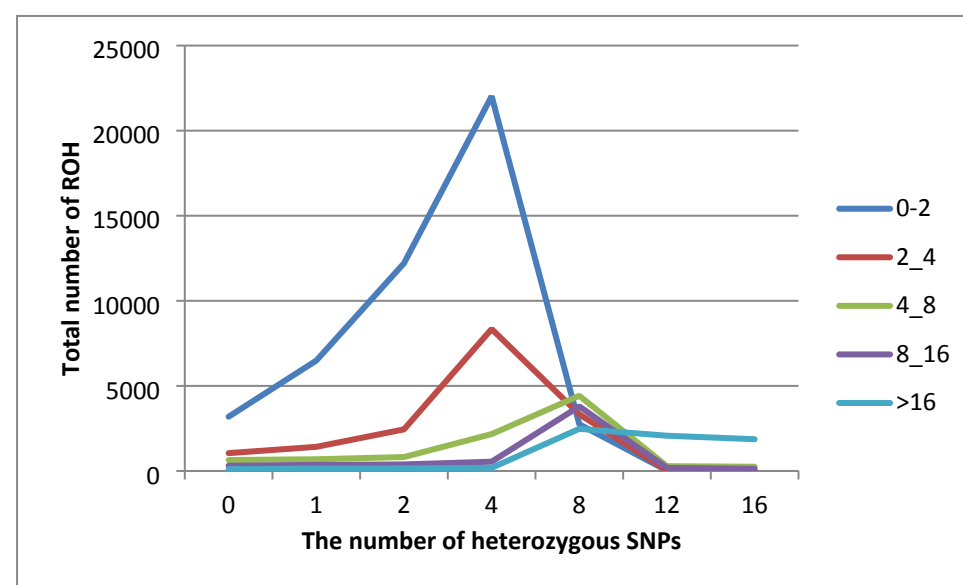
672   (PDF)

673   **S13 Fig. Q-Q plot based on sliding runs**.

674   Distribution of ROH segments from >16 Mb class.

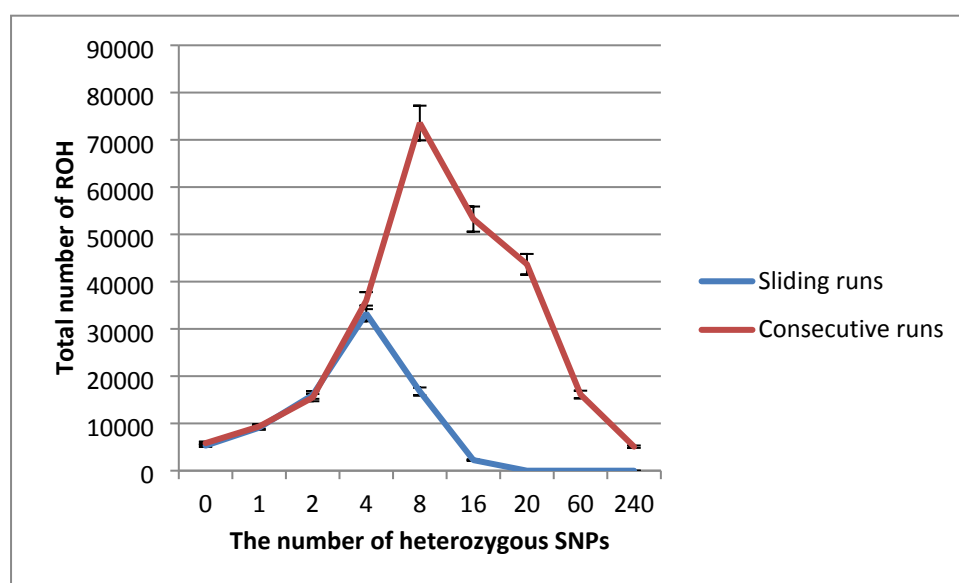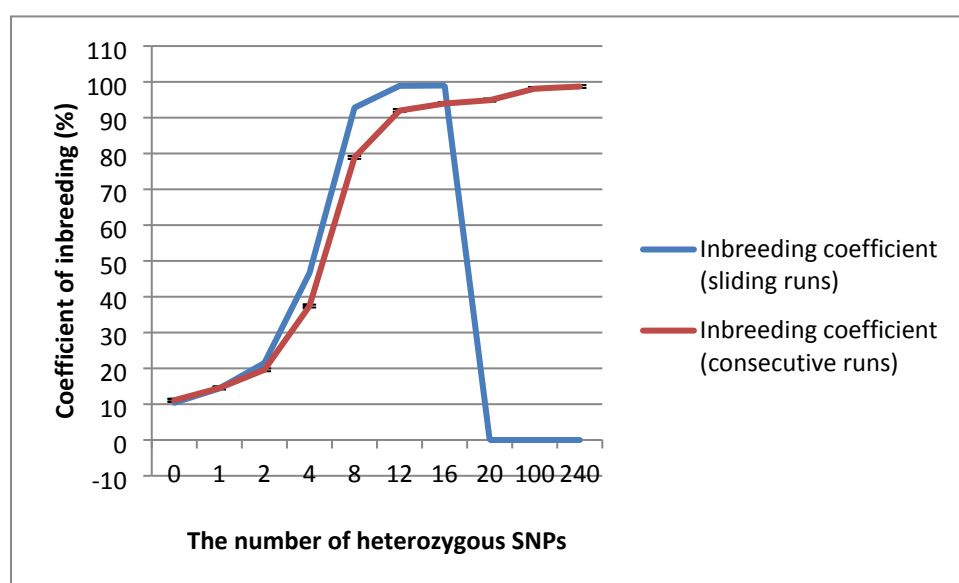675   **(PDF)**

676

677

678

679   **Figure 1. Distribution of ROH in five length classes based on consecutive runs**

680

681

682   Figure 2. Distribution of ROH in five length classes based on sliding runs

683

684

Figure 3. Total distribution of ROH in the cow genome

686

687

Figure 4. Dependence of the inbreeding coefficient on the number of heterozygous SNPs