Original scientific paper

UDK: 637.112.2

Comparison of pedigree-based (F_{PED}) and genomic inbreeding (F_{ROH}) in Croatian Holstein cattle using pedigree and SNP data

DOI: 10.15567/mljekarstvo.2026.0303

Nikola Raguž[1], Dragana Kuzmanović[1], Katarina Marić[2], Mario Shihabi[3], Tomislav Milković[2], Ino Curik[3,4], Marija Špehar[5], Tina Bobić[1], Boris Lukić[1]*

[1]Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek Vladimira Preloga 1, 31000 Osijek, Croatia
[2]Belje plus d.o.o., Svetog Ivana Krstitelja 1a, 31326 Darda, Croatia
[3]University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, 10000 Zagreb, Croatia
[4]Hungarian University of Agriculture and Life Sciences - Institute of Animal Sciences (MATE), Guba Sándor u. 40, 7400 Kaposvár, Hungary
[5]Croatian Agency for Agriculture and Food, Kardinala Alojzija Stepinca 17, 31000 Osijek, Croatia

Received: 15.04.2026. **Accepted:** 09.06.2026.

**Corresponding author: nraguz@fazos.hr*

Abstract

The two most commonly used approaches for estimating inbreeding and structural relationships in livestock populations are pedigree-based and genomic methods. These two measures are not directly equivalent, as pedigree-based inbreeding reflects expected identity by descent, whereas F_{ROH} provides a genomic estimate of realized autozygosity, although it may also be influenced by SNP density, ROH detection parameters, and genotype imputation accuracy. This research aimed to evaluate the pedigree structure in a Holstein population from a Croatian dairy farm and to compare pedigree-based and genomic estimates of inbreeding. Production and pedigree database records were obtained from the national production and pedigree databases. Following filtering, 6,337 cows were selected and analysed for the purpose of this research. These cows had a total of 14,732 records, and their pedigree was constructed through five generations. Thus, a total of 21,779 animals formed the basis of the pedigree. Pedigree analysis revealed increased inbreeding over time, with an average inbreeding coefficient of 0.014. The approximate rate of inbreeding was calculated to be about 0.002 (ΔF). As such, it appears that the population has moderate genetic diversity, but is also accumulating relatedness due to intensive selection. In contrast to the pedigree-based estimates of inbreeding, the genomic estimate using runs of homozygosity (F_{ROH}) indicated much higher inbreeding (mean $F_{ROH<1Mb}$ = 0.091). A direct comparison of individual estimates showed a weak but significant positive correlation between F_{PED} and F_{ROH} in animals with both types of information available (Pearson's r = 0.118, p = 0.007), indicating that the two approaches capture related but only partially overlapping aspects of inbreeding. Further insight into the distribution of ROH segments illustrated that nearly all genomic inbreeding resulted from short and intermediate ROHs ($F_{ROH1\text{-}2Mb}$ = 0.025, $F_{ROH2\text{-}4Mb}$ = 0.025, $F_{ROH4\text{-}8Mb}$ = 0.023). Since these ROH segments largely reflect older genetic relationships, relatively little evidence exists to suggest recent or ongoing inbreeding. Overall, it is evident that pedigree alone does not accurately depict the amount of inbreeding in certain dairy populations. By combining pedigree with genomic information, a more complete picture of both the structural makeup and history of a dairy population can be achieved. Furthermore, combined pedigree and genomic information can provide dairy producers with a valuable resource for managing genetic gain and the long-term sustainability of genetic diversity within Holstein cattle.

Keywords: Holstein cattle; pedigree structure; inbreeding; runs of homozygosity (ROH); genomic inbreeding (F_{ROH})

Introduction

Pedigree data historically served as the primary resource for both managing and evaluating the genetics of livestock breeds. Pedigrees of dairy cattle provide a way of tracing ancestry back through generations and thus allow researchers to estimate demographic characteristics of a breed, such as its average inbreeding (which is the ratio of the amount of genetically identical alleles from each parent to the total number of genes), its genetic diversity and effective population size (Falconer and Mackay, 1996; Mészáros et al., 2026). All three of these parameters are crucial for sustaining healthy and efficient breeding programs for dairy cattle, particularly when combined with large-scale milk recording databases that support the monitoring of production and health-related traits in Holstein populations (Gantner et al., 2024). Holstein dairy cattle represent the largest single breed of dairy cattle in terms of numbers bred throughout the world, and their genetics have been selectively improved for over a century, primarily for increased milk yield. The extensive use of artificial insemination and the widespread use of a relatively small number of elite sires across countries have increased genetic similarity among Holstein herds and populations (Cole et al., 2021). This issue is particularly relevant in small national dairy cattle populations, including the Croatian Holstein population, where the implementation of genomic selection further emphasizes the need for continuous monitoring of genetic diversity and inbreeding (Špehar et al., 2019). While the speed of genetic improvement has provided many benefits for producers and consumers, it has also resulted in higher levels of relatedness and inbreeding within the breed. Therefore, measuring inbreeding is now a critical part of modern dairy cattle breeding programs (Doekes et al., 2019).

Inbreeding has been measured using pedigree data since Wright first introduced his method of calculating the inbreeding coefficient (Wright, 1922). The inbreeding coefficient represents the probability that two randomly selected alleles at a specific locus are identical by descent, that is, inherited from a common ancestor. However, pedigree-based inbreeding estimates are largely dependent on the completeness and accuracy of genealogical information for each animal. If pedigree depth is insufficient or information about some distant ancestors is missing, pedigree-based estimates of inbreeding will likely be lower than the true level of genomic inbreeding (Leroy et al., 2013, Nishio et al., 2023, Bem et al., 2024; Halvonik et al., 2024; Luštrek et al., 2025). Additionally, pedigree analysis makes no distinction between different amounts of genetic contribution made by an animal's ancestors, nor does it take into consideration the random variation resulting from recombination and Mendelian sampling during reproduction.

Genomic information collected via high-density marker arrays is providing new alternatives for measuring inbreeding at the genomic level. Detection of runs of homozygosity (ROH) represents one of the best-known of these alternatives and refers to long stretches of homozygous DNA that are shared by multiple individuals and thus likely originated from a common ancestor (Curik et al., 2014; Ceballos et al., 2018). The F_{ROH}, or genomic inbreeding coefficient, is a direct measurement of genomic autozygosity and is therefore considered a better estimator of individual inbreeding than pedigree-based measures (Purfield et al., 2012; Kardos et al., 2015).

While genomic measures of inbreeding are becoming more commonly used in livestock populations, pedigree data still provide useful additional information. Comparing pedigree-based inbreeding coefficients with those determined from genomic data can help identify potential inaccuracies in pedigree records and provide insight into the historical structure of breeding populations (Doekes et al., 2019).

Thus, the purpose of this research was to analyze the pedigree structure of Holstein cows bred on a Croatian dairy farm using data provided from national databases. This research also aimed to determine how well pedigree-based estimates of inbreeding correlated with estimates derived from genomic data using ROH, allowing comparisons between the two types of inbreeding measurements.

Materials and methods

National pedigree and production data

In this research, we analysed a dataset of Holstein cattle from one Croatian farm, comprising 6,337 cows that represented all cows in production since 2010. Pedigree and production data were obtained from the national database provided by the Ministry of Agriculture, Forestry and Fisheries. Pedigree analysis was performed using all available records for cows originating from the Croatian dairy farm. Only a portion of the national production database (N= 576,251 records) corresponded to animals from the case study farm. After data cleaning, 14,732 production records for 6,337 individual cows remained available for further analysis. Production records were used only to identify the reference population of cows originating from the analyzed farm. These records were not directly included in the estimation of pedigree-based or genomic inbreeding parameters. The selected cows formed the reference population for further pedigree analysis. This was done because all animals had been born on the analysed farm since 2010 and therefore represented most recent generations, for which accurate production records were available. The pedigree data for the reference population were obtained from the national pedigree database, which contained 1,201,074 animals. Once the cows identified in the production database had been selected, their corresponding pedigree information was extracted from the national pedigree database. Extraction of pedigree information without restrictions on the number of previous generations resulted in a sub-pedigree of 23,326 animals. The sub-pedigree included the original reference cows and their known ancestors. Some individuals (n = 89) listed in the pedigree database as either sires or dams of other animals did not appear in the animal field of the database. Therefore, to maintain pedigree integrity, they were added to the database as founding animals. Founders received a sex designation based on their position in the pedigree and an unknown birthdate. For the purposes of this study, pedigree depth was limited to five generations in order to ensure comparable pedigree depth among animals and to reduce bias in pedigree-based inbreeding estimates caused by incomplete ancestral information in deeper generations. This was roughly equivalent to tracing the pedigree back to ancestors born prior to 1990. The final pedigree file used in this research consisted of 21,779 animals. Pedigree depth was additionally evaluated using CFC software (Sargolzaei et al., 2006) by calculating the average discrete generation equivalent, which represents the average equivalent number of fully traced ancestral generations. This value was used as an additional indicator of pedigree completeness and depth. A total of 1,547 animals were excluded from the analysis as a consequence of pedigree truncation to five generations rather than due to quality control criteria. This restriction was applied to ensure comparable pedigree depth and to avoid bias in inbreeding estimates arising from uneven genealogical information across individuals. Excluding these animals was not expected to affect the results substantially, because pedigree completeness generally decreases rapidly beyond five generations, whereas the majority of relevant genealogical relationships can still be captured within this depth. All data processing steps were performed in R (R Core Team, 2023) before conducting pedigree analysis. To protect the identity of the cows and farms involved in the study, all animals were recoded using newly generated random identifiers. The PopRep software package (Groeneveld et al., 2009) was used to analyze the pedigrees. Average pedigree relationships were calculated using PopRep from the available pedigree information. In this context, pedigree relationships represent the expected genealogical similarity among animals based on recorded pedigree links and were summarized by year of birth. An additional approach to assess temporal trends in inbreeding was based on transforming the inbreeding coefficient (F) using the natural logarithm of the non-inbred proportion, $\ln(1-F)$. This transformation was applied to linearize changes in inbreeding over time. The regression slope (b) of $\ln(1-F)$ on year of birth was used to estimate the annual change in inbreeding, while the rate of inbreeding per generation (ΔF) was approximated as $\Delta F = -bL$, where L represents the generation interval. The effective population size (Ne) was subsequently calculated as $Ne = 1/(2\Delta F)$ (Falconer and Mackay, 1996).

Genomic data and quality control

Following the procedure outlined by Usman et al. (2014) and Brajković et al. (2018), somatic cell extracts were obtained from the milk samples. The somatic cells were then used to extract genomic DNA using a commercially available extraction kit (DNeasy Blood & Tissue Kit, Qiagen, Germany). Samples containing at least the recommended amount of DNA (>1000 ng total DNA) were then subjected to low-pass whole genome sequencing (lpWGS). The lpWGS has an average coverage of 1x. All sequencing was completed in two batches: 283 samples were processed by Edinburgh Genetics (UK), and 357 samples were processed by Neogen Genomics (USA). The sequence-derived genotypes were first inferred using a reference panel, together with a combined Beagle (Browning et al., 2018) and GLIMPSE2 (Rubinacci et al., 2021) pipeline, to impute missing genotypes and improve overall genotype accuracy. The next step was to select single nucleotide polymorphisms (SNPs) that corresponded to the SNPs found on the Illumina BovineHD BeadChip (777K SNPs). The quality of a subsample of genotyped animals was evaluated by comparing the inferred genotypes with those obtained by direct genotyping using the BovineHD BeadChip. This evaluation resulted in a concordance rate of >99 %. Autosomal SNPs were selected from the dataset, which reduced it to 564,898 SNPs from 640 individuals. Genomic quality control (QC) was carried out using PLINK (Chang et al., 2015). All SNPs with a call rate below 90 % or deviating from Hardy-Weinberg equilibrium ($p<1\times10^{-7}$) were removed, as were all individuals with a call rate below 80 %. The HWE filtering criterion was retained as a precautionary quality-control step. However, in the final autosomal dataset used for ROH analysis, no SNPs were removed based on this threshold; therefore, HWE filtering did not affect ROH detection or F_{ROH} estimation. After QC, the dataset consisted of 562,225 SNPs and 638 individuals.

Genomic inbreeding (F_{ROH}) estimation

In order to determine runs of homozygosity (ROH), we used the "detectRUNS" package developed by Biscarini et al. (2019). A minimum ROH length of 1 Mb was used as the lower threshold for detecting autozygous segments. Given the high marker density of the final SNP dataset, a minimum of 15 consecutive SNPs was required to support each ROH call and reduce the probability of detecting spurious homozygous segments. The temporal interpretation of ROH was based primarily on segment length, whereas the SNP-number threshold was used as a technical criterion for reliable ROH detection. One heterozygous SNP and one missing genotype were allowed within a ROH segment to account for possible genotyping or imputation errors. As in previous studies, ROH were categorized to their length (1-2 Mb; 2-4 Mb; 4-8 Mb; 8-16 Mb, and >16Mb) in order to analyze the temporal patterns of the inbreeding. Shorter ROH segments, particularly those between 1 and 2 Mb, are commonly interpreted as resulting from more ancient inbreeding events, approximately 25-50 generations ago, whereas longer segments, particularly those >16Mb, reflect more recent common ancestry among individuals (Ferencakovic et al., 2013). F_{ROH} was also calculated for each individual as the proportion of the autosomal genome covered by ROH, both for all ROH combined and separately for each length category.

Comparison of pedigree-based and genomic inbreeding

To evaluate the relationship between pedigree-based and genomic estimates of inbreeding, individual pedigree-based inbreeding coefficients (F_{PED}) were calculated from the final five-generation pedigree using CFC software (Sargolzaei et al., 2006). These values were then merged with individual genomic inbreeding coefficients estimated from runs of homozygosity (F_{ROH}). The comparison was performed only for animals for which both F_{PED} and F_{ROH} estimates were available. The relationship between the two measures was assessed using Pearson's correlation coefficient. Spearman's rank correlation coefficient was additionally calculated as a non-parametric measure of association, given the non-normal distribution of F_{PED} values and the presence of many zero values. All statistical analyses and data merging were performed in R (R Core Team, 2023).

Results and discussion

Population dynamics

The overall pattern of growth in the number of breeding animals, pedigree completeness and genetic parameters provides an integrated view of the demographic and genetic development of the investigated Holstein population. The number of breeding animals increased continuously from the early 1990s up to about 2010 (Figure 1), reflecting both the development of the dairy sector and improvements in data collection. A strong increase observed in 2011 was related to farm management and the establishment of breeding operations within the production system, rather than solely to demographic changes. Specifically, the establishment of a large-scale dairy farm in 2011, housing approximately 1600 cows, contributed substantially to the observed peak in the number of breeding animals. Animals from such large farms can rapidly increase the number of recorded offspring and female breeding animals in a pedigree database. The number of breeding animals declined slightly after the expansion phase, particularly with regard to the number of sires. These tendencies are common in modern cattle breeding systems

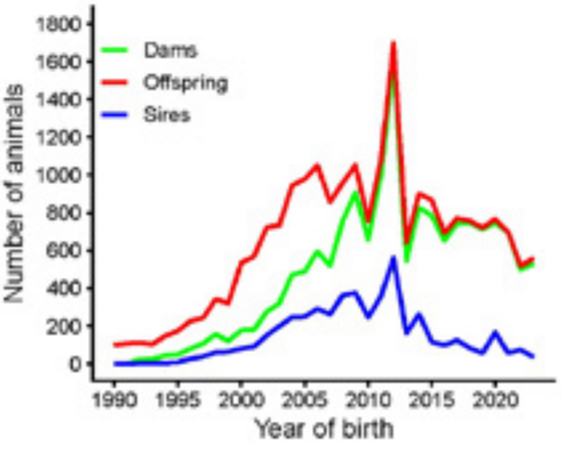

Figure 1. Number of breeding animals

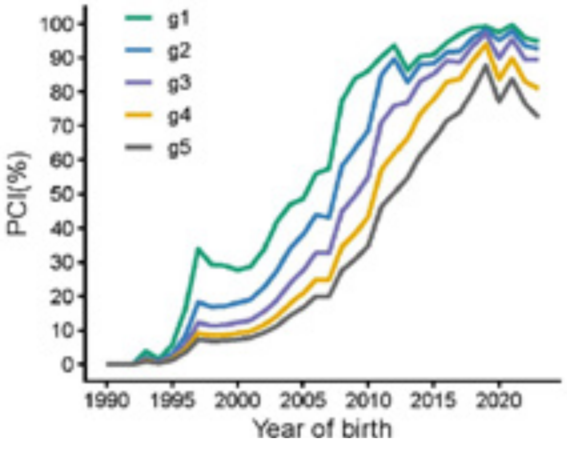

Figure 2. Average pedigree completeness for 1 to 5 generations

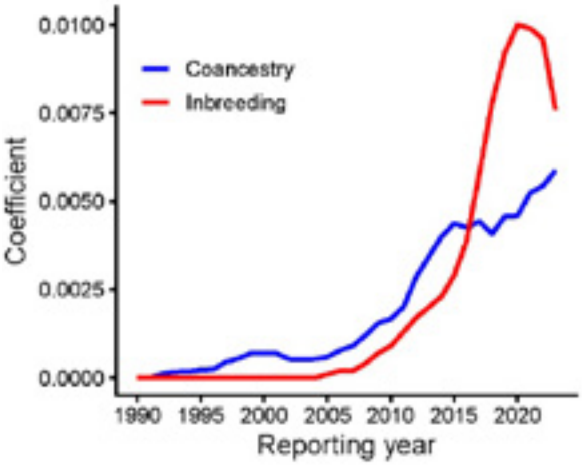

Figure 3. Average genetic relationship (coancestry) and inbreeding

using artificial insemination (AI) and the wide spread use of a relatively small number of highly valuable bulls (Mugambe et al., 2024; Sarviaho et al., 2023).

In parallel to the demographic developments, distinct positive changes in pedigree completeness were observed across generations (Figure 2). In the past, pedigree information was often insufficiently detailed, especially for deeper ancestral generations. This is quite common in historical livestock datasets. Pedigree completeness improved considerably over time. Nowadays, almost all pedigree information for recent generations is available. The average discrete generation equivalent was 2.33, indicating that, despite the five-generation pedigree restriction, the effective pedigree depth corresponded on average to approximately 2.3 fully traced ancestral generations. This confirms that pedigree information beyond the most recent generations was incomplete and supports the cautious interpretation of pedigree-based inbreeding estimates. Comparable developments have also been found in numerous cattle pedigree databases, where centrally managed data collection systems and digital databases for cattle breeders have greatly improved pedigree depth and quality (Vostrý et al., 2023). Pedigree incompleteness leads to underestimation of inbreeding, assuming that unknown ancestors are unrelated (Visscher et al., 2002; Cassell et al., 2003)

Both phenomena, improved pedigree quality and the broad application of AI, were also reflected in the trends in pedigree relationships and inbreeding coefficients (Figure 3). Pedigree relationships, expressed as coancestry, and inbreeding coefficients showed a general tendency to increase over time, thereby indicating an accumulation of genetic relatedness within the herd. Such patterns are typical of intensively selected livestock populations. Dairy cattle are particularly exposed to this type of selective pressure due to the reduced pool of high-value sires that may contribute disproportionately to the next generation (Salehi et al., 2025). The advantages of this approach include accelerated genetic gain. However, it may also result in loss of genetic diversity and an increased risk of inbreeding accumulation. Interestingly, the increase in genetic relatedness was less rapid than the increase in inbreeding, which may indicate specific mating patterns within the herd. This could be caused by population structure, within the population, for example, by decisions regarding herd selection or preferences for specific sire lines used on individual farms.

Structures within dairy cattle populations have been described repeatedly and may affect the relationship between inbreeding and estimates of coancestry (Sarviaho et al., 2023). Considering all the points mentioned above, we conclude that the examined Holstein population underwent a period of demographic expansion, followed by stabilisation, while pedigree recording simultaneously improved. Additionally, trends in additive genetic relationships and inbreeding indicate a continuous accumulation of genetic relatedness within the population, mainly influenced by intensive selection practices and the extensive use of elite sires. Therefore, continuous monitoring of pedigree-based parameters will be necessary to ensure maintenance of genetic diversity and the sustainability of this herd, as well as of the entire breeding program.

Inbreeding dynamics

Although there was an upward trend in inbreeding coefficients over time, the overall level of inbreeding remained relatively low. There was very little, if any, evidence of inbreeding during the early years of the data set. Because pedigree completeness was still low in the early years of the dataset (Figure 2), reflecting shallow genealogical depth, it was difficult to reliably identify whether animals shared common ancestors. By 2004, however, sufficient pedigree information had accumulated for inbreeding to be detected. At that point, the average inbreeding coefficient had risen to 0.0001 (Table 1), indicating that the animals' lineages were deep enough to reveal shared ancestry.

Table 1. Average and maximum inbreeding coefficient (F) in the population by selected years

Year	Number of animals	Average F_{PED}	Max F_{PED}
2004	942	0.0001	0.0625
2008	959	0.0009	0.2500
2012	1697	0.0020	0.2500
2015	867	0.0053	0.1260
2018	756	0.0142	0.2656
2023	559	0.0043	0.0381

The average inbreeding coefficient continued to rise over subsequent years. For example, it increased from 0.0009 in 2008 to 0.0020 in 2012 and then to 0.0053 in 2015. It peaked at 0.0142 in 2018, representing an average inbreeding level of approximately 1.42 % within the herd. This pattern is consistent with trends observed in many other intensively bred livestock herds, where the repeated use of high-merit parents contributes to the accumulation of relatedness within the herd (Leroy et al., 2013; Doekes et al., 2018).

Despite the increasing trend in inbreeding, the level in this herd remains considerably lower than that reported for many other highly selected dairy herds. Many Holstein herds worldwide report average inbreeding levels ranging from about 3 % to 7 %, depending on pedigree depth and selection intensity (Makanjuola et al., 2020). The relatively low pedigree-based inbreeding estimates likely reflect both the recent development of the analyzed herd and the limitations of the available pedigree depth. Although continuous use of AI and international germplasm exchange may have contributed to maintaining relatively low recent pedigree-based inbreeding, the higher F_{ROH} values, particularly those derived from short and intermediate ROH segments, indicate that older accumulated genomic autozygosity was not fully captured by the five-generation pedigree.

Another way to evaluate the structure of the population is to examine the percentage of animals classified as inbred. During the early years of this dataset, there was virtually no indication of inbred animals within the population. However, after 2010, the number of inbred animals began to increase rapidly. Specifically, only 14 animals were identified as inbred in 2008, whereas by 2021, 668 animals were classified as inbred (Table 2). In fact, by the most recent year of record, nearly all animals within this herd would be considered inbred based on pedigree data alone. An important consideration regarding this rapid increase in the number of inbred animals is that such a trend does not automatically imply a substantial loss of genetic diversity.

Rather, this growth is largely due to improvements in pedigree completeness and the improved ability to recognize shared ancestry in deeper pedigrees. The farther back one traces the ancestry of an animal, the more relationships are detected, resulting in higher estimates of the animal's inbreeding coefficient, even if the original mating strategy used to develop these animals has not changed significantly (Charlesworth and Willis, 2009; Leroy et al., 2013).

An additional method used to assess the temporal trend in inbreeding involved transforming the non-inbred proportion, (1-F), using its natural logarithm, ln(1-F), to linearize changes in inbreeding over time (Figure 4). The analysis of ln(1-F) across years of birth showed a clear downward trend, indicating a gradual accumulation of pedigree-based inbreeding within the herd. Because ln(1-F) decreases as F increases, the negative slope of this trend was used to derive the rate of inbreeding.

Table 2. *Number of all and inbred animals, and average F of inbred animals by selected years*

Year	Total animals	Inbred animals	Average F
2008	959	14	0.0603
2011	1076	69	0.0278
2015	867	294	0.0156
2018	756	537	0.0201
2021	698	668	0.0076
2023	559	517	0.0046

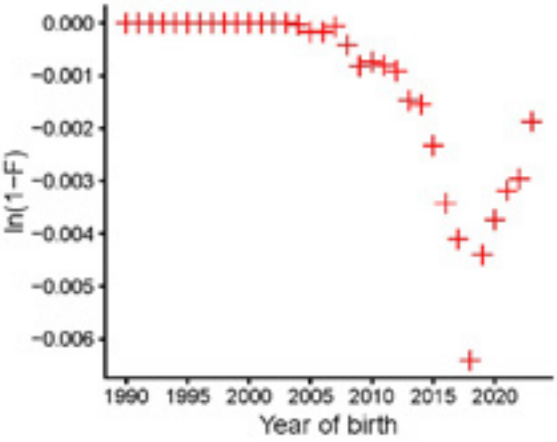

Figure 4. *Average ln(1-F) by year of birth*

The estimated rate of inbreeding per generation was approximately ΔF = 0.002 corresponding to a pedigree-based effective population size (Ne) of approximately 250 individuals. The Ne values determined here fall within the typical range for intensively managed dairy cattle populations that rely heavily on a small number of selected bulls for sire selection in each generation.

Although the Ne calculated here suggests that there is currently no immediate risk of substantial loss of genetic diversity, this trend indicates a long-term tendency toward the accumulation of inbreeding. Similar trends have historically been associated with modern selection methods and therefore serve as another reminder of the importance of continuous assessment and monitoring of genetic diversity. Additionally, because pedigree-based estimations of F are typically less accurate than ROH-based genomics assessments (F_{ROH}), especially when pedigree data are incomplete or otherwise inaccurate, these trends represent a useful starting point for comparison with genomic estimates.

Genomic inbreeding (F_{ROH})

The genomic data indicated the average level of autozygosity in 638 Holstein cows. Although individual F_{ROH} values varied considerably, the mean value was 0.091 with a standard deviation of 0.0563. F_{ROH} ranged from 0 to 0.221 (Table 3).

These values are generally consistent with previous studies in Holstein and other dairy cattle populations, where F_{ROH} typically ranges between approximately 0.05 and 0.15, depending on population structure and selection intensity (Ferenčaković et al., 2013, Makanjuola et al., 2020, Vostrý et al., 2023). F_{ROH} was divided into five ROH length categories.

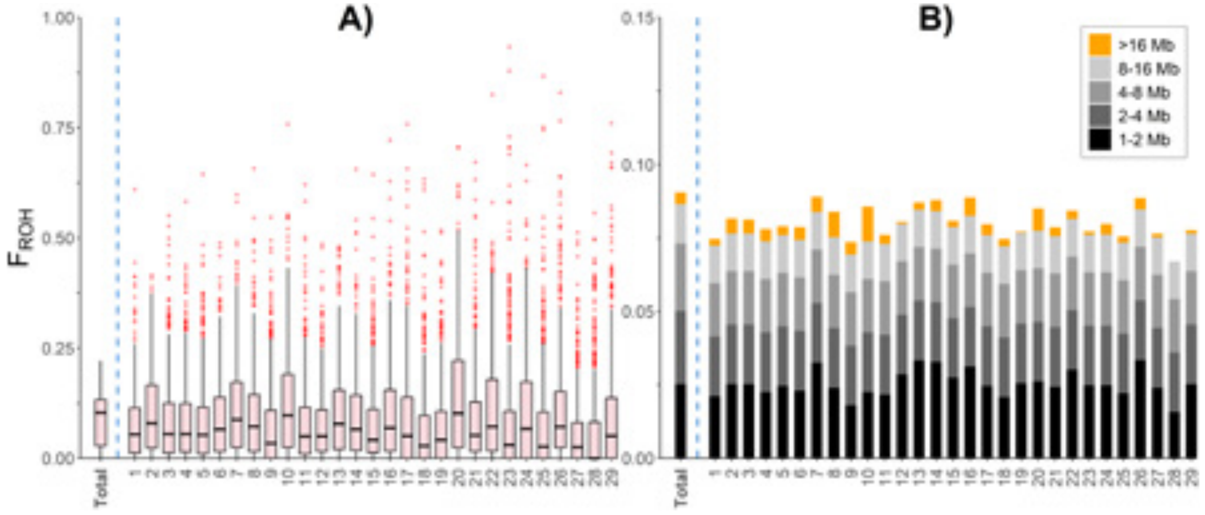

Figure 5. *Distribution of genomic inbreeding (F_{ROH}) in Holstein cows (A) and partitioning of F_{ROH} into ROH length categories (B), shown for the whole genome (Total) and by chromosome. The dashed line separates total from chromosome-specific estimates*

ROH segments in the 1-2 Mb, 2-4 Mb and 4-8 Mb categories had mean values of 0.025, 0.025 and 0.023, respectively. The two longer ROH categories, 8-16 Mb and >16 Mb, had much lower mean values of 0.013 and 0.004, respectively. These values indicate that nearly all of the detected genomic inbreeding resulted from short or intermediate ROH segments. Since these segments reflect older, less recent common ancestry, long ROH segments, which represent recent inbreeding, contributed little to overall inbreeding. Comparable patterns, characterized by the predominance of shorter ROH segments, have been reported in intensively managed dairy populations (Makanjuola et al., 2020, Vostrý et al., 2023). It should be noted that the detection of short ROH segments (<2-4 Mb) may be sensitive to SNP density and methodological settings. However, the use of a high-density SNP panel in this study enabled reliable detection of ROH segments as short as 1 Mb, allowing a more comprehensive characterization of ancient inbreeding compared with studies relying on medium-density chips (Purfield et al., 2012, Ferenčaković et al., 2013). Because genomic data were derived from low-pass whole-genome sequencing followed by imputation, the possibility that genotype uncertainty influenced ROH detection cannot be fully excluded. ROH-based estimates are sensitive to genotype accuracy, SNP density, and the parameters used for ROH calling. In particular, imputation errors may either break true ROH segments or create short artificial homozygous regions. However, the high concordance rate observed between imputed genotypes and direct BovineHD genotyping supports the reliability of the dataset. Nevertheless, F_{ROH} estimates should be interpreted in the context of the applied lpWGS imputation pipeline and ROH detection parameters. Figure 5A shows significant variation of $F_{ROH<1Mb}$ in each cow. Most cows displayed low to moderate levels of autozygosity, whereas fewer cows displayed higher levels of autozygosity. A similar distribution was observed at the chromosome level (Figure 5B). Shorter ROH categories contributed more to chromosome-specific F_{ROH} across almost all autosomes.

Using these values, chromosomes 10 and 20 displayed the highest median values and also exhibited the greatest variability as well as the most pronounced contributions of ROH segments longer than 16 Mb, which are indicative of recent inbreeding. The higher median F_{ROH} values and greater variability observed on chromosomes 10 and 20 may indicate chromosome-specific accumulation of autozygosity. Such patterns could arise from local differences in recombination rate, historical use of related ancestors, or selection acting on genomic regions associated with economically important traits. However, because the present study was not designed to identify selection signatures, this interpretation remains speculative and should be further investigated using dedicated genome-wide selection scan approaches.

Table 3. *Descriptive statistics of genomic inbreeding (FROH) and its partitioning into ROH length categories in Holstein cows.*

Statistic	F_{ROH}	1-2 Mb	2-4 Mb	4-8 Mb	8-16 Mb	>16 Mb
Mean	0.0905	0.0252	0.0249	0.0230	0.0134	0.0040
SD	0.0563	0.0150	0.0171	0.0202	0.0162	0.0088
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Max	0.2214	0.0814	0.0699	0.0820	0.0742	0.0619

Comparison of individual Fped and F_{ROH} estimates

To directly compare pedigree-based and genomic estimates of inbreeding, individual F_{PED} values were calculated from the final five-generation pedigree and merged with F_{ROH} estimates. This comparison was possible for 515 animals for which both pedigree-based and genomic information was available. The correlation between F_{PED} and F_{ROH} was positive but weak (Pearson's r=0.118, p=0.007; Spearman's ρ=0.108, p=0.014), indicating that the two measures capture related but only partially overlapping aspects of inbreeding. The correlation observed in the present study was lower than that reported by Cortés-Hernández et al. (2021), who found a low correlation between F_{PED} and F_{ROH} of approximately 0.30 in a small Holstein population. However, substantially higher correlations between pedigree-based and ROH-based inbreeding estimates have also been reported in other studies, reaching values up to 0.82 (Zhang et al., 2015). This wide range of reported correlations indicates that the agreement between F_{PED} and F_{ROH} is highly dependent on pedigree depth and completeness, population structure, SNP density, ROH detection criteria, and the extent to which genomic inbreeding is driven by recent or more ancient common ancestry.

In our case, the weak correlation can be explained by several factors. First, pedigree-based inbreeding was estimated from a pedigree restricted to five generations, whereas F_{ROH} also captures genomic autozygosity originating from more distant common ancestors, particularly through short and intermediate ROH segments. Second, 67 genotyped animals had F_{PED}=0, indicating that no common ancestors were detected for these animals within the available pedigree depth. In addition, the maximum F_{PED} among genotyped animals was 0.066, which was markedly lower than the maximum pedigree-based inbreeding observed in the complete pedigree dataset. This indicates that the genotyped subset did not include the most highly inbred animals according to pedigree records. Therefore, the weak but significant correlation between F_{PED} and F_{ROH} was not unexpected and supports the interpretation that pedigree-based and genomic inbreeding coefficients are complementary rather than directly interchangeable measures. These results suggest that the available pedigree captured only a limited proportion of the realized genomic autozygosity detected by ROH analysis.

Overall, our findings suggest that the Holstein population under study demonstrates a moderate but increasing amount of inbreeding, consistent with patterns observed in modern dairy cattle breeds due to the widespread use of AI and intensive selection based on elite bulls. Thus, continued observation of inbreeding within dairy cattle breeding programs will remain essential for maintaining the long-term sustainability of such programs, as well as for ensuring a balance between maximizing genetic gain and preserving genetic diversity.

Conclusion

Within the analyzed five-generation pedigree, both F_{PED} and pedigree relationships increased over time. However, this trend should be interpreted in the context of improved pedigree completeness in recent years, which enhanced the ability to detect shared ancestry and may partly explain the apparent increase in F_{PED}.

The higher F_{ROH} values compared with F_{PED} estimates suggest that the available five-generation pedigree captured only part of the genomic autozygosity present in the population. Therefore, the discrepancy between pedigree-based and genomic estimates should be interpreted in the context of pedigree depth, pedigree completeness, and ROH detection parameters. The weak but significant correlation between F_{PED} and F_{ROH} further indicates that these two measures capture related but only partially overlapping aspects of inbreeding. This finding supports the view that the available five-generation pedigree mainly reflects recent and detectable genealogical relationships, whereas F_{ROH} also captures older accumulated autozygosity. The predominance of short and intermediate ROH segments indicates that most genomic inbreeding originated from distant common ancestry, while the limited contribution of long ROH segments suggests that recent inbreeding was comparatively limited.

Overall, these findings highlight the importance of integrating pedigree and genomic information in breeding programs, particularly in populations where genomic data are not available for all individuals. In such cases, genomic information from a representative subset of animals can improve the accuracy of pedigree-based estimates and support more effective monitoring and management of inbreeding at the population level.

Funding

The results presented in this paper are the outcome of the European Union-NextGenerationEU project "Next Generation Animal Production", grant number NPOO.C3.2.R3-I1.04.0141 and the Croatian Science Foundation project PhenoGeno, grant number IP-2022-10-6914 - "Responding to innovations in phenomics and genomics in dairy cattle breeding".

Conflict of interest

The authors declare no conflict of interest.

Authors' contribution

Nikola Raguž conceived the study, performed the majority of the statistical analyses, wrote the manuscript, and supervised the overall research process; Dragana Kuzmanović contributed to the statistical analyses; Katarina Marić contributed to the interpretation of the results; Mario Shihabi performed part of the statistical analyses and contributed to manuscript writing; Tomislav Milković provided the data used in the study; Ino Curik coordinated the research, provided the data and supervised the preparation of the manuscript; Marija Špehar provided the data for the research and contributed to the preparation of the manuscript; Tina Bobić contributed to the interpretation of the results; Boris Lukić conceived the study and contributed to the statistical analyses, manuscript writing, and supervision of the overall research process. All authors read and approved the final manuscript.

ORCID ID

NR. Raguž https://orcid.org/0000-0001-7286-4383
MS. Shihabi https://orcid.org/0000-0001-7290-1957
IC. Curik https://orcid.org/0000-0001-7090-1654
MS. Spehar https://orcid.org/0000-0001-7600-5009
TB. Bobić https://orcid.org/0000-0001-9975-1258
BL. Lukic https://orcid.org/0000-0003-2384-5383

Usporedba inbridinga procijenjenog na temelju pedigrea (F_{PED}) i genomskog inbridinga (F_{ROH}) u hrvatskih holstein krava

Sažetak

Dvije najčešće korištene metode za procjenu inbridinga i strukturnih odnosa u populacijama domaćih životinja temelje se na podacima iz pedigrea i genomskim podacima. Ove dvije mjere nisu izravno ekvivalentne, budući da inbriding izračunat na temelju pedigrea odražava očekivanu identičnost alela po podrijetlu, dok F_{ROH} predstavlja genomsku procjenu ostvarene autozigotnosti, iako na njega mogu utjecati gustoća SNP markera, parametri detekcije ROH segmenata te točnost imputacije genotipova. Cilj ovog istraživanja bio je procijeniti strukturu pedigrea u populaciji holstein goveda s hrvatske mliječne farme te usporediti pedigre i genomske procjene inbridinga. Podaci su prikupljeni iz nacionalnih baza proizvodnih i pedigre podataka. Nakon filtriranja, u istraživanje je uključeno 6.337 krava s ukupno 14.732 proizvodna zapisa, a pedigre je rekonstruiran kroz pet generacija. Konačni pedigre obuhvaćao je 21.779 životinja. Pedigre analiza pokazala je porast inbridinga tijekom vremena, s najvećom prosječnom vrijednošću koeficijenta inbridinga od 0,014. Procijenjena stopa porasta inbridinga iznosila je približno 0,002 (ΔF), što upućuje na umjerenu razinu genetske raznolikosti, ali i na postupno nakupljanje srodnosti kao posljedicu intenzivne selekcije. Nasuprot procjenama na temelju pedigrea, genomska procjena inbridinga temeljena na homozigotnim sljedovima (F_{ROH}) pokazala je znatno višu razinu inbridinga (prosječni $F_{ROH<1Mb}$=0,091). Izravna usporedba individualnih procjena pokazala je slabu, ali statistički značajnu pozitivnu korelaciju između Fped i F_{ROH} u životinja za koje su bile dostupne obje vrste podataka (Pearsonov r=0,118, p=0,007), što upućuje na to da ova dva pristupa zahvaćaju povezane, ali samo djelomično preklapajuće aspekte inbridinga. Dodatna analiza raspodjele ROH segmenata pokazala je da je većina genomskog inbridinga potjecala od kratkih i srednje dugih ROH segmenata ($F_{ROH1–2Mb}$=0,025; $F_{ROH2–4Mb}$=0,025; $F_{ROH4–8Mb}$=0,023). Budući da takvi ROH segmenti uglavnom odražavaju stariju srodnost, dobiveni rezultati ukazuju na ograničen doprinos novijeg inbridinga. Sveukupno, rezultati pokazuju da sam pedigre ne prikazuje u potpunosti razinu inbridinga u određenim mliječnim populacijama. Kombiniranjem pedigre i genomskih informacija moguće je dobiti cjelovitiji uvid u strukturu i povijest mliječne populacije te osigurati vrijednu osnovu za upravljanje genetskim napretkom i dugoročnim očuvanjem genetske raznolikosti u holstein goveda.

Ključne riječi: holstein krave; struktura pedigrea; inbriding, homozigotni sljedovi (ROH); genomski inbriding (F_{ROH})

References

1. Bem, R.D., Benfica, L.F., Silva, D.A., Carrara, E.R., Brito, L.F., Mulim, H.A., Borges, M.S., Cyrillo, J.N.S.G., Canesin, R.C., Bonilha, S.F.M., Mercadante, M.E.Z. (2024): Assessing different metrics of pedigree and genomic inbreeding and inbreeding effect on growth, fertility, and feed efficiency traits in a closed-herd Nellore cattle population. *BMC Genomics* 24, 738. https://doi.org/10.1186/s12864-024-10641-3
2. Biscarini, F., Cozzi, P., Gaspa, G., Marras, G. (2019). detectRUNS: An R package to detect runs of homozygosity and heterozygosity in diploid genomes. CRAN Repository.
3. Brajković, V., Duvnjak, I., Ferenčaković, M., Špehar, M., Raguž, N., Lukić, B., Curik, I., Cubric-Curik, V. (2018): The effect of DNA quality on the sequencing success of cattle. *Journal of Central European Agriculture* 19 (4), 804-809. https://doi.org/10.5513/JCEA01/19.4.2340
4. Browning, B.L., Zhou, Y., Browning, S.R. (2018): A one-penny imputed genome from Next-Generation reference panels. *American Journal of Human Genetics* 103 (3), 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015
5. Cassell, B.G., Adamec, V., Pearson, R.E. (2003): Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holseins and Jerseys. *Journal of Dairy Science* 86 (9), 2967-2976. https://doi.org/10.3168/jds.S0022-0302(03)73894-6
6. Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M., Wilson, J.F. (2018): Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics* 19, 220-234. https://doi.org/10.1038/nrg.2017.109
7. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015): Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. https://doi.org/10.1186/s13742-015-0047-8
8. Charlesworth, D., Willis, J.H. (2009): The genetics of inbreeding depression. *Nature Reviews Genetics* 10 (11), 783-796. https://doi.org/10.1038/nrg2664
9. Cole, J.B., Dürr, J.W., Nicolazzi, E.L. (2021): Invited review: The future of selection decisions and breeding programs: What are we breeding for, and who decides? *Journal of Dairy Science* 104, 5111-5124. https://doi.org/10.3168/jds.2020-19777
10. Cortés-Hernández, J., García-Ruiz, A., Vásquez-Peláez, C.G., Ruiz-Lopez, F.J. (2021): Correlation of genomic and pedigree inbreeding coefficients in small cattle populations. *Animals* 12 (11), 3234. https://doi.org/10.3390/ani11113234
11. Curik, I., Ferenčaković, M., Sölkner, J. (2014): Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science* 166, 26-34. https://doi.org/10.1016/j.livsci.2014.05.034
12. Doekes, H.P., Veerkamp R.F., Bijma P., de Jong G., Hiemstra S.J., Windig J. J. (2019): Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein-Friesian dairy cattle. *Genetics Selection Evolution* 51, 54. https://doi.org/10.1186/s12711-019-0497-z
13. Falconer, D.S., Mackay, T.F.C. (1996): Introduction to quantitative genetics. 4th Edition, Prentice Hall, Essex.
14. Ferenčaković, M., Sölkner, J., Curik, I. (2013): Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* 45 (1), 42. https://doi.org/10.1186/1297-9686-45-42
15. Gantner, V., Steiner, Z., Jožef, I., Gantner, R., Solić, D., Potočnik, K. (2024): The variability in cow's recovery potential regarding the milk recording season. *Mljekarstvo* 74 (2), 156-165. https://doi.org/10.15567/mljekarstvo.2024.0206
16. Groeneveld E., Westhuizen B. van der, Maiwashe A., Voordewind F., Ferraz, J.B.S. (2009). POPREP: a generic report for population management. *Genetics and Molecular Research* 8 (3), 1158-1178. https://doi.org/10.4238/vol8-3gmr648

17. Halvonik, A., Moravčíková, N., Chalupková, M., Kasarda, R. (2025): Commonly used genomic estimators of individual inbreeding in livestock. *Czech Journal of Animal Science* 69 (7), 269-279.
https://doi.org/10.17221/91/2024-CJAS
18. Kardos, M., Luikart, G., Allendorf, F.W. (2015): Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* 115, 63-72.
https://doi.org/10.1038/hdy.2015.17
19. Leroy G., Mary-Huard T., Verrier E., Danvy S., Charvolin E., Danchin-Burge C. (2013): Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. *Genetics Selection Evolution* 45 (1), 1.
https://doi.org/10.1186/1297-9686-45-1
20. Luštrek, B., Šimon, M., Turk, K., Bogičević, S., Potočnik, K. (2025): Comparing genomic and pedigree inbreeding coefficients in the Slovenian Lipizzan horse as a case study for small closed populations. *Animals* 15, 2774.
https://doi.org/10.3390/ani15192774
21. Makanjuola B. O., Miglior F., Abdalla E. A., Maltecca C., Schenkel F. S., Raes, C.F. (2020): Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of Dairy Science* 103 (6), 5183-5199.
https://doi.org/10.3168/jds.2019-18013
22. Mugambe, J., Ahmed, R.H., Thaller, G., Schmidtmann, C. (2024): Impact of inbreeding on production, fertility and health traits in German Holstein dairy cattle. *Journal of Dairy Science* 107 (7), 4714-4725.
https://doi.org/10.3168/jds.2023-23728
23. Nishio, M., Inoue, K., Ogawa, S., Ichinoseki, K., Arakawa, A., Fukuzawa, Y., Okamura, T., Kobayashi, E., Taniguchi, M., Oe, M., Ishii, K. (2023): Comparing pedigree and genomic inbreeding coefficients, and inbreeding depression of reproductive traits in Japanese Black cattle. *BMC Genomics* 24, 376.
https://doi.org/10.1186/s12864-023-09480-5
24. Purfield, D.C., Berry, D.P., McParland, S., Bradley, D.G. (2012): Runs of homozygosity and population history in cattle. *BMC Genetics* 13, 70.
https://doi.org/10.1186/14712156-13-70
25. R Core Team (2023): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
26. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., Delaneau, O. (2021): Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics* 53 (1), 120-126.
https://doi.org/10.1038/s41588-020-00756-0
27. Salehi R., Javanmard, A., Mokhber, M., Alijani, S. (2025): Estimating linkage disequilibrium and effective population size across generations in Holstein cattle. *Veterinary Medicine and Science* 11 (6), e70684.
https://doi.org/10.1002/vms3.70684
28. Sargolzaei, M., H. Iwaisaki and J.J. Colleau. 2006. CFC: A tool for monitoring genetic diversity. Proc. 8th World Congress on Genetics Applied to Livestock Production. CD-ROM Communication 27-28.
29. Sarviaho, K., Uimari, P., Martikainen, K. (2023): Estimating inbreeding rate and effective population size in the Finnish Ayrshire population in the era of genomic selection. *Journal of Animal Breeding and Genetics* 140 (3), 343-353.
https://doi.org/10.1111/jbg.12762
30. Špehar, M., Ivkić, Z., Vranić, I., Crnčić, J., Dražić, M., Pašalić, D., Barać, Z. (2019): Opportunities of implementing the genomic selection in small populations - the Croatian case. *Mljekarstvo* 69 (2), 86-97.
https://doi.org/10.15567/mljekarstvo.2019.0201
31. Usman, T., Yu, Y., Liu, C., Fan, Z., Wang, Y. (2014): Comparison of methods for high quantity and quality genomic DNA extraction from raw cow milk. *Genetics and Molecular Research* 13 (2), 3319-3328.
https://doi.org/10.4238/2014.April.29.10

32. Visscher, P.M., Wooliams, J.A., Smith, D., Williams, J.L. (2002): Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of Dairy Science* 85, 2368-2375.
    https://doi.org/10.3168/jds.S0022-0302(02)74317-8
33. Vostrý L., Vostra-Vydrova, H., Moravcikova, N., Kasarda, R., Cubric-Curik, V., Brzakova, M., Solkner, J., Shihabi, M., Moreno, J.A.H., Spehar, M., Curik, I. (2023): Genomic diversity and population structure of the Czech Holstein cattle. *Livestock Science 273, 105261*.
    https://doi.org/10.1016/j.livsci.2023.105261
34. Wright, S. (1922): Coefficients of inbreeding and relationship. *The American Naturalist* 56 (645), 330-338.
35. Zhang, Q., Calus, M.P., Guldbrandtsen, B., Lund, M.S., Sahana, G. (2015): Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genetics* 16, 88.
    https://doi.org/10.1186/s12863-015-0227-7