Animal 17 (2023) 100793

Contents lists available at ScienceDirect

Animal
The international journal of animal biosciences

Estimation of inbreeding, between-breed genomic relatedness and definition of sub-populations in red-pied cattle breeds

H. Wilmot [a,b], T. Druet [a,c], I. Hulsegge [d,e], N. Gengler [a,b], M.P.L. Calus [d,*]

[a] National Fund for Scientific Research (F.R.S.-FNRS), Rue d'Egmont 5, B-1000 Brussels, Belgium
[b] TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, B-5030 Gembloux, Belgium
[c] Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34), Avenue de l'Hôpital 1, B-4000 Liège, Belgium
[d] Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands
[e] Centre for Genetic Resources, The Netherlands, Wageningen University & Research, Droevendaalsesteeg 1, 6700AH Wageningen, the Netherlands

ARTICLE INFO

Article history:
Received 12 October 2022
Revised 16 March 2023
Accepted 21 March 2023
Available online 30 March 2023

Keywords:
Across-breed
Genetic diversity
Local breeds
Reference population
Stratification

ABSTRACT

Currently, enhancing the collaboration between related breeds is of main importance to increase the competitivity and the sustainability of local breeds. One type of collaboration is the development of an across-breed reference population that will allow a better management of local breeds. For this purpose, the genomic relatedness between the local target breed and possible breeds to be included in the reference population should be estimated. In Europe, there are several local red-pied cattle breeds that would benefit from this kind of collaboration. However, how different red-pied cattle breeds from the Benelux are related to each other and can collaborate is still unclear. The objectives of this study were therefore: (1) to estimate the level of inbreeding of the East Belgian Red and White (**EBRW**), the Red-Pied of the Ösling (**RPO**) and Dutch red-pied cattle breeds; (2) to determine the genomic relatedness of several red-pied cattle breeds, with a special focus on two endangered breeds: the EBRW and the RPO, and (3) based on the second objective, to detect animals from other breeds that were genomically close enough to be considered as advantageous in the creation of an across-breed reference population of EBRW or RPO. The estimated inbreeding levels based on runs of homozygosity were relatively low for almost all the studied breeds and especially for the EBRW and RPO. This would imply that inbreeding is currently not an issue in these two endangered breeds and that their sustainability is not threatened by their level of inbreeding. The results from the principal component analysis, the phylogenetic tree and the clustering all highlighted that the EBRW and RPO breeds were included in the genomic continuum of the studied red-pied cattle breeds and can be therefore considered as genomically close to Dutch red-pied cattle breeds, highlighting the possibility of a collaboration between these breeds. Especially, EBRW animals were closely related to Deep Red and Improved Red animals while, to a lesser extent, the RPO animals were closely related to the Meuse-Rhine-Yssel breed. Based on these results, we could use distance measures, based either on the principal component analysis or clustering, to detect animals from Dutch breeds that were genomically closest to the EBRW or RPO breeds. This will finally allow the building of an across-breed reference population for EBRW or RPO for further genomic evaluations, considering these genomically closest animals from other breeds.

Implications

To stay competitive against mainstream breeds, the development of a genomic evaluation system for local breeds is necessary worldwide, which requires an across-breed reference population to be built. For this purpose, the genomic relatedness of the endangered East Belgian Red and White and the Red-Pied of the Ösling with red-pied cattle breeds from the Netherlands was estimated. These two breeds were part of a continuum of red-pied breeds from Benelux countries. Animals from Dutch breeds that were genomically close to them were detected. They can be considered when building an across-breed reference population for those endangered breeds.

* Corresponding author.
E-mail address: mario.calus@wur.nl (M.P.L. Calus).

https://doi.org/10.1016/j.animal.2023.100793

Introduction

Local breeds are of main importance because they are often seen as a reservoir of unique genetic material and as adapted to their environment (Medugorac et al., 2009). They are in general more resilient, fertile and with better health performances (e.g. udder health or calving ease) than specialised breeds. However, most of the time, local breeds are genetically inferior to mainstream breeds for production traits and the routine use of genomics in mainstream breeds, while still uncommon in local breeds with small reference populations, can increase the genetic gap between them. It is therefore of main importance to enhance the competitiveness of local breeds against mainstream breeds (Hiemstra et al., 2010; Wellmann and Bennewitz, 2019) and to improve their genetic gain (Marjanovic et al., 2021).

To improve the genetic gain of a breed, it can be advocated to develop a genomic evaluation system which relies on a relevant reference population. Reference populations should meet two criteria to allow precise and reliable genomic predictions: (1) The reference population should be of sufficient size (Goddard, 2009); and (2) The candidate animals to be evaluated should be genetically related to animals from the reference population (Wientjes et al., 2013). As the first criteria can hardly be met in the case of an endangered breed, building an across-breed (or joint) reference population can help to overcome this issue. However, to be effective, and meet the second criterion, an across-breed reference population should use related breeds. If the breeds are closely related, a higher increase in the accuracy of the genomic prediction is expected and, therefore, the addition of another breed in the reference population is more valuable (Hozé et al., 2014). Estimates of genetic relatedness can help to decide which breeds can be included in the reference population of the target breed. Especially, the effective number of chromosome segments has been commonly used because the number of animals from the related breed that is equivalent to the addition of one animal of the target breed in the reference population can be derived from it (Wientjes et al., 2016; Marjanovic and Calus, 2021). Another solution is to detect, inside the related breed, the animals that are the most genomically related to the target breed. One example is given by Rezende et al. (2020) who detected, based on the first principal component of a principal component analysis (**PCA**), animals from a local population that were genomically close enough to animals from the international population.

The development of an across-breed reference population can be interesting for the diverse local red-pied cattle breeds found in Europe. Indeed, these breeds have very different population sizes, from mainstream breeds such as the Meuse-Rhine-Yssel (**MRY**) to local breeds (van Breukelen et al., 2019; Schmidtmann et al., 2021; Wilmot et al., 2022). Several studies have demonstrated the genetic relatedness of Dutch red-pied breeds (van Breukelen et al., 2019; Marjanovic et al., 2021; Schmidtmann et al., 2021), but none of them included the endangered East Belgian Red and White (**EBRW**) and the Red-Pied of the Ösling (**RPO**), which are other red-pied breeds. Therefore, how the EBRW and RPO can collaborate with other red-pied breeds for the development of an across-breed reference population is currently unknown. The knowledge of the genetic diversity within and between these breeds will allow to get a better insight of their history as well as to improve their management for conservation purposes (Gómez-Romano et al., 2013), e.g. by conservation through utilisation (Slagboom et al., 2022). Among other analyses, PCA, fixation index values, phylogenetic trees or unsupervised clustering are commonly used to estimate between-breed diversity, to position studied breeds against each other, and to detect possible admixture events. For within-breed genetic diversity, the estimation of inbreeding coefficients is valuable. If inbreeding coefficients within a breed are high, it can be decided to crossbreed it with a related breed to lower its average inbreeding level and sustain it. Low inbreeding coefficients can then be related, if confirmed by other parameters, with admixture events. The proportion of genome involved in runs of homozygosity, which are stretches of homozygous markers, can give a valuable estimation of inbreeding coefficients. They can be estimated by a rule-based method, which means that empirical thresholds are used to characterise runs of homozygosity, e.g., by the definition of the minimum number of homozygous single nucleotide polymorphism (**SNP**), the maximum number of missing SNPs or heterozygous SNPs. The runs of homozygosity are often considered as pairs of haplotype segments inherited from one ancestor and therefore referred to as homozygous by descent segments. Probabilistic approaches based on a hidden Markov model have been proposed to identify homozygous by descent segments. For example, Druet and Gautier (2017) developed such a hidden Markov model with several states associated to homozygous by descent positions, corresponding to homozygous by descent segments of different lengths associated with different group of ancestors (present in different past generations).

The objectives of this study were therefore: (1) to compare inbreeding levels of eight red-pied cattle breeds using runs of homozygosity and homozygosity by descent probabilities, (2) to determine the genomic relatedness between the EBRW, the RPO and Dutch red-pied breeds, and (3) by the definition of subpopulations, to detect which animals from other breeds would be valuable to add when defining a reference population for the EBRW and RPO breeds for further genomic evaluations. The achievement of these objectives will allow to lay the foundation for a thoughtful collaboration of red-pied cattle breeds from Benelux countries.

Material and methods

Dataset and quality control

A total of eight cattle breeds were considered in this study: the Dutch belted (**DB**, n = 16), the Dutch Friesian (**DFR**, n = 51), the Deep Red (**DR**, n = 21), the EBRW (n = 226), the Groningen White Headed (**GWH**, n = 36), the Improved Red (**IR**, n = 21), the MRY (n = 292) and the RPO (n = 132). For this latter breed, all animals recorded in the Herd Book are coming from a single farm, and therefore, all genotyped animals were also coming from that farm. The studied RPO animals were therefore representing well the breed, as defined by its Herd Book. All the sampled animals had a red-pied colour pattern, even if, for the DB and DFR, black-pied animals also exist. Different chips were used for genotyping these animals: the BovineSNP50 Beadchip v2 and 3, the BovineHD Beadchip v12 and the EuroG MD v9-SI and v2 (Illumina, San Diego, CA, USA). Supplementary Table S1 provides the distribution of the different chips used for genotyping the different breeds. Genotype data from Dutch breeds (DB, DFR, DR, GWH and IR) were provided by the Centre of Genetic Resources (Wageningen, the Netherlands), those from the EBRW by the Walloon Breeders Association (Ciney, Belgium) and those from the RPO by the Administration of Technical Agricultural Services (Luxembourg, Luxembourg). The *Bos taurus* genome reference assembly ARS-UCD1.2 (Rosen et al., 2020), available on the website of Schnabel (2019), was used for mapping the available SNPs. The quality control was performed on the merged dataset, containing all the studied breeds. Before quality control, 40 695 SNPs mapped on autosomes were in common between the different chips and the lowest animal call rate was 87%. For quality control, SNPs with genotype call rate under 90% were discarded with Plink v.1.9 (Purcell and Chang, 2019). Then, the remaining missing SNP values were imputed with Beagle

v.4.0 (Browning and Browning, 2007) and SNPs with a predicted imputation accuracy (r-squared) lower than 0.8 were discarded with bcftools v.1.15 (Danecek et al., 2021). After quality control, a total of 39 967 SNPs remained and were used for further analysis.

Runs of homozygosity

To have an insight of the inbreeding status of each of the eight breeds under study, homozygous by descent segments and runs of homozygosity were computed. For these analyses, all the available samples were kept. A model-based approach using a hidden Markov model was used to compute homozygosity by descent. This approach relied on the multiple homozygous by descent class model proposed by Druet and Gautier (2017) and implemented in the RZooROH package v.0.3.1 (Bertrand et al., 2019). We used a model with nine homozygous by descent classes, modelled as nine nested layers of ancestors (Druet and Gautier, 2022), with the rates R_k set to {2, 4, 8, 16, 32, 64, 128, 256, 512}. The rate from homozygous by descent-class *k* corresponds approximately to twice the number of generations to the common ancestors associated with the homozygous by descent segments in that layer (Druet and Gautier, 2017). The last class regrouped non-homozygous by descent segments of the genome as well as more ancient homozygous by descent segments associated with more remote ancestors. This approach allowed an estimation of the proportion of the genome associated with each of the defined homozygous by descent classes as well as of the genomic inbreeding coefficient based on these classes.

A rule-based method was implemented using Plink v.1.9 (Purcell and Chang, 2019) by applying guidelines suggested by Meyermans et al. (2020) to estimate runs of homozygosity. Accordingly, there was no linkage disequilibrium pruning nor minor allele frequency pruning prior to the run of homozygosity analysis. The formula of Lencz et al. (2007), adapted by Purfield et al. (2012), was used to determine the minimal number of SNPs in runs of homozygosity, symbolised *L*, and the size of the scanning window:

$$L = \frac{\log_e \frac{\alpha}{n_s n_i}}{\log_e(1 - \overline{het})}$$

with α the percentage of false positive runs of homozygosity, set to 0.05, n_s the total number of SNPs available, n_i the number of individuals and $\overline{het}$ the mean heterozygosity across all SNPs.

The minimal number of SNPs in run of homozygosity and the size of the scanning window were therefore equal to 52. The maximum gap between two consecutive SNPs to be included in runs of homozygosity was fixed to 500 kilobases (**kb**), as suggested by Meyermans et al. (2020). Moreover, as the average marker density was around 15 SNPs/Mb and the minimal number of SNPs in run of homozygosity was equal to 52, the expected average size of run of homozygous segments was around 3.5 Mb. Therefore, the minimum size of run of homozygosity considered was fixed to 4 Mb. Using this minimum size is also warranted to limit the detection of false positive ROHs when using ~40 k SNP (Alemu et al., 2021). Heterozygotes in the scanning window and, thus, in runs of homozygosity were not allowed. As the genotypes were imputed, no missing genotypes were present, and there was no need to consider those in the run of homozygosity analyses. To be considered a run of homozygosity, segments must have on average one SNP per 75 kb, as suggested by Meyermans et al. (2020). The scanning window threshold (*t*) was computed according to Meyermans et al. (2020):

$$t = \text{floor}\left(\frac{N_{out} + 1}{L}, 3\right)$$

with *L* the scanning window previously defined and N_{out} the desired number of final outer SNPs on either side of the homozygous segment that should not be included in the final run of homozygosity. Three decimals were accounted for defining *t*. In this study, *t* equals 0.058 as a N_{out} of two was chosen. For visualisation of results, R v.4.2.2 (R Core Team, 2022) and Rstudio 2022.02.1 + 461 (R Studio Team, 2022) were used.

Genomic relatedness of breeds

The genomic relatedness of the eight breeds under study was analysed through three different approaches: a PCA, a fixation index and an admixture analysis. To avoid undesirable effects of oversampling on these three approaches (Lawson et al., 2018), a random sample of 50 animals of each of the four breeds with more than 50 samples (DFR, EBRW, MRY and RPO) was used, similar to Schmidtmann et al. (2021). The PCA was based on the matrix of correlations of genotypes and was computed with the FactomineR v.2.4 R package (Lê et al., 2008). The pairwise Weir & Cockerham's fixation index values (Weir and Cockerham, 1984) were computed with Plink v.1.9 (Chang et al., 2015; Purcell and Chang, 2019). Based on fixation index values, the unweighted pair group method with arithmetic mean was used to build a phylogenetic tree with the phangorn 2.7.1 R package (Schliep et al., 2017). Finally, an unsupervised clustering was performed with the ADMIXTURE v.1.3.0 software (Alexander et al., 2009). The optimised number of clusters was determined based on a 10-fold cross-validation for K = 1–10 clusters. To avoid bias in the unsupervised clustering, linkage disequilibrium pruning was performed using Plink v.1.9 (option --indep-pairwise 50 5 0.8) on the overall dataset, leading to a reduced number of 37 603 SNPs. The position of individuals on principal components, the phylogenetic tree and the proportions of individuals to each of the K clusters were visualised with R v.4.2.2 (R Core Team, 2022) and Rstudio 2022.02.1 + 461 (R Studio Team, 2022).

Detection of sub-populations

The concept behind the third objective is that there is a stratification within each breed and that not all the animals of a breed are equivalent in the development of an across-breed reference population. To evaluate which animals from other breeds could be included in reference populations of EBRW and RPO for future genomic evaluations, three different approaches of the determination of sub-populations were computed: (1) Weighted Euclidean distances of animals to EBRW and RPO centroids based on principal components explaining 95% of the variance, (2) Weighted Euclidean distances of animals to EBRW and RPO centroids based on the first four principal components, and (3) Euclidean distances of animals to EBRW and RPO mean proportions for ADMIXTURE defined-clusters. For each of these approaches, RPO animals were not considered when establishing the list of closest animals to the EBRW breed, and vice-versa.

The first approach was based on a PCA using the matrix of correlations of genotypes (option "scale.unit = TRUE" of the PCA function of the FactomineR v.2.4 R package, Lê et al., 2008), considering all the available samples and the principal components explaining altogether at least 95% of the total variance, i.e., in this study, 632 principal components. For EBRW and RPO, the mean coordinates of genotypes of all animals belonging to one of these two breeds were computed, considering the first 632 principal components. These mean coordinates were considered as centroids for EBRW and RPO. Centroids were computed as follows, for each of the two breeds separately and for j = 1–632 principal components:

$$centroid_j = \frac{\sum_{i=1}^{n} coordinate_{j,i}}{n}$$

where centroid$_j$ is the coordinate of the centroid for the principal component j, i is the animal considered, n is the total number of animals of the EBRW or RPO breeds and coordinate$_{j,i}$ is the coordinate of the animal i on the principal component j. Each centroid therefore had 632 coordinates, one for each principal component. Following this, the distances of each animal to these centroids were computed as follows:

$$d_i = \sqrt{\sum_{j=1}^{632} \left(\left(coordinate_{i,j} - centroid_j \right)^2 * \ var_j \right)}$$

where i is the animal for which the distance is computed, j is the principal component considered, coordinate$_{i,j}$ is the coordinate of the animal i on the principal component j, centroid$_j$ is the coordinate of the centroid for the principal component j and var$_j$ is the proportion of the variance explained by principal component j. The mean of the distances of genotypes from all animals, including EBRW and RPO animals, to centroids of EBRW and RPO was used as an empirical threshold to determine if animals from other breeds could potentially be included in the reference population of EBRW and RPO breeds. If the distance of the genotype of an animal to the centroid of EBRW or RPO was below the threshold, and if it does not belong to the EBRW or RPO breed, it was included in the list of animals that could be used in future reference populations of EBRW or RPO.

The second approach was also based on a PCA using the matrix of correlations of genotypes and considering all the available samples. However, for this second approach, only the first four principal components explaining altogether 7.78% of the total variance were used. The number of principal components to consider was evaluated based on the number of principal components after which there is a stabilisation of the percentage of the variance explained. Then, exactly the same steps as in the first approach were followed for the computation of centroids, distances and threshold. The only difference was that four principal components were considered, and not 632.

The last approach was based on ADMIXTURE results obtained when all the samples were considered and without any linkage disequilibrium pruning. It was indeed considered that SNPs in linkage disequilibrium would give important information about stratification within the breed and therefore about which animals could be included in the reference population of EBRW and RPO breeds. Mean ADMIXTURE proportions for K = 8 clusters were computed for the EBRW and the RPO breeds, as follows:

$$\bar{p}_k = \frac{\sum_{i=1}^{n} p_{k,i}}{n}$$

where $\bar{p}_k$ is the mean proportion of the ADMIXTURE cluster k, i is the animal considered, n is the total number of animals for the EBRW or RPO breeds, and $p_{k,i}$ is the proportion of the ADMIXTURE cluster k for the animal i. Similarly to the other approaches, Euclidean distances of each of the animals to these mean proportions were determined:

$$d_i = \sqrt{\sum_{k=1}^{8} \left(p_{i,k} - \bar{p}_k \right)^2}$$

where i is the animal for which the distance is computed, k is the cluster considered, $p_{i,k}$ is the proportion of the ADMIXTURE cluster k for the animal i and $\bar{p}_k$ is the mean proportion of the ADMIXTURE cluster k for the breed considered. A threshold was then defined similarly to the previous approaches (mean of distances). Animals to be included in the across-breed reference population of EBRW and RPO were listed based on this threshold. To compare the three different approaches, the overlap of selected animals by the different approaches was evaluated and presented in a Venn diagram.

Results

Runs of homozygosity

Two different approaches were used to estimate autozygosity in each of the eight breeds under study: (1) a hidden Markov model-based approach and (2) a rule-based approach. Fig. 1 shows, for each of the breeds, the proportion of the genome associated with the different homozygous by descent classes of our hidden Markov model. In general, inbreeding levels were relatively low in the different breeds. However, higher proportions of the genome were associated to homozygous by descent classes with rates R_k equal to 16, for GWH and DB, and to 32, for GWH (roughly equivalent to common ancestors present 8 and 16 generations in the past), in comparison to other breeds. Most breeds also showed higher proportions of homozygous by descent segments at rate 512 compared to other rates. However, considering the SNP density all over the genome, these homozygous by descent segments should be interpreted cautiously. Moreover, homozygous by descent segments at rate 512 would have been related to very ancient inbreeding events appearing before the creation of the breeds. The IR breed presented lower inbreeding levels in general. Supplementary Fig. S1 shows, for each breed, the average total length of runs of homozygosity in different run of homozygosity categories defined according to their length, based on the rules defined for the second approach. Similar patterns to Fig. 1 were observed, and similar conclusions could be drawn.

Fig. 2 allows to observe the variability of the genomic inbreeding coefficient based on homozygous by descent segments within each of the eight breeds for different base populations (by considering only classes with a rate R_k lower than a threshold T as homozygous by descent). As previously, higher inbreeding levels can be observed for the GWH animals while DR, EBRW, IR, MRY and RPO animals had on average at rate 512 a value around 0.10 for the genomic inbreeding coefficient based on homozygous by descent segments. In general, the individual genomic inbreeding coefficients based on homozygous by descent segments were relatively close to the average pattern observed for the breed. However, for the MRY breed, more variability was observed, partly because more samples were available than for most of the breeds,

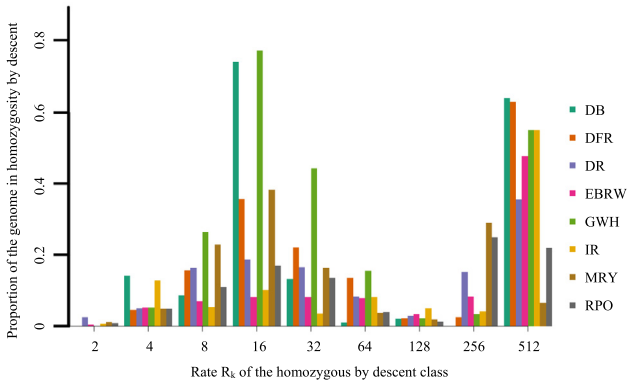

Fig. 1. Proportion of the genome in different homozygous by descent classes with different R_k rates for different cattle breeds. Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

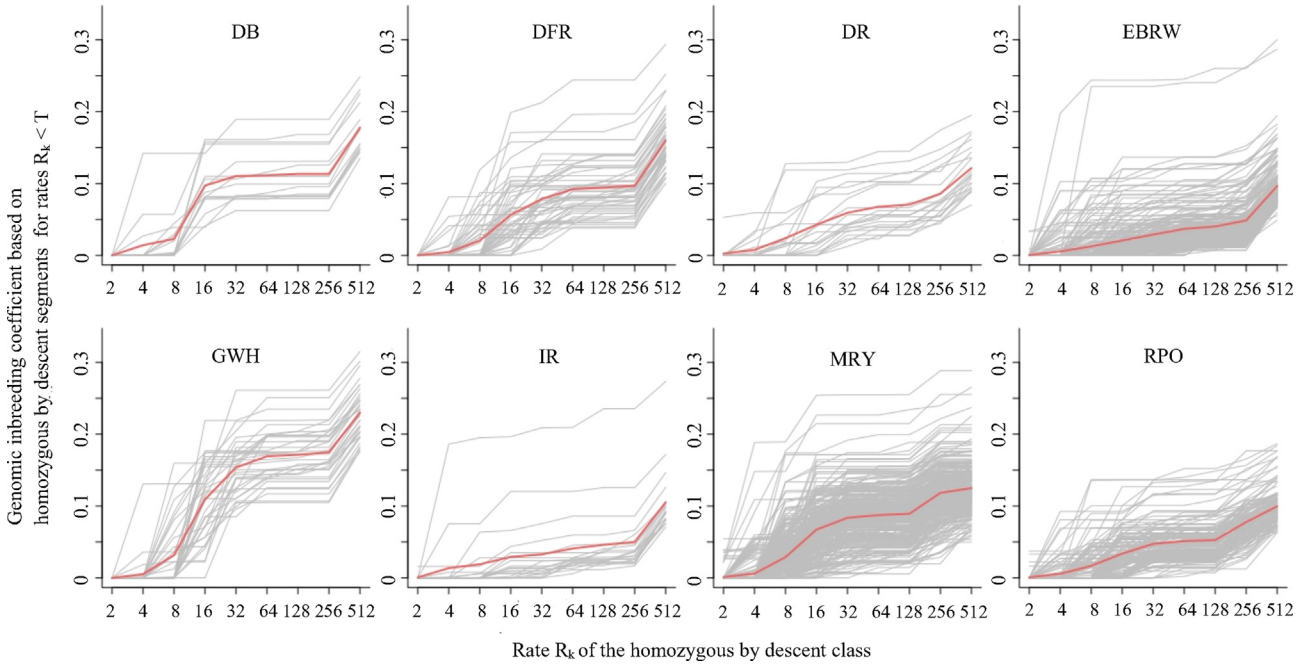

Fig. 2. Individual (in grey) and average (in red) genomic inbreeding coefficients per cattle breed. The genomic inbreeding coefficients were estimated with respect to different base populations by including only homozygous by descent classes with a rate R_k lower or equal than a threshold T (setting the reference population approximately 0.5*T generations in the past). Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

with the birthdate of MRY animals spanning a longer timeframe than other breeds, and partly because the breed is more mainstream, having higher headcounts. The EBRW and IR breeds showed two and one animals, respectively, with higher inbreeding levels compared to other animals of the same breed. Moreover, these homozygous by descent segments were longer, captured by the homozygous by descent classes with rates R_k equal to 4 or 8, suggesting recent inbreeding associated with ancestors present approximately two to four generations in the past.

Fig. 3 shows for each animal of each breed the number of runs of homozygosity comparatively to the genomic inbreeding coefficient, both estimated with the rule-based approach. Again, the MRY, EBRW and RPO breeds showed similar patterns, with most animals having a genomic inbreeding coefficient under 0.15 and less than 30 runs of homozygosity. For the MRY and EBRW breeds, a few animals have a genomic inbreeding coefficient of approximately 0.20. For EBRW, the two animals exhibiting recent inbreeding levels again clearly separated from the others. The IR breed also showed low genomic inbreeding coefficients and a similar pattern to MRY, EBRW and RPO, even if it is less obvious to observe as the sample size was smaller. The GWH breed showed again higher levels of inbreeding as only one animal from this breed had a genomic inbreeding coefficient lower than 0.05. The slope of the DFR animals in Fig. 3 appeared to be the steepest which would mean that high genomic inbreeding coefficients in this breed are due to more distant inbreeding than for other breeds, as confirmed by Figs. 1 and 2.

Genomic differentiation of breeds

To visualise the differentiation between the eight breeds under study, a PCA was realised (Fig. 4, Supplementary Fig. S2). The first component of this PCA explained 4.29% of the variation while the second one explained 3.45% of the variation. The first principal component allowed to clearly separate GWH animals from other breeds while the second component separated DB and DFR animals from other breeds. The DFR, DB and GWH also appeared to be clearly distinct from each of the other breeds on the phylogenetic tree (Fig. 5); the GWH breed being the most distant compared to other breeds, as already seen in Fig. 4.

The remaining breeds, namely IR, EBRW, DR, RPO and MRY, formed a continuum of breeds on the second principal component, highlighting their genomic proximity (Fig. 4). Similar conclusions can be drawn from the phylogenetic tree (Fig. 5); IR and EBRW were the most closely related breeds and DR animals appeared to be closely related to them. Another sub-group was composed of MRY and RPO animals, even if they appeared to be less closely related than the sub-group of IR, EBRW and DR animals. These two sub-groups were also closely related, forming the continuum of breeds previously seen on the PCA.

The results of the ADMIXTURE analysis can be observed in Fig. 6 for 2–8 clusters. The optimal number of clusters based on the 10-fold cross-validation was found to be eight, which is the number of studied breeds (Supplementary Fig. S3). When defining two clusters, the GWH appeared to be the first breed to be differentiated, supporting previous results. Three clusters allowed to differentiate the DFR from the other breeds while five clusters were necessary to differentiate the DB from other breeds. The MRY and RPO breeds showed similar patterns of admixture until the definition of four clusters. The IR and EBRW showed very similar patterns of admixture for any number of clusters while the DR started to show a different partitioning when using eight clusters. However, the DR breed was not clearly distinguished from IR and EBRW based on ADMIXTURE results. All of these results confirmed those obtained from the phylogenetic tree in Fig. 5.

Detection of sub-populations

Three different approaches were compared to determine sub-populations among the different breeds with the purpose to identify animals that could be included in the reference populations of EBRW and RPO for genomic evaluations. The EBRW animals were

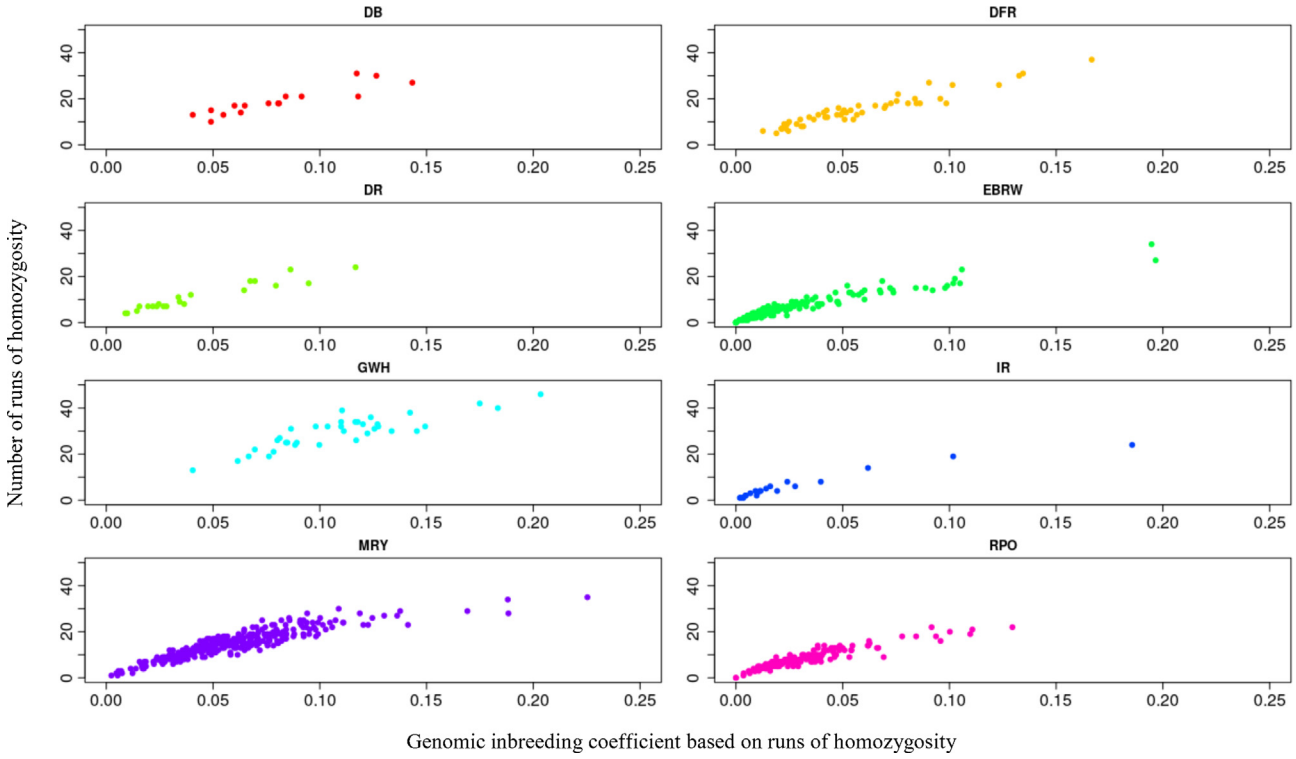

Fig. 3. For each of the eight studied cattle breeds, a scatter plot of the number of runs of homozygosity on genomic inbreeding coefficient, defined with the rule-based approach. Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

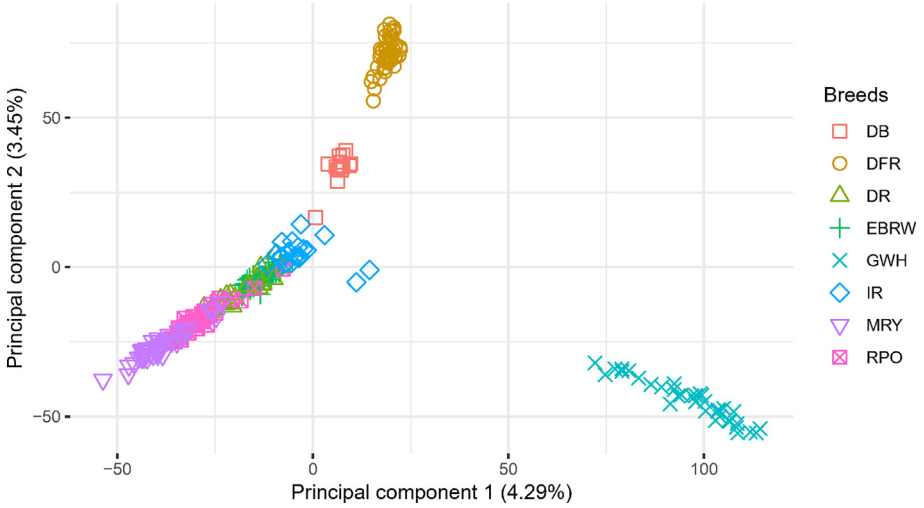

Fig. 4. First two dimensions of a principal component analysis for the eight cattle breeds under study. Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

excluded from the analysis for the RPO breed and vice-versa. The three approaches used were: (1) Weighted Euclidean distances from means of EBRW or RPO breeds on a PCA explaining 95% of the total variance, (2) A variation of the first approach using only the first four principal components explaining altogether 7.78% of the variance, and (3) Euclidean distances from proportion averages to K = 8 clusters defined by ADMIXTURE. Table 1 shows the proportion of each of the Dutch breeds that can be considered as genomically close to EBRW and RPO breeds, respectively, while Fig. 7 shows the overlap between the three different approaches for the EBRW and RPO breeds.

The three approaches were mostly in agreement regarding the list of animals to be considered as genomically related to the RPO and EBRW breeds. For the EBRW, the approach based on four principal components led to a lower number of animals considered as genomically close to this breed compared to the two other approaches. For this second approach, there was a higher overlap with the ADMIXTURE-based approach than with the other approach based on a PCA. This is probably related to the higher number of animals detected as close by the ADMIXTURE-based approach compared to the approach based on a PCA explaining 95% of the total variance. For RPO, animals selected by the

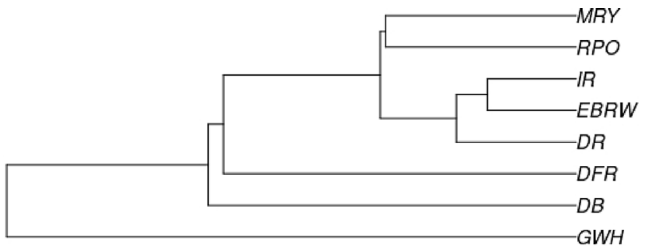

Fig. 5. Phylogenetic tree of the eight cattle breeds under study based on pairwise Weir & Cockerham's fixation index values. Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

approach based on four principal components were completely included in the approach based on a PCA explaining 95% of the total variance (Fig. 7). The overlap between the approach based on four principal components and the one based on ADMIXTURE was lower for the RPO than the EBRW breed.

From Table 1, it can be observed that all the DR and IR animals could potentially be included in a genomic reference population of the EBRW breed, regardless of the approach used. The proportion of DB animals to be included ranged from 37.50% with the first approach to 100% with the second approach. The proportion of MRY to be included in the across-breed reference population of EBRW ranged from almost 11% (second approach) to almost 19% (third approach). For the RPO breed, all the DR were considered as genomically close enough, as well as a very high proportion of the IR samples, for all the approaches. A high proportion of the MRY samples, ranging from a little less than 30% to a bit more than 40%, were considered as genomically close enough to be included in the reference population of the RPO breed. An increase of the MRY animals that were considered genomically close to the RPO breed compared to EBRW was expected based on the results obtained in the previous section. Most animals considered to be potentially included in the reference population of the RPO breed were MRY, and this was the case for all three approaches. It should also be noticed that, as animals selected by the second approach were completely included in the list of animals established by the first approach, the proportion of the MRY samples detected as genomically close enough to the RPO breed increased in the first approach compared to the second one.

Discussion

Inbreeding and between-breed genomic diversity

In general, inbreeding levels were relatively low, especially for the EBRW and RPO breeds, highlighting the possibility of admixture events which were confirmed by the clustering analysis. The PCA, the phylogenetic tree based on fixation index values and the ADMIXTURE clusters agreed about the relatedness of these eight breeds. A continuum of red-pied breeds was detected and, inside this continuum, the close relationship between EBRW, IR and DR, and to a lesser extent between RPO and MRY. These results are highlighting the possibility to develop an across-breed reference population for the EBRW and RPO breeds. For the three analyses, 50 animals for each breed with more than 50 samples were randomly sampled to avoid an undesirable effect of unbalanced sample size. This was particularly important for the ADMIXTURE analysis, which is highly sensitive to oversampling (Lawson et al., 2018). However, for the PCA and fixation index values, the results were approximately the same (results not shown), which means that the random samples were representative of the breed.

Inbreeding levels based on homozygous by descent segments and runs of homozygosity

The results were consistent between the two models used for the estimation of inbreeding coefficients (hidden Markov model and rule-based model). It therefore highlights that both methods provided similar estimators of the inbreeding coefficient in cattle populations when using a medium-density genotyping array, as previously observed by Alemu et al. (2021), Meyermans et al. (2020) or Solé et al. (2017). In the case of the rule-based method, adapted parameters should be chosen to optimise the results, as

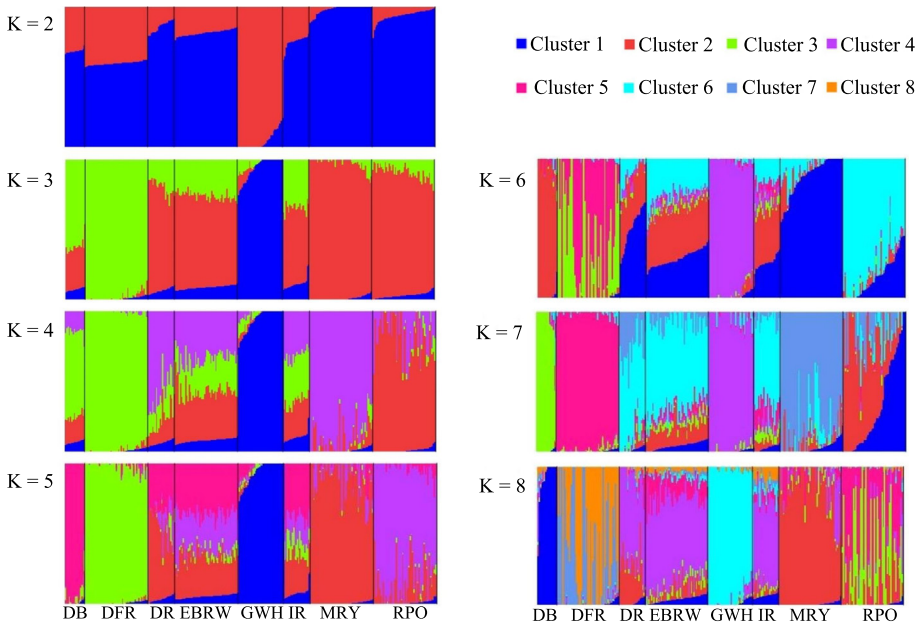

Fig. 6. Clustering results based on ADMIXTURE software for K = 2–8 clusters of the eight cattle breeds under study. Abbreviations: DB = Dutch belted; DFR = Dutch Friesian; DR = Deep Red; EBRW = East Belgian Red and White; GWH = Groningen White Headed; IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.

Table 1
Proportion (%) of the sampled Dutch breeds considered as genomically close to the East Belgian Red and White and Red-Pied of the Ösling cattle breeds based on the three approaches for defining sub-populations.

		Approaches		
Reference breeds	Breeds to be included[1]	PCA based on 95% of variance	PCA based on 4 PCs	ADMIXTURE-based
EBRW	DB	37.50	100	93.75
	DR	100	100	100
	IR	100	100	100
	MRY	17.81	10.96	18.84
	Absolute number of animals	100	90	112
RPO	DB	0	0	6.25
	DR	100	100	100
	IR	85.71	85.71	100
	MRY	42.12	28.77	28.08
	Absolute number of animals	162	123	125

Abbreviations: PCA = Principal component analysis; EBRW = East Belgian Red and White; DB = Dutch Belted; DR = Deep Red IR = Improved Red; MRY = Meuse-Rhine-Yssel; RPO = Red-Pied of the Ösling.
[1] Breeds that can be included in the reference population of the reference breed for further genomic evaluations.

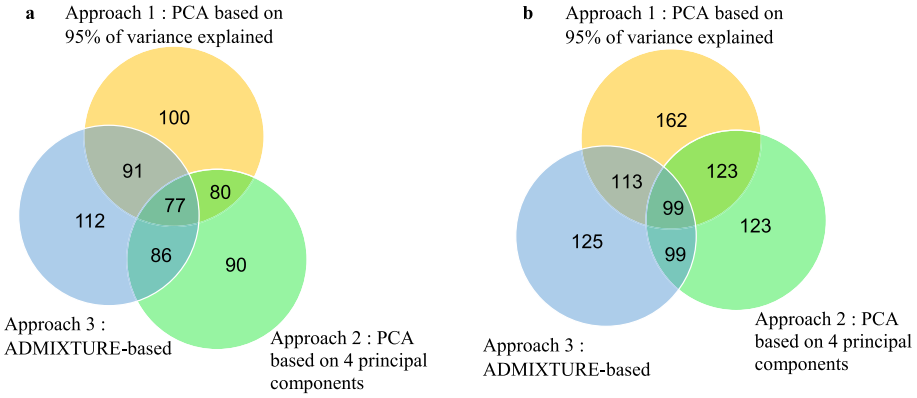

Fig. 7. Venn's diagram comparing the overlapping of the different approaches used to determine sub-populations for further defining reference populations for genomic evaluations of a) East Belgian Red and White and b) Red-Pied of the Ösling cattle. The size of the overlap is proportional to the number of animals. Abbreviations: PCA = principal component analysis.

we did according to Meyermans et al. (2020) and Ferenčaković et al. (2013). At lower density, the model-based approach is nevertheless more efficient (see for instance Solé et al., 2017 or Alemu et al., 2021).

Previous studies (Marjanovic et al., 2021; Schmidtmann et al., 2021) supported the values of the genomic inbreeding coefficient based on runs of homozygosity for the six Dutch Breeds under study (DB, DFR, DR, GWH, IR and MRY) as well as their ranking based on these values (Supplementary Table S2). The small differences between our study and those of Marjanovic et al. (2021) and Schmidtmann et al. (2021) can be explained by the parameters used to estimate the genomic inbreeding coefficient based on runs of homozygosity, the size of the segments to be used in this estimation and also the animals included in the sample. As the values of the genomic inbreeding coefficient based on runs of homozygosity obtained for the Dutch breeds were similar to other studies, we could infer that, for EBRW and RPO, this parameter was also correctly estimated in our study.

Figs. 1–3 agreed on the low inbreeding level found in EBRW and RPO. Considering the endangered status of these breeds, these results were relatively unexpected. If we compare genomic inbreeding coefficients based on runs of homozygosity and homozygous by descent segments of EBRW and RPO breeds with those of the endangered DR, that rooted from the same breed group and would have an equivalent breed history, they were even lower (Figs. 1–3, Supplementary Fig. S1). These low levels of inbreeding could be explained by several possible events: (1) The absence of selection of bulls and therefore the absence of a sire effect; (2) Exchanges of animals of the same breed but from different farms, with different breeding goals and/or definition of the breed. It is however difficult to measure the extent of this event; (3) Recent, or past, admixture with different related red-pied breeds from Europe, which is supported by the clustering results. The IR breed also had low levels of inbreeding and is known to be admixed (van Breukelen et al., 2019), making this hypothesis likely. The EBRW and RPO breeds were (and are still) historically characterised by their non-Holsteinisation. This does not mean that Red-Holstein animals were not used at all in mating of EBRW and RPO but the genetic influence of higher-yielding breeds is limited. However, a well-known fact supporting the third hypothesis is that the Luxembourgish breeder can have access to Rotbunte DN semen, another red-pied breed from Germany, and that breeders from Belgium and Luxembourg are often searching for new bulls, from abroad, to use for mating in their herds. Therefore, there were, and there are still, exchanges of red-pied cattle within and to Benelux countries but this needs to be organised in a manner that benefits the most to each of the involved breeds. These exchanges would imply some introgression events, which were unavoidable for these small breeds that remained without any official recognition and management for many years. However, this would imply a better sustainability of the EBRW and RPO breeds for the forthcoming years as risks related to inbreeding (e.g. lower reproduction performances, decrease of fitness, low resilience to changes in the market or the production environment, Hiemstra et al., 2010; Leroy, 2014; Bosse et al., 2019; Wellmann and Bennewitz, 2019) are currently limited.

However, it should be noticed that two animals of the EBRW breed showed a high level of recent inbreeding compared to other animals of the same breed (Fig. 2). For one of these animals, the dam was also the granddam, which supported our results. For the second one, only the maternal line could be traced back until generation three, meaning the pedigree inbreeding of this animal would be undetermined, which illustrates the benefit of genotyping. For this animal, the high inbreeding level found could be a strategy of the breeder but also, as pedigree records were scarce for the EBRW breed, due to undesirable inbred matings.

Even if inbreeding was generally not an issue in the endangered EBRW and RPO, it is still of main importance to monitor their inbreeding rate (Doublet et al., 2019) and to enhance their competitiveness against mainstream breeds (Wellmann and Bennewitz, 2019). For this purpose, one solution is to implement a genomic evaluation system in these breeds. The definition of an across-breed reference population for the EBRW and RPO breeds is the first step towards this solution and is targeted by the third objective of this study.

Principal component analysis and fixation index values

The genetic proximity of the EBRW, DR and IR was already pointed out by the PCA of François et al. (2017). Marjanovic et al. (2021) and Schmidtmann et al. (2021) studied the genomic diversity of the six Dutch breeds used in this study (DB, DFR, DR, GWH, IR and MRY) and they found similar relatedness for these breeds based on the PCA. The genetic proximity of the IR, MRY and DR breeds was also detected by van Breukelen et al. (2019) in their similarity matrix. Indeed, the IR and DR breeds are coming from the MRY breed and are representing now separated sub-populations with different breeding goals. Moreover, van Breukelen et al. (2019) showed that the IR breed diverged from those breeds on their neighbour joining tree and they explained it by the fact that sires from the Beef Belgian Blue cattle breed were used to improve the IR breed for beef production, as emphasised by its name. According to these authors, the DR breed was also derived from the MRY breed and despite different emphases on colour and the importance of beef production, they are both dual-purpose breeds. All these elements should explain their genetic relatedness but also their differences. van Breukelen et al. (2019), Marjanovic et al. (2021) and Schmidtmann et al. (2021) also pointed out the genetic distinctness of the GWH. Similar fixation index values than ours (Supplementary Table S3) were obtained by Marjanovic et al. (2021) for the six Dutch breeds under study. Therefore, all these studies are supporting our results.

Clustering

In Fig. 6, it can be observed that the purple cluster for K = 8 is mostly shared between EBRW, DR and IR, highlighting their genetic proximity. For the Dutch breeds, we obtained similar results than Schmidtmann et al. (2021) even if they did not study EBRW and RPO breeds and added other red and red-pied breeds to their analysis. van Breukelen et al. (2019), by using fastSTRUCTURE, observed similar clustering partition than us for IR, DFR and MRY. However, for the DR breed, they observed a similar clustering partition with the IR breed, while it was not the case in our study and this of Schmidtmann et al. (2021). The genetic origin of the IR breed involved crossbreeding and it resulted in a high genetic diversity that could explain the different results obtained. Even if they detected a similar genetic background for EBRW, IR and DR via clustering, François et al. (2017) obtained different clustering partitions than us for these breeds. This can be explained by the breeds included in their study, the animals sampled (i.e., for the IR breed), the different software used for clustering but also by the number of clusters chosen (K = 5, in the study of François et al. (2017) for clustering 10 breeds).

Fig. 6 allowed to observe a similar clustering between EBRW and IR. However, we know from Wilmot et al. (2022) that the differentiation between these breeds is possible with a high accuracy. The similar pattern observed can be explained by several known or possible events: (1) It is known that IR, DR and EBRW were rooted from the same group of breeds (François et al., 2017; van Breukelen et al., 2019). This can be seen by the purple clustering in Fig. 6; (2) It is also possible that animals from EBRW and IR were mated with animals from related red-pied breeds like DR, MRY or other red-pied breeds of Europe. It is for example known that EBRW animals were imported for mating with DR animals (Vereniging Het Brandrode Rund, 2022); (3) Exchanges between EBRW and IR could also be possible, even if it would probably be to a lesser extent.

Fig. 6 also allowed to identify some genetic contributions from MRY in RPO (red cluster), which were already well known. Two sub-populations can be identified in the RPO breed: one with the most green clustering proportion and a second one with the most pink clustering proportion. This can be due to different exchange strategies with similar red-pied breeds of Europe, or at least, as all samples were coming from the same farm, animals with different admixture levels from these other but similar red-pied breeds. It is for example known, as previously stated, that there were and, there are still, exchanges with the Rotbunte DN. Another hypothesis that is likely considering the within- and between-clusters average relationships based on the available pedigree of RPO animals (data not shown) is that these two sub-populations are representing two different sire lines.

Detection of sub-populations

Special attention was paid to the animals that can be included in the across-breed reference population of the EBRW and RPO breeds because one of the assumptions for genomic predictions to be accurate is that candidate animals should be genetically related to animals in the reference population (Wientjes et al., 2013). Moreover, differences in allele frequencies, in linkage disequilibrium between SNPs and quantitative trait loci, and in estimated SNP effects may exist between breeds (Meuwissen et al., 2016). This will be increasingly the case, with increasingly lower genetic correlations between breeds (Hozé et al., 2014). This is why we aimed to select, within related breeds, animals that seem to be more genomically related to the EBRW and RPO breeds.

Animals to be considered as genomically close to EBRW or RPO breeds were mostly the same considering the three approaches used for the detection of sub-populations. Namely, IR and DR animals were selected for the EBRW breed and MRY animals were selected for the RPO breeds, which is in agreement with the PCA and phylogenetic tree results previously obtained. However, the three approaches also detected DB samples as genomically close to the EBRW, which was not observed on the PCA and the phylogenetic tree. It may be due to the chosen threshold (mean of all distances), which might be a bit too liberal. Other thresholds were tested: mean of distances −0.5 or −1 times the SD of distances (results not shown). However, these two thresholds resulted in a limited number of detected animals because the SD is high compared to the mean. Therefore, finding an appropriate threshold to define which animals to select as genomically close instead of choosing an arbitrary number of animals may be difficult. It can be suggested to use the overlap of the three approaches to determine which animals can be included in the across-breed reference populations of EBRW or RPO. It is also planned in a future study to test the accuracy of genomic predictions when animals from other breeds detected by different thresholds are added to the across-breed reference population.

In the literature, estimates of relatedness, as the effective number of chromosomal segments, have often been proposed for the

building of across-breed reference populations (Wientjes et al., 2016; Marjanovic and Calus, 2021). Based on them, it is possible to determine how many animals from another breed would be equivalent to add one animal of the breed of interest in the reference population for further genomic evaluations (e.g. Marjanovic et al., 2021). In our study, one of the objectives was to detect animals from other breeds that could be used in an across-breed reference population, i.e., detection of stratification. For this purpose, animals that were more related to the breeds of interest, namely EBRW and RPO, than the average of their breed, were detected. This approach can be seen as a "reverse" approach to Rezende et al. (2020) that detected animals from a small-sized population that were the most closely related to animals from an international population. They also based the definition of the most closely related animals on a PCA but rather used coordinates on the first principal component as a threshold, which can be arbitrary. Another option for the detection of stratification within the breeds would be to use a fine-scaled approach such as ChromoPainter (Lawson et al., 2012). In contrast to ADMIXTURE, ChromoPainter has the advantage to not only consider if animals from different breeds share similar breed proportions at a genome-wide level but also to evaluate their genetic makeup throughout the genome. However, in our study, results obtained with ADMIXTURE- and PCA-based approaches for the detection of sub-populations were in agreement. The PCA is based on a different concept than ADMIXTURE and suggests that close animals on the PCA share a similar genetic makeup. Therefore, because of the high overlap between the PCA-based and the ADMIXTURE-based approaches, animals from other breeds detected as genomically close to EBRW and RPO by the ADMIXTURE-based approach probably shared a relatively high genetic makeup with these breeds.

Perspectives

The current study focused specifically on local red-pied breeds from the Benelux, drawing conclusions for these studied breeds. However, dual-purpose type animals, with no or little Holstein genes, from European low-land red-pied breeds still exist in Poland, Denmark, Germany and France, at least (Gengler and Wilmot, 2022). Through exports, they can even be found in Ireland and in some extra-European countries like Chile. The potential scope of this study should therefore be considered very international, going well over the three Benelux countries. Follow-up studies will need to extend the scope of populations involved.

This study laid down the foundations for an across-breed genomic evaluation system of red-pied cattle breeds by detecting animals that could be included in an across-breed reference population. The expected genomic prediction accuracy with such across-breed reference population can be obtained by using the equation presented by Wientjes et al. (2016), for any traits of interest. This equation requires estimated heritabilities for each breed, and estimated genetic correlations between breeds. Estimated heritabilities within breeds require sufficiently large numbers of available phenotypes per breed, which should then be combined with genomic data across the breeds to estimate the genetic correlations between those breeds. These data were not available to perform the current study. To allow the estimation of genetic correlations between the different breeds for a specific trait, while avoiding to exchange data, SNP effects could also be estimated separately for each breed and then the computed correlation between those SNP effects of different breeds could be used as a proxy for the genetic correlation between those breeds. As the reference population would be very small for endangered breeds, the accuracy of the estimated SNP effects for these breeds may however be a limiting factor to get reliable proxies for the genetic correlations between breeds.

Conclusion

A genomic diversity study of red-pied breeds was conducted, with a special focus on two local endangered breeds: the EBRW and RPO. These two breeds were part of a continuum of red-pied breeds and were genomically close to IR and DR in the case of EBRW and to MRY in the case of RPO. The EBRW breed appeared to be admixed if we consider the clustering analysis and shared a similar ADMIXTURE pattern with the DR breed. This analysis also revealed a stratification in the RPO breed as two clustering patterns showed up in this breed. Levels of inbreeding were relatively low for all breeds, except for the GWH breed. We detected animals from other breeds that could be valuable to add when defining a reference population for further genomic evaluations of EBRW and RPO. For this purpose, definitions of distances based on PCA and ADMIXTURE should be considered. However, it is still not clear which threshold should be used to determine if an animal is genomically close enough to be included in such a reference population. Further studies might explore this issue by comparing how the addition of animals selected by the different thresholds affects the accuracy of the genomic prediction. This will be the objective of a second study.

Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.animal.2023.100793.

Ethics approval

The SNP data for the animals included in this study were previously obtained on samples collected by concerned breeder associations based on relevant authorisation by the different local authorities.

Data and model availability statement

None of the data were deposited in an official repository. The data supporting the findings of this study cannot be made available as a whole. The corresponding author, upon reasonable request, will forward any requests to relevant data owners.

Author ORCIDs

Hélène Wilmot: https://orcid.org/0000-0002-1075-088X.
Tom Druet: https://orcid.org/0000-0003-1637-1706.
Ina Hulsegge: https://orcid.org/0000-0002-7015-6540.
Nicolas Gengler: https://orcid.org/0000-0002-5981-5509.
Mario Calus: https://orcid.org/0000-0002-3213-704X.

Author contributions

H. Wilmot: Conceptualisation, Methodology, Software, Investigation, Writing-Original Draft **T. Druet:** Software, Investigation, Writing-Review and Editing **I. Hulsegge:** Investigation, Writing-Review and Editing **N. Gengler:** Conceptualisation, Methodology, Investigation, Writing-Review and Editing, Supervision **M.P.L. Calus:** Conceptualisation, Methodology, Investigation, Writing-Review and Editing**,** Supervision.

Declaration of interest

The authors declare that they have no competing interests.

Acknowledgements

The authors acknowledge the technical support by the Walloon Breeders Association (elevéo), the Administration of Technical Agricultural Services (ASTA) of the Ministry of Agriculture, Viticulture and Rural Development of the Grand Duchy of Luxembourg and CONVIS, a breeders' cooperative from the Grand Duchy of Luxembourg. They are also acknowledged for providing genotypes of EBRW and RPO animals. The authors especially thank Jeanne Bormann (Administration of Agricultural Technical Services, Luxembourg, Grand Duchy of Luxembourg), Patrick Mayeres and Géry Glorieux (Walloon Breeders Association, Ciney, Belgium) for their help in understanding the history, the breeding and herd book management of RPO and EBRW, respectively. The Centre for Genetic Resources, the Netherlands (CGN) of Wageningen University and Research are acknowledged for providing the data for the Dutch breeds.

Financial support statement

H. Wilmot, as a current Research Fellow and as a recipient of a grant for "Accomplissement d'un séjour à l'étranger [V3]", T. Druet, as a Research Director, and N. Gengler, as a former Senior Research Associate, acknowledge the support of the Fonds de la Recherche Scientifique – FNRS (Brussels, Belgium). The University of Liège and the Wallonia-Brussels Federation (FW-B), through the subsidy "Missions scientifiques" for an abroad research stay, is acknowledged by H. Wilmot. The contributions of I. Hulsegge were funded by the Dutch Ministry of Agriculture, Nature and Food Quality (project KB-34-013-002). The Walloon Government (Service Public de Wallonie – Direction Générale Opérationnelle Agriculture, Ressources Naturelles et Environnement, SPW-DGARNE; Namur, Belgium) is acknowledged for its financial support. The authors also acknowledge the financial support of the Administration of Technical Agricultural Services (ASTA) of the Ministry of Agriculture, Viticulture and Rural Development of the Grand Duchy of Luxembourg, in particular through the CaLiCCo Project. The University of Liège – Gembloux Agro-Bio Tech (Liège, Belgium) supported computations through the technical platform Calcul et Modélisation Informatique (CAMI) of the TERRA Teaching and Research Centre, partly supported by the Fonds de la Recherche Scientifique – FNRS under grant no. T.0095.19 (PDR "DEEPSELECT"). Genotyping of Belgian and Luxembourgian animals was facilitated through the support of the Fonds de la Recherche Scientifique – FNRS under grant no. J.0174.18 (CDR "PREDICT-2").

References

Alemu, S.W., Kadri, N.K., Harland, C., Faux, P., Charlier, C., Caballero, A., Druet, T., 2021. An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity 126, 410–423. https://doi.org/10.1038/s41437-020-00383-9.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655–1664. https://doi.org/10.1101/gr.094052.109.

Bertrand, A.R., Kadri, N.K., Flori, L., Gautier, M., Druet, T., 2019. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. Methods in Ecology and Evolution 10, 860–866. https://doi.org/10.1111/2041-210X.13167.

Bosse, M., Megens, H.J., Derks, M.F.L., de Cara, Á.M.R., Groenen, M.A.M., 2019. Deleterious alleles in the context of domestication, inbreeding, and selection. Evolutionary Applications 12, 6–17. https://doi.org/10.1111/eva.12691.

Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. American Journal of Human Genetics 81, 1084–1097. https://doi.org/10.1086/521987.

Chang, C.C., Chow, C.C., Tellier, L.C.A.M., 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.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, 8. https://doi.org/10.1093/gigascience/giab008.

Doublet, A.-C., Croiseau, P., Fritz, S., Michenet, A., Hozé, C., Danchin-Burge, C., Laloë, D., Restoux, G., 2019. The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds. Genetics Selection Evolution 51, 52. https://doi.org/10.1186/s12711-019-0495-1.

Druet, T., Gautier, M., 2017. A model-based approach to characterize individual inbreeding at both global and local genomic scales. Molecular Ecology 26, 5820–5841. https://doi.org/10.1111/mec.14324.

Druet, T., Gautier, M., 2022. A hidden Markov model to estimate homozygous-by-descent probabilities associated with nested layers of ancestors. Theoretical Population Biology 145, 38–51. https://doi.org/10.1016/j.tpb.2022.03.001.

Ferenčaković, M., Sölkner, J., Curik, I., 2013. Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. Genetics Selection Evolution 45, 42. https://doi.org/10.1186/1297-9686-45-42.

François, L., Wijnrocx, K., Colinet, F.G., Gengler, N., Hulsegge, B., Windig, J.J., Buys, N., Janssens, S., 2017. Genomics of a revived breed: Case study of the Belgian campine cattle. PLoS ONE 12, e0175916.

Gengler, N., Wilmot, H., 2022. Rotbunte Rassen: Genetische und geschichtliche Betrachtung. Proceedings of the Kongress zur Zucht und Erhaltung alter und bedrohter einheimischer Nutztierrassen, 26–28 September 2022, Bonn, Germany, 11 pages.

Goddard, M., 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica 136, 245–257. https://doi.org/10.1007/s10709-008-9308-0.

Gómez-Romano, F., Villanueva, B., Rodríguez De Cara, M.Á., Fernández, J., 2013. Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. Genetics Selection Evolution 45, 38. https://doi.org/10.1186/1297-9686-45-38.

Hiemstra, S.J., de Haas, Y., Mäki-Tanila, A., Gandini, G., 2010. Local cattle breeds in Europe. Wageningen Academic Publishers, Wageningen, The Netherlands. Doi: 10.3920/978-90-8686-697-7.

Hozé, C., Fritz, S., Phocas, F., Boichard, D., Ducrocq, V., Croiseau, P., 2014. Efficiency of multi-breed genomic selection for dairy cattle breeds with different sizes of reference population. Journal of Dairy Science 97, 3918–3929. https://doi.org/10.3168/jds.2013-7761.

Lawson, D.J., Hellenthal, G., Myers, S., Falush, D., 2012. Inference of population structure using dense haplotype data. PLoS Genetics 8, e1002453.

Lawson, D.J., van Dorp, L., Falush, D., 2018. A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications 9, 3258. https://doi.org/10.1038/s41467-018-05257-7.

Lê, S., Josse, F., Husson, F., 2008. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software 25, 1–18. https://doi.org/10.18637/jss.v025.i01.

Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R., Malhotra, A.K., 2007. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America 104, 19942–19947. https://doi.org/10.1073/pnas.0710021104.

Leroy, G., 2014. Inbreeding depression in livestock species: Review and meta-analysis. Animal Genetics 45, 618–628. https://doi.org/10.1111/age.12178.

Marjanovic, J., Calus, M.P.L., 2021. Factors affecting accuracy of estimated effective number of chromosome segments for numerically small breeds. Journal of Animal Breeding and Genetics 138, 151–160. https://doi.org/10.1111/jbg.12512.

Marjanovic, J., Hulsegge, B., Calus, M.P.L., 2021. Relatedness between numerically small Dutch Red dairy cattle populations and possibilities for multibreed genomic prediction. Journal of Dairy Science 104, 4498–4506. https://doi.org/10.3168/jds.2020-19573.

Medugorac, I., Medugorac, A., Russ, I., Veit-Kensch, C.E., Taberlet, P., Luntz, B., Mix, H.M., Förster, M., 2009. Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. Molecular Ecology 18, 3394–3410. https://doi.org/10.1111/j.1365-294X.2009.04286.x.

Meuwissen, T., Hayes, B., Goddard, M., 2016. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers 6, 6–14. https://doi.org/10.2527/af.2016-0002.

Meyermans, R., Gorssen, W., Buys, N., Janssens, S., 2020. How to study runs of homozygosity using plink? A guide for analyzing medium density snp data in livestock and pet species. BMC Genomics 21, 94. https://doi.org/10.1186/s12864-020-6463-x.

Purcell, S., Chang, C., 2019. PLINK v1.9. Retrieved on 28 March 2022 from www.cog-genomics.org/plink/1.9/.

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/10.1186/1471-2156-13-70.

R Core Team, 2022. R: A language and environment for statistical computing. (3.6.3). R Foundation for Statistical Computing, Vienna, Austria.

R Studio Team, 2022. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA, USA.

Rezende, F.M., Haile-Mariam, M., Pryce, J.E., Peñagaricano, F., 2020. Across-country genomic prediction of bull fertility in Jersey dairy cattle. Journal of Dairy Science 103, 11618–11627. https://doi.org/10.3168/jds.2020-18910.

Rosen, B.D., Bickhart, D.M., Schnabel, R.D., Koren, S., Elsik, C.G., Tseng, E., Rowan, T. N., Low, W.Y., Zimin, A., Couldrey, C., Hall, R., Li, W., Rhie, A., Ghurye, J., McKay, S. D., Thibaud-Nissen, F., Hoffman, J., Murdoch, B.M., Snelling, W.M., MacDaneld, T. G., Hammond, J.A., Schwartz, J.C., Nadolo, W., Hagen, D.E., Dreischer, C., Schultheiss, S.J., Schroeder, S.G., Phillippy, A.M., Cole, J.B., Van Tassell, C.P., Liu, G., Smith, T.P.L., Medrano, J.F., 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience 9, 21. https://doi.org/10.1093/gigascience/giaa021.

Schliep, K., Potts, A.J., Morrison, D.A., Grimm, G.W., 2017. Intertwining phylogenetic trees and networks. Methods in Ecology and Evolution 8, 1212–1220. https://doi.org/10.1111/2041-210X.12760.

Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M., Hinrichs, D., Thaller, G., 2021. Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. Genetics Selection Evolution 53, 23. https://doi.org/10.1186/s12711-021-00613-6.

Schnabel, R.D., 2019. ARS-UCD1.2 Cow Genome Assembly: Mapping of all existing variants. Retrieved on 28 March 2022 from https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/.

Slagboom, M., Milkevych, V., Liu, H., Thomasen, J.R., Kargo, M., Schmidtmann, C., 2022. Conservation of local Red cattle breeds by collaboration with a mainstream Red dairy cattle breed. Livestock Science 260,. https://doi.org/10.1016/j.livsci.2022.104936 104936.

Solé, M., Gori, A.S., Faux, P., Bertrand, A., Farnir, F., Gautier, M., Druet, T., 2017. Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. Genetics Selection Evolution 49, 92. https://doi.org/10.1186/s12711-017-0370-x.

van Breukelen, A.E., Doekes, H.P., Windig, J.J., Oldenbroek, K., 2019. Characterization of genetic diversity conserved in the gene bank for dutch cattle breeds. Diversity 11, 229. https://doi.org/10.3390/d11120229.

Vereniging Het Brandrode Rund, 2022. Geschiedenis. Retrieved on 14 December 2022 from https://www.hetbrandroderund.nl/geschiedenis/.

Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370. https://doi.org/10.2307/2408641.

Wellmann, R., Bennewitz, J., 2019. Key genetic parameters for population management. Frontiers in Genetics 10, 667. https://doi.org/10.3389/fgene.2019.00667.

Wientjes, Y.C.J., Veerkamp, R.F., Calus, M.P.L., 2013. The effect of linkage disequilibrium and family relationships on the reliability of genomic prediction. Genetics 193, 621–631. https://doi.org/10.1534/genetics.112.146290.

Wientjes, Y.C.J., Bijma, P., Veerkamp, R.F., Calus, M.P.L., 2016. An equation to predict the accuracy of genomic values by combining data from multiple traits, populations, or environments. Genetics 202, 799–823. https://doi.org/10.1534/genetics.115.183269.

Wilmot, H., Bormann, J., Soyeurt, H., Hubin, X., Glorieux, G., Mayeres, P., Bertozzi, C., Gengler, N., 2022. Development of a genomic tool for breed assignment by comparison of different classification models - Application to three local cattle breeds. Journal of Animal Breeding and Genetics 139, 40–61. https://doi.org/10.1111/jbg.12643.