Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees

Moore, KL; Vilela, C; Kaseja, K; Mrode, R; Coffey, MP

Published in:
Journal of Animal Science

DOI:
10.1093/jas/sky407

First published: 20/10/2018

Document Version
Peer reviewed version

Running head: Genomic pedigree validation and discovery

Forensic use of the genomic relationship matrix to validate and discover livestock pedigrees

K. L. Moore*, C. Vilela*, K. Kaseja*, R, Mrode* and M. Coffey*

* Scotland's Rural College (SRUC), Easter Bush, Midlothian, Scotland, EH25 9RG, UK.

Acknowledgements: The authors wish to thank the British Limousin Cattle Society for providing access to their pedigree and genotype database.

Corresponding author: kirsty.moore@sruc.ac.uk

Abstract

Correct pedigree is essential to produce accurate genetic evaluations of livestock populations. Pedigree validation has traditionally been undertaken using microsatellites and more recently, based on checks on opposing homozygotes using Single Nucleotide Polymorphisms (SNPs). In this study, the genomic relationship matrix was examined to see if it was a useful tool to forensically validate pedigree and discover unknown pedigree. Using 5,993 genotyped Limousin animals which were imputed to a core set of 38,907 SNPs, the genomic relationships between animals were assessed to validate the reported pedigree. Using already pedigree verified animals, the genomic relationships between animals of different relationships were shown to be on average 0.58, 0.59, 0.32, 0.32, 0.19 and 0.14 between animals and their parents, full siblings, half siblings, grandparents, great grandparents and great great grandparents, respectively. Threshold values were defined based on the minimum genomic relationship reported between already pedigree verified animals; 0.46, 0.41, 0.17, 0.17, 0.07 and 0.05, respectively for animals and their parents, full siblings, half siblings, grandparents, great grandparents and great great grandparents. Using the wider population and the above genomic relationship threshold values, potential pedigree conflicts were identified within each relationship type. Pedigree error rates of between 0.9% (animal and great great grandparent) and 4.0% (full siblings) were identified. A forensic genomic pedigree validation and discovery system was developed to enable pedigree to be verified for individual genotyped animals. This system verifies not just the parents, but also a wide number of other genotyped relatives and can therefore identify more potential errors in the pedigree than current conventional methods. A novel aspect to this algorithm is that it can also be used to discover closely related animals on the basis of their genomic relationships although they are not recorded as such in the pedigree. This functionality enables missing pedigree information to be discovered and corrected in the pedigree of livestock populations.

The methods in this paper demonstrate that the genomic relationship matrix can be a useful tool in the validation and discovery of pedigree in livestock populations. However, the method does rely on being able to define threshold values appropriate to the specific livestock population, which will require sufficient number of animals to be genotyped and pedigree validated before it can be used.

Key words

genomic relationship matrix, pedigree discovery, pedigree verification

Introduction

Genetic evaluations in the UK are undertaken using Best Linear Unbiased Prediction (BLUP) techniques (Henderson, 1973). In addition to phenotypic information, a relationship matrix is constructed based on recorded pedigree information. Therefore, correct knowledge of pedigree is essential for accurate genetic evaluations. However, pedigree errors in livestock populations are common with significant error rates reported in sheep, beef and dairy populations (Kaseja et al., 2018; Spelman, 2002). Visscher et al., (2002) for UK dairy cows estimated an overall pedigree error rate of 10% and predicted this would result in a loss of selection response of 2 to 3%. For the same pedigree error rate, Israel and Weller, (2000) predicted a 4.3% loss in genetic response. Banos et al., (2001) showed that with 11% pedigree errors there was a reduction in the Estimated Breeding Values (EBVs) genetic trends of 11 to 18%.

To improve the accuracy of the pedigree, molecular techniques can be used for parentage verification. Until recently, microsatellite markers were the standard approach to parentage verification (Davis and DeNise, 1998). The international standard has been to use 12 International Society of Animal Genetic (ISAG) markers

(http://www.isag.us/Docs/CattleMMPTest_CT.pdf). With the introduction of Single Nucleotide Polymorphisms (SNP) and genomic selection (Meuwissen et al., 2001), SNP based parentage methods are now becoming the standard approach. The international standard has been to use the ISAG100 or ISAG200 SNP set, however McClure et al., (2015) has suggested that a panel with a minimum of 500 SNPs is more appropriate for parentage verification and prediction.

To date most pedigree verification has focused on the animal – parent relationship although Van Raden et al., (2013) and Wiggans et al., (2018) have both reported methods to assess the validity of animal – grandparent relationships using SNP based approaches. Considering relationships other than animal – parent could be advantageous as often the females in the population are not well genotyped but the maternal grandsires often are. Huisman (2017) used likelihood methods applied to SNP genotype markers to reconstruct pedigree in a number of simulated and empirical datasets of wildlife populations. This study found a wide range of relationship types useful to construct the pedigree, and developed an R package to do so. However the likelihood methods are computationally demanding and not able to compute for large datasets often observed in livestock populations. The study also showed generally strong positive correlations between the relationship matrix from the constructed pedigree and the genomic relationship matrix (GRM).

The GRM is required for genomic selection and much research attention has been on how to construct and invert the matrix and the impact of this on the resulting genetic evaluations and their accuracies (Chen et al., 2011; Habier et al., 2007; Jimenez-Montero et al., 2013; Koivula et al., 2012; Muir, 2007; Van Raden, 2007; 2008). However, little focus has been placed on whether the GRM could be a useful tool to validate and discover parentage for livestock populations. Grashei et al., (2018) considered the GRM in a simulation and assigned genomic relationship likelihood values to verify and discover sets of

parentage trios based on thresholds specific to genotype error rates of 1 and 3%. This approach assumed both parents were genotyped and considered verified parent – offspring relationships. In a chicken population, Wang et al., (2014) compared the GRM with the pedigree numerator relationship matrix (NRM). This study found where populations had long and complete pedigree recorded, clean genotypes and proper scaling applied to the GRM that the relationship coefficient from the NRM and GRM were in strong agreement. Recently, human forensic investigators have successfully used genomic relationships using DNA left at crime scenes and genotypes stored in human genealogical databases to identify suspects and solve previously unsolved cases (Ram et al., 2018). Often the perpetrator themselves do not have a genotype stored in these databases, but the suspect is identified based on identifying cousins and other close relatives – with relatives on the maternal and paternal side of the pedigree, this approach can identify a single family group to consider more closely to identify potential suspects.

The objective of this paper was to use genotypes from a UK beef population to construct a GRM and assess if it was a useful tool to forensically validate and discover missing pedigree to improve the accuracy of the pedigree, and thus ultimately the accuracy of genetic evaluations. In particular, we wanted to assess if the genomic relationships between more distantly related animals i.e. half sibs and grandparents could be used to verify pedigree involving un-genotyped parents.

Materials and Methods

After removing duplicate genotypes and genotypes with a call rate of less than 90%, 5,993 genotyped animals were available from a UK pedigree Limousin beef population. The dataset consisted of 1,942, 1,790, 1,494 and 767 animals genotyped with Illumina 50k, High density, International Dairy and Beef (IDB) 50k and IDB 14k SNP panels, respectively.

Previous unpublished work on this population undertook a principal component analysis which confirmed the genotyped population to be purebred without any cross bred and animals from another breed present in the genotyped population. Pedigree was available for these animals from a national bovine pedigree which included pedigree from pedigree Society databases, national British Cattle Movement Service (BCMS) data and milk recording organisations. On average, 7 generations (range = 1 to 14) of pedigree were available for genotyped animals. In almost all cases where pedigree is reported, both the sire and dam are reported as this is a breed society requirement. For 87% of the genotyped animals there were 4 or more generations of complete pedigree available. Inbreeding coefficients were computed using RelaX2 software (Stranden and Vuori, 2006) for all animals available in the national bovine pedigree, with no restriction placed on the number of generations of pedigree or genotype status. However the inbreeding results are reported only for the genotyped animals.

A panel of 116 USDA parentage SNPs was used to verify the reported parentage of the genotyped animals using opposing homozygotes (Hayes, 2011) where both parent and offspring were genotyped. Animal – parent combinations with more than 2 inconsistencies were considered to fail parentage verification.

All genotypes were imputed using the program Findhap Version 3 (Van Raden et al. 2011) to a core set of 38,907 SNPs currently used for the national genomic evaluations. These SNPs were selected based on minor allele frequencies greater than 0.05 and SNP call rates greater than 0.90, and included the parentage SNPs where they passed the inclusion criteria. The average minor allele frequency in the core SNP subset was 0.28. Using this set of imputed genotypes, a GRM was constructed using Van Raden's (2008) first method with the GRM scaled using the current population allele frequencies.

Analysis of GRM to validate pedigrees

Pairwise genomic relationship coefficients between genotyped animals that passed parentage verification using the SNP based opposing homozygote approach were extracted, summarized and reported for animals with their respective parents, grandparents, great grandparents, great great grandparents, full siblings and half siblings. The genomic relationship coefficients obtained gave a range of accepted genomic relationship coefficients for each of the different pedigree relationship categories. For example, to contribute to the animal – grandparent category the animal – parent and parent – grandparent relationship needed to be verified based on the SNP based opposing homozygote method. This was undertaken for all animals that met the criteria to contribute to the specific categories and then again using only animals where both animals in the pairwise comparison had inbreeding coefficients less than 7%.

This method was then applied to the wider genotyped population regardless of their pedigree verification status, provided both animals in the pair combination were genotyped. The pairwise relationship was deemed to have failed validation where the genomic relationship was lower than the minimum genomic relationship coefficient reported in the subset of genotyped animals pedigree verified from SNP based opposing homozygote method.

To verify un-genotyped sires and dams, genomic relationships within paternal and maternal half sibling family groups were compared, respectively. Again, the minimum genomic relationship coefficient reported for half siblings from the subset of previously pedigree verified genotyped animals was used to assess if the true relationship between the animals was in line with that of half siblings. This information, along with the number of genotyped animals in the half sibling family, was used to assess if the reported un-genotyped parent could be considered as being correct. An alternative method of assessing the accuracy

of a un-genotyped reported parent was to compare the genomic relationship between animals and their grandparents. Again the threshold for acceptance was the reported minimum genomic relationship for animal – grandparent from the study using only animals previously pedigree verified using SNP based methods.

For a given genotyped animal, all the genomic relationship coefficients between that animal and the wider genotyped population were used to produce a forensic genomic pedigree validation and discovery report. This report grouped animals based on the reported pedigree relationships into the following family groups; progeny, parents, grandparents, great grandparents, great great grandparents, full siblings, paternal and maternal half siblings, aunts/uncles, great aunts/uncles, great great aunts/uncles, nieces/nephews and 1st cousins. The genomic relationship coefficients between the given animal and their relatives were reported along with a marker showing if the genomic relationship coefficient is above or below the appropriate minimum relationship observed from the analysis using only animals previously pedigree verified. To assist with the forensic discovery of unknown pedigree, the report also ranked animals that were not in reported pedigree relationships or have genomic relationship coefficients inconsistent with the reported pedigree relationship into three candidate lists for consideration; 1. Likely to be a close relationship akin to grandparent, sibling, parents, 2. Those likely to be more distantly related, i.e. great grandparents and 3. Those not closely related. Studying these lists, in particular the close relationship list, can frequently lead to the discovery of missing pedigree information. To test potential candidates, the report has a function where parent information can be substituted, or set to unknown, and the genomic relationship coefficients of all genotyped relatives tested given the suspected true pedigree.

Results and Discussion

Using the 116 USDA parentage SNPs and opposing homozygotes, half (50.1%) of the genotyped animals in the dataset were able to be validated for the animal – parent relationship. In total 2,918 (48.7%) animals had the reported sire and/or dam confirmed with less than 2 SNP inconsistencies observed; the breakdown for these animals were 2,507 sire only, 162 dam only and 249 both sire and dam verified. There were 81 (1.4%) animals where the parentage was inconsistent with that reported in the pedigree; the breakdown for these animals were 77 sire only, 2 dam only and 2 both sire and dam inconsistent. With only 1.4% of animals having inconsistent pedigree reported, the level of pedigree errors for these genotyped animals was very low compared to levels reported in livestock populations (Kaseja et al., 2018; Spelman, 2002; Visscher et al., 2002). This can be attributed to the breed society policy requiring any bull sold at a society bull sale to be sire verified and any embryo transfer calf registered to have both sire and dam verified and correct pedigree reported in the database. The genomic relationship coefficient for the genotyped animals with themselves was on average 1.12 and ranged from 1.01 to 1.71. The pedigree based inbreeding coefficients for these animals averaged 0.01 and ranged from 0.0 to 0.33. The genomic relationship with self may be higher than 1.0 where an animal is inbred (Grashei et al., 2018) or there are SNPs that are identical by state rather than identical by descent. The genomic relationship between sires and dams for the 249 genotyped progeny where both parents were also genotyped was on average 0.09, but ranged from 0.02 to 0.30. The mating pairs were generally between non-related animals with only 14 of these progeny having a pedigree based inbreeding coefficient greater than 7%. The average inbreeding coefficient was 0.02 with a range of 0.0 to 0.14 for the 249 animals with both parents genotyped.

The pairwise genomic relationship coefficients were summarised for animals where the reported pedigree relationship was verified using the USDA parentage SNPs. This was undertaken for all verified animals and then for only those with pedigree inbreeding

coefficients less than 7% and these results are reported in Table 1. For all relationship type categories the average genomic relationship coefficient was higher than the value theoretically expected by between 7 and 9%. For example, animal – parent and animal – full sibling relationships are expected to have 50% of genes in common but in our study we saw the average genomic relationship ranging from 0.57 to 0.59. This increase is of the same magnitude to the genomic relationships between sires and dams from the 249 matings where both parents were genotyped. Animals that were inbred had higher genomic relationships compared to those that were not. However, there was no difference for the minimum genomic relationships observed within a relationship type category. It is these minimum genomic relationship coefficients that were used as threshold values to assess the validity of reported pedigree later in the study. Since inbreeding levels did not affect the minimum genomic relationship category it can be considered that the inbreeding level of the animals will not affect the conclusions drawn about the possibility of the reported pedigree. With only 83 full sibling pairs available, the minimum genomic relationship coefficient (0.46) was higher than that of animal – parent (0.41) relationships. This is likely to be due to the small sample size and not because of a true difference in ranges. Given the theoretical level of relatedness is the same for both relationship type categories and the low number of full siblings to establish a minimum threshold value, it is appropriate to use the minimum genomic relationship for animal – parents also for full siblings. The maximum genomic relationship coefficient within relationship type categories is not as robust to assess the likelihood of the reported pedigree being correct. This is because as seen in Table 1, inbreeding can inflate the genomic relationship coefficient but also the maximum coefficient is similar for the more distant relationships. For example, the maximum coefficient for animal – grandparent is similar to that of animal – great grandparent, while the minimum coefficients were sufficiently different. However, when looking at individual animals with the forensic genomic pedigree

validation and discovery report, comparing the reported coefficient with the appropriate relationship type maximum genomic relationship coefficient may be useful. The genomic relationship ranges reported in this paper are based on this population with population specific inbreeding and genetic diversity levels likely to affect the ranges observed. Therefore to apply this method to other populations, base line thresholds should first be assessed within the specific population.

For all reported pedigree relationships the genomic relationship coefficients are reported in Table 2. The average genomic relationship coefficient within relationship type categories were very similar to those reported in Table 1 for previously pedigree verified animals, as were the maximum genomic relationship values. For the full sibling category there was a set of identical twins, which as expected had a genomic relationship akin to that of the animal to itself. A pairwise comparison was considered inconsistent where the genomic relationship coefficient was below the minimum genomic relationship coefficient reported in Table 1. For example, there were 186 animal – grandparent pairs with a genomic relationship less than 0.17 and thus likely to be not be related at the animal – grandparent level. Across all relationship type categories there were between 0.9% (animal – great great grandparent) and 4.0% (full siblings) relationships that were considered to be inconsistent.

Un-genotyped sires and dams were potentially verified by examining the paternal and maternal half sibling family groups. Of the half sibling relationships reported in Table 2, 59,630 were the result of sharing the same sire and this represented 623 different sires with the number of progeny pairs ranging from 1 (2 progeny) to 14,365 (170 progeny). Using the minimum value for half siblings (0.17) reported in Table 1 there were 1,596 half sibling pairs which had a genomic relationship coefficient inconsistent with that reported in the pedigree. These inconsistencies involved 69 different sires and in some cases it was just 1 pair of half siblings involved and at the other extreme there were 245 pairs of half siblings for the sire

that were inconsistent. In this extreme case, the reported sire was a popular AI sire with 124 progeny genotyped generating 7,626 half sibling pairs to test. The 245 pairs that were inconsistent involved just 2 of his genotyped progeny. Although the sire himself was not genotyped, and thus it was not possible to test parentage using conventional methods, given the large volume of half siblings we can with reasonable confidence consider that the reported AI sire is not the true sire of the 2 animals involved in the failed half sibling pairs. However, this sire is likely to be the true sire for the other 122 genotyped progeny. For the maternal half sibling family groups there were 2,313 half sibling pairs to compare. These were the result of 529 different dams with the number of progeny pairs ranging from 1 (2 progeny) to 210 (21 progeny). There were 43 maternal half sibling pairs that were considered inconsistent, involving 17 dams. Again the number of inconsistent comparisons per dam ranged from 1 to 8. While the interpretation is identical for both paternal and maternal half sibling groups this analysis is better suited to verifying un-genotyped sires due to the larger size of paternal half sibling family groups compared to that for the maternal half sibling family groups. It was not clear exactly how many genotyped half siblings were needed to verify an un-genotyped parent. For those sires and dams with small family groups, this method alone may not be able to verify the pedigree but could identify which sires and dams need genotyping to confirm parentage if there are inconsistencies found. For those sires and dams with larger family groups, the reported parent may not need to be genotyped in order to draw conclusions about the true parentage of progeny. This is especially beneficial where DNA for the candidate parents is unable or too expensive to be obtained.

An alternative approach for verifying the pedigree of animals was to consider the animal – grandparent relationship. Table 2 shows that for the relationship type there were 186 (3.6%) animal – grandparent pairs that were below the threshold of 0.17. Having an inconsistent animal – grandparent genomic relationship coefficient does not automatically

mean that the reported parent is incorrect, as it could be that the reported parent is correct and the error is in fact between the parent – grandparent relationship. This approach can be applied equally to reported sires and dams, and in fact could be more beneficial for the maternal side of the pedigree as females are often not genotyped in the same volume as males. Testing the animal – grandparent relationship can also detect general issues with genotyping earlier. An example of where testing the animal – grandparent relationship can detect genotyping issues earlier is where samples for paternal half siblings are accidently swapped during the sampling and genotyping process. With animal – parent testing, both samples will be correctly parent verified as they share a common sire. However it will not be until the half siblings themselves have progeny, and the progeny subsequently fail the parentage testing process that the accidental genotype swap will be identified. Testing the animal – maternal grandparent relationship will detect that the maternal grandsire is not as reported and the issue can then be identified and resolved at the time of the animal being genotyped rather than when the next generation of animals are being genotyped and DNA from the sire potentially harder to obtain.

The forensic genomic pedigree validation and discovery report provides, for a single animal, information on related animals (those reported in the pedigree and those that are related but not recorded in the pedigree) and details of an example animal are provided in Table 3. For the animal being considered in Table 3, it was detected that despite the reported dam not being genotyped, there was an error on the maternal side of the pedigree and that the reported paternal pedigree appeared to be correct. Furthermore, discovering candidate maternal grandparents was possible which led to the discovery of the correct dam. The success of the report in forensically discovering and correcting pedigree is dependent of the size of the genotyped population – where there are more genotypes the more successful the process will be in identifying and correcting pedigree issues. The pedigree discovery process

312 also requires a level of interpretation and sense checking based on the year of birth and
313 gender of animals involved. There is also the potential for inferring a closer than actual
314 relationship if the genotyped animal is inbred with ancestors occurring several times in the
315 pedigree (i.e. double grandparent). This can be mitigated by considering all the relationships
316 reported in the report and being aware of the possibility of this occurring.

317 The presented methods for forensically validating and correcting pedigrees have been
318 shown to be useful tools for cleaning and enriching pedigrees used in genetic evaluations.
319 Despite this dataset having a relatively low number of parentage errors as a result of the breed
320 societies routine parentage testing scheme, there were still additional pedigree conflicts that
321 were identified in the genotyped dataset. It is likely that the number of pedigree conflicts
322 would be substantially higher in a livestock population that does not already have a stringent
323 pedigree verification scheme and it would be interesting to apply these methods to other
324 livestock populations for comparison. A limitation to the application of these methods in
325 other populations will be establishing robust minimum thresholds values that are used to
326 differentiate the different relationship types. While the thresholds have been robust during
327 testing for parent and grandparent relationship levels, with minimum threshold values of 0.07
328 and 0.05 reported for great and great great grandparents, respectively, a degree of caution
329 should be applied when interpreting the genetic relationships for more distant ancestors as it
330 is possible for unrelated animals to also have these genetic relationships.

331 The methods used to construct the GRM will also impact on the genomic relationship
332 coefficients. The NRM is constructed based on pedigree alone and assumes that the founder
333 animals in the recorded pedigree are unrelated, which is usually not the case. Whereas the
334 GRM is based only on the genotypes and captures the relationships between animals
335 regardless of what is recorded in a pedigree. This means that each method uses a different
336 base population which can result in different relationship coefficients (Wang et al., 2014).

The genomic relationship coefficients from the GRM are influenced by the SNP chip density and platform, the level of QA applied to the genotypes, in particular to the minor allele frequencies (Chen et al., 2011; Forni et al., 2011; Van Raden 2008; Wang et al., 2014). Applying appropriate QA to the genotypes and constructing the GRM so it is scaled using the observed allele frequencies should result in a GRM comparable to the NRM with differences in reported coefficients due to errors in the reported pedigree (Chen et al., 2011; Forni et al., 2011; Van Raden 2008).

Conclusion

This study has shown how analysis and interpretation of the genetic relationship coefficients reported from the genomic relationship matrix can be used to validate reported pedigree and in some cases discover the missing pedigree information. Pedigrees of un-genotyped relatives were also shown to be possible depending on the number of genotyped relatives available for comparisons. Applications of these methods to genotyped populations will be able to identify more pedigree errors than using the current animal – parent SNP based opposing homozygote approaches and this will ultimately improve the accuracy of genetic evaluations and thus increase the genetic gain achieved within these livestock populations.

Literature Cited

Banos, G., G. R. Wiggans, and R. L. Powell. 2001. Impact of Paternity Errors in Cow Identification on Genetic Evaluations and International Comparisons. Journal of Dairy Science. 84(11):2523–2529.

Chen, C. Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale. Journal of Animal Science. 89(9):2673–2679.

Davis, G. P., and S. K. DeNise. 1998. The impact of genetic markers on selection. Journal of Animal Science. 76(9):2331–2339.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. Genetics Selection Evolution. 43(1):1–7.

Grashei, K. E., J. Ødegård, and T. H. E. Meuwissen. 2018. Using genomic relationship likelihood for parentage assignment. Genetics Selection Evolution. 50(26):1–11.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 177(4):2389–2397.

Hayes, B. J. 2011. Technical note : Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. Journal of Dairy Science. 94(4):2114–2117.

Henderson, C.R. 1973. Sire evaluation and genetic trends. Proceedings of the Animal Breeding and Genetics Symposium in Honor of J.L. Lush. American Society for Animal Science. 10-41.

Huisman, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Molecular Ecology Resources. 17:1009–1024.

Israel, C., and J. I. Weller. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. Journal of Dairy Science. 83(1):181–187.

Jiménez-Montero, J. A., O. González-Recio, and R. Alenda. 2013. Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. Journal of Dairy Science. 96(1):625–34.

Kaseja, K., A. Mclaren, J. Yates, S. Mucha, G. Banos, and J. Conington. 2018. Estimation of breeding values for footrot and mastitis in UK Texel sheep. Proceedings of the 11th world Congress on Genetics Applied to Livestock Production. 11:552.

Koivula, M., I. Strandén, G. Su, and E.A. Mäntysaari. 2012. Different methods to calculate genomic predictions-comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). Journal of Dairy Science. 95(7):4065–4073.

McClure, M.C., J. McCarthy, P. Flynn, R. Weld, M. Keane, K. O'Connel, M.P. Mullen, S.Waters, and J.F. Kearney. 2015. SNP selection for nationwide parentage verification and identification in beef and dairy cattle. Proceedings, International Committee For Animal Recording Technical Series. June 2015:175-181.

Meuwissen, T. H. E., B. J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157(4):1819–1829.

Muir, W. M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. Journal of Animal Breeding and Genetics. 124(6):342–355.

Ram, B. N., C.J. Guerrini, and A.L. Mcguire. 2018. Genealogy databases and the future of criminal investigation. Science. 360(6393):1078–1080.

Spelman, R.J. 2002. Utilisation of molecular information in dairy cattle breeding. Proceedings of the 7th world Congress on Genetics Applied to Livestock Production. 22:1-7.

Stranden, I., and K. Vuori. 2006. RelaX2: Pedigree analysis program. Proceedings of the 8th world Congress on Genetics Applied to Livestock Production. 27-30.

Van Raden, P. M. 2007. Genomic Measures of Relationship and Inbreeding. Interbull Bulletin. 25(37):111–114.

Van Raden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science. 91(11):4414–23.

Van Raden, P. M., T.A. Cooper, G.R. Wiggans, J.R.O. Connell, and L.R. Bacheller. 2013. Confirmation and discovery of maternal grandsires and great-grandsires in dairy cattle. Journal of Dairy Science. 96(3):1874–1879.

VanRaden, P. M., J.R. O'Connell, G.R. Wiggans, and K.A. Weigel. 2011. Genomic evaluations with many more genotypes. Genetics, Selection, Evolution. 43(1):10.

Visscher, P. M., J. A. Woolliams, D. Smith, and J.L. Williams. 2002. Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. Journal of Dairy Science. 85(9):2368–2375.

Wang, H., I. Misztal, and A. Legarra. 2014. Differences between genomic-based and pedigree-based relationships in a chicken population, as a function of quality control and pedigree links among individuals. Journal of Animal Breeding and Genetics. 131(6):445–451.

Wiggans, G. R., P.M. Vanraden, and L.R. Bacheller. 2018. Methods for Discovering and Validating Relationships among Genotyped Animals. Interbull Bulletin. (53):10–13.

Table 1: Genomic relationship coefficients between all animals and those animals with inbreeding coefficients <7%, and that have been pedigree verified using Single Nucleotide Polymorphism based opposing homozygote methods.

Relationship Type	Theoretical relationship	Allowed inbreeding coefficient 0-100%					Allowed inbreeding coefficient 0-7%				
		N[1]	Avg[1]	Std[1]	Min[1]	Max[1]	N[1]	Avg[1]	Std[1]	Min[1]	Max[1]
Parents	0.5	3167	0.58	0.03	0.41	0.86	2991	0.58	0.03	0.41	0.71
Grandparents	0.25	1797	0.32	0.05	0.17	0.67	1684	0.32	0.04	0.17	0.49
Great grandparents	0.125	1083	0.19	0.05	0.07	0.7	1017	0.19	0.04	0.08	0.44
Great great grandparents	0.06	256	0.14	0.04	0.05	0.32	248	0.14	0.04	0.05	0.32
Full siblings	0.5	83	0.59	0.06	0.46	0.75	67	0.57	0.05	0.46	0.69
Half siblings	0.25	27625	0.32	0.04	0.17	0.57	24407	0.32	0.04	0.17	0.56

[1] N is the number of relationship pairs contributing to the category; Avg is the average genomic relationship coefficient; Std is the standard deviation genomic relationship coefficient; Min is the minimum genomic relationship coefficient; Max is the maximum genomic relationship coefficient.

433 **Table 2: Genomic relationship coefficients between all animals in the genotyped population based on the reported pedigree information.**

Relationship Type	Theoretical relationship	N[1]	Avg[1]	Std[1]	Min[1]	Max[1]	% below threshold[2]
Parents	0.5	3250	0.57	0.08	0.03	0.71	2.5
Grandparents	0.25	5184	0.32	0.07	0.01	0.86	3.6
Great grandparents	0.125	7819	0.19	0.05	0.02	0.76	1.7
Great great grandparents	0.06	8720	0.14	0.05	0.01	0.65	0.9
Full siblings	0.5	827	0.56	0.08	0.07	1.09	4.0
Half siblings	0.25	60289	0.30	0.06	0.01	0.80	2.9

434 [1] N is the number of relationship pairs contributing to the category; Avg is the average genomic relationship coefficient; Std is the standard
435 deviation genomic relationship coefficient; Min is the minimum genomic relationship coefficient; Max is the maximum genomic relationship
436 coefficient.

437 [2] the threshold applied is the minimum genomic relationship coefficient reported in Table 1 for each relationship type category.

438

439 **Table 3: Case study of available information in the pedigree verification and discovery report for an individual animal born in 2014.**

Relationship type	Information captured in the pedigree verification and discovery report and its interpretation
Progeny	There are 20 progeny, 1 of which is genotyped with a genomic relationship coefficient of 0.57 - which is above the minimum animal - parent threshold of 0.41.
Parents	None genotyped, but from paternal half sibling information there is reasonable confidence that the reported sire is correct.
Paternal half siblings	There are 61 paternal half siblings with genomic relationship coefficients ranging from 0.26 to 0.37, all these half siblings are above the minimum threshold of 0.17, supporting that they truly are half siblings. From this information we can then be reasonably confident that the reported sire is correct, even though we do not have the sire's genotype available to test.
Grandparents	Both paternal and maternal grandsires are genotyped with genomic relationship coefficients of 0.34 and 0.05, respectively. The lower than 0.17 threshold suggests that the reported maternal grandsire is not the true grandsire. This could be that the sire of the dam is incorrect, or that the dam has been incorrectly recorded.
Great grandparents	There are 4 in total genotyped. On the paternal side, both parents of the paternal grandsire are genotyped with genomic relationship coefficients of 0.18 and 0.23 for the great grand sire and great grand dam, respectively. On the maternal side, both great grand sires are genotyped and have genomic relationship coefficients of 0.07 and 0.06, both of which is lower than the threshold of 0.07 suggesting they may not be true great grandparents. This suggests that both the

	sire and dam of the animals dam are incorrect, or that the dam has been incorrectly reported.
Great great grandparents	There were 2 genotyped. On the paternal side, a great great grand sire had a genomic relationship coefficient of 0.15, and on the maternal sire, the great great grand sire genomic relationship coefficient =0.11. Both of these animals have values above the threshold of 0.05 suggesting that these may be the true relationships. However, at this distant a relationship it is also possible that they are not related since unrelated animals have been shown to have average genomic relationships of 0.09.
Half aunts/uncles	There were 56 genotyped aunts/uncles based on the pedigree. When tested, there were 45 with genomic relationship coefficients ranging from 0.125 to 0.32, and above the threshold of 0.125 (half aunt/uncle) and 11 which have genomic relationship coefficients of 0.04 to 0.08 and thus unlikely to be an aunt/uncle. A high level of failures here is expected when an grandparent has been incorrectly recorded.
Half niece/nephews	There were 11 genotyped niece/nephews based on the pedigree. When tested, there were 10 with genomic relationship coefficients ranging from 0.16 to 0.26, and above the threshold of 0.125 (half niece/nephews) and 1 which has a genomic relationship coefficient of 0.09 and thus unlikely to be an niece/nephews. A high level of failures here is expected when an parent has been incorrectly recorded.
Potential close relatives	There were 36 reported with genomic relationship values of 0.17 and higher, suggesting they are closer relatives. The top 4 animals in the list and the outcome of investigation is listed; 1. genomic relationship coefficient =0.40 – a female born in 1998. Given the age range and genetic relationship it is

possible that she is the dam, but more likely the grand-dam of animal.

2. genomic relationship coefficient =0.31 – a paternal sibling that was incorrectly recorded in the pedigree.
3. genomic relationship coefficient =0.30 – a paternal sibling that was incorrectly recorded in the pedigree.
4. genomic relationship coefficient =0.29 – a male born in 2001. Given the age range and genetic relationship it is possible that he is the grand-sire of animal.

After discussion with the breeder it was identified that matings between animals 1 and 4 on the list did occur and he supplied some candidate dams to test and it was confirmed that the pedigree recorded for the dam was incorrect, and after DNA verification was corrected to be the correct dam, which was a daughter of animals 1 and 4 in the above list.