

Technical note

# Revised calculation of Kalinowski's ancestral and new inbreeding coefficients

Harmen P. Doekes<sup>1,2\*</sup>, Ino Curik<sup>3</sup>, István Nagy<sup>4</sup>, János Farkas<sup>5</sup>, György Kövér<sup>5</sup>, Jack J. Windig<sup>1,2</sup>

<sup>1</sup> Animal Breeding and Genomics, Wageningen University & Research, P.O. Box 338, 6700 AH, Wageningen, the Netherlands

<sup>2</sup> Centre for Genetic Resources the Netherlands, Wageningen University & Research, P.O. Box 16, 6700 AA, Wageningen, the Netherlands

<sup>3</sup> University of Zagreb, Faculty of Agriculture, Department of Animal Science, Zagreb, Croatia

<sup>4</sup> Kaposvár University, Faculty of Agricultural and Environmental Sciences, Kaposvár, Hungary

<sup>5</sup> Kaposvár University, Faculty of Economic Science, Kaposvár, Hungary

\* Correspondence: harmen.doekes@wur.nl

Received: date; Accepted: date; Published: date

**Abstract:** To test for the presence of purging in populations, the classical pedigree-based inbreeding coefficient ( $F$ ) can be decomposed into Kalinowski's ancestral ( $F_{ANC}$ ) and new ( $F_{NEW}$ ) inbreeding coefficients. The  $F_{ANC}$  and  $F_{NEW}$  can be calculated by a stochastic approach known as gene dropping. However, the only publicly available algorithm for calculation of  $F_{ANC}$  and  $F_{NEW}$ , implemented in GRain v2.1 (and in the PEDIG software package), produced biased estimates. The  $F_{ANC}$  was systematically underestimated and, consequently,  $F_{NEW}$  was overestimated. To illustrate this bias, we calculated  $F_{ANC}$  and  $F_{NEW}$  by hand for simple example pedigrees. We revised the GRain program such that it now provides unbiased estimates. Correlations between biased and unbiased estimates of  $F_{ANC}$  and  $F_{NEW}$ , obtained for example data sets of Hungarian Pannon White rabbits and Dutch Holstein Friesian cattle, were high ( $> 0.96$ ). Although the magnitude of bias appeared to be small, results from studies based on biased estimates should be interpreted with caution. The revised GRain program (v2.2) is now available online and can be used to calculate unbiased estimates of  $F_{ANC}$  and  $F_{NEW}$ .

**Keywords:** ancestral inbreeding, new inbreeding, purging, gene dropping, inbreeding depression

## 1. Introduction

Inbreeding is the mating between (close) relatives and is unavoidable in genetically small populations. The degree of inbreeding is typically measured with inbreeding coefficients. Individuals with higher inbreeding coefficients show a lower phenotypic performance on average, a phenomenon known as inbreeding depression [1-3]. Inbreeding depression occurs because part of the genetic load in populations, known as inbreeding load, is only expressed in homozygotes [1]. Inbreeding depression is expected to be largely due to partial dominance, i.e. the existence of (partially) deleterious recessive alleles, although overdominance and epistasis may also play a role [1, 2, 4].

Inbreeding load in a population is not constant, but rather dynamic over time. New deleterious recessive alleles arise continuously by mutation and these alleles are eroded over time by (natural and/or artificial) selection and genetic drift [1]. Inbreeding increases the efficiency of selection against deleterious recessive alleles in a process called purging [1, 5].

To test for the existence of purging in populations, various pedigree-based methods have been proposed [6-8]. Ballou [6] introduced the ancestral inbreeding coefficient, which is the probability that a random allele in an individual has been previously exposed to inbreeding, i.e. that this allele has been identical-by-descent (IBD) in at least one ancestor. Kalinowski et al. [7] extended this concept

by considering the IBD-status of the individual itself as well. In the Kalinowski approach the total pedigree-based inbreeding coefficient is decomposed into an ancestral ( $F_{ANC}$ ) and a new ( $F_{NEW}$ ) inbreeding coefficient. The  $F_{ANC}$  is the probability that alleles are IBD in the individual while they were already IBD in at least one ancestor, while  $F_{NEW}$  is the probability that alleles are IBD for the first time in the individual's pedigree [7].

To calculate  $F_{ANC}$  and  $F_{NEW}$  (and other inbreeding coefficients), a gene dropping based algorithm has been developed and implemented in GRain v2.1 software [9]. The same algorithm has also been incorporated in the PEDIG package [10], versions 2007 and later. Various studies have used GRain v2.1 [11–15] and PEDIG [16–18] to calculate  $F_{ANC}$  and  $F_{NEW}$ .

The objective of this study was to demonstrate that the algorithm in GRain v2.1 (and PEDIG) produced biased estimates of  $F_{ANC}$  and  $F_{NEW}$ . For several simple pedigrees, we show how  $F_{ANC}$  and  $F_{NEW}$  can be calculated by hand. We also investigate the magnitude of the bias for two example data sets, one of Hungarian Pannon White rabbits and one of Dutch Holstein Friesian dairy cattle. A revised version of the GRain software (v2.2), which provides unbiased calculations of  $F_{ANC}$  and  $F_{NEW}$ , is now available online.

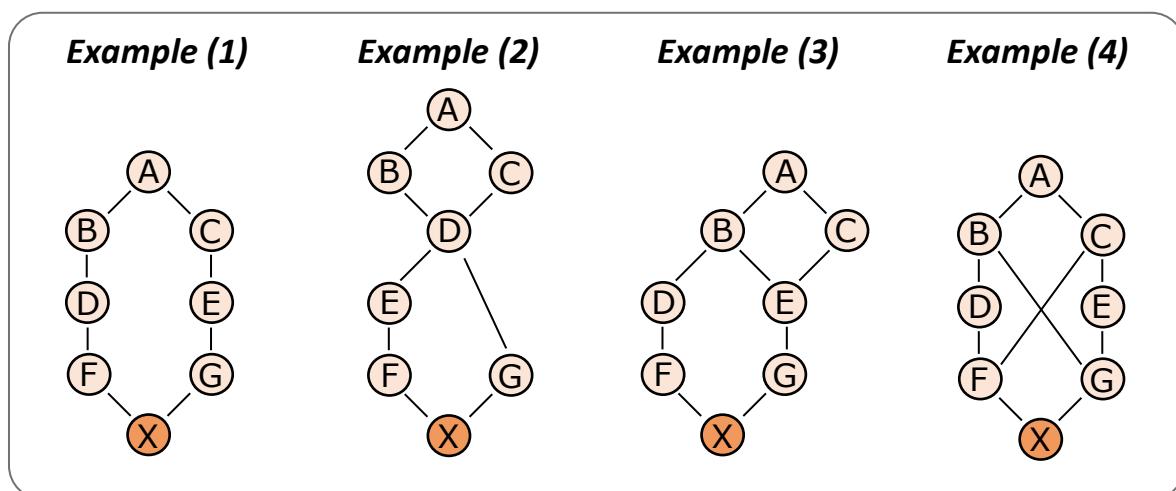
## 2. Calculation of ancestral and new inbreeding coefficients by hand

For simple pedigrees, Kalinowski's ancestral inbreeding ( $F_{ANC,X}$ ) and new inbreeding ( $F_{NEW,X}$ ) coefficients of an individual X can be calculated by hand, following Mendelian inheritance principles. First, the classical inbreeding coefficient ( $F_X$ ) has to be determined. The  $F_X$  is defined as the probability that the two alleles at a random locus in individual X are IBD, and is calculated as [19]:

$$F_X = \sum_{i=1}^n (1 + F_i) \left(\frac{1}{2}\right)^{k+1}$$

where  $n$  is the number of paths connecting one parent of X with the other parent of X through common ancestor  $i$ ,  $F_i$  is the inbreeding coefficient of the  $i^{th}$  common ancestor, and  $k$  is the number of steps in the path. Then,  $F_{ANC,X}$  is calculated as the probability that X is IBD for an allele, given that this allele was also IBD in at least one of the ancestors of X. Finally,  $F_{NEW,X}$  can be obtained by subtracting  $F_{ANC,X}$  from  $F_X$ , since the ancestral and new inbreeding sum up to the total inbreeding.

In Figure 1, four example pedigrees are shown. The corresponding inbreeding coefficients are provided in Table 1. In example (1), the  $F_X$  equals 0.0078, because it is the inbreeding on ancestor A ( $0.5^7$ ). The  $F_{ANC,X}$  for this example is 0, because none of the ancestors of X is inbred. Consequently,  $F_{NEW,X}$  is equal to  $F_X$  (so 0.0078).



**Figure 1.** Example pedigrees for calculation of classical and Kalinowski's inbreeding coefficients. Corresponding inbreeding coefficients are shown in Table 1.

**Table 1.** Inbreeding coefficients for four example pedigrees (Fig. 1), estimated with revised and previous version of GRain.

Pedigree	$F_X$	Revised version (v2.2)		Previous version (v2.1)		Difference in $F_{ANC,X}$
		$F_{ANC,X}$	$F_{NEW,X}$	$F_{ANC,X}$	$F_{NEW,X}$	
(1)	0.0078	0	0.0078	0	0.0078	0
(2)	0.0703	0.0156	0.0547	0.0156	0.0547	0
(3)	0.0390	0.0078	0.0312	0.0039	0.0351	0.0039
(4)	0.1641	0.0390	0.1250	0.0234	0.1406	0.0156

$F_X$ : classical inbreeding coefficient of individual X,  $F_{ANC,X}$ : Kalinowski's ancestral inbreeding coefficient of individual X,  $F_{NEW,X}$ : Kalinowski's new inbreeding coefficient of individual X.

In example (2), the  $F_X$  equals 0.0703, because it is the inbreeding on ancestor D ( $0.5^4$ ) multiplied with  $[1 + F_D]$ , where  $F_D$  is the inbreeding coefficient of ancestor D ( $0.5^3$ ). The  $F_{ANC,X}$  is calculated as the probability that X is IBD for an allele that was IBD in D as well. Since D is the only inbred ancestor, we do not need to consider the IBD status of all other ancestors. The probability that D is IBD for an allele from its grandparent A is 0.125 ( $0.5^3$ ). To obtain  $F_{ANC,X}$ , this probability has to be multiplied with the probabilities that the allele is transferred to X through both the paths D-E-F-X and D-G-X. The probability that E inherits the allele from D is simply 1, because D is IBD. The probability that F inherits the allele from E is 0.5 and that X inherits it from F is also 0.5, so the total probability for the path D-E-F-X is 0.25 ( $0.5^2$ ). Similarly, the probability for path D-G-X is 0.5. This gives a total probability of  $0.125 * 0.25 * 0.5 = 0.0156$  for  $F_{ANC,X}$ . Consequently,  $F_{NEW,X} = F_X - F_{ANC,X} = 0.0703 - 0.0156 = 0.0547$ .

In example (3), the  $F_X$  equals 0.0390 and is the sum of inbreeding on ancestor A ( $0.5^7$ ) and ancestor B ( $0.5^5$ ). The  $F_{ANC,X}$  is calculated as the probability that X is IBD for an allele that was IBD in ancestor E as well. Since ancestor E is the only inbred ancestor, we do not need to consider the IBD status of all other ancestors. The probability that E is IBD for an allele from its grandparent A is 0.125 ( $0.5^3$ ). This probability has to be multiplied by the probability that this allele is transferred to X through both the path E-G-X and B-D-F-X. The probability that G inherits the allele from E is 1, because E is IBD. The probability that X inherits the allele from G is 0.5, so the total probability for the path E-G-X is 0.5. The probability that B carries the allele is 1, otherwise E could not have been IBD. The probability that the allele is transferred from B to D, to F and to X is 0.125 ( $0.5^3$ ). This gives a total probability of  $0.125 * 0.125 * 0.5 = 0.0078$  for  $F_{ANC,X}$ . Consequently,  $F_{NEW,X} = F_X - F_{ANC,X} = 0.0390 - 0.0078 = 0.0312$ .

In example (4), the  $F_X$  equals 0.1641 and is the sum of inbreeding on ancestor A ( $0.5^7 + 0.5^5$ ), ancestor B ( $0.5^4$ ) and ancestor C ( $0.5^4$ ). The  $F_{ANC,X}$  in this example is the probability that X is IBD for an allele that was also IBD in F and/or G (since F and G are inbred ancestors). The  $F_{ANC,X}$  is the sum of the probabilities for three scenarios: (i) X is IBD for an allele that was IBD in both F and G, (ii) X is IBD for an allele that was IBD in F, but not in G, and (iii) X is IBD for an allele that was IBD in G, but not in F. The probability that F is IBD for an allele from A is 0.0625 ( $0.5^4$ ). In that case both B and C must be carriers of that same allele and the probability that G is also IBD for the same allele is 0.125 ( $0.5^3$ ). When F and G are IBD for the same allele, X has to be IBD for that allele as well. Therefore, the probability that scenario (i) happens is 0.0078 (i.e.  $0.0625 * 0.125 * 1$ ). If F is IBD for an allele from A, the probability that G carries two other "unknown" alleles is 0.375 (i.e.  $0.5 * [1 - 0.5^2]$ ), leaving  $1 - 0.125 - 0.375 = 0.5$  for the probability that G carries one copy of the allele and one copy of an unknown allele (scenario ii). In that case the probability that the allele is inherited by X from G is 0.5. The total probability for scenario (ii) is thus 0.0156 (i.e.  $0.0625 * 0.5 * 0.5$ ). Because of the symmetry in the pedigree, the probability for scenario (iii) is equal to that of scenario (ii), so also 0.0156. Thus, the total probability that X is IBD for an allele that was also IBD in F and/or G, i.e. the  $F_{ANC,X}$ , equals  $0.0078 + 0.0156 + 0.0156 = 0.0391$ . Consequently,  $F_{NEW,X} = F_X - F_{ANC,X} = 0.1641 - 0.0391 = 0.1250$ .

### 3. Underestimation of ancestral inbreeding by previous version of GRain

When  $F_{ANC,X}$  was computed with the previous version of GRain (v2.1), the  $F_{ANC,X}$  for examples (1), (2), (3) and (4) from Figure 1 equaled 0, 0.0156, 0.0039 and 0.0234, respectively (Table 1). Although the coefficients for examples (1) and (2) were correct, the  $F_{ANC,X}$  coefficients for examples (3) and (4) were underestimated. Note that example (3) is equivalent to the example used by McParland et al. [17], Figure 1a in their paper, for which they reported the incorrect  $F_{ANC,X}$  estimate of 0.0039.

The underestimation of  $F_{ANC,X}$  was caused by a sometimes incorrect tracking of IBD-status of ancestors throughout the pedigree. In the previous version of GRain (v2.1), each individual was given a flag that indicated whether one of the individual's ancestors had been IBD (1 if true, 0 if false). This flag was calculated as the sum of the flags of the parents, divided by two. Thus, when both parents had a flag of 1, the flag of the offspring would also be 1, which is correct. However, when only one of the parents had a flag of 1 (and the other 0), the offspring would get a value of 0.5, which is incorrect. In the revised version of GRain (v2.2), this issue was solved by obtaining the flag of an offspring as the maximum of the flags of its parents.

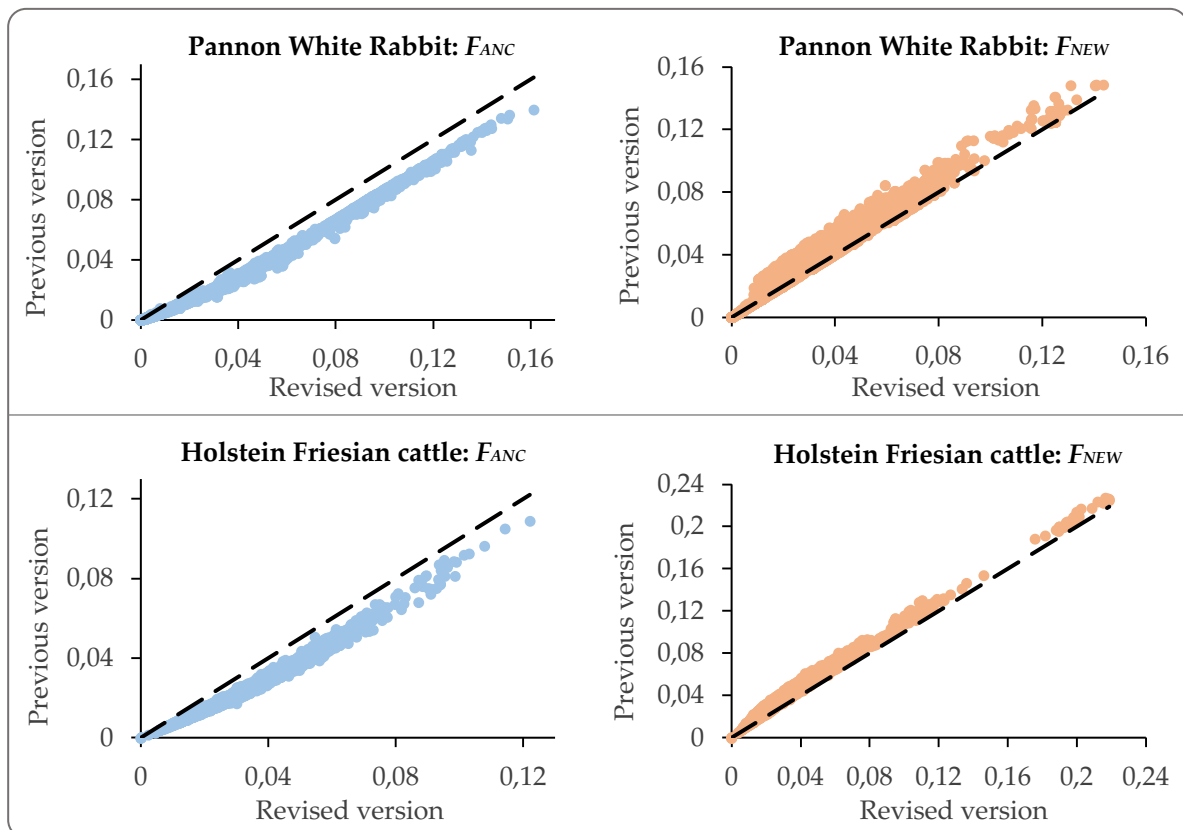
To clarify, in example (2) in Figure 1, whenever ancestor D was IBD, both parents F and G had a flag of 1 and X also got a flag of 1. Therefore, the  $F_{ANC,X}$  was estimated correctly. In example (3), however, whenever ancestor E was IBD, parent G had a flag of 1 and parent F had a flag of 0 and, as a result, X got a flag of 0.5. Therefore, the  $F_{ANC,X}$  for example (3) was underestimated by exactly a factor two. In example (4), whenever both F and G were IBD, X got a flag of 1. This happened in 0.0078 of the simulations (see explanation in the previous section for calculation by hand, scenario (i)). When only parent F or parent G was IBD, while the other parent was not, X got a flag of 0.5. This happened in  $0.0156 + 0.0156 = 0.0312$  of the simulations (see explanation in the previous section for calculation by hand, scenarios (ii) and (iii)). Therefore, the  $F_{ANC,X}$  for example (4) was underestimated by a factor between one and two. More precisely, the underestimated  $F_{ANC,X}$  was equal to  $0.0078 + (0.5 * 0.0312) = 0.0234$ .

### 4. Examples for Pannon White Rabbits and Holstein-Friesian cattle

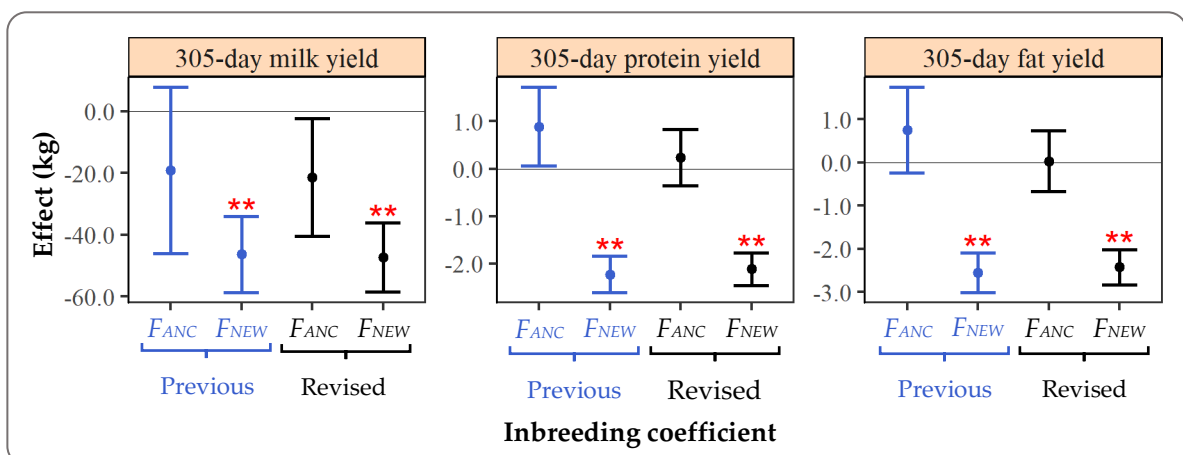
To investigate the impact of the incorrect estimation, we computed  $F_{ANC}$  and  $F_{NEW}$  for two example data sets, using both the previous and revised version of GRain, and  $10^6$  replications. The first data set was a pedigree of 22,781 rabbits of the Hungarian Pannon White (PW) breed. The second data set contained 37,061 Dutch Holstein-Friesian (HF) cows, which were part of a larger pedigree of 167,924 individuals and were used by Doekes et al. [20] to investigate the effects of ancestral and new inbreeding on various traits in HF cattle.

For both the PW and HF data set, the total inbreeding coefficients ( $F$ ) were identical between the previous and revised version of GRain (v2.1 versus v2.2). The  $F_{ANC}$  in the previous version, however, was generally underestimated and the  $F_{NEW}$  was overestimated (Figure 2). For the PW data set and for inbreeding coefficients above zero, the  $F_{ANC}$  from the previous version was on average 0.65 times the revised  $F_{ANC}$  (and the  $F_{NEW}$  was 1.27 times the revised  $F_{NEW}$ ). For the HF data set and inbreeding coefficients above zero, the  $F_{ANC}$  from the previous version was on average 0.71 times the revised  $F_{ANC}$  (and the  $F_{NEW}$  was 1.36 times the revised  $F_{NEW}$ ). Pearson correlation coefficients between coefficients estimated with the previous and revised version were high. For the PW data set, the correlations between the previous and revised version equaled 0.997 and 0.968 for  $F_{ANC}$  and  $F_{NEW}$ , respectively. For the HF data set, these correlations equaled 0.993 and 0.987, respectively. This indicates that the underestimation of  $F_{ANC}$  (and overestimation of  $F_{NEW}$ ) did not strongly affect the ranking of animals.

For the Holstein-Friesian data set, we also investigated potential differences in inbreeding depression estimates for  $F_{ANC}$  and  $F_{NEW}$  calculated with the previous and revised version of GRain. In general, differences were rather small (Figure 3). For example, the effects of  $F_{ANC}$  on 305-day milk yield were -19.1 kg for the previous version and -21.4 kg for the revised version. Standard errors for the inbreeding depression effects were smaller when the revised version of GRain was used to estimate  $F_{ANC}$  and  $F_{NEW}$  than when the previous version was used. The overall conclusion that new inbreeding is more harmful than ancestral inbreeding, however, was the same for both versions.



**Figure 2.** Relationship between Kalinowski's inbreeding coefficients calculated with previous (v2.1) and revised (v2.2) version of Grain, for two example data sets of Pannon White rabbits ( $n = 22,781$ ) and Holstein Friesian cattle ( $n = 37,061$ ). The dashed line indicates  $y = x$ , i.e. a relationship in which there is no difference in the estimation.  $F_{ANC}$ : Kalinowski's ancestral inbreeding coefficient.  $F_{NEW}$ : Kalinowski's new inbreeding coefficient.



**Figure 3.** Effect of a 0.01 increase in Kalinowski's ancestral ( $F_{ANC}$ ) and new ( $F_{NEW}$ ) inbreeding on yield traits in Dutch Holstein Friesian cattle ( $n = 37,061$ ), for  $F_{ANC}$  and  $F_{NEW}$  calculated with the previous (v2.1) and revised (v2.2) version of GRain. The same model was used as in Doekes et al. [20].

## 5. Conclusions

The previous version of GRain software (v2.1) systematically underestimated Kalinowski's ancestral inbreeding and, consequently, overestimated Kalinowski's new inbreeding coefficients. Although the magnitude of bias was rather small, results from studies based on biased estimates should be interpreted with caution. The GRain software has been revised and the revised version (v2.2), which provides unbiased estimates of Kalinowski's inbreeding coefficients, can be downloaded from <https://boku.ac.at/nas/nuwi/software/>.

**Author Contributions:** HD and JW performed the Holstein Friesian data analysis; JF and GK performed the Pannon White data analysis; HD prepared the manuscript. HD, IC, IN, JF, GK and JW participated in the interpretation of results and revision of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** Calculations performed on the Dutch Holstein Friesian population were conducted as part of the IMAGE project, which received funding from the European Union's Horizon 2020 Research and Innovation Programme under the grant agreement no 677353. Calculations performed on the Pannon White rabbit population were supported by the Hungarian Scientific Research Fund (OTKA) K 128177 project. The study was co-funded by the Dutch Ministry of Agriculture, Nature and Food Quality (KB-34-013-002).

**Acknowledgments:** The authors would like to thank the Dutch-Flemish cattle improvement cooperative (CRV) for providing the Holstein Friesian data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hedrick P.W., García-Dorado A. Understanding inbreeding depression, purging, and genetic rescue. *Trends in ecology & evolution*. **2016**, *31*, 940-952. doi:10.1016/j.tree.2016.09.005
- Howard J.T., Pryce J.E., Baes C., Maltecca C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *Journal of Dairy Science*. **2017**, *100*, 6009-6024. doi:10.3168/jds.2017-12787
- Leroy G. Inbreeding depression in livestock species: review and meta-analysis. *Animal genetics*. **2014**, *45*, 618-628. doi:10.1111/age.12178
- Charlesworth D., Willis J.H. The genetics of inbreeding depression. *Nature Reviews Genetics*. **2009**, *10*, 783. doi:10.1038/nrg2664
- García-Dorado A. Understanding and predicting the fitness decline of shrunk populations: inbreeding, purging, mutation, and standard selection. *Genetics*. **2012**, *190*, 1461-1476. doi:10.1534/genetics.111.135541
- Ballou J.D. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *Journal of Heredity*. **1997**, *88*, 169-178. doi:10.1093/oxfordjournals.jhered.a023085
- Kalinowski S.T., Hedrick P.W., Miller P.S. Inbreeding depression in the Speke's gazelle captive breeding program. *Conservation Biology*. **2000**, *14*, 1375-1384. doi:10.1046/j.1523-1739.2000.98209.x
- García-Dorado A., Wang J., López-Cortegano E. Predictive model and software for inbreeding-purging analysis of pedigreed populations. *G3: Genes, Genomes, Genetics*. **2016**, *6*, 3593-3601. doi:10.1534/g3.116.032425
- Baumung R., Farkas J., Boichard D., Mészáros G., Sölkner J., Curik I. GRAIN: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. *Journal of Animal Breeding and Genetics*. **2015**, *132*, 100-108. doi:10.1111/jbg.12145
- Boichard D. PEDIG: a fortran package for pedigree analysis suited for large populations. *Proceedings of the 7th world congress on genetics applied to livestock production*. **2002**, *32*, 525-528.
- Ács V., Bokor Á., Nagy I. Population Structure Analysis of the Border Collie Dog Breed in Hungary. *Animals*. **2019**, *9*, 250. doi:10.3390/ani9050250
- Addo S., Schäler J., Hinrichs D., Thaller G. Genetic Diversity and Ancestral History of the German Angler and the Red-and-White Dual-Purpose Cattle Breeds Assessed through Pedigree Analysis. *Agricultural Sciences*. **2017**, *8*, 1033. doi:10.4236/as.2017.89075
- Schäler J., Krüger B., Thaller G., Hinrichs D. Comparison of ancestral, partial, and genomic inbreeding in a local pig breed to achieve genetic diversity. *Conservation Genetics Resources*. **2018**, *1-10*. doi:10.1007/s12686-018-1057-5
- Todd E.T., Ho S.Y., Thomson P.C., Ang R.A., Velie B.D., Hamilton N.A. Founder-specific inbreeding depression affects racing performance in Thoroughbred horses. *Scientific Reports*. **2018**, *8*, 6167. doi:10.1038/s41598-018-24663-x
- Vostry L., Milerski M., Schmidova J., Vostra-Vydrova H. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. *Small Ruminant Research*. **2018**, *168*, 25-31. doi:10.1016/j.smallrumres.2018.09.004

16. Hinrichs D., Bennewitz J., Wellmann R., Thaller G. Estimation of ancestral inbreeding effects on stillbirth, calving ease and birthweight in German Holstein dairy cattle. *Journal of Animal Breeding and Genetics*. **2015**, 132, 59-67. doi:10.1111/jbg.12114
17. McParland S., Kearney F., Berry D.P. Purging of inbreeding depression within the Irish Holstein-Friesian population. *Genetics Selection Evolution*. **2009**, 41, 16. doi:10.1186/1297-9686-41-16
18. Roos L., Hinrichs D., Nissen T., Krieter J. Investigations into genetic variability in Holstein horse breed using pedigree data. *Livestock Science*. **2015**, 177, 25-32. doi:10.1016/j.livsci.2015.04.013
19. Wright S. Coefficients of inbreeding and relationship. *American Naturalist*. **1922**, 56, 330-338. doi:10.1086/279872
20. Doekes H.P., Veerkamp R.F., Bijma P., de Jong G., Hiemstra S.J., Windig J.J. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genetics Selection Evolution*. **2019**, 51, 54. doi:10.1186/s12711-019-0497-z



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).