

Genomic Methods for Rotational Crossbreeding in Local Dairy Cattle Breeds

Dissertation

To obtain the degree of Agricultural Sciences (Dr. sc. agr.)

Faculty of Agricultural Sciences
University of Hohenheim
Institute of Animal Science

Submitted by
Joana Stock
Born in Herford, Germany

2022

Date of the oral examination: 12.05.2022

Examination Committee:

Chairperson of the oral examination	Prof. Dr. Martin Hasselmann
Supervisor and Reviewer	Prof. Dr. Jörn Bennewitz
Co-Reviewer	Prof. Dr. Dirk Hinrichs
Additional examiner	Prof. Dr. Mizeck Chagunda

Table of Contents

Summary	1
Zusammenfassung	3
Chapter 1 General Introduction	5
Chapter 2 A Review of Genomic Models for the Analysis of Livestock Crossbred Data	9
Chapter 3 Improving the Accuracy of Multi-Breed Prediction in Admixed Populations by Accounting for the Breed Origin of Haplotype Segments	36
Chapter 4 Implementing a Genomic Rotational Crossbreeding Scheme to Promote Local Dairy Cattle Breeds – A Simulation Study	53
Chapter 5 Genomic Rotational Crossbreeding with Advanced Optimum Contribution Selection Methods Applied to Simulated German Angler and German Holstein Dairy Cattle Populations	74
Chapter 6 General Discussion	93
Acknowledgments	110
Curriculum Vitae	111
Eidesstattliche Versicherung	112

Summary

Local dairy breeds, such as German Angler, usually have small population sizes and thus a reduced genetic gain, compared to high-yielding breeds. Especially since genomic selection is widely used in the latter, the performance gap between local breeds and high-yielding breeds increased further, as it requires large reference populations in order to achieve accurate estimated breeding values. As a result, many farmers switched to high-yielding breeds. On the other hand, to increase the performance of local breeds the introgression of high-yielding breeds was a common strategy in the past, which resulted in high amounts of foreign genetic material in many of them. Much of the original genetic background got lost, however, they do not achieve the same performance level as high-yielding breeds. Local breeds are therefore faced with the risk of two types of extinction, i.e. a numerical extinction due to the small and decreasing numbers of breeding animals, and a genetic extinction due to massive introgression from high-yielding breeds.

To promote local dairy breeds, the implementation of a genomic rotational crossbreeding scheme can be a promising strategy. Local breeds can benefit from a genomic rotational crossbreeding scheme with a high-yielding breed due to 1) an enlarged reference population including both the local breed and crossbred animals, and 2) the increased performance level of crossbred animals. On the other hand, crossbreeding is particularly known to improve functional traits by the exploitation of heterosis. Thus, it appears to be an appealing option for high-yielding breeds, as well, as they tend to struggle with fitness related problems.

This thesis aimed to develop genomic methods for numerically small local dairy breeds in crossbreeding schemes in order to improve their genetic gain, genetic uniqueness, and their ability to compete with high-yielding breeds.

In **Chapter 2** a review study conducted a comparison of different genomic models which are suitable for crossbred data. Different additive models (such as the parental model, a model with breed-specific allele effects, and a single step model) and dominance models, which were either line-dependent, line-independent or included imprinting were discussed. It was concluded that the model choice needs to be made based on desired accuracies, computational possibilities, and data availability. In general, dominance models showed to result in higher accuracies compared to additive models.

A breed of origin of alleles model approach was introduced in **Chapter 3**, which assumes different SNP effects for different origins of haplotypes. This model is suitable for the multi-breed genomic prediction of breeding values of numerically small breeds (i.e. German Angler) that have experienced introgression from high-yielding breeds in the past. The breed of origin

of alleles model approach tended to be advantageous for Angler over multi-breed and within-breed genomic predictions with GBLUP.

Chapter 4 contains a simulation study about the implementation of a rotational crossbreeding scheme including German Angler x German Holstein, while introducing genomic selection in Angler. Different sizes and structures of growing reference populations and selection goals of Angler were examined. The results showed that crossbred animals had a small overall superiority to both Holstein and Angler populations. In addition, a reference population containing both Angler and crossbred animals, in combination with a selection based on the purebred performance of Angler, gave the highest response to selection in the purebred Angler population and in the crossbred population. The difference between selection methods for Angler individuals could only be observed in the long term, as the purebred-crossbred correlations decreased.

In **Chapter 5** a simulation study on rotational crossbreeding was performed including different Optimum Contribution Selection methods, in order to realize genetic gain while regaining the original genetic background of Angler. Different constraints regarding mean kinships, native kinships, and migrant contributions from Holstein were applied to investigate their effects on Angler, crossbred, and Holstein populations. Constraining the amount of migrant contribution in Angler increased their genetic uniqueness. However, it led to a notable reduction of genetic gain and thus a reduced superiority of the crossbred animals. The slowed rate of genetic gain and thus the large difference of the performance between the parental breeds could not be compensated by heterosis effects.

In **Chapter 6** the thesis ends with a general discussion about further genomic models for crossbreeding, and the practical relevance of crossbreeding in dairy cattle.

Zusammenfassung

Lokale Milchviehrassen, wie z.B. die Deutschen Angler, haben weitgehend kleine Populationsgrößen und einen entsprechend langsameren Zuchtfortschritt als Hochleistungsrassen. Durch den Einzug der genomischen Selektion in den Hochleistungsrassen wurde dieser Trend noch verstärkt, da sie eine große Tieranzahl in der Referenzpopulation voraussetzt um genaue Zuchtwerte schätzen zu können. Der Unterschied zwischen den Leistungen lokaler Rassen und Hochleistungsrassen wurde somit weiter vergrößert. Als Resultat wechselten viele Züchter von lokalen Milchviehrassen zu Hochleistungsrassen. Gleichzeitig war in der Vergangenheit die Einkreuzung von Hochleistungsrassen in lokale Rassen eine populäre Strategie um deren Zuchtfortschritt zu verbessern. Sie zeigen daher heute zum Teil einen hohen Fremdgenanteil – dennoch können sie nicht mit dem Leistungsniveau von Hochleistungsrassen mithalten. Lokale Milchviehrassen sind daher durch zwei Arten des Aussterbens bedroht: einerseits zahlenmäßig durch die immer geringer werdende Anzahl an (Zucht)tieren, und andererseits genetisch, durch die massive Einkreuzung von Hochleistungsrassen.

Um lokale Rassen zu fördern kann die Implementierung eines Rotationkreuzungsprogrammes von Nutzen sein, insbesondere wenn zeitgleich die genomische Selektion eingeführt wird. Lokale Rassen können so zum einen von der großen, gemeinsamen Referenzpopulation mit den Kreuzungstieren profitieren, und zum anderen von der Leistungssteigerung durch die Kreuzungstiere. Daneben ist die Kreuzungszucht, durch die Ausnutzung von Heterosiseffekten, insbesondere bei der Verbesserung von funktionalen Merkmalen von Bedeutung. Sie kann dadurch ebenfalls für Züchter von Hochleistungsrassen, die zu Fitnessproblemen neigen, eine attraktive Option sein um ihre Herde zu verbessern.

Das Ziel der Arbeit war es, genomische Methoden für lokale Milchviehrassen mit kleinen Populationsgrößen in Kreuzungszuchtprogrammen zu entwickeln, um ihren Zuchtfortschritt, ihre genetische Eigenständigkeit und ihre Wettbewerbsfähigkeit zu steigern.

In **Kapitel 2** wurden in einem Review Artikel verschiedene genomische Modelle für die Zuchtwertschätzung in Kreuzungszuchtprogrammen analysiert und verglichen. Dabei wurden additive Modelle, wie das Parental Modell, ein Modell mit rassspezifischen Alleleffekten, und ein Single-Step-Modell, mit Dominanzmodellen, die entweder populationsabhängig oder – unabhängig waren, bzw. Imprinting berücksichtigten, diskutiert. Die Wahl, welches genomisches Modell schließlich im Kreuzungszuchtprogramm zur Anwendung kommt, sollte je nach verfügbarer Datenmenge und -art, Rechenkapazitäten und gewünschten Genauigkeiten der geschätzten Zuchtwerte getroffen werden. Allgemein zeigten Dominanzmodelle höhere Genauigkeiten als Modelle, die lediglich additive Effekte berücksichtigten.

Ein genomisches Modell, welches den Ursprung von Allelen bezüglich der Elternrassen berücksichtigen kann, wurde in **Kapitel 3** vorgestellt. Das Modell nimmt an, dass SNP Effekte, abhängig vom Ursprung ihrer Haplotypen, unterschiedlich sein können. Es ist daher passend für die genomische Zuchtwertschätzung unter Einsatz von Referenzpopulationen, die mehrere Rassen beinhalten, um kleinere Rassen mit hohem Fremdgenanteil zu evaluieren. Die Ergebnisse zeigten, dass das vorgestellte Modell, unter Berücksichtigung einer Referenzpopulation von mehreren Rassen, genauere genomische Zuchtwerte schätzen kann als verschiedene andere genomische Ansätze.

Kapitel 4 beinhaltet eine Simulationsstudie zur Implementierung eines Rotationskreuzungsprogrammes mit den Rassen Angler und Holstein, während zeitgleich die genomische Selektion bei Anglern eingeführt wird. Es wurden verschiedene Arten von wachsenden Referenzpopulationen und Selektionszielen in der Anglerpopulation untersucht. Es konnte gezeigt werden, dass die Kreuzungstiere, unabhängig von der Referenzpopulation und dem Selektionsziel, den Holstein- und Angler-Reinzuchtlinien überlegen waren. Die Kombination aus einer die Referenzpopulation, die sowohl Angler, als auch Kreuzungstiere beinhaltete, und einer Selektion basierend auf der Reinzuchtleistung der Angler, zeigte den höchsten Zuchtfortschritt – sowohl in der Anglerpopulation, als auch bei den Kreuzungstieren. Die verschiedenen Selektionsmethoden der Angler zeigten lediglich in den späteren Generationen Unterschiede im Zuchtfortschritt, abhängig von der abnehmenden Korrelation von Reinzucht- und Kreuzungszuchtleistung.

In **Kapitel 5** wurde eine Simulationsstudie zur Implementierung eines Rotationskreuzungsprogrammes mit verschiedenen Arten von Optimum Contribution Selection Methoden durchgeführt. Das Ziel bestand dabei darin, Zuchtfortschritt zu erzielen, während die genetische Eigenständigkeit der Angler wiederhergestellt wird. Verschiedene Restriktionen bezüglich des Inzuchtlevels und Fremdgenanteils durch Holstein wurden angewendet und ihre Auswirkungen auf die Angler-, Kreuzungs-, und Holsteinpopulation untersucht. Die Restriktion des Fremdgenanteils in der Anglerpopulation zeigte eine Erhöhung der genetischen Eigenständigkeit der Rasse. Dennoch führte es zur deutlichen Verlangsamung des Zuchtfortschrittes und somit zu einer verringerten Überlegenheit der Kreuzungstiere. Der reduzierte Zuchtfortschritt in der Anglerpopulation, und somit die großen Leistungsunterschiede der Ausgangsrassen, konnte nicht durch die erzielten Heterosiseffekte ausgeglichen werden.

Die Thesis endet in **Kapitel 6** mit einer allgemeinen Diskussion zu weiteren genomischen Modellen, sowie zur praktischen Relevanz von Kreuzungszuchtprogrammen in der Milchrinderzucht.

Chapter 1

General Introduction

In the last century economic and social changes, urbanisation and policy factors lead to the intensification of livestock production systems (Hoffmann and Scherf, 2005) and therefore to high selection pressure on production traits in dairy cattle. This fact caused a remarkable improvement of genetic gain, especially in high yielding breeds, e.g. Holstein-Friesian. The focus was on milk production and conformation, whereas functional traits such as fertility, health and longevity were considered less. As a consequence deterioration of functional traits could be observed, which was even intensified by the antagonistic correlations between production and functional traits (Rauw *et al.*, 1998; Oldenbroek, 2007). In addition, insufficient health and fertility lead to decreasing milk yield which has an important effect on cow profitability.

On the other hand, several studies (e.g. (Gandini *et al.*, 2007; Curone, 2016, 2018) proved good functional characteristics of local lower yielding breeds. However, they are facing extinction, as their genetics have been replaced by high performance breeds due to economic reasons (Wellmann and Bennewitz, 2019). International markets for animal breeding and animal products threat this development of local production systems as well, as they provide only few different high yielding breeds (Hoffmann and Scherf, 2005). Since arguments for conserving local breeds are recognised by farmers, public and the scientific community, the goal is not to analyse whether it is necessary to conserve them, but how to do it best (Mendelsohn, 2003; Ovaska and Soini, 2017).

One possible solution can be rotational crossbreeding, as it also requires purebreeding of the parental breeds and thus would increase the importance of the local breed. Several studies showed the superiority of crossbred dairy cattle compared to purebreds as crossbred offspring are economically more efficient (Kargo *et al.*, 2021) and showed improved functional traits, such as longevity, fertility, and udder health, compared to their parents' breeds (Clasen *et al.*, 2017; 2019), independently from their management level (Kargo *et al.*, 2012). These advantages led to an increasing interest in crossbreeding in dairy cattle (Sørensen *et al.*, 2008).

In this project, German Angler and German Holstein were used. Angler is a dairy-focused dual purpose breed, originating from the northern part of Germany. In the past, the old Angler type was crossed frequently with Red Holstein and Holstein Friesian, in order to increase the performance. This resulted in today's admixed Angler type (RSHeG, 2021a). Due to their past introgression and good functional traits they therefore serve as a representative for local breeds. On the other hand, the German Holstein breed represents high-yielding breeds in this project.

The overall goal of this thesis was to develop genomic methods to promote local dairy breeds to increase their competitiveness to high-yielding breeds. In this way, they would contribute to the dairy business by exploiting their good functional characteristics, as well as to the biodiversity of livestock. Thus, the objective of this thesis was to investigate how to implement genomic selection and rotational crossbreeding in German Angler.

Genomic models which are suitable for crossbred data were reviewed in **Chapter 2**, in order to give an overview of the possibilities. Different additive models (such as the parental model, a model with breed-specific allele effects, and a single step model) and dominance models, which were either line-dependent, line-independent or included imprinting, were discussed regarding their accuracies, computational and data requirements.

Chapter 3 introduced an approach of the breed of origin of alleles model for multi-breed genomic prediction of breeding values suitable for numerically small breeds which experienced introgression from popular mainstream breeds in the past, such as German Angler. The model assumes different SNP effects for different origins of haplotype segments. It was compared to multi-breed and within-breed genomic predictions with GBLUP.

In **Chapter 4** a simulation study shows how a genomic rotational crossbreeding scheme including Angler and the high-yielding breed Holstein can be implemented, examining different sizes and structures of reference populations and selection goals of Angler. The reference populations included only Angler individuals, only crossbred individuals, or both Angler and crossbred individuals. The selection of Angler was either based on purebred performance, on crossbred performance, or on a combined purebred-crossbred performance.

In **Chapter 5** a rotational crossbreeding scheme of Angler x Holstein was extended to different Optimum Contribution Selection methods to evaluate the possibilities of realizing genetic gain while regaining the original genetic background of the Angler breed. Different constraints regarding the kinships and migrant contributions from Holstein were applied, to investigate their effects on Angler, crossbred and Holstein populations.

Finally, in **Chapter 6** further genomic models for crossbreeding are discussed shortly, followed by the practical relevance of crossbreeding in dairy cattle.

References

- Clasen JB, Fogh A and Kargo M 2019. Differences between performance of F1 crossbreds and Holsteins at different production levels. *Journal of Dairy Science* 102, 436–441. doi:10.3168/jds.2018-14975.
- Clasen JB, Norberg E, Madsen P, Pedersen J and Kargo M 2017. Estimation of genetic parameters and heterosis for longevity in crossbred Danish dairy cattle. *Journal of Dairy Science* 100, 6337–6342. doi:10.3168/jds.2017-12627.
- Curone G 2016. Milk Ketone Bodies Assessment in A Local Italian Cow Breed (Modenese) vs. Holstein and Characterization of its Physiological, Reproductive and Productive Performances. *International Journal of Health, Animal Science and Food Safety*, Vol 3, No 1s (2016). doi:10.13130/2283-3927/7072.
- Curone G, Filipe J, Cremonesi P, Trevisi E, Amadori M, Pollera C, Castiglioni B, Turin L, Tedde V, Vigo D, Moroni P, Minuti A, Bronzo V, Addis MF and Riva F 2018. What we have lost: Mastitis resistance in Holstein Friesians and in a local cattle breed. *Research in veterinary science* 116, 88–98. doi:10.1016/j.rvsc.2017.11.020.
- Gandini G, Maltecca C, Pizzi F, Bagnato A and Rizzi R 2007. Comparing local and commercial breeds on functional traits and profitability: the case of Reggiana dairy cattle. *Journal of Dairy Science* 90, 2004–2011. doi:10.3168/jds.2006-204.
- Hoffmann I and Scherf B 2005. Management of farm animal genetic diversity: opportunities and challenges. *WAAP Book of the Year*, 221–245.
- Kargo M, Clasen JB, Nielsen HM, Byskov K and Norberg E 2021. Short communication: Heterosis and breed effects for milk production and udder health traits in crosses between Danish Holstein, Danish Red, and Danish Jersey. *Journal of Dairy Science* 104, 678–682. doi:10.3168/jds.2019-17866.
- Kargo M, Madsen P and Norberg E 2012. Short communication: Is crossbreeding only beneficial in herds with low management level? *Journal of Dairy Science* 95, 925–928. doi:10.3168/jds.2011-4707.
- Mendelsohn R 2003. The challenge of conserving indigenous domesticated animals. *Ecological Economics* 45, 501–510. doi:10.1016/S0921-8009(03)00100-9.
- Oldenbroek K 2007. Utilisation and conservation of farm animal genetic resources. Wageningen Academic Publ, Wageningen.

-
- Ovaska U and Soini K 2017. Local Breeds - Rural Heritage or New Market Opportunities? Colliding Views on the Conservation and Sustainable Use of Landraces. *Sociologia Ruralis* 57, 709–729. doi:10.1111/soru.12140.
- Rauw WM, Kanis E, Noordhuizen-Stassen EN and Grommers FJ 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. *Livestock Production Science*, 15–33.
- RShEG 2021. Rassebeschreibung und Zuchtziele Angler. Retrieved on December 16, 2021, from <https://www.rsheg.de/de/zucht-herdbuch/rassebeschreibung-zuchtziele/angler-397.html>.
- Sørensen MK, Norberg E, Pedersen J and Christensen LG 2008. Invited review: crossbreeding in dairy cattle: a Danish perspective. *Journal of Dairy Science* 91, 4116–4128. doi:10.3168/jds.2008-1273.
- Wellmann R and Bennewitz J 2019. Key Genetic Parameters for Population Management. *Frontiers in Genetics* 10, 667. doi:10.3389/fgene.2019.00667.

Chapter 2

A Review of Genomic Models for the Analysis of Livestock Crossbred Data

Joana Stock¹, Jörn Bennewitz¹, Dirk Hinrichs², Robin Wellmann^{*1}

¹ Institute of Animal Science, Department of Animal Breeding and Genetics, University of Hohenheim, 70593 Stuttgart, Germany

² Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany

* Correspondence: Robin Wellmann, r.wellmann@uni-hohenheim.de

Published in *Frontiers in Genetics*, doi:10.3389/fgene.2020.00568

Abstract

Livestock breeding has shifted during the past decade towards genomic selection. For the estimation of breeding values in purebred breeding schemes, genomic best linear unbiased prediction has become the method of choice. Systematic crossbreeding with the aim to utilize heterosis and breed complementary effects is widely used in livestock breeding, especially in pig and poultry breeding. The goal is to improve the performance of the crossbred animals. Due to genotype-by-environment interactions, imperfect linkage disequilibrium, and the existence of dominance and imprinting, purebred and crossbred performances are not perfectly correlated. Hence, more complex genomic models are required for crossbred populations. This study reviews and compares such models. Compared to purebred genomic models, the reviewed models were of much higher complexity due to the inclusion of dominance effects, breed-specific effects, imprinting effects, and the joint evaluation of purebred and crossbred performance data. With the model assessment work conducted until now, it is not possible to come to a clear recommendation as to which existing method is most suitable for a specific breeding program and a specific genetic trait architecture. Since it is expected that a superior method includes all the different genetic effects in a single model, a dominance model with imprinting and breed-specific SNP effects is proposed. Further progress could be made by assuming realistic covariance structures between the genetic effects of the different breeding lines, and by using larger marker panels and mixture distributions for the SNP effects.

Keywords: genomic selection, genomic models, livestock, crossbreeding, heterosis

Introduction

The crossing of different lines or breeds is widely used in animal breeding with the main aim to produce superior offspring. This superiority results from heterosis and from breed complementary effects. Continuous and discontinuous crossbreeding schemes have been designed and are implemented in various livestock species (Samorè and Fontanesi, 2016; Lopez-Villalobos *et al.*, 2000). In discontinuous schemes, crossbred animals are used solely for production and are not selected as parents of the next generation. Breeding takes place in the parental breeds and the breeding goal is usually to improve crossbred performance. The level of organization in such a system is high and it is sometimes difficult to utilize by-products, such as male offspring of mother lines. These schemes can be predominantly found in livestock species with a high female reproduction rate such as pigs and poultry. In continuous breeding schemes, the female crossbreds are used as parents to breed the next generation. These systems are sometimes implemented in livestock species with a low female reproduction rate such as cattle. Since there are substantial non-additive effects for reproduction traits in dairy cattle (Jiang *et al.* 2017), the aims of crossbreeding in dairy cattle are to improve reproduction traits and other functional traits by exploiting heterosis and imprinting and by removing inbreeding depression (Sørensen *et al.*, 2008; Buckley *et al.*, 2014).

A further form of crossbreeding is the upgrading of low-performance breeds with high-yielding breeds. This introgression of genes from high-yielding breeds increases the production level in subsequent generations and reduces inbreeding depression by increasing the genetic diversity of the low-performance breed. This breeding system has frequently been applied to local breeds, such as the German Vorderwald cattle (Hartwig *et al.*, 2014, 2015). However, if upgrading is repeated over several generations, then the breed eventually goes extinct because the native alleles are removed from its gene pool. The formation of a synthetic breed can also be seen as a special form of crossbreeding. A well-known example is the establishment of the so-called Schwarzbuntes Milchrind in the former East Germany (Freyer *et al.*, 2008).

Livestock breeding has shifted towards genomic selection, which is now frequently implemented in large pure breeds. The core of the system that has been implemented in pure breeds is a reference population that consists of genotyped and phenotyped animals. The phenotypes are either the animal's own performance records, or deregressed conventional breeding values. The reference population is needed for the prediction of marker effects. The marker effects are then used for predicting genomic breeding values of the genotyped

selection candidates. The reliability of genomic breeding values depends on the size of the reference population, on the effective number of chromosome segments, and on the method used for the prediction of marker effects (Goddard, 2009).

Extensive research has been dedicated to develop statistical models for the prediction of marker effects. These statistical models include the SNP-BLUP model that assumes normally distributed SNP effects, various Bayesian models that assume more heavy-tailed distributions, as well as non-parametric and semi-parametric models (Meuwissen *et al.*, 2001; Gianola, 2013). More complex models assume different SNP variances, depending on the type of control region the SNP belongs to (MacLeod *et al.*, 2016). Some models avoid the prediction of marker effects by building a genomic relationship matrix based on SNP genotypes. The most prominent method based on genomic relationships is GBLUP, which is an equivalent model to SNP-BLUP (VanRaden, 2008; Goddard, 2009). The genotyped selection candidates are included in the model, and their genomic breeding values are calculated by utilizing their genomic relationships with the reference population. GBLUP assumes that all animals are genotyped, which is in general not the case. Therefore, the genomic breeding values are blended in a second step with pedigree-based breeding values to obtain genomically enhanced breeding values on which selection decisions are based. This two-step procedure can be avoided with so-called single-step GBLUP models (ssGBLUP). They were developed as extensions of GBLUP. Single-step models include genotyped and non-genotyped animals simultaneously (Legarra *et al.*, 2009; Aguilar *et al.*, 2010; Christensen and Lund, 2010; Legarra *et al.*, 2014) and assume a particular covariance structure for the breeding values that is computed from genomic and pedigree-based relationships. Fernando *et al.* (2014) extended the single step model towards non-normally distributed marker effects. In purebred routine application mostly additive effects are considered, with dominance being an integral part of the estimated breeding values. Some genomic models were extended towards accounting for dominance explicitly, but this increased the reliabilities of the breeding values only slightly (Su *et al.* 2012; Wellmann and Bennewitz, 2012; Azevedo *et al.*, 2015).

To summarize, it seems that in practical purebred genomic evaluations, GBLUP and ssGBLUP have and will become the models of choice, and non-additive gene effects are usually not an issue. The picture is however somewhat different if data from crossbred animals in combination with the parental purebred data is analyzed. The potential applications of genomic models with non-additive genetic effects have been reviewed by Varona *et al.* (2018). The main breeding goal is in this case to improve the performance of the crossbred animals. Due to genotype-by-environment interaction, imperfect LD, and the existence of dominance, epistasis and imprinting, purebred and crossbred performances (PP and CP, respectively) are not perfectly correlated (e.g. Wei and van der Werf, 1995; Dekkers, 2007; Zumbach *et al.*, 2007; Duenk *et*

al., 2019). Wientjes and Calus (2017) reviewed existing literature about purebred-crossbred correlations in pigs. The average from 201 reported correlation coefficients was 0.63 with 50% of the reported coefficients being between 0.45 and 0.87. The purebred-crossbred correlation affects the optimal design of the reference population (van Grevenhof and van der Werf, 2015) and the choice of an appropriate genomic model.

While genomic models are well established for pure breeds, much research has been conducted in the recent years to develop genomic models for the analysis of crossbred data. The aim of this study is to review genomic models for the prediction of crossbred performance that were recently developed and were evaluated either using simulated or real crossbred data.

Genomic Models

Genomic models for crossbred data are extensions of purebred models. The extensions were made in several directions. Most genomic models for the analysis of crossbred data are developed for two-way crosses. A two-way cross \mathcal{X} is created from a sire line \mathcal{A} and a dam line \mathcal{B} , which are usually not inbred. The pure lines have breeding values $\mathbf{a}_{\mathcal{A}}$ and $\mathbf{a}_{\mathcal{B}}$ for PP, and breeding values $\mathbf{c}_{\mathcal{A}}$ and $\mathbf{c}_{\mathcal{B}}$ for CP. Typically, some animals are genotyped, whereas others are not. The goal is to obtain accurate predictions of the breeding values for CP by utilizing phenotypic information from genotyped and ungenotyped purebred and crossbred animals. An overview over the considered models is given in Table 1.

Table 1: Additive and dominance models for the prediction of crossbred performance.

		Data requirements		
		Phenotyped crossbreds	Genotyped crossbreds	
Additive Models	Parental Model	x		
	BSAM / ASGM	x	x	
	Single step	x	(x)	
Dominance Models	Line-independent	(x)	(x)	Provide more accurate breeding values for CP than additive models
	Line-dependent	(x)	(x)	
	Dominance + Imprinting	x	x	
	(x) not necessarily needed but can be utilized			

The SNP alleles are usually assumed to be biallelic, so they may be coded as alleles 1 and 2. Most authors use centered allele content matrices as proposed by VanRaden (2008). The centering does not affect the predictions, but affects the model-based reliabilities (Strandén and Christensen, 2011). We denote with

$$\mathbf{Z}_G^{\mathcal{A}} = \mathbf{G}^{\mathcal{A}} - 2\mathbf{P}^{\mathcal{A}}$$

the centered allele content matrix for the genotyped animals from line \mathcal{A} , whereby the allele content $G_{im}^{\mathcal{A}} \in \{0,1,2\}$ of animal i from line \mathcal{A} is the number of copies of allele 2, animal i has at SNP m , and $P_{im}^{\mathcal{A}}$ is the frequency of allele 2 of SNP m in line \mathcal{A} . Moreover, we denote with

$$\mathbf{Z}_X^{\mathcal{A}} = \mathbf{G}_X^{\mathcal{A}} - \mathbf{P}^{\mathcal{A}}$$

the centered allele origin matrix for alleles from cross \mathcal{X} that originate from line \mathcal{A} . That is, $G_{Xim}^{\mathcal{A}} \in \{0,1\}$ is the number of copies of allele 2, crossbred animal i has obtained from sire line \mathcal{A} at SNP m . These matrices are needed to define genetic values of purebred and crossbred animals. The vector with breeding values for CP for animals from line \mathcal{A} has the representation

$$\mathbf{c}_{\mathcal{A}} = \mathbf{Z}_G^{\mathcal{A}} \boldsymbol{\alpha}^{\mathcal{A}}, \quad (1)$$

where $\boldsymbol{\alpha}^{\mathcal{A}}$ is the vector with allele substitution effects for CP. The vector with breeding values for PP has the representation

$$\mathbf{a}_{\mathcal{A}} = \mathbf{Z}_G^{\mathcal{A}} \tilde{\boldsymbol{\alpha}}^{\mathcal{A}}, \quad (2)$$

where $\tilde{\boldsymbol{\alpha}}^{\mathcal{A}}$ is the vector with allele substitution effects for PP. The equations for $\mathbf{a}_{\mathcal{B}}$ and $\mathbf{c}_{\mathcal{B}}$ are similarly.

Most genomic models for two-way crosses utilize, that the vector \mathbf{a}_X with additive genetic values of the crossbred animals can be decomposed into a contribution $\mathbf{c}_X^{\mathcal{A}}$ that comes from sire line \mathcal{A} , and a contribution $\mathbf{c}_X^{\mathcal{B}}$ that comes from dam line \mathcal{B} . That is,

$$\mathbf{a}_X = \mathbf{c}_X^{\mathcal{A}} + \mathbf{c}_X^{\mathcal{B}}, \quad (3)$$

where

$$\mathbf{c}_X^{\mathcal{A}} = \mathbf{Z}_X^{\mathcal{A}} \boldsymbol{\alpha}^{\mathcal{A}}, \text{ and } \mathbf{c}_X^{\mathcal{B}} = \mathbf{Z}_X^{\mathcal{B}} \boldsymbol{\alpha}^{\mathcal{B}}. \quad (4)$$

The contribution $\mathbf{c}_X^{\mathcal{A}}$ from line \mathcal{A} can be further decomposed into a contribution that comes from the breeding values $\mathbf{c}_{\mathcal{A}}$ for CP, and into a vector $\mathbf{m}_X^{\mathcal{A}}$ that contains the Mendelian sampling terms of the transmitted gametes (Wei and van der Werf, 1994). That is,

$$\mathbf{c}_X^{\mathcal{A}} = 0.5 \mathbf{Z}_{X\mathcal{A}} \mathbf{c}_{\mathcal{A}} + \mathbf{m}_X^{\mathcal{A}}, \quad (5)$$

where matrix $\mathbf{Z}_{X\mathcal{A}}$ assigns animals from line \mathcal{A} to their crossbred offspring.

Different models have been developed for predicting CP, which can broadly be classified into additive models and dominance models. While some models predict the breeding values for CP directly with Equation (5), others predict the vector $\alpha^{\mathcal{A}}$ with allele substitution effects for CP. In the latter case, the estimated breeding values $\hat{c}_{\mathcal{A}}$ for CP in line \mathcal{A} are obtained by substituting $\alpha^{\mathcal{A}}$ with the prediction $\hat{\alpha}^{\mathcal{A}}$ in Equation (1).

Additive Models

Different additive models have been proposed in the literature. Some models assume that the crossbred animals are genotyped, whereas others do not. The general additive model for a two-way cross is a trivariate model that has two equations for the parental lines, and one equation for the cross. It has the general representation

$$\mathbf{y}_{\mathcal{A}} = \mathbf{X}_{\mathcal{A}}\mathbf{b}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}} + \mathbf{a}_{\mathcal{A}} + \mathbf{E}_{\mathcal{A}}$$

$$\mathbf{y}_{\mathcal{B}} = \mathbf{X}_{\mathcal{B}}\mathbf{b}_{\mathcal{B}} + \mathbf{Z}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}} + \mathbf{a}_{\mathcal{B}} + \mathbf{E}_{\mathcal{B}}$$

$$\mathbf{y}_{\mathcal{X}} = \mathbf{X}_{\mathcal{X}}\mathbf{b}_{\mathcal{X}} + \mathbf{Z}_{\mathcal{X}}\mathbf{u}_{\mathcal{X}} + \dots + \mathbf{E}_{\mathcal{X}},$$

where $\mathbf{y}_{\mathcal{A}}, \mathbf{y}_{\mathcal{B}}, \mathbf{y}_{\mathcal{X}}$ are vectors with phenotypic records of the respective subpopulation, $\mathbf{b}_{\mathcal{A}}, \mathbf{b}_{\mathcal{B}}, \mathbf{b}_{\mathcal{X}}$ are vectors of fixed effects with design matrices $\mathbf{X}_{\mathcal{A}}, \mathbf{X}_{\mathcal{B}}, \mathbf{X}_{\mathcal{X}}$, and $\mathbf{u}_{\mathcal{A}}, \mathbf{u}_{\mathcal{B}}, \mathbf{u}_{\mathcal{X}}$ are vectors of non-genetic random effects with design matrices $\mathbf{Z}_{\mathcal{A}}, \mathbf{Z}_{\mathcal{B}}, \mathbf{Z}_{\mathcal{X}}$. Finally, $\mathbf{a}_{\mathcal{A}}, \mathbf{a}_{\mathcal{B}}$ are the breeding values for PP, and $\mathbf{E}_{\mathcal{A}}, \mathbf{E}_{\mathcal{B}}, \mathbf{E}_{\mathcal{X}}$ are the residual terms. The term “...” in the third equation depends on the respective model.

The first two model equations are needed because PP and CP are genetically correlated (Wientjes and Calus, 2017), so phenotypic records of purebred animals increase the reliabilities of the breeding values for CP.

The Parental Additive Model

The parental additive model is based on Equations (2), (3) and (5), and is suitable when the crossbred animals are not genotyped. The model assumes that the Mendelian sampling terms are part of the residuals, so the model equations become

$$\mathbf{y}_{\mathcal{A}} = \mathbf{X}_{\mathcal{A}}\mathbf{b}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}} + \mathbf{a}_{\mathcal{A}} + \mathbf{E}_{\mathcal{A}}$$

$$\mathbf{y}_{\mathcal{B}} = \mathbf{X}_{\mathcal{B}}\mathbf{b}_{\mathcal{B}} + \mathbf{Z}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}} + \mathbf{a}_{\mathcal{B}} + \mathbf{E}_{\mathcal{B}}$$

$$\mathbf{y}_{\mathcal{X}} = \mathbf{X}_{\mathcal{X}}\mathbf{b}_{\mathcal{X}} + \mathbf{Z}_{\mathcal{X}}\mathbf{u}_{\mathcal{X}} + 0.5 \mathbf{Z}_{\mathcal{X}\mathcal{A}}\mathbf{c}_{\mathcal{A}} + 0.5 \mathbf{Z}_{\mathcal{X}\mathcal{B}}\mathbf{c}_{\mathcal{B}} + \mathbf{E}_{\mathcal{X}},$$

where $\mathbf{a}_{\mathcal{A}} = \mathbf{Z}_{\mathcal{G}}^{\mathcal{A}}\tilde{\boldsymbol{\alpha}}^{\mathcal{A}}$, $\mathbf{a}_{\mathcal{B}} = \mathbf{Z}_{\mathcal{G}}^{\mathcal{B}}\tilde{\boldsymbol{\alpha}}^{\mathcal{B}}$, $\mathbf{c}_{\mathcal{A}} = \mathbf{Z}_{\mathcal{G}}^{\mathcal{A}}\boldsymbol{\alpha}^{\mathcal{A}}$, and $\mathbf{c}_{\mathcal{B}} = \mathbf{Z}_{\mathcal{G}}^{\mathcal{B}}\boldsymbol{\alpha}^{\mathcal{B}}$.

The BSAM and ASGM model

The model with breed-specific allele effects (BSAM) and the model with breed-independent allele effects (ASGM) are based on Equations (2), (3) and (4), and require that the crossbred animals are genotyped. While the ASGM model predicts one effect per SNP, the BSAM model predicts one effect for the paternal allele, and one for the maternal allele of the crossbred animals. Origin-specific allele effects may occur e.g. due to a different LD pattern between the marker and the QTL, different gene frequencies at the QTL, imprinting effects, or the epistatic effects may be different in the pure breeds. This results in different effects of the marker alleles and thus affects the estimated breeding values.

The first two equations of the BSAM and ASGM model are as above, whereas the third model equation becomes for the BSAM

$$\mathbf{y}_x = \mathbf{X}_x \mathbf{b}_x + \mathbf{Z}_x \mathbf{u}_x + \mathbf{Z}_x^A \boldsymbol{\alpha}^A + \mathbf{Z}_x^B \boldsymbol{\alpha}^B + \mathbf{E}_x.$$

An equivalent representation for the ASGM model is

$$\mathbf{y}_x = \mathbf{X}_x \mathbf{b}_x + \mathbf{Z}_x \mathbf{u}_x + \mathbf{a}_x + \mathbf{E}_x.$$

Ibáñez-Escriche *et al.* (2009) predicted CP of the parental lines from genotyped crossbred animals with BSAM and ASGM, whereby the breed-specific allele substitution effects of the BSAM model were a priori independent. The allele substitution effects were estimated with BayesB, which is a method that assumes that many of them are actually zero. An oligogene trait was simulated with breed-independent QTL effects. Although the SNP effects are expected to be breed-specific due to differences in LD between markers and QTL, the authors found that the BSAM model outperformed ASGM only if the number of markers was low, the number of records for training was high, and if the parental breeds were distantly related.

Lopes *et al.* (2017) used the BSAM model with normally distributed SNP effects to predict breeding values for CP from crossbred data, and compared the results with conventional GBLUP. The model provided similar prediction accuracies as conventional GBLUP for the traits litter size and gestation length in pigs. It may be not superior to GBLUP because the allele substitution effects of the different breeds were implicitly assumed to be uncorrelated, which is an assumption that is not likely to be fulfilled.

Sevillano *et al.* (2019) extended the BSAM and ASGM model towards a three-way cross and distinguished SNP that showed a strong trait association from all remaining SNP. For the trait associated SNP breed-specific effects were estimated, whereas for the remaining SNP one effect was estimated, regardless of the allele origin. This model was compared with the BSAM model and with the ASGM model for the trait daily gain by assuming normally distributed SNP

effects. Purebred as well as crossbred data was used for training. The results showed a superiority of their method only if the estimated genetic correlations between PP and CP for the trait associated SNPs and the remaining SNPs were unequal.

Vandenplas *et al.* (2017) derived equations for predicting the reliability of genomic breeding values for CP for BSAM and ASGM models and assumed normally distributed SNP effects. The authors found that BSAM outperformed ASGM for a specific parental line, if the effective number of chromosome segments in the crossbred reference animals that originate from the parental line is less than half the effective number of all chromosome segments that are independently segregating.

Additive Single Step Model

While BSAM has the disadvantage that all crossbred animals have to be genotyped, the parental additive model has the disadvantage, that the information provided by the Mendelian sampling terms cannot be utilized for prediction. These problems could be resolved by using a trivariate model of the form

$$\mathbf{y}_{\mathcal{A}} = \mathbf{X}_{\mathcal{A}}\mathbf{b}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}} + \mathbf{a}_{\mathcal{A}} + \mathbf{E}_{\mathcal{A}}$$

$$\mathbf{y}_{\mathcal{B}} = \mathbf{X}_{\mathcal{B}}\mathbf{b}_{\mathcal{B}} + \mathbf{Z}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}} + \mathbf{a}_{\mathcal{B}} + \mathbf{E}_{\mathcal{B}}$$

$$\mathbf{y}_{\mathcal{X}} = \mathbf{X}_{\mathcal{X}}\mathbf{b}_{\mathcal{X}} + \mathbf{Z}_{\mathcal{X}}\mathbf{u}_{\mathcal{X}} + \mathbf{c}_{\mathcal{X}}^{\mathcal{A}} + \mathbf{c}_{\mathcal{X}}^{\mathcal{B}} + \mathbf{E}_{\mathcal{X}}$$

that includes both, genotyped and phenotyped animals. Christensen *et al.* (2014) derived the joint covariance matrix $\mathbf{A}_{\mathcal{A}}$ of $\mathbf{c}_{\mathcal{X}}^{\mathcal{A}}$, $\mathbf{c}_{\mathcal{A}}$, and $\mathbf{a}_{\mathcal{A}}$ by using the pedigree-based model of Wei and van der Werf (1994) as a starting point. The authors derived the covariance matrix $\mathbf{A}_{\mathcal{A}}$ from pedigree relationships, and replaced it in a subsequent step by a covariance matrix $\mathbf{H}_{\mathcal{A}}$ that combines pedigree and genotype information.

Xiang *et al.* (2016a) validated the model of Christensen *et al.* (2014) in a two-way pig cross for the trait number of piglets born. The authors found that the inclusion of crossbred genomic information improved the model-based reliabilities for CP and reduced to some extent the bias of prediction.

Tusell *et al.* (2016) used a single-step model for two-way crossbred pigs and the sire line \mathcal{A} , so the model reduced to a bivariate model. The purebred animals were partly genotyped. Since the crossbred animals were not genotyped, the third equation in the model of Christensen *et al.* (2014) was replaced by a parental additive model equation, i.e. the Mendelian sampling terms were part of the residual. This resulted in a model equation of the form

$$\mathbf{y}_{\mathcal{A}} = \mathbf{X}_{\mathcal{A}}\mathbf{b}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}} + \mathbf{a}_{\mathcal{A}} + \mathbf{E}_{\mathcal{A}}$$

$$\mathbf{y}_{\mathcal{X}} = \mathbf{X}_{\mathcal{X}}\mathbf{b}_{\mathcal{X}} + \mathbf{Z}_{\mathcal{X}}\mathbf{u}_{\mathcal{X}} + 0.5 \mathbf{Z}_{\mathcal{X}\mathcal{A}}\mathbf{c}_{\mathcal{A}} + 0.5 \mathbf{Z}_{\mathcal{X}\mathcal{B}}\mathbf{c}_{\mathcal{B}} + \mathbf{E}_{\mathcal{X}}.$$

The authors evaluated six growth and meat traits and found that the genetic correlations between purebred and CP were larger than 0.69 for all traits. The accuracies of the genomic breeding values were higher than those obtained from univariate single-step models that took either purebred or CP into account, and also higher than those obtained with pedigree-based models.

Dominance models

Crossbreeding utilizes heterosis and breed complementarity. A widely accepted hypothesis is that heterosis arises predominantly from dominance effects. An animal carries a dominance effect only if it is heterozygous at a particular QTL. We denote with

$$\mathbf{Z}_H^{\mathcal{X}} = \mathbf{H}^{\mathcal{X}} - \bar{\mathbf{H}}^{\mathcal{X}}$$

the centered indicator matrix for heterozygosity. That is, $H_{im}^{\mathcal{X}} \in \{0,1\}$ equals one, if animal i is heterozygous at SNP m , and $\bar{H}_{im}^{\mathcal{X}}$ is the heterozygosity of SNP m in line \mathcal{X} . The dominance model assumes that the vector $\mathbf{g}_{\mathcal{X}}$ with genotypic values of the crossbred animals has the representation

$$\mathbf{g}_{\mathcal{X}} = \mu_{\mathcal{X}}\mathbf{1} + \mathbf{Z}_G^{\mathcal{X}}\mathbf{a}^{\mathcal{X}} + \mathbf{Z}_H^{\mathcal{X}}\mathbf{d}^{\mathcal{X}}, \quad (6)$$

where $\mu_{\mathcal{X}}$ is the population mean, $\mathbf{a}^{\mathcal{X}}$ is the vector with population-dependent additive effects, and $\mathbf{d}^{\mathcal{X}}$ is vector with population-dependent dominance effects. The genotypic values of purebred animals are defined accordingly. The trivariate dominance model for a two-way cross and the parental lines has therefore the representation

$$\mathbf{y}_{\mathcal{A}} = \mathbf{X}_{\mathcal{A}}\mathbf{b}_{\mathcal{A}} + \mathbf{Z}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}} + \mathbf{Z}_G^{\mathcal{A}}\mathbf{a}^{\mathcal{A}} + \mathbf{Z}_H^{\mathcal{A}}\mathbf{d}^{\mathcal{A}} + \mathbf{E}_{\mathcal{A}} \quad (7)$$

$$\mathbf{y}_{\mathcal{B}} = \mathbf{X}_{\mathcal{B}}\mathbf{b}_{\mathcal{B}} + \mathbf{Z}_{\mathcal{B}}\mathbf{u}_{\mathcal{B}} + \mathbf{Z}_G^{\mathcal{B}}\mathbf{a}^{\mathcal{B}} + \mathbf{Z}_H^{\mathcal{B}}\mathbf{d}^{\mathcal{B}} + \mathbf{E}_{\mathcal{B}}$$

$$\mathbf{y}_{\mathcal{X}} = \mathbf{X}_{\mathcal{X}}\mathbf{b}_{\mathcal{X}} + \mathbf{Z}_{\mathcal{X}}\mathbf{u}_{\mathcal{X}} + \mathbf{Z}_G^{\mathcal{X}}\mathbf{a}^{\mathcal{X}} + \mathbf{Z}_H^{\mathcal{X}}\mathbf{d}^{\mathcal{X}} + \mathbf{E}_{\mathcal{X}},$$

which we call the dominance model with line-dependent effects. The vector $\mathbf{c}_{\mathcal{A}}$ with breeding values for CP from breed \mathcal{A} has the representation given in Equation (1), but the vector with allele substitution effects for CP is

$$\mathbf{a}^{\mathcal{A}} = \mathbf{a}^{\mathcal{X}} + (1 - 2p^{\mathcal{B}}) \circ \mathbf{d}^{\mathcal{X}},$$

where \mathbf{p}^B is vector with allele frequencies in the opposite line, and the Hadamard product " \circ " is the component-wise product. The breeding values and allele substitution effects for line B are defined accordingly. Predictions $\hat{\mathbf{a}}^X$ and $\hat{\mathbf{d}}^X$ of \mathbf{a}^X and \mathbf{d}^X are needed to get predictions of the allele substitution effects for CP in line A with equation

$$\hat{\mathbf{a}}^A = \hat{\mathbf{a}}^X + (1 - 2\mathbf{p}^B) \circ \hat{\mathbf{d}}^X.$$

Some solvers are unable to account for the fact that $E(\mathbf{d}^X) = \mu_d^X \mathbf{1} \neq 0$ for most traits. As shown by Xiang *et al.* (2016b), one may write $\mathbf{d}^X = \mathbf{d}_*^X + \mu_d^X \mathbf{1}$ such that $E(\mathbf{d}_*^X) = 0$. Then, the term $\mathbf{Z}_H^X \mathbf{d}^X$ in Equation (7) equals $\mathbf{Z}_H^X \mathbf{1} \mu_d^X + \mathbf{Z}_H^X \mathbf{d}_*^X$, where μ_d^X is treated as an additional fixed effect. The same needs to be done for the parental lines. We can write $\mathbf{Z}_H^X \mathbf{1} \mu_d^X = \mu_d^X M(\hat{\mathbf{h}}_X - \bar{h}_X \mathbf{1})$, where M is the number of SNPs, $\hat{\mathbf{h}}_X$ is the vector with heterozygosities of the crossbred animals, and \bar{h}_X is the average heterozygosity of the crossbred animals. Hence, the value $-\mu_d^X M$ quantifies the inbreeding depression per unit of genomic inbreeding.

Vitezica *et al.* (2016) demonstrated how dominance models with normally distributed SNP effects can be transformed into equivalent dominance models with animal effects, whereby different covariance matrices are needed for the additive component and the dominance component of the animal effects. That is, if all SNP effects are normally distributed, then the SNP effects model can be replaced by the equivalent animal effects model

$$\mathbf{y}_A = \mathbf{X}_A \mathbf{b}_A + \mathbf{Z}_A \mathbf{u}_A + \mathbf{Z}_H^A \mathbf{1} \mu_d^A + \tilde{\mathbf{a}}_A + \tilde{\mathbf{d}}_A^* + \mathbf{E}_A$$

$$\mathbf{y}_B = \mathbf{X}_B \mathbf{b}_B + \mathbf{Z}_B \mathbf{u}_B + \mathbf{Z}_H^B \mathbf{1} \mu_d^B + \tilde{\mathbf{a}}_B + \tilde{\mathbf{d}}_B^* + \mathbf{E}_B$$

$$\mathbf{y}_X = \mathbf{X}_X \mathbf{b}_X + \mathbf{Z}_X \mathbf{u}_X + \mathbf{Z}_H^X \mathbf{1} \mu_d^X + \tilde{\mathbf{a}}_X + \tilde{\mathbf{d}}_X^* + \mathbf{E}_X$$

from which the SNP effects can be back-solved. Thereby, the animal effects satisfy $\tilde{\mathbf{a}}_X = \mathbf{Z}_G^X \mathbf{a}^X$, and $\tilde{\mathbf{d}}_X^* = \mathbf{Z}_H^X \mathbf{d}_*^X$, and so on. The joint covariance matrices of the animal effects are given in Christensen *et al.* (2019).

The SNP effects in Equation (7) were assumed to be line-dependent, which may be the case because the LD between SNP and QTL differs between lines. This may be neglected if the marker panel is sufficiently large. In this case, the SNP effects can be assumed to be line-independent, and we obtain the simplified model

$$\mathbf{y}_A = \mathbf{X}_A \mathbf{b}_A + \mathbf{Z}_A \mathbf{u}_A + \mathbf{Z}_G^A \mathbf{a} + \mathbf{Z}_H^A \mathbf{d} + \mathbf{E}_A$$

$$\mathbf{y}_B = \mathbf{X}_B \mathbf{b}_B + \mathbf{Z}_B \mathbf{u}_B + \mathbf{Z}_G^B \mathbf{a} + \mathbf{Z}_H^B \mathbf{d} + \mathbf{E}_B$$

$$y_x = X_x b_x + Z_x u_x + Z_G^x a + Z_H^x d + E_x,$$

which we call the dominance model with line-independent effects.

Vitezica *et al.* (2013) emphasized that two different parameterizations of the dominance model exist. The first parameterization, which is given by Equation (6), is suitable for two-way crosses, and includes the additive and dominant SNP effects. In contrast, the second parameterization includes the allele substitution effects and the dominance deviations of the SNP. Both parameterizations are equivalent, but their interpretation is different.

Model evaluation

Zeng *et al.* (2013) compared a Bayesian dominance model with the corresponding BSAM model and the corresponding ASGM model. A BayesC π type method was used to estimate the marker effects, so the prior assumption was that the SNP effects are either zero, or come from a normal distribution. The comparison was done for a simulated two-way crossbreeding program. A number of 20 generations of selection was simulated with the aim to improve CP in both parental lines. The marker effects were estimated only once in generation one from crossbred animals and used in all subsequent generations. The simulated traits showed a different degree of dominance variance, ranging from 'large' to 'realistic', or null. The dominance model was superior to the BSAM model and to the ASGM model. This superiority depended on the fraction of dominance and thus heterosis in the data, but even for situations where no dominance was simulated, the accuracy of the dominance model was similar to the additive model, indicating the robustness of the model. It can tentatively be concluded, that the use of a dominance model is in general advisable, even if dominance is not an important source of trait variability.

Xiang *et al.* (2016b) used a dominance model with line-dependent effects for a two-way cross and the parental breeds. The SNP effects were normally distributed, and the additive and dominance effects of the three different populations were correlated. The authors found that the increased predictive ability of the dominance model arose solely from capturing inbreeding depression. This suggests that dominance effects of individual QTL have not been captured. The reason may be that a 60K SNP panel is not sufficient for achieving high LD between markers and QTL, and that the normality assumption is unlikely to be fulfilled.

Esfandyari *et al.* (2016) compared a Bayesian dominance model with the corresponding Bayesian ASGM model at the example of litter size in a two-way pig cross, whereby BayesC of Habier *et al.* (2011) was used for prediction. Training was on the parental lines. The prediction accuracies for PP and CP obtained with the dominance model were both higher than those for PP obtained with the ASGM model.

Implications for breeding programs

All additive models for predicting CP rely on phenotypic data collected from crossbred animals. This can be problematic in situations where the crossbred animals are not individually identified and thus such data collection pipeline is not implemented. This is likely the case on many farms housing crossbred animals. While additive models require phenotypes from crossbred animals, this is not the case for dominance models because the breeding values for CP can be derived from additive and dominance effects that are predicted in the pure breed, and from the allele frequencies in the opposite breed. Esfandiyari *et al.* (2015a) proposed therefore to use dominance models for selecting purebred animals for CP based on purebred phenotypic and genotypic information only. They did a simulation study and estimated the marker effects with Bayesian LASSO (Park and Casella, 2008; los Campos *et al.*, 2009). The results showed that the gain in CP was higher when the purebreds were selected for CPs, which demonstrated the feasibility of the method even when no crossbred data is available. Moreover, combining several related lines into a single reference population increased the prediction accuracy. However, as shown by Esfandiyari *et al.* (2015b), training on crossbred animals leads to a higher selection response than training on purebred animals. A likely explanation is, that the level of heterozygosity was higher than in the purebred data.

Although genomic selection for CP is a promising strategy to increase selection response for CP in the short and medium term, Esfandiyari *et al.* (2018) found that genomic selection for CP leads eventually to lower CP in the long term than genomic selection on PP. This hold regardless of whether training was on purebred or crossbred animals.

Dominance model with imprinting

Dominance effects, as well as additive effects may depend on the breed of origin, which may be due to imprinting or breed complementarity. It could therefore be advantageous to account for imprinting explicitly. A dominance model with imprinting needs to distinguish between the paternal and the maternal allele. If an animal has received allele A_1 from line \mathcal{A} and allele A_2 from line \mathcal{B} , then we denote its genotype as A_1A_2 . The centered indicator matrix for genotype A_1A_2 is given by

$$\mathbf{W}_{\mathcal{X}}^{A_1A_2} = \mathbf{H}_{\mathcal{X}}^{A_1A_2} - \bar{\mathbf{H}}_{\mathcal{X}}^{A_1A_2},$$

where $H_{\mathcal{X}im}^{A_1A_2} \in \{0,1\}$ equals one, if animal i from cross \mathcal{X} has genotype A_1A_2 at SNP m , and $\bar{H}_{\mathcal{X}im}^{A_1A_2}$ is the proportion of animals from cross \mathcal{X} that have this genotype at SNP m .

The dominance model with imprinting assumes that the vector $\mathbf{g}_{\mathcal{X}}$ with genotypic values of the crossbred animals has the representation

$$\mathbf{g}_x = \mu_x \mathbf{1} + (\mathbf{W}_x^{21} + \mathbf{W}_x^{22})\mathbf{a}_{\mathcal{A}}^x + (\mathbf{W}_x^{12} + \mathbf{W}_x^{22})\mathbf{a}_{\mathcal{B}}^x + \mathbf{W}_x^{21}\mathbf{d}_{\mathcal{A}}^x + \mathbf{W}_x^{12}\mathbf{d}_{\mathcal{B}}^x, \quad (8)$$

where μ_x is the population mean, vectors $\mathbf{a}_{\mathcal{A}}^x$ and $\mathbf{a}_{\mathcal{B}}^x$ contain breed-of-origin dependent additive effects, and vectors $\mathbf{d}_{\mathcal{A}}^x$ and $\mathbf{d}_{\mathcal{B}}^x$ contain breed-of-origin dependent dominance effects. The model equation for the crossbred animals becomes

$$\mathbf{y}_x = \mathbf{X}_x \mathbf{b}_x + \mathbf{Z}_x \mathbf{u}_x + (\mathbf{W}_x^{21} + \mathbf{W}_x^{22})\mathbf{a}_{\mathcal{A}}^x + (\mathbf{W}_x^{12} + \mathbf{W}_x^{22})\mathbf{a}_{\mathcal{B}}^x + \mathbf{W}_x^{21}\mathbf{d}_{\mathcal{A}}^x + \mathbf{W}_x^{12}\mathbf{d}_{\mathcal{B}}^x + \mathbf{E}_x. \quad (9)$$

If imprinting in the parental lines is neglected, then the model equations for the parental lines remain as in Equation (7). The vector with allele substitution effects for CP of line \mathcal{A} is in this case

$$\boldsymbol{\alpha}^{\mathcal{A}} = \mathbf{a}_{\mathcal{A}}^x + (1 - \mathbf{p}^{\mathcal{B}}) \circ \mathbf{d}_{\mathcal{A}}^x - \mathbf{p}^{\mathcal{B}} \circ \mathbf{d}_{\mathcal{B}}^x, \quad (10)$$

where $\mathbf{p}^{\mathcal{B}}$ is the vector with allele frequencies in the opposite line. The proof is given in the electronic appendix. When the SNP effects in the cross do not depend on the breed of origin, then the model simplifies, and becomes identical to the dominance model with line-dependent effects.

Nishio and Satoh (2015) proposed two alternative parameterizations for models with dominance and imprinting and fitted them by assuming normally distributed SNP effects. Their first model includes an additive effect, a dominance effect, and an imprinting effect for the heterozygous genotype, while their second model includes a paternal and a maternal gametic effect, and a dominance effect. The models provided in a simulation study more accurate estimates of genotypic values than GBLUP. While the models of Nishio and Satoh (2015) have the advantage that only 3 effects are needed in the equivalent SNP model for modelling the contribution of each SNP to the genotypic value of an animal, the model in Equation (9) has the advantage that more rigorous prior assumptions can be made for the joint distribution of the effects. That is, if the paternal lines are closely related, then the additive effects $\mathbf{a}_{\mathcal{A}}^x$ and $\mathbf{a}_{\mathcal{B}}^x$ could assumed to be *a priori* highly correlated, as well as the dominance effects $\mathbf{d}_{\mathcal{A}}^x$ and $\mathbf{d}_{\mathcal{B}}^x$. However, the parameterization does not allow to predict the vectors $\mathbf{a}_{\mathcal{A}}^x, \mathbf{a}_{\mathcal{B}}^x, \mathbf{d}_{\mathcal{A}}^x$ and $\mathbf{d}_{\mathcal{B}}^x$ individually.

Esfandyari *et al.* (2015b) compared in a simulation study a Bayesian dominance model with imprinting with the corresponding dominance model with line-independent effects, but used a different parameterization. The model considered imprinting because it included a separate effect for each phased genotype. Compared to the model proposed above, it has the disadvantage that the effects have no direct interpretation as additive and dominance effects. The genetic effects of the parental breeds were *a priori* independent. Even though the authors

did not simulate imprinting, they found that the dominance model with imprinting was superior, if the reference population was sufficiently large, and if both lines were not closely related. The reason may be that the LD between markers and QTL was different in the cross and in the parental lines, so the additive effects and dominance effects were population-dependent.

Discussion

In this paper, genomic models for the analysis of discontinuous crossbred data were reviewed. Compared to purebred genomic models, the reviewed models were of much higher complexity due to the inclusion of dominance effects, breed-specific effects, imprinting effects, and the use of PP and / or CP data. In the following some additional aspects regarding the distribution of the SNP effects and the model choice are considered.

Distribution of SNP effects

The normal distribution is the most common assumption about the distribution of SNP effects. Such models have the advantage, that they have an equivalent representation as animal models with genomic covariance matrices for which fast solvers exist, such as DMU (Madsen *et al.*, 2010), WOMBAT (Meyer, 2007), ASReml (Gilmour *et al.*, 2009), blupf90 (Miszta, 1999), or MiX99 (Vuori *et al.*, 2006). Although the assumption of a normal distribution is not likely to be fulfilled when large marker panels are used, the experience with purebred data suggest that the reliabilities of the breeding values are only slightly worse than those obtained with non-normally distributed marker effects. However, the situation in crossbreeding is different because the parental lines are commonly distantly related, and it may be envisaged to evaluate all lines simultaneously in order to increase the reliabilities of the breeding values. This requires that all QTL are in high LD with at least one marker, which implies the necessity to use a large marker panel. However, if the marker panel is large, then only few markers are needed to capture the effect of any QTL. Consequently, the true effects of most markers are actually zero. The model for genomic selection should account for this and assume as a prior distribution for the SNP effects a mixture of two distributions. One component provides the distribution for markers that are in strong LD with a QTL, and the other one is actually zero. In this case, a random-variable γ_m is commonly introduced, which indicates whether the effects of an SNP m are different from zero. Well-known examples are BayesB (Meuwissen *et al.*, 2001), BayesC (Habier *et al.*, 2011), and BayesR (Erbe *et al.*, 2012). Such algorithms are usually implemented with MCMC algorithms, which results in long computation times. However, alternative and faster implementations are available for some models (e.g. Meuwissen, 2009; Shepherd *et al.*, 2010).

For models with additive and dominance effects, an important aspect is, whether these effects are a priori independent or not. It may be advantageous to assume that all effects of a particular

SNP m are of the same order of magnitude. This is possible if all effects of a particular SNP m have conditionally on the common covariance matrix $\gamma_m \sigma_m^2 \Sigma$ a normal distribution, where $\sigma_m^2 \sim \text{Inv-}\chi^2(v, s)$ and Σ is an appropriately chosen covariance matrix. For the dominance model with line-dependent effects, this means that

$$(a_m^A, a_m^B, a_m^X, d_m^A, d_m^B, d_m^X) | \sigma_m^2, \gamma_m \sim N(0, \gamma_m \sigma_m^2 \Sigma).$$

It can be shown that in this case, all effects of SNP m would have for $\gamma_m = 1$ a t -distribution with v degrees of freedom, and are for $\gamma_m = 0$ equal to zero. Moreover, the magnitude of the effect size would be similar for all effects of a given SNP m , which reduces the proportion of overdominant SNP. Developing a fast algorithm for such a model is an area for future research.

Model choice

The most suitable model for a breeding program depends on the achievable accuracies for the breeding values of the selection candidates, and on the available data. Among the additive models, the parental model provided the least accurate predictions for CP, which is because the Mendelian sampling terms are part of the residual and can therefore not be utilized for prediction. It has, however, the advantage that the crossbred animals do not need to be genotyped and may therefore be suitable for animals with low economic value.

The BSAM and ASGM models provided similar results in most cases. The BSAM model, however, needs the trace of the alleles from the purebred parent breed to the crossbred end product, which is a source of potential errors. This might even be more a problem when more complex crossbred structures are involved, e.g. three- or four-way crossbred data. Vandenplas *et al.* (2016) and Sevillano *et al.* (2016) developed a statistical pipeline for this purpose and applied it to a three-way crossbred pig data set.

The reviewed papers suggest that the dominance models provide more accurate genomic breeding values for CP than the additive models. Although Xiang *et al.* (2016b) showed that this gain in accuracy results in the case of normally distributed SNP effects almost solely from capturing inbreeding depression, this may be not the case when large marker panels and appropriate Bayesian models are used for evaluation. Dominance models have the additional advantage that breeding values for crossbred performance can be obtained from purebred animals, so phenotyping and genotyping crossbred individuals may not be necessary. However, as shown by Esfandyari *et al.* (2015b), the accuracy of the breeding values can be increased when phenotyped and genotyped crossbred individuals are included in the reference population.

Three different dominance models have been applied to crossbred data, which are the dominance model with line-independent effects, the dominance model with line-dependent effects, and the dominance model with imprinting. The dominance model with line-dependent effects is likely to be inferior to the model with line-independent effects if the SNP effects of the different lines are falsely assumed to be statistically independent, the reference population is small, and the lines are closely related. This could be avoided by specifying a covariance between the SNP effects of the different lines.

When imprinting is relevant, then a dominance model with imprinting is of interest. For example, Jiang *et al.* (2017) found that there is substantial imprinting for reproduction traits in dairy cattle. The application of imprinting models requires that the crossbred animals are genotyped and that the alleles are traced from the parental lines to the crossbred animals. Unfortunately, to the best of our knowledge, these models are not well analyzed yet. More research should be done in this area, which includes to analyze all models with common data sets.

Conclusion

Genomic models for crossbred data are of much higher complexity than models for purebred data, which results from the inclusion of dominance effects, breed-specific effects, imprinting effects, and from the joint evaluation of PP and CP. Although much research has already been done to develop genomic models for crossbred data, it can be expected that further progress can be made by developing statistical models that include all the different genetic effects in a single model, assume realistic covariance structures between the genetic effects of different breeding lines, use large marker panels, and assume realistic distributions for the SNP effects. The comparisons made in the reviewed papers are not sufficiently comprehensive to come to a clear recommendation as to which existing method is most suitable for a specific breeding program and a specific genetic trait architecture. Some papers suggested a superiority of dominance models. In the reviewed papers, the focus was on discontinuous crossbreeding schemes. This was because, to our best knowledge, no genomic models have been published that are specifically designed for continuous crossbreeding schemes.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

RW did the formal model comparison, JS, RW, and JB wrote the paper, DH contributed to the writing.

Funding

This work was financially supported by the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2817ERA10D. The project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 696231 - ReDiverse (Biodiversity within and between European Red Dairy Breeds). Joana Stock was partly supported by the H. Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged.

Acknowledgments

The manuscript has benefitted from the critical and helpful comments of two reviewers.

Supplementary Material

The Supplementary Material for this article can also be found online at:

<https://www.frontiersin.org/articles/10.3389/fgene.2020.00568/full#supplementary-material>

References

- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S and Lawlor TJ 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93, 743–752. doi:10.3168/jds.2009-2730.
- Azevedo CF, de Resende MDV, e Silva FF, Viana JMS, Valente MSF, Resende MFR, and Muñoz P 2015. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genetics* 16, 105. doi: 10.1186/s12863-015-0264-2
- Buckley F, Lopez-Villalobos N and Heins BJ 2014. Crossbreeding: implications for dairy cow fertility and survival. *Animal* 8 Suppl 1, 122–133. doi:10.1017/S1751731114000901.
- Christensen OF and Lund MS 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42, 2. doi:10.1186/1297-9686-42-2.
- Christensen OF, Madsen P, Nielsen B and Su G 2014. Genomic evaluation of both purebred and crossbred performances. *Genetics Selection Evolution* 46, 23. doi:10.1186/1297-9686-46-23.
- Christensen OF, Nielsen B, Su G, Xiang T, Madsen P, Ostersen T, Velander I and Strathe AB 2019. A bivariate genomic model with additive, dominance and inbreeding depression effects for sire line and three-way crossbred pigs. *Genetics Selection Evolution* 51, 45. doi:10.1186/s12711-019-0486-2.
- Dekkers JCM 2007. Marker-assisted selection for commercial crossbred performance. *Journal of Animal Science* 85, 2104–2114. doi:10.2527/jas.2006-683.
- Duenk P, Calus MPL, Wientjes YCJ, Breen VP, Henshall JM, Hawken R and Bijma P 2019. Estimating the purebred-crossbred genetic correlation of body weight in broiler chickens with pedigree or genomic relationships. *Genetics Selection Evolution* 51, 6. doi:10.1186/s12711-019-0447-9.
- Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, Mason BA and Goddard ME 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* 95, 4114–4129. doi:10.3168/jds.2011-5019.
- Esfandyari H, Berg P and Sørensen AC 2018. Balanced selection on purebred and crossbred performance increases gain in crossbreds. *Genetics Selection Evolution* 50, 8. doi:10.1186/s12711-018-0379-9.

-
- Esfandyari H, Bijma P, Henryon M, Christensen OF and Sørensen AC 2016. Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genetics Selection Evolution* 48, 40. doi:10.1186/s12711-016-0220-2.
- Esfandyari H, Sørensen AC and Bijma P 2015a. Maximizing crossbred performance through purebred genomic selection. *Genetics Selection Evolution* 47, 16. doi:10.1186/s12711-015-0099-3.
- Esfandyari H, Sørensen AC and Bijma P 2015b. A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genetics Selection Evolution* 47, 76. doi:10.1186/s12711-015-0155-z.
- Fernando RL, Dekkers JC, Garrick DJ 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. *Genetics Selection Evolution*. doi: 10.1186/1297-9686-46-50.
- Freyer G, König S, Fischer B, Bergfeld U and Cassell BG 2008. Invited review: crossbreeding in dairy cattle from a German perspective of the past and today. *Journal of Dairy Science* 91, 3725–3743. doi:10.3168/jds.2008-1287.
- Gianola D 2013. Priors in whole-genome regression: the bayesian alphabet returns. *Genetics* 194, 573–596. doi:10.1534/genetics.113.151753.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R. 2009. AS-Reml User Guide. Release 3.0. VSN International, Hemel Hempstead, UK.
- Goddard, ME 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257. doi:10.1007/s10709-008-9308-0
- Habier D, Fernando RL, Kizilkaya K and Garrick DJ 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186. doi:10.1186/1471-2105-12-186.
- Hartwig S, Wellmann R, Emmerling R, Hamann H and Bennewitz J 2015. Short communication: Importance of introgression for milk traits in the German Vorderwald and Hinterwald cattle. *Journal of Dairy Science* 98, 2033–2038. doi:10.3168/jds.2014-8571.
- Hartwig S, Wellmann R, Hamann H and Bennewitz J 2014. The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. *Journal of Animal Breeding and genetics* 131, 496–503. doi:10.1111/jbg.12099.
- Ibáñez-Escriche N, Fernando RL, Toosi A and Dekkers JCM 2009. Genomic selection of purebreds for crossbred performance. *Genetics Selection Evolution* 41, 12. doi:10.1186/1297-9686-41-12.

- Jiang J, Shen B, O'Connell JR, VanRaden PM, Cole JB, and Ma L 2017. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics* 18:425. doi: 10.1186/s12864-017-3821-4
- Legarra A, Aguilar I and Misztal I 2009. A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92, 4656–4663. doi:10.3168/jds.2009-2061.
- Legarra A, Christensen OF, Aguilar I and Misztal I 2014. Single Step, a general approach for genomic selection. *Livestock Science* 166, 54–65. doi:10.1016/j.livsci.2014.04.029.
- Lopes MS, Bovenhuis H, Hidalgo AM, van Arendonk JAM, Knol EF and Bastiaansen JWM 2017. Genomic selection for crossbred performance accounting for breed-specific effects. *Genetics Selection Evolution* 49, 51. doi:10.1186/s12711-017-0328-z.
- Lopez-Villalobos N, Garrick DJ, Holmes CW, Blair HT, and Spelman RJ 2000. Profitabilities of some mating systems for dairy herds in New Zealand. *Journal of Dairy Science* 83:144–153. doi: 10.3168/jds.S0022-0302(00)74866-1
- los Campos G de, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K and Cotes JM 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi:10.1534/genetics.109.101501.
- MacLeod IM, Bowman PJ, Vander Jagt CJ, Haile-Mariam M, Kemper KE, Chamberlain AJ, Schrooten C, Hayes BJ, Goddard ME (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* 17 (144). doi: 10.1186/s12864-016-2443-6
- Madsen, P ., Su, G ., Labouriau, R . & Christensen, F 2010. DMU-a package for analyzing multivariate mixed models. In *Proc. 9th World Congress on Genetics Applied to Livestock Production* 137, 137. Leipzig, Germany.
- Meuwissen TH, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Meuwissen THE 2009. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* 41, 35. doi:10.1186/1297-9686-41-35.
- Meyer K 2007. WOMBAT—A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML)*. *Journal of Zhejiang University. Science. B* 8, 815–821. doi:10.1631/jzus.2007.B0815.

- Misztal I 1999. Complex Models, More Data: Simpler Programming? Proceedings of the Interbull Workshop Computers and Cattle Breeds, Tuusula, Finland. Interbull Bulletin 20, 33-42.
- Nishio M and Satoh M 2015. Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation. *Genetics Selection Evolution*. 47:32. DOI 10.1186/s12711-015-0091-y
- Park T and Casella G 2008. The Bayesian Lasso. *Journal of the American Statistical Association* 103, 681–686. doi:10.1198/016214508000000337.
- Samorè AB and Fontanesi L 2016. Genomic selection in pigs: state of the art and perspectives. *Italian Journal of Animal Science* 15, 211–232. doi:10.1080/1828051X.2016.1172034.
- Sevillano CA, Bovenhuis H and Calus MPL 2019. Genomic Evaluation for a Crossbreeding System Implementing Breed-of-Origin for Targeted Markers. *Frontiers in Genetics* 10, 418. doi:10.3389/fgene.2019.00418.
- Sevillano CA, Vandenplas J, Bastiaansen JWM and Calus MPL 2016. Empirical determination of breed-of-origin of alleles in three-breed cross pigs. *Genetics Selection Evolution* 48, 55. doi:10.1186/s12711-016-0234-9.
- Shepherd RK, Meuwissen THE and Woolliams JA 2010. Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics* 11, 529. doi:10.1186/1471-2105-11-529.
- Sørensen MK, Norberg E, Pedersen J and Christensen LG 2008. Invited review: crossbreeding in dairy cattle: a Danish perspective. *Journal of Dairy Science* 91, 4116–4128. doi:10.3168/jds.2008-1273.
- Strandén I, Christensen OF 2011. Allele coding in genomic evaluation. *Genetics Selection Evolution* 43, 25. doi: 10.1186/1297-9686-43-25
- Su G, Christensen OF, Ostensen T, Henryon M and Lund MS 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS ONE* 7, e45293. doi:10.1371/journal.pone.0045293.
- Tusell L, Gilbert H, Riquet J, Mercat M-J, Legarra A and Larzul C 2016. Pedigree and genomic evaluation of pigs using a terminal-cross model. *Genetics Selection Evolution* 48, 32. doi:10.1186/s12711-016-0211-3.

-
- van Grevenhof IEM and van der Werf JHJ 2015. Design of reference populations for genomic selection in crossbreeding programs. *Genetics Selection Evolution* 47, 14. doi:10.1186/s12711-015-0104-x.
- Vandenplas J, Calus MPL, Sevillano CA, Windig JJ and Bastiaansen JWM 2016. Assigning breed origin to alleles in crossbred animals. *Genetics Selection Evolution* 48, 61. doi:10.1186/s12711-016-0240-y.
- Vandenplas J, Windig JJ and Calus MPL 2017. Prediction of the reliability of genomic breeding values for crossbred performance. *Genetics Selection Evolution* 49, 43. doi:10.1186/s12711-017-0318-1.
- VanRaden PM 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91, 4414–4423. doi:10.3168/jds.2007-0980.
- Varona L, Legarra A, Toro MA, Vitezica ZG 2018. Non-additive Effects in Genomic Selection. *Frontiers in Genetics*. 9:78. doi: 10.3389/fgene.2018.00078
- Vitezica ZG, Varona L, Elsen J-M, Misztal I, Herring W and Legarra A 2016. Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genetics Selection Evolution* 48, 6. doi:10.1186/s12711-016-0185-1.
- Vitezica ZG, Varona L and Legarra A 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* 195, 1223–1230. doi:10.1534/genetics.113.155176.
- Vuori, K., Strandén, I., Lidauer, M., and Mäntysaari, E. A. 2006. MiX99 – Effectivesolver for large and complex linear mixed models. *Proc. 8thWorld Congr. Genet. Appl. Livest. Prod.* Belo Horizonte, Minas Gerais, Brazil.
- Wei M and van der Werf JHJ 1995. Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. *Journal of Animal Science* 73, 2220–2226. doi:10.2527/1995.7382220x.
- Wei M and van der Werf JHJ 1994. Maximizing genetic response in crossbreds using both purebred and crossbred information. *Animal Production* 59, 401–413. doi:10.1017/S0003356100007923.
- Wellmann R and Bennewitz J 2012. Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genetics Research* 94, 21–37. doi:10.1017/S0016672312000018.
- Wientjes YCJ and Calus MPL 2017. BOARD INVITED REVIEW: The purebred-crossbred correlation in pigs: A review of theory, estimates, and implications. *Journal of Animal Science* 95, 3467-3478. doi:10.2527/jas2017.1669.

-
- Xiang T, Nielsen B, Su G, Legarra A and Christensen OF 2016a. Application of single-step genomic evaluation for crossbred performance in pig. *Journal of Animal Science* 94, 936–948. doi:10.2527/jas.2015-9930.
- Xiang T, Christensen OF, Vitezica ZG and Legarra A 2016b. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genetics Selection Evolution* 48, 92. doi:10.1186/s12711-016-0271-4.
- Zeng J, Toosi A, Fernando RL, Dekkers JCM and Garrick DJ 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics Selection Evolution* 45, 11. doi:10.1186/1297-9686-45-11.
- Zumbach B, Misztal I, Tsuruta S, Holl J, Herring W and Long T 2007. Genetic correlations between two strains of Durocs and crossbreds from differing production environments for slaughter traits. *Journal of Animal Science* 85, 901–908. doi:10.2527/jas.2006-499.

Supplementary Material

The dominance model with imprinting

Robin Wellmann

February 2, 2020

The set of SNP is denoted as \mathcal{M} . All SNP are assumed to be biallelic and coded as alleles 1 and 2. A two-way cross \mathcal{X} with sire line \mathcal{A} and dam line \mathcal{B} is considered. The cross is obtained by random mating of males from line \mathcal{A} with females from line \mathcal{B} .

Definition 1. If a crossbred individual i has received allele A_1 from sire line \mathcal{A} and allele A_2 from dam line \mathcal{B} , then its genotype is denoted as A_1A_2 . The **centered genotype indicator** for genotype A_1A_2 is

$$W_{\mathcal{X}im}^{A_1A_2} = H_{\mathcal{X}im}^{A_1A_2} - \bar{H}_{\mathcal{X}m}^{A_1A_2},$$

where $H_{\mathcal{X}im}^{A_1A_2} \in \{0, 1\}$ equals one, if individual i from cross \mathcal{X} has genotype A_1A_2 at SNP m , and $\bar{H}_{\mathcal{X}m}^{A_1A_2}$ is the proportion of individuals from cross \mathcal{X} with genotype A_1A_2 at SNP m .

Definition 2. The **dominance model with imprinting** assumes that the genotypic value of an individual i from cross \mathcal{X} has the representation

$$g(i) = \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} ((W_{\mathcal{X}im}^{21} + W_{\mathcal{X}im}^{22})a_{\mathcal{A}m}^{\mathcal{X}} + (W_{\mathcal{X}im}^{12} + W_{\mathcal{X}im}^{22})a_{\mathcal{B}m}^{\mathcal{X}} + W_{\mathcal{X}im}^{21}d_{\mathcal{A}m}^{\mathcal{X}} + W_{\mathcal{X}im}^{12}d_{\mathcal{B}m}^{\mathcal{X}}),$$

where

- $a_{\mathcal{A}m}^{\mathcal{X}}$ is the additive effect of allele 2 at SNP m in cross \mathcal{X} if allele 2 was received from line \mathcal{A} ,
- $d_{\mathcal{A}m}^{\mathcal{X}}$ is the dominance effect of allele 2 at SNP m in cross \mathcal{X} if allele 2 was received from line \mathcal{A} .

The SNP effects $a_{\mathcal{B}m}^{\mathcal{X}}$ and $d_{\mathcal{B}m}^{\mathcal{X}}$ are defined accordingly.

Lemma 1. In the dominance model with imprinting is $\mu_{\mathcal{X}}$ the mean genotypic value of cross \mathcal{X} .

Proof: Let individual i_b be randomly chosen from cross \mathcal{X} . Conditionally on the SNP effects, we have

$$\begin{aligned}
 & E(g(i_b)) \\
 = & \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} (E(W_{\mathcal{X}im}^{21} + W_{\mathcal{X}im}^{22})a_{\mathcal{A}m}^{\mathcal{X}} + E(W_{\mathcal{X}im}^{12} + W_{\mathcal{X}im}^{22})a_{\mathcal{B}m}^{\mathcal{X}} + E(W_{\mathcal{X}im}^{21})d_{\mathcal{A}m}^{\mathcal{X}} + E(W_{\mathcal{X}im}^{12})d_{\mathcal{B}m}^{\mathcal{X}}) \\
 = & \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} (0a_{\mathcal{A}m}^{\mathcal{X}} + 0a_{\mathcal{B}m}^{\mathcal{X}} + 0d_{\mathcal{A}m}^{\mathcal{X}} + 0d_{\mathcal{B}m}^{\mathcal{X}}) \\
 = & \mu_{\mathcal{X}}
 \end{aligned}$$

□

Definition 3. The **centered allele content** of individual i from line \mathcal{A} is

$$Z_{Cim}^{\mathcal{A}} = C_{im}^{\mathcal{A}} - 2p_m^{\mathcal{A}},$$

where $C_{im}^{\mathcal{A}} \in \{0, 1, 2\}$ is the number of copies of allele 2, individual i has at SNP m .

Definition 4. The **breeding value** of individual i from line \mathcal{A} for **crossbred performance** with respect to mating partners from line \mathcal{B} is

$$c_{\mathcal{A}}(i) = 2(E(g(o_i)) - \mu_{\mathcal{X}}),$$

where the mating partner is chosen at random from line \mathcal{B} , and o_i is a randomly chosen offspring of individual i . The expectation is taken conditionally on the SNP effects.

Theorem 1. The breeding value of individual i from line \mathcal{A} for crossbred performance with respect to mating partners from \mathcal{B} is

$$c_{\mathcal{A}}(i) = \sum_{m \in \mathcal{M}} Z_{Cim}^{\mathcal{A}} \alpha_{\mathcal{A}m}^{\mathcal{X}},$$

where the allele substitution effects for crossbred performance are

$$\alpha_{\mathcal{A}m}^{\mathcal{X}} = a_{\mathcal{A}m}^{\mathcal{X}} + (1 - p_m^{\mathcal{B}})d_{\mathcal{A}m}^{\mathcal{X}} - p_m^{\mathcal{B}}d_{\mathcal{B}m}^{\mathcal{X}}$$

for $m \in \mathcal{M}$.

Proof: Let o_i be a randomly chosen offspring of individual i from line \mathcal{A} . The individual has a paternal haplotype $\tilde{\mathcal{H}}_i$ and a maternal haplotype $\tilde{\mathcal{P}}_{\mathcal{B}}$, where

- haplotype $\tilde{\mathcal{H}}_i$ originates from animal i , and
- haplotype $\tilde{\mathcal{P}}_{\mathcal{B}}$ originates from population \mathcal{B} .

Let $\mathcal{H}_i = \tilde{\mathcal{H}}_i - 1$, $\mathcal{P}_{\mathcal{B}} = \tilde{\mathcal{P}}_{\mathcal{B}} - 1$, and $m \in \mathcal{M}$. Since $\mathcal{H}_{im}, \mathcal{P}_{\mathcal{B}m} \in \{0, 1\}$, the uncentered genotype indicators for offspring o_i are

$$\begin{aligned} H_{\mathcal{X}o_im}^{11} &= (1 - \mathcal{H}_{im})(1 - \mathcal{P}_{\mathcal{B}m}) \\ H_{\mathcal{X}o_im}^{12} &= (1 - \mathcal{H}_{im})\mathcal{P}_{\mathcal{B}m} \\ H_{\mathcal{X}o_im}^{21} &= \mathcal{H}_{im}(1 - \mathcal{P}_{\mathcal{B}m}) \\ H_{\mathcal{X}o_im}^{22} &= \mathcal{H}_{im}\mathcal{P}_{\mathcal{B}m}. \end{aligned}$$

Since the cross is obtained by random mating of individuals from line \mathcal{A} with individuals from line \mathcal{B} , the centered genotype indicators are

$$\begin{aligned} W_{\mathcal{X}o_im}^{11} &= (1 - \mathcal{H}_{im})(1 - \mathcal{P}_{\mathcal{B}m}) - (1 - p_m^{\mathcal{A}})(1 - p_m^{\mathcal{B}}) \\ W_{\mathcal{X}o_im}^{12} &= (1 - \mathcal{H}_{im})\mathcal{P}_{\mathcal{B}m} - (1 - p_m^{\mathcal{A}})p_m^{\mathcal{B}} \\ W_{\mathcal{X}o_im}^{21} &= \mathcal{H}_{im}(1 - \mathcal{P}_{\mathcal{B}m}) - p_m^{\mathcal{A}}(1 - p_m^{\mathcal{B}}) \\ W_{\mathcal{X}o_im}^{22} &= \mathcal{H}_{im}\mathcal{P}_{\mathcal{B}m} - p_m^{\mathcal{A}}p_m^{\mathcal{B}}. \end{aligned}$$

For brevity, we write $W_{o_im}^{A_1 A_2} = W_{\mathcal{X}o_im}^{A_1 A_2}$ in the following. Conditionally on the SNP effects we have

$$\begin{aligned} &E(g(o_i)) \\ &= E\left(\mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} ((W_{o_im}^{21} + W_{o_im}^{22})a_{\mathcal{A}m}^{\mathcal{X}} + (W_{o_im}^{12} + W_{o_im}^{22})a_{\mathcal{B}m}^{\mathcal{X}} + W_{o_im}^{21}d_{\mathcal{A}m}^{\mathcal{X}} + W_{o_im}^{12}d_{\mathcal{B}m}^{\mathcal{X}})\right) \\ &= \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} (E(W_{o_im}^{21} + W_{o_im}^{22})a_{\mathcal{A}m}^{\mathcal{X}} + E(W_{o_im}^{12} + W_{o_im}^{22})a_{\mathcal{B}m}^{\mathcal{X}} + E(W_{o_im}^{21})d_{\mathcal{A}m}^{\mathcal{X}} + E(W_{o_im}^{12})d_{\mathcal{B}m}^{\mathcal{X}}). \end{aligned}$$

Animal o_i carries with probability 0.5 the paternal allele and with probability 0.5 the maternal allele of parent i , so

$$E(\mathcal{H}_{im}) = \frac{C_{im}^A}{2},$$

and

$$E(\mathcal{H}_{im}) - p_m^A = \frac{Z_{Cim}^A}{2}.$$

Thus,

$$\begin{aligned} E(W_{o_i m}^{21} + W_{o_i m}^{22}) &= E(\mathcal{H}_{im}(1 - \mathcal{P}_{Bm}) - p_m^A(1 - p_m^B) + \mathcal{H}_{im}\mathcal{P}_{Bm} - p_m^A p_m^B) \\ &= E(\mathcal{H}_{im} - p_m^A) \\ &= \frac{Z_{Cim}^A}{2}, \\ E(W_{o_i m}^{12} + W_{o_i m}^{22}) &= E((1 - \mathcal{H}_{im})\mathcal{P}_{Bm} - (1 - p_m^A)p_m^B + \mathcal{H}_{im}\mathcal{P}_{Bm} - p_m^A p_m^B) \\ &= E(\mathcal{P}_{Bm} - p_m^B) \\ &= 0, \end{aligned}$$

Since the mating partner of individual i was chosen at random from line \mathcal{B} , we have

$$\begin{aligned} E(W_{o_i m}^{21}) &= E(\mathcal{H}_{im}(1 - \mathcal{P}_{Bm}) - p_m^A(1 - p_m^B)) \\ &= E(\mathcal{H}_{im})(1 - p_m^B) - p_m^A(1 - p_m^B) \\ &= E(\mathcal{H}_{im} - p_m^A)(1 - p_m^B) \\ &= \frac{Z_{Cim}^A}{2}(1 - p_m^B), \\ E(W_{o_i m}^{12}) &= E((1 - \mathcal{H}_{im})\mathcal{P}_{Bm} - (1 - p_m^A)p_m^B) \\ &= (1 - E(\mathcal{H}_{im}))p_m^B - (1 - p_m^A)p_m^B \\ &= (p_m^A - E(\mathcal{H}_{im}))p_m^B \\ &= (E(\mathcal{H}_{im}) - p_m^A)(-p_m^B) \\ &= \frac{Z_{Cim}^A}{2}(-p_m^B). \end{aligned}$$

Thus,

$$\begin{aligned} E(g(o_i)) &= \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} \left(\frac{Z_{Cim}^A}{2} a_{Am}^{\mathcal{X}} + 0 a_{Bm}^{\mathcal{X}} + \frac{Z_{Cim}^A}{2} (1 - p_m^B) d_{Am}^{\mathcal{X}} + \frac{Z_{Cim}^A}{2} (-p_m^B) d_{Bm}^{\mathcal{X}} \right) \\ &= \mu_{\mathcal{X}} + \sum_{m \in \mathcal{M}} \frac{Z_{Cim}^A}{2} (a_{Am}^{\mathcal{X}} + (1 - p_m^B) d_{Am}^{\mathcal{X}} + (-p_m^B) d_{Bm}^{\mathcal{X}}) \\ &= \mu_{\mathcal{X}} + \frac{1}{2} \sum_{m \in \mathcal{M}} Z_{Cim}^A (a_{Am}^{\mathcal{X}} + (1 - p_m^B) d_{Am}^{\mathcal{X}} - p_m^B d_{Bm}^{\mathcal{X}}). \end{aligned}$$

□

Chapter 3

Improving the Accuracy of Multi-Breed Prediction in Admixed Populations by Accounting for the Breed Origin of Haplotype Segments

Markus Schmid^{1*}, Joana Stock¹, Jörn Bennewitz¹, Robin Wellmann¹

¹ Institute of Animal Science, Department of Animal Genetics and Breeding, University of Hohenheim, Garbenstrasse 17, 70599 Stuttgart, Germany.

* Correspondence:

Markus Schmid, Markus_schmid@uni-hohenheim.de

Published in *Frontiers in Genetics*, doi: 10.3389/fgene.2022.840815

Abstract

Numerically small breeds have often been upgraded with mainstream breeds. This historic introgression predisposes the breeds for joint genomic evaluations with mainstream breeds. The LD structure differs between breeds. The marker effects of a haplotype segment may, therefore, depend on the breed from which the haplotype segment originates. An appropriate method for genomic evaluation would account for this dependency. This study proposes a method for the computation of genomic breeding values for small admixed breeds that incorporate phenotypic and genomic information from large introgressed breeds by considering the breed origin of alleles (BOA) in the evaluation. The proposed BOA model classifies haplotype segments according to their origins and assumes different but correlated SNP effects for the different origins. The BOA model was compared in a simulation study to conventional within-breed genomic best linear unbiased prediction (GBLUP) and conventional multi-breed GBLUP models. The BOA model outperformed within-breed GBLUP as well as multi-breed GBLUP in most cases.

Keywords: Admixed population, multi-breed genomic prediction, BOA model, cattle, allele origin

Introduction

The efficiency of breeding programs for local breeds is often compromised by the limited number of individuals and has resulted in a decreasing competitiveness with high yielding breeds, especially with the advent of genomic selection (GS). In GS large reference populations are required to accurately predict breeding values of the individuals (Goddard and Hayes 2009) and are therefore difficult to establish in small local breeds. In order to improve the performance of local breeds, sires of closely related high-yielding breeds were frequently used in the past and genetic gain has been generated by introgression. Such strategies increase the genetic relatedness between breeds because a certain number of alleles of the high yielding breed segregate within the target breed after introgression.

Several studies were conducted using different approaches to enable GS in numerically small breeds using the reference population of a second breed (across-breed prediction) or extending the own reference population by adding the reference population of the second breed (multi-breed prediction) as reviewed by Lund et al. (2014, 2016). The major findings were that across-breed prediction is often not suitable to improve the accuracy of prediction and that the benefit of multi-breed reference populations strongly depend on the relatedness between the breeds and density of the SNP panels. A substantial increase in accuracy can only be expected when the breeds are closely related and the number of SNPs is high to capture across-breed linkage disequilibrium (LD) between markers and QTLs. However, variation of LD as well as differences of allelic effects across populations limit the application of such approaches. Different models were proposed accounting for breed-specific effects (e.g. Makgahlela et al. 2012, Thomassen et al. 2013, El Hamidi and Rekaya 2015, van den Berg et al. 2020) and differences in LD (Rahimi et al. 2020) in the field of multi-breed dairy cattle evaluation. One way is to assign the breed origin of alleles (BOA) (Wellmann 2019, Vandenplas et al. 2016) that allows for models assuming SNP effects to be different but correlated across breeds. Such models were applied to simulated and real datasets of crossbred or admixed populations in cattle (Karaman et al. 2021) as well as other livestock (e.g. Duenk et al. 2019) or plant species (Rio et al. 2020) and are reviewed in Stock et al. (2020) and Duenk et al. 2021) The studies have shown that considering BOA has the potential to increase the accuracy of multi-breed GS.

In many numerical small dairy cattle breeds sires from a large and high yielding breed were used in order to speed up genetic gain in the small breed. This resulted in some cases in a substantial amount of introgressed genes and in a mosaic-like haplotype pattern with a mix of native and introgressed haplotypes. For example, in the German Angler breed located in the northern part of Germany, admixture plays a substantial role in the population structure and the proportion of

migrant alleles from other breeds is remarkable (Addo et al. 2019, Wang et al. 2017a,b, Schmidtmann et al. 2021). A very close relationship to the Holstein Friesian breed, especially the Red Holstein breed, was observed (Wang et al. 2017b). A similar level of admixture was observed for the German Vorderwald breed, where the genetic progress was mostly driven by the introgressed genes (Hartwig et al. 2014, 2015). For these kinds of breeds, a genomic model that considers the mosaic pattern of the haplotype structure would be beneficial in multi-breed genomic evaluations.

This study proposes a method for the computation of genomic breeding values for small admixed breeds that incorporates phenotypic and genomic information from large introgressed breeds. A multi-breed BOA model is derived for multi-breed genomic selection that is suitable for application when the individuals have fragmented genomes. It classifies haplotype segments according to their origins and assumes different SNP effects for the different origins.

For validation, it was compared with models that did not consider the breed-origin of QTL alleles. All models were applied to simulated datasets. In the simulation, the genotypes of the small admixed breed were derived from German Angler cattle, while the genotypes of the introgressed breed were derived from German Holstein cattle. Different scenarios were investigated in which the number of genotyped animals of the target breed, i.e. the numerically small Angler breed, varied, while the number of genotyped animals of the large introgressed breed, i.e. the German Holstein, remained constant.

Material and Methods

Simulation

The data basis for the simulation study were 50k SNP-chip (Illumina BovineSNP50 BeadChip, Illumina Inc., San Diego, CA) genotypes of Angler (AN) (Wang et al. 2017a) and Holstein (HF) (Streit et al. 2013) individuals from the German population. Starting with the base generation, one further generation was simulated for each breed according to the simulation protocol of Stock et al. (2021) with R-package x-breed (Esfandyari and Sørensen 2017). The resulting simulated HF dataset (simHF) consisted of 6,000 individuals and the simulated Angler data set contained 3,000 individuals.

Several subsets of the total Angler data set were sampled to mimic different population sizes for the small breed. Subset simAN1 consisted of 750 individuals, simAN2 consisted of 1,500 individuals and simAN3 contained all 3,000 simulated Angler individuals. The different simAN

populations are referred to as breed size scenarios. The sample sizes represent 12.5%, 25% and 50% of the number of simHF individuals. In each of the subsets, all sires had the same number of offspring.

Table 2: Overview of scenarios.

Scenario	n	Validation Set	Reference Set 1 (within-breed)	Reference Set 2 (multi-breed)	simAN % (multi-breed)
simAN1	750	150	600	6,600	9.09
simAN2	1,500	300	1,200	7,200	16.67
simAN3	3,000	600	2,400	8,400	28.57

The size of the simulated Angler data set (n), the number of simulated Angler individuals in each validation set (Validation Set), the number of simulated Angler individuals in each reference set for within-breed prediction (Reference Set 1), the number of simulated Angler and simulated Holstein Friesian individuals in the reference set for multi-breed prediction (Reference Set 2), and the proportion of simulated Angler individuals in the multi-breed reference set (simAN %).

From the 23,448 SNPs that segregated in both breeds, 1,000 SNPs were randomly selected as QTLs, while the remaining SNPs were used as markers for genomic prediction.

The QTL effects for the simAN datasets and the simHF dataset were correlated. The additive effects $a_q^{\text{HF}}, a_q^{\text{AN}}$ of QTL q were correlated and normally distributed with

$$\begin{pmatrix} a_q^{\text{HF}} \\ a_q^{\text{AN}} \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_A^2 \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix} \right)$$

Hence, the correlation of QTL effects between the two simulated breeds was 0.95. Dominance was not modelled. The additive effects were scaled to represent a trait with an additive variance of $V_A \approx 0.3$ in each of the pure breeds. The additive genetic variance was calculated using standard formulas (Falconer and Mackay 1996). The phenotypes were obtained by adding normally distributed errors to the true genotypic values. The error variance was chosen to obtain a phenotypic variance of $V_P = 1$. The entire simulation was independently repeated ten times.

True Breeding Values

The true breeding values (TBV) were calculated as

$$TGV_{ik} = \sum_{q \in Q} \sum_{k \in \{\text{AN}, \text{HF}\}} Z_{Aiq}^k a_q^k$$

where a_q^k is the additive effect of QTL q when the haplotype segment containing the QTL originated from breed $k \in \{\text{AN}, \text{HF}\}$. The allele content $Z_{Ai q}^k \in \{0, 1, 2\}$ equals the number of copies of the alternative allele in individual i that originate from breed k . The R-package optiSel (Wellmann 2019) was used to assign the QTL alleles from the simAN data set to either the Angler or the Holstein breed in a segment-based approach. Considered segments consisted of minimum 20 consecutive markers and had a length of $\geq 1.5\text{Mb}$. No Angler cattle were introgressed into Holstein cattle, so the summand corresponding to $k = \text{AN}$ equals zero in the simHF data set.

Genomic Prediction

The BOA model was compared with two conventional methods for the prediction of genomic breeding values, which are within-breed prediction with GBLUP for the simulated Angler cattle, and a multi-breed prediction with GBLUP.

The BOA model: It is assumed that genotypes and phenotypes from several breeds or crosses are available, which includes the target breed. The number of SNP is denoted as M , and the total number of individuals as N . The genotypes are phased, so each individual i has a maternal haplotype h_i° and a paternal haplotype h_i^σ . The binary coded alleles of individual i at SNP m are denoted as $h_{im}^\circ, h_{im}^\sigma \in \{0, 1\}$. The origins $o_{im}^\sigma, o_{im}^\circ \in \{1, \dots, K\}$ of all positions m in all haplotypes h_i°, h_i^σ are determined, whereby K denotes the number of possible origins. Each origin can be considered as a genetic group, whereby the first genetic group is the target breed. The indicators for genetic group k are denoted as

$$\delta_{kim}^\sigma = \begin{cases} 1 & \text{if } o_{im}^\sigma = k \\ 0 & \text{otherwise} \end{cases} \text{ and}$$

$$\delta_{kim}^\circ = \begin{cases} 1 & \text{if } o_{im}^\circ = k \\ 0 & \text{otherwise} \end{cases}.$$

The model equation for the phenotypic value of individual i is

$$y_i = \sum_{k=1}^K c_{ik} \beta_k + \sum_{k=1}^K \sum_{m=1}^M (h_{im}^\sigma \delta_{kim}^\sigma + h_{im}^\circ \delta_{kim}^\circ) a_m^k + e_i$$

where $c_{ik} \in [0, 1]$ is the genetic contribution, individual i has from genetic group k , β_k is the fixed effect of genetic group k , a_m^k is the normally distributed additive effect of marker m in genetic group k , and e_i is the residual. The model equation in matrix form is

$$y = X\beta + \sum_{k=1}^K Z_A^k a_k + e$$

where $\beta = (\beta_1, \dots, \beta_K)^T$ is the vector of fixed effects with $N \times K$ design matrix X , where $X_{ik} = c_{ik}$. The M -vector a_k of random SNP effects for genetic group k has design matrix Z_A^k , where $Z_{Aim}^k = h_{im}^\sigma \delta_{kim}^\sigma + h_{im}^\varphi \delta_{kim}^\varphi$ is the number of copies of the alternative allele that originate from genetic group k . The vector e with residuals has covariance matrix $R = \sigma_e^2 I$. The covariance matrix of the KM -vector $a = (a_1^T, \dots, a_K^T)^T$ is

$$D = Cov(a) = \begin{pmatrix} \sigma_{A1}^2 I & \cdots & \sigma_{A1K} I \\ \vdots & \ddots & \vdots \\ \sigma_{A1K} I & \cdots & \sigma_{AK}^2 I \end{pmatrix} = \Sigma \otimes I$$

with

$$\Sigma = \begin{pmatrix} \sigma_{A1}^2 & \cdots & \sigma_{A1K} \\ \vdots & \ddots & \vdots \\ \sigma_{AK1} & \cdots & \sigma_{AK}^2 \end{pmatrix}.$$

Alternative representations of the model and the mixed model equations are given in the appendix. The BOA model requires the breed origins o_{im}^σ and o_{im}^φ of the haplotypes as input parameters. The R-package optiSel (Wellmann 2019) was used to assign all marker haplotype segments from the simAN data set to either the Angler or the Holstein breed. Considered segments consisted of minimum 20 consecutive markers and had a length of ≥ 1.5 Mb.

The covariance matrix Σ of the marker effects of the two pure breeds was needed as an input parameter. The correlation was determined by a grid search for each replicate and breed size. A 5-fold cross-validation was conducted for each candidate value, and the correlation was chosen for further analyses that maximized the accuracy of the predictions with the BOA model. As the resulting correlation values were similar across all replicates and breed size scenarios and the mean value was approximately 0.75, this value ($r = 0.75$) was chosen as input parameter for all subsequent evaluations. It was expected that such an approach would avoid an overfitting of the model. The correlation matrix for the marker effects was multiplied with the estimated variance of the allele substitution effects to obtain the covariance matrix of the marker effects.

Alternative models: For within-breed genomic prediction we used the model

$$y = \beta \mathbf{1} + Z_A a + e,$$

where the M -vector a of random SNP effects has covariance matrix $D = \sigma_a^2 I$ and design matrix Z_A , where Z_{Aim} is the number of copies of the alternative allele at marker m in individual i . The vector e with residuals has covariance matrix $R = \sigma_e^2 I$. For multi-breed genomic prediction, the same model was used. We did not include a breed effect because no breed effect was simulated. The SNP markers for genomic prediction were chosen as follows. From the 22,448 SNPs that were not chosen as QTLs, all SNPs that segregated with a minor allele frequency (MAF) < 0.03 within one of the simulated breeds and SNPs that did not segregate in both breeds were omitted. Across all replicates, on average 21,670 SNPs remained and were used for genomic prediction.

Cross Validation

The genomic predictions were done separately for each breed-size scenario and each replicate. The accuracies of prediction were assessed by a 5-fold cross validation. The individuals of the respective simAN dataset were assigned to 5 different classes such that individuals from different classes had no sires in common. Hence, each class included the offspring of 10 sires. In each cross-validation cycle, one class was used as the validation set, and the four remaining classes were used as the reference population.

For multi-breed GBLUP and for the BOA model, the respective simAN reference set was joined with the simHF individuals. Consequently, the number of individuals from the simAN population in the reference population varied, while the number of simHF individuals was constant.

An overview on the sample sizes is given in Table 1. The reference populations for within-breed prediction consisted of 600, 1,200, and 2,400 simAN individuals, respectively. The reference populations for multi-breed prediction were enlarged by the 6,000 simHF individuals. The proportions of simAN individuals in the multi-breed reference population were thus 9%, 17% and 29% for the simAN1, simAN2 and simAN3 scenario, respectively.

The marker effects were estimated with all three models for each cross-validation cycle.

Estimated Breeding Values

The genomic breeding values of the individuals in the validation set were computed for the BOA model as

$$GEBV_i = \sum_{m \in M} \sum_{k \in \{AN, HF\}} Z_{Aim}^k \hat{a}_m^k,$$

where M is the set of SNP markers, \hat{a}_m^k is the estimated SNP effect of marker m that is used for haplotype segments originating from genetic group k , and the allele content $Z_{Aim}^k \in \{0, 1, 2\}$ equals

the number of copies of the alternative allele in individual i that originate from genetic group k . For the other models, the genomic breeding values were estimated as

$$GEBV_i = \sum_{m \in M} Z_{Aim} \hat{a}_m,$$

where \hat{a}_m is the estimated SNP effect of marker m , and the allele content $Z_{Aim} \in \{0,1,2\}$ equals the number of copies of the alternative allele in individual i at marker m .

Prediction Accuracies

For each method, the prediction accuracy was calculated as the correlation between the GEBVs and the TBVs of the validation individuals. The accuracies presented in the results are the averages, taken over all cross-validation cycles and replicates.

Results

The mean proportion of SNPs with Holstein origin across all replicates was 0.157 ± 0.007 . Table 2 shows the results of the model comparison for all investigated breed-size scenarios. In general, the prediction accuracies increased with increasing size of the reference population. The BOA model provided the highest accuracies for simAN1 and simAN2, whereas it showed the same mean accuracy as within-breed GBLUP for simAN3. Multi-breed GBLUP was inferior to the other models in simAN1 and simAN3. Within-breed GBLUP resulted in the lowest accuracies in the medium-sized reference population scenario simAN2. The standard deviations (SD) of the accuracies were highest using the small reference set, while it showed the smallest SD values in the medium-sized reference set. The standard errors of the accuracies were relatively small (0.011-0.017).

Table 3: Mean accuracies of genomic prediction in the simulated breed size scenarios.

Model	Reference set	simAN1			simAN2			simAN3		
		mean	sd	se	mean	sd	se	mean	sd	se
BOA	simAN + simHF	0.407	0.110	0.016	0.492	0.074	0.011	0.546	0.085	0.012
GBLUP	simAN + simHF	0.387	0.118	0.017	0.486	0.078	0.011	0.541	0.092	0.013
GBLUP	simAN	0.401	0.114	0.016	0.477	0.080	0.011	0.546	0.083	0.012

The mean, standard deviation (sd) and standard error (se) of the 5-fold cross-validation across the 10 simulated replicates are given for the breed size scenarios simAN1, simAN2 and simAN3 and for the BOA and the GBLUP models. Column 2 indicates whether a multi-breed reference set or a within-breed reference set was used.

Discussion

It was shown in this study that a multi-breed genomic prediction with the proposed BOA model can increase the accuracies of the GEBVs for numerically small admixed populations over multi-breed and within-breed genomic predictions with GBLUP under certain conditions. The model considers the breed origin of each haplotype in individuals with admixed genomes consisting of native and introgressed haplotype segments. This makes it especially interesting for numerical small breeds with historic introgression from high yielding breeds, as it was observed e.g. in the German Angler or Vorderwald breed (Addo et al. 2019, Wang et al. 2017a,b, Schmidtman et al. 2021, Hartwig et al. 2014, 2015).

The simulated QTL positions were the same in both breeds. Recent mutations were neglected that could have created new QTLs. The QTL effects of both genetic groups were assumed to be highly correlated with a correlation of 0.95. The QTL positions were chosen from the 50k chip, so the simulated QTLs are common variants whereas a large fraction of the additive variance is expected to come from rare QTL variants (e.g. Kemper and Goddard 2012, Visscher et al. 2017). This can compromise a direct transition of the simulation results to real data.

The LD structures in the simulation are expected to be similar to those investigated in real populations (Qanbari et al. 2010, Addo et al. 2019, Schmidtman et al. 2021) because only one generation was simulated, so recombination could occur only within one meiotic division. Although the QTL effects were highly correlated, the correlation of the marker effects was only 0.75. The reason for the relatively low correlation of marker effects is that the QTLs were excluded from the marker set. Therefore, the effect of a single QTL is captured by several markers and the LD between markers and QTLs is different in both genetic groups. The shorter ranges of LD in admixed populations like the German Angler compared with other breeds (Addo et al. 2019, Schmidtman et al. 2021) contributes to the observed low correlation. A higher correlation between marker effects might be observed when more dense markers would be used and a heavy-tailed distribution of marker effects would be assumed.

This paper focused on methods to improve GS in small admixed populations. It compared the prediction accuracies of various methods at the example of a simulated population that had a similar LD structure as the target breeds. A detailed quantification of the impact of influencing factors (e.g. LD and its consistency across populations, or the relatedness and genetic correlation between the populations) on the accuracies was beyond the scope of the study.

In this study, the multi-breed BOA model was compared with conventional multi-breed GBLUP and within-breed GBLUP. The multi-breed BOA approach led to an increase in the accuracy of the genomic breeding values when the number of genotyped AN individuals was small and

medium, and showed similar results as the within-breed GBLUP method for the large reference sets. The difference between the prediction accuracies of the models, however, tend to decrease with an increasing number of genotyped AN individuals. The multi-breed prediction with GBLUP was not superior to within-breed prediction when the number of genotyped AN individuals was large. The reason is possibly that the multi-breed GBLUP model assumes a perfect correlation between the marker effects of both breeds. This assumption was certainly violated in the simulation. The BOA model, which accounts for the correlation between marker effects, could improve upon single-breed evaluations and outperformed multi-breed GBLUP in all cases.

For the prediction of genomic breeding values of Angler in practice, to date, a joint reference population of several Scandinavian red dairy breeds (i.e. Danish Red, Norwegian Red, Swedish Red, and Finnish Ayrshire) is used. To increase the accuracies of the GEBVs for Angler, about 170 genotyped and progeny-tested German Angler bulls have been included to this reference set as well (private communication RSheG, 2021). Hence, the findings of the study in scenario simAN1 might be most relevant for the current Angler cattle breeding program.

In the past decades, the Angler breed has been upgraded with other breeds, such as Red Holstein and Holstein Friesian to increase its economic value. This has led to relatively high kinships between them (Wang et al. 2017b). However, in this study only Holstein Friesian genotypes were available and considered, and thus the total amount of introgression was probably not detected completely. In addition, the available Holstein Friesian genotypes originated from the current population, which might have also biased the categorization of the native parts of the genome. Generally, multi-breed prediction is increasingly beneficial when applied to high density marker information or whole-genome sequence data (Lund et al. 2014), however, such datasets are mostly not available in cost-efficient breeding programs of small local cattle populations.

Conclusion

A multi-breed genomic prediction with the proposed BOA model increased the accuracies of the estimated genomic breeding values for numerically small admixed populations over multi-breed and within-breed genomic predictions with GBLUP. The BOA model assumes that the additive effect of an allele depends on the genetic group from which the respective haplotype segment originates. It is of special interest for multi-breed genomic predictions for numerical small breeds with past introgression from high yielding breeds.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

RW, JB and MS designed the study. RW developed the proposed BOA model. JS and MS simulated the data. MS did the statistical analyses. All authors drafted and approved the manuscript.

Funding

JS was partly supported by the H.Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged.

Data Availability Statement

The genotype data used in this study was published and was cited in the manuscript. Simulated data can be made available at reasonable request.

Acknowledgments

Not applicable.

References

- Addo, S., S. Klingel, D. Hinrichs, and G. Thaller. 2019. "Runs of Homozygosity and NetView Analyses Provide New Insight into the Genome-Wide Diversity and Admixture of Three German Cattle Breeds." *PLoS ONE* 14 (12): 1–20.
<https://doi.org/10.1371/journal.pone.0225847>.
- Berg, I. van den, I. M. MacLeod, C. M. Reich, E. J. Breen, and J. E. Pryce. 2020. "Optimizing Genomic Prediction for Australian Red Dairy Cattle." *Journal of Dairy Science* 103 (7): 6276–98. <https://doi.org/10.3168/jds.2019-17914>.
- Duenk, P., M.P.L. Calus, Y.C.J. Wientjes, V. P. Breen, J. M. Henshall, R. Hawken, and P. Bijma. 2019. "Validation of Genomic Predictions for Body Weight in Broilers Using Crossbred Information and Considering Breed-of-Origin of Alleles." *Genetics Selection Evolution* 51 (1): 1–12. <https://doi.org/10.1186/s12711-019-0481-7>.
- Duenk, P., P. Bijma, Y.C.J. Wientjes, and M.P.L. Calus. 2021. "Review: Optimizing Genomic Selection for Crossbred Performance by Model Improvement and Data Collection." *Journal of Animal Science* 99 (8): 1–24. <https://doi.org/10.1093/jas/skab205>.
- Esfandyari, H. and A. C. Sørensen. 2017. "Xbreed: An R Package for Genomic Simulation of Purebreds and Crossbreds." In *68th Annual Meeting of the European Federation of Animal Science*. Tallinn, Estonia.
- Falconer, D. S., and T. F. C. Mackay. 1996. "Introduction to Quantitative Genetics." In , 4th ed. London: Longman Group Ltd.
- Goddard, M. E., and B. J. Hayes. 2009. "Mapping Genes for Complex Traits in Domestic Animals and Their Use in Breeding Programmes." *Nature Reviews Genetics* 10 (6): 381–91. <https://doi.org/10.1038/nrg2575>.
- Hamidi, H. E. and R. Rekaya. 2015. "A Multi-Compartment Model for Genomic Selection in Multi-Breed Populations." *Livestock Science* 177: 1–7.
<https://doi.org/10.1016/j.livsci.2015.03.027>.
- Hartwig, S., R. Wellmann, R. Emmerling, H. Hamann, and J. Bennewitz. 2015. "Short Communication: Importance of Introgression for Milk Traits in the German Vorderwald and Hinterwald Cattle." *Journal of Dairy Science* 98 (3): 2033–38.
- Hartwig, S., R. Wellmann, H. Hamann, and J. Bennewitz. 2014. "The Contribution of Migrant

-
- Breeds to the Genetic Gain of Beef Traits of German Vorderwald and Hinterwald Cattle.” *Journal of Animal Breeding and Genetics* 131: 496–503.
- Karaman, E., G. Su, I. Croue, and M. S. Lund. 2021. “Genomic Prediction Using a Reference Population of Multiple Pure Breeds and Admixed Individuals.” *Genetics Selection Evolution* 53 (1): 1–15. <https://doi.org/10.1186/s12711-021-00637-y>.
- Kemper, K. E., and M. E. Goddard. 2012. “Understanding and Predicting Complex Traits: Knowledge from Cattle.” *Human Molecular Genetics* 21 (R1): 45–51. <https://doi.org/10.1093/hmg/dds332>.
- Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. “Review: How to Improve Genomic Predictions in Small Dairy Cattle Populations.” *Animal* 10 (6): 1042–49. <https://doi.org/10.1017/S1751731115003031>.
- Lund, M. S., G. Su, L. Janss, B. Guldbrandtsen, and R. F. Brøndum. 2014. “Invited Review: Genomic Evaluation of Cattle in a Multi-Breed Context.” *Livestock Science* 166 (1): 101–10. <https://doi.org/10.1016/j.livsci.2014.05.008>.
- Makgahlela, M. L., E. A. Mäntysaari, I. Strandén, M. Koivula, U. S. Nielsen, M. J. Sillanpää, and J. Juga. 2012. “Across Breed Multi-Trait Random Regression Genomic Predictions in the Nordic Red Dairy Cattle.” *Journal of Animal Breeding and Genetics* 130 (1): 10–19. <https://doi.org/10.1111/j.1439-0388.2012.01017.x>.
- Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, a. R. Sharifi, and H. Simianer. 2010. “The Pattern of Linkage Disequilibrium in German Holstein Cattle.” *Animal Genetics* 41 (4): 346–56. <https://doi.org/10.1111/j.1365-2052.2009.02011.x>.
- Rahimi, S. M., A. Rashidi and H. Esfandyari. 2020. “Accounting for Differences in Linkage Disequilibrium in Multi-Breed Genomic Prediction.” *Livestock Science* 240 (June): 104165. <https://doi.org/10.1016/j.livsci.2020.104165>.
- Rio, S., L. Moreau, A. Charcosset and T. Mary-Huard. 2020. “Accounting for Group-Specific Allele Effects and Admixture in Genomic Predictions: Theory and Experimental Evaluation in Maize.” *Genetics* 216 (1): 27–41. <https://doi.org/10.1534/genetics.120.303278>.
- RShEG 2021. Personal Communication (Email, K. Breuer, 12.11.2021).
- Schmidtman, C., A. Schoenherz, Bernt Guldbrandtsen, J. Marjanovic, M.P.L. Calus, D. Hinrichs, and G. Thaller. 2021. “Assessing the Genetic Background and Genomic

-
- Relatedness of Red Cattle Populations Originating from Northern Europe." *Genetics Selection Evolution*. Vol. 53. <https://doi.org/10.1186/s12711-021-00613-6>.
- Stock, J., H. Esfandyari, D. Hinrichs and J. Bennewitz. 2021. "Implementing a Genomic Rotational Crossbreeding Scheme to Promote Local Dairy Cattle Breeds - A Simulation Study." *Journal of Dairy Science* 104. <https://doi.org/10.3168/jds.2020-19927>.
- Stock, J., Bennewitz, J., D. Hinrichs, and R. Wellmann. 2020. "A Review of Genomic Models for the Analysis of Livestock Crossbred Data." *Frontiers in Genetics* 11 (June): 1–10. <https://doi.org/10.3389/fgene.2020.00568>.
- Streit M, R. Wellmann, F. Reinhardt, G. Thaller, H.-P. Piepho and J. Bennewitz. 2013. Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle. *G3 (Bethesda, Md.)* 3, 1085–1093. doi:10.1534/g3.113.006536.
- Thomassen, J. R., A. C. Sørensen, Guosheng Su, Per Madsen, Mogens Sandø Lund, and Bernt Guldbraendtsen. 2013. "The Admixed Population Structure in Danish Jersey Dairy Cattle Challenges Accurate Genomic Predictions." *Journal of Animal Science* 91 (7): 3105–12. <https://doi.org/10.2527/jas.2012-5490>.
- Vandenplas, J., M.P.L. Calus, C. A. Sevillano, J. J. Windig, and J. W. M. Bastiaansen. 2016. "Assigning Breed Origin to Alleles in Crossbred Animals." *Genetics Selection Evolution* 48 (1): 1–22. <https://doi.org/10.1186/s12711-016-0240-y>.
- Visscher, P.M., N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. 2017. "10 Years of GWAS Discovery: Biology, Function, and Translation." *The American Journal of Human Genetics* 101 (1): 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>.
- Wang, Y., J. Bennewitz, and R. Wellmann. 2017a. "Novel Optimum Contribution Selection Methods Accounting for Conflicting Objectives in Breeding Programs for Livestock Breeds with Historical Migration." *Genetics Selection Evolution* 49 (1): 1–12. <https://doi.org/10.1186/s12711-017-0320-7>.
- Wang, Y., D. Segelke, R. Emmerling, J. Bennewitz, and R. Wellmann. 2017b. "Long-Term Impact of Optimum Contribution Selection Strategies on Local Livestock Breeds with Historical Introgression Using the Example of German Angler Cattle." *G3: Genes, Genomes, Genetics*. Vol. 7. <https://doi.org/10.1534/g3.117.300272>.

Wellmann, R. 2019. "Optimum Contribution Selection for Animal Breeding and Conservation: The R Package OptiSel." *BMC Bioinformatics*. Vol. 20. <https://doi.org/10.1186/s12859-018-2450-5>.

Appendix

The BOA model has the alternative animal-model representation

$$y = X\beta + \sum_{k=1}^K u_k + e,$$

where $u_k = Z_A^k a_k$ is the N -vector with animal effects from genetic group k . The KN -vector $u = (u_1^T, \dots, u_K^T)^T$ with all animal effects has the representation

$$u = \begin{pmatrix} Z_A^1 a_1 \\ \vdots \\ Z_A^K a_K \end{pmatrix} = \tilde{Z}_A a$$

with $KN \times KM$ -matrix

$$\tilde{Z}_A = \begin{pmatrix} Z_A^1 & 0 & \dots & 0 \\ 0 & Z_A^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & Z_A^K \end{pmatrix}.$$

The covariance matrix of u is

$$\begin{aligned} G = \text{Cov}(u) &= \text{Cov}(\tilde{Z}_A a) = \tilde{Z}_A \text{Cov}(a) \tilde{Z}_A^T = \tilde{Z}_A (\Sigma \otimes I) \tilde{Z}_A^T \\ &= \begin{pmatrix} Z_A^1 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & Z_A^K \end{pmatrix} \begin{pmatrix} \sigma_{A1}^2 I & \dots & \sigma_{A1K} I \\ \vdots & & \vdots \\ \sigma_{AK1} I & \dots & \sigma_{AK}^2 I \end{pmatrix} \begin{pmatrix} Z_A^{1T} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & Z_A^{KT} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{A1}^2 Z_A^1 Z_A^{1T} & \dots & \sigma_{A1K} Z_A^1 Z_A^{KT} \\ \vdots & & \vdots \\ \sigma_{A1K} Z_A^K Z_A^{1T} & \dots & \sigma_{AK}^2 Z_A^K Z_A^{KT} \end{pmatrix} \end{aligned}$$

Thus, the mixed linear model has the alternative representation

$$y = X\beta + Zu + e$$

with $N \times KN$ -matrix $Z = (I, \dots, I)$. Henderson's mixed model equations are

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{pmatrix},$$

where $R = \sigma_e^2 I$. They can be simplified as

$$\begin{pmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \sigma_e^2 G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^T y \\ Z^T y \end{pmatrix}.$$

The alternative model representation as an SNP-model is

$$y = X\beta + Z_A a + e,$$

where $Z_A = (Z_A^1, \dots, Z_A^K)$, which provides the following alternative mixed model equation:

$$\begin{pmatrix} X^T X & X^T Z_A \\ Z_A^T X & Z_A^T Z_A + \sigma_e^2 D^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} X^T y \\ Z_A^T y \end{pmatrix},$$

where $D^{-1} = \Sigma^{-1} \otimes I$. The expanded version of this equation is

$$\begin{pmatrix} X^T X & X^T Z_A^1 & \dots & X^T Z_A^K \\ Z_A^{1T} X & Z_A^{1T} Z_A^1 + \sigma_e^2 \Sigma_{11}^{-1} I & \dots & Z_A^{1T} Z_A^K + \sigma_e^2 \Sigma_{1K}^{-1} I \\ \vdots & \vdots & \dots & \vdots \\ Z_A^{KT} X & Z_A^{KT} Z_A^1 + \sigma_e^2 \Sigma_{K1}^{-1} I & \dots & Z_A^{KT} Z_A^K + \sigma_e^2 \Sigma_{KK}^{-1} I \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{a}_1 \\ \vdots \\ \hat{a}_K \end{pmatrix} = \begin{pmatrix} X^T y \\ Z_A^{1T} y \\ \vdots \\ Z_A^{KT} y \end{pmatrix}$$

Chapter 4

Implementing a Genomic Rotational Crossbreeding Scheme to Promote Local Dairy Cattle Breeds – A Simulation Study

J. Stock,^{*1} H. Esfandyari,[†] D. Hinrichs,[‡] J. Bennewitz^{*}

^{*} Department of Animal Genetics and Breeding, University of Hohenheim, 70599 Stuttgart, Germany

[†] TYR, Storhamargata 44, 2317 Hamar, Norway

[‡] Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany

¹Joana Stock, Department of Animal Genetics and Breeding, University of Hohenheim, Garbenstraße 17, 70599 Stuttgart, Germany, +49 711 459 22479, joana.stock@uni-hohenheim.de

Published in *Journal of Dairy Science*, doi:10.3168/jds.2020-19927

Abstract

In dairy cattle breeding, there is a clear trend towards the use of only a few high-yielding breeds. One main reason is that efficient breeding programs require a certain population size. Since some numerically small breeds are well known for their functional traits, they might be an interesting crossing partner for high yielding breeds with the aim to utilize heterosis. This simulation study investigated the transition period of a small cattle population for the implementation of genomic selection and rotational crossbreeding with a high-yielding breed. Real SNP chip genotype data from the numerically small red dairy breed Angler and the high-yielding breed Holstein Friesian were used to simulate the base generations, from which rotational crossbreeding was conducted for 10 generations. A polygenic trait with many QTL with additive and directional dominance effects was simulated. Different selection methods for Angler sires (purebred performance, crossbred performance, and weighted purebred-crossbred performance) and different sizes and structures of the reference population (Angler, crossbred animals, and Angler plus crossbred animals) were considered. The results showed that the implementation of a genomic rotational crossbreeding scheme is an appealing option to promote the numerically small Angler breed. The growing reference population consisting of Angler and crossbred individuals maximized the genetic gain for Angler and the performance

level for the crossbred individuals. Selection for purebred performance, crossbred performance or a weighted combination of both hardly affected the results, and differences between selection scenarios were observed only in the long term with decreasing purebred-crossbred correlations.

Keywords: rotational crossbreeding, genomic prediction, purebred and crossbred performance, heterosis

Introduction

In dairy cattle breeding, there is a clear trend towards the use of a few high-yielding breeds to the detriment of many local breeds. This can probably best be observed in the Holstein Friesian population, which has become by far the most important dairy cattle breed in many countries during recent decades. One main reason is that efficient breeding programs require a certain population size. This holds true especially for genomic selection, where a large reference population is of central importance for the estimation of reliable genomic breeding values of young selection candidates (Meuwissen *et al.*, 2016). Once this large reference population is established, genomic selection has the potential to double the genetic gain (Schaeffer, 2006). Because a large reference population can only be established in large populations, accelerating genetic gain by genomic selection is problematic in numerically small populations. Thus, the performance gap between large high-yielding breeds and breeds with small population sizes will continue to increase. As long as the numerically small breed has found its economic niche, e.g., as a source of special breed products or for use in landscape maintenance, this might be less of a problem. However, if the numerically small breed is used to produce products that are interchangeable with products from high-yielding breeds, the increasing performance gap will inevitably reduce the economic competitiveness, accelerating the trend of the decline of numerically small breeds.

Several options have been investigated in the past to involve numerically small breeds in the genetic gain of high-yielding breeds and to increase their competitiveness. Genomic selection was attempted by using marker effects estimated in a large and related population or by building a common reference population with this large population. The results of these joint efforts were highly dependent on the genetic distance between the breeds, previous genetic admixture of the breeds, and their similarities in LD structure. They were not encouraging in most cases and thus genomic selection has not been widely adopted in numerically small breeds (Lund *et al.*, 2014).

A very efficient strategy to increase production yield is to introgress genetic material from high-yielding breeds into numerically small breeds, which was frequently performed in the past

(Zhang *et al.*, 2018). For example, Hartwig *et al.* (2014) reported the introgression of Montbeliarde and Red Holstein into Vorderwald cattle, a numerically small breed located in southern Germany. This introgression resulted in a substantial increase in milk yield and in an increase in genetic diversity within the Vorderwald breed. However, this strategy comes at the costs of the genetic uniqueness of the breed, which reduces its conservation potential and might result in the genetic extinction of the numerically small breed by replacement breeding in the long term (Wang *et al.*, 2017b; 2019).

Some local and numerically small breeds have been selected for many years for fertility, udder health, and other functional traits. Improvement of these trait complexes was one of the main reasons for conducting crossbreeding in dairy cattle breeding (Freyer *et al.*, 2008; Sørensen *et al.*, 2008). For example, Heins *et al.* (2006a, 2006b, 2006c) reported the results of a large crossbreeding experiment with Normande, Montbeliarde, and Scandinavian Red as the crossing partners of Holstein Friesian. Significant improvements in some functional traits were observed. Thus, a further option for local breeds might be their use as crossing partners for high-yielding breeds with the aim of utilizing breed complementary effects and heterosis, especially for functional traits. Naturally, rotational crossbreeding is the only suitable method for systematically crossbreeding dairy cattle due to their low female reproductive capacity and long generation interval. For the establishment of such a breeding design, several questions must be answered, for example, about the design of the reference population (van Grevenhof and van der Werf, 2015), the statistical model (reviewed in Stock *et al.*, 2020), or the selection criteria regarding purebred or crossbred performance (Esfandyari *et al.*, 2018).

The implementation of a rotational crossbreeding scheme might be accompanied by the introduction of genomic selection in the numerically small breed. Therefore, the aim of the present study was to analyze the transition period of a small breed becoming a crossbreeding partner for a large high-yielding breed while genomic selection is introduced with a initially small but continuously growing reference population size. Stochastic simulations were performed considering different types of selection criteria and reference populations. The simulation protocol was designed to mimic the Angler dairy cattle breed, a numerically small dairy breed located in the northern part of Germany, and German Holstein, a high-yielding breed. Real SNP chip genotype data from these two breeds were used to simulate the two populations. A typical polygenic quantitative trait with additive genetic and dominance variance was simulated.

Material and Methods

Simulation Protocol

A rotational crossbreeding program, including the local red dairy cattle breed Angler and the high-yielding breed German Holstein, was simulated for 10 discrete generations using the R package *xbreed* (Esfandyari and Sørensen, 2017). The base populations of both breeds were distinct and were generated using real data from 268 Angler cattle (Wang *et al.*, 2017b) and 1,935 Holstein cattle (Streit *et al.*, 2013), including 50K genotypes and pedigree information. After applying standard quality control parameters and excluding SNPs that were not available for both breeds, 23,448 SNPs in total remained for the analysis. Due to computation time, the simulated genomes consisted only of the first five chromosomes (BTA1-BTA5). Of the 5,843 SNPs located on these 5 chromosomes, 500 were randomly chosen as QTL. The QTL alleles were assumed to be the same in both breeds, which resulted in high genetic correlations across populations in the first generations. Therefore, the QTL effects were jointly simulated by sampling additive effects from a gamma distribution (0.4, 1.66) (Meuwissen *et al.*, 2001) and by calculating dominance effects by $d_i = h_i \cdot |a_i|$, with h_i as the degree of dominance, sampled from a normal distribution, $h_i = N(0.5, 1)$, leading to a directional dominance and therefore to positive heterosis. This way of simulating the QTL effects ensured that QTL with high values of additive effects were expected to have large dominance effects, as predicted by theory ((Wellmann & Bennewitz 2011) and frequently observed in real data (e.g. Bennewitz and Meuwissen (2010). Additionally, we expected crossbred offspring to show heterosis and a higher degree of heterozygosity compared to the parental breeds.

A single trait with additive variance (0.3), dominance variance (0.15) and environmental variance (0.55) was specified for both breeds following the genetic architecture found in a large-scale experiment by Bolormaa *et al.* (2015). Additive and dominance effects were scaled to reach the desired additive and dominance variances of 0.3 and 0.15, respectively. Phenotypes were generated by adding a standard normal residual to the genotypic value.

The founder generation was obtained from the real data within two generations without selection. Here, the aim was solely to increase the size to 3,000 Angler and 6,000 Holstein. From these founder generations, the first crossing was purebred Angler × purebred Holstein. In the following 9 generations, the crossbred dams were mated alternately to purebred Angler bulls and purebred Holstein bulls. Therefore, the simulated breeding scheme included purebreeding lines of both Holstein and Angler cattle (see Figure 1). The number of animals in each population was kept constant. It was assumed that each dam can have 3 offspring per generation. In each generation of purebreeding in Angler, 1,000 dams and 26 sires were selected as parents, to have a sufficient size of N_e , resulting in 3,000 (1,500 female and 1,500 male) purebred offspring. In each generation of purebreeding in Holstein, 2,000 dams and 26

sires were selected, resulting in 6,000 (3,000 female and 3,000 male) offspring. For crossbreeding, 1,000 crossbred dams and 26 purebred bulls, alternating Angler or Holstein, were selected. The Angler bulls selected for crossbreeding were the same as those selected for purebreeding. This resulted in 3,000 (1,500 female, 1,500 male) crossbred offspring per generation. Discrete generations were considered for both pure and crossbreeding schemes. Allocation of sires and dams was random.

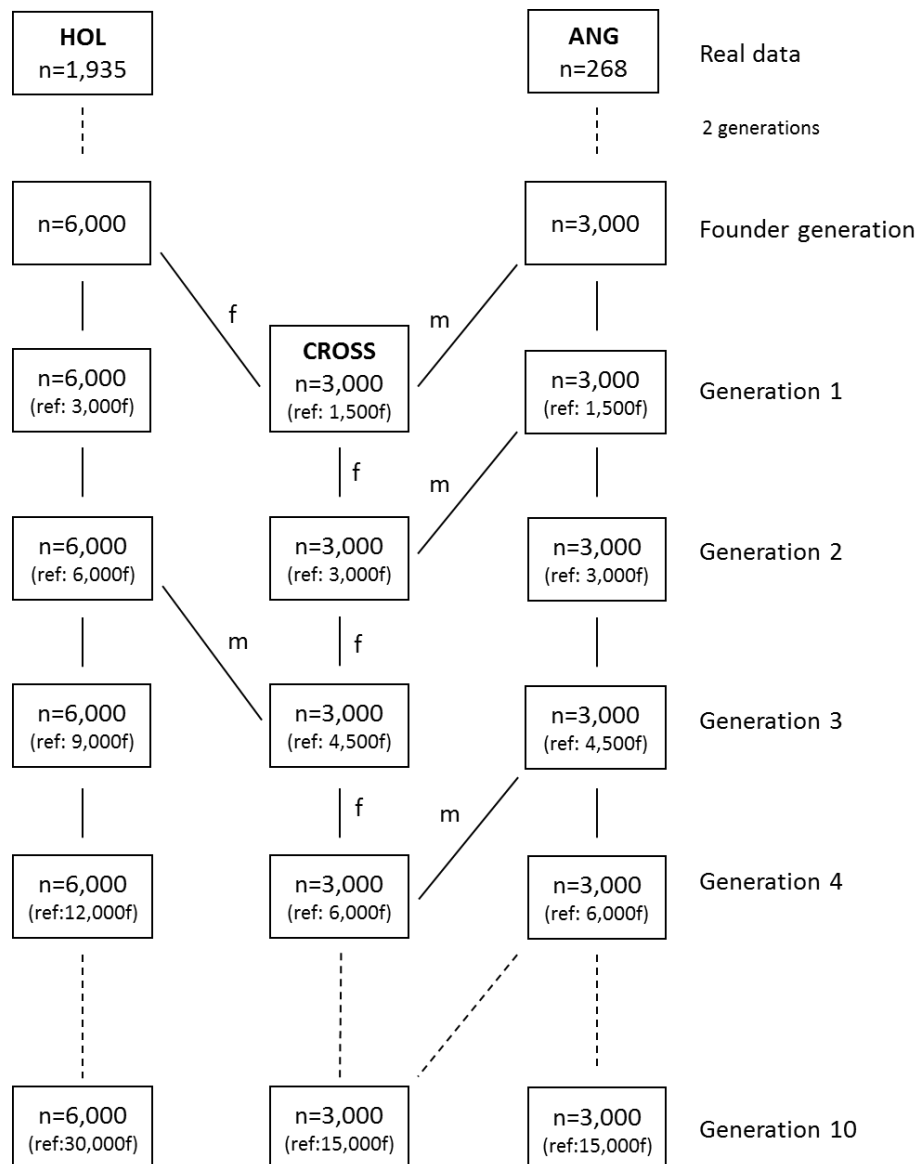


Figure 1. Rotational crossbreeding scheme. HOL = Holstein, CROSS = crossbred, ANG = Angler, ref = reference population, f = female, m = male. In all even generations Angler were used as sires for crossbreds, in odd generations Holstein were used as sires for crossbreds.

Table 1. Overview of the nine simulated scenarios

Scenario	Selection method	Reference population
1a	Genomic estimated breeding value for purebred performance (<i>GEBV_p</i>)	Angler
1b		Crossbreds
1c		Angler + Crossbreds
2a	Genomic estimated breeding value for crossbred performance (<i>GEBV_c</i>)	Angler
2b		Crossbreds
2c		Angler + Crossbreds
3a	Weighted genomic estimated breeding value (<i>GEBV_w</i>)	Angler
3b		Crossbreds
3c		Angler + Crossbreds

Genomic Evaluations

True breeding values (*TBV*) were calculated for both purebred (*TBV_p*) and crossbred performance (*TBV_c*) as the expected genotypic value of offspring in the next generation of a parent mated randomly to its own breed or to another breed, respectively. Thus, to calculate the *TBV* for purebred performance of animal *i* belonging to breed *r*,

$$TBVp_{ir} = \sum_{j=1}^{500} [(x_{ij})(p_{jr}a_j + q_{jr}d_j)] + [(1 - x_{ij})(-q_{jr}a_j + p_{jr}d_j)],$$

where x_{ij} indicates the genotype of animal *i* at QTL *j*, which can be either 1 for *AA*, 0.5 for *Aa* or 0 for *aa*. p_{jr} and q_{jr} are the allele frequencies for *A* and *a* at QTL *j* in breed *r*, respectively, and a_j and d_j are the true additive and dominance effects of the *j*th QTL, respectively.

Analogous to that, the *TBV_c* was calculated but using the allele frequencies for *A* and *a* of the other population (*r'*), which are denoted as $p_{jr'}$ and $q_{jr'}$:

$$TBVc_{ir} = \sum_{j=1}^{500} [(x_{ij})(p_{jr'}a_j + q_{jr'}d_j)] + [(1 - x_{ij})(-q_{jr'}a_j + p_{jr'}d_j)].$$

This means that for the Angler *TBV_c* the other population *r'* was Holstein in the first generation and the crossbred population to which Angler were mated in the later generations. Notably, Holstein were solely selected for purebred performance; thus, genomic values or breeding values for crossbred performance were not an issue. Calculating the genomic estimated breeding values for purebred (*GEBV_p*) and crossbred performance (*GEBV_c*) was similar, but

SNP genotypes were used instead of QTL genotypes, and estimated effects, denoted as \hat{a}_j and \hat{d}_j , were used instead of true additive and dominance effects as follows.

$$GEBV_{p_{ir}} = \sum_{j=1}^{4,843} [(x_{ij})(p_{jr}\hat{a}_j + q_{jr}\hat{d}_j)] + [(1 - x_{ij})(-q_{jr}\hat{a}_j + p_{jr}\hat{d}_j)].$$

Analogous to that, the $GEBV_c$ was calculated but again using the allele frequencies for A and a of the other breed (r'), denoted as $p_{jr'}$ and $q_{jr'}$, respectively:

$$GEBV_{c_{ir}} = \sum_{j=1}^{4,843} [(x_{ij})(p_{jr'}\hat{a}_j + q_{jr'}\hat{d}_j)] + [(1 - x_{ij})(-q_{jr'}\hat{a}_j + p_{jr'}\hat{d}_j)].$$

To predict the effects associated with each SNP (QTL were excluded), Bayesian ridge regression (BRR) implemented in the *BGLR* R package by Pérez and los Campos (2014) was the chosen method following the model

$$y_i = \mu + H_i b + \sum X_{ij} a_j + \sum Z_{ij} d_j^* + e_i,$$

where y_i is the phenotypic value of animal i in the reference population, μ is the overall mean, H_i is observed heterozygosity for individual i considered as covariate with effect b , X_{ij} indicates the genotype of animal i at SNP j , which can be either 0 for aa , 1 for Aa or 2 for AA . The random unknown additive effect for SNP j is denoted by a_j , and Z_{ij} indicates the heterozygosity of animal i at SNP j , which is either 0 if animal i is homozygous (aa/AA) or 1 if animal i is heterozygous (Aa) at marker j . The term d_j^* is the random unknown dominance effect at SNP j , and e_i is the residual effect for animal i . As directional dominance effects were simulated for the trait architecture, to account for non-zero mean for dominance effects, the individual observed heterosis was considered as a covariate (H_i) in the model as proposed by Xiang *et al.* (2016). Subsequently, dominance effects (d_j) were obtained as $d_j = d_j^* + b/N$, where N is number of SNP.

BRR includes prior assumptions of Gaussian distributed effects, with a mean of zero and residual variance of $\sigma_e^2 \sim X^{-2}(df_e, S_e)$ which is assigned to scaled-inverse chi-squared densities and indexed by a degree of freedom ($df_e = 5$) and scale ($S_e = 0.33$). The corresponding algorithm uses a Gibbs sampler. We chose 10,000 iterations and discarded the first 2,000 as burn-ins.

True genomic values (TGV) were calculated as

$$TGV_i = \sum_{j=1}^{500} (a_{ij} + (2 - x_{ij})d_{ij})x_{ij},$$

using the above defined parameters.

Genetic differentiation between Angler and Holstein in each generation was calculated by F_{ST} values, according to equation (8) of Weir and Cockerham (1984).

Scenarios

Genetic gain in Angler and their rotational crossbred progeny with Holstein was examined in 9 different scenarios (Table 1). The simulated scenarios differed in the selection criteria and structure of the reference population used for the estimation of marker effects. Each scenario was replicated 10 times. Referring to Angler, 3 different selection methods were used to select animals as sires and dams. They were either selected based on their estimated breeding value for purebred performance ($GEVP$, scenario 1), crossbred performance ($GEVC$, scenario 2) or combined weighted purebred and crossbred performance ($GEVW$, scenario 3). When selection was based on $GEVW$, the weight (w) was 50% on purebred and 50% on crossbred performance as $GEVW = (1 - w) \cdot GEVP_i + w \cdot GEVC_i$, with $w = 0.5$. In all implemented scenarios, Holstein sires and dams were selected based on their purebred performance ($GEVP$). Crossbred animals were always selected randomly to be dams.

In addition, 3 different female reference populations were used to predict the marker effects for Angler. In each generation, they accumulated and consisted of all available females, which resulted in an increasing reference population from generation to generation. Depending on the scenario, they were either Angler (scenario a), crossbreds (scenario b), or combined Angler + crossbreds (scenario c). In each scenario, purebred Holstein built the reference population for Holstein. It was assumed that Angler competed with Holstein; thus, no common reference population with Angler and Holstein was considered.

Results

Genetic Gain

The trends of the true genomic values (TGV) over the ten simulated generations and nine scenarios are shown in Figure 2 and Figure 3 for crossbreds, Angler and Holstein. The averages are presented along with standard deviations for generations five and ten in Table 4. In general, the population of crossbred animals was superior to both purebred Angler and Holstein populations. Starting with small differences in first generations, the differences within and between populations increased over time and therefore showed how the scenarios

affected the genetic gain. The genetic gain in Angler improved smoothly, whereas in crossbred animals it improved in waves depending on which breed was used as sire. Peaks occurred in even generations, when Angler were used as sires to produce the crossbred offspring.

Regarding Angler, the highest response to selection was found in scenarios 1c, 2c, and 3c, where the reference population included both Angler and crossbred animals. Scenario 1c, with selection based on $GEBV_p$, was slightly superior to 2c (selection on $GEBV_c$) and 3c (selection on $GEBV_w$). The lowest improvement over time was found in scenario 2b, where the reference population only included crossbred animals and selection was based on $GEBV_c$. The highest genetic gain in crossbred animals was in scenario c, followed by a and b, which was similar to the results in Angler. While increasing of TGV in Angler was smooth, it developed in waves in crossbred animals. The amplitudes increased over time, with the largest differences between the generations in scenario a and b and the smallest in scenario c. In general, there were very small or lack of differences between scenarios a and b. In scenario 2 (selection based on $GEBV_c$), all a, b and c were very close to each other.

Correlations between Purebred and Crossbred Performance (r_{pc})

Correlations between purebred and crossbred performance (r_{pc}) were calculated as correlations between TBV_p and TBV_c of Angler. They are shown in Figure 4. Up to generation 4, r_{pc} was > 0.98 in all scenarios, which was expected because homogeneous QTL effects were simulated and the SNPs used were informative in both breeds. The F_{ST} values were 0.01 in the first generations, showing little genetic differentiation between Angler and Holstein. It increased up to 0.08 in generation 10, with little differences between the scenarios (not shown elsewhere). In later generations, a small decrease of r_{pc} was observed, with some waves. The amplitudes of the waves increased over time. The largest decrease was found in scenario 2c, with the lowest r_{pc} in generation 9 (0.825). In general, the lowest decrease was found in scenarios 1a, 1b, and 1c with the selection criterion of Angler based on $GEBV_p$.

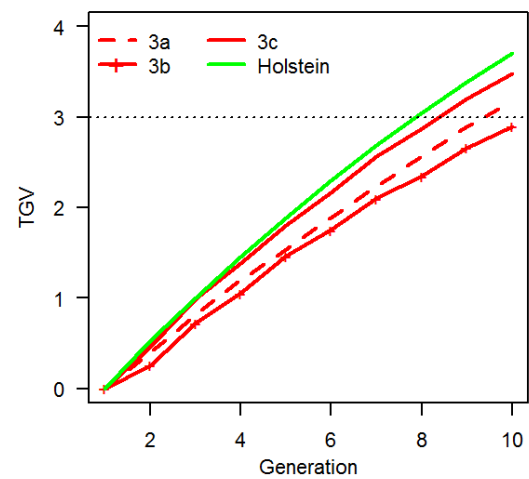
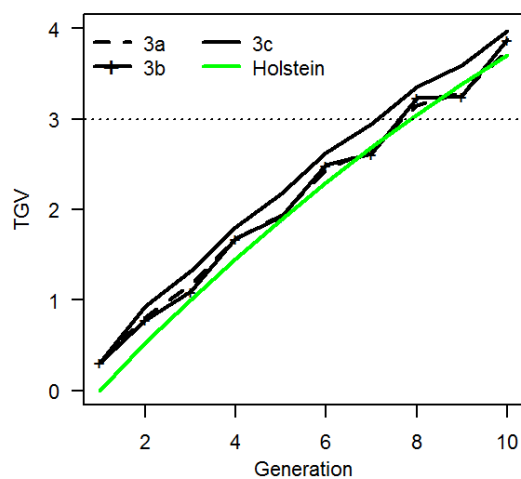
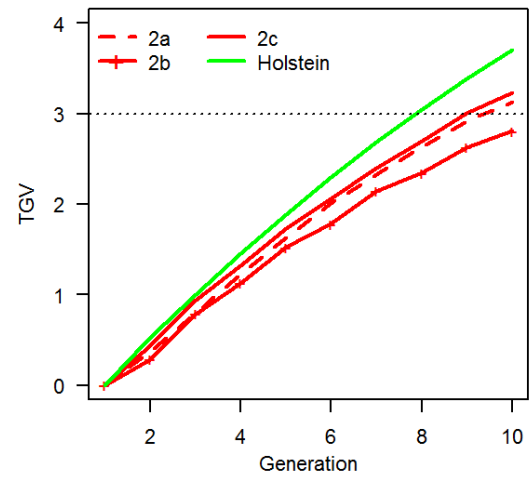
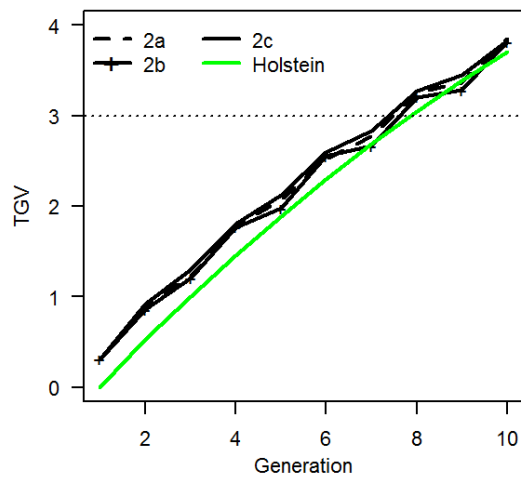
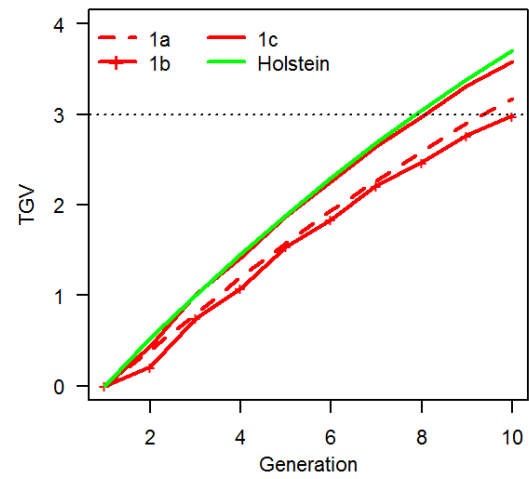
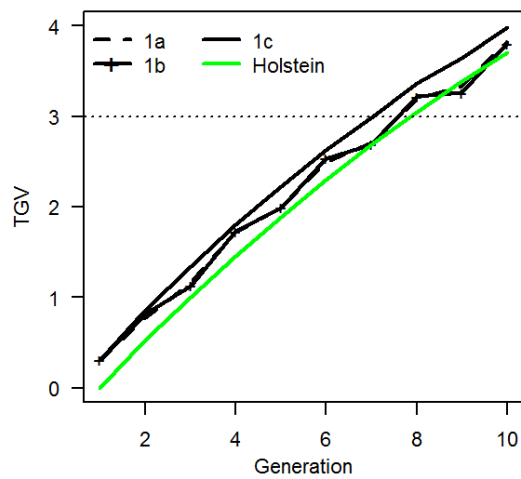


Figure 2. True genomic values (TGV) of crossbred animals and Holstein over 10 generations of simulation. Scenarios grouped by selection criterion. From top to bottom: GEBVp, GEBVc, and GEBVw. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals.

Figure 3. True genomic values (TGV) of purebred Angler and Holstein over 10 generations of simulation. Scenarios grouped by selection criterion. From top to bottom: GEBVp, GEBVc, and GEBVw. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals.

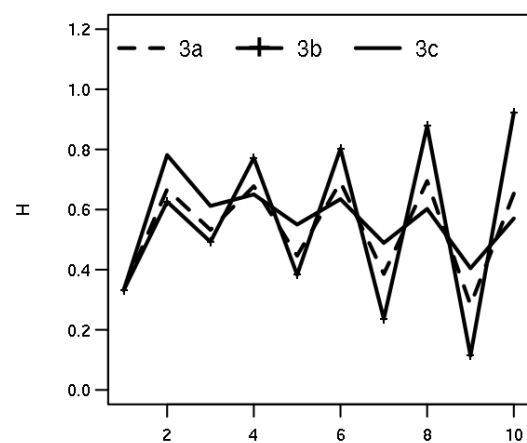
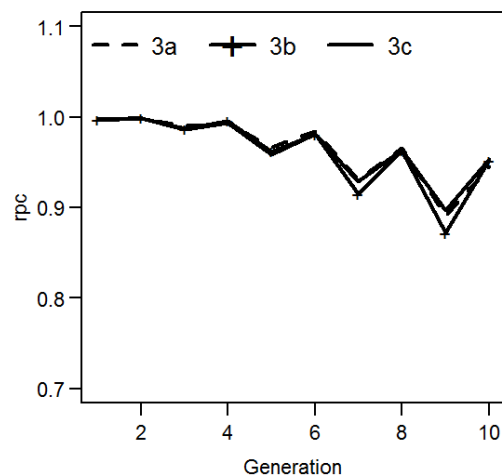
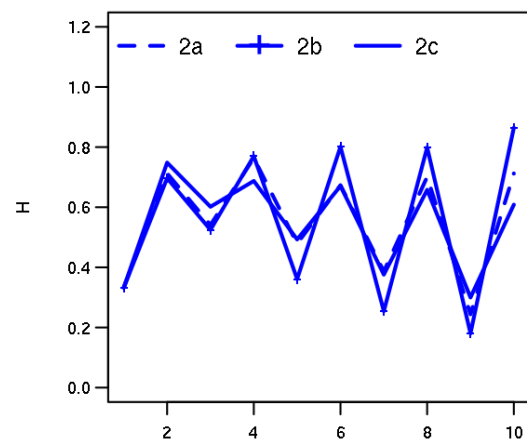
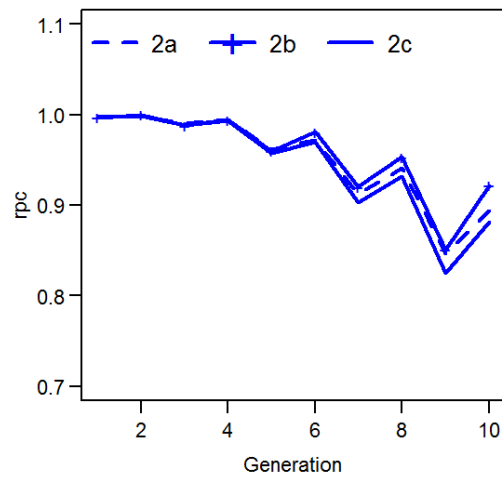
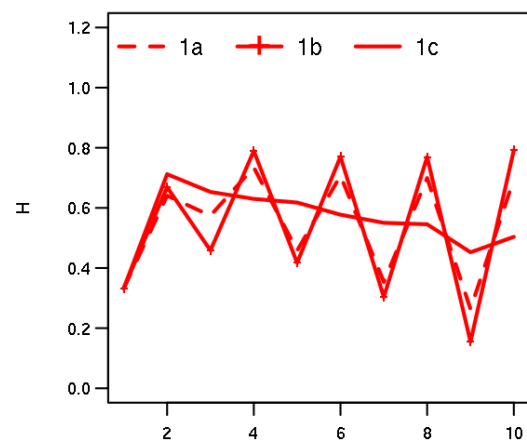
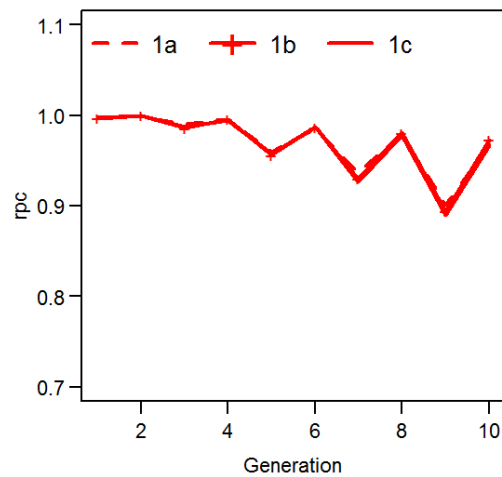


Figure 4. Correlations of purebred and crossbred performance (r_{pc}) of purebred Angler. Scenarios grouped by selection criterion. From top to bottom: *GEBV_p*, *GEBV_c*, and *GEBV_w*. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals.

Figure 5. Heterosis (H) in crossbred animals over 10 generations of simulation. Scenarios grouped by selection criterion. From top to bottom: *GEBV_p*, *GEBV_c*, and *GEBV_w*. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals.

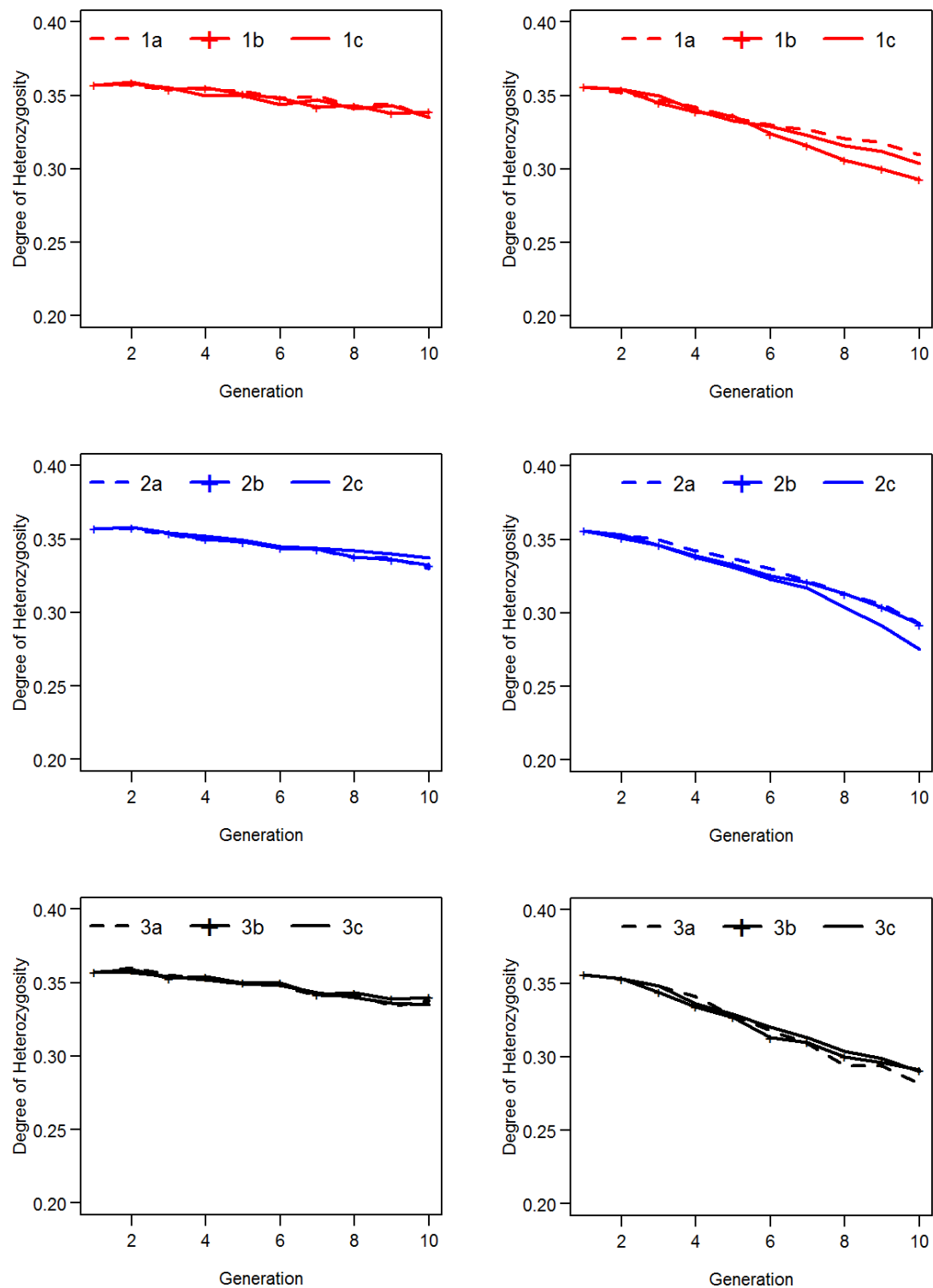


Figure 6. Degree of heterozygosity in crossbred animals (left) and purebred Angler (right) over 10 generations of simulation. Scenarios grouped by selection criterion. From top to bottom: GEBVp, GEBVc, and GEBVw. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals.

Table 4. True genomic values (SD) of Angler, crossbreds and Holstein in generations 5 and 10. All scenarios shown. Scenario 1 = selection on GEBV_p, scenario 2 = selection on GEBV_c, scenario 3 = selection on GEBV_w. a = reference population consisting of Angler, b = reference population consisting of crossbred animals, and c = reference population consisting of Angler and crossbred animals. Holstein animals are not affected by scenarios, thus results are the same in each scenario.

Scenario	Angler		Crossbreds		Holstein (not affected by scenarios)	
	Generation 5	Generation 10	Generation 5	Generation 10	Generation 5	Generation 10
1a	1.575 (0.600)	3.176 (0.493)	1.990 (0.601)	3.814 (0.487)	1.888 (0.591)	3.706 (0.477)
1b	1.541 (0.602)	2.982 (0.495)	1.989 (0.616)	3.806 (0.481)		
1c	1.878 (0.599)	3.580 (0.477)	2.221 (0.598)	3.981 (0.488)		
2a	1.631 (0.618)	3.134 (0.470)	2.072 (0.594)	3.853 (0.488)		
2b	1.522 (0.596)	2.808 (0.497)	1.977 (0.621)	3.817 (0.499)		
2c	1.733 (0.576)	3.237 (0.474)	2.126 (0.586)	3.834 (0.476)		
3a	1.541 (0.603)	3.168 (0.490)	1.939 (0.601)	3.746 (0.484)		
3b	1.464 (0.619)	2.896 (0.497)	1.916 (0.616)	3.878 (0.503)		
3c	1.805 (0.587)	3.479 (0.473)	2.172 (0.590)	3.969 (0.485)		

Heterosis and Degree of Heterozygosity

Heterosis (H) of crossbred animals was observed over 10 generations by the difference between TGV of crossbred animals (CP) and the average TGV of their parents' populations (PA), such as $H = CP - PA$. The results (Figure 5) show most differences between type of reference and between generations in scenario 3, where selection was based on $GEVW$. Within scenario 3, we found the highest amplitudes when the reference population included crossbred animals only (scenario 3b). The lowest differences were found in scenario 1c. In general, the amplitudes increased over time, with peaks when the sires were Angler and low values when the sires were Holstein.

The degree of heterozygosity in crossbred animals and Angler was calculated as the average proportion of heterozygous loci to the total number of loci. The results showed a decreasing amount of heterozygosity over time (Figure 6). In general, the decrease in crossbreds was smaller. It evolved in waves, particularly in scenarios 1c and 3b. At the same time, the decline of heterozygosity in Angler was fastest in these scenarios.

Discussion

The transition period of a numerically small cattle population for the implementation of genomic selection and rotational crossbreeding with a high-yielding breed was simulated using real SNP chip genotypes from the small Angler and large German Holstein populations. The transition period was mimicked by an increasing reference population starting with 1,500 simulated individuals in the first Angler purebred and crossbred reference populations increasing to 15,000 simulated individuals in the two reference populations after 10 generations. Simultaneously, rotational crossbreeding was introduced by a first cross of Angler \times Holstein, followed by alternating crossings of Angler \times crossbreds and Holstein \times crossbreds. The simulation of these first 10 generations revealed how to select Angler, how to establish the reference population during transition and how to proceed on the long run.

Only the first five chromosomes were considered, equivalent to approximately 1/3 of the full genome (Thomsen *et al.*, 2000). Following the scaling arguments in Meuwissen (2009), on a full-genome scale, the simulated individuals in the two reference populations at the beginning and after ten generations correspond to 4,500 and 45,000 individuals, respectively. The results from the last generations can be used to identify suitable strategies for the long term.

We investigated response to selection under different selection criteria of Angler, which was either purebred, crossbred or combined purebred-crossbred performance. Additionally, SNP effects were estimated either based on purebred Angler, crossbred animals or Angler and

crossbred animals in the reference population. In general, crossbred animals showed a slight superiority compared to purebred Angler and Holstein, regardless of the choice of reference population or selection criteria. This fact can make them, and therefore the implementation of a rotational crossbreeding scheme, attracting to breeders of both Angler and Holstein.

The highest genetic gain in Angler was achieved in scenario 1c, where the selection criterion was based on $GEBV_p$ and the reference population consisted of Angler and crossbred animals. In addition, crossbred animals performed best in this scenario, as well as their TGV had the smallest amplitudes. In contrast, in scenarios 1b, 2b and 3b Angler and crossbred animals had the least genetic gain and amplitudes of crossbred animals were highest. From a practical and economically point of view these amplitudes are undesired because they introduce another source of variability making the herd stock more unequal. The use of the combined Angler and crossbred reference population maximised the crossbred and the purebred Angler performance and reduced the waves substantially. Therefore, 1c would be preferable over all other scenarios and genotyping crossbred cows in the long run is desired to increase the size of reference population with these cows.

The superiority of crossbred animals was due to the heterosis effects, which were positive under all considered scenarios (Figure 5). The waves observed in the genetic values in crossbred animals (Figure 3) correspond to the waves of heterosis (Figure 5), as peaks occurred in the same generations and the degree of amplitudes was similar in the respective scenarios. Especially for scenario 1b, 2b and 3b (crossbred reference population), the waves increased over the generations. Intuitively, this goes against the expectation of maximum heterosis in the first cross generation and expectations from literature were not met, which assume heterosis in a two-way rotation to be highest in F1 and then to be settled at 67% of F1 over time (Sørensen *et al.*, 2008; Buckley *et al.*, 2014). This can be attributed to past gene flow from the Holstein to the Angler (Wang *et al.*, 2017b). During the simulations, no further gene flow between the populations was assumed. Consequently, the F_{ST} values were low at the beginning, but increased over time. The performance waves in the crossbred individuals peaked in those generations where Angler were used as sires (see Figure 1). In these generations, the degree of heterozygosity was higher than that in the generations in which Holstein were used as sires (left panel of Figure 6). The heterozygosity in Angler decreased, especially for the scenarios where waves were observed (right panel Figure 6).

Esfandyari *et al.* (2015) reported higher prediction accuracies of dominance effects using a crossbred instead of a purebred reference population, which was due to the higher level of heterozygosity. van Grevenhof and van der Werf (2015) studied the effect of adding crossbred individuals to a purebred reference population. If selection is for crossbred performance, the

addition of crossbred individuals results in an increase in genetic gain, but the effect of a replacement of purebred with crossbred individuals is small, especially for large r_{pc} . In our study, r_{pc} was high during the first generations and decreased only later (Figure 4). This can be explained by their small genetic distance at the beginning. Thus, no effects could be observed for the performance of crossbreds (Figure 2) when using a crossbred reference population (scenario b) instead of a purebred Angler reference population (scenario a). Nevertheless, as r_{pc} might decrease further in following generations, a crossbred reference population could be advantageous in the long term.

In this study, the combination of structure and size of the reference population had more impact on the genetic gain than the selection method, shown by overall highest values when the reference population included both Angler and crossbred animals. Especially in the first generations the selection method had almost no impact. In later generations we found slightly increasing differences between $GEVp$, $GEVc$ and $GEVw$ in terms of the Angler performance and crossbred performance, respectively. Hence, a marked differentiation of the breeding goal regarding purebred or crossbred performance might be particularly an issue in the long term, if increasing distances between parent populations and thus decreasing r_{pc} occur. In a study by Esfandyari *et al.* (2018) they showed in a simulated crossbreeding scheme for pigs that selection in purebreds for $GEVc$ increased crossbred performance only in the short term, whereas selection for $GEVp$ was beneficial even for crossbred performances in the long term. The benefit of $GEVc$ over $GEVp$ for short-term response was because heterosis was maximized. Selection for $GEVp$ drove beneficial QTL alleles faster to fixation in the pure breeds, which reduced heterosis in crossbreds, but this fixation was beneficial in the long term. As both short- and long-term responses are important, the authors suggested using $GEVw$ as a selection criterion.

It is well known that marker density matters in regard to multibreed genomic evaluations (Erbe *et al.* 2012; Lund *et al.* 2014). A high marker density is also desired for the application of more advanced genomic models that, e.g., account for the breed origin of alleles or for the modeling of dominance (Ibáñez-Escriche *et al.* 2009; Wellmann & Bennewitz 2012; Christensen *et al.* 2014; Sevillano *et al.* 2017). On the other hand, it is unlikely that in a transition period the individuals of a small breed will be genotyped with dense SNP chips, and most high-yielding breeds are genotyped with a standard medium density SNP chip. Therefore, we relied on 50K SNP chip data during the simulation and applied a rather simple genomic model. To analyze the effect of a higher marker density, we additionally ran the genomic model for some scenarios without excluding the true QTL from the SNP panel. The results did not alter the general picture (results not shown); thus, for the model applied, the SNP density appeared to be sufficient.

The results showed that the implementation of the genomic rotational crossbred scheme is a twofold interesting option for the Angler population. First, the reference population can be augmented with individuals from the crossbred population; second, the population can be bred to become an interesting crossing partner for the high-yielding breed Holstein. However, the limits of the simulation scheme must be acknowledged. One trait with equal performance in both breeds was simulated. It is well known that average milk production performance is much higher in Holstein; on the other hand, Angler are well known for their functional traits. Hence, extending this study towards multiple traits and trait-specific population parameters (e.g., breed means) and genetic parameters (i.e., heritability, dominance variance) is needed to make inferences for the total merit index. In addition, it was assumed that the establishment of the crossbred population did not come at the cost of the number of Angler individuals. Translated into practice, this means that Holstein breeders must start crossbreeding, whereas at least most of the Angler breeders should continue purebreeding.

Additionally, to further increase heterosis, the crossbreeding scheme could be extended towards three breeds. Recently, Shonka-Martin *et al.* (2019a; 2019b) showed that crossbred animals of Montbeliarde × Viking Red × Holstein were superior to purebred Holstein regarding feed efficiency, fertility and longevity; thus, it might be interesting to evaluate the potential of such a three-way rotational crossbreeding scheme including Angler.

Conclusions

The results showed that the implementation of a genomic rotational crossbreeding scheme could be an interesting option to promote the numerically small Angler breed. The growing reference population consisting of Angler and crossbred individuals maximized the genetic gain for Angler and the performance level for the crossbred individuals. Selection for purebred performance, crossbred performance or a weighted combination of both hardly affected the results, and the effects were observed only in the long term with decreasing purebred-crossbred correlations.

Acknowledgments

This work was financially supported by the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2817ERA10D. The project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 696231 - ReDiverse (Biodiversity within and between European Red Dairy Breeds). Joana Stock was partly

supported by the H. Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged. The manuscript has benefitted from the critical remarks of two anonymous reviewers and of the editor.

References

- Bennewitz J. & Meuwissen T.H.E. (2010) The distribution of QTL additive and dominance effects in porcine F2 crosses. *Journal of Animal Breeding and Genetics*, 127, 171–9.
- Bolormaa S., Pryce J.E. & Zhang Y. *et al.* (2015) Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. *Genetics, Selection, Evolution: GSE*, 47, 26.
- Buckley F., Lopez-Villalobos N. & Heins B.J. (2014) Crossbreeding: implications for dairy cow fertility and survival. *Animal: an international journal of animal bioscience*, 8 Suppl 1, 122–33.
- Esfandyari H., Berg P. & Sørensen A.C. (2018) Balanced selection on purebred and crossbred performance increases gain in crossbreds. *Genetics, Selection, Evolution: GSE*, 50, 8.
- Esfandyari H. & Sørensen A.C. (2017) xbreed: an R package for genomic simulation of purebreds and crossbreds.: [R package xbreed version 1.0.1.1]. *68th Annual Meeting of the European Federation of Animal Science. Tallinn, Estonia.*
- Esfandyari H., Sørensen A.C. & Bijma P. (2015) A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genetics, Selection, Evolution: GSE*, 47, 76.
- Freyer G., König S., Fischer B., Bergfeld U. & Cassell B.G. (2008) Invited review: crossbreeding in dairy cattle from a German perspective of the past and today. *Journal of Dairy Science*, 91, 3725–43.
- Hartwig S., Wellmann R., Hamann H. & Bennewitz J. (2014) The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. *Journal of Animal Breeding and Genetics*, 131, 496–503.
- Heins B.J., Hansen L.B. & Seykora A.J. (2006a) Calving difficulty and stillbirths of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *Journal of Dairy Science*, 89, 2805–10.
- Heins B.J., Hansen L.B. & Seykora A.J. (2006b) Fertility and survival of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *Journal of Dairy Science*, 89, 4944–51.

-
- Heins B.J., Hansen L.B. & Seykora A.J. (2006c) Production of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *Journal of Dairy Science*, 89, 2799–804.
- Lund M.S., Su G., Janss L., Guldbbrandtsen B. & Brøndum R.F. (2014) Genomic evaluation of cattle in a multi-breed context. *Livestock Science*, 166, 101–10.
- Meuwissen T.H.E. (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics, Selection, Evolution: GSE*, 41, 35.
- Meuwissen T.H.E., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819–29.
- Meuwissen T.H.E., Hayes B.J. & Goddard M.E. (2016) Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers*, 6, 6–14.
- Pérez P. & los Campos G. de (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198, 483–95.
- Schaeffer L.R. (2006) Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123, 218–23.
- Shonka-Martin B.N., Hazel A.R., Heins B.J. & Hansen L.B. (2019a) Three-breed rotational crossbreds of Montbéliarde, Viking Red, and Holstein compared with Holstein cows for dry matter intake, body traits, and production. *Journal of Dairy Science*, 102, 871–82.
- Shonka-Martin B.N., Heins B.J. & Hansen L.B. (2019b) Three-breed rotational crossbreds of Montbéliarde, Viking Red, and Holstein compared with Holstein cows for feed efficiency, income over feed cost, and residual feed intake. *Journal of Dairy Science*, 102, 3661–73.
- Sørensen M.K., Norberg E., Pedersen J. & Christensen L.G. (2008) Invited review: crossbreeding in dairy cattle: a Danish perspective. *Journal of Dairy Science*, 91, 4116–28.
- Stock J., Bennewitz J., Hinrichs D. & Wellmann R. (2020) A review of genomic models for the analysis of livestock crossbred data. *Frontiers in Genetics*, 11, 568.
- Streit M., Wellmann R., Reinhardt F., Thaller G., Piepho H.-P. & Bennewitz J. (2013) Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle. *G3 (Bethesda, Md.)*, 3, 1085–93.
- Thomsen H., Reinsch N. & Xu N. *et al.* (2000) A male bovine linkage map for the ADR granddaughter design. *Journal of Animal Breeding and Genetics*, 117, 289–306.

-
- van Grevenhof I.E.M. & van der Werf J.H.J. (2015) Design of reference populations for genomic selection in crossbreeding programs. *Genetics, Selection, Evolution: GSE*, 47, 14.
- Wang Y., Bennewitz J. & Wellmann R. (2019) Managing genomes of conserved livestock breeds with historical introgression to decrease genetic overlap with other breeds. *Journal of Animal Breeding and Genetics*, 136, 505–17.
- Wang Y., Segelke D., Emmerling R., Bennewitz J. & Wellmann R. (2017) Long-term impact of Optimum Contribution Selection strategies on local livestock breeds with historical introgression using the example of German Angler cattle. *G3 (Bethesda, Md.)*, 7, 4009–18.
- Weir B.S. & Cockerham C.C. (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution; international journal of organic evolution*, 38, 1358.
- Wellmann R. & Bennewitz J. (2011) The contribution of dominance to the understanding of quantitative genetic variation. *Genetics Research*, 93, 139–54.
- Wellmann R. & Bennewitz J. (2012) Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genetics Research*, 94, 21–37.
- Xiang T., Nielsen B., Su G., Legarra A. & Christensen O.F. (2016) Application of single-step genomic evaluation for crossbred performance in pig. *Journal of Animal Science*, 94, 936–48.
- Zhang Q., Calus M.P.L., Bosse M., Sahana G., Lund M.S. & Guldbrandtsen B. (2018) Human-mediated introgression of haplotypes in a modern dairy cattle breed. *Genetics*, 209, 1305–17.

Chapter 5

Genomic Rotational Crossbreeding with Advanced Optimum Contribution Selection Methods Applied to Simulated German Angler and German Holstein Dairy Cattle Populations

J. Stock¹, H. Esfandyari², D. Hinrichs³, R. Wellmann¹, J. Bennewitz¹

¹ Institute of Animal Genetics and Breeding, University of Hohenheim, 70599 Stuttgart, Germany

² TYR, 2317 Hamar, Norway

³ Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany

In Revision

Abstract

Many local dairy cattle breeds are facing genetic extinction due to a large proportion of foreign genes which have been introgressed in the past. In addition, the performance gap to popular high-yielding breeds is increasing, resulting in a risk of numeric extinction. In the present simulation study, a genomic rotational crossbreeding scheme with the high yielding German Holstein breed and the numerically small German Angler breed was analysed with the aim to utilize heterosis effects in the crossbred animals, and simultaneously to control inbreeding, and to reduce the amount of Holstein migrant contribution observed in the Angler breed. Different scenarios of implementing OCS methods for Angler individuals were evaluated, which differed in their restrictions regarding conventional and native kinship, as well as the amount of migrant contributions from German Holstein. The results showed that rotational crossbreeding can result in superior crossbred offspring compared to the purebred parental lines, while simultaneously restricting the increase of inbreeding and reducing migrant contributions with OCS methods.

However, reducing the amount of migrant contributions in Angler turned out to be the most compromising restriction with regard to genetic gain, and resulted in waves of heterosis effects

and genomic values observed in the crossbred individuals over the generations. Further, the reduction of migrant contribution did not increase heterosis effects in the crossbred offspring, and, thus did not result in superior crossbred offspring in general.

Keywords: Rotational crossbreeding, advanced Optimum Contribution Selection, local dairy cattle breeds

Introduction

The landscape of dairy cattle breeds is characterized by only few numerically large high-yielding breeds and several numerically small local breeds. Since genetic progress can be easier achieved in large breeds, especially using genomics selection, the performance gap between high yielding breeds and local breeds has increased further over the last decades. This has reduced the economic competitiveness of local breeds and, in many cases, resulted in a decline of their population sizes and in a threat by numerical extinction.

In order to improve the performance of numerically small breeds, high-yielding breeds were frequently introgressed in the past, sometimes to a high extent. For example, the introgression of mainly Montbéliarde and Red Holstein reduced the proportion of native genes in the German Vorderwald breed to 40% (Hartwig *et al.*, 2014). Comparable amounts of introgressed genes were reported for the German Angler and the Red and White Dual Purpose breed (Schmidtman *et al.*, 2021). Thus, although this introgression has improved performance levels (Hartwig *et al.*, 2014), it has revealed another possibility of local breed extinction, namely extinction of the native gene pool or genetic extinction (Wellmann and Bennewitz, 2019).

Amador *et al.* (2011) reported that even a small amount of introgressed material can reach a high level in the breed within several generations and that removing this is a challenging task. In order to act against the threat of genetic extinction, the proportion of native genes has to be increased in the breed. The method of advanced Optimum Contribution Selection (aOCS) was developed for this purpose (Wellmann, 2019). This selection method can be used for selecting dams and sires as parents of the next generation in order to decrease the level of migrant contributions and thus recover the native parts of a breed, while simultaneously achieving genetic gain and controlling the rate of inbreeding (Wang *et al.*, 2017a). However, the costs for the additional constraint of reducing introgressed genes in the objective function of aOCS is a reduced genetic gain, compared to classical OCS without this extra constraint (Wang *et al.*, 2017b).

Many local breeds are well-known for their superiority in functional traits and robustness compared to high-yielding breeds, which makes them interesting partners in systematic crossbreeding schemes with high-yielding breeds (Buckley *et al.*, 2014). The reasons for these crossbreeding schemes are to use breed complementary effects and heterosis. For example, heterosis effects of 4.4 % were observed for milk and fat yield in a cross of Danish Jersey and Holstein cattle (Kargo *et al.*, 2021). In the dominance model, the amount of heterosis is maximised when the parental breeds are divergent fixated at the QTN, dominance variance is substantial, and the dominance effects of heterozygote genotypes are on average positive (Falconer and Mackay, 1996). Thus, if a crossbreeding scheme is considered with a numerically small breed and a high-yielding breed which was introgressed recently in the small breed, it can be expected that the application of aOCS will increase heterosis effects due to the removal of introgressed genes.

In a recent study (Stock *et al.*, 2021) we showed how a genomic rotational crossbreeding scheme could be established for a small local breed and a high yielding breed at the example of simulated Angler and Holstein populations. However, due to the high level of similarity of both breeds due to recent introgression of Holstein genes into the Angler (Wang *et al.* 2017b, Schmidtman *et al.* 2021), the level of heterosis was limited. The aim of the present simulation study was to analyse rotational crossbreeding schemes at the example of these two breeds with the application of aOCS in the Angler breed. Several and conflicting objectives were considered, i.e. the increase of native gene frequencies in the Angler, controlling inbreeding, genetic gain in the Angler for purebred and for crossbred performance, and maximisation of crossbred offspring performance.

Material and Methods

A two-way rotational crossbreeding scheme for the numerically small local dairy breed Angler and the high-yielding German Holstein breed was simulated for 8 discrete generations, using the R package *xbreed* (Esfandyari and Sørensen, 2017). The genotypes (Illumina BovineSNP50 BeadChip, 50k) and pedigree data from 268 Angler cattle (Wang *et al.*, 2017a) and 1935 German Holstein (Streit *et al.*, 2013) were used as input to create the founder generations of 3000 Angler and 6000 Holstein animals. The founder generations were obtained from the real data within one generation without selection, solely aiming to increase the size of animals to 3000 and 6000, respectively. The first crossing was purebred Angler x purebred Holstein, followed by 7 generations of the rotational crossbreeding scheme, in which Angler and Holstein bulls were mated alternately to crossbred dams. Thus, purebred lines of Angler and Holstein were generated, as well (see Figure 1, crossbreeding scheme). The

number of animals in each population was kept constant, either 3000 or 6000, to mimic a large Holstein population and a small Angler population. Allocation of sires and dams was always random. One offspring was generated per mating and the overall sex-ratio of offspring was 50% female and 50% male.

From the SNPs that were segregating in both breeds, 1000 were randomly chosen as QTL. The QTL effects were jointly simulated for both breeds by sampling additive effects from a gamma distribution $\Gamma(0.4, 1.66)$ (Meuwissen *et al.*, 2001) and by calculating dominance effects as $d_i = h_i \cdot |a_i|$, where the dominance degree h_i was sampled from the normal distribution, $N(0.5, 1)$. This led to a directional dominance and therefore to positive heterosis. This way of simulating the QTL effects ensured that QTL with high values of additive effects were expected to have large dominance effects, as predicted by theory (Wellmann and Bennewitz, 2011). A single trait with additive variance (0.3), dominance variance (0.15), and phenotypic variance (1) was specified for both breeds. Additive and dominance effects were scaled to reach the desired variances, respectively. The simulated dominance variance was chosen to be large for demonstration purpose. Nevertheless, they are within the range as estimated by Bolormaa *et al.*, (2015) for a wide range of traits in cattle.

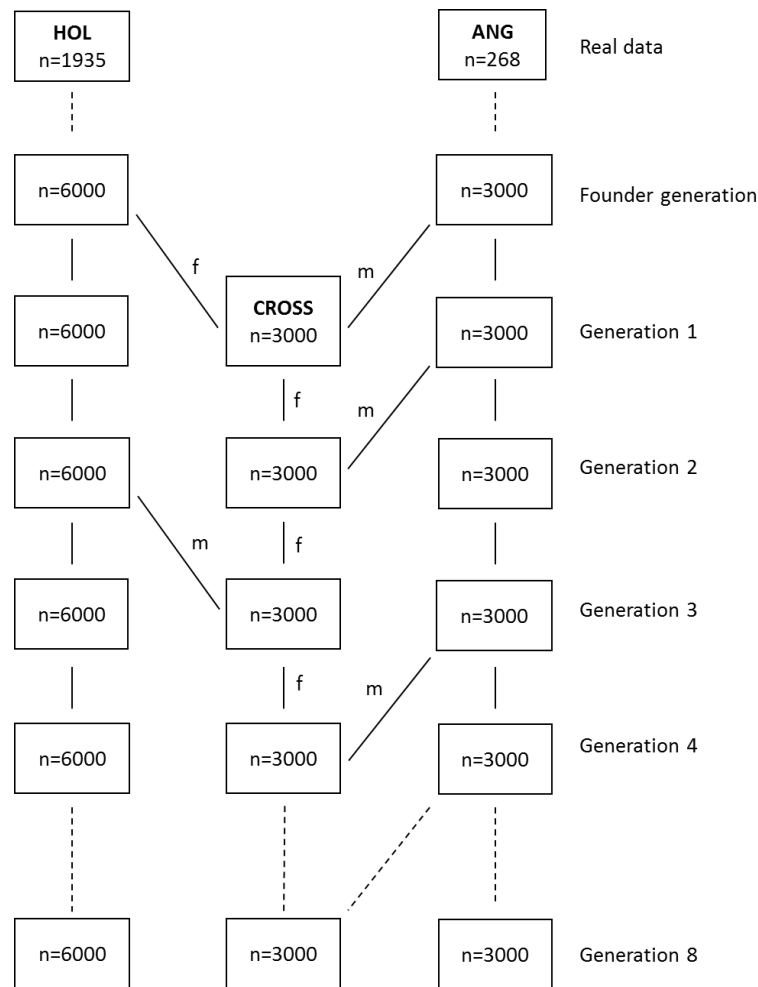


Figure 1. Rotational crossbreeding scheme. HOL = Holstein, CROSS = crossbred, ANG = Angler, f = female, m = male. In all even generations Angler were used as sires for crossbreds, in odd generations Holstein were used as sires

Scenarios

Genetic gain, mean kinship, native kinship and migrant contributions in Angler and their rotational crossbred progeny with Holstein were examined in 3 different scenarios. The simulated scenarios differed in the selection criteria of Angler. Selection of individuals to become parents of the next generation was based on classical OCS or aOCS methods as implemented in the R package *optiSel* (Wellmann 2019). The classical OCS scenario maximized genetic gain and restricted the increase of mean kinship in the population. The aOCS scenario additionally restricted the increase of the native kinship and the amount of migrant contributions (MC) in the population. The kinship was computed as the segment-based

kinship f_{SEG} from marker data. The kinship between two individuals equals the probability that two alleles randomly taken from two individuals belong to identical segments (Cara *et al.*, 2013). We considered segments consisting of at least 20 consecutive markers and a length of at least 1.5Mb. Native segments in Angler were determined by their frequencies in Holstein. If a segment containing the allele had a frequency < 0.01 in Holstein, then it was classified as native in Angler. The migrant contribution (MC) of an Angler individual was calculated as the proportion of its genome classified to be native in Holstein. Hence, the migrant contribution calculated in this paper contains contributions from Holstein, but not contributions from other introgressed breeds. The native kinship $f_{SEG|N}$ in Angler equals the conditional probability that two alleles randomly taken from two individuals belong to identical segments, given that they are native (Wellmann *et al.*, 2012). The R package *optiSel* (Wellmann, 2019) was also used for the identification of the origin of alleles, computation of MC, f_{SEG} and $f_{SEG|N}$. Exact mathematical formulas for these calculations can be found in Wellmann and Bennewitz (2019).

The overall goal of all scenarios was to maximize genetic gain in the Angler breed selected for a combined purebred and crossbred performance. The considered scenarios are summarized in Table 1.

In Scenario 1 the upper bound for the mean kinship in generation $t + 1$ was calculated as

$$ub.f_{SEG_{t+1}} = \mu_{f_{SEG_t}} + (1 - \mu_{f_{SEG_t}}) \Delta F$$

with ΔF being the desired rate of inbreeding, calculated as $\Delta F = 0.5N_e$ (Falconer and Mackay, 1996) and a desired N_e of 100. The term $\mu_{f_{SEG_t}}$ is the mean kinship in the population in generation t . In Scenario 2 and 3, the upper bound for the native kinship was calculated by replacing f_{SEG_t} with $f_{SEG|N_t}$ in the above equation. The upper bound for MC was either set as the mean MC from the founder generation (Scenario 2) or decreased by 1% each generation (Scenario 3). Angler sires selected for crossbreeding were the same as those selected for purebreeding. Selection of Holstein was always based on classical OCS and EBV for purebred performance. Selection of crossbred dams was always random. Based on the optimization goal and respective constraints, the optimum genetic contributions of individuals were calculated and used for the selection of sires and dams to breed the next generation. Each of the 3 different Scenarios was replicated 10 times.

Table 1. Overview of simulated scenarios.

Scenario	Selection method of Angler (optimization goal)	Constraints
1	OCS (EBV_w)	$ub.f_{SEG}$
2	aOCS (EBV_w)	$ub.f_{SEG N}$, $ub.MC$ (as in founder generation)
3	aOCS (EBV_w)	$ub.f_{SEG N}$, decreasing $ub.MC$ (-1% each generation)

EBV_w = combined weighted estimated breeding values for purebred and crossbred performance, $ub.f_{SEG}$ = upper bound of segment-based kinship, $ub.f_{SEG|N}$ = upper bound of segment based native kinship, $ub.MC$ = upper bound of migrant contribution.

Genomic evaluations

True breeding values (TBV) were calculated for both purebred (TBV_p) and crossbred performance (TBV_c) as the expected genotypic value of offspring in the next generation of a parent mated randomly to its own breed or to another breed, respectively. Thus, to calculate the TBV for purebred performance of animal i belonging to breed r ,

$$TBV_{p_{ir}} = \sum_{j=1}^{1000} [(x_{ij})(p_{jr}a_j + q_{jr}d_j)] + [(1 - x_{ij})(-q_{jr}a_j + p_{jr}d_j)],$$

where x_{ij} indicates the genotype of animal i at QTL j , which can be either 1 for AA , 0.5 for Aa or 0 for aa , p_{jr} and q_{jr} are the allele frequencies for alleles A and a at QTL j in breed r , respectively, and a_j and d_j are the true additive and dominance effects of the j th QTL, respectively.

The true breeding value for crossbred performance (TBV_c) was calculated analogously, but using the frequencies for the alleles A and a of the other population (r'), which are denoted as $p_{jr'}$ and $q_{jr'}$:

$$TBV_{c_{ir}} = \sum_{j=1}^{1000} [(x_{ij})(p_{jr'}a_j + q_{jr'}d_j)] + [(1 - x_{ij})(-q_{jr'}a_j + p_{jr'}d_j)].$$

This means that for the Angler $TBVc$ the other population (r') was Holstein in the first generation and the crossbred population in the following generations.

Estimated breeding values for purebred performance ($EBVp$), with reliability (r^2) of 0.75, were simulated according to Wang et al. (2017b):

$$EBVp_i = \mu_{TBVp} + r^2(TBVp_i - \mu_{TBVp}) + E_i,$$

where μ_{TBVp} represents the mean of the true breeding values of the corresponding generation and E_i is the residual term sampled from a normal distribution with mean 0 and variance $\sigma_E^2 = r^2(1 - r^2)\sigma_{TBVp}^2$. Accordingly, estimated breeding values for crossbred performance ($EBVc$) were simulated with the same r^2 of 0.75, using $TBVc$ rather than $TBVp$.

As shown by previous studies (i.e. Esfandyari *et al.*, 2018) selection for crossbred performance meets short term goals in crossbreeding, whereas selection for purebred performance showed to be advantageous in crossbreeding in the long term. In this study a combination of purebred and crossbred performance as the breeding goal was considered, to fit both short- and long-term demands. Thus, weighted EBV ($EBVw$) were used, which was a weighted combination of $EBVp$ (50%) and $EBVc$ (50%), such as

$$EBVw_i = (1 - w) \cdot EBVp_i + w \cdot EBVc_i, \text{ with } w = 0.5.$$

True genomic values (TGV) were calculated as

$$TGV_i = \sum_{j=1}^{1000} (a_{ij} + (2 - x'_{ij})d_{ij})x'_{ij},$$

with $x'_{ij} = 0, 1$, or 2 , depending on the respective genotype.

The amount of heterosis (H) in crossbred animals was calculated as the difference between the crossbreds' TGV (C) and the average TGV of the parent's populations (PA), such as $H = C - PA$. PA in generation n was calculated depending on the actual breed composition of the crossbred dam, e.g. for generation $n = 4$ it was

$$PA_n = \frac{1}{2}TGVsire_{n-1} + \frac{1}{4}TGVsire_{n-2} + \frac{1}{8}TGVsire_{n-3} + \frac{1}{16}(TGVf_s + TGVf_d),$$

where $TGVsire_{n-1}$ is the mean TGV of the sire breed in generation $n - 1$, which is Holstein, in $n - 2$ it is Angler, and in $n - 3$ it is Holstein. $TGVf_s$ and $TGVf_d$ are the mean TGV of the sires and dams from the founder generation, respectively. Thus, the general representation of PA for all generations is

$$PA_n = \sum_{k=1}^{n-1} \left(\frac{1}{2^k} TGV_{sire_{n-k}} \right) + \frac{1}{2^n} (TGV_{f_s} + TGV_{f_D}) \quad (1)$$

Results

The genetic gain (shown by TGV) of Angler, Holstein and crossbred animals in 3 different scenarios are presented in Figure 2 and Table 2. Angler showed the highest genetic gain in Scenario 1, in which classical OCS was applied, closely followed by Scenario 2, where aOCS with a constant threshold of migrant contribution was used. Scenario 3 (aOCS with decreasing threshold of migrant contribution) resulted in the slowest genetic gain in Angler. TGV of Holstein was superior to the Angler TGV in all cases.

Crossbred animals outperformed Angler and Holstein, regardless of the scenario, up to generation 4. They performed best in Scenario 1 and 2, whereas Scenario 3 they showed the lowest improvement. Their genetic gain evolved in waves, with increasing amplitudes, particularly in Scenario 3. The peaks occurred when Holstein were used as sires. Crossbred animals outperformed Angler and showed a slight superiority to Holstein in the first generations. However, in the following generations their performance improved not as fast and thus in Scenario 1 and 2 it was similar to Holstein. In Scenario 3 crossbred animals showed a lower genetic gain than Holstein from generation 5 onwards.

Table 2. True genomic values (TGV) of Angler, Holstein and crossbred animals in generation 4 and 8.

	Angler				Crossbred Animals				Holstein (not affected by scenarios)			
	Gen 4		Gen 8		Gen 4		Gen 8		Gen 4		Gen 8	
	TGV	SD	TGV	SD	TGV	SD	TGV	SD	TGV	SD	TGV	SD
Scenario 1	1.899	1.012	3.928	0.949	2.356	1.004	4.426	0.939	2.136	1.005	4.381	0.941
Scenario 2	1.847	1.005	3.867	0.951	2.417	1.008	4.458	0.952				
Scenario 3	1.458	1.041	2.422	1.049	2.337	1.018	4.111	0.967				

The results over the time period of 8 generations (Figure 4) showed a general decrease of heterosis, while some notable differences between the generations could be observed in Scenario 3. Starting with a high amount of heterosis from the initial crossing of Angler x Holstein, it dropped after generation 2, before it increased again in generation 4. In the following generations heterosis dropped and rose alternately with increasing amplitudes in Scenario 3, while the amplitudes decreased in Scenario 1 and 2. Again, peaks were found in the generations where Holstein were used as sires for the crossbred population. The additive genetic variance (V_a), estimated as the variance of the TBV, showed remarkable differences between Scenario 3 and the remaining scenarios in Angler (Figure 6). Within 8 generations it decreased only 20% in Scenario 3, whereas it decreased about 60% in Scenario 1 and 2.

The amount of migrant contribution (Figure 5) from Holstein in Angler decreased only marginally in Scenarios 1 and 2, but notably in Scenario 3, as expected. Beginning with 17.5% MC in each scenario in generation 1, in generation 8 we found 15% MC in Scenario 1, in which it was not constrained, and 10.7% MC in Scenario 3, in which the amount of MC was constrained to decrease 1% each generation. The mean kinship and the native kinship (Figure 3) in the Angler population increased in a similar pattern over the 8 generations of simulation. For both kinship measures, Scenario 1 showed the fastest increase. In Scenario 3 the mean kinship was increasing the least. Scenarios 2 and 3 showed the same increase of the native kinship.

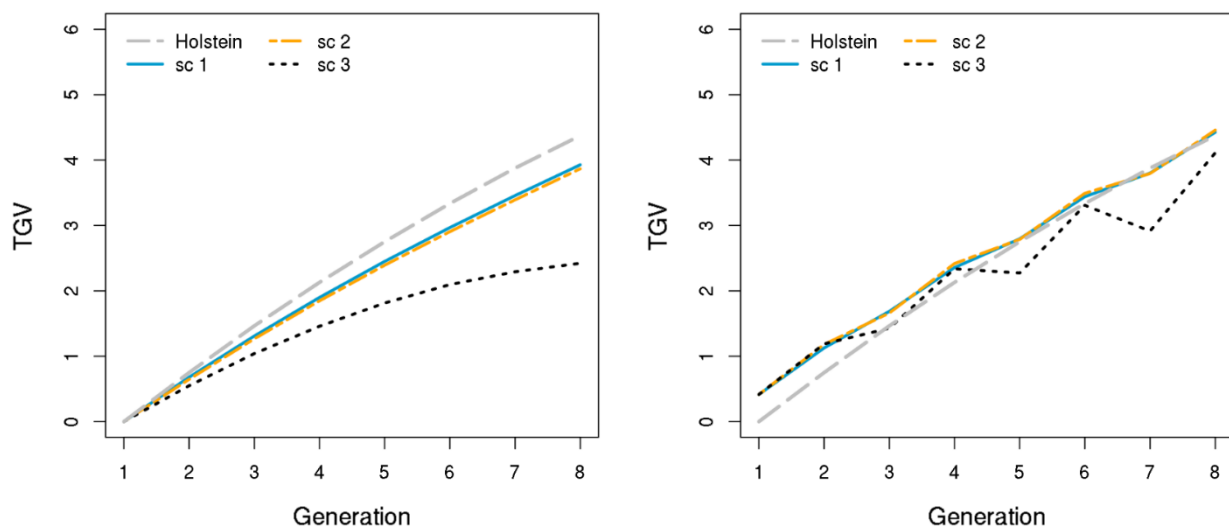


Figure 2. True genomic values (TGV) of Angler (left), Holstein and crossbred animals (right) over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1 (OCS; $ub.f_{SEG}$), sc 2 = Scenario 2 (aOCS; $ub.f_{SEG|N}$, $ub.MC$ as in founder generation), sc 3 = Scenario 3 (aOCS; $ub.f_{SEG|N}$, decreasing $ub.MC$ -1% each generation)

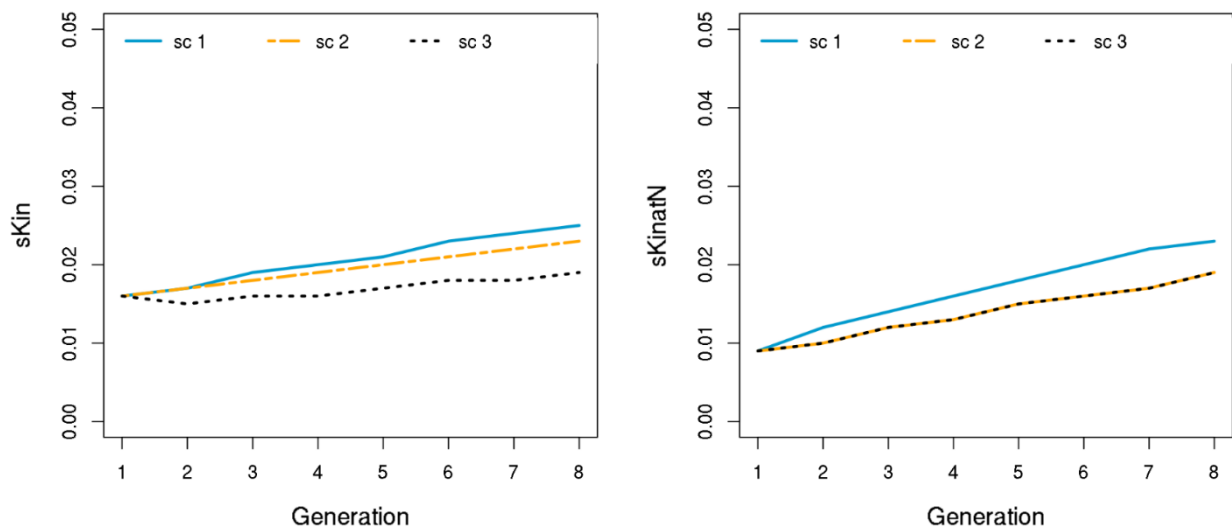


Figure 3. Amount of mean kinship (sKin) and native kinship (sKinatN) in Angler over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1 (OCS; $ub.f_{SEG}$), sc 2 = Scenario 2 (aOCS; $ub.f_{SEG|N}$, $ub.MC$ as in founder generation), sc 3 = Scenario 3 (aOCS; $ub.f_{SEG|N}$, decreasing $ub.MC$ -1% each generation)

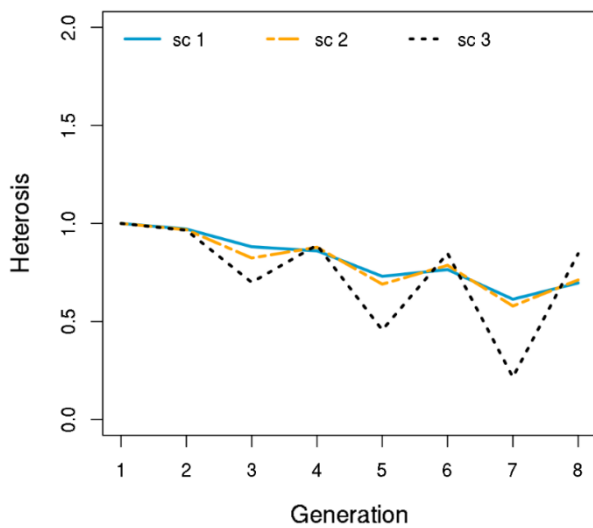


Figure 4. Heterosis in crossbred animals over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1 (OCS; $ub.f_{SEG}$), sc 2 = Scenario 2 (aOCS; $ub.f_{SEG|N}$, $ub.MC$ as in founder generation), sc 3 = Scenario 3 (aOCS; $ub.f_{SEG|N}$, decreasing $ub.MC$ -1% each generation)

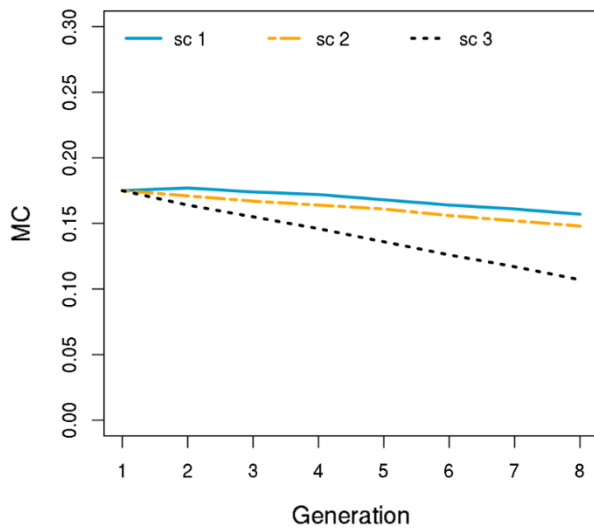


Figure 5. Amount of migrant contribution (MC) in Angler over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1 (OCS; $ub.f_{SEG}$), sc 2 = Scenario 2 (aOCS; $ub.f_{SEG|N}$, $ub.MC$ as in founder generation), sc 3 = Scenario 3 (aOCS; $ub.f_{SEG|N}$, decreasing $ub.MC$ -1% each generation)

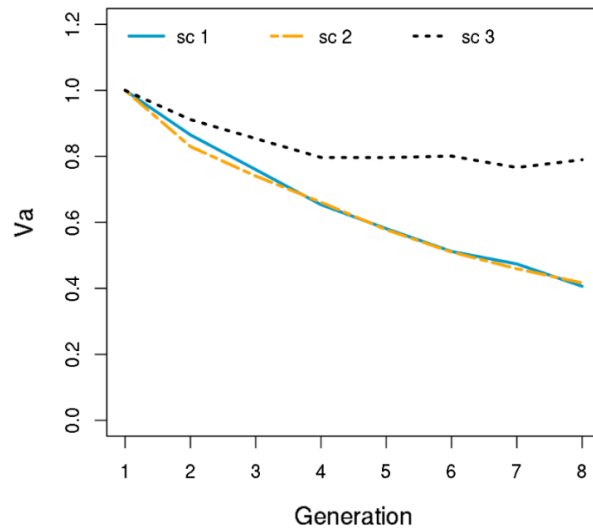


Figure 6. Additive variance (Va) in Angler over the simulated time period of 8 generations of rotational crossbreeding. Sc1 = Scenario 1 (OCS; $ub.f_{SEG}$), sc 2 = Scenario 2 (aOCS; $ub.f_{SEG|N}$, $ub.MC$ as in founder generation), sc 3 = Scenario 3 (aOCS; $ub.f_{SEG|N}$, decreasing $ub.MC$ -1% each generation)

Discussion

Crossbreeding has received increasing attention in dairy cattle breeding and research, mainly to utilize breed complementary and heterosis effects (Sørensen et al., 2008). Rotational crossbreeding is especially practical in dairy cattle, mainly due to the long generation interval and low female reproductive capacity, although terminal crossbreeding schemes are also considered (Clasen *et al.*, 2021a, b). In this study a genomic rotational crossbreeding scheme with the high yielding German Holstein breed and the numerically small German Angler breed was analysed that aims to utilize heterosis effects in the crossbred animals, and simultaneously to control inbreeding, and to reduce the migrant contribution in the Angler breed. Different scenarios of implementing OCS methods for Angler individuals were evaluated, which differed in their restrictions regarding conventional and native kinship, as well as the amount of migrant contributions from Holstein. It was shown that rotational crossbreeding can result in superior crossbred offspring compared to the purebred parental lines, when OCS methods are used.

However, reducing the amount of migrant contributions in Angler slowed down their genetic progress and did not result in higher heterosis effects in their crossbred offspring.

Homogeneous QTL effects were simulated for both breeds and complementary breed effects were not considered. Thus, the benefit of applying such a rotational crossbreeding for the Holstein breed were not analysed. This has to rely on practical data where next to heterosis also breed complementary effects are considered. For example, the average milk yield (protein yield) is 9,500 kg (330 kg) for the Holsteins and 8,300 kg (300 kg) for the Angler in the Northern part of Germany (LKV, 2020), underlying strong breed effects for milk production traits. For functional or health traits, no clear pattern regarding the differences between these two breeds reared in this region could be detected (LKV, 2020) although it is frequently said by farmers that the latter trait complexes are superior in the Angler.

The performance of Angler was best in Scenario 1 and worst in Scenario 3, especially in later generations (Figure 2), which confirmed that genetic gain is in conflict with the reduction of migrant contribution and the restriction of conventional and native kinship. Crossbreds slightly outperformed Angler and Holstein in early generations due to heterosis. Since heterosis is decreasing over generations (Figure 4), the crossbred performance was not superior compared to the Holstein population from generation 5 onwards. Note, that in Holsteins the genetic gain is higher as in the Angler (Figure 2) due to the larger population size, and heterosis could not compensate for this in later generations.

The flattened genetic gain observed for Angler in Scenario 3 caused also a reduced increase in TGV of the crossbreds and the TGV fluctuated with increasing amplitudes over generations. The amplitudes peaked in those generations where Holstein sires were used to produce the crossbred offspring. This can be explained by the overall higher genetic gain of Holstein compared to Angler (Figure 2), and by the heterosis effects (Figure 4). Removing Holstein genes means to favour animals for breeding that have low Holstein contributions. After few generations, only short Holstein segments remain in the population, which are difficult to remove. Consequently, the constraint that forces the Holstein contributions to reduce by 1% each generation becomes more serious in each generation. This constraint reduces the pool of animals that are eligible for breeding and thus reduces the genetic gain that can be achieved for the trait under selection. In general, a high marker density is desired to identify introgressed haplotype segments (Wang *et al.*, 2017a).

A moderately high historic effective population size of the Angler population in Germany was mainly achieved by introgression in the past. Some Angler breeders might have used introgression more extensively than others, so Angler breeders who did not want to use

introgression extensively had to choose their breeding animals from a smaller gene pool than the others. This implies that animals with small Holstein contributions tend to have similar Angler cattle in their pedigrees. Consequently, they have a higher native kinship. In this study, restricting the increase of the native kinship was, therefore, a slightly more serious constraint than restricting the increase of the conventional kinship (see Figure 2). Moreover, removing the Holstein genes from the Angler population while restricting the increase of the native kinship in Scenario 3 required strong selection on the maintenance of native allele diversity, which allowed only for a low selection intensity on the trait of interest, and thus resulted in a higher genetic diversity within Angler compared to Scenario 1 and 2, and also compared to Holstein. This can be seen by Figure 6, which shows that the additive genetic variance is substantially higher for Scenario 3 in later generations. This high variance caused the Angler animals in Scenario 3 to carry substantial heterosis effects by themselves.

In Scenario 3, the heterosis effect observed in the crossbreds was the lowest (Figure 4). The opposite was expected because removing Holstein genes from Angler cattle would make the breeds genetically more dissimilar. The reason for the smaller heterosis is as follows. The heterosis was calculated as the difference between the parent average and the TGV. The more genetically diverse each breeding line is, the higher the parent average is. In Scenario 3, Angler cattle were substantially more genetically diverse than in the other scenarios. Consequently, the heterosis of the cross was the lowest.

Applied to real animal data, the removal of Holstein haplotypes would even more compromise the genetic gain of the Angler, because it can be assumed that the introgressed Holstein haplotypes would contain favourable QTL alleles, due to the breed complementary effects for milk production traits and fertility traits illustrated above. Thus, in further simulation studies, it would be beneficial to identify the introgressed favourable QTL alleles in the Angler breed and to include them in the optimization process of aOCS. Future research is needed in order to go in this direction, as especially mapping the relevant QTL in such a small population is a challenging task. Further, it is well known, that not only Holstein but also other breeds were introgressed in the Angler breed (Schmidtman *et al.*, 2021). Hence, reducing the amount of MC could also include contributions from these other breeds. In this way the contribution coming from Holstein would not decrease as fast, yet the total amount of MC would be reduced.

When a high genetic diversity of livestock breeds is politically intended, farmers need incentives that compensate for the lower competitiveness of endangered breeds. However, they might be less dependent on incentives when they can achieve a higher herd profitability by including crossbred animals that are superior to their local breed.

The heterosis effect cause two crossbred animals to be economically superior to the same number of purebred animals, whereby one animal is from the endangered breed and the other is from the high-yielding breed. Consequently, implementing a rotational crossbreeding program would lower the need for incentives when running a breeding program for the endangered breed. Thus, farms with local breeds and breeders engaged in conservation programs could be saved.

The present paper showed that the gain in heterosis which arose from the removal of genetic contributions originating from the German Holstein could not compensate for the resulting lower breeding progress in the local Angler breed. Consequently, a breeding program that aims at increasing the proportion of the genome that is native requires incentives that compensate for the resulting lower breeding progress.

Conclusions

The results confirmed that the implementation of a genomic rotational crossbreeding scheme could be an interesting option to promote numerically small breeds like German Angler, as their superior crossbred offspring can lead to a higher herd performance, given that a suitable partner breed is available. Thus, rotational crossbreeding might increase the attractiveness of the local breed, and eventually might reduce their risk of numeric extinction. The results further showed that it is in general possible to combine a genomic rotational crossbreeding scheme with the maximization of genetic gain, while simultaneously restricting the increase of inbreeding and reducing migrant contributions. However, the latter one turned out to be the most compromising restriction with regard to genetic gain in purebreds and resulted in waves of heterosis effects observed in the crossbred individuals. The reduction of migrant contribution with the main aim to increase heterosis-effects in crossbreds was, therefore, not a recommendable breeding strategy in this study. This is because the possible gain in heterosis could not compensate for the reduced genetic gain imposed by the additional OCS restrictions.

Acknowledgements

This work was financially supported by the German Federal Ministry of Food and Agriculture (BMEL) through the Federal Office for Agriculture and Food (BLE), grant number 2817ERA10D. The project has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 696231 – ReDiverse Biodiversity within and between European Red Dairy Breeds). Joana Stock was partly supported by the H. Wilhelm Schaumann Foundation, Hamburg, Germany, which is gratefully acknowledged.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

JS, HE, RW, DH and JB designed the study. JS and HE simulated the data. JS did the statistical analyses. JS with the help of all other authors wrote the manuscript. All authors approved the manuscript.

References

- Buckley F, Lopez-Villalobos N and Heins BJ 2014. Crossbreeding: implications for dairy cow fertility and survival. *Animal : an international journal of animal bioscience* 8 Suppl 1, 122–133. doi:10.1017/S1751731114000901.
- Cara MÁR de, Villanueva B, Toro MÁ and Fernández J 2013. Using genomic tools to maintain diversity and fitness in conservation programmes. *Molecular Ecology* 22, 6091–6099. doi:10.1111/mec.12560.
- Clasen JB, Fikse WF, Kargo M, Rydhmer L, Strandberg E and Østergaard S 2020. Economic consequences of dairy crossbreeding in conventional and organic herds in Sweden. *Journal of Dairy Science* 103, 514–528. doi:10.3168/jds.2019-16958.
- Clasen JB, Kargo M, Fikse WF, Strandberg E, Wallenbeck A, Østergaard S and Rydhmer L 2021a. Conservation of a native dairy cattle breed through terminal crossbreeding with commercial dairy breeds. *Acta Agriculturae Scandinavica, Section A — Animal Science* 136, 1–12. doi:10.1080/09064702.2020.1867632.
- Clasen JB, Kargo M, Østergaard S, Fikse WF, Rydhmer L and Strandberg E 2021b. Genetic consequences of terminal crossbreeding, genomic test, sexed semen, and beef semen in dairy herds. *Journal of Dairy Science* 104, 8062–8075. doi:10.3168/jds.2020-20028.
- Esfandyari H, Berg P and Sørensen AC 2018. Balanced selection on purebred and crossbred performance increases gain in crossbreds. *Genetics, Selection, Evolution: GSE* 50, 8. doi:10.1186/s12711-018-0379-9.
- Esfandyari H and Sørensen AC 2017. xbreed: an R package for genomic simulation of purebreds and crossbreds. [R package xbreed version 1.0.1.1]. 68th Annual Meeting of the European Federation of Animal Science. Tallinn, Estonia.
- Falconer DS and Mackay TFC 1996. *Introduction to quantitative genetics*. 4th edition. Longmans Green, Harlow.
- Fernández J, Meuwissen THE, Toro MA and Mäki-Tanila A 2011. Management of genetic diversity in small farm animal populations. *Animal : an international journal of animal bioscience* 5, 1684–1698. doi:10.1017/S1751731111000930.
- Hartwig S, Wellmann R, Hamann H and Bennewitz J 2014. The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. *Journal of Animal Breeding and Genetics* 131, 496–503. doi:10.1111/jbg.12099.

- Kargo M, Clasen JB, Nielsen HM, Byskov K and Norberg E 2021. Short communication: Heterosis and breed effects for milk production and udder health traits in crosses between Danish Holstein, Danish Red, and Danish Jersey. *Journal of Dairy Science* 104, 678–682. doi:10.3168/jds.2019-17866.
- LKV Landeskontrollverband Schlesweig-Holstein e.V., Jahresbericht 2020. <https://www.lkv-sh.de/home/archiv/384-jahresbericht-2020>
- Meuwissen THE, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Oldenbroek K 2007. Utilisation and conservation of farm animal genetic resources. Wageningen Academic Publ, Wageningen.
- Schmidtman C, Schönherz A, Guldbrandtsen B, Marjanovic J, Calus M, Hinrichs D and Thaller G 2021. Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. *Genetics, Selection, Evolution: GSE* 53, 6. doi:10.1186/s12711-021-00613-6.
- Shonka-Martin BN, Heins BJ and Hansen LB 2019. Three-breed rotational crossbreds of Montbéliarde, Viking Red, and Holstein compared with Holstein cows for feed efficiency, income over feed cost, and residual feed intake. *Journal of Dairy Science* 102, 3661–3673. doi:10.3168/jds.2018-15682.
- Sørensen MK, Norberg E, Pedersen J and Christensen LG 2008. Invited review: crossbreeding in dairy cattle: a Danish perspective. *Journal of Dairy Science* 91, 4116–4128. doi:10.3168/jds.2008-1273.
- Stock J, Esfandyari H, Hinrichs D and Bennewitz J 2021. Implementing a genomic rotational crossbreeding scheme to promote local dairy cattle breeds-A simulation study. *Journal of Dairy Science* 104, 6873–6884. doi:10.3168/jds.2020-19927.
- Streit M, Wellmann R, Reinhardt F, Thaller G, Piepho H-P and Bennewitz J 2013. Using genome-wide association analysis to characterize environmental sensitivity of milk traits in dairy cattle. *G3 (Bethesda, Md.)* 3, 1085–1093. doi:10.1534/g3.113.006536.
- Thompson JR, Everett RW and Hammerschmidt NL 2000. Effects of Inbreeding on Production and Survival in Holsteins. *Journal of Dairy Science* 83, 1856–1864. doi:10.3168/jds.S0022-0302(00)75057-0.
- Wang Y, Bennewitz J and Wellmann R 2017a. Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds

with historical migration. *Genetics, Selection, Evolution: GSE* 49, 45. doi:10.1186/s12711-017-0320-7.

Wang Y, Segelke D, Emmerling R, Bennewitz J and Wellmann R 2017b. Long-term impact of Optimum Contribution Selection strategies on local livestock breeds with historical introgression using the example of German Angler cattle. *G3 (Bethesda, Md.)* 7, 4009–4018. doi:10.1534/g3.117.300272.

Wellmann R 2019. Optimum contribution selection for animal breeding and conservation: the R package optiSel. *BMC Bioinformatics* 20, 25. doi:10.1186/s12859-018-2450-5.

Wellmann R and Bennewitz J 2011. The contribution of dominance to the understanding of quantitative genetic variation. *Genetics Research* 93, 139–154. doi:10.1017/S0016672310000649.

Wellmann R and Bennewitz J 2019. Key Genetic Parameters for Population Management. *Frontiers in Genetics* 10, 667. doi:10.3389/fgene.2019.00667.

Wellmann R, Hartwig S and Bennewitz J 2012. Optimum contribution selection for conserved populations with historic migration. *Genetics, Selection, Evolution: GSE* 44. doi:10.1186/1297-9686-44-34.

Chapter 6

General Discussion

The main goal of this thesis was to develop methods for promoting local dairy breeds as crossing partners for high-yielding breeds in a rotational crossbreeding scheme.

Chapter 2 reviewed genomic models that are most suitable for the analysis of crossbred data, depending on data availabilities, desired accuracies and computational possibilities. Different additive models (such as the parental model, a model with breed-specific allele effects, and a single step model) and dominance models, which were either line-dependent, line-independent or included imprinting, were discussed. It was concluded, that the model choice needs to be made based on desired accuracies, computational possibilities, and data availability. In general, dominance models showed to result in higher accuracies compared to additive models without considering dominance.

A breed of origin of alleles (BOA) model approach was introduced in **Chapter 3**, which assumes different SNP effects for different origins of haplotypes, and thus fits especially local breeds with historical introgression from other breeds (i.e. Angler). The application of the BOA approach using a multi-breed reference set resulted in higher accuracies of genomic estimated breeding values for Angler compared to multi-breed and within-breed genomic predictions with GBLUP.

In the simulation study of **Chapter 4** different selection strategies were analysed when it comes to the implementation of genomic selection and rotational crossbreeding of Angler x Holstein, including different kinds of reference populations and selection goals. It was shown that crossbred animals tend to have a small overall superiority to both Holstein and Angler populations. Furthermore, a reference population containing both Angler and crossbred animals, in combination with a selection based on the purebred performance of Angler, resulted in the highest response to selection in the purebred Angler population and in the crossbred population.

In order to realize genetic gain while regaining the original genetic background of Angler, a simulation study on rotational crossbreeding with Holstein was performed, including different Optimum Contribution Selection methods strategies for the Angler population (**Chapter 5**). Different constraints regarding the kinships and migrant contributions from Holstein were applied to investigate their effects on Angler, crossbred, and Holstein populations. Constraining the amount of migrant contribution in Angler increased their genetic uniqueness. However, it led to a notable reduction of genetic gain and thus a reduced superiority of the

crossbred animals. The reduced genetic gain and thus the large difference of the performance between the parental breeds could not be compensated by heterosis effects.

In the following general discussion further genomic models for crossbreeding are discussed shortly, followed by the main part about the practical relevance of crossbreeding in dairy cattle, including different strategies of implementations, specific requirements of small sized populations, and the practical relevance of the studies' findings in this thesis. Finally, the transition period of shifting from a purebreeding system to rotational crossbreeding is discussed at the farm level.

Genomic Models

To counteract the negative consequences of past selection schemes, crossbreeding demonstrated to be an effective strategy to improve the long-term profitability and sustainability of dairy herds (Buckley *et al.*, 2014). The two reasons for superiority of crossbred animals is the utilization of different additive genetic levels between breeds and therefore new combinations of additive genetic components (Falconer and Mackay, 1996), and the exploitation of heterosis, due to non-additive genetic effects which is most pronounced for functional traits (Sørensen *et al.*, 2008). However, non-additive effects used to be ignored in the past in genetic evaluation of animal breeding for several reasons, i.e. 1) no sufficient informative pedigree data, such as full sib families, 2) complex calculations, or 3) the difficult use of dominant values in practical mate allocation (Hill, 2010; Varona *et al.*, 2018). Since dense SNP panels are available, several studies re-examined the importance of considering non-additive effects again and concluded that it can improve the accuracy of genomic predictions under certain conditions (Wellmann and Bennewitz, 2012; Zeng *et al.*, 2013; Nishio and Satoh, 2014; Sun *et al.*, 2014; Aliloo *et al.*, 2016; Esfandyari *et al.*, 2016; Lopes *et al.*, 2016).

The goal of genomic prediction in crossbreeding is to improve the performance of crossbred offspring while selection takes place in the purebred parental lines. Typically, purebred reference populations are used, despite the fact that purebred-crossbred correlations, called r_{pc} , are usually lower than one. Thus the genetic gain and response to selection of crossbred animals cannot be maximized (Duenk, 2018). Several studies (Esfandyari *et al.*, 2015; van Grevenhof and van der Werf, 2015; Xiang *et al.*, 2016; Lopes *et al.*, 2017; Duenk, 2018; Duenk *et al.*, 2019) found that the inclusion of crossbred animals in the reference population can increase the response to selection in crossbred offspring in typical crossbreeding schemes for pigs and broilers. Furthermore, Duenk *et al.* (2021) investigated the role of r_{pc} and its components, such as dominance, epistasis, and G x E interactions more detailed in order to optimize genomic selection for crossbred performance. They pointed out, that the most

important factor to know about is the r_{pc} and how it is affected. Therefore they suggested to not only obtain information about the size and structure of the reference population, but also the distance between the reference animals and the selection candidates, the genetic distance between the parental lines, and the heritability of the trait in order to get the highest accuracies of GEBV and response to selection. Based on this information the most suitable genomic model can be defined, and thus the required phenotype and genotype data. However, their work focused on terminal crossbreeding schemes and is therefore predominantly suitable for breeding strategies in pigs or poultry.

Recently, some studies focused particularly on systematic crossbreeding in dairy cattle. Systematic crossbreeding in dairy cattle often includes rotational crossbreeding, thus different amounts of contributions from the founder breeds can be observed in the crossbred individuals from generation to generation. Therefore, it is beneficial to consider the actual composition of the crossbred individuals while predicting their GEBV. In VanRaden *et al.* (2020) it was shown that using marker effects from each parental breed and then weighting them depending on their contribution to the crossbred animals was advantageous in order to improve the accuracies of genomic predictions. However, Eiríksson *et al.* (2021) stated that this approach might be only appropriate for the initial crossing, as the local ancestry varies throughout the genome. Furthermore, the origin of the alleles was not taken into account by VanRaden *et al.* (2020), which is of interest for heterozygous loci. Hence, Eiríksson *et al.* (2021) recommended a genomic model which considers the breed of origin of alleles (BOA). In this way, the difference of marker effects due to differences in LD between markers and QTL can be accounted for in the prediction of GEBV for rotational crossbred animals. They concluded, that their approach could be a promising option to provide GEBV for crossbred dairy cows without the need for a large reference population of crossbreds, and thus could support efficient dairy farming. In practice, this is especially of relevance in countries in which genotype data of crossbred animals is sparsely available. These findings are in accordance to Karaman *et al.* (2021) and the main results in Chapter 3, in which it was stated that the application of a BOA model in combination with breed specific SNP effects can improve genomic estimation of breeding values for rotational crossbred animals, as well as for populations with subpopulations or former introgression.

Dairy Crossbreeding in Practice

Heterosis is fully expressed in the F1 generation, as the level of heterozygosity is highest because genes at a locus originate from different breeds. Thus, the more different the parental breeds are, the more heterosis effects can be expected (Mäki-Tanila, 2007). But as dairy cattle have a low reproduction rate, a long generation interval and a relatively high economic value

per animal, continuous breeding schemes, such as two- or three- way rotational crossbreeding, fits better. Heterosis then settles at 67% or 86%, respectively (Buckley *et al.*, 2014). By exploiting heterosis effects functional traits can be improved to generate animals which show enhanced profitability and performance in comparison to their parents average, and to remove negative effects of inbreeding (Sørensen *et al.*, 2008; Buckley *et al.*, 2014). However, systematic crossbreeding of dairy cattle doesn't play a big role in most countries (Buckley *et al.*, 2014), besides the exception of New Zealand - there, crossbreeding showed its popularity in dairy production systems with a proportion of crossbred cows of around 50% in 2019 (Livestock Improvement Corporation and Dairy NZ, 2020).

The most common crossbreeding strategies in dairy cattle are:

- Two-way crossbreeding: cows of the F1 generation are mated with a high genetic merit sire of one of the parents' breeds. Average heterosis settles at 66.6% over time.
- Three-way crossbreeding: cows of the F1 generation are mated with a high genetic merit sire of a third breed. Heterosis is 100%. In the following generations, when the sire of the third breed is reintroduced again, average heterosis settles at 85.7%.
- Synthetic crossing: using high genetic merit crossbred sires to produce a new breed. Average heterosis settles at ~50% (Buckley *et al.*, 2014).

Rotational Crossbreeding

To assess economic effects resulting from changes in performance for milk yield, milk solid contents, reproduction, health and replacement, induced by rotational crossbreeding Holstein with other breeds, a simulation model was developed by Dezetter *et al.* (2017). The performance of purebred Holsteins were compared to the performance of Holstein x Montbéliarde, Holstein x Montbéliarde x Normande and Holstein x Montbéliarde x Scandinavian Red. After 15 years of simulation all crossbreds showed reduced performance for milk yield compared to pure Holsteins. On the other hand, the performance concerning fitness was improved in all three crossbreeding schemes, which led to improved margins over variable costs, especially under a high prevalence of reproductive and health disorders. In a study by Shonka-Martin *et al.* (2019) real animal data of purebred Holsteins were compared with rotational crossbred animals during their first three lactations. Crossbred animals showed the same fat plus protein production as pure Holsteins while having a lower milk volume, a lower dry matter intake and a higher body condition score. Since an optimal body condition score is related to improved fertility and reduced health disorders (Roche *et al.*, 2009), and in addition, low dry matter intakes lead to

reduced feed costs, Shonka-Martin *et al.* (2019) concluded that crossbred cows may have a higher profitability in commercial milk production than purebred Holsteins. This is in accordance to the results of several other studies which showed the economic superiority of rotational crossbreeding schemes compared to a pure breeding strategy (Lopez-Villalobos *et al.*, 2000; Heins *et al.*, 2012; Clasen *et al.*, 2020; Hazel *et al.*, 2020).

Furthermore, even in different environments and management levels, such as low- and high-input, milk production benefits from crossbreeding, as shown by several Scandinavian studies. To reject the myth about crossbreeding only being beneficial for low production levels and poor managed production systems, Kargo *et al.* (2012) investigated how the effect of heterosis is expressed in different management levels in crossbred Jersey herds. They concluded that heterosis for milk yield, protein yield, and fat yield was smallest in the low producing herds and showed the highest effect in the intermediate groups. Thus, they recommended crossbreeding as a strategy to consider for any dairy producer. These findings were supported by Clasen *et al.* (2019) who additionally included functional traits in their analyses. Their results of comparing Holstein with Holstein x Nordic Red demonstrated that heterosis effects in production traits were largest in high-yielding herds, whereas it was independent from the production level for fitness related traits, such as fertility, udder health, stillbirth, and survival. Further, in a following study, Clasen *et al.* (2020) could prove that crossbreeding Swedish Red x Swedish Holstein has economic advantages over purebred Swedish Holstein herds. This applied to both organic and conventional dairy production systems and was most successful in rotational crossbreeding.

Composite Breeds

In Latin America a common strategy in the beef cattle industry is to use composite breeds in order to combine the favorable fitness and adaptability traits from indicine cattle (*Bos indicus*) and the production and carcass traits from European cattle (*Bos taurus*) (Buzanskas *et al.*, 2017). The composite breed Brangus was developed in order to meet the needs of meat production in tropical and subtropical environments. Thus, adaptability, disease resistance, maternal instinct and general robustness traits from Zebu (mainly Brahman) were combined with fertility and meat quality traits from Angus. This resulted in a founder breed composition of 62.5% Angus and 37.5% Zebu (Goszczynski *et al.*, 2018). Other composite breeds, such as Canchim (Charolais x Zebu) or Braford (Brahman x Hereford) are used for meat production, as well (Orellana *et al.*, 2009; Buzanskas *et al.*, 2017), whereas Girolando (Holstein x Gyr) was developed to improve dairy production systems (Canaza-Cayo *et al.*, 2016). Also in Germany a synthetic breed was developed in the past - during the 1970s and 1980s the composite breed Schwarzbuntes Milchrind (SMR) was used in former East Germany for commercial milk production. To generate this dual-purpose breed with focus on milk production, a three-breed

crossing including native East German Black and White, Danish Jersey and Canadian Holstein-Friesian was done (Freyer *et al.*, 2008). The three-breed crossbreeding, originally suggested by Schönmutz (1963), started in the 1960s by crossing East German Black and White dams and Jersey sires, to generate F1 dams. Then, in the following years, Holstein sires were mated with F1 dams, to achieve a long-term breeding goal defined as an annual milk production of 5000 to 6000kg with 4% fat, sufficient muscling for 600kg body weight and around 130cm in height at withers. In 1975 the breeding of SMR in pure lines began. SMR showed to be a useful breed for milk production, even under suboptimal management conditions and feeding which were found in former East Germany. After the reunification of Germany in 1990, the conditions in East Germany improved and thus the need for SMR decreased. As a result, it was decided to incorporate the SMR population into the German Holstein breed (Freyer *et al.*, 2008).

Terminal Crossbreeding

The “Beef-on-dairy” strategy and the potential of beef-on-dairy production systems was reviewed in Berry (2021). In several countries, such as New Zealand (Morris, 2008), Sweden (Federation of Swedish Farmers, 2019), and Finland (Niemi and Ahlstedt, 2013), 60-80% of the total beef output is produced in dairy herds. This meat originates directly (such as culled cows) or indirectly (such as surplus calves of beef x dairy matings) from dairy production systems. Especially the latter is of growing interest, as it can provide overall advantages. In combination with sexed semen and the use of dairy-beef breeding indexes, it can be an appealing option for dairy farmers to extend their businesses. In addition, from the consumers increasing awareness of environmental and ethical issues of animal production systems, beef x dairy calves provide improvement for both (Berry, 2021). Assuming an environmental footprint attributed to kg of the end product, beef from dairy herds showed to yield lower (i.e. in terms of global warming potential, acidification potential, eutrophication potential, energy usage, and land usage) compared to beef produced in suckler-based systems. This can be explained by the fact that in suckler-based systems the whole environmental impact is allocated to the production of beef, whereas in dairy-based systems it is allocated to milk and beef. Thus, the impact is distributed more widely and it is lower per kg of output. In addition, the advantage of beef from dairy herds holds true for both concentrate-based and roughage-based systems and was observed to show further potential to mitigate environmental impact when dual-purpose cattle or dairy x beef crosses were used – caused by faster growth rates and improved feed efficiency (de Vries *et al.*, 2015). In order to meet the societies’ concerns about worthless bobby calves, which may be slaughtered at a very young age as they are neither suitable for replacement of the herd nor for profitable fattening, a beef-on-dairy production might be a promising strategy. In this way, the number of bobby calves could be reduced remarkably, particularly in combination with sexed

semen. However, beef x dairy animals tend to have poorer performance (i.e. primal cut yield) compared to beef animals, as well as they can have high labor and specialized infrastructure requirements in their pre-weaning period. This might lead to low margins, which might be a major weakness of the system under certain conditions (Berry, 2021).

Despite the fact of long generation intervals and low reproduction rates, terminal crossbreeding can also be advantageous in dairy breeding systems in terms of total economic benefits. Clasen *et al.* (2021b) investigated whether breeding tools, such as sexed semen, genomic testing, beef semen, and terminal crossbreeding can improve the profitability at herd level. They concluded that breeding tools in general led to high economic returns, while a combination of terminal crossbreeding, sexed semen and genomic testing resulted in the highest profitability. Nevertheless, Clasen *et al.* (2021a) showed another advantageous aspect of terminal crossbreeding by examining the conservation potential of native breeds at farm level when terminal crossbreeding is applied. Due to higher herd economies caused by crossbred animals, and lowered associated costs, farms with local breeds could be kept viable. Hence, the number of breeders engaged in conservation schemes could be saved. In addition, the marketing of niche products from the local breed might be a solution to further increase the benefits of keeping them. However, it is of crucial importance to control the crossbreeding and to have a well-managed conservation scheme, in order to not exterminate the local breeds by inappropriate crossbreeding schemes (leading to displacement of the original genes) or high inbreeding levels (leading to inbreeding depression).

Genetic Diversity in Small Sized Populations

Another threat to local breeds can be a special form of crossbreeding called “upgrading”. In this way cows of moderate or low performance breeds are crossed with sires of high yielding breeds to increase the production level of subsequent generations in the low performance breed. At the same time it increases the genetic diversity within the breed and counteracts potential inbreeding problems. Upgrading was frequently applied in low-sized and local breeds in the past, e.g. in the Vorderwald cattle in southern Germany (Hartwig *et al.*, 2014) or German Angler (Schmidtmann *et al.*, 2021). However, if this upgrading is repeated over several generations, the original breed might become extinct due to the replacement of the native genes (Wang *et al.*, 2017; 2019).

Domestic breeds often have small effective population sizes and a low genetic diversity within the breeds. Thus, much of the genetic diversity is found between breeds. On the other hand, to ensure future selection response breeds need a certain level of within-breed diversity. Hence, it is required to maintain both within- and between-breed genetic diversity in order to meet future demands for robust and food-efficient breeds, such as the ability to adapt to climate

changes or new diseases (Wellmann and Bennewitz, 2019). For numerically small breeds with historic introgression the focus should be therefore on the between-breed diversity and thus the original genetic background. To realize the latter, or in other words, to regain the genetic uniqueness of a breed, advanced Optimum Contribution Selection (aOCS) showed to be the method of choice (Wang *et al.*, 2019; Kohl *et al.*, 2020a, 2020b). In this way, genetic gain can be achieved while the rate of inbreeding and the amount of migrant contributions is controlled. However, the optimization goal and the corresponding constraints of aOCS methods should meet the main goal of the breeding program, as there can be a large deficit in genetic gain, as soon as the amount of foreign genetic material is restricted (see Chapter 5). This deficit then might lead to a very low profitability of the breed and thus very little interest of farmers to keep it. A breeding program for local breeds with the main goal to remove introgression from mainstream breeds or to conserve endangered breeds is therefore economically depended on subsidies, as stated in Wellmann and Bennewitz (2019) and in Chapter 5.

Practical Relevance of Studies' Findings

In practice, GEBV for Angler are available since 2016. The corresponding reference population is provided by the Danish breeding organization VikingGenetics and contains the following Scandinavian red dairy breeds: Danish Red, Norwegian Red, Swedish Red and Finnish Ayrshire. It includes about 5000 bulls and 30.500 cows. To increase the accuracies for the GEBV of Angler, Angler bulls have been added to the reference population, as well (August 2021: 174 bulls) (RSHeG, 2021b). Hence, the first step to realize a genomic rotational crossbreeding scheme in practice could be to start with simplified conditions, such as selecting Angler and Holstein sires for crossing based on their available (genomic) breeding values (see Scenario 1a in Chapter 4). However, as soon as the rotational crossbreeding program has established and showed evident benefits, breeding organizations might be willing to invest money for further improvement, such as genotyping more animals of the local breeds. In this way the advantages of additional data (i.e. enlarging the reference population in general or increasing the contribution of the target breed in the reference population) could be used to increase the accuracies of GEBV and the response to selection. As a next step to improve such a breeding scheme, phenotype and genotype data from crossbred animals could be collected and used for more complex genomic models. As a consequence, both crossbred animals and local breeds could profit from further increased accuracies and thus genetic gain. Given that the performance of Angler is high, implementing aOCS methods could be the final step to improve the overall breeding scheme. Depending on financial support, such as subsidies, and the considered foreign breeds that caused migrant contributions, a remarkable

regain of the original genetic background could be realized without suffering from economic disadvantages.

Implementation of Rotational Crossbreeding on Farm

In general, to make the implementation of a rotational crossbreeding scheme possible, breeders must be willing to participate. Introducing a new breeding scheme, such as changing from purebreeding to rotational crossbreeding is not self-evident as it leads to a higher overall effort in farm management in the transition period (Quénon *et al.*, 2020).

However, there seem to be common factors that drive farmers to consider crossbreeding, as studied by Magne and Quénon (2021) in France. The drivers were clustered in three main groups containing several indicators such as

- I) Counteracting breeding and health problems of high-yielding cows
 - Improving performance level of purebred herds
 - Improving functional traits, i.e. fertility, health, robustness
 - Increasing protein and fat content in the milk
 - Managing the inbreeding level of purebred herds
- II) Shifting towards a more sustainable and resilient dairy production system
 - Transition to a grassland based / low-input / organic farming system
 - Improving farm economic profitability
 - Improving working conditions
- III) Regaining decision-making autonomy in their farm management
 - Countering the separation of knowing about animal breeding and actually performing it

The majority of farmers referred to improving performance and functional traits (group I) as the main reason to shift their breeding strategy, and to a lesser extent group II) and III). Once they decided to shift, they were faced with challenges, such as lack of knowledge about the possible crossbreeding schemes or practical implementation, but also with cultural and technological barriers and concerns about low-economic return. In addition, many French stakeholders who are involved in dairy breeding (i.e. breeding societies, research institutions, consultants, livestock traders) were not able or willing to support rotational crossbreeding, which might even enhanced the challenges of the dairy farmers (Magne and Quénon, 2021).

This could be the reason why rotational crossbreeding is still not commonly used in dairy production systems, even though its popularity has been increased over the last two decades (Quénon *et al.*, 2020). Nevertheless, once the decision about starting a rotational crossbreeding scheme was made, Quénon *et al.* (2020) observed typical technical pathways

and whole-farm transitions when it came to the practical transition in French production systems. By interviewing and retrospectively analyzing, they showed three different pathways of adapting rotational crossbreeding as follows:

- I) Customizing and implementing a rotational crossbreeding scheme **quickly** while redesigning the whole farm, such as converting to organic or low-input farming systems.
- II) Customizing and implementing a rotational crossbreeding scheme **progressively** while redesigning the whole farm, such as converting to organic or low-input farming systems.
- III) Implementing a relatively well-known rotational crossbreeding scheme **quickly** to fix fertility issues without changing the overall farming system.

Farmers of pathways I) and II) customized the breeding scheme in a trial-and-error phase to find a scheme which fits their herd and personal objectives best, before they started the actual crossbreeding. Thus, they gained autonomy in decision-making over dairy genetics. In contrast, in pathway III) farmers relied on existing breeding schemes (i.e. the commercial three-way rotational crossbreeding scheme 'ProCROSS') which reduced uncertainty and provided technical advising services. A consequence of the fast transition in pathway I) and III) was a change in the herd demographics shown by excess heifers. Strategies to manage these animals were increasing sales and stopping the purchase of heifers, replacing some dairy crossings with beef crossings, and increasing the overall herd size. However, regardless of the chosen pathway, the transition must be accompanied by feedback loops between herd management, their effect on herd performances, herd demographics, and technical choices (Quénon *et al.*, 2020). In addition, to support dairy farmers who are considering to start rotational crossbreeding, it is of crucial importance to understand how former transitions have been performed by others and to develop informal peer networks to overcome doubts and impediments arising during transition (Quénon *et al.*, 2020; Magne and Quénon, 2021). Nevertheless, to support the popularity of crossbreeding in order to improve dairy production systems, the corresponding knowledge should be spread at all levels, including all kinds of different stakeholders, as well as farmers.

Relating the results from France to the findings of this thesis, the shift from purebreeding to rotational crossbreeding in dairy cattle appear to be highly complex. Thus, providing the methodical answers for the implementation (i.e. GEBV for crossbred animals, reference population design, selection methods) should be interpreted as one part of the big picture, rather than an overall solution.

Concluding Remarks

Despite the knowledge about the advantages of systematic crossbreeding in dairy cattle it is still not widely adapted in Europe. Challenges include complex genomic models to obtain accurate genomic prediction of crossbred animals, as they require a certain amount of genotype and phenotype data. However, up to now, routine genomic evaluation is neither implemented widely for crossbred dairy nor for numerically small breeds. Thus, to implement a genomic rotational crossbreeding scheme in local breeds, perhaps even including OCS methods, lacking animal data would be the main barrier to overcome. A further challenge might be of practical relevance. Farmers must be willing to shift to a new breeding scheme and accept possible uncertainties during the process. Therefore, it is of importance that dairy breeding stakeholders are supporting them. To achieve solutions for the above mentioned challenges, further activities might focus on technological topics, i.e. feasible data collection and feasible data processing in genomic models, and public relations work in the dairy breeding sector, in order to obtain an overall progress.

References

- Aliloo H, Pryce JE, González-Recio O, Cocks BG and Hayes BJ 2016. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genetics, Selection, Evolution: GSE* 48, 8. doi:10.1186/s12711-016-0186-0.
- Berry DP 2021. Invited review: Beef-on-dairy-The generation of crossbred beef × dairy cattle. *Journal of Dairy Science* 104, 3789–3819. doi:10.3168/jds.2020-19519.
- Buckley F, Lopez-Villalobos N and Heins BJ 2014. Crossbreeding: implications for dairy cow fertility and survival. *Animal : an international journal of animal bioscience* 8 Suppl 1, 122–133. doi:10.1017/S1751731114000901.
- Buzanskas ME, Ventura RV, Seleguim Chud TC, Bernardes PA, Santos DJdA, Regitano LCdA, Alencar MMd, Mudadu MdA, Zanella R, da Silva MVGB, Li C, Schenkel FS and Munari DP 2017. Study on the introgression of beef breeds in Canchim cattle using single nucleotide polymorphism markers. *PLoS ONE* 12, e0171660. doi:10.1371/journal.pone.0171660.
- Canaza-Cayo AW, Cobuci JA, Lopes PS, Almeida Torres R de, Martins MF, dos Santos Daltro D and Barbosa da Silva MVG 2016. Genetic trend estimates for milk yield production and fertility traits of the Girolando cattle in Brazil. *Livestock Science* 190, 113–122. doi:10.1016/j.livsci.2016.06.009.
- Clasen JB, Fikse WF, Kargo M, Rydhmer L, Strandberg E and Østergaard S 2020. Economic consequences of dairy crossbreeding in conventional and organic herds in Sweden. *Journal of Dairy Science* 103, 514–528. doi:10.3168/jds.2019-16958.
- Clasen JB, Fogh A and Kargo M 2019. Differences between performance of F1 crossbreds and Holsteins at different production levels. *Journal of Dairy Science* 102, 436–441. doi:10.3168/jds.2018-14975.
- Clasen JB, Kargo M, Fikse WF, Strandberg E, Wallenbeck A, Østergaard S and Rydhmer L 2021a. Conservation of a native dairy cattle breed through terminal crossbreeding with commercial dairy breeds. *Acta Agriculturae Scandinavica, Section A — Animal Science* 136, 1–12. doi:10.1080/09064702.2020.1867632.

-
- Clasen JB, Kargo M, Østergaard S, Fikse WF, Rydhmer L and Strandberg E 2021b. Genetic consequences of terminal crossbreeding, genomic test, sexed semen, and beef semen in dairy herds. *Journal of Dairy Science* 104, 8062–8075. doi:10.3168/jds.2020-20028.
- de Vries M, van Middelaar CE and Boer IJM de 2015. Comparing environmental impacts of beef production systems: A review of life cycle assessments. *Livestock Science* 178, 279–288. doi:10.1016/j.livsci.2015.06.020.
- Dezetter C, Bareille N, Billon D, Côrtes C, Lechartier C and Seegers H 2017. Changes in animal performance and profitability of Holstein dairy operations after introduction of crossbreeding with Montbéliarde, Normande, and Scandinavian Red. *Journal of Dairy Science* 100, 8239–8264. doi:10.3168/jds.2016-11436.
- Duenk P 2018. Accuracy of genomic estimated breeding values for crossbred performance in broilers using a purebred or crossbred reference population. *Proceedings of the World Congress on Genetics Applied to Livestock Production*.
- Duenk P, Bijma P, Wientjes YCJ and Calus MPL 2021. Review: optimizing genomic selection for crossbred performance by model improvement and data collection. *Journal of Animal Science* 99. doi:10.1093/jas/skab205.
- Duenk P, Calus MPL, Wientjes YCJ, Breen VP, Henshall JM, Hawken R and Bijma P 2019. Validation of genomic predictions for body weight in broilers using crossbred information and considering breed-of-origin of alleles. *Genetics, Selection, Evolution: GSE* 51, 38. doi:10.1186/s12711-019-0481-7.
- Eiríksson JH, Karaman E, Su G and Christensen OF 2021. Breed of origin of alleles and genomic predictions for crossbred dairy cows. *Genetics, Selection, Evolution: GSE* 53, 84. doi:10.1186/s12711-021-00678-3.
- Esfandyari H, Bijma P, Henryon M, Christensen OF and Sørensen AC 2016. Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genetics, Selection, Evolution: GSE* 48, 40. doi:10.1186/s12711-016-0220-2.
- Esfandyari H, Sørensen AC and Bijma P 2015. A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genetics, Selection, Evolution: GSE* 47, 76. doi:10.1186/s12711-015-0155-z.

-
- Falconer DS and Mackay TFC 1996. Introduction to quantitative genetics. 4th edition. Longmans Green, Harlow.
- Federation of Swedish Farmers 2019. LRFs Statistikplattform - not. Gris Och Lamm, Federation of Swedish Farmers, Stockholm, Sweden.
- Freyer G, König S, Fischer B, Bergfeld U and Cassell BG 2008. Invited review: crossbreeding in dairy cattle from a German perspective of the past and today. *Journal of Dairy Science* 91, 3725–3743. doi:10.3168/jds.2008-1287.
- Goszczynski DE, Corbi-Botto CM, Durand HM, Rogberg-Muñoz A, Munilla S, Peral-Garcia P, Cantet RJC and Giovambattista G 2018. Evidence of positive selection towards Zebuine haplotypes in the BoLA region of Brangus cattle. *Animal : an international journal of animal bioscience* 12, 215–223. doi:10.1017/S1751731117001380.
- Hartwig S, Wellmann R, Hamann H and Bennewitz J 2014. The contribution of migrant breeds to the genetic gain of beef traits of German Vorderwald and Hinterwald cattle. *Journal of Animal Breeding and Genetics* 131, 496–503. doi:10.1111/jbg.12099.
- Hazel AR, Heins BJ and Hansen LB 2020. Health treatment cost, stillbirth, survival, and conformation of Viking Red-, Montbéliarde-, and Holstein-sired crossbred cows compared with pure Holstein cows during their first 3 lactations. *Journal of Dairy Science* 103, 10917–10939. doi:10.3168/jds.2020-18604.
- Heins BJ, Hansen LB and Vries A de 2012. Survival, lifetime production, and profitability of Normande × Holstein, Montbéliarde × Holstein, and Scandinavian Red × Holstein crossbreds versus pure Holsteins. *Journal of Dairy Science* 95, 1011–1021. doi:10.3168/jds.2011-4525.
- Hill WG 2010. Understanding and using quantitative genetic variation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365, 73–85. doi:10.1098/rstb.2009.0203.
- Karaman E, Su G, Croue I and Lund MS 2021. Genomic prediction using a reference population of multiple pure breeds and admixed individuals. *Genetics, Selection, Evolution: GSE* 53, 46. doi:10.1186/s12711-021-00637-y.

- Kargo M, Madsen P and Norberg E 2012. Short communication: Is crossbreeding only beneficial in herds with low management level? *Journal of Dairy Science* 95, 925–928. doi:10.3168/jds.2011-4707.
- Kohl S, Wellmann R and Herold P 2020a. Advanced optimum contribution selection as a tool to improve regional cattle breeds: a feasibility study for Vorderwald cattle. *Animal : an international journal of animal bioscience* 14, 1–12. doi:10.1017/S1751731119001484.
- Kohl S, Wellmann R and Herold P 2020b. Implementation of advanced Optimum Contribution Selection in small-scale breeding schemes: prospects and challenges in Vorderwald cattle. *Animal : an international journal of animal bioscience* 14, 452–463. doi:10.1017/S1751731119002295.
- Livestock Improvement Corporation and Dairy NZ 2020. New Zealand dairy statistics 2019-2020.
- Lopes MS, Bastiaansen JWM, Janss L, Knol EF and Bovenhuis H 2016. Genomic prediction of growth in pigs based on a model including additive and dominance effects. *Journal of Animal Breeding and Genetics* 133, 180–186. doi:10.1111/jbg.12195.
- Lopes MS, Bovenhuis H, Hidalgo AM, van Arendonk JAM, Knol EF and Bastiaansen JWM 2017. Genomic selection for crossbred performance accounting for breed-specific effects. *Genetics, Selection, Evolution: GSE* 49, 51. doi:10.1186/s12711-017-0328-z.
- Lopez-Villalobos N, Garrick DJ, Blair HT and Holmes CW 2000. Possible Effects of 25 Years of Selection and Crossbreeding on the Genetic Merit and Productivity of New Zealand Dairy Cattle. *Journal of Dairy Science* 83, 154–163. doi:10.3168/jds.S0022-0302(00)74866-1.
- Magne M-A and Quénon J 2021. Dairy crossbreeding challenges the French dairy cattle sociotechnical regime. *Agronomy for Sustainable Development* 41, 1. doi:10.1007/s13593-021-00683-2.
- Mäki-Tanila A 2007. An overview on quantitative and genomic tools for utilising dominance genetic variation in improving animal production. *Agricultural and Food Science* 16, 188. doi:10.2137/145960607782219337.
- Morris ST 2008. A review of beef cross dairy bred cattle as beef breeding cows. Report prepared for Meat and Wood, New Zealand.

- Niemi J and Ahlstedt J 2013. Finnish agriculture and rural industries 2013. Publication 114a. MTT Agrifood Research Finland, Helsinki, Finland.
- Nishio M and Satoh M 2014. Including dominance effects in the genomic BLUP method for genomic evaluation. PLoS ONE 9, e85792. doi:10.1371/journal.pone.0085792.
- Orellana C, Peña F, García A, Perea J, Martos J, Domenech V and Acero R 2009. Carcass characteristics, fatty acid composition, and meat quality of Criollo Argentino and Braford steers raised on forage in a semi-tropical region of Argentina. Meat science 81, 57–64. doi:10.1016/j.meatsci.2008.06.015.
- Quénou J, Ingrand S and Magne M-A 2020. Managing the transition from purebred to rotational crossbred dairy cattle herds: three technical pathways from a retrospective case-study analysis. Animal : an international journal of animal bioscience 14, 1293–1303. doi:10.1017/S1751731119003458.
- Roche JR, Friggens NC, Kay JK, Fisher MW, Stafford KJ and Berry DP 2009. Invited review: Body condition score and its association with dairy cow productivity, health, and welfare. Journal of Dairy Science 92, 5769–5801. doi:10.3168/jds.2009-2431.
- RShEG 2021. Genomic selection in the German Angler population. Personal Communication (Email, K. Breuer, 12.11.2021).
- Schmidtman C, Schönherz A, Guldbrandtsen B, Marjanovic J, Calus M, Hinrichs D and Thaller G 2021. Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. Genetics, Selection, Evolution: GSE 53, 6. doi:10.1186/s12711-021-00613-6.
- Schönmuth G 1963. Zur Züchtung eines milchbetonten Zweinutzungsrides mit hohem MilCHFett- und Eiweißgehalt und bestem Euter. Archives Animal Breeding, 79–92.
- Shonka-Martin BN, Heins BJ and Hansen LB 2019. Three-breed rotational crossbreds of Montbéliarde, Viking Red, and Holstein compared with Holstein cows for feed efficiency, income over feed cost, and residual feed intake. Journal of Dairy Science 102, 3661–3673. doi:10.3168/jds.2018-15682.
- Sørensen MK, Norberg E, Pedersen J and Christensen LG 2008. Invited review: crossbreeding in dairy cattle: a Danish perspective. Journal of Dairy Science 91, 4116–4128. doi:10.3168/jds.2008-1273.

-
- Sun C, VanRaden PM, Cole JB and O'Connell JR 2014. Improvement of Prediction Ability for Genomic Selection of Dairy Cattle by Including Dominance Effects. *PLoS ONE* 9. doi:10.1371/journal.pone.0103934.
- van Grevenhof IEM and van der Werf JHJ 2015. Design of reference populations for genomic selection in crossbreeding programs. *Genetics, Selection, Evolution: GSE* 47, 14. doi:10.1186/s12711-015-0104-x.
- VanRaden PM, Tooker ME, Chud TCS, Norman HD, Megonigal JH, Haagen IW and Wiggins GR 2020. Genomic predictions for crossbred dairy cattle. *Journal of Dairy Science* 103, 1620–1631. doi:10.3168/jds.2019-16634.
- Varona L, Legarra A, Toro MA and Vitezica ZG 2018. Non-additive Effects in Genomic Selection. *Frontiers in Genetics* 9, 78. doi:10.3389/fgene.2018.00078.
- Wang Y, Bennewitz J and Wellmann R 2019. Managing genomes of conserved livestock breeds with historical introgression to decrease genetic overlap with other breeds. *Journal of Animal Breeding and Genetics* 136, 505–517. doi:10.1111/jbg.12405.
- Wang Y, Segelke D, Emmerling R, Bennewitz J and Wellmann R 2017. Long-term impact of Optimum Contribution Selection strategies on local livestock breeds with historical introgression using the example of German Angler cattle. *G3 (Bethesda, Md.)* 7, 4009–4018. doi:10.1534/g3.117.300272.
- Wellmann R and Bennewitz J 2012. Bayesian models with dominance effects for genomic evaluation of quantitative traits. *Genetics Research* 94, 21–37. doi:10.1017/S0016672312000018.
- Wellmann R and Bennewitz J 2019. Key Genetic Parameters for Population Management. *Frontiers in Genetics* 10, 667. doi:10.3389/fgene.2019.00667.
- Xiang T, Nielsen B, Su G, Legarra A and Christensen OF 2016. Application of single-step genomic evaluation for crossbred performance in pig. *Journal of Animal Science* 94, 936–948. doi:10.2527/jas.2015-9930.
- Zeng J, Toosi A, Fernando RL, Dekkers JCM and Garrick DJ 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genetics, Selection, Evolution: GSE* 45, 11. doi:10.1186/1297-9686-45-11.

Acknowledgments

Nie geplant und doch so schön. Das sind wahrscheinlich die zutreffendsten Worte für meine Doktorandenzeit hier am Fachgebiet für Tiergenetik und Züchtung. Angefangen hat alles mit meiner Masterarbeit, gemeinsamen Mittagspausen in der Mensa und der Info über eine frei gewordene Doktorandenstelle. Dann nahmen die Dinge ihren Lauf und ich möchte nichts davon missen.

Allen voran gilt mein Dank Prof. Dr. Jörn Bennewitz, meinem Doktorvater, für die tolle Betreuung, fürs Nachhaken und gemeinsame Überlegen, fürs Fordern und Fördern, für das Eröffnen von Möglichkeiten, über mich selbst hinaus zu wachsen.

Auch möchte ich mich herzlich bei Prof. Dr. Dirk Hinrichs und Prof. Dr. Mizeck Chagunda für die Übernahme der 2. und 3. Prüfung bedanken, und der dazu aufgebrauchten Zeit und Muße.

Weiterhin danke ich der H. Wilhelm Schaumann Stiftung für das Promotionsstipendium, ohne welches dieses Promotionsprojekt nicht hätte realisiert werden können.

Auch ohne die Co-Autoren wäre Vieles nicht möglich gewesen. Ein ganz großer Dank gebührt daher auch ihnen: Dr. Dr. Robin Wellmann, der stets mit konstruktivem Feedback und Ideen zur Stelle war; Dr. Markus Schmid, oft mein erster Ansprechpartner, mit dem immer offenes Ohr für kleinere oder größere Probleme. And last but not least, Dr. Hadi Esfandyari. I've never expected that the very first Email I sent to you would turn out into such a nice collaboration. Thank you so much for your support and time!

Further, I want to thank Christopher Childs for the language editing and not turning my thesis into a scifi novel. However, I might keep it in mind for the future.

Ganz besonders möchte ich mich abschließend bei den Doktorand*innen und beim gesamten Team des Fachgebietes bedanken, die mich auf diesem Weg begleitet haben. Von euch lebt der Teamspirit, den ich hier so genossen habe. Die Zeit mit euch war fabelhaft!

Curriculum Vitae

Personal Details

Name	Joana Stock
Date of Birth	20.03.1990
Place of Birth	Herford, Germany

Education

10/2018 – to date	Ph.D., Department of Animal Genetics and Breeding University of Hohenheim, Germany
10/2015 – 09/2018	M.Sc. Animal Sciences University of Hohenheim, Germany
10/2012 – 09/2015	B.Sc. Agricultural Sciences University of Hohenheim, Germany
03/2012 – 07/2012	Internship at organic farm Biohof Bobbert Bielefeld, Germany

Hohenheim, 20. Januar 2022



Joana Stock

Eidesstattliche Versicherung

gemäß § 8 Absatz 2 der Promotionsordnung der Universität Hohenheim zum Dr.sc.agr.

1. Bei der eingereichten Dissertation zum Thema „Genomic Methods for Rotational Crossbreeding in Local Dairy Cattle Breeds“

handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Ich habe nicht die Hilfe einer kommerziellen Promotionsvermittlung oder -beratung in Anspruch genommen.

4. Die Bedeutung der eidesstattlichen Versicherung und der strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Die Richtigkeit der vorstehenden Erklärung bestätige ich. Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

Hohenheim, 20. Januar 2022



Joana Stock