Aus dem Institut für Tierzucht und Haustiergenetik
Professur für Tierzüchtung
der Justus-Liebig-Universität Gießen

Breeding for polledness in cattle – Exploring potential side effects and sustainable breeding strategies in quantitative genetic, genomic and simulation studies.

INAUGURAL-DISSERTATION
zur Erlangung des Doktorgrades (Dr. agr.)
im Fachbereich Agrarwissenschaften, Ökotrophologie und
Umweltmanagement der Justus-Liebig-Universität Gießen

vorgelegt von

Carsten Scheper
aus Recklinghausen

Gießen, 2023

Mit Genehmigung des Fachbereiches Agrarwissenschaften, Ökotrophologie und Umweltmanagement der
Justus-Liebig-Universität Gießen

Dekan: Prof. Dr. Klaus Eder

Prüfungskommission

1. Gutachter:	Prof. Dr. Sven König
2. Gutachter:	Prof. Dr. Kay-Uwe Götz
Prüfer:	Prof. Dr. Dirk Hinrichs
Prüfer:	Prof. Dr. Horst Brandt
Vorsitzende:	Prof. Dr. Gesine Lühken

Tag der Disputation: 19.09.2023

Diese Arbeit wurde aus Mitteln des Zweckvermögens des Bundes bei der landwirtschaftlichen Rentenbank im Rahmen des Projektes „Zuchtprogramme für Hornlosigkeit beim Milch- und Zweinutzungsrind in Deutschland“ gefördert (Förderkennzeichen: Z 20141).

"Without prejudice to what the future may disclose whether by way of limitation or extension of [the] Mendelian method, it can be declared with confidence and certainty that we have now the means of beginning an analysis of living organisms, and distinguishing many of the units or factors which essentially determine and cause the development of their several attributes.
Briefly put, the essence of Mendelism lies in the discovery of the existence of unit characters or factors."

– William Bateson, Problems of Genetics, 1913

for Hanna, Fine and Fredi

PUBLICATIONS AND CONFERENCE CONTRIBUTIONS

Publications

Scheper, C., Wensch-Dorendorf, M., Yin, T., Dressel, H., Swalve, H. H., König, S. 2016. Evaluation of breeding strategies for polledness in dairy cattle using a newly developed simulation framework for quantitative and Mendelian traits. Genet Sel Evol 48, 50. https://doi.org/10.1186/s12711-016-0228-7

Scheper, C., Emmerling, R., Götz, KU., König, S. 2021. A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle. Genet Sel Evol 53, 60. https://doi.org/10.1186/s12711-021-00652-z

Conference contributions

Scheper, C., König, S. 2014. Ansätze zur Evaluierung von Selektionsstrategien für das Merkmal Hornlosigkeit mittels Simulationsstudien. Presentation at the DGFZ working group - „Genetisch-Statistische Methoden“ meeting. 01.10.2014, Bovenden, Germany

Scheper, C., Yin, T., König, S. 2015. Inclusion of polled geno- or phenotype into breeding goals: Impact on genetic gain and inbreeding. Presentation at the 66th annual EAAP meeting in Warszaw 31.08–04.09.2015.

Scheper, C., Dressel, H., Yin, T., Wensch-Dorendorf, M., Swalve, H. H., König, S. 2015. Ein Programmierungsansatz zur Evaluierung von Selektionsstrategien für das Merkmal Hornlosigkeit beim Rind. Presentation at the DGfZ und GfT meeting 16.-17. September 2015 in Berlin, Germany.

Scheper, C., Emmerling, R., Götz, K.-U., Swalve, H.H., König, S. 2016. Genetische Varianz- und Kovarianzkomponenten für qualitative Merkmale am Beispiel Hornlosigkeit beim Deutschen Fleckvieh. Presentation at the DGfZ und GfT meeting, 20.-21. September Hannover, Germany

Scheper, C. and König, S. 2017. Using a MQTL matrix to test for pleiotropic effects of Mendelian trait loci on quantitative traits. Presentation at the 68th annual EAAP meeting in Talinn 28.08.-01.09.2015.

Scheper, C., Emmerling, R., Götz, K.-U., Swalve, H.H., König, S. 2017. Prüfung pleiotroper QTL-Effekte des „POLLED“-locus auf Leistungs- und Fruchtbarkeitsmerkmale beim Deutschen Fleckvieh. Presentation at the DGfZ und GfT meeting, 20.-21. September 2017 Hohenheim, Germany

Scheper, C. and König, S. 2018. Selektionssignaturen auf BTA1 im Vergleich hornloser und behornter Holstein Besamungsbullen auf Basis von SNP-Daten. Presentation at the DGfZ und GfT meeting, 12.-13.September 2018 in Bonn, Germany

SUMMARY

Breeding for genetic polledness represents one of the alternatives to routine dehorning in dairy and beef cattle farming. Due to the increasingly critical perception of dehorning in the ongoing animal welfare debate, the search for alternatives is becoming more urgent. Since most dairy and dual-purpose cattle raised in Germany today are currently dehorned, intensifying the breeding for polledness is the most pragmatic alternative. As a result, polled breeding in the largest dairy and dual-purpose breeds in Germany, German Holsteins and Simmental, is experiencing increased demand and ongoing integration into existing breeding programs. Starting from very low allele frequencies, the proportion of polled animals has increased sharply over the past 10 years. Traditionally existing breeding value deficits of polled insemination bulls have decreased during this time.

While most economically relevant traits in cattle breeding are quantitative traits, polledness is one of the few relevant qualitative traits. The inheritance of the trait at the polled locus on chromosome 1 of the bovine genome was described early on and has been intensively investigated and increasingly elucidated by molecular genetics over the past 20 years. However, recent research results show an increasingly complex picture with different structural or allelic variants at the locus. In addition, the precise influence of the known variants at the polled locus on physiological processes of horn growth are still unclear. In addition, there are further phenotypic phenomena such as the occurrence of scurs and double rows of cilia, which contribute to the complexity of the trait.

The overall objective of this thesis was to investigate the phenotypic and genetic relationships between the trait polledness and other traits in light of the intensification of hornless breeding in the German Holstein and Simmental breeds. Besides performance traits, which are strongly relevant in current breeding programs, available functional traits were also considered. Classical quantitative genetic models (chapters 2 and 3) as well as methods adapted to genomic data (chapter 4) were used. In addition, stochastic simulation studies (chapter 5) were performed to evaluate selection strategies for intensification of polled breeding and their possible consequences on important parameters of breeding success.

In chapters 2 and 3, single-gene effects of the polled locus were estimated for secondary traits in the Simmental breed with the application of adapted quantitative-genetic models, which take the monogenic structure of the trait into account. In both univariate and bivariate models, possible pleiotropic effects of the polled locus could be investigated in this way. While no direct effects of the polled locus were found for the majority of the studied performance and functional traits, a significant effect of the polled locus was found for the trait milk protein percentage. However, the genetic correlation estimated from bivariate models does not indicate an unfavorable genetic

relationship in this regard.

On the genomic level, the results obtained from the quantitative genetic analyses could be further confirmed (chapter 4). Using genomic data from the German Holstein breed, a comparison of polled and horned insemination bulls revealed selection signatures mainly in the proximal region of chromosome 1 near the polled locus. Significant associations to secondary traits could not be detected in genome-wide association studies based on breeding values.

Within the framework of a stochastic simulation study (chapter 5), it could be shown that a rapid intensive selection for polledness is associated with substantial losses in genetic gain if the existing initial status-quo of the polled population is taken into account. Furthermore, due to the low initial allele frequencies, a completely polled breeding population can realistically only be achieved after > 5-10 generations. Regarding possible selection strategies, it could be shown that a sex-specific differentiated selection for the hornless phenotype on the female side and specifically for the hornless genotype (targeted homozygosity) on the male side seems to be promising. In summary, an intensification of polled breeding should strive for a balance between increasing the allele frequency of polledness while securing the genetic progress by adjusting the selection intensity and strategy.

The present work contributes to a further in-depth study of polledness in cattle. While the molecular genetic structure has already been extensively studied, there has been a lack of work focusing on possible secondary effects of increased breeding for polledness. Even though the progressive intensification of polled breeding has just begun, some important questions concerning this process could be investigated on an already larger data base. In this sense, in the summary of the results of this thesis, indications that significant (negative) secondary effects of breeding for polledness are not to be expected, become stronger.

ZUSAMMENFASSUNG

Die Zucht auf genetische Hornlosigkeit stellt eine der Alternativen zur routinemäßigen Enthornung in der Milch- und Fleischrinderhaltung dar. Aufgrund der zunehmend kritischen Wahrnehmung der Enthornung im Zuge der anhaltenden Tierwohldebatten wird die Suche nach Alternativen drängender. Da der Großteil der heute in Deutschland gehaltenen Milch- und Zweinutzungsrinder aktuell hornlos gehalten wird, ist die Intensivierung der Zucht auf genetische Hornlosigkeit die pragmatischste Alternative. Infolgedessen erfährt die Hornloszucht in den zahlenmäßig größten Milch- und Zweinutzungsrassen in Deutschland, Deutsche Holsteins und Fleckvieh, eine verstärkte Nachfrage und Integration in die bestehenden Zuchtprogramme. Ausgehend von sehr niedrigen Allelfrequenzen hat sich der Anteil hornloser Tiere in den genannten Populationen in den letzten 10 Jahren stark gesteigert. Traditionell bestehende Zuchtwertdefizite hornloser Besamungsbullen haben sich in dieser Zeit verringert.

Während der Großteil der heute züchterisch und ökonomisch relevanten Merkmale in der Rinderzucht quantitativ geprägt ist, ist die genetische Hornlosigkeit eines der wenigen züchterisch relevanten qualitativen Merkmale. Der Erbgang des Merkmals am Hornlos-Locus auf Chromosom 1 des Rindergenoms wurde bereits frühzeitig beschrieben und in den letzten 20 Jahren molekulargenetisch intensiv untersucht und zunehmend aufgeklärt. Die jüngsten Forschungsergebnisse zeigen aber auch ein zunehmend komplexes Bild mit verschiedenen Struktur- bzw. Allelvarianten am Hornlos-Locus und die Beeinflussung der physiologischen Prozesse hin zu einem Ausbleiben des Hornwachstums sind weiterhin ungeklärt. Hinzu kommen weitergehende phänotypische Phänomene wie bspw. das Auftreten von Wackelhörnern und doppelten Wimpernreihen, die zur Komplexität des Merkmals beitragen.

Übergeordnetes Ziel der vorliegenden Arbeit war es vor dem Hintergrund der Intensivierung der Hornloszucht in den Rassen Deutsche Holsteins und Fleckvieh, die phänotypischen und genetischen Beziehungen zwischen dem Merkmal Hornlosigkeit und weiteren Merkmalen gezielt zu untersuchen. Neben den in den Zuchtprogrammen stark relevanten Leistungsmerkmalen wurden auch funktionale Merkmale betrachtet. Methodisch wurden sowohl klassische quantitativ-genetische Verfahren (Kapitel 2 und 3), als auch an genomische Daten angepasste Verfahren (Kapitel 4), verwendet. Darüber hinaus wurden im Rahmen von stochastischen Simulationsstudien (Kapitel 5) Selektionsstrategien für eine Intensivierung der Hornloszucht und deren mögliche Folgen auf wichtige Parameter des Zuchterfolgs evaluiert.

In Kapitel 2 und 3 konnte mit der Anwendung angepasster quantitativ-genetischer Modelle, die die monogene Struktur des Merkmals berücksichtigen, Einzelgen-Effekte des Hornlos-Locus für Sekundarmerkmale in der Rasse Fleckvieh geschätzt werden. Sowohl in univariaten als auch

bivariaten Modellen konnten auf diesem Weg mögliche pleiotrope Effekte des Hornlos-Locus untersucht werden. Während für den Großteil der etablierten Leistungsmerkmale und funktionalen Merkmale keine direkten Effekte des Hornlos-Locus gefunden wurden, ergab sich für das Merkmal Milcheiweißgehalt ein signifikanter Effekt des Hornlos-Locus. Die aus bivariaten Modellen geschätzte genetische Korrelation weist hierbei aber nicht auf eine züchterisch ungünstige genetische Beziehung hin.

Auf genomischer Ebene konnte das aus den quantitativ-genetischen Analysen gewonnene Bild weitergehend bestätigt werden (Kapitel 4). Anhand genomischer Daten aus der Rasse Deutsche Holsteins konnten im Vergleich hornloser und horntragender Besamungsbullen Selektionssignaturen vor allem im proximalen Bereich von Chromosom 1 nahe des Hornlos-Locus gezeigt werden. Signifikante Assoziationen zu Sekundärmerkmalen konnten im Rahmen von genomweiten Assoziationsstudien auf Zuchtwertbasis nicht nachgewiesen werden.

Im Rahmen einer umfangreichen stochastischen Simulationsstudie (Kapitel 5) konnte gezeigt werden, dass eine schnelle intensive Selektion auf Hornlosigkeit mit substanziellen Zuchtwertverlusten verbunden ist, wenn die bestehende züchterische Ausgangssituation der Hornlospopulation berücksichtigt wird. Zudem ist infolge der geringen Ausgansfrequenzen eine gänzlich hornlose Zuchtpopulation realistischerweise erst nach > 5-10 Generationen zu erreichen. Mit Blick auf mögliche Selektionsstrategien konnte gezeigt werden, dass eine geschlechtsspezifisch differenzierte Selektion auf den Hornlos-Phänotyp auf weiblicher Seite und gezielt auf Hornlos-Genotyp (angestrebte Homozygotie) auf männlicher Seite empfehlenswert erscheint. In der Zusammenschau sollte eine Intensivierung der Hornloszucht einen mittelfristigen Ausgleich zwischen der Erhöhung der Allelfrequenz der Hornlosigkeit bei Sicherung des Zuchtfortschrittes über eine Anpassung der Selektionsintensität und -strategie anstreben.

Die vorliegende Arbeit leistet einen Beitrag zur weiteren und eingehenden Untersuchung der genetischen Hornlosigkeit beim Rind. Während die molekulargenetische Struktur bereits umfassend untersucht wurde, fehlte es bislang an Arbeiten, die auf mögliche Sekundärwirkungen einer verstärkten Zucht auf Hornlosigkeit fokussieren. Auch wenn die Entwicklung einer fortschreitenden Intensivierung der Hornloszucht zeitlich immer noch relativ am Anfang steht, konnten einige wichtige Fragen zu diesem Prozess auf einer bereits größeren Datenbasis untersucht werden. In diesem Sinne verdichten sich in der Zusammenfassung der Ergebnisse die Hinweise darauf, das signifikante (negative) Sekundäreffekte der Zucht auf Hornlosigkeit nicht zu erwarten sind.

TABLE OF CONTENTS

ABBREVIATIONS

AI	artificial insemination
BP	base pair position (on chromosome)
BRS	Bundesverband Rind und Schwein
BTA	bos taurus autosome
BWH	Black and White Holstein
DFS	days to first service
DNA	desoxyribonucleic acid
DO	days open
EBV	estimated breeding values
F%	fat percentage
FY	fat yield
GMOs	genetically modified organisms
GWAS	genome-wide association studies
GZW	total net merit in German Simmental cattle
h^2	heritability
IBD	identity by descent
LD	linkage disequillibrium
LRT	likelihood ratio test
MAF	minor allele frequency
MAS	marker assisted selection
MY	milk yield
NRR	non return rate
OGC	optimum genetic contribution
OMIA	Online Mendelian Inheritance in animals
P%	protein percentage
p(P)	polled allele frequency
PC	Celtic polled variant
PCA	Principal component analysis
PF	Friesian polled variant
pp	homozygous horned genotype
Pp	heterozygous polled genotype
PP	homozygous polled genotype
PY	protein yield

QQ	quantile-quantile
QTL	quantitaitve trait loci
RH	Red Holstein
RNA	ribonucleic acid
RZG	total net merit in German Holstein cattle
SCS	somatic cell score
SD	standard deviation
SE	standard error
SNP	Single Nucleotide Polymorphism
VC	variance components
xp-EHH	accross population extended haplotype homozygosit

CHAPTER 1

General Introduction

Polledness is one of only a few examples of favorable or beneficial genetic characteristics with Mendelian inheritance in cattle breeding. While managing a growing number of detrimental genetic characteristics is a complex challenge in current dairy cattle breeding (Cole, 2015; Segelke et al., 2016), the integration of polledness in breeding goals in the major German cattle breeds Holstein-Friesian and Simmental appears to be a rather simple and manageable task at first sight. With the general availability of valid direct or SNP-based gene tests to securely identify polled animals and a growing demand for and supply of polled sires by farmers, there is a consequent increase in the allele frequency of polledness due to active selection. A closer look however reveals several unanswered questions on potential side effects and sustainable strategies concerning enhanced breeding for polledness. This is especially important if the cattle breeding sector ultimately targets the fixation of the causative variants for polledness. For example, potentially functional genomic regions in strong linkage with the causative alleles will therefore also be fixated implying at least a risk for secondary effects. Considering the frequent, mainly anecdotal, reports of disadvantages or impairments of polled animals in the past which are still repeated until today, a thorough scientific review is imperative. The following thesis tries to contribute to this ongoing review based on data from German Holstein and Simmental cattle and focusses on the evaluation of potential side-effects of selection for polledness and suitable as well as sustainable breeding strategies.

Polledness in the context of routine non-curative procedures in livestock management.

Non-curative procedures in livestock such as dehorning or disbudding in cattle, tail docking in sheep and beak trimming in chicken represent long-standing routine procedures in livestock management typically used to prevent welfare and health related problems in intensive production conditions (e.g. lesions due to horn blows). As of now these procedures are mainly performed very early in the animal's lifespan and are regulated by animal welfare legislation in many countries (Adcock, 2021). It is well established that the mentioned procedures do have direct adverse effects (i.e. pain) of varying degree related to the performed methods (Faulkner and Weary, 2000; Morisse et al., 1995; Stafford and Mellor, 2005). Further secondary impairing effects for example on the animals behaviour and other traits are also well documented (Morisse et al., 1995; Graf and Senn, 1999; Lutz et al., 2019) including potential long-term phenotypic effects (Adcock and Tucker, 2020).

Although adverse effects of non-curative procedures can be limited by improving the used methods

(e.g. suitable analgesia treatment and/or anti-inflammatory medication), they cannot be fully and sustainably avoided. Given the growing concern and ongoing discussion of farm animal welfare related topics in Germany and other western societies, it is very likely that novel animal welfare regulations will prohibit additional, if not all, non-curative procedures in livestock in the future. Hence, research initiatives aiming at effective and sustainable alternatives to the mentioned procedures cover a broad palette of topics from livestock housing, management and handling to breeding and genetics. In this regard polledness in cattle is one of only a few examples, where breeding and selection can play a major role in consequently superseding the non-curative procedures of dehorning and disbudding in dairy and beef calves.

Evolutionary and phylogenetic aspects on the development of headgear and polledness in cattle

The development of horns in bovids (i.e. cattle, goat and sheep) reflects the advantages of skull attachments that can be used as weapons or tools as well as social organs during natural selection (Simon et al., 2022). From a phylo- and molecular genetic point of view, the evolution of headgear and consequently horns is a complex and challenging research question due to the complexity of the underlying molecular and cellular mechanisms and likely involves hundreds of genes (Allais-Bonnet et al., 2021). This appears to contrast with the ostensibly simple genetic structure of the trait polledness with Mendelian inheritance at a single locus at first sight. However, recent research from the phenotypic as well as genetic perspective shed further light on the fact that the trait polledness, as we know it today in context of modern cattle breeding, is most likely only a part of a much more complex trait "regulation of horn development" rather than the classical binary trait model (e.g. polled or horned) (Gehrke et al., 2020a; Hennig et al., 2022a; Hennig et al., 2022b). Although it is rather likely that the initially domesticated cattle population was entirely or almost entirely horned, there are also indicators that show that polled cattle was not fully uncommon also in the early phases of domestication and probably also before (Schafberg and Swalve, 2015). In addition, polled animals are common but often less frequent in most domesticated bovids apart from cattle (Simon et al., 2022).

History and status-quo of breeding for polledness in German Holstein-Friesian and Simmental cattle

Breeding for polled animals has a longer tradition in beef compared to dairy cattle in Germany as well as worldwide (Schafberg and Swalve, 2015; Windig et al., 2015; Rowan et al; Windig and Eggen, 2009; Randhawa et al., 2021). In Germany, the major dairy cattle populations in the Holstein-Friesian and Simmental breeds are still characterized by rather small proportions of polled animals (see Figure 1 a). However, the last decade is marked by a significant increase of the polled allele frequency, the supply of heterozygous as well as homozygous polled sires (see-Figure 1 b) and

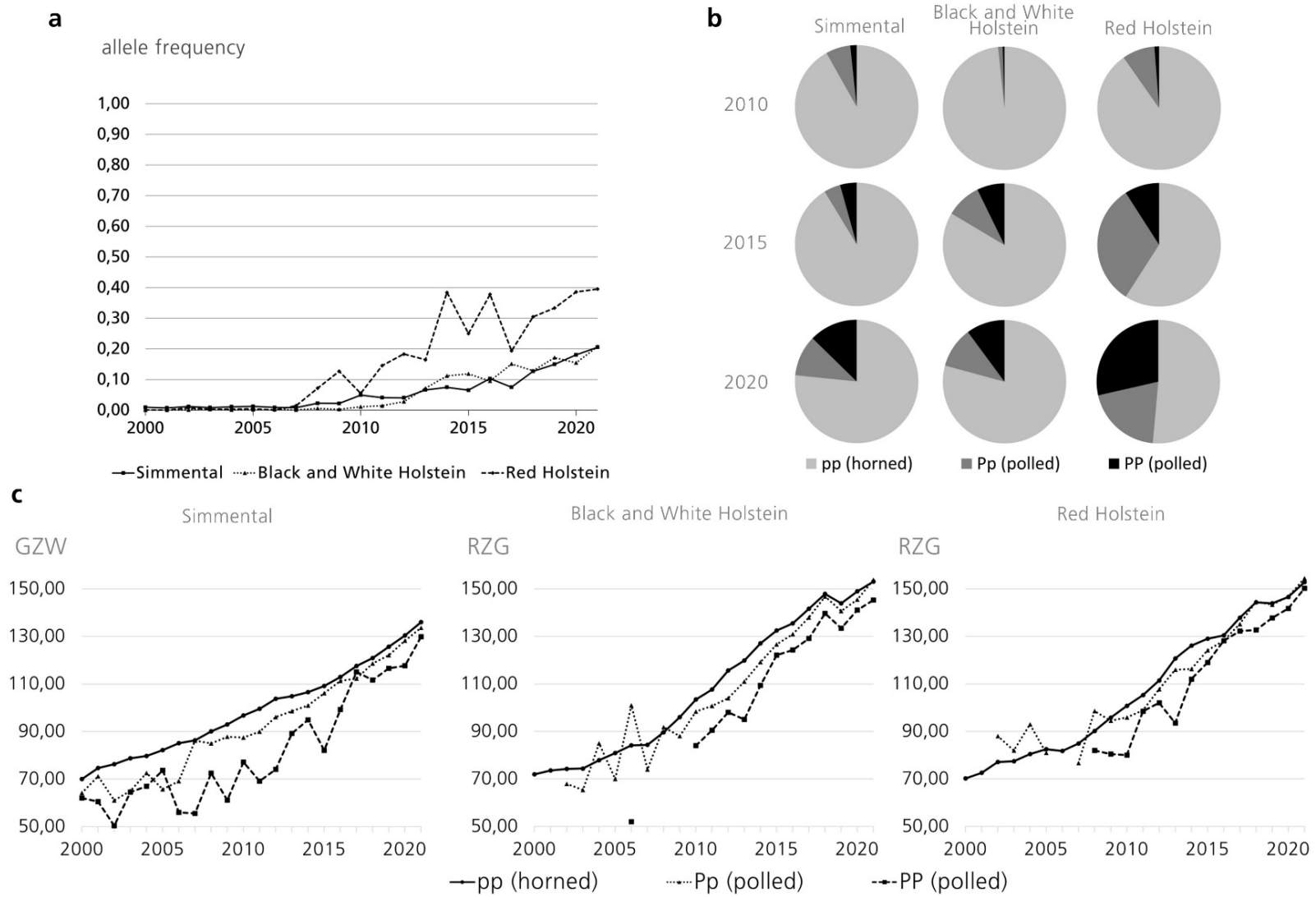

Figure 1 Descriptive statistics for the development of the polled allele frequency (a), percentage of heterozygous and homozygous polled bulls (b) and the genetic level of polled bulls compared to horned bulls (c) in the population of German artificial insemination sires in Simmental and Holstein cattle. GZW = total net merit in German Simmental cattle; RZG = total net merit in German Holstein cattle.

sharply growing demand for these sires from farmers. Historically, breeding for polledness in the German Holstein and Simmental populations was founded on the initiative of a limited number of motivated breeders focusing on that trait (Schafberg and Swalve, 2015; Götz et al., 2015; Specht, 2008). Only in the last decade, a growing number of breeding organizations have identified the potential to specifically market polled sires leading to rising numbers of available polled Holstein bulls worldwide and in Germany. This development was particularly noticeable in the rather small Red Holstein population, due to the early availability of competitive bull sires (e.g. the "Lawn Boy" effect, see Figure 1 a and c). In general, however, due to the rather limited number of founders, groups of polled Holstein and Simmental individuals displayed significantly lower average breeding values mostly in performance traits in the past, and higher average kinship in comparison to the horned population (Spurlock et al., 2014; Segelke et al., 2013; Windig et al., 2015; Windig and Eggen, 2009). This deficit in genetic merit of especially homozygous polled sires across breeds (see Figure 1 c) in the past lessened in the younger birth year cohorts of bulls, especially in Red Holstein. The demand for polled sires by farmers across breeds is continuously growing. This is reflected by recent reports of disproportionately high numbers of inseminations with semen from polled sires compared to their share in the total bull population. For example, Krogmeier and Luntz reported that 38.5 percent of all inseminations in 2018 with young genomic bulls used semen from polled bulls, while only 10-15 percent of all sires born in 2016 and 2017 were polled (Krogmeier and Luntz, 2020). This was further confirmed also for the total bull population (i.e. genomic and proven) in 2020, with a similar percentage of around 38% of inseminations with polled bulls, although only 10-25% of all registered bulls per birth year were polled (LfL, 2023). Similar values were also reported recently for both Black and White Holstein and Red Holstein cattle (BRS, 2022). Although reviewed scientific studies on inseminations are not available, the presented values do offer a good representation of the present situation with only a risk of considerable marketing bias in reporting.

Molecular genetic background of polledness in cattle

Bovine polledness is a Mendelian trait (OMIA 000483-9913) that was one of the first animal specific Mendelian characteristics described by Bateson and Saunders in 1902 (Bateson and Saunders, 1902) upon the rediscovery and acceptance of Mendel's Laws. The trait is controlled by one locus located at the proximal end of Bos taurus autosome 1 (BTA1). There are currently four known dominant allelic variants that cause polled phenotypes and scurs (loosely attached hornlike formations at the skull) and the recessive wild-type variant causes the horned phenotype in cattle (Long and Gregory, 1978; Medugorac et al., 2017; Randhawa et al., 2020; Stafuzza et al., 2018; Medugorac et al., 2012). The identified "Celtic" (PC, predominant in e.g. Simmental, Limousin, Charolais) and "Friesian" (PF, predominant in e.g. Holstein and Jersey) variants are predominant

in polled animals from European dairy and dual-purpose cattle breeds (Medugorac et al., 2012; Rothammer et al., 2014). Valid direct and single nucleotide polymorphisms (SNP) based gene-tests are currently available in almost all cattle breeds kept in Germany (Gene Control GmbH, 2023). The Mendelian inheritance pattern at the polled locus is proven (Medugorac et al., 2012), however older inheritance models proposing a limited number of additional loci next to the polled locus could not be validated considering the complex diversity of phenotypes related to polledness (White and Ibsen, 1936; Long and Gregory, 1978). Considering the position of all known polled variants in DNA segments not known to be coding or regulatory (Allais-Bonnet et al., 2013; Mariasegaram et al., 2010), the actual genetic and physiological effect remains unclear. Recent research in context of efforts to utilize gene-editing methods in breeding polled animals have substantiated the potential regulatory function likely due to strong linkage between the polled variants and Long noncoding RNA on BTA1 in close proximity to the polled locus (Aldersey et al., 2020; Hennig et al., 2022a).

Recently, an oligogenic model of inheritance was discussed to explain the remaining complexity of horn-related phenotypes, including scurs and polledness, in which the polled locus has a presumed epistatic suppressive function (Gehrke et al., 2020a). In this regard, the polled phenotype is part of a much more complex discrete or even continuous distribution of the trait "horn growth", where the phenotypic states "fully polled" and "fully horned" mark the tails of the distribution.

Complementing the picture of the rather complex genetic and phenotypic heterogeneity of the trait polledness (or likely better "horn growth" as pointed out before) are repeated reports of new mutations or variants (including de-novo mutations) leading to polled phenotypes in animals that do carry the wild-type variant at the polled locus ("i.e. normally horned individuals") (Gehrke et al., 2020b; Capitan et al., 2011; Capitan et al., 2012). These cited cases also show other highly detrimental phenotypes that do not affect polled animals carrying one of the four known polled variants. Additionally, there are also frequent reports of non-detrimental phenotypes observed only in polled animals (i.e. animals carrying one of the four polled variants), for example double eyelashes (Aldersey et al., 2020). The variety and complexity of the reports may offer an explanation why concerns for side-effects of breeding for polledness are persistent, although there is clear evidence from scientific studies as well as practical experience that polled animals are not generally impaired or necessarily carry additional detrimental traits due to their polledness.

Outlook for polled breeding in German Holstein-Friesian and Simmental cattle

Considering the well documented animal-welfare related negative effects of routine dehorning, the public perception of this practice is increasingly negative (Morisse et al., 1995; Stafford and Mellor, 2011; Stafford and Mellor, 2005). Although currently used dehorning procedures aim on animal welfare improvements (Faulkner and Weary, 2000; Graf and Senn, 1999; Heinrich et al., 2010;

Guatteo et al., 2012) it is clear that further steps to exacerbate the German animal welfare legislation are likely and will almost certainly include a prohibition of routine dehorning. Therefore, there is an increasing demand for sustainable alternatives to current dehorning practices. Even though keeping horned cattle appears to be the simplest solution to this problem, the proportion of dehorned or hornless individuals compared to horned individuals in German Holstein and Simmental cattle clearly shows that the majority of farmers prefer hornless cattle (Krogmeier and Luntz, 2020; Cozzi et al., 2015). Among the variety of reasons why farmers prefer to dehorn their cattle, safety concerns surrounding the potential use of horns as weapons against other animals or handlers stand out (Cozzi et al., 2015). Hence, selection for polled animals is likely the most practicable alternative to dehorning based on farmers preferences and current and future animal welfare legislation.

It is already apparent that there is a continuously growing demand for polled bulls in Germany and that the German Holstein and Simmental breeding associations have adapted to this demand by intensifying their selection for polledness in recent years (Segelke et al., 2016). In addition, there is growing evidence that breeding for polledness does not lead to significant side effects in performance and functional traits (Cole et al., 2016; Scheper et al., 2021).

As a sidenote, recently novel gene-editing methods were used to achieve an intraspecies polled allele introgression in cattle to produce two living homozygous polled individuals (Carlson et al., 2016). As of today, implementation of gene-editing methods in European dairy and beef cattle breeding programs is highly regulated by current legislation.

Scientific results from quantitative genetic and genomic studies on effects of breeding for polledness

Traditionally, polled animals and especially polled artificial insemination (AI) bulls were associated with lower phenotypic performances and breeding values in production and functional (especially fertility) traits across beef and dairy breeds. In the literature, lower average breeding values in production traits in homo- and heterozygous polled animals have been reported for various breeds (Cole and Null, 2019; Dressel et al. 2016; Gehrke et al., 2016; Gehrke, 2020; Götz et al., 2015; Lamminger et al., 2000; Frisch et al., 1980; Goonewardene et al., 1999a; Goonewardene et al., 1999b). In contrast, for reproduction traits, mainly neutral or even positive effects were reported (Cole and Null, 2019; Gehrke et al., 2016; Lamminger et al., 2000). Some authors attribute the breeding value inferiority in dairy or dual-purpose breeds for example in polled German Simmental to the introgression of polled alleles from beef populations and therefore to genetic drift (i.e. a random increase of less favorable alleles) rather than to pleiotropic effects (Götz et al., 2015). Other authors suggest that the initial performance inferiority of polled individuals might be due to a persisting selection advantage of their horned counterparts (Windig et al., 2015).

Simulation studies on breeding programs for polledness

Until today there is only a limited number of simulation studies focusing on selection for polledness or breeding programs for polledness. The available studies focused on conventional breeding as well as genomic selection schemes to increase the polled allele frequency and phenotype dairy populations, while also evaluating negative impacts on inbreeding and genetic merits (Spurlock et al., 2014; Gaspa et al., 2015; Windig et al., 2015; Segelke et al., 2016). The results are highly consistent in the conclusion that a fast transition to a completely polled population substantially reduces the rate of genetic gain compared to reference scenarios without selection for polledness. However, they also show that there are suitable strategies that limit the loss of genetic gain while managing increasing inbreeding and still achieving a reasonably fast gain in polled allele frequency. The most recent simulation studies focus on the potential of novel gene-editing methods to overcome the disadvantages of classical introgression approaches as reflected in the aforementioned older simulation studies (Mueller et al., 2019; Mueller et al., 2021; Bastiaansen et al., 2018). Compared to a classic introgression approach the results suggest an advantage of gene-editing methods from a breeding perspective regarding genetic gain, inbreeding and the targeted increase in polled allele frequency.

Suitable methods to study the trait polledness

Quantitative genetic methods

Associations between genetic traits, regardless of the trait architectures, are due to either pleiotropy or linkage (Falconer & Mackay, 1996). The estimation of differences in means of genotype groups at single marker loci is a traditional method to assess pleiotropic effects and associations, often using controlled trial designs and model organisms (Falconer & Mackay, 1996; Sax, 1923). These methods were initially also utilized when studying differences between polled and horned cattle (Cole & Null, 2019; Dressel et al., 2016; Gehrke et al., 2016; Götz et al., 2015; Lamminger et al., 2000). However, using error prone phenotypic field data and an initial lack of a valid gene test to identify polled animals potentially introduces bias and error that cannot be validly controlled using these traditional methods. Hence, methods evaluating differences in means for genotype groups lead to incoherent results and no clear answer if and to what extent pleiotropic effects exist.

More advanced variance component (VC) estimation methods based on additive linear models are flexible enough to incorporate single or multiple marker genotype data in order to infer QTL effects at a given chromosomal segment or position (George, Visscher, & Haley, 2000; van Arendonk et al., 1998). They allow modelling random QTL allele effects using for example marker-based gametic or numerator relationship matrices next to random polygenic effects (e.g. based on pedigree information) (e.g. van Arendonk, Tier, & Kinghorn, 1994). In Principle, this enables a separate estimation of additive variances for major genes or QTLs (based on single marker data) and all

other additive polygenic effects. Comparable VC estimation methods to separate polygenic and single-marker based QTL effects were developed mainly in the pre-genomic selection era in the initial search for suitable Marker-assisted-selection (MAS) approaches. They have also been extended to multivariate models to increase the power of QTL detection (George et al., 2000; Sørensen, Lund, Guldbrandtsen, Jensen, & Sorensen, 2003). In principle, these methods should also be suitable to utilize for the special case of Mendelian traits, however, there are no reports in the literature to my knowledge that show their suitability in practice.

Methods utilizing genomic marker data

Methods to detect diverging selection are traditionally based on an evaluation of the allele frequency spectrum across the genome. Causal loci influenced by selection or loci that are in linkage with causal variants are expected to have significant allele frequency differences compared to loci not influenced by selection. Various methods have been developed to map these differences based on different marker information, e.g. SNP as markers, typically by comparing divergent populations (Qanbari & Simianer, 2014; Sabeti et al., 2006). Traditional methods such as the Fst value estimation (Weir & Cockerham, 1984) detect mainly long-term selection events such as evolutionary changes. Methods focusing on short-term selection events focus on the evaluation of the length and structure of haplotypes. A widely used and well-established method is the analysis or comparison of "extended haplotype homozygosity" between populations (xp-EHH) and its derivates (Sabeti et al., 2002; Sabeti et al., 2006, Tang, Thornton, & Stoneking, 2007). The mentioned methods have also been successfully used to study divergently selected sub-populations comparable to the polled and horned sub-populations in cattle (e.g. (Avila et al., 2018)).
The methodology for genome-wide association studies (GWAS) has evolved significantly over time. In the past decade rather complex statistical methods based on linear mixed model methodology have gained popularity compared to the classical approaches in cohort or case-control designs using relatively simple statistical tests. Now widely used program packages such as GCTA (Yang et al., 2011) enable association studies with several thousand SNP markers while controlling the population stratification by considering genomic relationships among all studied individuals.

Simulation studies

Simulation methods are well established tools to evaluate long-term selection responses. They are widely used to study evolutionary as well as anthropogenic processes, and have been consistently adapted parallel to the development of genomic selection methods (e.g. Daetwyler et al, 2013; Hoban, Bertorelle, & Gaggiotti, 2011). Deterministic simulations use an equation based prediction of average genetic gain and average inbreeding development, typically per generation, to evaluate different selection strategies and breeding program structures; they can be adapted to highly complex breeding schemes and are very resource efficient in these situations. In contrast, stochastic

simulation methods explicitly simulate individuals in a breeding program, including phenotypes, genotypes and mating relationships (Sargolzaei & Schenkel, 2009). They offer a lot of flexibility to simulate a wide range of genetic and genomic architectures and population structures in livestock. Considering multiple traits in a simulation study, regardless if deterministic or stochastic simulation techniques are used, requires genetic (co)variance components for both traits. Hence, both strategies rely on assumptions if no estimated values from real datasets are available.

Outline and research questions in this thesis

The overall aim of this thesis was to investigate potential side effects of (i.e. focussing on pleiotropy and selection signatures) and suitable breeding strategies for polledness based on data from German Holstein and Simmental cattle.

Chapter 2 focusses on the evaluation of suitable methods to estimate variance components for the trait polledness based on simulated and real data from German Simmental cattle.

Chapter 3 aims on the evaluation of pleiotropic or linked QTL effects of the Mendelian polled locus on production and reproduction traits in German Simmental cattle using univariate and bivariate variance component estimation.

Chapter 4 assesses long-term and recent selection signatures and chromosome-wide associations with performance and functional traits comparing horned and polled Holstein subpopulations in a GWAS framework utilizing chromosome- and genomewide SNP markers.

Chapter 5 utilizes the simultaneous selection for a quantitative and qualitative trait using a variety of selection strategies for the polled trait in a stochastic simulation framework to study the effects of different selection schemes on the polled allele frequency, genetic gain and inbreeding in a long term perspective.

References

Adcock, S. J. J. 2021. Early Life Painful Procedures: Long-Term Consequences and Implications for Farm Animal Welfare. Front. Anim. Sci. 2:78. https://doi.org/10.3389/fanim.2021.759522.

Adcock, S. J. J., and C. B. Tucker. 2020. The effect of early burn injury on sensitivity to future painful stimuli in dairy heifers. PLOS ONE 15(6):e0233711. https://doi.org/10.1371/journal.pone.0233711.

Aldersey, J. E., T. S. Sonstegard, J. L. Williams, and C. D. K. Bottema. 2020. Understanding the effects of the bovine POLLED variants. Animal Genetics 51(2):166–176. https://doi.org/10.1111/age.12915.

Allais-Bonnet, A., C. Grohs, I. Medugorac, S. Krebs, A. Djari, A. Graf, S. Fritz, D. Seichter, A. Baur, I. Russ, S. Bouet, S. Rothammer, P. Wahlberg, D. Esquerré, C. Hoze, M. Boussaha, B. Weiss, D. Thépot, M.-N. Fouilloux, M.-N. Rossignol, E. van Marle-Köster, G. E. Hreiðarsdóttir, S. Barbey, D. Dozias, E. Cobo, P. Reversé, O. Catros, J.-L. Marchand, P. Soulas, P. Roy, B. Marquant-Leguienne, D. Le Bourhis, L. Clément, L. Salas-Cortes, E. Venot, M. Pannetier, F. Phocas, C. Klopp, D. Rocha, M. Fouchet, L. Journaux, C. Bernard-Capel, C. Ponsart, A. Eggen, H. Blum, Y. Gallard, D. Boichard, E. Pailhoux, and A. Capitan. 2013. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. PLOS ONE 8(5):e63512. https://doi.org/10.1371/journal.pone.0063512.

Allais-Bonnet, A., A. Hintermann, M.-C. Deloche, R. Cornette, P. Bardou, M. Naval-Sanchez, A. Pinton, A. Haruda, C. Grohs, J. Zakany, D. Bigi, I. Medugorac, O. Putelat, O. Greyvenstein, T. Hadfield, S. B. Jemaa, G. Bunevski, F. Menzi, N. Hirter, J. M. Paris, J. Hedges, I. Palhiere, R. Rupp, J. A. Lenstra, L. Gidney, J. Lesur, R. Schafberg, M. Stache, M.-D. Wandhammer, R.-M. Arbogast, C. Guintard, A. Blin, A. Boukadiri, J. Rivière, D. Esquerré, C. Donnadieu, C. Danchin-Burge, C. M. Reich, D. G. Riley, E. van Marle-Koster, N. Cockett, B. J. Hayes, C. Drögemüller, J. Kijas, E. Pailhoux, G. Tosser-Klopp, D. Duboule, and A. Capitan. 2021. Analysis of Polycerate Mutants Reveals the Evolutionary Co-option of HOXD1 for Horn Patterning in Bovidae. Molecular biology and evolution 38(6):2260–2272. https://doi.org/10.1093/molbev/msab021.

Avila, F., Mickelson, J.R., Schaefer, R.J., McCue, M.E. 2018. Genome-wide signatures of selection reveal genes associated with performance in American quarter horse subpopulations. Frontiers in Genetics 9:249.

Bastiaansen, J. W. M., H. Bovenhuis, M. A. M. Groenen, H.-J. Megens, and H. A. Mulder. 2018. The impact of genome editing on the introduction of monogenic traits in livestock. Genet Sel Evol 50(1):18. https://doi.org/10.1186/s12711-018-0389-7.

Bateson, W., and E. R. Saunders. 1902. The facts of heredity in the light of Mendel's discovery. Reports to the Evolution Committee of the Royal Society(I. Pp):125–160.

BRS. 2022. Polled Genetics: On the rise (translated from German). Milchrind, 2022:14–15.

Capitan, A., A. Allais-Bonnet, A. Pinton, B. Marquant-Le Guienne, D. Le Bourhis, C. Grohs, S. Bouet, L. Clément, L. Salas-Cortes, E. Venot, S. Chaffaux, B. Weiss, A. Delpeuch, G. Noé, M.-N. Rossignol, S. Barbey, D. Dozias, E. Cobo, H. Barasc, A. Auguste, M. Pannetier, M.-C. Deloche, E. Lhuilier, O. Bouchez, D. Esquerré, G. Salin, C. Klopp, C. Donnadieu, C. Chantry-Darmon, H. Hayes, Y. Gallard, C. Ponsart, D. Boichard, and E. Pailhoux. 2012. A 3.7 Mb deletion encompassing ZEB2 causes a novel polled and multisystemic syndrome in the progeny of a somatic mosaic bull. PloS one 7(11):e49084. https://doi.org/10.1371/journal.pone.0049084.

Capitan, A., C. Grohs, B. Weiss, M.-N. Rossignol, P. Reversé, and A. Eggen. 2011. A newly described bovine type 2 scurs syndrome segregates with a frame-shift mutation in TWIST1. PloS one 6(7):e22242. https://doi.org/10.1371/journal.pone.0022242.

Carlson, D. F., C. A. Lancto, B. Zang, E.-S. Kim, M. Walton, D. Oldeschulte, C. Seabury, T. S. Sonstegard, and S. C. Fahrenkrug. 2016. Production of hornless dairy cattle from genome-edited cell lines. Nat Biotechnol 34(5):479–481. https://doi.org/10.1038/nbt.3560.

Cole, J. B. 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genet Sel Evol 47(1):94. https://doi.org/10.1186/s12711-015-0174-9.

Cole, J. B., and D. J. Null. 2019. Short communication: Phenotypic and genetic effects of the polled haplotype on yield, longevity, and fertility in US Brown Swiss, Holstein, and Jersey cattle. Journal of dairy science 102(9):8247–8250. https://doi.org/10.3168/jds.2019-16530.

Cole, J. B., D. J. Null, and P. M. VanRaden. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. Journal of dairy science 99(9):7274–7288. https://doi.org/10.3168/jds.2015-10777.

Cozzi, G., F. Gottardo, M. Brscic, B. Contiero, N. Irrgang, U. Knierim, O. Pentelescu, J. J. Windig, L. Mirabito, F. Kling Eveillard, A. C. Dockes, I. Veissier, A. Velarde, C. Fuentes, A. Dalmau, and C. Winckler. 2015. Dehorning of cattle in the EU Member States: A quantitative survey of the current practices. Livestock Science 179:4–11. https://doi.org/10.1016/j.livsci.2015.05.011.

Daetwyler, H.D., Calus, M.P., Pong-Wong, R., de Los Campos, G., Hickey, J.M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193(2):347-365. doi:10.1534/genetics.112.147983

Dressel, H., Wensch-Dorendorf, M., and H. H. Swalve. 2016. Genetic relationship between polledness and other relevant traits in Holstein dairy cattle. p. 491. *In* Book of abstracts of the

67th Annual Meeting of the European Federation of Animal Science: Belfast, United Kingdom, 29 August - 2 September 2016. Wageningen Academic Publishers, Wageningen.

Falconer, D.S., Mackay, T.F.C. 1996. Introduction to quantitative genetics. Harlow: Longman

Faulkner, P. M., and D. M. Weary. 2000. Reducing pain after dehorning in dairy calves. Journal of Dairy Science 83(9):2037–2041. https://doi.org/10.3168/jds.S0022-0302(00)75084-3.

Frisch, J. E., H. Nishimura, K. J. Cousins, and H. G. Turner. 1980. The inheritance and effect on production of polledness in four crossbred lines of beef cattle. Anim. Sci. 31(2):119–126. https://doi.org/10.1017/S000335610002434X.

Gaspa, G., R. F. Veerkamp, M. P. Calus, and J. J. Windig. 2015. Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation. Livestock Science 179:86–95. https://doi.org/10.1016/j.livsci.2015.05.020.

Gehrke, L. 2020. Investigations of the genetic architecrure of horns in cattle (bos taurus) for improved breeding strategies for polledness. Phd Thesis. Phd Thesis.

Gehrke, L., D. Seichter, I. Ruß, I. Medugorac, J. Tetens, and G. Thaller. 2016. Impact of the Friesian POLLED Mutation on Milk Production Traits in Holstein Friesian. p. 434. *In* Book of abstracts of the 67th Annual Meeting of the European Federation of Animal Science: Belfast, United Kingdom, 29 August - 2 September 2016. Wageningen Academic Publishers, Wageningen.

Gehrke, L. J., A. Capitan, C. Scheper, S. König, M. Upadhyay, K. Heidrich, I. Russ, D. Seichter, J. Tetens, I. Medugorac, and G. Thaller. 2020a. Are scurs in heterozygous polled (Pp) cattle a complex quantitative trait? Genet Sel Evol 52(1):6. https://doi.org/10.1186/s12711-020-0525-z.

Gehrke, L. J., M. Upadhyay, K. Heidrich, E. Kunz, D. Klaus-Halla, F. Weber, H. Zerbe, D. Seichter, A. Graf, S. Krebs, H. Blum, A. Capitan, G. Thaller, and I. Medugorac. 2020b. A de novo frameshift mutation in ZEB2 causes polledness, abnormal skull shape, small body stature and subfertility in Fleckvieh cattle. Scientific reports 10(1):17032. https://doi.org/10.1038/s41598-020-73807-5.

Gene Control GmbH. 2023. Sonstige Tests. Accessed Feb 11, 2023. https://www.genecontrol.de/leistungen/rinder/sonstige-tests/.

George, A.W., Visscher, P.M., Haley, C.S. 2000. Mapping quantitative trait loci in complex pedigrees: A two-step variance component approach. Genetics 156:2081–92.

Goonewardene, L. A., H. Pang, R. T. Berg, and M. A. Price. 1999°. A comparison of reproductive and growth traits of horned and polled cattle in three synthetic beef lines. Can. J. Anim. Sci. 79(2):123–127. https://doi.org/10.4141/A98-096.

Goonewardene, L. A., M. A. Price, M. F. Liu, R. T. Berg, and C. M. Erichsen. 1999b. A study of growth and carcass traits in dehorned and polled composite bulls. Can. J. Anim. Sci. 79(3):383–385. https://doi.org/10.4141/A98-121.

Götz, K.-U., B. Luntz, J. Robeis, C. Edel, R. Emmerling, J. Buitkamp, H. Anzenberger, and J. Duda. 2015. Polled Fleckvieh (Simmental) cattle – Current state of the breeding program. Livestock Science 179:80–85. https://doi.org/10.1016/j.livsci.2015.05.019.

Graf, B., and M. Senn. 1999. Behavioural and physiological responses of calves to dehorning by heat cauterization with or without local anaesthesia. Applied Animal Behaviour Science 62(2-3):153–171. https://doi.org/10.1016/S0168-1591(98)00218-4.

Guatteo, R., O. Levionnois, D. Fournier, D. Guémené, K. Latouche, C. Leterrier, P. Mormède, A. Prunier, J. Servière, C. Terlouw, and P. Le Neindre. 2012. Minimising pain in farm animals: the 3S approach - 'Suppress, Substitute, Soothe'. Animal an international journal of animal bioscience 6(8):1261–1274. https://doi.org/10.1017/S1751731112000262.

Heinrich, A., T. F. Duffield, K. D. Lissemore, and S. T. Millman. 2010. The effect of meloxicam on behavior and pain sensitivity of dairy calves following cautery dehorning with a local anesthetic. Journal of dairy science 93(6):2450–2457. https://doi.org/10.3168/jds.2009-2813.

Hennig, S. L., B. R. McNabb, J. F. Trott, A. L. van Eenennaam, and J. D. Murray. 2022°. LincRNA#1 knockout alone does not affect polled phenotype in cattle heterozygous for the celtic POLLED allele. Sci Rep 12(1):7627. https://doi.org/10.1038/s41598-022-11669-9.

Hennig, S. L., J. R. Owen, J. C. Lin, B. R. McNabb, A. L. van Eenennaam, and J. D. Murray. 2022b. A deletion at the polled PC locus alone is not sufficient to cause a polled phenotype in cattle. Sci Rep 12(1):2067. https://doi.org/10.1038/s41598-022-06118-6.

Hoban, S., Bertorelle, G., Gaggiotti, O.E. 2012. Computer simulations: tools for population and evolutionary genetics. Nat Rev Genet. 13(2):110-122. Published 2012 Jan 10. doi:10.1038/nrg3130

Krogmeier, D., and B. Luntz. 2020. Untersuchungen zur Entwicklung der Zucht auf natürliche Hornlosigkeit bei Braunvieh und Fleckvieh in Bayern.

Lamminger, A., H. Hamann, G. Röhrmoser, Rosenberger, E., Kräusslich, H., and O. Distl. 2000. Relationships between polledness and traits used in the breeding objectives for German Fleckvieh: (in german, english abstract available). Züchtungskunde 72(5):325–339.

LfL. 2023. Breeding of polled Simmental animals in dariy farms (translated from german). Accessed Feb 11, 2023. https://www.lfl.bayern.de/itz/rind/019880/index.php.

Long, C. R., and K. E. Gregory. 1978. Inheritance of the horned, scurred, and polled condition in cattle. J Hered 69(6):395–400. https://doi.org/10.1093/oxfordjournals.jhered.a108980.

Lutz, J., J.-B. Burla, L. Gygax, B. Wechsler, H. Würbel, and K. Friedli. 2019. Horned and dehorned dairy cows differ in the pattern of agonistic interactions investigated under different space

allowances. Applied Animal Behaviour Science 218:104819. https://doi.org/10.1016/j.applanim.2019.05.008.

Mariasegaram, M., A. Reverter, W. Barris, S. A. Lehnert, B. Dalrymple, and K. Prayaga. 2010. Transcription profiling provides insights into gene pathways involved in horn and scurs development in cattle. BMC Genomics 11(1):370. https://doi.org/10.1186/1471-2164-11-370.

Medugorac, I., A. Graf, C. Grohs, S. Rothammer, Y. Zagdsuren, E. Gladyr, N. Zinovieva, J. Barbieri, D. Seichter, I. Russ, A. Eggen, G. Hellenthal, G. Brem, H. Blum, S. Krebs, and A. Capitan. 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nature genetics 49(3):470–475. https://doi.org/10.1038/ng.3775.

Medugorac, I., D. Seichter, A. Graf, I. Russ, H. Blum, K. H. Göpel, S. Rothammer, M. Förster, and S. Krebs. 2012. Bovine polledness—an autosomal dominant trait with allelic heterogeneity. PLOS ONE 7(6):e39477. https://doi.org/10.1371/journal.pone.0039477.

Morisse, J. P., J. P. Cotte, and D. Huonnic. 1995. Effect of dehorning on behaviour and plasma cortisol responses in young calves. Applied Animal Behaviour Science 43(4):239–247. https://doi.org/10.1016/0168-1591(95)00569-E.

Mueller, M. L., J. B. Cole, N. K. Connors, D. J. Johnston, I. A. S. Randhawa, and A. L. van Eenennaam. 2021. Comparison of Gene Editing Versus Conventional Breeding to Introgress the POLLED Allele Into the Tropically Adapted Australian Beef Cattle Population. Frontiers in genetics 12:593154. https://doi.org/10.3389/fgene.2021.593154.

Mueller, M. L., J. B. Cole, T. S. Sonstegard, and A. L. van Eenennaam. 2019. Comparison of gene editing versus conventional breeding to introgress the POLLED allele into the US dairy cattle population. Journal of dairy science 102(5):4215–4226. https://doi.org/10.3168/jds.2018-15892.

OMIA 000483-9913. Polled/Horns in Bos taurus. Accessed Feb 11, 2023. https://omia.org/OMIA0483/9913/.

Qanbari, S., Simianer, H. 2014. Mapping signatures of positive selection in the genome of livestock. Livestock Science 166:133–43.

Randhawa, I. A. S., B. M. Burns, M. R. McGowan, L. R. Porto-Neto, B. J. Hayes, R. Ferretti, K. M. Schutt, and R. E. Lyons. 2020. Optimized Genetic Testing for Polledness in Multiple Breeds of Cattle. G3 (Bethesda, Md.) 10(2):539–544. https://doi.org/10.1534/g3.119.400866.

Randhawa, I. A. S., M. R. McGowan, L. R. Porto-Neto, B. J. Hayes, and R. E. Lyons. 2021. Comparison of Genetic Merit for Weight and Meat Traits between the Polled and Horned Cattle in Multiple Beef Breeds. Animals 11(3):870. https://doi.org/10.3390/ani11030870.

Rothammer, S., A. Capitan, E. Mullaart, D. Seichter, I. Russ, and I. Medugorac. 2014. The 80-kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. Genet Sel Evol 46(1):44. https://doi.org/10.1186/1297-9686-46-44.

Rowan, T. N., H. J. Durbin, C. M. Seabury, R. D. Schnabel, and J. E. Decker. 2021. Powerful detection of polygenic selection and evidence of environmental adaptation in US beef cattle. PLOS Genetics 17(7):e1009652. https://doi.org/10.1371/journal.pgen.1009652.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–7.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., et al. 2006 Positive natural selection in the human lineage. Science 312:1614–20. doi:10.1126/science.1124309.

Sargolzaei, M., Schenkel, F.S. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics 25(5):680-681. doi:10.1093/bioinformatics/btp045

Sax, K. 1923. The association of size differences with seed-coat pattern and pigmentation in Phaseolus vulgaris. Genetics 8:552–60.

Schafberg, R., and H. H. Swalve. 2015. The history of breeding for polled cattle. Livestock Science 179:54–70. https://doi.org/10.1016/j.livsci.2015.05.017.

Scheper, C., R. Emmerling, K.-U. Götz, and S. König. 2021. A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle. Genet Sel Evol 53(1):60. https://doi.org/10.1186/s12711-021-00652-z.

Segelke, D., H. Täubert, F. Reinhardt, and G. Thaller. 2013. Chances and limits of breeding polled cattle Deutsche Holstein. Züchtungskunde 85(4):253–269.

Segelke, D., H. Täubert, F. Reinhardt, and G. Thaller. 2016. Considering genetic characteristics in German Holstein breeding programs. Journal of dairy science 99(1):458–467. https://doi.org/10.3168/jds.2015-9764.

Simon, R., C. Drögemüller, and G. Lühken. 2022. The Complex and Diverse Genetic Architecture of the Absence of Horns (Polledness) in Domestic Ruminants, including Goats and Sheep. Genes 13(5). https://doi.org/10.3390/genes13050832.

Sørensen, P., Lund, M.S., Guldbrandtsen, B., Jensen, J., Sorensen, D. 2003. A comparison of bivariate and univariate QTL mapping in livestock populations. Genet Sel Evol. 35:605–22.

Specht, L. W. 2008. Polled Holstein History. Accessed Feb 11, 2023. https://nebula.wsimg.com/339065745033f44d872cb82122ef7ce2?AccessKeyId=27933CBAC73165BF9625&disposition=0&alloworigin=1.

Spurlock, D. M., M. L. Stock, and J. F. Coetzee. 2014. The impact of 3 strategies for incorporating polled genetics into a dairy cattle breeding program on the overall herd genetic merit. Journal of dairy science 97(8):5265–5274. https://doi.org/10.3168/jds.2013-7746.

Stafford, K. J., and D. J. Mellor. 2005. Dehorning and disbudding distress and its alleviation in calves. Veterinary journal (London, England 1997) 169(3):337–349. https://doi.org/10.1016/j.tvjl.2004.02.005.

Stafford, K. J., and D. J. Mellor. 2011. Addressing the pain associated with disbudding and dehorning in cattle. Applied Animal Behaviour Science 135(3):226–231. https://doi.org/10.1016/j.applanim.2011.10.018.

Stafuzza, N. B., Silva, Rafael Medeiros de Oliveira, E. Peripolli, L. A. F. Bezerra, R. B. Lôbo, C. d. U. Magnabosco, F. A. Di Croce, J. B. Osterstock, D. P. Munari, D. A. L. Lourenco, and F. Baldi. 2018. Genome-wide association study provides insights into genes related with horn development in Nelore beef cattle. PloS one 13(8):e0202978. https://doi.org/10.1371/journal.pone.0202978.

Tang, K., Thornton, K.R., Stoneking, M. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5:e171.

van Arendonk, J.A.M., Tier, B., Bink, M.C.A.M., Bovenhuis, H. 1998. Restricted maximum likelihood analysis of linkage between genetic markers and quantitative trait loci for a granddaughter design. J Dairy Sci. 81:76–84.

van Arendonk, J.A.M., Tier, B., Kinghorn, B. 1994. Use of multiple genetic-markers in prediction of breeding values. Genetics 137:319–29.

Weir, B.S., Cockerham, C.C. 1984. Estimating F-Statistics for the analysis of population structure. Evolution 38:1358.

White, W. T., and H. L. Ibsen. 1936. Horn inheritance in Galloway-Holstein cattle crosses. Journ. of Genetics 32(1):33–49. https://doi.org/10.1007/BF02982500.

Windig, J. J., and A. Eggen. 2009. D.2.2.2. Report on the assessment of breeding strategies in relation to the introduction of the polled gene. ALCASDE: Study on the Improved Methods for Animal-Friendly Production. In Particular on Alternatives to the Dehorning of Cattle. Appendix23. http://ec.europa.eu/food/animal/welfare/farm/docs/calves_alcasde_d-2-2-2.pdf.

Windig, J. J., R. A. Hoving-Bolink, and R. F. Veerkamp. 2015. Breeding for polledness in Holstein cattle. Livestock Science 179:96–101. https://doi.org/10.1016/j.livsci.2015.05.021.

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M. 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 88:76–82.

CHAPTER 2

Preliminary study to validate the variance component estimation approach to infer pleiotropic effects of the polled locus using simulated and real data

Carsten Scheper, Reiner Emmerling, Kay-Uwe Götz and Sven König

The contents of Chapter 2 were published in two separate Supplementary Files to Scheper, C., Emmerling, R., Götz, KU., König, S. 2021. A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle. Genet Sel Evol 53, 60 (2021) as presented in Chapter 3 of this thesis.

The originally published versions of the Supplementary Files are available online:

https://static-content.springer.com/esm/art%3A10.1186%2Fs12711-021-00652-z/MediaObjects/12711_2021_652_MOESM3_ESM.pdf

https://static-content.springer.com/esm/art%3A10.1186%2Fs12711-021-00652-z/MediaObjects/12711_2021_652_MOESM4_ESM.pdf

Own contribution to the publication:

Study design:	substantial
Collection of data:	participated
Statistical analysis:	autonomous
Manuscript:	autonomous

Abstract

A preliminary study was designed to study the suitability of a variance component estimation approach utilizing uni- and bivariate linear animal models including a putative QTL effect for the trait polledness. The QTL effect for the trait polledness was modelled using a marker-based numerator relationship matrix calculated based on real and simulated polled genotypes. In a first step univariate linear models for the trait polledness based on a real test dataset were used to evaluate variance component estimation for mono- and polygenic components and heritability. In a second step uni- and bivariate linear models including a simulated Mendelian trait (representing polledness) with a varying pleiotropic QTL-effect on a secondary simulated quantitative trait were used to evaluate variance component estimation for pleiotropic QTL- and polygenic effect detection. In summary the results showed the suitability of the approach in reference to the expected values for variance components and heritability for a Mendelian trait (i.e. $h^2 = 1$), if numerically coded polled genotypes were used as phenotypes. In addition, the approach was able to validly capture the simulated pleiotropic QTL effects of the Mendelian trait on a secondary quantitative trait. In conclusion, the preliminary study was successful to prove the general suitability of the chosen methodic approach for variance component estimation in a bigger dataset.

Background

To initially test the methodological variance component estimation (VC) approach outlined in Chapter 3 of this thesis (Scheper et al., 2021) we estimated VC for the trait polledness in a preliminary study based on simulated and real test data. Although the outlined VC approach appeared straightforward for our research question, to our knowledge there are no available studies proving the applicability based on either simulated or real data in context of Mendelian traits.

For the first step of validation, we used a reduced real dataset based on the full dataset as presented in Chapter 3 (Scheper et al., 2021). We reconstructed the genotypes at the polled locus in the pedigree of the full dataset according to the method described in Chapter 3 (Scheper et al., 2021) before reducing the dataset to the test size. Due to the small size of the dataset, we only estimated univariate models for polledness based on the real test dataset in the first step.

As flexible stochastic simulation packages to simulate precise genomic trait architectures are readily available (Sargolzaei M and Schenkel FS, 2009; Faux et al., 2016), we then decided to further validate our approach based on given parameters adapted to the hypothesized pleiotropic effect of the polled locus by simulation.

Methods

Step 1 – Univariate linear models for the trait polledness based on real test data

Test dataset

For the preliminary study we selected 12 farms and 1796 cows from the final dataset of 24 farms used in the full manuscript (see Chapter 3). To infer the genotypes at the polled locus we followed the steps described in detail in Chapter 3 of this thesis. After inference, the entire pedigree was trimmed to 8624 animals based on the 1796 polled target cows with records for performance traits including 5 generations of ancestors. Table 1 gives an overview of the absolute frequencies of genotype labels in the dataset after the inference procedures.

Table 1 Descriptive statistics for the polled trait in the dataset for preliminary analysis before and after inference of genotypes.

Dataset		**Animals**	**pp**	**Pp**	**PP**	***p*(P)**
Animals with phenotypes	**initially registered**	n = 1796	1285	493	18	0.147
	after inference	n = 1796	1246	544	6	0.155
Full pedigree		n = 8624	7560	993	71	0.066

Calculation of A_v for inferred real genotypes

The inferred genotypes were used to compute the probabilities of inheriting the paternal or maternal alleles from sire and dam at the polled locus starting from founders in the pedigree. Identical-by-descent (IBD) probabilities between the alleles at the polled locus of any two founders were assumed to be zero.

The G_v matrix, representing a gametic relationship matrix based on computed IBD probabilities at the polled locus was subsequently computed using the algorithm by van Arendonk et al (van Arendonk et al., 1994) with a self-written R function (see Additional file 3 in Chapter 3). After full computation, G_v was scaled down to the dimensions of a marker based numerator relationship matrix A_v using the following matrix transformation $A_v = \frac{1}{2} K G_v K'$, with $K = I_n \otimes [1,1]; n =$ number of animals). The inverse of A_v was then used in VC estimation of QTL effects in the univariate test models for the trait polledness.

Models

During Step 1 we also evaluated the effect of the genotype inference using simple univariate animal models including pedigree relationships by comparing models based on phenotypes before

(polled_raw) and after (polled_inf) inference. Animals without any registered polled genotype in the raw dataset were set to horned. Polled phenotypes were defined as either binary (i.e. pp = 0, Pp and PP = 1, pol_bin) or numerically coded genotype labels based on the inferred polled genotypes (i.e. pp = 0, Pp = 1, PP = 2, pol_num) representing the allele content at the polled locus. Both phenotypes were evaluated using linear models. For pol_bin, we also applied threshold models.
The basic linear models (M_{Basic}) for pol_bin and pol_num were defined as:

$$y = Z_a a + e,$$

where y is a vector of phenotypes, u is a vector of additive polygenic effects, Z_a is an incidence matrix relating animals to phenotypes, and e is a residual vector. The random effects a and e are assumed to be uncorrelated and distributed as univariate normal densities as follows: $a \sim N_q\,(0, A\sigma_a^2)$ and $e \sim N_m(0, R\sigma_e^2)$, where σ_a^2 and σ_e^2 are the polygenic variance and the residual variance, respectively. A is the standard additive genetic relationship matrix and R is a known diagonal matrix.
Differing from the description above, an additonal threshold model using a logit link function was defined for pol_bin. Due to the proven Mendelian inheritance at the polled locus, environmental effects do not affect the phenotype by definition. Therefore, no fixed effects were incorporated. Variance components for all models were estimated using DMU (Madsen et al. 2006).
Going further, we then tested the approach to incorporate QTL effects at the polled locus using A_v as described in detail in Chapter 3. Hence, in addition to the basic linear and threshold models as described above, we estimated variance components from the following models for all defined traits. The extended QTL models ($M_{QTL\,ext}$) were defined as

$$y = Z_a a + Z_v v + e,$$

with the same properties as described for the basic model adding v, a vector of additive QTL effects with a distribution of $v \sim N_q\,(0, A\sigma_v^2)$ and Z_v an incidence matrix relating animals to phenotypes. Finally, reduced QTL models ($M_{QTL\,red}$) without the standard additive genetic relationships were also defined as

$$y = Z_v v + e,$$

with v, a vector of additive QTL effects with a distribution of $v \sim N_q\,(0, A\sigma_v^2)$ and Z_v an incidence matrix relating animals to phenotypes.

Step 2 - Simulation study

Dataset

The software package QMSim (Sargolzaei & Schenkel, 2009) was used to simulate 2 quantitative traits with differing heritabilites of 0.3 (SimTrait 1) and 0.05 (SimTrait 2), which reflect the

spectrum of relevant traits in the breeding goal for German Simmental cattle. Both quantitative traits were simulated as a female sex-limited polygenic trait.

To create a pleiotropic effect between a Mendelian trait mimicking polledness and the simulated quantitative traits, a single QTL with varying effects (i.e. QTL heritabilites) on the respective trait was simulated with different QTL effect size scenarios. Table 2 displays all simulated QTL scenarios in SimTrait 1 and SimTrait 2. All additional genetic variation apart from the contribution of the simulated QTL was set to be polygenic. The QTL was positioned at the proximal end of one simulated chromosome reflecting the size of the bovine chromosome 1 (i.e. 158cM). The QTL was simulated with 2 alleles, i.e. one allele representing the wild-type allele for hornedness and one allele representing a causal mutation for polledness. Allele frequencies for both alleles were set to 0.5 at the beginning of the simulation. The simulated genotypes were used to code a binary phenotype representing horned and polled animals. The simulated genotypes were further used for the calculation of an A_v matrix according to van Arendonk et al (van Arendonk et al., 1994) for use in the variance component estimation as described below.

To initiate the simulation, a historical population of 1000 females and 100 males was simulated for 100 generations with generation 100 subsequently serving as a founder population. Based on the founder population 4 subsequent generations were simulated. The female reproductive rate was limited to one progeny per female, with an equal probability for either male or female progeny. Selection and mating were set to random to avoid fixation of the simulated QTL and ensure segregation. The sire and dam replacement rates were set to 1 and 0.5 respectively. Males without progeny were discarded from the dataset. Per scenario only one repetition was simulated. The final dataset in each QTL scenario (see Table 1) consisted of 3400 animals including 3000 females with phenotypic records for the simulated quantitative traits. In contrast to the simulated quantitative traits, all animals had phenotypes for the simulated Mendelian trait based on the QTL genotypes as described above to mimic the realistic situation in the trait polledness.

Table 2 Overview of simulated traits and QTL effects for the validation of the variance component estimation approach.

	QTL effects (QTL-h^2)		
SimTrait 1 (h^2 = 0.3)	**0.1**	**0.05**	**0.025**
SimTrait 2 (h^2 = 0.05)	**0.025**	**0.01**	

Calculation of A_v for simulated genotypes

The simulated genotypes were used to compute the probabilities of inheriting the paternal or

maternal alleles from sire and dam at the simulated polled locus starting from founders. Identical-by-descent (IBD) probabilities between the alleles at the simulated polled locus of any two founders were assumed to be zero.

The G_v matrix based on computed IBD probabilities at the simulated locus was subsequently computed using the algorithm by van Arendonk et al (van Arendonk et al., 1994) with a self-written R function (see Additional file 3). After full computation, G_v was scaled down to the dimensions of a marker based numerator relationship matrix A_v using the following matrix transformation $A_v = \frac{1}{2} K G_v K'$, with $K = I_n \otimes [1,1]; n = \text{number of animals})$. The inverse of A_v was then used in VC estimation of QTL effects.

Models

Variance components for all models were estimated using DMU (Madsen et al. 2006). In addition to the univariate models for the simulated quantitative traits, we also used bivariate models including the simulated polled trait as a dependent variable.

The basic linear model without QTL effects (M_{Basic}) for the simulated quantitative trait was defined as:

$$y = Z_a a + e,$$

where y is a vector of phenotypes, a is a vector of additive polygenic effects, Z_a is an incidence matrix relating animals to phenotypes, and e is a residual vector. The random effects a and e are assumed to be uncorrelated and distributed as univariate normal densities as follows: $a \sim N_q(0, A\sigma_a^2)$ and $e \sim N_m(0, R\sigma_e^2)$, where σ_a^2 and σ_e^2 are the polygenic variance and the residual variance, respectively. A is the standard additive genetic relationship matrix and R is a known diagonal matrix.

The extended linear QTL model (M_{QTL}) was defined as

$$y = Z_a a + Z_v v + e,$$

with the same properties as described for the basic model adding v, a vector of additive QTL effects with a distribution of $v \sim N_q(0, Av\sigma_v^2)$ and Z_v an incidence matrix relating animals to phenotypes.

Test statistics

Hypothesis tests for the presence of pleiotropic QTL-effects of the simulated polled locus were based on the asymptotic distribution of the likelihood ratio test (LRT) statistic,

$$LRT = -2ln(L_{BASIC} - L_{QTL}),$$

where L_{BASIC} and L_{QTL} are the maximized likelihoods under M_{Basic} and M_{QTL} respectively. Under

regularity conditions, the asymptotic distribution of the likelihood ratio test statistic follows a X^2-distribution, with degrees of freedom equal to the difference in the number of independent parameters between the models tested (Sorensen et al. 2003). LRT tests were only calculated for the univariate estimations.

Results and Discussion

Step 1 – Real test dataset

Table 3 Estimated variance components for different phenotype datasets and definitions of the trait polledness.

Trait definition	Model type	Dataset					
		polled_raw			polled_inf		
		σ^2_a	σ^2_e	h^2 (SE)	σ^2_a	σ^2_e	h^2 (SE)
pol_num	**linear**	0.178	0.043	0.802 (0.041)	0.089	0.001	0.995 (0.005)
pol_bin	**linear**	0.154	0.041	0.790 (0.041)	0.069	0.006	0.919 (0.009)
	logit	1.802	3.290	0.354 (0.027)	3.228	3.290	0.495 (0.013)

The heritability estimates displayed in Table 3, show a positive effect of the inference of all polled geno- and phenotypes in the pedigree and their consideration. In general, low standard errors in all models indicate a valid model fit for both trait definitions and model types.

Although there are, to our knowledge, no examples of quantitative-genetic analysis of qualitative traits with Mendelian inheritance after successful mapping in the literature, the estimates should be close to or around $h^2 = 1$ based on the trait's genetic architecture at the polled locus. The results show that such values were achieved in the applied linear models only by considering all available (reconstructed) genotypes from the entire pedigree and based on numerically coded polled genotypes (i.e. gene content) as phenotypes (pol_num) in animal models based on additive pedigree relationships. However, estimation based on binary phenotypes also led to a high heritability around 0.90 in the linear model. The heritability estimate from the threshold model on the other hand was only moderate.

Breeding value correlations comparing the datasets polled_raw and polled_inf were 0.75 for pol_num, 0.80 for pol_bin from linear models and 0.78 from threshold models. Correlations for the different phenotypic definitions in the dataset polled_inf were 0.97 for pol_num and pol_bin linear, 0.78 for pol_num and pol_bin logit and 0.79 for pol_bin linear and pol_bin logit.

Figure 2 shows a direct comparison of estimated individual additive genetic effects from the

pol_num models as presented in Table 3 plotted against the input phenotypes (i.e. unchanged numerically coded raw genotypes or inferred genotypes). Figure 1 b clearly shows a substantial number of polled animals registered as horned, which leads to rather poorly accurate estimated additive effects explaining the rather low heritability. Inference of missing and falsely registered polled genotypes greatly improves the accordance between the input genotypes (i.e. numerically coded genotypes as phenotypes) and the estimated additive effects.

It should be noted that the analysis of the trait polledness in our study is not focused on estimation of breeding values but more importantly best fit to available pheno- and genotype data while reflecting gametic relationships between animals at the polled locus. Hence, as a practical conclusion from the preliminary study, using completely inferred genotypes for the full pedigree and defining the polled phenotype as numerically coded genotypes representing the allele content at the polled locus (as in pol_num) in linear models appears to be most suitable to capture the expected genetic variance of the polled trait based on pedigree relationships. Therefore, all following analysis were based on inferred geno- and phenotypes in the full pedigree.

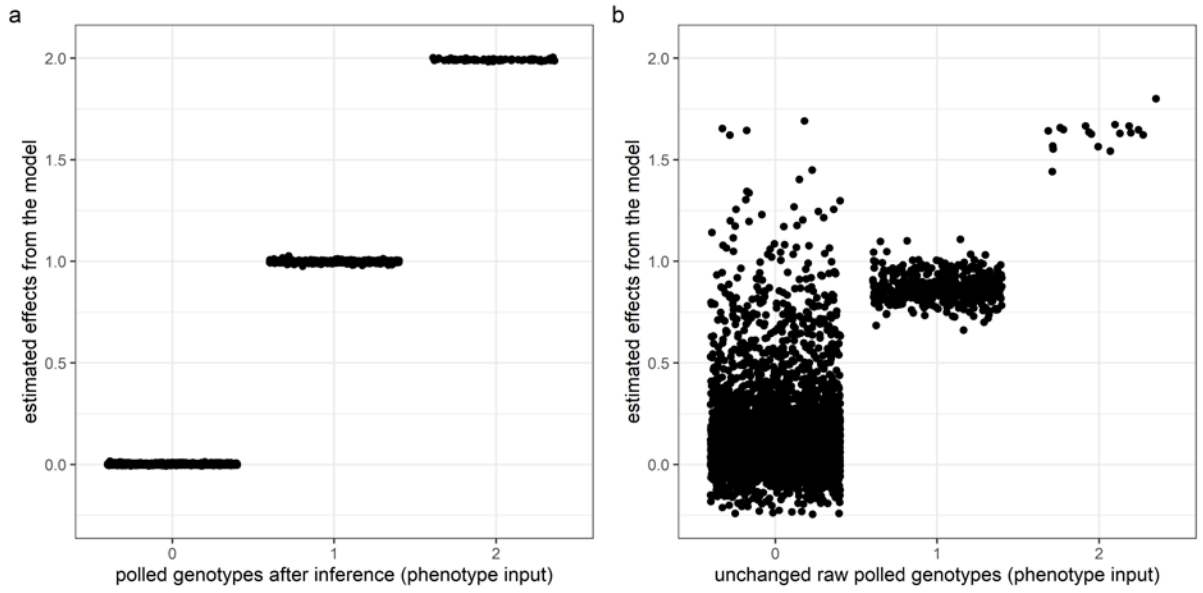

Figure 2 Estimated individual additive effects from the variance component models compared to the input phenotypes (a = inferred genotypes as input phenotypes (polled_inf), b = unchanged raw genotypes as input phenotypes (polled_raw)). Genotypes are coded as follows: 0 = pp, 1 = Pp, 2 = PP.

The results shown in Table 4 reflect that the incorporation of the single locus effect based on gametic relationships at the polled locus (random effect v modelled with MQTL matrix A_v) was successful. Hence, in the extended models ($M_{QTL\,ext}$), almost the entire genetic variance is transferred to the single locus effect, with only marginal variance remaining polygenic (random effect a modelled with A) with phenotypes pol_num and pol_bin in the logit model. In the linear model based on pol_bin however the remaining polygenic variance was still remarkably higher. In

addition, heritability estimates continue to approach the theoretical expectation value of 1 in models including v in models for pol_num.

Considering the monogenic structure of the trait polledness the model $M_{QTL\,red}$ for pol_num therefore appears to best fit the data as well as realistically capturing the genetic variance at the polled locus. Although a very small fraction of polygenic variance remains in the model $M_{QTL\,ext}$, the results reflect a comparable fit compared to $M_{QTL\,red}$. Therefore $M_{QTL\,ext}$ appears to be preferable for the planned bivariate analysis to be in line with methods for bivariate QTL analysis described in the literature (Sørensen et al., 2003). In addition, the results can be interpreted as a first validation of our approach to incorporate a single locus effect using the MQTL matrix A_v in the context of linear animal models. Hence, using the MQTL matrix A_v in the analysis of further traits could be suitable to model and map potential direct or closely linked QTL effects of the polled locus.

Table 4 Estimated variance components for different phenotype definitions of the trait polledness comparing different models incorporating QTL effects. $\boldsymbol{\sigma^2_a}$ = additive genetic variance, $\boldsymbol{\sigma^2_v}$ = single locus variance, $\boldsymbol{\sigma^2_e}$ = residual variance, **h^2 (SE)** = heritability, standard error in brackets, **pol_num** = numerically coded genotypes (i.e. gene content) as phenotypes, **pol_bin** = binary coded phenotype,

Trait definition	Model	σ^2_a	σ^2_v	σ^2_e	h^2 (SE)
pol_num (linear)	M_{Basic}	0.089		0.001	0.995 (0.005)
	$M_{QTL\,ext}$	0.343e-06	0.018	0.288e-07	1.000 (0.010)
	$M_{QTL\,red}$		0.018	0.100e-06	1.000 (0.004)
pol_bin (linear)	M_{Basic}	0.069		0.006	0.919 (0.009)
	$M_{QTL\,ext}$	0.002	0.014	0.004	0.818 (0.019)
	$M_{QTL\,red}$		0.015	0.004	0.789 (0.008)
pol_bin (logit)	M_{Basic}	3.228		3.290	0.495 (0.013)
	$M_{QTL\,ext}$	0.100e-9	3.327	3.290	0.503 (0.091)
	$M_{QTL\,red}$		3.327	3.290	0.502 (0.019)

Step 2 - Simulation study

For SimTrait 1 with moderate simulated heritability (h2 = 0.30), QTL effects were detected in all models incorporating A_v, but were either over- or underestimated (see Table 5). Estimates scattered around the expected values with high deviations. Two out of 3 univariate QTL models also had

significantly better model fits compared to the null models based on the performed likelihood ratio tests. Total trait heritablity estimates for all models deviated only slightly from the predefined heritablility with generally low standard values < 0.05.

Results for SimTrait 2 with smaller simulated heritability (h2 = 0.05) also proved that all models incorporating A_v were able to detect the simulated QTL effects (see Table 6). However, in contrast to SimTrait 1 QTL effects were highly overestimated in all models in relation to the overall heritability of the simulated trait. Hence, all genetic variance of the simulated trait was falsely attributed to the QTL effect, and QTL models showed no better fit compared to the Basic models based on likelihood ratio tests. Total trait heritablity estimates for all models in both traits deviated only slightly from the predefined heritability with generally low standard errors (SE) < 0.05 indicating good model fit despite the rather small dataset.

Although our goal to validate the chosen VC estimation approach to dissect direct pleiotropic effects of Mendelian trait loci from other polygenic effects via stochastic simulation was successful, a few questions remain open. Based on our results it is not possible to clarify the reasons for the observed over- and underestimation of simulated QTL effects. Running the chosen scenarios with a sufficient number of repetitions per scenario could give a clearer picture of the variation of realized QTL effects in relation to the initial settings of the simulation. We opted to use only single repetitions in the present study to limit computation times while focusing on the general validity of the chosen VC estimation approach.

In addition to the aforementioned aspect, our simulation approach also only realized a simplified structure of a Mendelian trait locus with only two effective variants. In contrast, the polled locus is more heterogeneous with 4 effective variants with differing genomic structure and potentially varying effects on other traits. While other recent simulation studies have shown that suitable selection and mate allocation strategies allowing for a significant increase in the frequency of the desired alleles while preserving high genetic gain exist (Scheper et al., 2016; Cole, 2015; Gaspa et al., 2015; Spurlock et al., 2014), none of these studies considered varying variant specific pleiotropic effects as a possible scenario. Although our results suggest that based on our dataset from German Simmental direct pleiotropic effects of the polled locus do not exist, the presented simulation approach could also readily be applied to novel, yet undiscovered, Mendelian traits where pleiotropic effects exist.

Conclusions

The variance component estimation approach to evaluate pleiotropic effects of the polled locus on quantitative traits using a marker-based numerator relationship matrix was successfully validated using a real test dataset and stochastic simulation. The simulated QTL effects were validly detected with a tendency towards overestimation in simulated quantitative traits with moderate and low heritability.

Table 5 VC estimation results for SimTrait 1 with h2 = 0.30 for differing QTL effects.

Trait h^2 0.30	**Model**	σ^2_a	σ^2_v	σ^2_e	**QTL-h²**	**h² (SE)** (polygenic + QTL)	**LRT p (λ)**
QTL-h^2 = 0.10	*Basic*	0.303e-02		0.709e-02		0.283 (0.036)	1 (-20.294)
	QTL	0.043e-02	0.238e-02	0.725e-02	0.237	0.280 (0.055)	
QTL-h^2 = 0.05	*Basic*	0.316e-02		0.643e-02		0.330 (0.037)	< 0.01 (10.314)
	QTL	0.125e-02	0.168e-02	0.706e-02	0.168	0.293 (0.036)	
QTL-h^2 = 0.025	*Basic*	0.297e-02		0.720e-02		0.292 (0.037)	0.016 (4.571)
	QTL	0.191e-02	0.095e-02	0.724e-02	0.094	0.283 (0.051)	

σ^2_a = additive genetic variance based on $\boldsymbol{A}$, σ^2_v = additive genetic variance based on $\boldsymbol{A_v}$, σ^2_e = residual variance, **QTL-h²** = QTL heritability calculated as $\sigma^2_v/\sigma^2_a+\sigma^2_v+\sigma^2_e$, **h² (SE)** = overall heritability and standard error (in brackets) calculated as $\sigma^2_v+\sigma^2_a / \sigma^2_a+\sigma^2_v+\sigma^2_e$, **LRT p (λ)** = p- and lambda values from likelihood ratio tests.

Table 6 VC estimation results for the simulated quantitative trait with h2 = 0.05 with differing QTL effects.

Trait h^2 0.05	**Model**	σ^2_a	σ^2_v	σ^2_e	**QTL-h²**	**h² (SE)** (polygenic + QTL)	**LRT p (λ)**
QTL-h^2 = 0.025	*Basic*	0.117		2.911		0.039 (0.021)	1 (-2034.207)
	QTL	0.354e-06	0.141	2.878	0.047	0.047 (0.028)	
QTL-h^2 = 0.01	*Basic*	0.145		2.741		0.050 (0.021)	1 (-2012.018)
	QTL	0.125e-05	0.141	2.736	0.049	0.049 (-)*	

σ^2_a = additive genetic variance based on $\boldsymbol{A}$, σ^2_v = additive genetic variance based on $\boldsymbol{A_v}$, σ^2_e = residual variance, **QTL-h²** = QTL heritability calculated as $\sigma^2_v/\sigma^2_a+\sigma^2_v+\sigma^2_e$, **h² (SE)** = overall heritability and standard error (in brackets) calculated as $\sigma^2_v+\sigma^2_a / \sigma^2_a+\sigma^2_v+\sigma^2_e$, **LRT p (λ)** = p- and lambda values from likelihood ratio tests, *the model did not fully converge, calculation of SE was not possible.

References

Cole, J. B. 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genetics, selection, evolution GSE 47:94. https://doi.org/10.1186/s12711-015-0174-9.

Faux, A.-M., G. Gorjanc, R. C. Gaynor, M. Battagin, S. M. Edwards, D. L. Wilson, S. J. Hearne, S. Gonen, and J. M. Hickey. 2016. AlphaSim: Software for Breeding Program Simulation. The plant genome 9(3). https://doi.org/10.3835/plantgenome2016.02.0013.

Gaspa, G., R. F. Veerkamp, M. P. Calus, and J. J. Windig. 2015. Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation. Livestock Science 179:86–95. https://doi.org/10.1016/j.livsci.2015.05.020.

Madsen, P., P. Sørensen, G. Su, L. H. Damgaard, H. Thomsen, and R. Labouriau. 2006. DMU - a package for analyzing multivariate mixed models. p. 11. *In* Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006.

Sargolzaei M, and Schenkel FS. 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics 25(5):680–681. https://doi.org/10.1093/bioinformatics/btp045.

Scheper, C., R. Emmerling, K.-U. Götz, and S. König. 2021. A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle. Genetics, selection, evolution GSE 53(1):60. https://doi.org/10.1186/s12711-021-00652-z.

Scheper, C., M. Wensch-Dorendorf, T. Yin, H. Dressel, H. Swalve, and S. Konig. 2016. Evaluation of breeding strategies for polledness in dairy cattle using a newly developed simulation framework for quantitative and Mendelian traits. Genetics, selection, evolution GSE 48(1):50. https://doi.org/10.1186/s12711-016-0228-7.

Sørensen, P., M. S. Lund, B. Guldbrandtsen, J. Jensen, and D. Sorensen. 2003. A comparison of bivariate and univariate QTL mapping in livestock populations. Genetics, selection, evolution GSE 35(6):605–622. https://doi.org/10.1051/gse:2003042.

Spurlock, D. M., M. L. Stock, and J. F. Coetzee. 2014. The impact of 3 strategies for incorporating polled genetics into a dairy cattle breeding program on the overall herd genetic merit. Journal of dairy science 97(8):5265–5274. https://doi.org/10.3168/jds.2013-7746.

van Arendonk, J. A., B. Tier, and Kinghorn. B. 1994. Use of multiple genetic-markers in prediction of breeding values. GENETICS 137(1):319–329.

CHAPTER 3

A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle

Carsten Scheper, Reiner Emmerling, Kay-Uwe Götz and Sven König

Genetics Selection Evolution (2021) 53:60
https://doi.org/10.1186/s12711-021-00652-z

Additional files not shown in this thesis, due to their extent are directly available in the online version of this article at:
https://gsejournal.biomedcentral.com/articles/10.1186/s12711-021-00652-z#Sec12

Own contribution to the publication:

Study design:	substantial
Collection of data:	participated
Statistical analysis:	autonomous
Manuscript:	mostly autonomous

Scheper *et al. Genet Sel Evol* (2021) 53:60
https://doi.org/10.1186/s12711-021-00652-z

GSE Genetics Selection Evolution

RESEARCH ARTICLE Open Access

A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle

Carsten Scheper[1*], Reiner Emmerling[2], Kay-Uwe Götz[2] and Sven König[1]

Abstract

Background: Managing beneficial Mendelian characteristics in dairy cattle breeding programs implies that the correlated genetic effects are considered to avoid possible adverse effects in selection processes. The Mendelian trait polledness in cattle is traditionally associated with the belief that the polled locus has unfavorable effects on breeding goal traits. This may be due to the inferior breeding values of former polled bulls and cows in cattle breeds, such as German Simmental, or to pleiotropic or linkage effects of the polled locus.

Methods: We focused on a variance component estimation approach that uses a marker-based numerator relationship matrix reflecting gametic relationships at the polled locus to test for direct pleiotropic or linked quantitative trait loci (QTL) effects of the polled locus on relevant traits. We applied the approach to performance, health, and female fertility traits in German Simmental cattle.

Results: Our results showed no evidence for any pleiotropic QTL effects of the polled locus on test-day production traits milk yield and fat percentage, on the mastitis indicator 'somatic cell score', and on several female fertility traits, i.e. 56 days non return rate, days open and days to first service. We detected a significant and unfavorable QTL effect accounting for 6.6% of the genetic variance for protein percentage only.

Conclusions: Pleiotropy does not explain the lower breeding values and phenotypic inferiority of polled German Simmental sires and cows relative to the horned population in the breed. Thus, intensified selection in the polled population will contribute to increased selection response in breeding goal traits and genetic merit and will narrow the deficit in breeding values for production traits.

Background

Bovine polledness is a Mendelian trait (OMIA 000483-913) that was discovered as early as 1902 [1] and is controlled by one locus located at the proximal end of *Bos taurus* autosome 1 (BTA1). The four dominant allelic variants cause polled phenotypes and scurs, and the recessive wild-type variant causes the horned phenotype [2–6]. The identified "Celtic" (P_C) and "Friesian" (P_F) variants are predominant in polled animals from European dairy and dual-purpose cattle breeds [3, 7]. The Mendelian inheritance pattern at the polled locus is proven [3]. Nevertheless, an oligogenic model of inheritance may explain the remaining complexity of

*Correspondence: carsten.scheper@agrar.uni-giessen.de
[1] Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Giessen, Germany
Full list of author information is available at the end of the article

horn-related phenotypes, including scurs and polledness, in which the polled locus has a presumed epistatic suppressive function [8].

The broad availability and use of the high-throughput genotyping technology in cattle have enhanced the discovery of new Mendelian genetic characteristics during the past decade [9]. Most of these relevant genetic characteristics are lethal recessive monogenic disorders (e.g. cholesterol deficiency (CDH) [10]) or detrimental recessive haplotypes with unfavorable effects on production and functional traits (e.g. [11]). However, there are a few examples of favorable or beneficial genetic characteristics such as the red factor in Holstein cattle [12] or polledness [3]. Managing the growing number of genetic characteristics is a major challenge in current dairy cattle breeding programs [13, 14]. The management of beneficial genetic characteristics implies an increase in the frequency of the desired causative alleles by selection, ultimately until fixation, and the genomic regions that are linked to the causative alleles will also be fixed. Thus, it is important to examine the pleiotropic or linked effects of the desired alleles before aiming for their fixation.

The basis of genetic trait associations is either pleiotropy or linkage [15]. From the perspective of the Mendelian trait polledness, genetic associations with other traits may be caused by direct or linked effects of causative variants for polledness, in analogy to a quantitative trait locus (QTL). Differences in the means of a quantitative trait between genotype groups at a single marker locus is a classic example of a potential pleiotropic QTL effect [15, 16]. Lower means for the breeding values of production traits in homo- and heterozygous polled animals, which are in part statistically significant, have been reported for various breeds [17–21]. In contrast, for reproduction traits, neutral or even positive effects have been estimated [17, 18, 20]. One explanation for breeding value inferiority in polled German Simmental is the history of introgression from Simmental beef populations, which possibly result in conserved chromosomal segments with detrimental effects that are distant from the polled locus [21]. Consequently, this breeding value inferiority could also be caused by genetic drift (i.e. a random increase of less favorable alleles), because all polled animals descend from only a few polled founder animals [21]. To date, simple comparisons of breeding values do not explain the genetic mechanisms that underlie the associations of polledness with other quantitative traits. In this study, our aim was to separate direct QTL effects of the polled locus from associated polygenic effects (i.e., genetic variation in genomic locations that are distant from the polled locus on BTA1 or other chromosomes) by using available genomic and pedigree information.

Variance component (VC) estimation methods that incorporate single or multiple marker genotype information to infer QTL effects at a given chromosomal segment or position appear to be suitable to test for correlated effects of the polled locus [22, 23]. In this context, random QTL allele effects can be modelled using either marker-based gametic (hereafter called $\mathbf{G_v}$) or numerator (hereafter called $\mathbf{A_v}$) relationship matrices [24]. These matrices reflect expected identity-by-descent (IBD) relationships between individuals inferred from genotypes at given marker loci, e.g., the polled locus. In contrast to $\mathbf{A_v}$, the pedigree-based relationship matrix $\mathbf{A}$ reflects the expected IBD relationships across the entire genome that are inferred from known pedigree relationships and is traditionally used in VC models to model the polygenic additive genetic variance. The combination of $\mathbf{A_v}$ and $\mathbf{A}$ in a univariate linear model for a trait of interest, in our case, traits that are potentially affected by pleiotropic effects of the polled locus, allows separation of additive QTL effects of the polled locus from the remaining additive polygenic effects. The outlined VC estimation approach to separate polygenic and single-locus QTL effects has been extended to multivariate models to increase the power of QTL detection [22, 25]. When applied to the main question of our study, a bivariate approach including phenotypes for the polled trait will also enable the estimation of the genetic correlations based on $\mathbf{A_v}$, which would reflect the direction of the pleiotropic effects.

Our aim was to derive complete polled genotypes in a complex pedigree to be able to calculate the IBD probabilities and to test for pleiotropic or linked QTL effects of the Mendelian polled locus on production and reproduction traits in German Simmental cattle via univariate and bivariate VC estimation.

Methods

Phenotypes and genotypes for polledness

The polled status in German Simmental cattle is routinely recorded via a collaboration between farmers and the Bavarian milk recording organization [21]. Only polled animals are specifically registered in the database using the following labels: PP = homozygous polled; Pp = heterozygous polled; PS = heterozygous polled, scurred; P = polled, exact genotype unknown. Animals with a genotype test result are marked with an asterisk (*) that is added to the polled label. Animals without any label in the database are assumed to be horned. Sorting and selection of the polled animals from the German Simmental database was performed in two steps. First, we identified 89 farms for which between 25 and 75% of the animals were registered as polled at the end of 2015, in order to generate an almost balanced sampling

design of polled and horned animals in target farms. As the number of newborn polled calves has only recently started to increase [21], we included a second filtering step that selected farms with at least one polled calf born per year during the 2007–2013 period. Thus, we created a final dataset including 24 farms with 2420 polled and horned cows for which phenotypes for production and fertility traits from multiple generations were available. Genotype frequencies prior to the inference of missing and falsely-registered polled genotypes in the final dataset are in Table 1. Based on the 2420 selected cows, the full pedigree included 13,256 animals traced back five generations.

Inference of polled genotypes in the full pedigree to calculate the A_V matrix

Available polled labels for ancestors of the selected cows in the final dataset were taken from the literature [17, 21, 26, 27], and extracted from public databases for Simmental and other breeds such as Holstein and Brown Swiss. The method to infer missing and falsely-registered polled genotypes in the full pedigree consisted of different procedures and tests, which were all performed using self-written R [28] functions and scripts (see Additional file 1). The main steps of the algorithm were: (i) iterative reconstruction of missing parent genotypes, (ii) correction of Mendelian inheritance errors, and (iii) derivation of polled genotypes based on progeny genotype statistics. For further detailed information (see Additional file 1). Table 1 gives an overview of the frequencies of genotype labels in the dataset including ancestors after the inference procedures.

A comparison of the unchanged initial genotypes as reported from the routine records and the genotypes after inference showed a significant amount of non-registered (e.g. progeny of gene- or progeny-tested homozygous polled sires or dams) and also falsely-registered (e.g. polled progeny from horned matings) animals (see Table 1) and (see Additional file 3). Hence, in a preliminary study that focused on variance component models for the trait polledness, the inference of polled genotypes greatly improved the model fit compared to the expectations for a Mendelian trait (see Additional file 3).

Table 1 Descriptive statistics for the polled trait in the analyzed real dataset before and after inference of genotypes

Dataset	Number of animals	pp	Pp	PP	*p*(P)
Animals with phenotypes					
Initially registered	2420	1605 (28)	771 (65)	44 (4)	0.177
After inference	2420	1532	863	25	0.189
Full pedigree	13,256	11,737 (28)	1428 (264)	91 (68)	0.061

pp: horned, Pp: heterozygous polled, PP: homozygous polled, *p*(P): polled allele frequency

The number of animals with a gene-test result at the polled locus is given between brackets

Calculation of A_V

Inferred genotypes at the polled locus from all 13,256 animals in the pedigree were used to compute the probabilities of inheriting the paternal or maternal alleles from the sire and dam at the polled locus, starting from the founders. Identical-by-descent (IBD) probabilities between the alleles at the polled locus of any two founders were assumed to be zero. Genotypes at the polled locus were treated as a tri-allelic marker. Hence, polled alleles of founder animals from breeds that predominantly carry the "Celtic" polled allelic variant such as Simmental, Brown Swiss and most beef breeds are initially coded differently than those of founder animals from breeds that predominantly carry the "Friesian" polled allelic variant, such as Holstein and Jersey [3]. However, this differentiation is only effective for informative matings, thus for uninformative matings both polled allelic variants are treated as the same allele.

The $\mathbf{G_V}$ matrix based on previously computed IBD probabilities at the polled locus was computed using the algorithm reported by van Arendonk et al. [24] with a self-written R function (see Additional file 2). $\mathbf{G_V}$ as a gametic relationship matrix has an order twice the number of animals. Therefore, after completing the computation, $\mathbf{G_V}$ was scaled down to the dimensions of a marker-based numerator relationship matrix $\mathbf{A_V}$ using the matrix transformation $\mathbf{A_V} = \frac{1}{2}\mathbf{K}\mathbf{G_V}\mathbf{K}'$ with $\mathbf{K} = \mathbf{I}_n \otimes [\mathbf{1}, \mathbf{1}]$ and n the number of animals. The inverse of $\mathbf{A_V}$ was computed using the function solve() from the base package in R [28], and was then used in the VC estimations of QTL effects.

Cow traits

Test-day production traits included milk yield (MY), protein percentage (P%), fat percentage (F%) and the health indicator somatic cell score (SCS). In total, 58,262 test-day records from lactations 1 to 7 recorded from 2005 to 2015 were included for parameter estimation. Female fertility traits were binary 56 days non-return-rate (NRR-56), days open (DO) and days to first service (DFS) for cows. NRR-56, DFS and DO were calculated from routine insemination data collected from 2004 to 2014. Only the first four lactations were considered for parameter estimation of female fertility traits. Descriptive statistics for all analyzed cow traits are in Table 2.

Table 2 Descriptive statistics for all evaluated production, SCS and fertility traits for the three polled genotype groups

Trait	Total number of records	Horned			Polled					
		pp			Pp			PP		
		n	Mean	SD	n	Mean	SD	n	Mean	SD
Milk yield/d	58,262	38,335	24.94	8.39	19,566	22.23	8.52	361	20.54	8.11
Protein percentage	58,262	38,335	3.54	0.39	19,566	3.46	0.40	361	3.42	0.41
Fat percentage	58,262	38,335	4.09	0.72	19,566	4.04	0.73	361	4.19	0.64
Somatic cell score	57,538	37,810	2.68	1.68	19,370	2.48	1.65	358	2.85	1.66
Non return rate 56	3333	2227	0.34	0.47	1093	0.30	0.46	13	0.23	0.44
Days open	4223	2824	31.46	51.37	1383	28.38	44.53	16	26.12	45.28
Days to first service	4223	2824	93.91	55.80	1383	91.59	49.86	16	90.19	50.34

Statistical models

Variance components for all cow traits were estimated in single trait animal models using the software package DMU and the implemented AI-REML algorithm [29]. The basic linear model without QTL effects (M_{Basic}) for production, fertility and health traits was defined as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of cow traits, $\boldsymbol{\beta}$ is a vector of fixed effects, $\mathbf{u}$ is a vector of random additive polygenic effects, $\mathbf{p}$ is a vector of random permanent environment effects, and $\mathbf{e}$ is a vector of random residual effects. $\mathbf{X}$, $\mathbf{Z_a}$ and $\mathbf{Z_p}$ are the incidence matrices for fixed, additive-genetic and permanent environmental effects, respectively. The random effects $\mathbf{a}$, $\mathbf{p}$ and $\mathbf{e}$ were assumed to be uncorrelated and to follow univariate normal distributions as follows: $\mathbf{u} \sim N_q(\mathbf{0}, \mathbf{A}\sigma_a^2)$, $\mathbf{p} \sim N_m\left(\mathbf{0}, \mathbf{I}\sigma_p^2\right)$ and $\mathbf{e} \sim N_m(\mathbf{0}, \mathbf{R}\sigma_e^2)$, with σ_a^2, σ_p^2 and σ_e^2 being the polygenic variance, permanent environmental variance and residual variance, respectively. $\mathbf{A}$ is the standard additive genetic relationship matrix and $\mathbf{I}$ is and identity matrix and $\mathbf{R}$ is a known diagonal matrix.

Fixed effects considered in the models for the test-day traits were lactation, herd-test day and calving season. Days pregnant and calving age (linear regression) and days in milk (Legendre polynomials of order 3) were considered as covariates.

A threshold model using a logit link function was defined for the NRR-56 trait without changing the structure of fixed and random effects described above. Fixed effects in the models for the fertility traits were the combined effects of herd-year, type of insemination-year, and lactation-calving age.

The extended linear univariate QTL models (M_{QTL}) were defined as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z_a a} + \mathbf{Z_v v} + \mathbf{Z_p p} + \mathbf{e},$$

for production and fertility traits, respectively, with the same properties as described for M_{Basic} but adding $\mathbf{v}$, a vector of additive QTL effects with the distribution $\mathbf{v} \sim N_q\left(\mathbf{0}, \mathbf{A_v}\sigma_v^2\right)$, and the corresponding incidence matrix $\mathbf{Z_v}$.

For the bivariate analyses, we defined the following models based on the basic and extended QTL model as described above. The basic linear models without QTL effects ($M_{Basic\ bivariate}$) for the quantitative traits and the Mendelian trait polledness were defined as:

$$\mathbf{y}_1 = \mathbf{X\beta} + \mathbf{Z_a a} + \mathbf{Z_p p} + \mathbf{e},$$

$$\text{and } \mathbf{y}_2 = \mathbf{Z_a a} + \mathbf{e},$$

where $\mathbf{y}_1$ is a vector of phenotypes for the respective quantitative trait, and $\mathbf{y}_2$ is a vector of phenotypes for the Mendelian trait polledness. Based on preliminary studies that focused on the identification of suitable VC models for the analysis of the Mendelian polled trait, we chose numerically-coded genotype labels, which represent the allele content at the polled locus as phenotypes for the trait polledness (i.e., 0 for horned animals, 1 for heterozygous polled animals and 2 for homozygous polled animals). All effect categories were identical to those in the univariate models. Due to the Mendelian inheritance of the polledness trait, environmental effects do not affect the phenotype. Thus, fixed effects were excluded from VC estimation for polledness in the bivariate models.

The extended linear QTL models ($M_{QTL\ bivariate}$) were defined as:

$$\mathbf{y}_1 = \mathbf{X\beta} + \mathbf{Z_a a} + \mathbf{Z_v v} + \mathbf{Z_p p} + \mathbf{e},$$

$$\text{and } \mathbf{y}_2 = \mathbf{Z_a a} + \mathbf{Z_v v} + \mathbf{e},$$

with the same properties as described for the basic models, and adding $\mathbf{v}$, a vector of additive QTL effects with the distribution $\mathbf{v} \sim N_q\left(\mathbf{0}, \mathbf{A_v}\sigma_v^2\right)$, and $\mathbf{Z_v}$, an incidence

matrix relating animals to phenotypes as in the univariate QTL models.

Based on covariance estimates from the bivariate models, we estimated the genetic correlation based on QTL effects ($r_{g\ v}$) (i.e., the "monogenic" genetic correlation between the polledness trait and the evaluated quantitative traits), the genetic correlation based on additive polygenic effects ($r_{g\ a}$) and the phenotypic correlation (r_p).

Test statistics

The hypothesis test for the presence of pleiotropic QTL effects at the polled locus was based on the asymptotic distribution of the likelihood ratio test (LRT) statistic:

$$\mathrm{LRT} = -2ln\left(L_{\mathrm{BASIC}} - L_{\mathrm{QTL}}\right),$$

where L_{BASIC} and L_{QTL} were the maximized likelihoods under $\mathrm{M_{Basic}}$ and $\mathrm{M_{QTL}}$, respectively. The asymptotic distribution of the likelihood ratio test statistic follows a χ^2-distribution, with the number of degrees of freedom equal to the difference between the number of independent parameters of both models [20]. LRT tests for the presence of pleiotropic QTL-effects at the polled locus were only calculated for the univariate models after VC estimation.

Although the outlined VC approach appears to be straightforward for our research question, to our knowledge no previous studies have proved its applicability using either simulated or real data. As flexible stochastic simulation packages to simulate precise genomic trait architectures are available [30, 31], we decided to validate our approach by simulation. Hence, we performed a small preliminary stochastic simulation study prior to the analyses of the real data. The results of the preliminary simulation study are in Additional file 4.

Results

Table 2 includes the descriptive statistics for all the evaluated traits separated by polled genotype groups. The means for the production traits clearly reflect the phenotypic inferiority of polled animals, e.g. horned cows (pp) produced 2.71 kg and 4.40 kg more milk on average than polled Pp and PP animals, respectively. In contrast, the means for SCS and fertility traits reflect an advantage of the polled compared to the horned cows. Standard deviations were in similar ranges across all traits and groups.

Standard errors of the heritabilities for all production and fertility traits from both models $\mathrm{M_{Basic}}$ and $\mathrm{M_{QTL}}$ were quite small and in the range from 0.018 to 0.049 for the univariate and from 0.014 to 0.048 for the bivariate models (see Tables 3 and 4). Estimates of the permanent environmental variances, residual variances and

Table 3 Variance components for test-day production traits and SCS in the real dataset

Trait	Model	σ^2_a	σ^2_v	σ^2_{PE}	σ^2_e	QTL − h^2	h^2(SE) (polygenic + QTL)	LRT p (λ)
MY (kg)	Basic (univariate)	5.919		5.007	14.862		0.230 (0.025)	1 (− 0.1e−03)
	QTL (univariate)	5.919	0.778e−05	5.007	14.862	0.302e−06	0.230 (0.027)	
	Basic (bivariate)	6.000		4.922	14.867		0.233 (0.025)	
	QTL (bivariate)	6.012	0.816e−04	4.914	14.867	0.316e−05	0.233 (0.027)	
F (%)	Basic (univariate)	0.109		0.014	0.283		0.268 (0.018)	0.597 (0.279)
	QTL (univariate)	0.106	0.002	0.015	0.283	0.005	0.267 (0.022)	
	Basic (bivariate)	0.109		0.014	0.283		0.269 (0.019)	
	QTL (bivariate)	0.109	0.460e−03	0.014	0.283	0.001	0.268 (0.020)	
P (%)	Basic (univariate)	0.033		0.006	0.040		0.414 (0.025)	0.047 (3.941)
	QTL (univariate)	0.030	0.002	0.006	0.040	0.023	0.411 (0.033)	
	Basic (bivariate)	0.032		0.006	0.040		0.410 (0.026)	
	QTL (bivariate)	0.032	0.245e−04	0.006	0.040	0.313e−03	0.410 (0.028)	
SCS	Basic (univariate)	0.224		0.636	1.543		0.093 (0.019)	0.405 (0.692)
	QTL (univariate)	0.211	0.008	0.639	1.543	0.004	0.091 (0.020)	
	Basic (bivariate)	0.236		0.626	1544		0.098 (0.019)	
	QTL (bivariate)	0.233	0.576e−03	0.626	1.544	0.240e−03	0.097 (0.022)	

Corresponding variance components for the polledness trait from the bivariate models are included in Additional file 5

MY: milk yield; F: fat, P: protein; SCS: somatic cell score

σ^2_a = additive genetic variance based on *A*; σ^2_v = additive genetic variance based on $\mathbf{A_v}$; σ^2_{PE} = permanent environment variance; σ^2_e = residual variance; QTL − h^2 = QTL heritability calculated as $\sigma^2_v/\sigma^2_a + \sigma^2_v + \sigma^2_{PE} + \sigma^2_e$; h^2 (SE) = overall heritability and standard error (in brackets) calculated as $\sigma^2_v + \sigma^2_a/\sigma^2_a + \sigma^2_v + \sigma^2_{PE} + \sigma^2_e$; LRT p (λ) = p- and lambda values from likelihood ratio tests

Table 4 Variance components for fertility traits in the real dataset

Trait	Model	σ_a^2	σ_v^2	σ_{PE}^2	σ_e^2	QTL − h^2	h^2(SE) (polygenic + QTL)	LRT p (λ)
NRR-56 (binary)	Basic (univariate)	0.075		0.409	3.290		0.020 (0.038)	1 (− 0.407)
	QTL (univariate)	0.047	0.018	0.416	3.290	0.005	0.017 (0.049)	
	Basic (bivariate)	0.061		0.232	3.290		0.017 (0.030)	
	QTL (bivariate)	0.061	0.531e−03	0.228	3.290	0.148e−03	0.017 (0.048)	
DFS* (days)	Basic (univariate)	54.142		209.418	2098.704		0.023 (0.018)	0.663 (0.190)
	QTL (univariate)	39.286	12.249	210.966	2097.825	0.005	0.022 (0.024)	
	Basic (bivariate)	45.011		220.952	2096.554		0.019 (0.016)	
	QTL (bivariate)	38.886	0.825	220.753	2097.187	0.350e−03	0.017 (0.014)	
DO (days)	Basic (univariate)	55.600		102.212	1931.358		0.027 (0.018)	0.434 (0.613)
	QTL (univariate)	26.347	25.290	102.586	1930.923	0.012	0.025 (0.027)	
	Basic (bivariate)	46.174		112.101	1930.572		0.022 (0.016)	
	QTL (bivariate)	31.215	11.429	110.559	1931.022	0.005	0.020 (0.024)	

Corresponding variance components for the polled trait from the bivariate models are included in Additional file 5

NRR-56: non return rate 56; DFS: days to first service; DO: days open

σ_a^2 = additive genetic variance based on **A**; σ_v^2 = additive genetic variance based on $\mathbf{A_v}$; σ_{PE}^2 = permanent environment variance; σ_e^2 = residual variance; QTL − h^2 = QTL heritability calculated as $\sigma_v^2/\sigma_a^2 + \sigma_v^2 + \sigma_{PE}^2 + \sigma_e^2$; h^2 (SE) = overall heritability and standard error (in brackets) calculated as $\sigma_v^2 + \sigma_a^2/\sigma_a^2 + \sigma_v^2 + \sigma_{PE}^2 + \sigma_e^2$; LRT p (λ) = p- and lambda values from likelihood ratio tests

*Since the full bivariate model including **A** and $\mathbf{A_v}$ for the trait DFS did not fully converge, we present the results for a model excluding **A** in the model for polledness.

heritabilities for the same traits from the different univariate and bivariate models were almost identical. The estimated heritabilities for the performance traits ranged from 0.23 for MY to 0.41 for P%. The estimated heritability for SCS was 0.09. The estimated heritabilities for the female fertility traits ranged from 0.02 for NRR-56 to 0.03 for DO.

With regard to the M_{QTL} applications, we detected a statistically significant QTL effect of the polled locus that contributed up to 6.6% of the genetic variance and 2.3% of the phenotypic variance for P%, respectively. However, the estimated QTL effect in the corresponding bivariate model was substantially smaller. The genetic correlation based on the estimated direct QTL effect was slightly negative (see Table 5).

The estimated direct QTL effects of the polled locus on production traits MY, F% and on SCS were consistently low and not significant. Comparison of the univariate with the bivariate QTL models indicates that the QTL effects from the bivariate models are smaller for the moderately heritable production traits. All the estimates of QTL effects for the fertility traits were non-significant and those from the corresponding bivariate models were considerably smaller, thus following the same trend as for the production traits.

The estimates for phenotypic and genetic correlations (Table 5) support the very small and non-significant estimated direct QTL effects of the polled locus. The genetic correlations estimated from the direct effect of the polled locus ($r_{g\,v}$) reflect no antagonistic but rather neutral relationships with test-day production traits. This is also the case for P% for which we detected a small QTL effect in the univariate models. Genetic correlations based on direct effects of the polled locus with SCS and all the fertility traits were favorable from a breeding perspective. Interestingly, the genetic correlations estimated from the separated polygenic effects based on the pedigree ($r_{g\,a}$) were close to zero for all traits, which further disprove any antagonistic relationships associated with the polled status.

Table 5 Phenotypic and genetic correlations between polledness, test-day production traits, SCS and fertility traits

Trait	r_P	$r_{g\,v}$(SE)	$r_{g\,a}$(SE)
MY	− 0.004	− 0.003 (0.068)	− 0.005 (0.068)
F%	− 0.004	− 0.010 (0.056)	0.002 (0.055)
P%	− 0.005	− 0.010 (0.052)	0.003 (0.051)
SCS	− 0.015	− 0.048 (0.101)	0.001 (0.096)
NRR-56	− 0.020	− 0.093 (0.372)	0.002 (0.340)
DFS	− 0.019	− 0.029 (0.042)	/*
DO	− 0.017	− 0.120 (0.248)	0.002 (0.221)

Standard errors for the genetic correlations were calculated based on heritability estimates and corresponding standard errors according to [15]

r_P: phenotypic correlation calculated based on variance estimates from the QTL models; $r_{g\,v}$ (SE): genetic correlation and standard errors (in brackets) calculated based on variance estimates for *v*; $r_{g\,a}$ (SE): genetic correlation and standard errors (in brackets) calculated based on variance estimates for *a*

*Since the full bivariate model including **A** and $\mathbf{A_v}$ for the trait DFS did not fully converge, we present the results for a model excluding **A** in the model for polledness; $r_{g\,a}$. cannot be estimated from this model.

The results for the preliminary simulation study showed a general suitability of our approach for the detection of predefined QTL effects (see Additional file 4). However, the estimated QTL effects for the simulated trait with a low heritability were statistically non-significant, in contrast to the results for the simulated moderately heritable trait. Interestingly, the bivariate QTL models generally performed better than the univariate models and showed the highest accuracy to detect the simulated QTL effects. The estimates of the heritability of polledness obtained with the bivariate models for the numerically-coded allele content phenotype were close to 1, and thus close to the expected value of 1 for a monogenic Mendelian trait (see Additional file 5).

Discussion

The descriptive statistics in our dataset are in line with the reported inferiority of polled animals for production traits and the benefits for health and fertility traits [17, 21]. The moderate estimates of the heritability obtained for the test-day production traits MY, F% and P% and the low estimate of the heritability for test-day SCS are in line with previously reported estimates in the German Simmental population [32, 33]. The low estimates for heritability of fertility traits are also in line with those found in the literature for the German and Austrian Simmental populations [34, 35].

Estimates of the variance component for the production traits, SCS and the female fertility traits revealed no significant QTL effects, apart from a moderately large contribution (i.e. ~ 5%) of the polled locus to the genetic variance of P%. In consequence, we refuted any concerns for possible unfavorable effects of the polled locus on production, udder health and female fertility traits in German Simmental cattle. At first sight, this contrasts with previous studies in the Simmental population based on the comparison of breeding values between polled and horned sires [17, 21]. However, to our knowledge, our study is the first attempt to separate direct pleiotropic effects of the polled locus from other negative polygenic effects present in polled families.

Based on our results, there is no general antagonistic or detrimental effect of the polled alleles on productivity, udder health and female fertility. The formerly general and presently partial inferiority of polled animals is most likely due to genetic drift caused by a small number of polled Simmental founders or ancestors from beef type cattle with inferior genetic values in dairy traits. Strong associations due to pleiotropy or tight linkage would imply unidirectional effects across breeds that are unaffected by selection over time. In contrast, associations due to genetic drift have a random nature and can be altered by selection and recombination. Thus, the closing gap in production traits between polled and horned animals in Simmental [21] and Holstein [36] due to selection clearly points to genetic drift as the reason for the initial breeding value inferiority of polled animals. This is further supported by the striking analogy that polled animals in the Simmental and Holstein breeds both descend from very small inferior polled founder pools [21, 36]. Recently published results in beef cattle breeds that are under intensive selection for polledness since the last 20 years provide additional evidence that there are no systematic detrimental effects on growth and carcass traits across breeds in this regard [37]. Therefore, the remaining breeding value inferiority of polled sires can be overcome by selective breeding, i.e., continued targeted mating of superior young polled sires with superior horned dams while eliminating polled selection candidates. Nonetheless, linkage based on long-range LD might also contribute partly to the remaining deficit in production traits of polled animals, even if no direct pleiotropic effects exist.

Our results also give further support to previously published simulation results on breeding strategies for polledness (e.g. [13, 14, 38–41]), since all the studies implicitly ignored potential persisting pleiotropic effects. In summary, a moderate selection intensity for polledness of genomically-selected sire candidates appears to be the best compromise to narrow down the gap in production traits even more between polled and horned animals in the population while preserving population-wide maximal genetic gain [38, 39]. In German Simmental, highly intensified selection, especially among cows, which for example could be based on an index incorporating polledness, involves the risk of losing genetic gain because of the current inferiority of polled animals [13, 14, 38]. Recent simulation studies that applied novel gene-editing methods showed the potential of this technology to potentially overcome such a loss of genetic gain [42, 43], and polledness is one of the first traits for which the technology was successfully applied in cattle [44–46].

It should be noted that selective breeding for polledness in German Simmental is an ongoing dynamic process that has significantly changed and still changes the frequency of the polled allele in the population. Therefore, given that the estimates of QTL effect and variance depend on allele frequency, we recommend monitoring the effects of the polled locus in the future, when the proportion of polled animals has substantially increased. In addition, although non-significant, the estimated QTL effects for NRR-56, DFS and DO reflect significant proportions of the overall estimated genetic variance for each of these traits. This may be due to a lack of power to accurately estimate the effects of the polled locus on

these traits, and a larger sample size should be used in future studies.

Nonetheless, we were able to show that a relationship matrix based on polled genotypes ($\mathbf{A_v}$) allowed us to estimate direct QTL effects of the polled trait. However, with the broad availability of imputed data on polled genotypes from commercially-used SNP chips [3, 47], we are confident that larger datasets from polled animals will become broadly available to further validate our results. Certainly, in addition to this methodological aspect, broad genotyping and genomic selection will also practically improve selection of superior polled animals.

Given that the number of polled animals in the population is constantly growing and that young animals are much more frequently genotyped in the current genomic breeding schemes, it will therefore be much easier to analyze extensive datasets in German Simmental cattle and other breeds by using the methodological framework developed here. In this regard, explicitly modelling and dissecting more precisely the potentially diverging effects of the known polled variants such as P_c and P_f based on genotype data could help to answer the remaining questions concerning different family effects in association with the polled trait [18, 48].

Conclusions

Our results reveal no direct pleiotropic or linked QTL effects of the polled locus on the studied production traits MY and F%, on the udder health indicator SCS and on the female fertility traits, NRR56, DO and DFS, in German Simmental cattle. Only one statistically significant direct QTL effect of the polled locus on P%, was detected. Thus, further selection on polledness is not expected to result in negative side effects on breeding goal traits. Based on our results, we conclude that any remaining inferiority of polled cows and bulls will be reduced by increasing the dissemination of the polled alleles and by intensive simultaneous selection on breeding goal traits and polled genotypes in the German Simmental population.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12711-021-00652-z.

Additional file1. Rscripts used for the Inference of polled genotypes in the full pedigree.

Additional file 2. Rscripts used for the calculation of $\mathbf{A_v}$.

Additional file 3. Preliminary study on variance component models for polledness. Preliminary study on variance component estimation for the trait polledness including a QTL relationship matrix based on inferred genotypes from the pedigree in univariate models.

Additional file 4. Preliminary simulation study. Preliminary stochastic simulation study to validate the variance component estimation approach.

Additional file 5. Variance component estimation results from bivariate analysis for simulated and real data for the trait polledness.

Acknowledgements
The authors gratefully thank the Landwirtschaftliche Rentenbank for funding this project.

Authors' contributions
CS designed the study, performed all the analyses and drafted the manuscript. RE and KUG prepared the phenotypic data. SK initiated the primary research. SK and KUG assisted in writing the manuscript. All authors read and approved the final manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. The present study was supported by funds of the German Government´s Special Purpose Fund held at Landwirtschaftliche Rentenbank.

Availability of data and materials
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable. All phenotypes used in this study were recorded from animals during routine performance testing without special testing requirements.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interest.

Author details
[1]Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Giessen, Germany. [2]Bavarian State Research Center for Agriculture, Institute of Animal Breeding, Prof. Dürrwaechter-Platz 1, 85586 Poing-Grub, Germany.

Received: 23 February 2021 Accepted: 29 June 2021
Published online: 14 July 2021

References

1. Bateson W, Saunders ER. The facts of heredity in the light of Mendel's discovery. Rep Evol Comm R Soc. 1902;1:125–60.
2. Long CR, Gregory KE. Inheritance of the horned, scurred and polled condition in cattle. J Hered. 1978;69:395–400.
3. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Gopel KH, et al. Bovine polledness–an autosomal dominant trait with allelic heterogeneity. PLoS One. 2012;7:e39477.
4. Medugorac I, Graf A, Grohs C, Rothammer S, Zagdsuren Y, Gladyr E, et al. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nat Genet. 2017;49:470–5.
5. Stafuzza NB, Silva RMO, Peripolli E, Bezerra LAF, Lôbo RB, Magnabosco CU, et al. Genome-wide association study provides insights into genes related with horn development in Nelore beef cattle. PLoS One. 2018;13:e0202978.
6. Randhawa IAS, Burns BM, McGowan MR, Porto-Neto LR, Hayes BJ, Ferretti R, et al. Optimized genetic testing for polledness in multiple breeds of cattle. G3 (Bethesda). 2020;10:539–44.
7. Rothammer S, Capitan A, Mullaart E, Seichter D, Russ I, Medugorac I. The 80-kb DNA duplication on BTA1 is the only remaining candidate

mutation for the polled phenotype of Friesian origin. Genet Sel Evol. 2014;46:44.
8. Gehrke LJ, Capitan A, Scheper C, König S, Upadhyay M, Heidrich K, et al. Are scurs in heterozygous polled (Pp) cattle a complex quantitative trait? Genet Sel Evol. 2020;52:6.
9. Nicholas FW, Hobbs M. Mutation discovery for Mendelian traits in non-laboratory animals: a review of achievements up to 2012. Anim Genet. 2014;45:157–70.
10. Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, et al. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. J Dairy Sci. 2016;99:8915–31.
11. Cole JB, Null DJ, VanRaden PM. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. J Dairy Sci. 2016;99:7274–88.
12. Dorshorst B, Henegar C, Liao X, Sallman Almen M, Rubin CJ, Ito S, et al. Dominant red coat color in Holstein cattle is associated with a missense mutation in the *coatomer protein complex, subunit alpha* (*COPA*) gene. PLoS One. 2015;10:e0128969.
13. Cole JB. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genet Sel Evol. 2015;47:94.
14. Segelke D, Taubert H, Reinhardt F, Thaller G. Considering genetic characteristics in German Holstein breeding programs. J Dairy Sci. 2016;99:458–67.
15. Falconer DS, Mackay TFC. Introduction to quantitative genetics. Harlow: Longman; 1996.
16. Sax K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics. 1923;8:552–60.
17. Lamminger A, Hamann H, Röhrmoser G, Rosenberger E, Kräusslich H, Distl O. Relationships between polledness and traits used in the breeding objectives for German Fleckvieh. Züchtungskunde. 2000;72:325–39.
18. Gehrke L, Seichter D, Ruß I, Medugorac I, Tetens J, Thaller G. Impact of the Friesian POLLED mutation on milk production traits in Holstein Friesian. In: Proceedings of the 67th Annual Meeting of the European Federation of Animal Science: 29 August - 2 September 2016; Belfast; 2016. p.434.
19. Dressel, H., Wensch-Dorendorf, M., Swalve HH. Genetic relationship between polledness and other relevant traits in Holstein dairy cattle. In: Proceedings of the 67th Annual Meeting of the European Federation of Animal Science: 29 August - 2 September 2016; Belfast; 2016. p.491.
20. Cole JB, Null DJ. Short communication: Phenotypic and genetic effects of the polled haplotype on yield, longevity, and fertility in US Brown Swiss, Holstein, and Jersey cattle. J Dairy Sci. 2019;102:8247–50.
21. Götz KU, Luntz B, Robeis J, Edel C, Emmerling R, Buitkamp J, et al. Polled Fleckvieh (Simmental) cattle – Current state of the breeding program. Livest Sci. 2015;179:80–5.
22. George AW, Visscher PM, Haley CS. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics. 2000;156:2081–92.
23. van Arendonk JAM, Tier B, Bink MCAM, Bovenhuis H. Restricted maximum likelihood analysis of linkage between genetic markers and quantitative trait loci for a granddaughter design. J Dairy Sci. 1998;81:76–84.
24. van Arendonk JAM, Tier B, Kinghorn B. Use of multiple genetic-markers in prediction of breeding values. Genetics. 1994;137:319–29.
25. Sørensen P, Lund MS, Guldbrandtsen B, Jensen J, Sorensen D. A comparison of bivariate and univariate QTL mapping in livestock populations. Genet Sel Evol. 2003;35:605–22.
26. Brem G, Karnbaum B, Rosenberger E. Zur Vererbung der Hornlosigkeit beim Fleckvieh. Bayerisches landwirtschaftliches Jahrbuch. 1992;59:688–95.
27. Lange H, Brem G, Utz J, Gottschalk A, Karnbaum B, Kräußlich H. Untersuchungen ueber Hornlosigkeit beim Deutschen Fleckvieh. Bayerisches landwirtschaftliches Jahrbuch. 1990:16–68.
28. Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
29. Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau R. DMU - a package for analyzing multivariate mixed models. In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production: 13–18 August 2006; Belo Horizonte; 2006.
30. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 2009;25:680–1.
31. Faux A-M, Gorjanc G, Gaynor RC, Battagin M, Edwards SM, Wilson DL, et al. AlphaSim: Software for breeding program simulation. Plant Genome. 2016. https://doi.org/10.3835/plantgenome2016.02.0013.
32. Gerber A, Krogmeier D, Emmerling R, Götz K-U. Analysis of genotype by environment interaction for milk yield traits in first lactation of Simmental cattle. J Anim Breed Genet. 2008;125:382–9.
33. Emmerling R, Lidauer M, Mäntysaari E. Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. Interbull Bull. 2002;29:111–7.
34. Fuerst C, Gredler B. Genetic evaluation for fertility traits in Austria and Germany. Interbull Bull. 2009;40:3–9.
35. Gredler B, Fuerst C, Sölkner J. Analysis of new fertility traits for the joint genetic evaluation in Austria and Germany. Interbull Bull. 2007;37:152–5.
36. Windig JJ, Hoving-Bolink RA, Veerkamp RF. Breeding for polledness in Holstein cattle. Livest Sci. 2015;179:96–101.
37. Randhawa IAS, McGowan MR, Porto-Neto LR, Hayes BJ, Lyons RE. Comparison of genetic merit for weight and meat traits between the polled and horned cattle in multiple beef breeds. Animals (Basel). 2021;11:870.
38. Scheper C, Wensch-Dorendorf M, Yin T, Dressel H, Swalve H, Konig S. Evaluation of breeding strategies for polledness in dairy cattle using a newly developed simulation framework for quantitative and Mendelian traits. Genet Sel Evol. 2016;48:50.
39. Gaspa G, Veerkamp RF, Calus MPL, Windig JJ. Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation. Livest Sci. 2015;179:86–95.
40. Cole JB. Management of Mendelian traits in breeding programs by gene editing: A simulation study. bioRxiv. 2017. https://doi.org/10.1101/116459.
41. Spurlock DM, Stock ML, Coetzee JF. The impact of 3 strategies for incorporating polled genetics into a dairy cattle breeding program on the overall herd genetic merit. J Dairy Sci. 2014;97:5265–74.
42. Mueller ML, Cole JB, Connors NK, Johnston DJ, Randhawa IAS, van Eenennaam AL. Comparison of gene editing versus conventional breeding to introgress the POLLED allele into the tropically adapted Australian beef cattle population. Front Genet. 2021;12:593154.
43. Mueller ML, Cole JB, Sonstegard TS, van Eenennaam AL. Comparison of gene editing versus conventional breeding to introgress the POLLED allele into the US dairy cattle population. J Dairy Sci. 2019;102:4215–26.
44. Young AE, Mansour TA, McNabb BR, Owen JR, Trott JF, Brown CT, et al. Genomic and phenotypic analyses of six offspring of a genome-edited hornless bull. Nat Biotechnol. 2020;38:225–32.
45. Carlson DF, Lancto CA, Zang B, Kim ES, Walton M, Oldeschulte D, et al. Production of hornless dairy cattle from genome-edited cell lines. Nat Biotechnol. 2016;34:479–81.
46. Schuster F, Aldag P, Frenzel A, Hadeler KG, Lucas-Hahn A, Niemann H, et al. CRISPR/Cas12a mediated knock-in of the Polled Celtic variant to produce a polled genotype in dairy cattle. Sci Rep. 2020;10:13570.
47. GeneControl GmbH. GeneControl. 2017. http://www.genecontrol.de/e_hornlosigkeit.html. Accessed 1 Oct 2017.
48. Gehrke L. Investigations of the genetic architecrure of horns in cattle (*Bos taurus*) for improved breeding strategies for polledness. PhD Thesis, University of Kiel. 2020.

Publisher's Note

CHAPTER 4

Assessing selection signatures and genomic associations to performance and functional traits on BTA1 in polled and horned Holstein sire groups

Carsten Scheper and Sven König

submitted to Genetics Selection Evolution

Own contribution to the publication:

Study design:	mostly autonomous
Collection of data:	participated
Statistical analysis:	autonomous
Manuscript:	mostly autonomous

Abstract

Background:

Intensified selection for polledness is ongoing in the German Holstein cattle population due to increasing farmer demands. Although general detrimental associated effects can be ruled out with growing certainty, deeper insights on selection signatures and associations in close genomic proximity to the polled locus on BTA1 are needed.

Methods:

The analyzed dataset consisted of 314 polled (142 Red and White Holstein (RH), 172 Black and White Holstein (BWH)) and 1846 horned (248 RH and 1598 BWH) bulls from birth years 2010-2016. After quality control, 2817 SNPs located on BTA1 and 41,600 SNPs located on BTA2 to BTA29 from all bulls were used for genomic analyses. Fst and xp-EHH were calculated to assess selection signatures separately for RH and BWH comparing polled and horned subpopulations. Available phenotypes for association analysis included 3 performance traits (milk yield, protein and fat yield) and 2 functional traits (non-return rate and somatic cell score). Association analyses for these traits were carried out on BTA1 using the mixed model approach implemented in GCTA. Potential candidate genes near selection or association signals were annotated based on the SNP position ±25kb.

Results:

The results from selection signature analyses indicate that the region on BTA1 directly influenced by recent or past selection for polledness is limited from 0-25Mb. A total of 89 (Fst) and 29 (XP-EHH) SNPs showed significant selection signals, with the majority of these SNPs located at the proximal end of BTA1. 29 genes were annotated to the detected SNPs. On all other chromosomes (BTA2 to BTA29), a total of 68 (Fst) and 36 (XP-EHH) additional significant selection signals were identified, corresponding to 28 annotated genes. No significant association signals according to chromosome-wide and genome-wide Bonferroni corrected significance thresholds were found for the analyzed production and functional traits on BTA1.

Conclusions:

The segment directly influenced by selection on BTA1 in the Holstein breed is limited to the proximal end of BTA1 (0-25Mb) as confirmed by selection signature analysis. In this region, no significant associations with other phenotypic features were detected. Hence, there are no further indications that significant direct effects of selection at the polled locus affect secondary traits due to selection in combination with linkage in polled Holstein cattle. However, further selection signatures on other chromosomes were found when comparing polled and horned subgroups.

Background

There is a continuously growing demand for polled bulls in Germany as animal welfare focused legislation increasingly restricts dehorning practices. The German Holstein breeding associations have adapted to this demand by intensifying the selection for polledness in recent years, leading to a growing allele frequency in the population, especially in Red Holstein (RH) [1]. Generally, across breeds, there is growing evidence that increased selection for polledness does not lead to significant detrimental effects in performance and functional traits [2–4]. Polledness (i.e., the absence of horns) is a phenotype determined by 4 dominant variants located at the proximal end of BTA1 in the region from 1.5-2.5Mb [5–7]. The Mendelian character of the trait and its clarified structure therefore offer potential to investigate the structural effects of selection for polledness with focus on BTA1.

Diverging selection causes directional changes in the allele frequency spectrum. Hence, causal loci or loci that are in linkage with causal variants influenced by selection are expected to have significant allele frequency differences compared to loci not influenced by selection. These differences can be mapped as genomic selection signatures when using, e.g., single nucleotide polymorphisms (SNP) as markers [8, 9]. For the analysis of selection signatures, various methods have been developed based on human and general population genetics. The traditional starting point to evaluate selection signatures is the study of the difference in allele frequencies between different populations. Long lasting divergent selection causes sustained differences in the allele frequency spectrum in separate populations [8]. The Fst value introduced by Weir and Cockerham [10] is one of the standard measures for the evaluation of selection signatures based on allele frequencies between populations. However, short-term past selection events often show no clear signals solely based on allele frequencies, since sustained allele frequency changes in the context of selection only occur over a longer period. Therefore, for the analysis and detection of recent selection events, additional methods have been developed that take the length and structure of the haplotypes into account. A widely used and well-established method is the analysis or comparison of "extended haplotype homozygosity" between populations (xp-EHH) [8, 11]. This methodology has been further developed to better represent selection effects and selection signals that lead to a fixation or approximate fixation of alleles [12]. The corresponding parameter in this regard, Rsb, is calculated based on the EHH patterns [12]. Although most selection signature studies use the mentioned methods to compare evidently separated populations, there are also examples that studied selective sweeps in divergently selected sub-populations comparable to the polled and horned sub-populations in cattle (e.g. [13]).

Traditionally, inferiority of phenotypic performances and genetic values for production and fertility traits was attributed to the polled Holstein cows and bulls. In contrast to these mostly anecdotical reports, the available scientific studies focusing on correlation and association analyses between

polledness and other traits did not support any antagonistic relationships [2, 4, 14, 15]. Nonetheless, especially homozygous polled animals tend to have lower breeding values. Further association studies utilizing dense genomic marker data are not available but may contribute to a deeper understanding in this regard. Hence, a more detailed insight into effects of selection for polledness on chromosome 1 combined with a similarly detailed overview of associations to further traits might help to understand group differences from the past and indicate potential effects in a long-term perspective.

The methodology for genome-wide association studies (GWAS) has evolved significantly over the past 10 years. Statistical methods based on linear mixed models are now a widely used alternative to classical cohorts or case-control design studies using relatively simple statistical tests. Implementations such as in the program package GCTA [16] enable association studies with several thousand SNP markers while controlling the population stratification by considering genomic relationships among all studied individuals.

Consequently, the objectives of this study were: (1) to assess long-term and recent selection signatures in comparison between horned and polled Holstein subpopulations, and (2) to analyze chromosome-wide associations of polled and horned subpopulations with performance and functional traits in a GWAS framework utilizing chromosome-wide SNP marker data from Holstein bulls on chromosome 1.

Methods

Phenotypic and genomic dataset and quality control

The overall dataset consisted of 314 polled (142 RH, 172 Black and White Holstein (BWH)) and 1846 horned (248 RH and 1598 BWH) active bulls from the birth cohorts 2010-2016 selected for artificial insemination. For all bulls in the dataset, genomic marker data for 45,613 SNPs across the genome, were available.

Sub-populations for the analyses were defined based on the known genotypes at the polled locus, which were available for all animals. Animals with the genotype pp were assigned to the horned population, while animals with the genotypes Pp and PP were assigned to the polled population. The selection signature analyses were additionally performed separately for RH and BWH, as well as in a combined dataset.

Preparation and quality control of genomic data was carried out using the software packages R (Version 3.3.2, [17]) and PLINK (Version 1.9, [18]). After quality control (i.e., filtering of genotypes for MAF < 0.01, individual call rate and SNP call rate < 0.01), 44,356 SNP markers remained in the first step. Subsequently, the dataset was split into the 2,756 SNPs located on BTA1 and 41,600 SNPs located on BTA2 to BTA29. After quality control, a principal component analysis was performed for BTA1 as well as for the whole genome using the --pca command in PLINK to

generally assess the genetic diversity in the dataset. Phasing and haplotype reconstruction, which are important prerequisites for the analysis of selection signatures, were subsequently carried out with the program package BEAGLE (Version 3.3; [19]).

Phenotypic information for association analyses were available through breeding values from the routine national genetic evaluation for the traits milk yield (MY), protein yield (PY), fat yield (FY), non-return rate (NRR, binary trait) and somatic cell score (SCS). Due to the rather small dataset and the closely related breeds, the association analyses were performed only in the combined dataset with RH and BWH. To generally validate the approach, we also performed association analyses for the traits polledness (binary phenotypes) and milk yield on BTA14 to confirm the well described DGAT1 QTL.

Selection signature analyses

All selection signature analyses were carried out separately for BTA1 and BTA2 to BTA29. Fst values were calculated using PLINK (Version 1.9, [18]). The –fst function implemented in PLINK estimates the Fst values for each SNP in comparison of the two populations (i.e., horned vs. polled subpopulations) using the method of Weir and Cockerham [10]. According to Hartl and Clark [20], Fst values > 0.15 indicate a moderate to high genetic differentiation between two populations. SNPs that show an Fst value above 0.15 are therefore treated as selection signature candidates. The calculation of xp-EHH was carried out for each SNP in comparison of the two populations (i.e., horned vs. polled subpopulations) using the R package "rehh" [21]. In this regard, for a given candidate SNP, the position-specific extended haplotype homozygosity (EHHS) is defined as the probability that two randomly chosen chromosomes (evaluated on the basis of all SNPs) are identical in origin in a given chromosomal region ("identity-by-descent"). For a candidate SNP *s* over a chromosomal interval extending to the SNP *t*, EHHS is calculated as follows:

The xp-EHH statistic is calculated for a given SNP *s* as follows:

$$xp - EHH(s) = \frac{LRiES(s) \quad - med_{LRiES}}{\sigma_{LRiES}}$$

where *LRiES* is calculated as follows:

$$LRiES(s) \quad = \log\left(\frac{iES_{pop1}(s)}{iES_{pop2}(s)}\right).$$

iES(s) is defines as the integrated Extended Haplotype Homozygosity at SNP *s* with *LRiES(s)* representing the respective Log Ratio values of *iES* [21]. The parameters med_{LRiES} and σ_{LRiES} represent the median and standard deviation of $LRiES(s)$, respectively, calculated over all analyzed SNPs.

The *xp-EHH* statistic is constructed to follow an approximate Normal distribution and can therefore be transformed to the p-value scale as follows:

$$p_{xp-EHH} = 1 - 2 \mid \Phi(xp - EHH) - 0.5 \mid)$$

The chromosome-wide significance threshold for *xp-EHH* on BTA1 was calculated using a Bonferroni correction: 0.05 / 2817 Marker = 0.017e-04. SNPs that show p-values below this threshold are treated as significant selection signatures for BTA1.

The genome-wide significance threshold for *xp-EHH* for analyses on BTA2 to BTA29 was calculated using a Bonferroni correction: 0.05 / 44,356 Marker = 0.011e-06. SNPs that show p-values below this threshold are treated as significant selection signatures for BTA2 to BTA29.

For the graphical representation of the results, the R packages "rehh" [21] and ggplot2 [22] were used. Following Fst and XP-EHH calculations, genes localized in the regions of identified selection signatures were annotated based on the SNP position ±25kb. The Ensembl database was used for gene annotations (release 90, [23]).

Association analyses

The association analyses based on the breeding values used as phenotypes were carried out with the program package GCTA [16], and applying the following mixed model:

$$y = 1\mu + xb + u + e$$

with y is the vector of breeding values; μ is the general mean; b is the fixed additive effect of the candidate SNP that is tested for association; x is the vector of genotypes at the candidate SNP; $u \sim N(0, G\sigma_u^2)$ is the vector with random polygenic effects, with G being the genomic relationship matrix calculated using the genome-wide SNP genotypes, and σ_u^2 the polygenic additive-genetic variance estimated in a first step based on a so-called zero model (i.e. $y = 1\mu + u + e$) and then fixed in the second step, i.e., the test on association for the candidate-SNP; and $e \sim N(0, I\sigma_e^2)$ is the vector of random residual effects with I an identity matrix and the residual variance σ_e^2. Additional fixed effects were the breed (RH or BWH) and the year of birth.

As significance thresholds, both a genome-wide Bonferroni corrected P-value of 0.05 / 44,356 = 0.011e-06 and the chromosome wide (based on chromosome 1) Bonferroni corrected P-value of 0.05 / 2905 = 0.017e-04, were used for the correction of the multiple testing problem. Since the Bonferroni correction is a very conservative procedure, associations that fell below a threshold of p = 0.0001, were treated as candidates to take into account even weak associations with potentially small effects.

Results and Discussion

Principal component analysis (PCA)

Figure 1 displays the results from the PCA analysis based on SNP data only from BTA1 and from all chromosomes. The different breeds (RH, BWH) and the polled and horned subpopulations are displayed in different colors. Overall, the separation based on the first and second principal component between RH and BWH as well as between horned and polled individuals is small. Comparing the results based only on data from BTA1 and all chromosomes, BTA1 shows stronger genetic homogeneity compared to the whole genome.

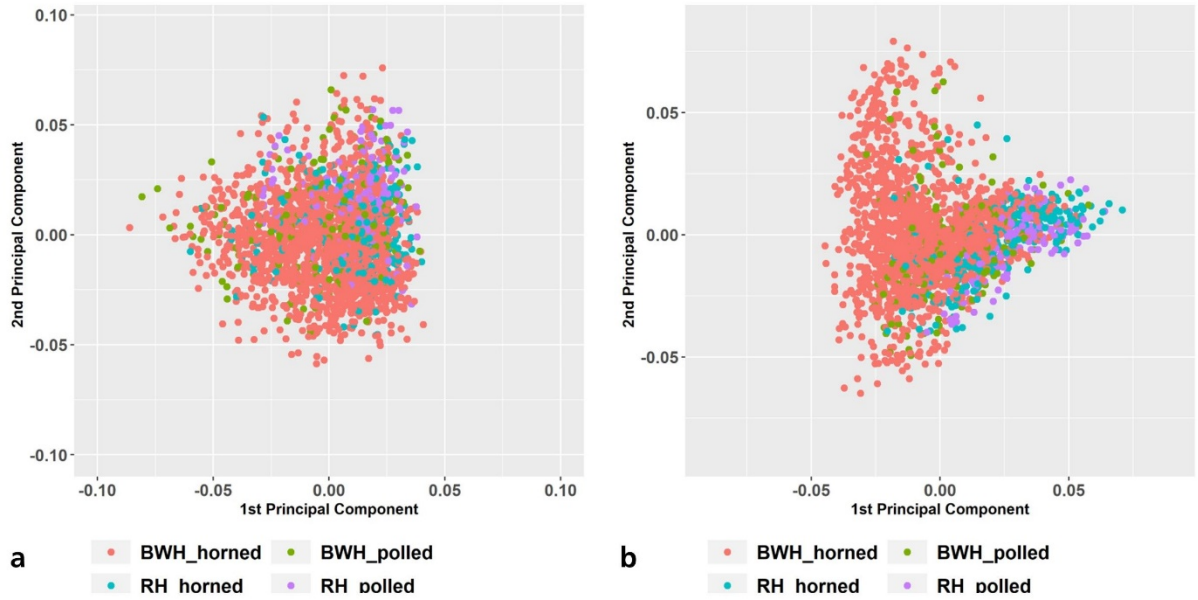

Figure 1 PCA-Plots for RH and BWH polled/horned subpopulations based on genomic data from only BTA1 (a) and the full genome (b).

Selection signatures on BTA1

Figures 2a and 2c show the results of the estimated F_{St} values for the comparison of the horned and polled samples for RH and BWH, respectively. According to Hartl and Clark [20], Fst values > 0.15 indicate a moderate to high genetic differentiation between two populations. The estimated values clearly indicate a high degree of differentiation at the proximal end of BTA1 for a large number of markers (in total 89 SNPs > 0.15 Fst, BWH and RH combined). The region of high differentiation based on Fst can be narrowed down to the range of 0 - ~30Mb, which contains the causal mutation for the trait polledness (~2Mb). Further selection signatures were not identified on BTA1 on the basis of Fst. Only the range of 120-140Mb indicates an increased degree of differentiation in individual regions for RH. There is no comparable signal in the BWH subpopulation.

The results for XP-EHH show a more differentiated picture compared to Fst (see Figure 2b and 2d). A total of 29 SNPs with significant selection signals were found in the RH and BWH breeds.

The region of 0-5Mb shows the highest density of significant signals in both subpopulations comprising a total of 13 SNPs. Compared to the Fst values, the segment that contains the causal mutation and indicates clear signals, is less extensive. In addition to some individual signals, especially at 120MB in the RH breed, there is a clear signature that extends over 4 significant and spatially closely associated SNPs. The results presented suggest that the genomic range directly influenced by selection for polledness is spatially limited to 0-30Mb. The haplotype-based method in particular was able to detect further signatures on BTA1 compared to Fst reflecting ongoing recent selection.

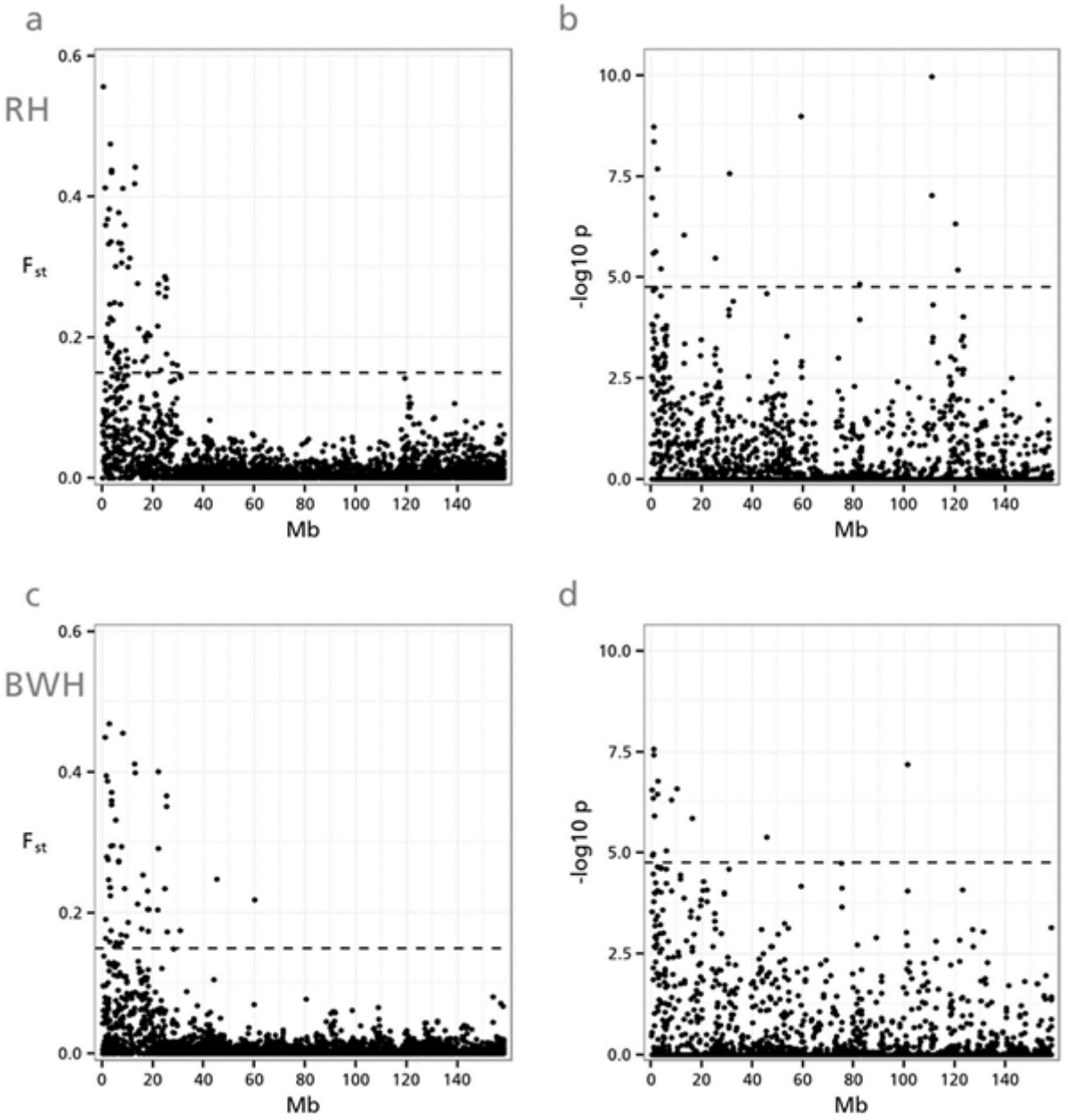

Figure 2 Manhattan plots for Fst (a, c) and -log10 p(XP-EHH) values (b,d) calculated for 2817 SNPs on BTA1 comparing polled and horned bulls from RH and BWH, respectively.

A total of 29 genes could be annotated for the SNPs identified on the basis of Fst and XP-EHH

(sign. SNP ± 25kb). Table 1 lists all SNPs with annotated genes including the results for the selection signature statistics if significant (see also Additional File 1 for a full list of all SNPs identified as selection signals).

Table 1 Significantly associated SNPs on BTA1 with genes annotated based on the SNP position ±25kb.

		FST			XP-EHH			
SNP	**BP**	**significant in subset**	**FST value RH**	**FST value BWH**	**significant in subset**	**p XP-EHH RH**	**p XP-EHH BWH**	**Annotated Genes (Ensembl ID and Name)**
rs29015852	516404				RH, BWH	1.09402E-07	3.16114E-12	ENSBTAG00000026260 (KCNE2)
rs109709783	654413				BWH		1.15658E-05	ENSBTAG00000012594 (MRPS6)
rs110064393	1009504				RH	2.61586E-06		ENSBTAG00000021997 (ITSN1)
rs41635940	1209308				RH, BWH	1.92496E-09	3.74546E-08	ENSBTAG00000008482 (SON)
Hapmap53766-ss46526150	1359951	BWH		0.16421				ENSBTAG00000043399 (SNORA20) ENSBTAG00000009187 (DNAJC28)
ARS-BFGL-NGS-39992	1668494	BWH		0.394895				ENSBTAG00000047288 (HIST1H4G)
rs108992364	1673525	RH	0.199802					ENSBTAG00000047288 (HIST1H4G)
rs41622765	2128924	RH,BWH	0.368115	0.387873				ENSBTAG00000003063 (SYNJ1) ENSBTAG00000003064 (PAXBP1)
rs110296879	2291153	RH,BWH	0.218997	0.275803				ENSBTAG00000017310 (EVA1C)
rs110490165	2313042	RH	0.178425					ENSBTAG00000017310 (EVA1C)
rs43207817	2771830	RH,BWH	0.382572	0.469363	BWH		1,68225E-07	ENSBTAG00000020762 (HUNK)
BTA-32603-no-rs	3459530	BWH		0.617417				ENSBTAG00000017839 (TIAM1)
rs110593395	3541738	RH	0.336484					ENSBTAG00000017839 (TIAM1)
rs41633082	4052161	RH	0.155211					ENSBTAG00000015812 (KRTAP8-1)
rs109430859	5034441	RH,BWH	0.249418	0.151818				ENSBTAG00000039820 (CLDN8)

rs41622772	5541297	RH,BWH	0.300916	0.331957				ENSBTAG00000034854 (GRIK1)
rs41580510	5610335	RH,BWH	0.173831	0.158134				ENSBTAG00000034854 (GRIK1)
rs43215599	6383413	RH	0.169264					ENSBTAG00000007444 (MAP3K7CL)
rs41583696	6526757	RH,BWH	0.334811	0.274259				ENSBTAG00000020121 (RWDD2B) ENSBTAG00000000201 (LTN1)
rs29024165	8837296	RH,BWH	0.359414	0.234903				ENSBTAG00000000648 (ADAMTS5)
ARS-BFGL-NGS-113570	9601018	BWH		0.167083				ENSBTAG00000017753 (APP)
rs109729245	9942902	RH	0.159016					ENSBTAG00000017753 (APP)
Hapmap25334-BTA-160518	10274301	BWH		0.186318				ENSBTAG00000042261 (U6)
rs41671573	14587033	RH	0.212697					ENSBTAG00000045128 (7SK)
BTB-01511695	18099231	BWH		0.230807				ENSBTAG00000000597 (TMPRSS15)
BTA-06080-no-rs	18254887	BWH		0.173685				ENSBTAG00000000588 (CHODL)
BTB-00010729	22074275	BWH	0.203901					ENSBTAG00000002623 (SAMSN1)
rs41646592	22110938	RH,BWH	0.275759	0.291963				ENSBTAG00000002623 (SAMSN1)
rs43219642	22148647	RH,BWH	0.262752	0.401057				ENSBTAG00000002623 (SAMSN1)
BTB-00009232	25607948	BWH		0.173026				ENSBTAG00000009851 (ROBO1)
rs110136403	31013633				RH	2.69026E-08		ENSBTAG00000043669 (U6)
rs43240216	59272723				RH, BWH	1.04386E-09	1,02141E-14	ENSBTAG00000002132 (DRD3)
BTA-31643-no-rs	60254737	BWH		0.218235				ENSBTAG00000011928 (ZBTB20)
rs43111100	120326805				RH	4.82647E-07		ENSBTAG00000000456 (pCPB)

Selection signatures on BTA2 to BTA29

Figures 3a and 3c show the results of the estimated F_{St} values for the comparison of the horned and polled samples for RH and BWH, respectively, for the other chromosomes apart from BTA1. In RH, 4 signals only slightly above the threshold of 0.15, were detected on BTAs 3, 4, 6 and 25. In BWH, a total of 64 SNPs were identified as selection signatures on BTAs 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26 and 28. The largest number of individual signals were found on BTA6 (11), BTA16 (10) and BTA4 (8).

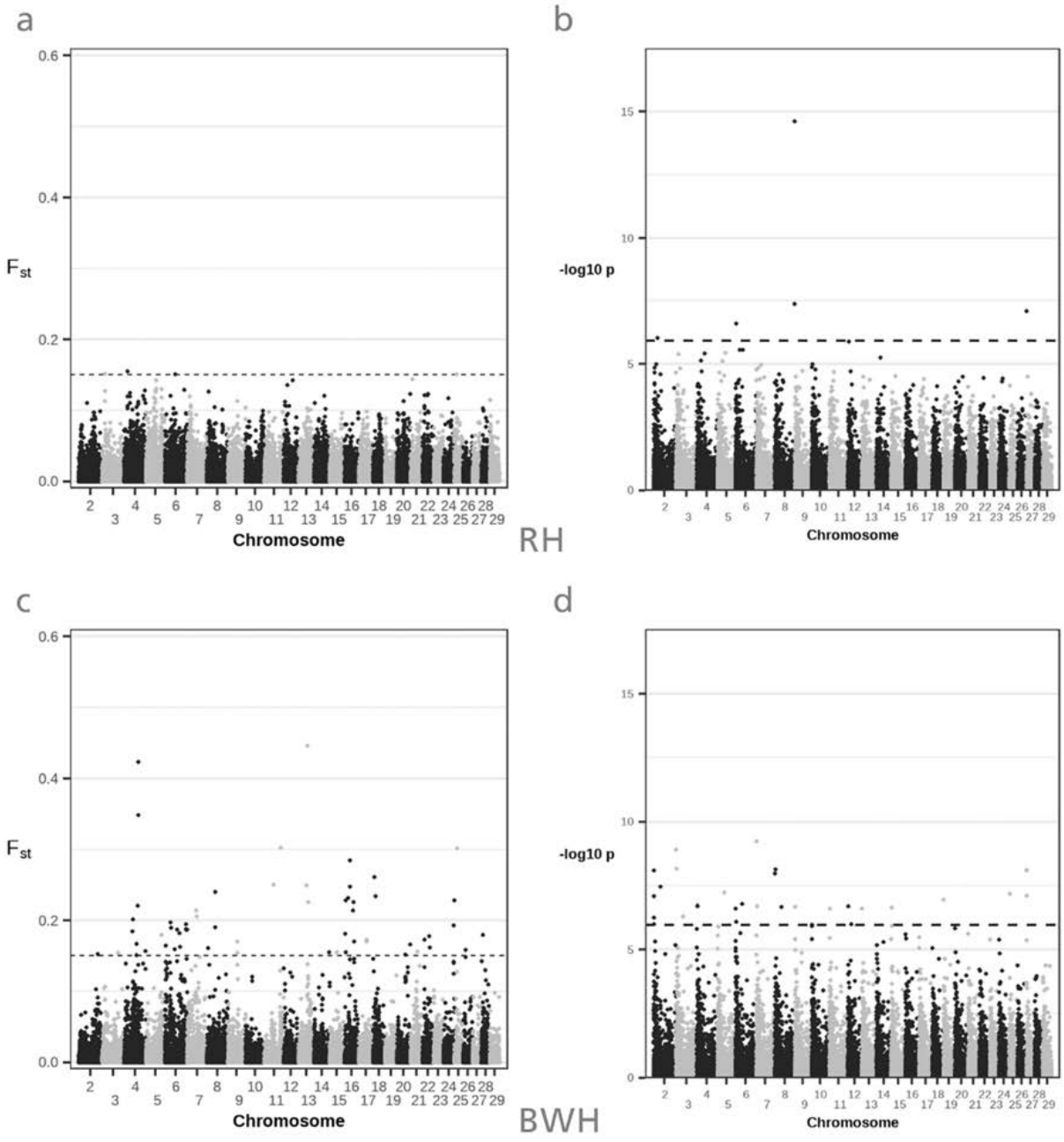

Figure 3 Manhattan plots for Fst (a, c) and -log10 p(XP-EHH) values (b,d) calculated for 44,356 SNPs on BTA2 to BTA29 comparing polled and horned bulls from RH and BWH

Again, as on BTA1, the results for XP-EHH show a more differentiated picture compared to Fst. A total of 36 SNPs with significant selection signals were found in the RH and SBT breeds. In contrast to BTA1 comparing the methods, the XP-EHH analysis revealed a smaller amount of selection signals compared to Fst for the remaining genome. In RH, genome-wide significant signals were detected on BTAs 2, 6, 8 and 26 with the strongest signal on BTA8 (see Figure 3b). In BWH, genome-wide significant signals were detected on BTAs 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 19, 25 and 27. The largest number of individual signals was detected on BTA2 with 5 genome-wide significant SNPs.

Based on the identified selection signals on BTA2 to BTA29, 28 additional genes were annotated for the respective SNPs. Annotated genes were identified on BTAs 3, 4, 6, 7, 8, 11, 12, 15, 16, 18, 22, 25 and 28. Additional File 2 lists all SNPs identified as selection signals including annotated genes.

Association analyses

Validation based on polledness and DGAT1

Figure 4 displays the results of the association analyses aimed at a validation of the approach. For polledness, the expected result was a single isolated association signal at the proximal end on BTA1 and, in addition, no further associations due to the monogenic structure of the trait. The results fit this expectation with high accordance (see Figure 4a). In addition, we also used the locus for the DGAT1 QTL for fat yield (see, for example, [24]) at the proximal end of BTA14 for validation. Similar to the trait polledness, it is expected that the association analysis of the breeding values for the performance characteristics reflects a clear association signal at the proximal end of BTA14. Figure 4c shows that this expectation can be validly fulfilled with the chosen method.

Performance traits: milk yield, fat yield, protein yield

Figure 5 shows the association results on BTA1 for the breeding values for performance traits. No significant associations were detected for one of the performance traits. Especially in the proximal region of BTA1, where the clearest selection signatures could be detected, there are no detectable associations. Individual regions in the medial and distal region of BTA1 show slight trends to higher association signals, but these do not reach the chromosome-wide or the candidate significance threshold. In this context, the adjacent QQ plots (see Figure 5b, d, f) indicate that there was no stratification within the analysis, supporting validity of the result.

Functional traits: somatic cell count and non-return rate

Figure 6 shows the association results on BTA1 for the breeding values for functional traits. Comparable to the results for the performance traits, no significant associations were detected. Again, especially in the proximal region of BTA1, there are no detectable associations. For somatic

cell count, one small peak in the medial region indicates a stronger association signal but does not reach the chromosome-wide or candidate significance threshold. Again, the QQ plots (see Figure 6b, d, f) indicate no sign of stratification.

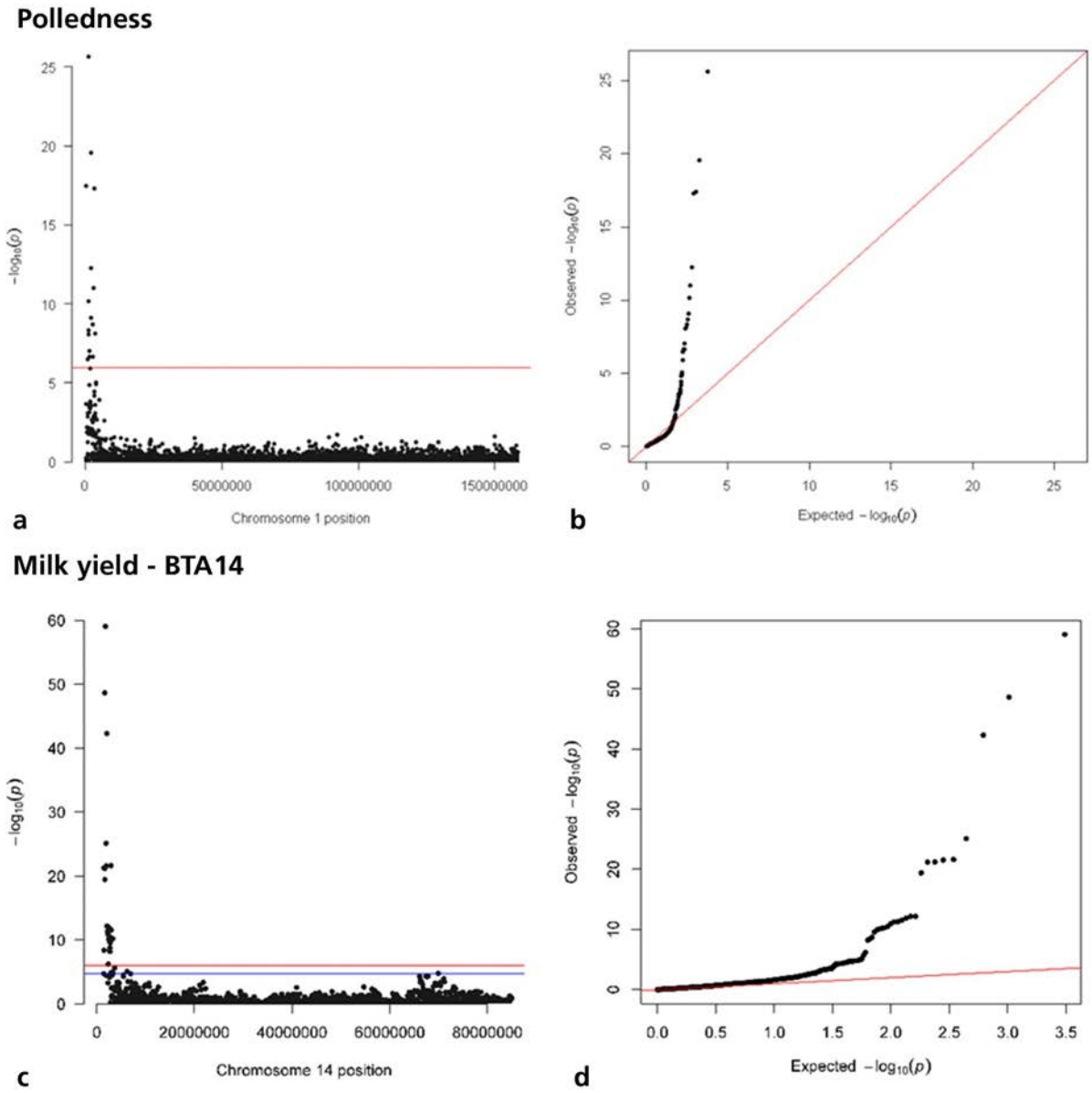

Figure 4 Association analyses for the traits polledness (a Manhattan plot ,b QQ-PLot, based on BTA1) and milk yield (c Manhattan plot, d QQ-Plot, based on BTA14).

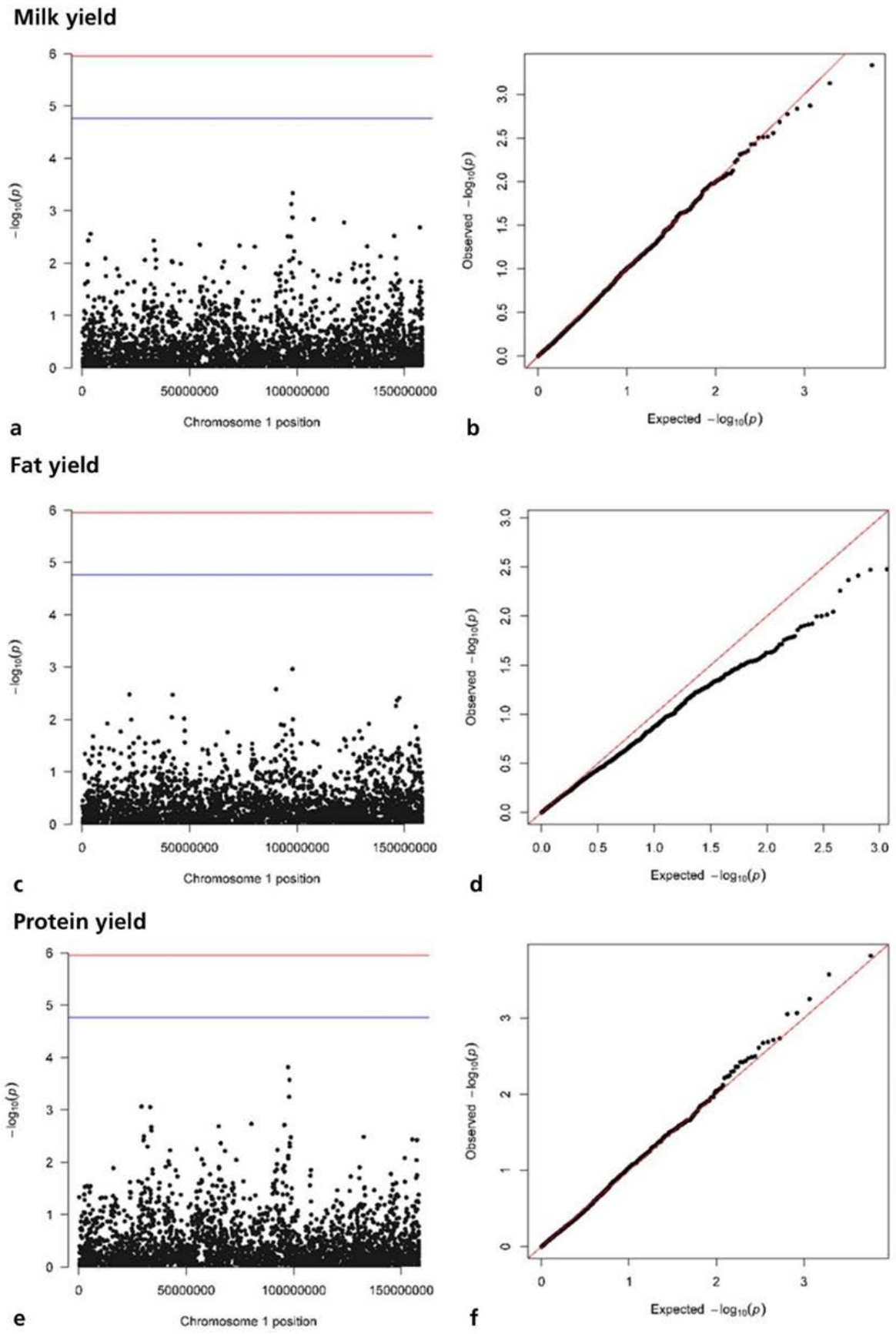

Figure 5 Association analysis for the traits milk yield (a Manhattan plot, b QQ-Plot), fat yield (c Manhattan Plot, d QQ-Plot) and protein yield (e Manhattan Plot, f QQ-Plot) on BTA1. Red line: genome-wide significance threshold (0.011e-06); blue line: chromosome wide significance threshold (0.017e-04)

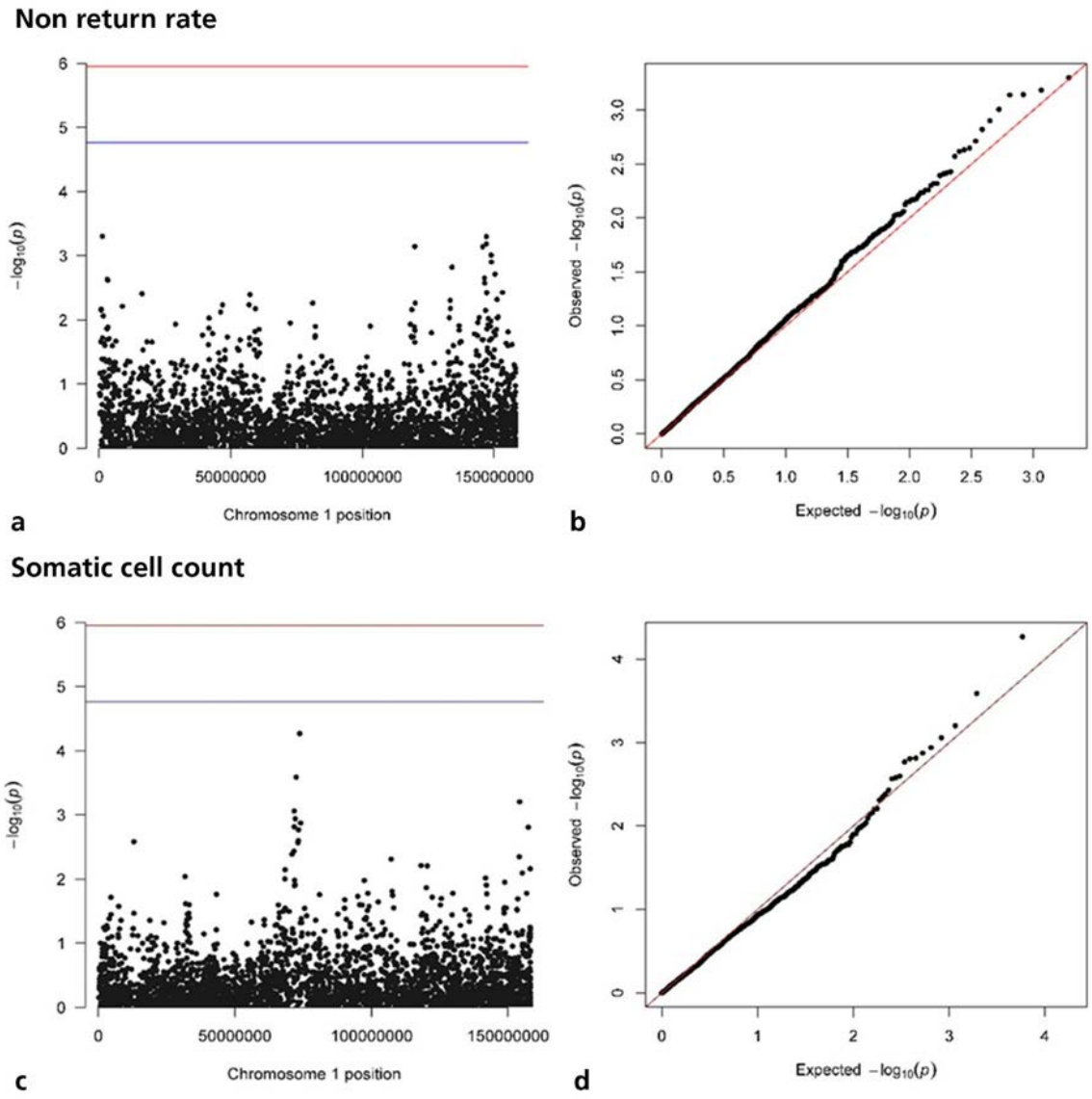

Figure 6 Association analysis for the traits non return rate (a Manhattan Plot, b QQ-Plot) and somatic cell score (c Manhattan Plot, d QQ-Plot) on BTA1. Red line: genome-wide significance threshold (0.011e-06); blue line: chromosome wide significance threshold (0.017e-04)

Overlap between selection signals and association analyses

Interpreting the results for selection signatures and associations in a shared context, a few aspects can be highlighted. No significant associations were found for the tested phenotypes. The results for the selection signatures on BTA1 clearly show that the effect of selection on polledness is not limited to a very small segment in proximity to the polled locus at the proximal end of the chromosome, but indicate a measurable influence on the first 25Mb of the chromosome. Since no clear quantitative trait associations have been found in this segment, it can be assumed that the direct effects of the polled locus on secondary features, e.g. via linkage, are negligible or if present, are rather small. Visual trends to associations could only be detected for the traits milk and protein yield and non-return rate. Accordingly, it could also be argued that all analyzed traits are generally only influenced to a small extent by variants on BTA1.

The results presented above should be seen in close connection with the results at the quantitative genetic level [3, 4, 15]. Our results indicate that there is only a small direct genetic relationship between polledness and, in particular, milk performance traits. At the same time, however, it should also be noted that the region directly affected by selection of polledness spans roughly 15% of BTA1 and clearly contains potentially affected functional genes as presented in the annotation results. Hence, further research and monitoring is recommended to rule out or manage potential, although most likely small, side effects.

In addition, the selection signature analysis focusing on the other genomic regions apart from BTA1 indicates, especially in BWH, further regions of genomic differentiation between the polled and horned subgroups. Interestingly, consistent for both methods used, the amount of differentiation between polled and horned subgroups is much smaller in RH. The RH subset is more balanced in the ratio of polled and horned animals compared to BWH. In addition, it is well known [25] that a large number of polled BWH sires originally descend from RH (e.g. Lawn Boy) due to selection on polledness. Hence, a higher differentiation in BWH might partly reflect general differences between the RH and BWH subgroups apart from polledness or diverging selection in polled and horned cattle. This could also explain that the number of signals based on Fst is higher in the BWH subset, which typically reflects less recent selection compared to XP-EHH.

Conclusions

The segment directly influenced by selection on BTA1 spans about 25Mb. In this segment, no significant associations with other phenotypic features could be detected. The results of the genomic analyses are in line with reported quantitative-genetic analyses and confirm that there are no indications that significant direct effects of selection at the polled locus affect secondary traits. However, we also found evidence, that there are additional genomic regions on other chromosomes showing selection signatures that reflect genomic differentiation between polled and horned animals due to diverging selection in the past.

Competing interests

The authors declare that they have no competing interest.

Acknowledgments

The authors gratefully thank the Landwirtschaftliche Rentenbank for funding this project.

Funding

The present study was supported by funds of the German Government´s Special Purpose Fund held at Landwirtschaftliche Rentenbank

Authors' contributions

CS designed the study, performed all analyses and drafted the manuscript. SK initiated the primary research and assisted in writing the manuscript. All authors have read and approved the manuscript.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Segelke D, Täubert H, Reinhardt F, Thaller G. Considering genetic characteristics in German Holstein breeding programs. J Dairy Sci. 2016;99:458–67. doi:10.3168/jds.2015-9764.
2. Cole JB, Null DJ. Short communication: Phenotypic and genetic effects of the polled haplotype on yield, longevity, and fertility in US Brown Swiss, Holstein, and Jersey cattle. J Dairy Sci. 2019;102:8247–50. doi:10.3168/jds.2019-16530.
3. Scheper C, Emmerling R, Götz K-U, König S. A variance component estimation approach to infer associations between Mendelian polledness and quantitative production and female fertility traits in German Simmental cattle. Genet Sel Evol. 2021;53:60. doi:10.1186/s12711-021-00652-z.
4. Gehrke L, Seichter D, Ruß I, Medugorac I, Tetens J, Thaller G. Impact of the Friesian POLLED mutation on milk production traits in Holstein Friesian. In: Proceedings of the 67th Annual Meeting of the European Federation of Animal Science: 29 August - 2 September 2016; Belfast; 2016. p.434 2016.
5. Wiedemar N, Tetens J, Jagannathan V, Menoud A, Neuenschwander S, Bruggmann R, et al. Independent polled mutations leading to complex gene expression differences in cattle. PLoS One. 2014;9:e93435. doi:10.1371/journal.pone.0093435.
6. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Göpel KH, et al. Bovine polledness—an autosomal dominant trait with allelic heterogeneity. PLoS One. 2012;7:e39477. doi:10.1371/journal.pone.0039477.
7. Allais-Bonnet A, Grohs C, Medugorac I, Krebs S, Djari A, Graf A, et al. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. PLoS One. 2013;8:e63512. doi:10.1371/journal.pone.0063512.
8. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. Science. 2006;312:1614–20. doi:10.1126/science.1124309.
9. Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. Livestock Science. 2014;166:133–43. doi:10.1016/j.livsci.2014.05.003.
10. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38:1358. doi:10.2307/2408641.
11. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002;419:832–7. doi:10.1038/nature01140.
12. Tang K, Thornton KR, Stoneking M. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 2007;5:e171. doi:10.1371/journal.pbio.0050171.
13. Avila F, Mickelson JR, Schaefer RJ, McCue ME. Genome-Wide Signatures of Selection Reveal

Genes Associated With Performance in American Quarter Horse Subpopulations. Frontiers in Genetics. 2018;9:249. doi:10.3389/fgene.2018.00249.

14. Windig JJ, Hoving-Bolink RA, Veerkamp RF. Breeding for polledness in Holstein cattle. Livestock Science. 2015;179:96–101. doi:10.1016/j.livsci.2015.05.021.
15. Dressel, H., Wensch-Dorendorf, M., Swalve HH. Genetic relationship between polledness and other relevant traits in Holstein dairy cattle. In: Proceedings of the 67th Annual Meeting of the European Federation of Animal Science: 29 August - 2 September 2016; Belfast; 2016. p.491 2016.
16. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82. doi:10.1016/j.ajhg.2010.11.011.
17. Team RC. R: A language and environment for statistical computing: Team RC. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019 2019.
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. doi:10.1186/s13742-015-0047-8.
19. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet. 2009;84:210–23. doi:10.1016/j.ajhg.2009.01.005.
20. Hartl DL, Clark AG. Principles of population genetics. 4th ed. Sunderland, Mass: Sinauer Associates; 2007.
21. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017;17:78–90. doi:10.1111/1755-0998.12634.
22. Wickham H. ggplot2: Elegant graphics for data analysis. Cham: Springer international publishing; 2016.
23. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018;46:D754-D761. doi:10.1093/nar/gkx1098.
24. Thaller G, Kühn C, Winter A, Ewald G, Bellmann O, Wegner J, et al. DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. Animal Genetics. 2003;34:354–7. doi:10.1046/j.1365-2052.2003.01011.x.
25. Specht L. Polled Holstein History. 2008. https://nebula.wsimg.com/339065745033f44d872cb82122ef7ce2?AccessKeyId=27933CBAC73165BF9625&disposition=0&alloworigin=1. Accessed 5 Oct 2022.

Additional Files

Additional File 1. Full list of significantly associated SNPs on BTA1 based on FST or XP-EHH including information on genes annotated based on the SNP position ±25kb.

		FST			XP-EHH			
SNP	**BP**	**significant in subset**	**FST value RH**	**FST value BWH**	**significant in subset**	**p XP-EHH RH**	**p XP-EHH BWH**	**Annotated Genes (Ensembl ID and Name)**
rs29026917	434180	RH, BWH	0,556841	0,639956				ENSBTAG00000026259
rs29015852	516404				RH, BWH	1,09402E-07	3,16114E-12	ENSBTAG00000026260 (KCNE2)
rs110082431	533815				BWH		2,79824E-07	
rs109709783	654413				BWH		1,15658E-05	ENSBTAG00000012594 (MRPS6)
rs108981857	845494				BWH		4,514E-07	
rs110467572	883895				BWH		1,07314E-05	
rs110064393	1009504				RH,	2,61586E-06		ENSBTAG00000021997 (ITSN1)
rs110950216	1189382				RH, BWH	4,36845E-09	2,71283E-08	
rs41635940	1209308				RH, BWH	1,92496E-09	3,74546E-08	ENSBTAG00000008482 (SON)
rs109671440	1288510	RH,BWH	0,412992	0,449474	BWH		1,24567E-06	
Hapmap53766-ss46526150	1359951	BWH		0,16421				ENSBTAG00000043399 (SNORA20) ENSBTAG00000009187 (DNAJC28)
rs109818851	1582828	RH, BWH	0,3602	0,191208				ENSBTAG00000019404
ARS-BFGL-NGS-39992	1668494	BWH		0,394895				ENSBTAG00000047288 (HIST1H4G)

rs108992364	1673525	RH,	0,199802					ENSBTAG00000047288 (HIST1H4G)
rs109130352	1896112				RH,	2,89721E-07		
rs109797076	1921756				RH,	2,329E-06		
rs110875985	2049400	RH,BWH	0,194171	0,279614				ENSBTAG00000043993 (C1H21orf62)
rs41622765	2128924	RH,BWH	0,368115	0,387873				ENSBTAG00000003063 (SYNJ1) ENSBTAG00000003064 (PAXBP1)
rs110296879	2291153	RH,BWH	0,218997	0,275803				ENSBTAG00000017310 (EVA1C)
rs110490165	2313042	RH,	0,178425					ENSBTAG00000017310 (EVA1C)
rs108978478	2595577				RH, BWH	2,08487E-08	3,47162E-07	
rs110930058	2714141	RH,BWH	0,333091	0,247043				
rs43207817	2771830	RH,BWH	0,382572	0,469363	BWH		1,68225E-07	ENSBTAG00000020762 (HUNK)
rs43211381	3116101	RH,BWH	0,247514	0,236248				ENSBTAG00000018854
rs109362109	3148834	RH,	0,190958	0,159106				
rs42381983	3197378	RH,	0,186512					
rs29012842	3249057	RH,	0,227402					
rs43180934	3319070	RH,BWH	0,227402	0,236609				
rs41582551	3404571	RH,	0,226895					
rs110602426	3433611	RH,	0,475191	0,224281				
BTA-32603-no-rs	3459530	BWH		0,617417				ENSBTAG00000017839 (TIAM1)
rs110593395	3541738	RH,	0,336484					ENSBTAG00000017839 (TIAM1)

ARS-BFGL-NGS-43005	3568205	BWH		0,295371				
Hapmap59750-rs29024375	3601437	BWH		0,174941				ENSBTAG00000046072 (MIR2284X)
rs43215834	3628012	RH,	0,186512					
rs110253059	3792863	RH, BWH	0,190347	0,371221				ENSBTAG00000004530
rs43212062	3873380	RH, BWH	0,437887	0,359547	RH,	6,18144E-06		
rs43211655	3913742	RH,BWH	0,434272	0,353935				
rs41633082	4052161	RH,	0,155211					ENSBTAG00000015812 (KRTAP8-1)
rs43204112	4311365	RH,	0,224045					
ARS-BFGL-NGS-65157	4348137	BWH		0,296194				
rs109430859	5034441	RH,BWH	0,249418	0,151818				ENSBTAG00000039820 (CLDN8)
rs42845302	5179453	RH,	0,169718					
rs41622772	5541297	RH,BWH	0,300916	0,331957				ENSBTAG00000034854 (GRIK1)
rs41580510	5610335	RH,BWH	0,173831	0,158134				ENSBTAG00000034854 (GRIK1)
rs42612145	5997487				BWH		9,03143E-06	
rs109691080	6307443	RH,	0,163724					
rs43215599	6383413	RH,	0,169264					ENSBTAG00000007444 (MAP3K7CL)
rs41638872	6492449	RH,BWH	0,377456	0,272641				ENSBTAG00000014233
rs41583696	6526757	RH,BWH	0,334811	0,274259				ENSBTAG00000020121 (RWDD2B) ENSBTAG00000000201 (LTN1)

rs41639125	6882069	RH,BWH	0,179045	0,155365				
rs110666334	7344117	RH,	0,156666					
rs41633309	7405025	RH,	0,246938					
rs42612494	7530524	RH,BWH	0,333714	0,169194				
rs41606310	7669386	RH,BWH	0,306118	0,15784				
rs41616911	7728571	RH,BWH	0,324488	0,29418				
rs41600061	8148041	RH,	0,412119					
rs43707416	8169172	BWH	0,455519		BWH		4,87279E-07	
rs29024165	8837296	RH,BWH	0,359414	0,234903				ENSBTAG00000000648 (ADAMTS5)
rs109122047	9375752	RH,	0,159345					
rs109918255	9404794	RH,	0,181414					
ARS-BFGL-NGS-113570	9601018	BWH		0,167083				ENSBTAG00000017753 (APP)
rs109729245	9942902	RH,	0,159016					ENSBTAG00000017753 (APP)
rs41576751	9982398	RH,	0,169641					
Hapmap25334-BTA-160518	10274301	BWH		0,186318				ENSBTAG00000042261 (U6)
rs110829815	10369662	RH,BWH	0,300265	0,668375	BWH		2,64513E-07	
rs29019235	10936036	RH,	0,312402					ENSBTAG00000038900
rs110416269	12805406	RH,	0,165592					
rs110305292	12996787	RH,BWH	0,418883	0,411745				

rs41665389	13019983	RH,BWH	0,441977	0,399036	RH,	9,06832E-07		
rs41566591	13979316	RH,BWH	0,276985	0,212286				
rs41671573	14587033	RH,	0,212697					ENSBTAG00000045128 (7SK)
BTB-00569345	15646164	BWH		0,177206				
rs42237543	16245203	RH,BWH	0,166114	0,25414	BWH		1,41101E-06	
rs41924236	16923285	RH,	0,200714					
rs41923266	17203157	RH,	0,195365					
rs43217526	17448939	RH,	0,172714					
rs43109937	18002981	RH,BWH	0,206108	0,205078				
BTB-01511695	18099231	BWH		0,230807				ENSBTAG00000000597 (TMPRSS15)
BTA-06080-no-rs	18254887	BWH		0,173685				ENSBTAG00000000588 (CHODL)
rs110982370	18353847	RH,	0,20239					
BTB-01511592	18390386	BWH		0,205078				
rs110543362	18909161	RH,	0,202236					ENSBTAG00000047943
rs109596755	22031001	RH,	0,216204					
BTB-00010729	22074275	BWH	0,203901					ENSBTAG00000002623 (SAMSN1)
rs41646592	22110938	RH,BWH	0,275759	0,291963				ENSBTAG00000002623 (SAMSN1)
rs43219642	22148647	RH,BWH	0,262752	0,401057				ENSBTAG00000002623 (SAMSN1)
rs42668898	23115110	RH,	0,154127					ENSBTAG00000046369

rs43227184	24757983	RH,	0,286673					
ARS-BFGL-NGS-85045	24779790	BWH	0,234382					
rs29017639	24982221	RH,	0,258297					
rs43225545	25321721	RH,	0,282498					
ARS-BFGL-BAC-16299	25471181	BWH	0,366811					
rs109098557	25502991	RH,	0,270114		RH,	3,32817E-06		
rs43223792	25576064	RH,BWH	0,176951	0,35106				
BTB-00009232	25607948	BWH		0,173026				ENSBTAG00000009851 (ROBO1)
rs42413552	27786985	RH,	0,163359					
rs43219334	29561198	RH,	0,16077					
Hapmap50687-BTA-41950	30865885	BWH		0,175119				
rs110136403	31013633				RH,	2,69026E-08		ENSBTAG00000043669 (U6)
BTB-01691456	45157959	BWH		0,248205				
rs41592484	45864540				BWH		4,17639E-06	
rs43240216	59272723				RH, BWH	1,04386E-09	1,02141E-14	ENSBTAG00000002132 (DRD3)
BTA-31643-no-rs	60254737	BWH		0,218235				ENSBTAG00000011928 (ZBTB20)
rs41635208	82546418				RH,	1,47639E-05		ENSBTAG00000007296
rs110373410	101330295				BWH		6,50484E-08	
rs109126926	111030092				RH,	9,5787E-08		

rs41611561	111039014				RH,	1,09456E-10		
rs43111100	120326805				RH,	4,82647E-07		ENSBTAG00000000456 (pCPB)
rs41578204	121363361				RH,	6,66305E-06		

Additional File 2. Full list of significantly associated SNPs on BTA2-29 based on FST or XP-EHH including information on genes annotated based on the SNP position ±25kb.

			FST			XP-EHH			
SNP	**CHR**	**BP**	**significant in subset**	**FST value RH**	**FST value BWH**	**significant in subset**	**p XP-EHH RH**	**p XP-EHH BWH**	**Annotated Genes (Ensembl ID and Name)**
ARS-BFGL-NGS-6050	2	1216101				BWH	8,0004E-08		ENSBTAG00000014714 (TUBGCP5)
ARS-BFGL-NGS-18261	2	1896078				BWH	1,00069E-06		ENSBTAG00000000941 (PLEKHB2)
ARS-BFGL-NGS-108007	2	1926143				BWH	5,39738E-07		
ARS-BFGL-NGS-53839	2	3197666				BWH	7,81885E-09		
Hapmap38677-BTA-46746	2	22322927				RH	8,97721E-07		
BTA-46613-no-rs	2	37358592				BWH	3,38458E-08		ENSBTAG00000004177 (TANC1)
Hapmap30864-BTA-157170	2	115227024	BWH		0,152708				
ARS-BFGL-NGS-654	3	1937626				BWH			
ARS-BFGL-NGS-36590	3	3743710				BWH			ENSBTAG00000012025 (LMX1A)
ARS-BFGL-NGS-13158	3	19337220	RH	0,152028					
BTA-20774-no-rs	3	47138507				BWH			
BTB-00143880	3	92600960	BWH		0,154772				ENSBTAG00000010717 (SSBP3)
ARS-BFGL-NGS-81557	4	4282334				BWH		1,91376E-07	
ARS-BFGL-NGS-112260	4	5061724				BWH		1,95751E-07	ENSBTAG00000017086 (GRB10)
ARS-BFGL-NGS-25363	4	20004627	RH	0,155326					

ARS-BFGL-NGS-102463	4	47358350	BWH		0,184753				
BTA-91631-no-rs	4	51330208	BWH		0,201612				ENSBTAG00000000169 (ASZ1)
ARS-BFGL-NGS-57945	4	68599650	BWH		0,166671				
ARS-BFGL-NGS-67967	4	69539994	BWH		0,15092				
BTA-71368-no-rs	4	73837632	BWH		0,220522				ENSBTAG00000046430 (ZNF804B)
Hapmap39800-BTA-102932	4	78319582	BWH		0,423832				
BTB-01096725	4	78440044	BWH		0,348427				
BTB-00213455	4	119437182	BWH		0,157022				
ARS-USMARC-636	5	45774499				BWH		5,67951E-08	
ARS-BFGL-NGS-39379	5	106269362	BWH		0,179693				ENSBTAG00000016649 (CCND2)
Hapmap46514-BTA-122322	6	1091047				BWH		2,39668E-07	
BTA-16306-no-rs	6	3755618	BWH		0,164307				
BTA-101239-no-rs	6	5177464				RH, BWH	2,39953E-07	8,20268E-07	
ARS-BFGL-NGS-102813	6	28827723	BWH		0,197118				
Hapmap27817-BTC-032401	6	33059996	BWH		0,18984				
Hapmap32211-BTC-035717	6	40355102				BWH		1,62389E-07	
BTB-01446390	6	51478350	BWH		0,159648				
Hapmap51102-BTA-97964	6	55462479	RH	0,151114					

BTB-00260624	6	65675884	BWH		0,187478				
Hapmap56688-rs29025335	6	81767374	BWH		0,182014				
BTA-86049-no-rs	6	83919789	BWH		0,162858				
BTA-66511-no-rs	6	115953817	BWH		0,187867				ENSBTAG00000031497 (FGFBP1)
BTB-00283498	6	116041509	BWH		0,194468				ENSBTAG00000013736 (PROM1)
BTA-05567-rs29019726	6	116864937	BWH		0,187357				ENSBTAG00000014058 (LDB2)
ARS-BFGL-NGS-34909	6	128444286	BWH		0,186922				
ARS-BFGL-NGS-61223	7	1635703				BWH		5,65524E-10	ENSBTAG00000019868
ARS-BFGL-NGS-35459	7	2792296				BWH		2,01918E-07	
ARS-BFGL-NGS-37891	7	52977921	BWH		0,213868				
BTA-108606-no-rs	7	56290162	BWH		0,205442				ENSBTAG00000019472 (NR3C1)
BTA-111263-no-rs	8	1569883				BWH		1,01807E-08	ENSBTAG00000035293
ARS-BFGL-NGS-8051	8	4047354				BWH		7,0058E-09	
ARS-BFGL-NGS-59383	8	8340612	BWH		0,161475				ENSBTAG00000000500 (PINX1)
ARS-BFGL-NGS-15063	8	42100669				BWH		2,10965E-07	ENSBTAG00000018517 (VLDLR)
Hapmap41169-BTA-117725	8	48788585	BWH		0,190149				ENSBTAG00000011545 (GDA)
BTB-00347022	8	49457771	BWH		0,239991				
ARS-BFGL-NGS-19723	8	113301391				RH	2,35246E-15		
ARS-BFGL-NGS-11454	8	113301392				RH	4,11072E-08		

ARS-BFGL-NGS-104763	9	1859781				BWH		2,12809E-07	
Hapmap44223-BTA-26639	9	56548919	BWH		0,154497				
ARS-BFGL-NGS-80765	9	56904690	BWH		0,170468				
ARS-BFGL-NGS-35593	10	2777001				BWH		1,05644E-06	
ARS-USMARC-Parent-DQ786766-rs29012070	10	3530271				BWH		1,17323E-06	
BTB-00449812	11	1310834				BWH		2,43994E-07	ENSBTAG00000004297 (ACOXL)
Hapmap56580-rs29022336	11	45641384	BWH		0,249942				ENSBTAG00000005614 (UXS1)
ARS-BFGL-NGS-97969	11	89497577	BWH		0,302439				
ARS-BFGL-NGS-74419	12	2015196				BWH		1,96097E-07	
ARS-BFGL-NGS-59877	12	23991213				BWH		9,94991E-07	ENSBTAG00000009405 (TRPC4)
Hapmap58162-rs29021214	13	1799876				BWH		2,50446E-07	
Hapmap44369-BTA-32763	13	44912266	BWH		0,249468				
ARS-BFGL-NGS-75040	13	53550674	BWH		0,445849				
ARS-BFGL-BAC-839	13	57320572	BWH		0,225478				
Hapmap43506-BTA-19779	14	82380321	BWH		0,154975				
Hapmap27045-BTA-37616	15	3857845				BWH		1,17225E-06	
ARS-BFGL-NGS-84790	15	4992087				BWH		2,18644E-07	
ARS-BFGL-BAC-27769	15	40560264	BWH		0,155603				ENSBTAG00000032657 (TEAD1)

ARS-BFGL-NGS-83148	16	3177193	BWH		0,181658				ENSBTAG00000004952 (MFSD4A)
BTB-02010595	16	6523545	BWH		0,155501				ENSBTAG00000016729 (KCNT2)
BTB-01199899	16	6557355	BWH		0,155501				ENSBTAG00000016729 (KCNT2)
ARS-BFGL-NGS-107190	16	8261831	BWH		0,22822				
Hapmap42077-BTA-114779	16	23037476	BWH		0,231726				
BTB-00632775	16	34052345	BWH		0,285057				
BTB-00635702	16	34994367	BWH		0,247562				
BTB-00649201	16	52822104	BWH		0,214012				
BTA-39180-no-rs	16	53777067	BWH		0,225676				ENSBTAG00000015606 (KAZN)
ARS-BFGL-NGS-34518	16	57954537	BWH		0,170468				
ARS-BFGL-NGS-69554	17	43966063	BWH		0,170476				
BTA-40973-no-rs	17	45065620	BWH		0,17283				
ARS-BFGL-NGS-13478	18	7181692	BWH		0,261522				
ARS-BFGL-NGS-86599	18	14359452	BWH		0,233993				ENSBTAG00000016006 (ANKRD11)
Hapmap44238-BTA-42338	18	14401871	BWH		0,233993				ENSBTAG00000016006 (ANKRD11)
ARS-BFGL-NGS-24837	18	14503218	BWH		0,233993				
ARS-BFGL-NGS-42215	19	3793023				BWH		1,12817E-07	
Hapmap36599-SCAFFOLD120897_8821	20	46941646	BWH		0,15157				
ARS-BFGL-NGS-118389	20	69528142	BWH		0,166581				

ARS-BFGL-NGS-42955	21	39597049	BWH		0,156116				
ARS-BFGL-NGS-11959	22	8475203	BWH		0,173119				
Hapmap44388-BTA-54275	22	38047255	BWH		0,178247				ENSBTAG00000000265 (SYNPR)
ARS-BFGL-NGS-12851	22	41437950	BWH		0,162123				ENSBTAG00000014418 (FHIT)
Hapmap49230-BTA-58719	24	59307071	BWH		0,193155				
BTA-105360-no-rs	24	62569209	BWH		0,228317				
ARS-BFGL-NGS-103099	25	1188901				BWH		6,49098E-08	ENSBTAG00000033389
UA-IFASA-3143	25	11564340	RH	0,150649					ENSBTAG00000007142 (CPPED1)
BTA-59468-no-rs	25	15398821	BWH		0,301835				
Hapmap41670-BTA-87679	26	16091850	BWH		0,158348				
ARS-BFGL-NGS-17816	26	50800200				RH	7,68217E-08		
ARS-BFGL-NGS-13221	27	500				BWH		7,51639E-09	
ARS-BFGL-NGS-95953	27	550				BWH		8,04367E-09	
ARS-BFGL-NGS-47559	27	700				BWH		7,69354E-08	
Hapmap49252-BTA-64616	28	8421029	BWH		0,179826				
ARS-BFGL-NGS-79829	28	8501840	BWH		0,179826				ENSBTAG00000016804 (LYST)

CHAPTER 5

Evaluation of breeding strategies for polledness in dairy cattle using a newly developed simulation framework for quantitative and Mendelian traits

Carsten Scheper, Monika Wensch-Dorendorf, Tong Yin, Holger Dressel, Herrmann Swalve and Sven König

Genetics Selection Evolution (2016) 48:50
https://doi.org/10.1186/s12711-016-0228-7

Additional files not shown in this thesis, due to their extent are directly available in the online version of this article at:
https://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0228-7#Sec14

Own contribution to the publication:

Study design:	mostly autonomous
Collection of data:	autonomous
Statistical analysis:	autonomous
Manuscript:	mostly autonomous

Scheper *et al. Genet Sel Evol (2016) 48:50*
DOI 10.1186/s12711-016-0228-7

Genetics Selection Evolution

RESEARCH ARTICLE Open Access

Evaluation of breeding strategies for polledness in dairy cattle using a newly developed simulation framework for quantitative and Mendelian traits

Carsten Scheper[1*], Monika Wensch-Dorendorf[2], Tong Yin[1], Holger Dressel[2], Herrmann Swalve[2] and Sven König[1]

Abstract

Background: Intensified selection of polled individuals has recently gained importance in predominantly horned dairy cattle breeds as an alternative to routine dehorning. The status quo of the current polled breeding pool of genetically-closely related artificial insemination sires with lower breeding values for performance traits raises questions regarding the effects of intensified selection based on this founder pool.

Methods: We developed a stochastic simulation framework that combines the stochastic simulation software QMSim and a self-designed R program named QUALsim that acts as an external extension. Two traits were simulated in a dairy cattle population for 25 generations: one quantitative (QMSim) and one qualitative trait with Mendelian inheritance (i.e. polledness, QUALsim). The assignment scheme for qualitative trait genotypes initiated realistic initial breeding situations regarding allele frequencies, true breeding values for the quantitative trait and genetic relatedness. Intensified selection for polled cattle was achieved using an approach that weights estimated breeding values in the animal best linear unbiased prediction model for the quantitative trait depending on genotypes or phenotypes for the polled trait with a user-defined weighting factor.

Results: Selection response for the polled trait was highest in the selection scheme based on genotypes. Selection based on phenotypes led to significantly lower allele frequencies for polled. The male selection path played a significantly greater role for a fast dissemination of polled alleles compared to female selection strategies. Fixation of the polled allele implies selection based on polled genotypes among males. In comparison to a base breeding scenario that does not take polledness into account, intensive selection for polled substantially reduced genetic gain for this quantitative trait after 25 generations. Reducing selection intensity for polled males while maintaining strong selection intensity among females, simultaneously decreased losses in genetic gain and achieved a final allele frequency of 0.93 for polled.

Conclusions: A fast transition to a completely polled population through intensified selection for polled was in contradiction to the preservation of high genetic gain for the quantitative trait. Selection on male polled genotypes with moderate weighting, and selection on female polled phenotypes with high weighting, could be a suitable compromise regarding all important breeding aspects.

*Correspondence: cscheper@uni-kassel.de
[1] Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany
Full list of author information is available at the end of the article

Background

The routinely-used practice of dehorning in dairy and beef cattle worldwide has attracted increased negative public perception, and contributes to conflicts between modern intensive livestock management and animal welfare. Undoubtedly, it is well documented that dehorning of calves is associated with stress, pain and temporary negative impact on calf growth [1–3]. Currently, the dehorning procedures used aim at improving animal welfare [4–6] by alleviating or even eliminating pain reactions which has led to the development of procedures regulated by legal prohibition [7]. Hence, the urgent need for alternatives to cattle dehorning is strengthened. Simply avoiding dehorning by keeping naturally-horned cattle is considered as one possibility [8], but implies substantial adaptation of housing conditions and does not contribute to reduce the risk of injuries among cattle and animal keepers [9, 10]. In this regard, targeted selection and dissemination of genetically-polled animals into naturally-horned cattle breeds (i.e. via introgression of polled alleles), appears to be the most practicable and sustainable solution.

Two loci, "*polled*" and "*scurs*", determine the variety of phenotypes that are associated with the polled trait in cattle [11–18]. While the precise molecular structure and specific inheritance pattern of the *scurs* locus is not conclusively clarified, the *polled* locus in cattle has been mapped to the proximal end of bovine autosome 1 (BTA1, BTA for *Bos taurus*). The *polled* locus is characterized by (1) autosomal dominant inheritance of mutant alleles and (2) structural heterogeneity that depends on the origin of a breed. Two breed-specific haplotypes were identified at the *polled* locus. The Celtic allele present in Angus, Simmental, Limousin, Charolais, etc. is a complex insertion-deletion (indel), whereas breeds of Friesian origin (e.g. Holstein and Jersey) share a 80 kbp duplication as the most likely causative variant [11, 13, 14]. Since these variants are localized in non-coding DNA regions [15, 18], they are assumed to have rather a regulatory than directly a functional effect. Consequently, the identification of the molecular structure at the *polled* locus has led to the development of a validated direct gene test, which allows precise genotyping as required for substantiated selection decisions [19].

As reflected by the present number of entirely polled breeds and breeds with a significant ratio of polled individuals, breeding and selection for polled animals has a longer tradition in beef than in dairy cattle [20–22]. Thus, the most prevalent dairy cattle breeds in Europe (i.e. Holstein and Jersey) are characterized by a small proportion of polled animals, which is due to the initiative of a limited number of motivated "polled breeders" [20, 23]. It is only during the last decade that a slow but steadily increasing demand for polled artificial insemination (AI) sires has resulted in increasing the numbers of available polled Holstein AI bulls worldwide. Due to the limited number of polled founders, groups of polled Holstein individuals display lower average breeding values and a higher average kinship than horned individuals [21, 22, 24, 25]. These findings were recently confirmed by own evaluations based on the database from the German national genetic evaluation for Holstein AI sires [26]. Quite similar results were reported for dual-purpose German Simmental cattle, while no differences between polled and horned groups were found with regard to health, growth and reproductive traits in beef cattle [10, 27–29]. Evidence accumulated for polled German Simmental cattle, as well as more recent advances in the Holstein breed, further indicate that the initial inferior performance of polled individuals might be due to a selection advantage of their horned pendants rather than an inevitable genetic disadvantage [22, 30].

Based on these assumptions and comparisons between groups of horned and polled animals using estimated breeding values (EBV), it is essential to evaluate a wide variety of polled breeding strategies in terms of long-term selection response and future true genetic relationships by applying simulation techniques. Simulation studies have a long tradition in population genetics to evaluate the effects of evolutionary as well as anthropogenic processes, and have gained additional importance with the rapid development of genomic methods and the increased availability of powerful computer systems [31, 32]. Nonetheless, the availability of specialized software packages using deterministic as well as stochastic approaches that are developed to tackle issues directly targeting animal breeding combined with mating systems is rather limited [33–36]. Deterministic simulations allow equation-based prediction of average genetic gain and average inbreeding level without considering specific individuals. Results from deterministic simulations that addressed selection for the polled trait clearly showed a loss in genetic gain, and steady or decreasing average inbreeding depending on the chosen selection strategy [24]. Inbreeding reduction following selection for polledness was recently confirmed by stochastic simulations [37]. However, short-term inbreeding reduction due to the use of polled sires that are very related between each other, but not so strongly related to the horned populations, will be eroded with high probability in a long-term breeding perspective [22, 38]. Traditionally, for a multiple-trait approach, both deterministic and stochastic simulation techniques require genetic (co)variance components for both traits. Regarding the situation with polledness, only assumptions can be made since results are not available yet.

QMSim [36] is a powerful whole-genome stochastic simulation program that was designed to simulate a wide range of genetic and genomic architectures and population structures, particularly in livestock. Nonetheless, QMSim is limited to the simulation of a single quantitative trait, but includes an interface that can be used for, e.g., the external estimation of breeding values. On the basis of QMSim and with the intention to use the mentioned interface, we developed an R program as an external extension to simulate an additional qualitative polled trait. To our knowledge, there is no stochastic simulation software package available that, simultaneously, simulates a quantitative trait combined with a Mendelian trait (such as the *polled* allele) within the framework of complex dairy cattle breeding programs with multi-trait selection.

Based on the aforementioned simulation technique requirements and the practical need for breeding polled populations, the objectives of this study were: (1) to extend the functionality of the stochastic simulation software QMSim by developing a self-designed R-program for the simulation of an additional qualitative trait; (2) to enable simultaneous selection for the quantitative and qualitative trait using a variety of selection strategies for the polled trait; (3) to evaluate the effects of different selection schemes on the allele frequency of *polled*, genetic gain and inbreeding in a long term perspective.

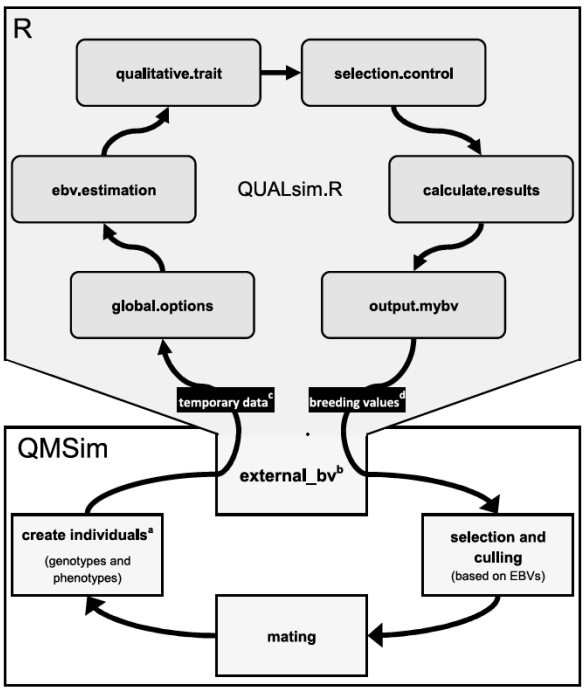

Fig. 1 Simulation flow diagram for the combination of QMSim and R_QUALsim. [a]Start point of the simulation in generation 0. [b]Option external_bv serves as the interface to connect QMSim and R_QUALsim. [c]QMSim produces a temporary dataset that is read by R_QUALsim.R after its initiation. [d]After the successful execution of R_QUALsim, the program creates a file containing the estimated breeding values used for selection in QMSim

Methods

General programming structure and simulation flow

The presented simulation framework is a combination of the whole-genome simulation software QMSim [36] for the simulation of a quantitative trait in a dairy cattle population, and an own R algorithm named QUALsim that simulates Mendelian inheritance for a qualitative trait. In the present analysis, QUALsim serves as an extension to simulate polledness in the population initiated by QMSim, and enables simultaneous selection for both traits using various selection strategies. QUALsim and its components were developed and tested using R version 3.2.0 [39]. Programming and testing was performed using the TinnR Editor for the R environment and the associated R package [40]. QUALsim is based on R base functions and functions from community-contributed packages [41, 42]. A detailed technical description of QUALsim and instructions for the usage of QUALsim are provided as a "Technical Note" (see Additional file 1). All the necessary files to run QUALsim with QMSim are in Additional file 2. Figure 1 illustrates the simulation and data flow between QMSim and QUALsim. To date, we have tested QUALsim on Windows and Linux OS systems. The simulation results presented in this study were obtained on a desktop computer system with the following characteristics: operating system (OS) Windows 7 (64 bit); CPU Intel Core i7-4770 3.4 Ghz; 16 GB Ram.

Population simulation with QMSim

The initial simulation of a dairy cattle population for the quantitative trait was performed by applying QMSim. A quantitative trait ("milk yield") was simulated as a female sex-limited true polygenic trait with a predefined heritability of 0.3 and a phenotypic variance of 1. True breeding values in generation 0 were set to a mean of 0 with a genetic standard deviation (SD) of 0.54 (i.e. square root of 0.3). Accordingly, no QTL or markers and no historical populations were simulated. The founder generation consisted of 250 male and 50,000 female individuals. From generations 1 to 5, the number of females in the population increased by 12.5 % to reflect the growth of superior breeding lines. After generation 5, the size of the population was kept constant with 250 sires and 75,000 dams. Thirty subsequent generations under selection were simulated. Replacement rates for sires and dams were 50 and 25 % per generation, respectively. Estimated breeding values (EBV) were used as selection and culling criteria

for both sires and dams. Selected sires and dams were mated at random. An equal use of selected sires implies 300 offspring per sire and generation. The female reproductive rate was limited to one progeny per female, with an equal probability for either male or female progeny. Each breeding scenario included 20 repetitions.

Simulation extensions with QUALsim.R

QUALsim.R extends the functionality of QMSim with regard to the subject of our study by processing three main tasks consecutively: (1) EBV estimation for the quantitative trait, (2) simulation of the qualitative trait polledness, and (3) weighting of EBV depending on the simulated polled genotypes or phenotypes.

Breeding values were based on simulated phenotypes from QMSim and estimated by using the external software package DMU [43] (module DMU5). The following animal model was applied:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Za} + \mathbf{e},$$

where $\mathbf{y}$ is a vector of observations for the quantitative trait, μ is the overall mean of the observations, $\mathbf{a}$ is a vector of random additive genetic effects, $\mathbf{e}$ is a vector of random residual effects, and $\mathbf{Z}$ is the associated incidence matrix for genetic effects.

The polled trait considered as the qualitative trait of interest in this study is assumed to be controlled by a dominant mutant allele at a single locus that determines the polled phenotype. The current status quo for the black-and-white (BWH) polled Holstein population is characterized by high percentages of heterozygous polled individuals, a lower genetic level and higher genetic relatedness, compared to the horned population [21, 22, 24, 26]. Such a realistic genetically-related polled population was initiated by simulating five generations under selection for the quantitative trait. Qualitative trait genotypes and phenotypes were generated by assigning alleles *P* for *polled* and *p* for *horned* at one locus in the progeny of generation 5. Hereafter, the five generations that precede that for which polled genotypes were assigned, will be labeled as generations −5 to −1. Accordingly, generations under selection for both simulated traits are labeled as generations 0 to 25. The polled genotype assignment algorithm implemented in QUALsim allowed for lower breeding values for the quantitative trait, and higher average genetic relationships among polled individuals. The realized polled allele frequency in generation 0 was equal to 0.03 in all simulated breeding scenarios. Average genetic relationships and average true breeding values (TBV) in generation 0 for the polled and horned group reflect the characteristics of the German and international Holstein populations [22, 24] and were similar across scenarios (Table 1). Allele inheritance at the simulated *polled* locus after generation 0 was computed by simulating random combination of parental alleles during mating. Possible evolutionary factors such as recurrent mutations, effects of crossing-over or possible linkage effects in relation to the quantitative trait, were neglected.

Due to the fact that selection and culling of sires and dams in QMSim are strictly based on the EBV, i.e. in our case, estimates from the DMU software package, we developed an alternative approach that allows simultaneous selection for both simulated traits across generations. Our approach weights EBV (allowing a user-defined weighting factor) for the quantitative trait based on individual genotype or phenotype for the qualitative trait. In the present study, initially we used a weighting factor that reflected one genetic SD of the EBV for the quantitative trait (≈weighting factor 0.5 for a quantitative trait with mean = 0 and SD = 1). In the context of the polled breeding scenario evaluations, the weighting factor can be interpreted as an economic weight for the polled trait, i.e., by mimicking a simplified index which includes the polled status of a given individual. We designed two general polled selection strategies.

The first selection strategy GENO weights EBV using the following formula:

$$EBV_w = EBV_{quant} + \left(wf * n_P\right),$$

where EBV_{quant} is the predicted EBV for the quantitative trait of a given individual, wf is the chosen weighting factor of 0.5, n_P is the number of *polled* alleles *P* of the individual, and EBV_w is the final weighted EBV given back to QMSim. Hence, the selection strategy GENO refers to marker-assisted selection of polled individuals based on gene test results. GENO implies that all animals are genetested at the *polled* locus, and homozygous *polled* individuals are preferably selected.

The second selection strategy PHENO weights EBV using the following formula:

$$EBV_w = EBV_{quant} + \left(wf * PT_{polled}\right),$$

where EBV_{quant} is the predicted EBV for the quantitative trait, wf is the chosen weighting factor of 0.5, PT_{polled} is the binary coded polled phenotype (0 = horned, 1 = polled) of an individual, and EBV_w is the final weighted EBV given back to QMSim. PHENO mimics selection of polled individuals based only on phenotypic information.

While selection strategies that focus on *polled* genotypes (i.e. selection strategy GENO) rely on valid gene tests [19], selection strategies that focus on the polled phenotype (i.e. selection strategy PHENO) might be influenced by different phenotyping errors. In particular, heterozygous polled individuals may develop

Table 1 Average pedigree relationship coefficients, inbreeding coefficients and true breeding values ± SD in selected progeny from generation 0 after the assignment of the *polled* allele

	Phenotype			
	Horned		Polled	
	Male	Female	Male	Female
Average relationship coefficient	0.0863 ± 0.0243	0.0338 ± 0.0112	0.1434 ± 0.0504	0.0347 ± 0.0116
Average inbreeding coefficient	0.0233 ± 0.0098	0.0124 ± 0.0049	0.0423 ± 0.0261	0.0130 ± 0.0050
Average true breeding values	2.1111 ± 0.1028	1.5763 ± 0.0743	1.7393 ± 0.1834	1.5273 ± 0.0728

Results were similar across scenarios

horn-like skull attachments of variable types, the so-called scurs [11, 44], and at an early calf stage, it can be difficult to phenotypically distinguish between scurs and horns. Such possible phenotyping errors were taken into account when simulating PHENO breeding scenarios. Specifically, we simulated a general phenotyping error rate rather than directly simulating the *scurs* locus as responsible for a second separate qualitative trait for two reasons. First, the precise underlying genetic mechanism of the *scurs* locus is not yet clarified. Second, for both important German cattle breeds Holsteins [22] and Simmental [30], recent evaluations lack detailed information with regard to the allele frequencies of *scurs*. Following our simplified error term strategy, 2 % of all polled progeny in each generation were randomly selected and assigned the horned phenotype, although they were genetically polled.

Polled breeding scenarios

For the two general selection strategies GENO and PHENO, we designed different sub-selection strategies by imposing EBV weighting to additional constraints, such as sex-specific weighting scenarios. The breeding strategies evaluated here and hereafter referred to as scenarios, comprise a broad range of possible polled selection strategies that include both theoretical scenarios but also scenarios based on practical implementations of commercial farms and breeding organizations. The reference scenario for the comparisons of polled selection scenarios is a base scenario CONTROL, in which there is no targeted selection for the polled trait. Hence, the qualitative polled trait is simulated according to the described methods, but without EBV weighting. Accordingly, selection in scenario CONTROL is strictly based on unweighted EBV for the quantitative trait. Selection scenarios GENO-ALL and PHENO-ALL apply the corresponding general polled selection strategy as explained above in both sexes. Scenarios GENO-M, GENO-F, PHENO-M, PHENO-F are gender-dependent polled selection strategies by weighting EBV only in one sex (M = only among males, F = only among females). In addition, the scenario GENO-M-PHENO-F applies GENO selection among males and PHENO selection among females.

Results and discussion

Allele, genotype and phenotype frequencies

The CONTROL scenario, which reflects the traditional breeding and selection strategy applied in black-and-white Holstein cattle, is characterized by a further decrease of the initial allele frequency for *polled* from $f_p = 0.03$ in generation 0 to 0.02 in generation 25 (Fig. 2). In several CONTROL runs, the *polled* allele is even totally eliminated from the active population as reflected by the SD from 20 replicates. The decrease in allele frequency for *polled* with the CONTROL scenario is due to the lower genetic level of polled individuals, as achieved through the initial assignment scheme. Thus, inferior polled individuals in the active population are replaced by superior horned individuals since selection is based strictly on EBV for the quantitative trait regardless of the polled status of an individual.

Both overall polled selection strategies GENO-ALL and PHENO-ALL resulted in phenotypically complete polled active populations. Moreover, application of GENO-ALL resulted in full fixation of the *polled* allele after 18 generations. Coherently, prioritized selection of homozygous individuals was reflected by the genotype frequencies. All polled individuals in scenario GENO-ALL were homozygous polled ongoing from generation 17 ($f_{pp} = 1$). In contrast, selection for the polled phenotype in scenario PHENO-ALL, regardless of the precise genotypes, retained a significantly larger number of heterozygous individuals in the active population, which resulted in a significantly lower allele frequency for *polled* [p(P) = 0.8022] after 25 generations. Nevertheless, the realized genotype frequencies in scenario PHENO-ALL show that a strict selection on polled individuals based only on phenotype also substantially increased the number of homozygous polled individuals within a time span of 25 generations.

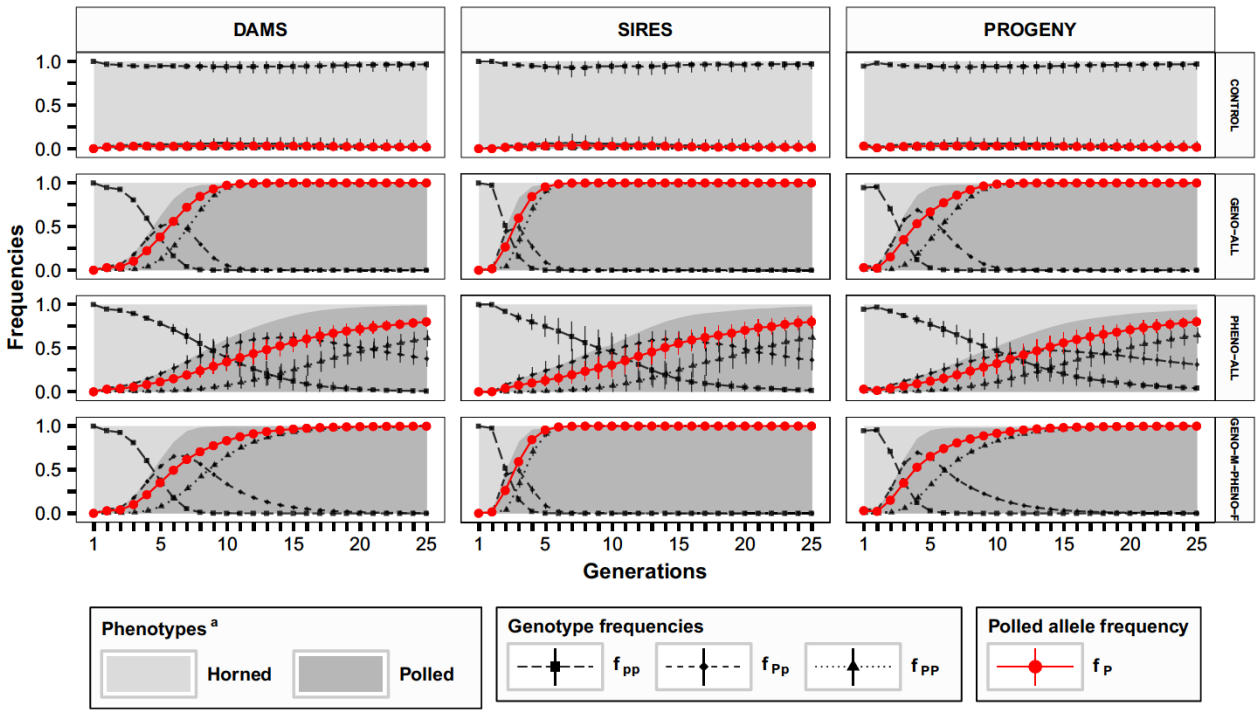

Fig. 2 Evolution of genotype, allele and phenotype frequencies across 25 generations in the active population and progeny. Active population = selected individuals. [a]Sizes of the colored areas are proportional to the percentages of phenotypes in the active population and progeny

Due to the higher selection intensity, the male selection path was significantly more efficient in breeding polled populations compared to the female selection path (scenarios PHENO-M and GENO-M versus PHENO-F and GENO-F, respectively). The results show that the transition from the low initial allele frequency for *polled* of 0.03 in generation 0 to a high percentage of polled sires and dams in the active population was clearly faster in scenarios GENO-M and PHENO-M compared to female scenarios GENO-F and PHENO-F, respectively (see Additional file 3). Hence, active selection of polled sires accelerated the desired selection effects among dams due to the faster dissemination of *polled* alleles in new born selection candidates.

Remarkably, within only five generations, the sex restricted scenario GENO-M led to polled selection effects among the dams that were equivalent to those with the corresponding scenario with active selection in both sexes (GENO-ALL). In contrast, scenario PHENO-M resulted in a moderate increase in the number of polled dams, but the allele frequency of *polled* among the dams after 25 generations was substantially lower than in the corresponding scenario PHENO-ALL (see Fig. 2; see Additional file 3). Restricting selection for polledness to females (scenarios GENO-F and PHENO-F) only moderately increased the number of heterozygous polled dams, with minor associated selection effects on polled sires. Selection strategies PHENO-F and GENO-F reflect traditional polled selection strategies in Holstein cattle for which polledness is mainly transmitted through the female path [22, 23]. The changes in allele and genotype frequencies for these scenarios were almost identical (see Additional file 3). Therefore, application of polled gene tests for cows kept in commercial herds yields no extra response in the allele frequency of *polled* compared to a selection strategy based on female polled phenotypes.

The sex-dependent combination of both general selection strategies (scenario GENO-M-PHENO-F) led to a phenotypically completely polled active population with an allele frequency of 0.99 for *polled* and 99 % of homozygous polled individuals, as in scenarios GENO-ALL and GENO-M.

Genetic gain

In the preceding five generations (=generations −5 to −1) before polledness was simulated, the rate of genetic gain was similar and positive in all breeding scenarios (see Additional file 4). For all scenarios, because of the higher selection intensity, the genetic levels of the sires were generally higher than those of the dams. Scenarios with the highest TBV correspond to those with the lowest allele frequencies for *polled* (see also Fig. 2). Accordingly,

scenarios with either no or only slight increases in the allele frequency for *polled* (CONTROL, GENO-F, and PHENO-F) displayed higher average TBV than those with significant increases in the allele frequency for *polled* (GENO-ALL, GENO-M, PHENO-ALL and GENO-M-PHENO-F). In the latter scenarios, genetic gain after 25 generations was reduced by 4–2 % compared to the CONTROL scenario (Table 2).

Active selection for the polled trait among males (GENO-ALL, GENO-M, PHENO-ALL, PHENO-M and GENO-M-PHENO-F) reduced the average rate of genetic gain per generation in selected sires and dams compared to the CONTROL scenario within 10 generations after the polled allele was assigned. For the later generations 10–25, the rates of average genetic gain for all scenarios were quite constant (see Additional file 4). Final differences in TBV that resulted from reduced rates of genetic gain compared to that of the CONTROL scenario were larger when selection strategies were based on male genotypes (GENO-ALL, GENO-M and GENO-M-PHENO-F) than on male phenotypes (PHENO-ALL and PHENO-M).

The recent evaluations reported by Windig et al. [22] and results for German Red Holstein cattle [24] support the genetic improvement of polled bulls using PHENO strategies. Gaspa et al. [37] applied a moderate PHENO strategy, and found a rather low genetic improvement for polled homo- and heterozygote new-born progeny in a time span of 12 years under conventional BLUP selection and larger losses in rates of genetic gain per year using stochastic simulation. However, the initial parameters that they used, i.e. a rather moderate allele frequency of ~0.10 for *polled,* the significantly lower genetic levels of the polled individuals, the lack of consideration of their relationship level, and the small simulated population, probably explain the small improvements that they observed compared to a more realistic situation. Nonetheless, they identified the potential of further improvements via PHENO strategies when implementing genomic selection [37].

Less overall genetic gain for active selection on polledness is mainly due to long-term selection of male polled selection candidates with lower breeding values for the quantitative trait. Accordingly, a strict preferential selection of homozygous polled sires through higher weighting of their EBV compared to heterozygous polled individuals in scenarios GENO-ALL, GENO-M and GENO-M-PHENO-F retained individuals with lower EBV for the quantitative trait, and excluded genetically-superior horned selection candidates. Polled selection restricted to the female selection path in scenarios GENO-F and PHENO-F showed a comparable effect with increased losses in genetic gain among the active dams compared to the CONTROL dams (see Additional file 4). Thus, practically, polled selection strategies that are restricted to the female pathway may be potentially advantageous for herd performance levels following strict selection of inferior polled cows. Furthermore, allele frequencies for *polled* indicate that active selection for polled males is necessary to achieve sufficient selection responses for polled also in females from a whole population perspective. Accordingly, if only small numbers of polled bulls are available for AI, a moderate selection should be applied to commercial herds using PHENO strategies until more and better polled sires are available from polled breeding programs.

The permanent exclusion of genetically-superior horned selection candidates would not only reduce the genetic potential of the population definitely, but would also unnecessarily decrease the genetic variability of the population. Thus, our simulation results clearly indicate that polled GENO selection strategies should only be applied partially, with moderate intensity and mainly in the male selection pathway by using approaches such

Table 2 Mean frequency of the *polled* allele, true breeding values and inbreeding coefficients for 20 replicates ± SD in generation 25

Scenario	*Polled* allele frequency ± SD	True breeding value ± SD	Inbreeding coefficients ± SD
CONTROL	0.0180 ± 0.0354	8.2280 ± 0.1864	0.1249 ± 0.0142
GENO-ALL	1 ± 0	7.9233 ± 0.1810	0.1084 ± 0.0150
GENO-M	0.9983 ± 0.0002	7.9244 ± 0.1635	0.1132 ± 0.0175
GENO-F	0.3228 ± 0.2050	8.0845 ± 0.1854	0.1300 ± 0.0227
PHENO-ALL	0.8022 ± 0.0504	8.0900 ± 0.1127	0.1191 ± 0.0155
PHENO-M	0.5369 ± 0.0782	8.1564 ± 0.1457	0.1274 ± 0.0151
PHENO-F	0.3052 ± 0.1802	8.0813 ± 0.1350	0.1193 ± 0.0165
GENO-M_PHENO-F	0.9986 ± 0.0002	7.9217 ± 0.1434	0.1132 ± 0.0155

as genomic selection [37] and optimum genetic contribution (OGC) theory [22] in future polled breeding programs.

Inbreeding

Ranking of scenarios according to average inbreeding coefficients generally corresponded to rankings according to TBV, but differences in average inbreeding coefficients among scenarios were quite small (Table 2). The variation of inbreeding coefficients among replicates indicated a substantial impact of individual matings on the actual inbreeding level. A general and similar increase in average inbreeding coefficients as the number of generations increased was observed for all scenarios and for both sexes, with higher levels of inbreeding in bulls than in cows. Average inbreeding rates per generation (ΔF) after generation 0 ranged from 0.312 to 0.576 % (see Additional file 5). Such increases are consistent with recently reported values for the German Holstein and international Holstein populations in the pre-genomic era, these values ranging from 0.44 [45] to 0.95 % [38]. One reason for these slightly lower average inbreeding coefficients in the simulated data could be that the chosen population structure had a relatively small number of active cows compared to the number of active sires, which differs from current practical dairy cattle breeding programs [38]. Nevertheless, we aimed at producing valid inbreeding comparisons across the various polled breeding scenarios, because all scenarios were based on founder populations with the same parameters.

Average inbreeding in the CONTROL scenario showed a consistent linear increase over generations (see Additional file 5). Interestingly, when selecting for the polled trait based on male genotypes (scenarios GENO-ALL, GENO-M and GENO-M-PHENO-F), average inbreeding coefficients after 25 generations were lower than those obtained with the CONTROL scenario. However, the average inbreeding coefficients that were obtained indicated that the lower average inbreeding reached in scenarios GENO-ALL, GENO-M and GENO-M-PHENO-F was mainly due to reduced inbreeding rates in generations 0 to 10. In contrast, in generations 20 to 25, inbreeding rate increased more rapidly, especially among the sires, in scenarios GENO-M and GENO-M-PHENO-F with an assumed impact beyond 25 generations. Average inbreeding coefficients in generation 25 in scenarios PHENO-ALL, PHENO-M, GENO-F and PHENO-F are consistent with those of the CONTROL scenario. In contrast, Gaspa et al. [37] found lower inbreeding rates for a PHENO polled selection strategy using conventional BLUP selection.

Selection based on breeding values from BLUP animal models that combine all the information from relatives contributes to increase the co-selection of related animals with an associated increase in inbreeding [46, 47]. The temporary decrease in average inbreeding as a result of selection for male polled genotypes is partly explained by the selection effects due to the BLUP animal model. Thus, selecting initially only a few individuals and continuously increasing the numbers of polled male and female progeny, decreases average relatedness in the active population by replacing superior and more closely-related horned selection candidates. Such an "alleviation effect" is irrelevant in a long-term perspective with larger proportions of selected polled individuals.

In practice, the group of polled founders (i.e. available polled dams and AI sires already in the population) that could potentially act as donors of the *polled* allele during selection, are highly related [22, 24]. In addition, as shown above, a strict GENO selection strategy cannot be applied in practice because of the implications for genetic gain and performance in the population. Hence, the decrease in inbreeding due to selection based on male polled genotypes (scenarios GENO-ALL, GENO-M and GENO-M-PHENO-F) should not be interpreted as a realistic possibility to reduce inbreeding levels concurrent to polled selection. Instead, the results for the currently practiced PHENO-M selection strategy should be evaluated critically. The high relatedness between potential donors of the *polled* allele could in reality lead to higher inbreeding levels in the long term following an intensified selection for the polled trait [22].

Important aspects for practical polled breeding

With regard to practical selection decisions, our results strongly suggest that application of GENO selection strategies among males will maximize selection response for polled. The corresponding scenarios GENO-M and GENO-M-PHENO-F resulted in a completely polled active population in a reasonable time span with reduced costs and efforts for genotyping and phenotyping. Scenario GENO-ALL led to a similar result, but the broad genotyping of commercial milking cows at the *polled* locus using the available gene test cannot be carried out in practice due to the current genotyping costs (e.g., 27€ per cow, [19]). As an alternative, imputation of polled genotypes based on marker and pedigree data with low error rates might contribute to broader genotyping activities at an acceptable cost level [24]. From a practical breeding perspective and also considering the costs of genotyping, scenarios GENO-M and GENO-M-PHENO-F seem to be the most efficient strategies to increase the frequency of the *polled* allele in the population. In this context, additional genotyping of females (e.g., as for scenarios GENO-ALL and GENO-M-PHENO-F) resulted only in minor gains regarding the final frequency of the

polled allele. Commercial herds should focus on balanced selection strategies with regard to the use of available polled AI and elite horned AI sires following traditional selection strategies [23]. Such a strategy requires that potential new born polled progeny be carefully phenotyped, in order to introgress the *polled* allele into the herd. Nevertheless, Segelke et al. [24] suggested an active selection of elite polled females (e.g. potential polled bull dams), which complements the intensive selection among polled males. This suggested strategy contributes to a faster increase of both evaluation criteria i.e. frequency of the *polled* allele and genetic gain among potential polled AI bull selection candidates.

In Holstein AI programs, male polled selection candidates were generally outperformed by horned sires. There are only a few exceptions, e.g. the polled sires Lawn Boy in red Holstein and Mitey P in black-and-white Holstein that disseminated the *polled* allele through the male pathway of selection. Continuous use of only a small number of available polled AI sires has resulted in a population of closely-related polled individuals and in higher inbreeding in the polled subpopulation [22, 24]. The comparison of our results from the simulation with current practical developments indicates that the reported increase in the frequency of the *polled* allele in dairy cattle breeds [22, 30] is consistent with the trend observed in scenarios PHENO-ALL and PHENO-M. Our findings from the GENO scenarios are supported by previously published simulation results [37] and both studies recommend the continued use of gene-tested polled AI sires to achieve high overall frequencies of the *polled* allele within a reasonable time span. The success of the polled AI breeding program in Simmental cattle is exemplarily in this regard [30]. The numerical increase of polled AI sires in black-and-white Holstein in recent years reflects the efforts of the German as well as the international Holstein breeding organizations to broaden the polled sire breeding pool, and to create a basis for structured polled breeding programs [22].

For calf dehorning to be completely abandoned requires 100 % phenotypically-polled new born progeny, which was achieved in scenarios GENO-ALL, GENO-M, GENO-M-PHENO-F within 10 generations, respectively (Fig. 2). However, to maintain a 100 % polled population in the long term requires full fixation of the *polled* allele through selection, which implies a completely homozygous polled active population. In scenario GENO-ALL, all new born progeny are homozygous *polled* ongoing from generation 17. In scenarios GENO-M and GENO-M-PHENO-F, we observed a small number of heterozygous polled progeny up to generation 25. In addition, the results in Fig. 2 clearly illustrate that a selection strategy that includes the genotyped males (GENO-ALL, GENO-M, GENO-M-PHENO-F) is essential to achieve complete polledness in new born progeny. In contrast, selection based on polled phenotypes (PHENO-ALL and PHENO-M) will result in a substantial number of horned progeny still present in generation 25. Specific assortative mating schemes for genotyped polled individuals have the potential to accelerate the breeding process towards polled progeny [25], but in practice, assortative mating schemes are only defined by elite breeders, and with limited applications in commercial herds [48]. Nonetheless, an increasing number of available polled AI bulls with valid gene test results [22, 24, 30] including homozygous *polled* sires, allows commercial farmers to apply assortative polled matings for a faster dissemination of the *polled* allele in their herds.

Other specific polled breeding applications

We focused on scenario GENO-M-PHENO-F for a further extension of the presented simulation approach aiming at reducing the loss in genetic gain concurrent to the increase in frequency of the *polled* allele (see Fig. 2). For that reason, we changed the weighting factor for GENO selection among males to a lower value of 0.1, while maintaining the high weighting factor of 0.5 for PHENO selection among females. Reducing the male weighting factor (wf-M-0.1) significantly decreased the desired selection response for the polled trait in sires as well as in dams in the first generations (see Table 3; see Additional file 6) compared to wf-M-0.5 (i.e. being the originally simulated GENO-M-PHENO-F scenario). However, the final average frequency of the *polled* allele after 25 generations was equal to 0.93 for dams and 0.98 for sires, which indicated a progressive acceleration of selection response for the qualitative trait. The final overall frequency of the polled phenotype (0.99) also indicated that nearly all the individuals in the active population and new born progeny were polled after 25 generations.

Reducing the weighting factor among males (wf-M-0.1) limited the loss in overall genetic gain for the quantitative trait in sires and dams to 2 % compared to CONTROL (see Table 3; Additional file 7a). Reducing the male weighting factor in scenario wf-M-0.1 resulted in similar average genetic merits for different sire genotypes. In contrast, we found a remaining small deficit in the genetic value of selected polled dams compared to selected horned dams in generation 25. Reducing the weighting factor among males (wf-M-0.1) led to similar inbreeding levels compared to the CONTROL scenario (see Table 3; Additional file 7b).

A fast transition to a completely polled active population and furthermore completely polled progeny is opposed to the preservation of high genetic gain in the quantitative trait. Nonetheless, results from the

Table 3 Further application: mean frequency of the *polled* allele, true breeding values and inbreeding coefficients for 20 replicates ± SD in generation 25

Scenario	*Polled* allele frequency ± SD	True breeding value ± SD	Inbreeding coefficients ± SD
CONTROL	0.0180 ± 0.0354	8.2280 ± 0.1864	0.1249 ± 0.0142
wf-M-0.1	0.9348 ± 0.0353	8.0555 ± 0.1649	0.1210 ± 0.0158
wf-M-0.5	0.9986 ± 0.0002	7.9217 ± 0.1434	0.1132 ± 0.0155

simulations in which sex-dependent weighting factors were applied, indicate that this decline in genetic gain for the quantitative trait can be limited in combination with significant increases in the proportion of polled individuals. As a compromise, we suggest an approach that takes the described sex-dependent structurally driven effects into account. Hence, it is essential that intensified selection for the polled trait aims at improving the genetic level of polled selection candidates (homozygous as well as heterozygous polled progeny). Such a suggested rather mild selection strategy for the polled trait among male AI candidates is possible with GENO-M and a moderate weighting factor, combined with more intensive selection among females based on polled phenotypes (PHENO-F). This strategy reflects current practical breeding programs that use assortative elite mating schemes and genomic selection, which results in improved EBV for polled Holstein AI bulls [22, 25, 37].

Conclusions

A fast and lasting dissemination and fixation of the *polled* allele across 25 generations implies a strict selection strategy based on *polled* genotypes. Considering the current characteristics of the available polled AI bulls in most dairy cattle breeds, simulation results indicate that such a strategy is coupled with significant decreases in genetic gain for quantitative performance traits. Selection strategies based only on phenotypic information for the polled trait also led to high frequencies of the *polled* allele, but without its fixation after 25 generations. Such strategies based on phenotype information result in significant increases in the number of heterozygous individuals remaining in the population and in the number of horned progeny born up to the final generations. Therefore, abandoning completely dehorning is not possible when selection is based on polled phenotypes only. The application of polled selection strategies based on gene tests for the male selection pathway combined with moderate weighting of polled genotypes during selection, and a phenotypic polled selection strategy for females using high weighting of polled phenotypes, appears to be the optimal compromise regarding all important evaluation criteria.

Additional files

Additional file 1. AF1_TechnicalNote_QUALsim.

Additional file 2. AF2_QUALsim.

Additional file 3. Evolution of genotype, allele and phenotype frequencies across 25 generations in the active population and progeny: Additional scenarios. Active population = selected individuals; [a]sizes of the colored areas are proportional to the percentages of phenotypes in the active population and progeny.

Additional file 4. Average true breeding values (TBV) for active sires and dams across 25 generations. Results for average true breeding values (TBV) of active sires and dams in the CONTROL scenario (labeled as SIRES_CONTROL and DAMS_CONTROL) are included in each plot as a reference.

Additional file 5. Average inbreeding coefficients in the active population across 25 generations. Results for average inbreeding coefficients of active sires and dams in the CONTROL scenario (labelled as SIRES_CONTROL and DAMS_CONTROL) are included in each plot as a reference.

Additional file 6. Further application – Evolution of genotype, allele and phenotype frequencies across 25 generations in the active population for scenario wf-M-0.1 (GENO-M-PHENO-F). Active population = selected individuals. [a] the sizes of the colored areas are proportional to the percentages of phenotypes in the active population.

Additional file 7. Further applications (a) Average true breeding values (TBV) for active sires and dams over 25 generations for scenario wf-M-0.1 (GENO-M-PHENO-F) and (b) Average inbreeding coefficients for active sires and dams over 25 generations. Results for average TBV of active sires and dams in the CONTROL scenario (labelled as SIRES_CONTROL and DAMS_CONTROL) are included in each plot as a reference.

Authors' contributions
CS designed the R-program, performed all analyses and drafted the manuscript. SK contributed to the design of the study. TY, MWD and HD assisted in technical support. SK and HS initiated the polled simulation study and assisted in writing the manuscript. All authors read and approved the final manuscript.

Author details
[1] Department of Animal Breeding, University of Kassel, 37213 Witzenhausen, Germany. [2] Institute of Agricultural and Nutritional Sciences, University of Halle, 06099 Halle, Germany.

Acknowledgements
The present study was supported by funds from the German Government´s Special Purpose Fund held at Landwirtschaftliche Rentenbank. The authors gratefully thank the Landwirtschaftliche Rentenbank for funding this project. The authors also thank the editor and two anonymous reviewers for their constructive comments, which helped to improve the manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 30 October 2015 Accepted: 18 June 2016
Published online: 29 June 2016

References

1. Stafford KJ, Mellor DJ. Dehorning and disbudding distress and its alleviation in calves. Vet J. 2005;169:337–49.
2. Stafford KJ, Mellor DJ. Addressing the pain associated with disbudding and dehorning in cattle. Appl Anim Behav Sci. 2011;135:226–31.
3. Morisse JP, Cotte JP, Huonnic D. Effect of dehorning on behaviour and plasma cortisol responses in young calves. Appl Anim Behav Sci. 1995;43:239–47.
4. Faulkner PM, Weary DM. Reducing pain after dehorning in dairy calves. J Dairy Sci. 2000;83:2037–41.
5. Heinrich A, Duffield TF, Lissemore KD, Millman ST. The effect of meloxicam on behavior and pain sensitivity of dairy calves following cautery dehorning with a local anesthetic. J Dairy Sci. 2010;93:2450–7.
6. Graf B, Senn M. Behavioural and physiological responses of calves to dehorning by heat cauterization with or without local anaesthesia. Appl Anim Behav Sci. 1999;62:153–71.
7. Guatteo R, Levionnois O, Fournier D, Guémené D, Latouche K, Leterrier C, et al. Minimising pain in farm animals: the 3S approach—'Suppress, Substitute, Soothe'. Animal. 2012;6:1261–74.
8. Waiblinger S, Menke C. D.2.3.1. Report on practical recommendations at farm level for keeping horned cattle and on the use of genetic solutions. Alternatives to castration and dehorning (ALCASDE; SANCO/2008/D5/018); 2009.
9. Menke C, Waiblinger S, Fölsch DW, Wiepkema PR. Social behaviour and injuries of horned cows in loose housing systems. Anim Welf. 1999;8:243–58.
10. Goonewardene LA, Price MA, Liu MF, Berg RT, Erichsen CM. A study of growth and carcass traits in dehorned and polled composite bulls. Can J Anim Sci. 1999;79:383–5.
11. Wiedemar N, Tetens J, Jagannathan V, Menoud A, Neuenschwander S, Bruggmann R, et al. Independent polled mutations leading to complex gene expression differences in cattle. PLoS One. 2014;9:e93435.
12. Tetens J, Wiedemar N, Menoud A, Thaller G, Drögemüller C. Association mapping of the scurs locus in polled Simmental cattle—evidence for genetic heterogeneity. Anim Genet. 2015;46:224–5.
13. Rothammer S, Capitan A, Mullaart E, Seichter D, Russ I. The 80-kb DNA duplication on BTA1 is the only remaining candidate mutation for the polled phenotype of Friesian origin. Genet Sel Evol. 2014;46:44.
14. Medugorac I, Seichter D, Graf A, Russ I, Blum H, Göpel KH, et al. Bovine polledness—an Autosomal dominant trait with allelic heterogeneity. PLoS One. 2012;7:e39477.
15. Mariasegaram M, Reverter A, Barris W, Lehnert SA, Dalrymple B, Prayaga K. Transcription profiling provides insights into gene pathways involved in horn and scurs development in cattle. BMC Genomics. 2010;11:370.
16. Mariasegaram M, Harrison BE, Bolton JA, Tier B, Henshall JM, Barendse W, et al. Fine-mapping the POLL locus in Brahman cattle yields the diagnostic marker CSAFG29. Anim Genet. 2012;43:683–8.
17. Capitan A, Grohs C, Gautier M, Eggen A. The scurs inheritance: new insights from the French Charolais breed. BMC Genet. 2009;10:33.
18. Allais-Bonnet A, Grohs C, Medugorac I, Krebs S, Djari A, Graf A, et al. Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. PLoS One. 2013;8:e63512.
19. GeneControl GmbH. 2015. http://www.genecontrol.de/e_hornlosigkeit.html. Accessed 11 Jun 2015.
20. Schafberg R, Swalve HH. The history of breeding for polled cattle. Livest Sci. 2015;179:54–70.
21. Windig JJ, Eggen A. D.2.2.2. Report on the assessment of breeding strategies in relation to the introduction of the polled gene. ALCASDE: Study on the improved methods for animal-friendly production. In particular on alternatives to the dehorning of cattle. Appendix 23. 2009. http://ec.europa.eu/food/animal/welfare/farm/docs/calves_alcasde_d-2-2-2.pdf.
22. Windig JJ, Hoving-Bolink RA, Veerkamp RF. Breeding for polledness in Holstein cattle. Livest Sci. 2015;179:96–101.
23. Specht L. Polled Holstein history. 2008. 10.1016/j.livsci.2015.05.017. http://extension.psu.edu/animals/dairy/documents/polled-holsteins-history. Accessed 4 May 2015.
24. Segelke D, Täubert H, Reinhardt F, Thaller G. Chances and limits of breeding polled cattle Deutsche Holstein. Züchtungskunde. 2013;85:253–69.
25. Spurlock DM, Stock ML, Coetzee JF. The impact of 3 strategies for incorporating polled genetics into a dairy cattle breeding program on the overall herd genetic merit. J Dairy Sci. 2014;97:5265–74.
26. Scheper C, Yin T, König S. Inclusion of polled geno- or phenotype into breeding goals: Impact on genetic gain and inbreeding. In: Proceedings of the 66th annual meeting of the European association for animal production, 31 August–4 September 2015, Warsaw; 2015.
27. Lamminger A, Hamann H, Röhrmoser G, Rosenberger E, Kräusslich H, Distl O. Relationships between polledness and traits used in the breeding objectives for German Fleckvieh. Züchtungskunde. 2000;72:325–39.
28. Goonewardene LA, Pang H, Berg RT, Price MA. A comparison of reproductive and growth traits of horned and polled cattle in three synthetic beef lines. Can J Anim Sci. 1999;79:123–7.
29. Frisch JE, Nishimura H, Cousins KJ, Turner HG. The inheritance and effect on production of polledness in four crossbred lines of beef cattle. Anim Prod. 1980;31:119–26.
30. Götz KU, Luntz B, Robeis J, Edel C, Emmerling R, Buitkamp J, et al. Polled Fleckvieh (Simmental) cattle—current state of the breeding program. Livest Sci. 2015;179:80–5.
31. Hoban S, Bertorelle G, Gaggiotti OE. Computer simulations: tools for population and evolutionary genetics. Nat Rev Genet. 2011;13:110–22.
32. Daetwyler HD, Calus MPL, Pong-Wong R, de Los Campos G, Hickey JM. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics. 2013;193:347–65.
33. Peng B, Chen H, Mechanic LE, Racine B, Clarke J, Clarke L, et al. Genetic simulation resources: a website for the registration and discovery of genetic data simulators. Bioinformatics. 2013;29:1101–2.
34. Täubert H, Reinhardt F, Simianer H. ZPLAN+: A new software to evaluate and optimize animal breeding programs. In: Proceedings of the 9th world congress on genetics applied to livestock production, 1–6 August 2010, Leipzig; 2010.
35. Rutten MJM. SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. J Hered. 2002;93:456–8.
36. Sargolzaei M, Schenkel FS. QMSim: a large-scale genome simulator for livestock. Bioinformatics. 2009;25:680–1.
37. Gaspa G, Veerkamp RF, Calus MPL, Windig JJ. Assessment of genomic selection for introgression of polledness into Holstein Friesian cattle by simulation. Livest Sci. 2015;179:86–95.
38. Koenig S, Simianer H. Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. Livest Sci. 2006;103:40–53.
39. R Core Team. R: a language and environment for statistical computing. 2015. http://www.R-project.org/. Accessed 01 Feb 2015.
40. Faria JC. Resources of Tinn-R GUI/Editor for R Environment 2012. Ilheus, Bahia, Brasil.
41. Coster A. Pedigree: pedigree functions. 2012. http://CRAN.Rproject.org/package=pedigree. Accessed 01 Feb 2015.
42. Wickham Hadley. The split-apply-combine strategy for data analysis. J Stat Softw. 2011;40:1–29.
43. Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau R. DMU—a package for analyzing multivariate mixed models. In: Proceedings of the 8th world congress on genetics applied to livestock production, 13–18 August, 2006, Belo Horizonte; 2006.
44. Long CR, Gregory KE. Inheritance of the horned, scurred and polled condition in cattle. J Hered. 1978;69:395–400.
45. Stachowicz K, Sargolzaei M, Miglior F, Schenkel FS. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. J Dairy Sci. 2011;94:5160–75.
46. Verrier E, Colleau JJ, Foulley JL. Long-term effects of selection based on the animal model BLUP in a finite population. Theor Appl Genet. 1993;87:446–54.
47. Bijma P, Woolliams JA. Prediction of rates of inbreeding in populations selected on best linear unbiased prediction of breeding value. Genetics. 2000;156:361–73.
48. Wensch-Dorendorf M, Yin T, Swalve HH, König S. Optimal strategies for the use of genomic selection in dairy cattle breeding programs. J Dairy Sci. 2011;94:4140–51.

CHAPTER 6

General Discussion

In the light of the aim of this thesis, which is to investigate potential side effects of polledness and appropriate breeding strategies, the following chapter attempts to coherently discuss the main findings of the included research studies. In addition, other practical considerations of breeding for polledness that could not be covered in the previous chapters are discussed. Finally, an extended outlook is presented, considering current developments in modern breeding programs, as well as technological and scientific advances that may influence the results and dynamics of breeding for polledness.

How likely are secondary effects of breeding for polledness based on the current updated knowledge?

Summarizing the results of this thesis and current results from the literature (Gehrke et al., 2016; Cole et al., 2016; Cole and Null, 2019; Gehrke, 2020; Gehrke et al., 2018; Randhawa et al., 2021), it can be assumed with increasing certainty, that breeding for polledness does not have any fundamental negative effects. The results for Simmental presented in **Chapter 3** show, based on a sufficiently large test data set and suitable methodology, that pleiotropic effects of the polled locus are present in single traits (i.e. protein percentage), but without significant negative consequences in general. The statistically significant QTL effect of the polled locus estimated for protein percentage contributed to 6.6% of the genetic variance and 2.3% of the phenotypic variance in this regard. The respective genetic correlation between polledness and protein percentage however was neutral. Especially in traits anecdotally reported to be negatively influenced by the trait polledness, such as fertility traits, no pleiotropic effects were found. This is in line with current results in the Holstein breed (Gehrke, 2020), where family specific effects were found that contradict a pleiotropic effect, but point to linkage to putative QTLs for secondary traits as a potential reason for side effects of polledness. The results presented in **Chapter 4** also point in the same direction when considering identified selection signatures which are located in candidate genes with functions for secondary traits and in proximity to the polled locus.

Major effects on secondary traits in the course of breeding for polledness cannot be ruled out, but obviously have diverse and more complex reasons, which are currently not fully understood. This in turn corresponds to the perception of a significantly higher complexity in the genetic structure of polledness, in part contradicting the view of polledness as a classical Mendelian trait. Hence, the results of Gehrke et al. (2020) indicate that the polled locus is part of a much more complex genetic trait structure for horn formation rather than a simple and isolated trait. In addition, the growing

diversity of causative variants at the polled locus also bears new potential to explain diverging effects, for example when comparing different breeds (Medugorac et al., 2012; Medugorac et al., 2017; Utsunomiya et al., 2019). The presented results in **Chapter 4** also point in this direction. Further and new insights into the trait polledness and its association to other traits can therefore only be expected if future studies address this complexity on the phenotypic as well as genetic level. The recent utilization of novel gene-editing methods leading to new insights into the physiological and functional pathways involved in horn development show the potential in this regard (Aldersey et al., 2020; Schuster et al., 2020). In addition, the variety of still poorly understood associated phenotypic (and potentially also genetic) phenomena prevalent in polled individuals such as double-eyelashes (Gehrke, 2020) also need to be addressed in further research.

Practical considerations and implications for strategies to sustainably intensify breeding for polledness in Holstein and Simmental cattle

From a practical breeding perspective however, the mentioned implications and complexity should be addressed by adjusted selection strategies to identify and select only superior polled sires and cows based on available information (i.e. gene tests, genotyping and genomic breeding values) while limiting potential side effects at various levels. In this regard, the aims and requirements of breeding organizations and farmers to breed polled animals successfully and sustainably are not always easy to balance. Farmers naturally have a high demand primarily for homozygous polled sires as all offspring are definitely polled in order to achieve a fast transition to a polled herd. Based on the results from the stochastic simulations (**Chapter 5**)**,** additional simulations further highlight this practical aspect. In this regard, we simulated two exemplary mating scenarios, i.e., scenario 1 – exclusive use of heterozygous (Pp) polled sires, and scenario 2 – exclusive use of homozygous (PP) polled sires, for three different herd sizes (50, 100, 500 cows). A targeted selection of polled female calves inside the herds was not simulated. In order to achieve realistic herd characteristics, all scenarios were based on similar age structures (average age of cows in the herd = 5 years) and replacement rates (i.e. 0.25). Each active cow had 1 calf per year, and the probability of a calf being born male or female was 0.5. Male calves were not considered in the results. Per scenario, 100 repetitions were simulated for 50 years. All scenarios started with a base polled allele frequency of 0.05, representing a farmer that only recently started to focus on breeding for polledness.

Figure 1 shows the results for an active herd size of 100 cows, the other herd sizes are not shown because of similar results. The simplified single herd simulation study reveals quite clearly that a fast transition to a fully polled herd is only possible by exclusively using homozygous polled sires. A fast transition in 8-10 years to a fully polled herd can only be achieved when using homozygous polled sires exclusively. If heterozygous polled sires are exclusively used, the average percentage of polled animals (Pp and PP) after 15 years is 69.1 %, and after 50 years 75.7 %.

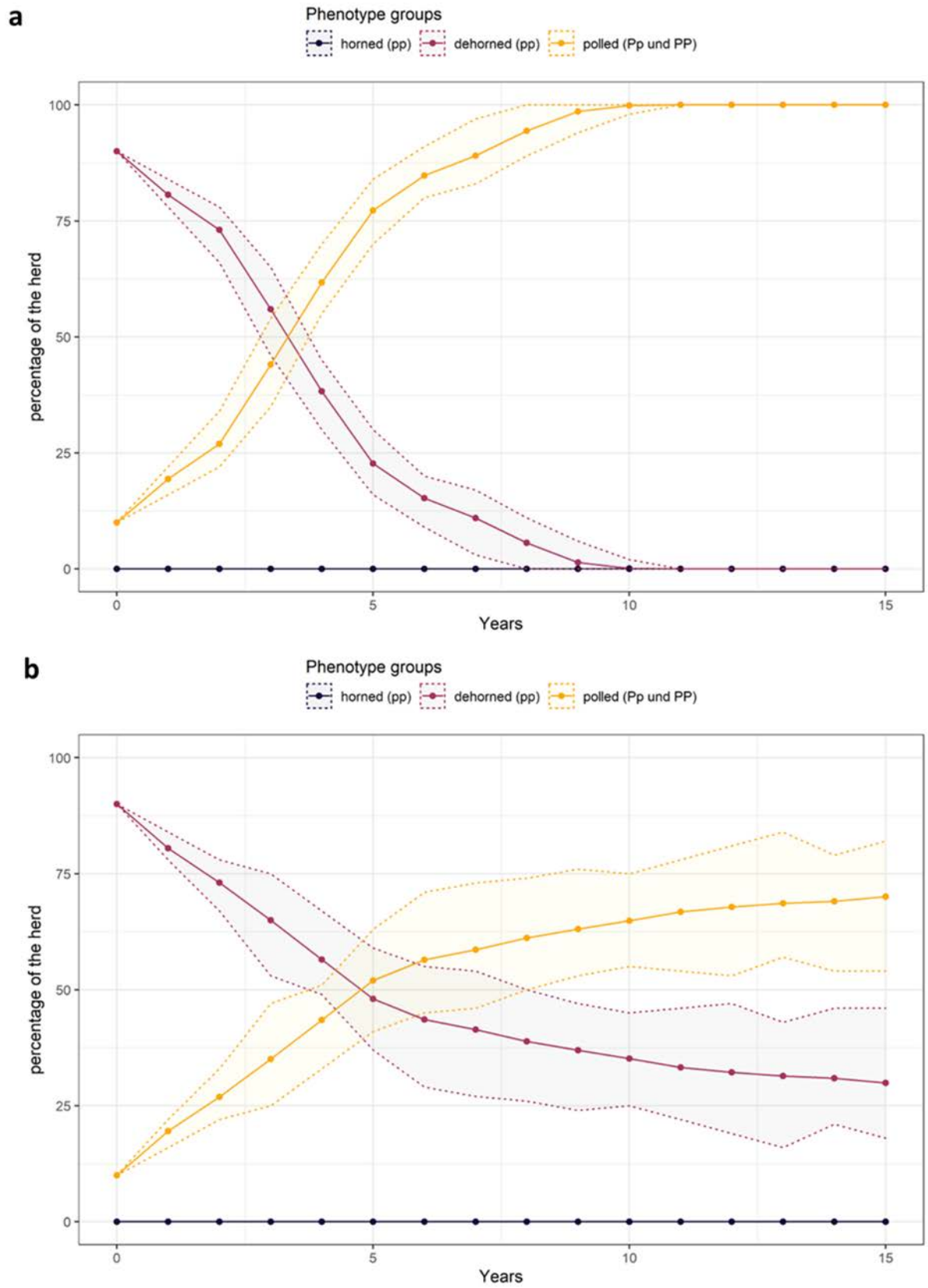

Figure 1 Average percentage of different phenotype groups for 15 years after exclusive usage of homozygous polled sires (a) or heterozygous polled sires (b) in a single herd simulation. Solid lines represent the average percentages over 100 replicates. Dashed lines represent the results from maximal and minimal replicates among all replicates.

The main aim in this rather simplified simulation approach was further to reflect a situation when

a rapid policy change on animal welfare regulations forces the majority of farmers to breed for polledness as quickly as possible. This is for example already in parts the case in organic dairy farming in Germany (European Union, 2023b) and highlights the need to already provide a sufficient number of homozygous polled sires while balancing overall genetic gain and genetic diversity in the breeding programs.

Unfavorably in this regard, homozygous polled sires traditionally had considerably lower average breeding values in Simmental and Holstein cattle. Figure 1 (also shown similarly in the General Introduction) clearly reflects that although the deficit in average total net merit of hetero- and homozygous polled sires has become smaller in recent years, especially homozygous polled sires still show lower average total net merits.

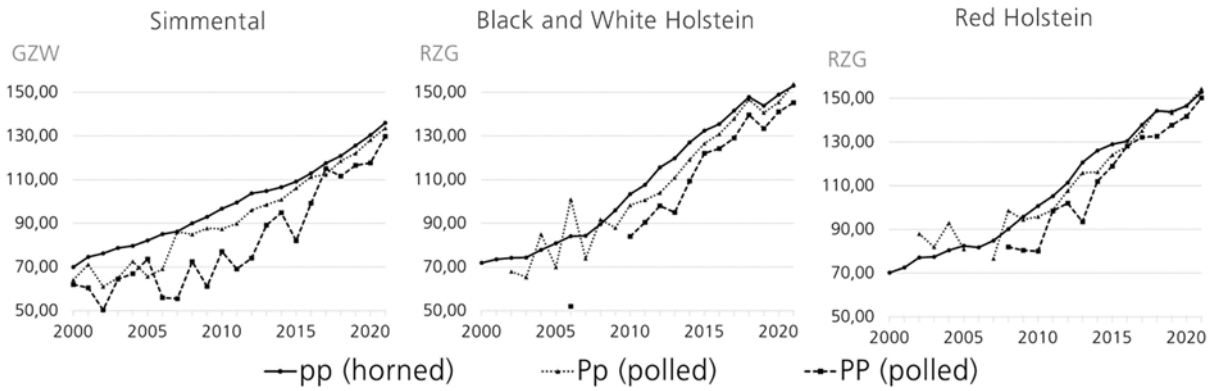

Figure 2 Development of average total net merit breeding values for horned and polled sires in the Simmental and Holstein cattle in Germany. GZW = total net merit in German Simmental cattle, RZG = total net merit in German Holstein cattle

The breeding organizations can only meet the high demand for homozygous polled sires by continuously focusing on high genetic gain for polledness (i.e. a fast polled allele frequency increase) among male selection candidates. However, as shown in **Chapter 5**, such a selection strategy (i.e. scenarios GENO-ALL and GENO-M-PHENO-F in the simulations study) potentially leads to considerably less genetic gain in economically important quantitative traits if implemented too intensively. The development in recent years as shown in Figure 2 (see also Figure 1 in the general introduction) reflects that the breeding organizations are successful in sustainably enhancing the percentage of polled sires while preserving high genetic gain in the population. At this point, with ~20-25% selected polled sires in Simmental and Black and White Holstein and ~50% selected polled sires in Red Holstein in new birth years no general drawbacks in genetic gain should be expected from the future development.

Potential influence of developments in breeding programs and technological advance on breeding for polledness

One of the major drawbacks for quantitative genetic and genomic studies focusing on secondary

effects of the polled locus in the past up until today was the lack of available valid single- gene (i.e. at the polled locus) or whole-genome genotypes for a large number of animals. In the past, available datasets typically consisted of only low numbers of animals with available genotype data, often complemented with animals only with phenotype data for the trait polledness. This was also the case for the dataset presented in **Chapter 2**. These basic conditions made large-scale studies focusing on pleiotropic effects of the polled locus including sufficient numbers of hetero- as well as homozygous polled cows difficult, apart from the fact that the overall number of polled animals in Holstein and Simmental cattle was low, because polledness only recently gained broad popularity among farmers. With the dynamic development of genotyping projects enabling farmers to routinely genotype large numbers of females in the population, the basic conditions to further study potential side-effects due to pleiotropy or linkage of the polled locus will be improved in the near future. The available methods as presented in **Chapter 2 and 3** of this thesis are tested and can be quickly applied to study larger datasets. Other approaches directly developed for or in the context of whole-genome genotype data are also available and show the potential for further advances (Legarra and Vitezica, 2015). Similar to the approach presented in **Chapter 2 and 3** Legarra and Vitezica showed that linear multi-trait models including "gene content" (i.e. the number of alleles at a gene locus) as a trait are suitable to cover major gene effects on quantitative traits if some individuals are not genotyped. Hence, they also proved that an imputation (or a "manual" reconstruction as in the dataset used in this thesis) is not generally necessary to cover the major gene effect if a significant portion of the studied population is genotyped. It should however be noted that their study did not focus on a qualitative trait as a source of a major gene effect. Nonetheless, they generally proved the potential to routinely include major gene effects on quantitative traits during routine genetic evaluation including modern Single-Step applications.

As a positive subsidiary effect to enhanced routine genotyping in general, farmers can use a valid selection tool for the targeted selection of female homozygous polled animals during routine SNP-chip genotyping of calves, supporting the polled selection strategy from a farmers perspective.

The second field characterized by a very dynamic technological and methodological development that could influence breeding for polledness are novel gene editing methods. Simulation studies have already shown the potential of these methods to overcome the already presented conflict to balance overall genetic gain, genetic diversity and increase in polled allele frequency in the "conventional" way of breeding for polledness (Mueller et al., 2019; Mueller et al., 2021; Bastiaansen et al., 2018). Given these simulation results, gene-editing methods were more efficient and cost-effective for the introgression of the polled allele in breeds with only a low initial polled allele frequency. When comparing the economics of gene-editing methods to conventional breeding, it was found that gene editing was more expensive initially but had lower long-term costs. Vice versa, conventional breeding was less expensive initially, but had higher long-term costs

(Mueller et al., 2021; Mueller et al., 2019). Studies actually utilizing gene-editing methods to create polled animals are already published (Schuster et al., 2020; Aldersey et al., 2020; Carlson et al., 2016). However, to the knowledge of the author, gene-editing methods are not yet part of any actual commercial cattle breeding program. In addition, under current European legislation (European Union, 2023a; Court of justice of the european union, 2023), gene-editing methods are regulated synonymously to genetically modified organisms (GMOs), heavily restricting their usage and commercialization in plants and animals. In the legal context, there will be a re-evaluation of the decision to regulate gene-editing methods synonymously to GMOs soon, but the outcome and its consequences for gene-editing methods in European breeding programs is completely uncertain.

General conclusions and practical recommendations

Summing up all major results from the separate chapters of this thesis, the conclusions are as follows:

- in German Simmental cattle no evidence for direct pleiotropic or linked QTL effects of the polled locus on the production traits milk yield and fat percentage, the udder health indicator SCS and on female fertility traits non-return rate 56, days open and days to first service.
- in German Simmental cattle we found only one statistically significant and moderately large direct QTL effect of the polled locus on protein percentage. The respective genetic correlation between polledness and protein percentage however was neutral.
- further selection on polledness implies no negative side effects on breeding goal traits in German Simmental. We conclude that any remaining inferiority of polled cows and bulls will be reduced by increasing the dissemination of the polled alleles in the population followed by intensive simultaneous selection on breeding goal traits and polled genotypes in the German Simmental population.
- Selection signature analysis in German Holstein cattle revealed a region of 25Mb at the proximal end of BTA1 including the polled locus as directly influenced by historic and recent selection. In this segment, no significant associations with other phenotypic traits could be detected based on a genome-wide association. We found no indications that significant direct effects of selection at the polled locus affect secondary traits.
- A fast transition to a fixated polled population through intensified selection for polledness is in contradiction to the preservation of high genetic gain for a simulated quantitative trait.
- Fast transitions to a fixated polled herd imply the exclusive usage of homozygous polled sires.
- The selection strategy displaying the best compromise regarding all important breeding aspects (genetic gain in the overall breeding goal, genetic diversity/inbreeding and genetic

gain in polledness) consisted of selection on male polled genotypes with moderate weighting in selection, and selection on female polled phenotypes with high weighting.

References

Aldersey, J. E., T. S. Sonstegard, J. L. Williams, and C. D. K. Bottema. 2020. Understanding the effects of the bovine POLLED variants. Animal Genetics 51(2):166–176. https://doi.org/10.1111/age.12915.

Bastiaansen, J. W. M., H. Bovenhuis, M. A. M. Groenen, H.-J. Megens, and H. A. Mulder. 2018. The impact of genome editing on the introduction of monogenic traits in livestock. Genet Sel Evol 50(1):18. https://doi.org/10.1186/s12711-018-0389-7.

Carlson, D. F., C. A. Lancto, B. Zang, E.-S. Kim, M. Walton, D. Oldeschulte, C. Seabury, T. S. Sonstegard, and S. C. Fahrenkrug. 2016. Production of hornless dairy cattle from genome-edited cell lines. Nat Biotechnol 34(5):479–481. https://doi.org/10.1038/nbt.3560.

Cole, J. B., and D. J. Null. 2019. Short communication: Phenotypic and genetic effects of the polled haplotype on yield, longevity, and fertility in US Brown Swiss, Holstein, and Jersey cattle. Journal of dairy science 102(9):8247–8250. https://doi.org/10.3168/jds.2019-16530.

Cole, J. B., D. J. Null, and P. M. VanRaden. 2016. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. Journal of dairy science 99(9):7274–7288. https://doi.org/10.3168/jds.2015-10777.

Court of justice of the european union. 2023. JUDGMENT OF THE COURT (Grand Chamber). Accessed Mar 12, 2023. https://curia.europa.eu/juris/document/document.jsf?text=&docid=204387&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=1865887.

European Union. 2023a. DIRECTIVE 2001/18/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Accessed Mar 12, 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02001L0018-20210327&from=DE.

European Union. 2023b. REGULATION (EU) 2018/848 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2018 on organic production and labelling of organic products and repealing Council Regulation (EC) No 834/2007. Accessed Mar 12, 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02018R0848-20220101&from=DE.

Gehrke, L. 2020. Investigations of the genetic architecrure of horns in cattle (bos taurus) for improved breeding strategies for polledness. Phd Thesis. Phd Thesis.

Gehrke, L., D. Seichter, I. Ruß, I. Medugorac, J. Tetens, and G. Thaller. 2016. Impact of the Friesian POLLED Mutation on Milk Production Traits in Holstein Friesian. p. 434. *In* Book of abstracts of the 67th Annual Meeting of the European Federation of Animal Science: Belfast, United Kingdom, 29 August - 2 September 2016. Wageningen Academic Publishers, Wageningen.

Legarra, A., and Z. G. Vitezica. 2015. Genetic evaluation with major genes and polygenic

inheritance when some animals are not genotyped using gene content multiple-trait BLUP. Genet Sel Evol 47(1):89. https://doi.org/10.1186/s12711-015-0165-x.

Lilian Johanna Gehrke, Doris Seichter, Ingolf Ruß, Dierck Segelke, Friedrich Reinhardt, and Georg Thaller. 2018. Is the Friesian polled mutation affecting performance traits in Holstein Friesian cattle? Proceedings of the World Congress on Genetics Applied to Livestock Production Electronic Poster Session - Genetic gain - Breeding Objectives and Economics of Selection Schemes 2:518.

Medugorac, I., A. Graf, C. Grohs, S. Rothammer, Y. Zagdsuren, E. Gladyr, N. Zinovieva, J. Barbieri, D. Seichter, I. Russ, A. Eggen, G. Hellenthal, G. Brem, H. Blum, S. Krebs, and A. Capitan. 2017. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nature genetics 49(3):470–475. https://doi.org/10.1038/ng.3775.

Medugorac, I., D. Seichter, A. Graf, I. Russ, H. Blum, K. H. Göpel, S. Rothammer, M. Förster, and S. Krebs. 2012. Bovine polledness—an autosomal dominant trait with allelic heterogeneity. PLOS ONE 7(6):e39477. https://doi.org/10.1371/journal.pone.0039477.

Mueller, M. L., J. B. Cole, N. K. Connors, D. J. Johnston, I. A. S. Randhawa, and A. L. van Eenennaam. 2021. Comparison of Gene Editing Versus Conventional Breeding to Introgress the POLLED Allele Into the Tropically Adapted Australian Beef Cattle Population. Frontiers in genetics 12:593154. https://doi.org/10.3389/fgene.2021.593154.

Mueller, M. L., J. B. Cole, T. S. Sonstegard, and A. L. van Eenennaam. 2019. Comparison of gene editing versus conventional breeding to introgress the POLLED allele into the US dairy cattle population. Journal of dairy science 102(5):4215–4226. https://doi.org/10.3168/jds.2018-15892.

Randhawa, I. A. S., M. R. McGowan, L. R. Porto-Neto, B. J. Hayes, and R. E. Lyons. 2021. Comparison of Genetic Merit for Weight and Meat Traits between the Polled and Horned Cattle in Multiple Beef Breeds. Animals 11(3):870. https://doi.org/10.3390/ani11030870.

Schuster, F., P. Aldag, A. Frenzel, K.-G. Hadeler, A. Lucas-Hahn, H. Niemann, and B. Petersen. 2020. CRISPR/Cas12a mediated knock-in of the Polled Celtic variant to produce a polled genotype in dairy cattle. Sci Rep 10(1):1–9. https://doi.org/10.1038/s41598-020-70531-y.

Utsunomiya, Y. T., R. B. P. Torrecilha, M. Milanesi, S. d. C. Paulan, A. T. H. Utsunomiya, and J. F. Garcia. 2019. Hornless Nellore cattle (Bos indicus) carrying a novel 110 kbp duplication variant of the polled locus. Animal Genetics 50(2):187–188. https://doi.org/10.1111/age.12764.

STATEMENT OF ORIGINALITY / ERKLÄRUNG

Ich erkläre:
Ich habe die vorgelegte Dissertation selbständig und ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der "Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis" niedergelegt sind, eingehalten.

_____ _____

Ort, Datum Carsten Scheper