Aus dem Institut für Tierzucht und Haustiergenetik

Professur für Tierzüchtung

der Justus-Liebig-Universität Gießen

Analysen und Evaluierungen im Rahmen der Entwicklung eines rassespezifischen SNP-Chips für das Deutsche Schwarzbunte Niederungsrind: Genomweite Assoziationen, genetische Parameter, genomische Vorhersagen und Selektionsschemata von Elitetieren

INAUGURAL-DISSERTATION

zur Erlangung des Doktorgrades (Dr. agr.)

im Fachbereich Agrarwissenschaften, Ökotrophologie und Umweltmanagement der Justus-Liebig-Universität Gießen

vorgelegt von

MANUEL JOHANNES WOLF

aus Unna

Gießen, Juni 2024

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. Gutachterin:	Prof. Dr. Gudrun A. Brockmann
Prüfer:	Prof. Dr. Horst Brandt
Prüfer:	Prof. Dr. Marc F. Schetelig
Vorsitzender:	Prof. Dr. Gesine Lühken

Tag der Disputation:
20.09.2024

Diese Arbeit entstand aus dem “DSN-SNP” Projekt zur Entwicklung eines SNP-Chips für die Zuchtwertschätzung und zum Diversitätsmanagement für das Deutsche Schwarzbunte Niederungsrind. Die Förderkennzeichen lauten 2818BM090 und 2818BM091. Die Förderung des Vorhabens erfolgte aus Mitteln des Bundesministeriums für Ernährung und Landwirtschaft (BMEL) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgte über die Bundesanstalt für Landwirtschaft und Ernährung (BLE).

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

ABKÜRZUNGSVERZEICHNIS

BLAD	Bovine Lekozytenadhäsionsdefizienz
BTA	*Bos taurus* Autosom
CTFS	Rastzeit (engl.: „calving to first service interval“)
DSN	Deutsches Schwarzbuntes Niederungsrind
DSN200K	Rassespezifischer Genotypisierungschip mit ~200.000 Markern
FPR	Fett-Eiweiß-Quotient (engl.: „fat-to-protein ratio”)
FV	Fleckvieh
GBLUP	Genomic Best Linear Unbiased Prediction
GEBV	Genomischer Zuchtwert (engl.: „genomic estimated breeding value“)
GWAS	Genomweite Assoziationsstudien (engl.: „genome-wide association studies”)
HF	Holstein Frisean
MHS	Malignes Hyperthermie Syndrom
NR56	Non-Return-Rate nach 56 Tagen
OGC	Optimum Genetic Contribution
Pro%	Proteinanteil in der Milch
QTL	Quantitative Trait Locus
SCS	Somatischer Zellscore (engl.: „somatic cell score”)
SNP	Einzelnukleotid-Polymorphismus (engl.: „single nucleotide polymorphism”)
STAT	Körpergröße (engl.: „stature”)
SV	Sequenzvarianten
WGS	Vollgenomsequenz (engl.: „whole genome sequence“)

ZUSAMMENFASSUNG

Ein nachhaltiger Erhalt bedrohter Nutztierrassen kann nur realisiert werden, wenn Rassen hinsichtlich ihrer Leistung und/oder Alleinstellungsmerkmale im Vergleich zu den konventionellen Rassen konkurrenzfähig bleiben. In den großen Populationen wie Holstein Friesian (HF) oder Fleckvieh (FV) hat die Einführung der genomischen Selektion zu einer weitgehenden Optimierung der Zuchtprogramme mit rasantem Zuchtfortschritt für Merkmale der konventionellen Leistungsprüfung beigetragen. Vom Aussterben bedrohte Rassen mit geringen Populationsgrößen, wie das Deutsche Schwarzbunte Niederungsrind (DSN) profitieren von diesen hocheffizienten Zuchtverfahren und vom aktuellen Entwicklungsfortschritt auf dem Gebiet der Schätzungen von genomischen Zuchtwerten (GEBVs; engl.: „genomic estimated breeding values“) trotz des offensichtlichen Potenzials bisher wenig. Basis für die GEBVs sind kommerzielle Genotypisierungs-Chips auf Basis von Einzelnukleotid-Polymorphismen (SNPs; engl.: „single nucleotide polymorphisms“), die speziell auf die populären Rassen „zugeschnitten“ sind. Bei der Entwicklung der kommerziellen SNP-Chip Technologie spielen die genomische Architektur der Hochleistungsrassen und die konventionellen Zuchtzielmerkmale (insbesondere Produktionsmerkmale) eine entscheidende Rolle.

Die Inhalte dieser Dissertation begleiten die Entwicklung und Evaluierung eines rassespezifischen SNP-Chips für das DSN. Dabei werden neben genomweiten Assoziationsstudien (GWAS; engl.: „genome-wide association studies“), genomische Methoden zur Zuchtwertschätzung und zum Diversitätsmanagement angewandt, um die genetische Vielfalt zu erhalten und die Produktivität dieser bedrohten Rasse zu verbessern. Die Grundlage für die Entwicklung ist der Aufbau einer aussagekräftigen Lernstichprobe mit genomischen Daten für die Rasse DSN. Für die Analysen sind Genotypinformationen einer repräsentativen Stichprobe der DSN-Population aus Vollgenomsequenzierung (WGS, engl. „whole-genome sequencing“), dem kommerziellen Illumina BovineSNP50 BeadChip (im Folgenden als 50K bezeichnet) und einem rassespezifischen 200K-Chip (im Folgenden als DSN200K bezeichnet) verfügbar, um den potenziellen Mehrwert des rassespezifischen DSN200K SNP-Chips zu bewerten. Neben der Anwendung von GWAS zur Aufdeckung der genomischen Architektur von Leistungsmerkmalen (Milchleistungsmerkmale), funktionalen Merkmalen (Fruchtbarkeit) und neuen Merkmalen der Tiergesundheit (Endoparasitenresistenz) werden auch GEBVs für die Rasse DSN geschätzt. Um das gesamte Potenzial der neugewonnenen Informationen auszuschöpfen, soll abschließend auf Basis der *Optimum Genetic Contribution* (OGC) Methode ein Monitoringverfahren mit Anpaarungsempfehlung getestet werden, um neben der Maximierung von Zuchtfortschritt und

somit einem Zugewinn an Konkurrenzfähigkeit auch die Diversität innerhalb der Rassen weitestgehend zu erhalten.

Die Arbeit ist in insgesamt 5 Kapitel unterteilt. Nach einer allgemeinen Einleitung folgt der Hauptteil in Kapitel 2-4, der zwei veröffentlichte und einen eingereichten wissenschaftlichen Artikel enthält. Abschließend werden die Inhalte der Arbeit in Kapitel 5 diskutiert.

Kapitel 1 betont die zentrale Rolle der Tierzucht in der globalen Landwirtschaft und die kontinuierliche Verbesserung der genetischen Qualität und Leistungsfähigkeit von Nutztieren durch selektive Zucht. Es wird auf die Fortschritte in den genetischen und genomischen Analysen der letzten Jahrzehnte eingegangen, die tiefere Einblicke in die genetische Architektur verschiedener Merkmale ermöglicht haben. Die DSN-Rasse wird näher charakterisiert. Hierbei handelt es sich um eine robuste Zweinutzungsrasse, die sowohl für die Milch- als auch für die Fleischproduktion verwendet wird. Historische und aktuelle Herausforderungen der DSN-Rasse, insbesondere infolge des Wettbewerbs mit leistungsstärkeren Rassen wie HF, werden dargestellt. Das Kapitel unterstreicht die Notwendigkeit, ein Gleichgewicht zwischen Produktivitätssteigerungen und der Erhaltung der genetischen Vielfalt zu finden. Es wird die potenzielle Bedeutung eines rassespezifischen SNP-Chips für DSN und die Anwendung von WGS Daten aufgezeigt, die wichtige genetische Informationen für die Zucht liefern könnten.

In **Kapitel 2** werden die Ergebnisse einer GWAS auf Basis von WGS zur Identifikation von Sequenzvarianten und Kandidatengenen, die mit Fruchtbarkeit, Gesundheit und Endoparasitenresistenzen beim DSN assoziiert sind, präsentiert. Für die Studie werden 304 sequenzierte DSN-Rinder verwendet, um die Genotypen von 1797 DSN auf WGS zu imputieren. Der finale Datensatz umfasst 11.413.456 Sequenzvarianten von 1886 Kühen. Untersucht werden die Merkmale Rastzeit (CTFS; engl.: „calving to first service interval“), die Non-Return-Rate nach 56 Tagen (NR56), der somatische Zellscore (SCS; engl.: „somatic cell score“), der Fett-Eiweiß-Quotient (FPR; engl.: „fat-to-protein ratio“) und drei vorkorrigierte Indikatormerkmale für Endoparasitenresistenz. Insgesamt werden 40 signifikante Sequenzvarianten (SVs) bestehend aus SNPs und INDELs identifiziert, die mit CTFS und NR56 assoziiert sind, sowie drei wichtige potenzielle Kandidatengene annotiert (*ARHGAP21*, *MARCH11* und *ZNF462*). Für SCS werden die meisten Assoziationen auf *‚Bos taurus* Autosom (BTA) 25 beobachtet. Die GWAS ergibt 61 signifikante SVs, ein Cluster von zehn Kandidatengenen auf BTA 13 und BTA 7 Pfade für FPR, einschließlich wichtiger Mediatoren der Milchfettsynthese. Die stärksten Assoziationen für Infektionen mit gastrointestinalen Nematoden und *Dictyocaulus viviparus* (großer Lungenwurm) werden auf BTA 8 bzw. BTA 24 gefunden. Für *Fasciola hepatica* (großer Leberegel) Infektionen liegen die stärksten assoziierten SVs auf BTA 4 und BTA 7. Insgesamt werden 200 Gene für das Merkmal

Endoparasitenresistenz identifiziert, die mit 16 Pfaden im Zusammenhang mit der Immunantwort des Wirts während der Infektion in Verbindung stehen.

Die genetisch-statistische Evaluierung des neuen DSN200K SNP-Chips, der speziell für das DSN entwickelt wurde, wird in **Kapitel 3** dargestellt. Die Evaluierung umfasst den Vergleich von GEBVs auf Basis verschiedener Marker-Panels: der kommerzielle 50K SNP-Chip von Illumina, der spezifisch für die DSN-Rasse entworfenen DSN200K SNP-Chip, sowie zufällig aus WGS-Daten generierte 200K Genotypendatensätze und WGS-Daten. Die Studie zielt darauf ab, die Effektivität dieser unterschiedlichen genetischen Datensätze zu bewerten, um die Genauigkeit der genomischen Vorhersagen in einer Population mit begrenzter Größe zu verbessern. Es werden 305-Tage-Produktionsmerkmale, FPR und SCS am ersten Laktations-Testtag nach dem Kalben, sowie Exterieurmerkmale in der Analyse berücksichtigt. Es werden gemischte Modelle zur Schätzung der genetischen Parameter und wiederholte zufällige Subsampling-Validierung zur Bestimmung der Genauigkeit der GEBVs angewandt. Die Ergebnisse zeigen, dass die Heritabilitätsschätzungen basierend auf den WGS-Daten tendenziell höher im Vergleich zu den SNP-Chip-basierten Analysen sind. Allerdings sind die Unterschiede marginal. Die Genauigkeiten der GEBVs sind bei den meisten untersuchten Merkmalen am höchsten, wenn WGS-Daten oder der spezifische DSN200K-Chip verwendet werden. Jedoch sind die Unterschiede in der Genauigkeit zwischen den verschiedenen Marker-Panels gering und statistisch nicht signifikant. In Anbetracht der marginalen Verbesserungen in den Genauigkeiten und den hohen Kosten, die durch die DSN200K- und WGS-Genotypisierung entstehen, ist die Anwendung des kommerziellen 50K Chips für die Schätzung der GEBVs gerechtfertigt. Trotzdem sind WGS und der DSN200K-Chip wertvoll für die Identifikation kausaler genetischer Mechanismen innerhalb der lokalen DSN-Population. Die Studie hebt die Bedeutung der Entwicklung spezifischer Werkzeuge für die genetische Bewertung lokaler Rinderrassen hervor und trägt zum besseren Verständnis der genetischen Struktur spezialisierter Rassen bei.

Die in **Kapitel 4** inkludierte Studie wird beim *Journal of Animal Breeding und Genetics* eingereicht und befindet sich zum jetzigen Zeitpunkt im Begutachtungsverfahren. In der Studie wird die OGC-Methode zur Selektion von DSN-Elitetieren angewandt. Die für die OGC-Methode benötigten GEBVs werden mit einer Single-Step-Methode für die Merkmale FPR, Proteinanteil (Pro%) und Körpergröße (STAT) geschätzt. Als Datengrundlage für die Schätzungen werden 50K Illumina und DSN200K SNP-Chips verwendet, um hier einen Vergleich bis zur finalen Anpaarungsplanung durchzuführen. Bei der Single-Step-Methode werden genomische und Pedigree-Informationen kombiniert. Die OGC-Methode zielt darauf ab, den genetischen Fortschritt zu maximieren und gleichzeitig die Zunahme der Inzucht zu minimieren, was für die Erhaltung der genetischen Vielfalt in kleinen Populationen entscheidend ist. Dabei werden verschiedene Zuchtstrategien bewertet, um den optimalen

genetischen Beitrag selektierter Eliteelterntiere zur nächsten Jungbullenkohorte zu bestimmen. Die Ergebnisse zeigen, dass die Datengrundlage mit 50K oder 200K bei den Schätzverfahren keinen statistisch relevanten Unterschied macht. Eine weniger strenge Inzuchtbeschränkung über einen längeren Zeitraum ist mit positiven Effekten auf die durchschnittlichen GEBVs für Pro%, FPR und STAT sowie einer abnehmenden Anzahl ausgewählter Bullenväter verbunden. Die jeweiligen Effekte sind marginal, wenn die Inzuchtbeschränkungen auf 6% oder mehr gelockert werden. Es wird empfohlen, die OGC-Anwendung beim DSN mit einer Inzuchtbeschränkung von 6% anzuwenden, was weiterhin zu Zuchtfortschritten bei Erhaltung der genetischen Diversität durch den Einsatz einer erhöhten Anzahl von Bullenvätern führt.

Abschließend werden die wichtigsten Ergebnisse dieser Arbeit in **Kapitel 5** zusammenhängend diskutiert. Hier wird insbesondere darauf eingegangen, dass die in der Arbeit angewandten genomischen Analysen und Methoden wichtige Werkzeuge für den Erhalt bedrohter Rassen sein können.

SUMMARY

Sustainable conservation of local breeds with small population size is supported through the identification of phenotypic and genetic characteristics, which are not common in commercial breeds. The implementation of genomic selection in large-scale populations such as Holstein Friesian (HF) or Fleckvieh (FV) implied substantial increases of genetic gain for conventional breeding traits. Theoretically, also in local breeds such as in the Deutsche Schwarzbunte Niederungsrind (DSN), there is potential through the development of novel genomic breeding schemes and selection based on genomic breeding values (GEBV), but practical applications are limited. The basis for GEBVs are commercial chips considering single nucleotide polymorphisms (SNPs), which are specifically designed for the popular breeds with large population size, i.e., taking into account the genomic architecture of these breeds and their breeding goal traits mostly reflection productivity. Consequently, distances in genetic gain between local breeds with HF or FV will increase in the future when basing breeding and selection on GEBV, implying a loss in economic competitiveness of the local breeds.

Consequently, the present thesis focuses on the utilization and evaluation of a recently developed breed-specific DSN SNP chip with a marker density of ~200K (DSN200K) in a breeding and selection context. This includes genome-wide association studies (GWAS), methods for breeding value estimations, and the management of genetic diversity, ultimately aiming for a balance between genetic gain and genetic variability in the long term. A prerequisite for all genomic breeding efforts in this regard is the implementation of a DSN training set, including animals with genotypes and phenotypes. For a sub-sample of animals, whole genome sequences (WGS), commercial 50K data from the Illumina BovineSNP50 BeadChip, and DSN-specific 200K data were available to study the additional benefits of the breed-specific DSN200K SNP chip in genomic applications. In addition to performing GWAS to infer specific patterns of genetic architectures for production traits and respective genomic predictions, the focus was on functional traits (fertility) and novel health traits (endoparasite resistance). Finally, the optimum genetic contribution (OGC) approach was applied, considering the genomic estimated breeding values (GEBV) and genetic relationships to monitor genetic diversity and to define mating strategies aimed at maximizing both long-term genetic gain and genetic diversity simultaneously.

The thesis comprises 5 chapters, including a general introduction and a main part in chapters 2-4. The main part is based on 2 already published scientific papers and one submitted manuscript for peer-review. Content of chapter 5 is an overall discussion considering the main aspects from the previous studies.

In **chapter 1**, the central role of animal breeding in a global agricultural context and the necessary continuous improvement of the genetic quality and productivity of livestock through selective breeding, is outlined. The general advances with regard to genetic and genomic analyses in the past decades contributing to deeper insights into genomic architectures of various traits, are described. In this chapter, the local and robust dual-purpose DSN breed utilised for milk and meat production, is characterised. Past and current challenges address the competition with high-yielding specialised breeds, indicating the necessity to consider genetic variability in the context of productivity increases. In this regard, a breed-specific DSN SNP chip and WGS might be valuable instruments to enhance breeding strategies. Such prospects are clearly addressed in this chapter.

In **chapter 2**, results from GWAS based on WGS are used for the detection of sequence variants and to annotate potential candidate genes being associated with fertility, health and endoparasite resistance in the DSN breed. Data basis for this study are 305 whole-genome sequenced DSN cattle, used, e.g., for the imputation of SNP-genotypes of further 1797 DSN cattle on WGS level. After imputation, the final dataset for comprised 11,413,456 structural variants (SV) from 1886 DSN cows. Traits considered in the WGS analyses were the interval from calving to first service (CTFS), non-return rate after 56 days (NR56), somatic cell score (SCS), fat-to-protein ratio (FPR) and 3 pre-corrected indicator traits for endoparasite resistance. 40 significant SVs were associated with CTFS and NR56, which were annotated with 3 potential candidate genes (*ARHGAP21, MARCH11* und *ZNF462*). For SCS, most significant associations were detected on *bos taurus* autosome (BTA) 25. GWAS contributed to the detection of 61 significant SVs for FPR, with one cluster including ten candidate genes on BTA 13 and BTA 7, and seven corresponding physiological pathways including mediators of milk fat synthesis. The most pronounced association signals for gastrointestinal nematodes and *Dictyocaulus viviparous* were inferred on BTA 8 and BTA 24. For *Fasciola hepatica*, most significant associations were located on BTA 4 and BTA 7. In total, 200 potential candidate genes were related with endoparasite resistance, contributing to 16 physiological pathways in the context of immune responses of the host.

The genetic-statistical evaluation of the newly developed DSN200K SNP chip in **chapter 3** focusses on GEBVs. Comparisons included the commercial 50K SNP chip from Illumina, the specifically developed DSN200K SNP chip, randomly generated (on the basis of WGS) 200K genotype data and WGS. Most important evaluation criterion was the accuracy of genomic predictions in sub-populations of limited population size. Considered traits were 305-d lactation yields for production traits, FPR and SCS from the first test-day after calving as well as linear conformation traits. Genetic parameters were estimated via linear-mixed model applications, and a repeated random sub-sampling validation were used to assess the prediction accuracy of GEBV. There was a general trend for slightly higher heritabilities based on WGS compared

to SNP chip applications, but differences were quite small. Accordingly, prediction accuracies were higher based on WGS and the DSN200K data compared to 50K data, but again, differences were not significant. The only marginal improvements and associated higher costs justify the ongoing utilization of the commercial 50K chip in genetic evaluations for DSN. Nevertheless, WGS and the DSN200K are very valuable with regard to the identification of causal genetic mechanisms in the endangered local DSN breed. The study in chapter 3 highlights the importance to develop specific novel tools for the genetic characterization of local breeds, contributing to a deeper understanding of genetic structures and mechanisms.

The study in **chapter 4** is under review in the *Journal of Animal Breeding und Genetics*. Main topic of this study is the application of OGC for the selection of DSN elite animals (bull sires and bull dams). Input data for OGC were GEBV for FPR, protein percentage (Pro%) and stature (STAT), which were estimated by applying single-step methodology. Again, evaluations focused on the comparison 50K versus DSN200K. The single-step methodology combined genomic relationships with pedigree data. Single-step GEBVs were used in OGC approaches, i.e., aiming on the maximization of genetic gain by restricting inbreeding, being a major criterion to conserve long-term genetic diversity. The OGC of elite animals to generate the next cohort of young bulls was determined in various breeding scenarios. Again, utilization of either the 50K or the DSN200K data basis had negligible effects on GEBVs as well as on the final elite animal selection plans. A relaxed inbreeding constraint implied increased genetic gain for Pro%, FPR and STAT, which was associated in a declining number of selected bull sires. However, the positive effects on genetic gain were quite small for inbreeding constraints beyond 6%. In consequence, we suggested an OGC application in DSN with a long-term inbreeding constraint of 6%, still contributing to genetic gain and increased genetic diversity through a larger number of utilized bull sires.

An overall discussion including the most relevant aspects from the previous chapters is content of **chapter 5**. The final statements emphasize the importance of genetic analyses based on novel genomic data are imperative for long-term conservation of local cattle breeds.

KAPITEL 1

Einleitung

Die Tierzucht spielt eine zentrale Rolle in der globalen Landwirtschaft und ist entscheidend für die Sicherung der Ernährung der wachsenden Weltbevölkerung. Durch die gezielte Auswahl und Paarung von Tieren mit gewünschten Eigenschaften strebt die Tierzucht danach, die genetische Qualität und Leistungsfähigkeit landwirtschaftlicher Nutztiere kontinuierlich zu verbessern. Diese Verbesserungen umfassen insbesondere eine erhöhte Produktivität und Produktqualität (Boichard & Brochard, 2012). In den letzten Jahrzehnten haben genetische und genomische Analysen revolutionäre Fortschritte in der Tierzucht ermöglicht, indem sie tiefe Einblicke in die genetische Architektur verschiedener Zuchtmerkmale liefern, wie z.B. Fellfarbe, Milchfettanteil (Hayes et al., 2010), Kalbeeigenschaften (Purfield et al., 2020) und Klimaadaption (Porto-Neto et al., 2014). Genomische Werkzeuge, wie z.B. Genotypisierungs-Chips, ermöglichen es, genetische Informationen schneller und kosteneffizienter als je zuvor zu sammeln, was die Selektion auf Basis präziser genetischer Marker erlaubt und damit die Zuchtfortschritte beschleunigt (Weller et al., 2017).

Trotz der soeben skizzierten Fortschritte stehen Tierzüchter[1] vor der Herausforderung, das Gleichgewicht zwischen Produktivitätssteigerungen und der Erhaltung der genetischen Vielfalt zu finden. Diese Balance ist entscheidend, um die Anfälligkeit von Zuchtpopulationen für Krankheiten zu minimieren und ihre Fähigkeit zur Anpassung an zukünftige Veränderungen zu sichern. In den letzten Jahrzehnten wurden Einschnitte in der genetischen Vielfalt, sowohl innerhalb als auch zwischen den Rassen, beobachtet. Die Verluste von lokalen Rassen tragen unmittelbar dazu bei (Biscarini et al., 2015). Nun sollen genetische Analysen es Züchtern ermöglichen, die genetische Diversität innerhalb und zwischen Rassen zu bewerten und zu erhalten, was für den langfristigen Erhalt und der Produktivität von Zuchtpopulationen entscheidend ist (Notter, 1999). Insbesondere bei lokal angepassten Rassen, wie dem Deutschen Schwarzbunten Niederungsrind (DSN), ist ein tiefes Verständnis der genetischen Basis wünschenswerter Merkmale essenziell, um zielgerichtete und nachhaltige Zuchtstrategien zu entwickeln.

Herkömmliche genutzte Genotypisierungs-Chips beinhalten einen großen Anteil von SNPs, die durch Sequenzierungen von ökonomisch relevanten Fleisch- und Milchrassen entdeckt wurden (Van Tassel et al. 2008). Obwohl aus historischer Sicht behaupten wird, dass eine enge Verwandtschaft zwischen DSN und HF vorliegt, konnten Naderi et al. (2020) zeigen, dass die durchschnittliche genetische Verwandtschaft zwischen den heutigen DSN und HF Populationen mit 0,1% recht gering ist. Es ist denkbar, dass mit z.B. Milchleistungsmerkmalen assoziierte Marker auf kommerziellen Genotypisierungs-Chips hauptsächlich für hochleistende Rassen relevant sind. Darüber hinaus sind neuartige Merkmale wie Robustheit,

[1] Aufgrund der besseren Lesbarkeit wird im Text das generische Maskulinum verwendet. Gemeint sind jedoch immer alle Geschlechter.

Krankheitsresistenzen und weitere funktionelle Merkmale, die lokal angepasste Rassen charakterisieren, bei der Entwicklung der Genotypisierungsarrays nicht berücksichtigt wurden. Das im Folgenden beschriebene Projekt beschäftigt sich mit der These, dass Informationen von rassespezifischen Genotypisierungschips für bedrohte Rassen einen Mehrwert in genetischen Analysen und Vorhersagen hervorbringen und somit zum Erhalt der Rassen beitragen können.

Im Rahmen des Projektes „Entwicklung eines SNP-Chips zur Zuchtwertschätzung und zum Diversitätsmanagement für das Deutsche Schwarzbunte Niederungsrind“ mit dem Akronym „DSN-SNP“, wurde ein der Rasse DSN angepasster Custom-SNP-Chip-Array für den Einsatz in genomischen Zuchtwertschätzungen sowie weitergehende Anwendungen zum Diversitätsmanagement entwickelt. Das Vorhaben soll einen Modellcharakter für die züchterische Weiterentwicklung und Sicherung der genetischen Diversität in bedrohten Rassen haben. Dabei werden modernste Verfahren der Molekulargenetik, Bioinformatik und Zuchtwertschätzung angewandt. Die Innovation besteht darin, auf Basis von engmaschigen genomischer Informationen, kombiniert mit modernsten genetisch-statistischen Verfahren der Bioinformatik, Zielvorstellungen zur züchterischen Rasseverbesserung zu realisieren, ohne auf Fremdgensequenzen bzw. genmanipulative Maßnahmen zurückgreifen oder ausweichen zu müssen.

Nach einer Übersicht zur Rasse DSN, einer Einleitung zur SNP-Chip-Genotypisierung, Zuchtwertschätzung in Rassen mit geringen Populationsgrößen und Methoden zum Diversitätsmanagement folgen drei Publikationen, die den Hauptteil dieser Arbeit bilden und maßgeblich zum erfolgreichen Abschluss des oben beschriebenen Projektes beigetragen haben.

1.1 Die Rasse: Deutsches Schwarzbuntes Niederungsrind

Die robuste Zweinutzungsrasse DSN wird zur Milch- und Fleischproduktion genutzt. Historisch stellt sie eine bedeutende Rasse in der europäischen Landwirtschaft dar, dessen Ursprünge und Entwicklung eng mit der kulturellen und ökonomischen Geschichte des Kontinents verwoben sind (Brade & Brade, 2013). Dieser Abschnitt gibt einen Überblick über die Herkunft, charakteristische Merkmale, die Entwicklung der Populationszahlen, den Gefährdungsstatus sowie die aktuelle Forschungslage zu dieser Rasse.

Ihren Ursprung hat die Rasse DSN in den Marsch- und Niederungsgebieten der Nordseeregion. Im 18. Jahrhundert gab es in den dänischen, holländischen und deutschen Küstengebieten viele schwarzbunte Landschläge mit relativ hohen Milchleistungen, welche den Ursprung der heutigen Rasse DSN bildeten (Mügge, 1999). Ab Mitte des 19. Jahrhunderts wurden viele Rinder mit englischen Shorthornrindern gekreuzt, wodurch die Milchleistung der schwarzbunten Kühe zurückging. Das hatte zur Folge, dass wieder zur Reinzucht der

schwarzbunten Rinder zurückgefunden wurde. So entstand das erste Zuchtbuch der Rasse im Jahr 1876 (Haller, 2015). DSN wurde für seine Robustheit und Anpassungsfähigkeit an die lokal vorherrschenden harschen klimatischen Bedingungen geschätzt. Neben ihrer Produktivität bei der Milcherzeugung weisen sie bis heute eine gute Fleischqualität auf, was sie zu einer wertvollen Zweinutzungsrasse macht. Im 19. Jahrhundert wurden schwarzbunte Tiere der Nordseeregion in die USA exportiert, wo sie auf eine hohe Milchleistung selektiert wurden und schon bald die Rasse HF bildeten. Die heutige geringe Populationsgröße der Rasse DSN ist auf eine Verdrängungskreuzung durch HF in den 1960er Jahren zurückzuführen. Damals wurden stark HF-Bullen aus Amerika eingesetzt, um hohe Milchleistungen zu erreichen.
Die ursprüngliche Zuchtrichtung der Rasse DSN wurde in der damaligen DDR als Genreserve gehalten. Dort wurden DSN-Kühe mit dänischen Jersey-Bullen gekreuzt. Mit deren F1-Generation wurde durch Kreuzung mit der Rasse HF das Schwarzbunte Milchrind gezüchtet, dass als dominierende Milchrasse in der DDR etabliert wurde (Jaeger, 2018).

Charakteristisch für DSN sind die erhöhten Milchfettanteile, hohe Dauerleistungen, gute Anpassungsfähigkeit an die Haltungsumwelt, Fruchtbarkeit und ein ruhiges Temperament. Auch bei reiner Grundfutterzufuhr weisen DSN gute Mastleistungen auf und sind daher besonders für Grünlandstandorte geeignet (Jaeger, 2018). Das äußere Erscheinungsbild der Rasse ist durch einen mittleren Rahmen und eine mittlere Bemuskelung geprägt (Abbildung 1). Weibliche Tiere werden bei einem durchschnittlichen Gewicht von 550 bis 650 kg im Schnitt 1,30 bis 1,42 m groß. Mit einer durchschnittlichen Größe von 1,50 bis 1,62 m und 1000 bis 1150 kg sind die männlichen Tiere etwas größer und deutlich schwerer (Rahmann, 2011). Laut Rassedefinition sind DSN regelmäßig schwarz-weiß gescheckt und genetisch horntragend. Schwarze Zeichnungen finden sich an Kopf, Hals, Brust und Becken. Weiße Scheckung haben die Tiere an Bauch, Euter und an den Beinen bis zum Karpal- bzw. Sprunggelenk. Als Zuchtziel soll die Doppelnutzung aufrechterhalten werden, jedoch soll das genetische Milchleistungspotential auf über 7000 kg mit einem Fettgehalt von 4,30% und Eiweißgehalt 3,70% gesteigert werden (ZBH, 2019). Außerdem sollen Zuchttiere aufgrund der gemeinsamen Zuchtgeschichte mit HF maximal einen HF-Anteil von 10% aufweisen.

Abbildung 1: Deutsches Schwarzbuntes Niederungsrind (Foto: L. Aufmhof)

Trotz ihrer historischen Bedeutung und genetischen Anpassungsfähigkeit, die DSN für verschiedene Umweltbedingungen geeignet macht, steht die Rasse vor Herausforderungen. Ein starker Rückgang der Populationszahlen, verursacht durch Konkurrenz mit international verbreiteten Hochleistungsrassen wie HF, hat zu einem kritischen Gefährdungsstatus geführt. Aus dieser Entwicklung resultiert eine zunehmende Homogenisierung der genetischen Vielfalt in der Milchviehwirtschaft und stellte die langfristige Überlebensfähigkeit lokaler Rassen in Frage. Im Jahr 2021 umfasste die lebende DSN-Population 2514 weibliche und 27 männliche Zuchttiere. Das entspricht einer effektiven Populationsgröße (N_e) von 106,85. Somit handelt es sich bei der DSN-Population nach der Definition der Bundesanstalt für Landwirtschaft und Ernährung um eine Erhaltungspopulation ($N_e < 200$).

Um den Bestand der Rasse DSN zu sichern, wurde 1989 der „Verein zum Erhalt und Förderung des alten Deutschen Schwarzbunten Niederungsrind“ gegründet. Beim Friedrich-Loeffler-Institut in Mariensee gibt es eine Kryoreserve bestehend aus 300 Embryonen der Rasse DSN. Diese dient dazu, ein Inzuchtmanagement zu ermöglichen und die genetische Diversität zu bewahren (Jaeger, 2018).

Jüngste Forschungen sind zum größten Teil dem oben benannten „DSN-SNP“ Projekt zuzuordnen. Dazu gehören ebenfalls die Publikationen dieser Arbeit. Die Forschenden haben es sich unter anderem zum Ziel gemacht, die genetische Basis verschiedener Merkmale in der Rasse DSN zu verstehen. Dazu wurden insbesondere Analysen zu den verschiedenen Milchleistungsmerkmalen (Korkuć et al., 2021; Korkuć et al., 2023), Endoparasitenresistenzen (May et al., 2019, Wolf et al. 2021), Furchtbarkeitsmerkmalen und Gesundheitsindikatoren (Wolf et al., 2021) sowie Mastitisresistenz (Meier et al., 2020) durchgeführt. Des Weiteren wurde auf bioinformatischer Ebene anhand einer DSN Lernstichprobe zwei

Imputationsverfahren verglichen (Korkuć et al., 2019), ein rassespezifischer DSN SNP-Chip designt und anschließend erprobt (Neumann et al., 2021; Wolf et al., 2023). Es wurden Untersuchungen zur genetischen Diversität und Verwandtschaft durchgeführt (Neumann et al., 2023a), und die OGC-Methode wurde angewendet, um eine Anpaarungsstrategie vorzuschlagen, die eine Steigerung der Zuchtwerte bei gleichzeitiger Überwachung des Verwandtschaftsgrades in der Nachkommengeneration gewährleistet (Wolf et al., 2024). Auf Basis der vorliegenden WGS Daten wurden unmapped short reads entdeckt, die auf potenzielle Infektionserreger beim DSN hinweisen (Neumann et al., 2023b) und der Nutzen von Genotyp- und Sequenzdaten für die genetische Verbesserung bei Rassen mit kleiner Populationsgröße am Beispiel von DSN herausgearbeitet (May et al., 2023). Zusammenfassend wurde in den letzten Jahren durch die Einsätze von Genotypisierungs-Chips (50K, 200K) und Whole-Genome-Sequencing (WGS) Ansätzen, Anstrengungen unternommen, die genetischen Ressourcen des DSN zu charakterisieren, Strategien für seinen Erhalt und seine nachhaltige Nutzung zu entwickeln.

1.2 Entwicklung der Genotypisierung

Die Genotypisierungstechniken haben sich in den letzten Jahrzehnten erheblich weiterentwickelt und ermöglichen zunehmend detailliertere Einblicke in die genetische Architektur von Nutztieren. Im Folgenden wird die Entwicklung von Mikrosatelliten über SNP-Chip-Genotypisierung bis hin zu WGS dargestellt, einschließlich ihrer wissenschaftlichen und praktischen Implikationen.

Mikrosatelliten-Analyse (1990er Jahre):

Mikrosatelliten, auch als Short Tandem Repeats bekannt, sind kurze, wiederholte DNA-Sequenzen in den Genomen vieler Arten. Aufgrund ihrer hohen Variabilität wurden sie in den 1990er Jahren populär für genetische Studien, einschließlich der genetischen Kartierung, Verwandtschaftsanalysen und der Populationsgenetik. Mikrosatelliten bieten eine hohe Auflösung aufgrund ihrer Vielfältigkeit, sind jedoch durch den Bedarf an spezifischen Primern für jede Sequenz und die potenzielle Komplexität der Datenanalyse begrenzt (Weber & May, 1989).

SNP-Chip-Genotypisierung (2000er Jahre):

Mit der Einführung der SNP-Chip Technologie Anfang der 2000er Jahre wurde die genetische Forschung revolutioniert. SNPs sind einzelne Basenpaaränderungen in der DNA, die sich als äußerst funktional für die Untersuchung genetischer Variationen und Assoziationsstudien erwiesen haben. SNP-Chips ermöglichen die gleichzeitige Analyse tausender SNPs, wodurch genetische Studien auf eine neue Stufe der Effizienz und Genauigkeit gehoben wurden. Diese Technologie hat insbesondere im Rahmen von genomweiten Assoziationsstudien (GWAS) und in der praktischen Nutztierzucht (genomische Selektion) umfangreiche Anwendungen

gefunden (Schork et al., 2000; Hayes & Goddard, 2010). SNP-Chip-Marker sind für viele Nutztierarten verfügbar und werden in verschiedenen Markerdichten angeboten. Für Rinder werden z.B. Produkte von der Firma Illumina mit den Markerdichten 7000 (BovineLD) 54000 (BovineSNP50) und 777000 (BovineHD) angeboten.

Vollgenomsequenzierung (2010er Jahre bis heute):

Die seit den 2010er Jahren zugängliche Hochdurchsatzsequenzierung (Next Generation Sequencing), ermöglichte es Forschenden die komplette DNA-Sequenz eines Organismus zu lesen. Im Vergleich zu früheren Methoden, bei denen nur ausgewählte Abschnitte des Genoms analysiert wurden, bietet WGS eine umfassendere Basis für Entscheidungen und die Möglichkeit, auch komplexe Merkmale zu verbessern, die von mehreren Genen beeinflusst werden (Daetwyler et al., 2014). Der detaillierte Blick in genomische Informationen ermöglichen genaue Analysen der genetischen Diversität von Populationen. Durch die Identifizierung von genetisch einzigartigen Individuen oder Linien können Züchter gezielt Strategien zur Erhaltung der genetischen Vielfalt implementieren, was für die langfristige Anpassungsfähigkeit und Gesundheit der Populationen von entscheidender Bedeutung ist (FAO, 2015).

Zukunftsperspektiven und Herausforderungen:

Die Zukunft der Genotypisierung bei Nutztieren verspricht durch die Entwicklung neuer Technologien wie *CRISPR-Cas9* für die gezielte Genom-Editierung und Einzelzell-Sequenzierung weitere Fortschritte. Diese Entwicklungen könnten zu einer noch präziseren Zuchtstrategie führen und neue Möglichkeiten zur Verbesserung der Tiergesundheit und Leistung eröffnen. Gleichzeitig werfen sie wichtige ethische Fragen auf, betreffen die gesellschaftliche Akzeptanz gentechnisch veränderter Organismen (GVO) und den Schutz genetischer Ressourcen (Tabor et al., 2021; Müller et al., 2022).

1.3 Fortschritte in der Tierzucht

Die Entwicklungen in der Technik der Genotypisierung, wie etwa die Einführung von SNP-Chip-Technologien und Hochdurchsatzsequenzierung, führten zu einer Revolution in den tierzüchterischen Praktiken. Durch den Zugang zu präzisen genetischen Informationen konnten Züchter selektivere Zuchtentscheidungen treffen, die auf spezifischen genetischen Merkmalen und Leistungskennzahlen basieren. Dies ermöglichte eine gezielte Verbesserung der Zuchtprogramme, einschließlich der Zucht auf Krankheitsresistenz, Wachstumsrate, Futtereffizienz und andere wirtschaftlich wichtige Eigenschaften. Diese Fortschritte in der Zuchttechnologie haben nicht nur zu einer Steigerung der Effizienz und Produktivität in der Nutztierhaltung geführt, sondern auch zu einer verbesserten Tiergesundheit und Wohlbefinden beigetragen, indem genetische Prädispositionen für Krankheiten minimiert wurden.

Frühe Anfänge und konventionelle Zucht (bis 1980er Jahre):

Die traditionelle Nutztierzucht, die bis in die 1980er Jahre praktiziert wurde, basierte vorrangig auf der Auswahl von Tieren anhand phänotypischer Merkmale und Stammbauminformationen. Diese Methoden, obwohl grundlegend, waren zeitaufwendig und oft mit Ungenauigkeiten behaftet, da viele bedeutende Eigenschaften phänotypisch nicht offensichtlich sind und der direkte Einfluss des Genotyps auf den Phänotyp nicht erfasst werden konnte (Hill, 2014).

Aufkommen molekularer Marker (1980er-1990er Jahre):

Mit der Entdeckung molekularer Marker in den 1980er und 1990er Jahren begann eine neue Ära in der Nutztiergenetik. Diese Marker ermöglichten es Forschenden, DNA-Sequenzen zu identifizieren, die mit bestimmten phänotypischen Merkmalen korrelierten, was eine präzisere Selektion und Zucht von Tieren mit gewünschten genetischen Eigenschaften erlaubte. Ein Schlüsselwerkzeug war die Entwicklung der Polymerasen-Kettenreaktion in den 1980er Jahren, die eine rasche Vervielfältigung spezifischer DNA-Abschnitte und somit vereinfachte genetische Tests ermöglichte. Mit der sogenannten markergestützten Selektion wurde erfolgreich gegen problematische Allele gezüchtet, die etwa mit dem malignen Hyperthermie Syndrom (MHS) bei Schweinen (Brem & Brenig 1992) oder der bovinen Leukozytenadhäsionsdefizienz (BLAD) bei Rindern (Zhang et al. 2012) assoziiert sind. Heute werden diese Allele oftmals standardmäßig in den genomischen Zuchtprogrammen berücksichtigt.

Genomische Selektion (2000er Jahre bis heute):

Die Einführung der genomischen Selektion Anfang der 2000er Jahre markierte einen Wendepunkt in der Nutztiergenetik (Meuwissen et al. 2001). Durch die Analyse tausender genetischer Marker über das gesamte Genom hinweg konnte der genetische Wert eines Tieres für komplexe Merkmale präzise vorhergesagt werden. Diese Innovation stützte sich auf den Nachweis, dass viele Eigenschaften durch die kumulative Wirkung zahlreicher Gene im gesamten Genom beeinflusst werden (Hayes et al., 2009). Durch die genomische Selektion konnte eine Verkürzung des Generationsintervalls vorgenommen werden, da die Wartebullenhaltung von Bullen zur Nachkommenprüfung eliminiert wurde (Bengtsson et al., 2020). Die Fortschritte in den statistischen Methodiken der genomischen Selektion führen zu immer höheren Vorhersagegenauigkeiten von bis zu 50% (Knol, et al., 2016; Wolc et al., 2016). So ist es möglich, basierend auf Genomdaten und Beziehungsinformationen zwischen Individuen, fehlende Informationen ohne Vorliegen von eigenen Leistungsprüfungsergebnissen, zu schätzen.

1.4 Genomische Methoden und Anwendungen in vom Aussterben bedrohten Nutztierrassen

Die Anwendung genomischer Methoden in vom Aussterben bedrohten Nutztierrassen hat sich als ein entscheidender Schritt in Richtung der Erhaltung ihrer genetischen Vielfalt und der Verbesserung ihrer Lebensfähigkeit erwiesen. Diese Techniken bieten tiefe Einblicke in die genetische Struktur und Variation innerhalb dieser Populationen und ermöglichen eine präzisere Selektion und Zucht. Die Forschung in diesem Bereich hat nicht nur zum besseren Verständnis der genetischen Grundlagen wichtiger Merkmale geführt, sondern auch Strategien für den Erhalt und die nachhaltige Nutzung genetischer Ressourcen bedrohter Rassen bereitgestellt (Theissinger et al., 2023).

Neben den beschriebenen molekulargenetischen Fortschritten (wie Mikrosatelliten, SNP-Genotypisierung und WGS), die es ermöglichen, immer kosteneffizienter und detaillierter genomische Informationen zu erhalten, spielen auch die Weiterentwicklungen in den statistisch-genetischen Methoden eine entscheidende Rolle. Diese Methoden sind erforderlich, um die wachsende Datenmenge zu verarbeiten, präzisere Strukturen im Genom aufzudecken und schließlich die Selektion geeigneter Zuchttiere zu unterstützen.

Im Folgenden wird auf die Methoden der genomweiten Assoziation von Markern und Merkmalen, genomische Zuchtwertschätzung und die OGC-Methode eingegangen. Dabei wird Bezug zu dem Sachstand bei der Rasse DSN genommen.

1.4.1 *Genomweite Assoziationsstudien*

Die Idee genomweit Gene und Quantitative Trait Loci (QTLs) zu identifizieren, die komplexe Merkmale beeinflussen, wird bereits auf die 1990er Jahre datiert. Das Kartieren von QTLs basierte zu der Zeit auf Mikrosatteliteninformationen (Lipkin et al., 1998). Mit der Entwicklung von SNP-Chip Genotypisierung war es möglich, Zusammenhänge von tausenden Markern mit bestimmten Merkmalsausprägungen gleichzeitig zu untersuchen. Durch eine genomweite Abdeckung mit SNP-Markern haben sich GWAS für die Kartierung von QTLs ökonomisch wichtiger Merkmale durchgesetzt. GWAS haben in der Nutztierzucht erhebliche Fortschritte erbracht, indem sie tiefere Einblicke in die genetischen Grundlagen von Produktionsmerkmalen, Krankheitsresistenzen und anderen wichtigen Eigenschaften gaben. Hier sind einige bedeutende Meilensteine in der Anwendung von GWAS in der Nutztierzucht aufgeführt:

- Nachdem die Methode in der Humangenetik etabliert war, wurde sie in den 2000er Jahren in der Schweine-, Rinder- und Hühnerzucht angewandt, um wichtige Produktionsmerkmale wie Milchleistung, Fleischqualität und Wachstumsraten zu untersuchen (Hayes et al., 2009).

- Für wichtige landwirtschaftliche Tierarten wurden spezielle SNP-Chips entwickelt. Beispielsweise wurde der *Illumina BovineSNP50 BeadChip* für Rinder im Jahr 2007 eingeführt. Solche Chips ermöglichen eine effizientere und kostengünstigere Genotypisierung für GWAS (Matukumalli et al., 2009). Neben Produktionsmerkmalen wurden Studien zur genetischen Aufklärung von Krankheitsresistenzen und funktionellen Merkmalen wie Fruchtbarkeit durchgeführt (Sahana et al., 2023).
- Ergebnisse aus den GWAS haben zur Weiterentwicklung der genomischen Selektion beigetragen, bei der genomische Informationen direkt in die Zuchtwertschätzung einfließen. Dies führte wiederum zu schnelleren genetischen Fortschritten und präziseren Vorhersagen der Zuchtwerte (Eggen, 2012).
- Mit der Verringerung der Kosten für genetische Analysen wurden GWAS auch auf Rassen mit geringerer Populationsgrößere und weniger wirtschaftlich bedeutende Tierarten sowie auf seltene und einheimische Rassen ausgeweitet, um deren genetische Ressourcen zu konservieren und zu nutzen (Kijas et al., 2012).
- Die Kombination von GWAS mit anderen omics-Daten, wie Transkriptomik und Epigenomik trug dazu bei, die funktionellen Mechanismen hinter GWAS-Signalen zu verstehen und die biologische Interpretation der Ergebnisse zu verbessern (Lindholm-Perry et al., 2017).

Die oben aufgezeigten Entwicklungen und Anwendungen von GWAS zeigen, wie vielfältig die Einsatzmöglichkeit der Methodik über verschiedene Tierarten hinweg ist.

Die Anwendung von GWAS auf lokale Nutztierrassen bietet viele Möglichkeiten und Chancen, bringt jedoch auch spezifische Einschränkungen mit sich. Durch die Anwendung von GWAS können Gene und Marker identifiziert werden, die mit wichtigen Merkmalen, wie Krankheitsresistenz, Fruchtbarkeit oder Produktivität verbunden sind. Somit können auch einzigartige Charakteristika dieser Rassen untersucht werden. Mittels GWAS können wertvolle genetische Informationen gesammelt werden, die für Erhaltungsprogramme genutzt werden können, um die genetische Diversität zu überwachen und damit die Überlebensfähigkeit der Rasse zu sichern (Kijas et al., 2012). Viele dieser Rassen haben sich an spezifische lokale Umweltbedingungen angepasst. GWAS helfen, diese genetischen Anpassungen zu verstehen um somit gezielt Rassen zu identifizieren, die besonders für bestimmte Umweltbedingungen geeignet sind. Die Anwendungen von GWAS in lokalen Rassen implizieren jedoch auch einige Herausforderungen. So haben lokale Rassen oft begrenzte Populationsgrößen, was zu geringer genetischer Variation und geringerer statistischer Power in GWAS führt. Zudem sind kleine Populationen anfälliger für Inzucht und genetische Drift, was zu einer erhöhten Frequenz unerwünschter Allele führen kann. Dies kann die korrekte Interpretation von GWAS-Ergebnissen verkomplizieren, da Marker, die durch Inzucht entstandene Allelanhäufungen

anzeigen, fälschlicherweise mit untersuchten Merkmalen assoziiert werden könnten (Charlesworth & Willis 2009; Keller & Waller 2002). Unterschiedliche Rassen verfügen nicht über die gleiche genomische Infrastruktur (z.B. Referenzgenome oder spezifische SNP-Chips) wie größere kommerziell eingesetzte Rassen, was die Durchführung und Genauigkeit von GWAS einschränken kann. Daher sind GWAS mit WGS in Rassen mit kleiner Populationsgröße eine Möglichkeit, Abschnitte im Genom mit Merkmalen assoziieren, die durch die geringere Abdeckung kommerzieller Chips unentdeckt bleiben.

1.4.2 *Genomischen Zuchtwertschätzung*

Um GEBVs zu erhalten und um die genetische Leistung eines Individuums vorherzusagen, wird die gleichmäßige Verteilung genetischer Marker über das gesamte Genom genutzt. Dieser Prozess hat die Zuchtpraxis revolutioniert, da er eine präzisere und schnellere Bewertung genetischer Potenziale ermöglicht, insbesondere bevor die Tiere eigene Leistungsdaten vorweisen können. Meuwissen et al. (2001) erläuterten das Grundprinzip der genomischen Selektion und erklärten das Konzept, wie dichte Markerinformationen zur Vorhersage des genetischen Zuchtwerts genutzt werden können. Für die genomische Zuchtwertschätzung werden genetische Markerdaten und phänotypische Daten zu verschiedenen Merkmalen von Tieren einer ausreichend großen Referenzpopulation benötigt. Auf Basis der genomischen Daten wir eine Beziehungsmatrix erstellt, die die genetische Verwandtschaft zwischen allen Individuen im Datensatz darstellt. Nun können mithilfe statistischer Modelle die Einzeleffekte der genetischen Marker auf ausgewählte Merkmale geschätzt werden. Hier liegen verschiedene Methoden zugrunde, wie etwa der Ansatz einer Genomic Best Linear Unbiased Prediction (GBLUP) oder Bayes'sche Ansätze (VanRaden, 2008). Auf Basis der geschätzten SNP-Effekte und Genotypinformationen können GEBVs und somit das genetische Potenzial von Individuen ausgewiesen werden. Der methodische Ansatz der Imputation von Genotypdaten hilft dabei, Kosten der Genotypisierung zu senken und gleichzeitig eine hohe Genauigkeit der genomischen Vorhersage beizubehalten (Goddard & Hayes, 2007). Mittels der Berücksichtigung von genomischen Informationen wird die Genauigkeit der Zuchtwerte im Vergleich zu traditionellen Methoden, die sich auf Pedigrees und phänotypische Daten stützen, erhöht. Eine Selektionsentscheidung zur Auswahl geeigneter Zuchttiere kann bereits in einem sehr jungen Alter gefällt werden (Wolc et al., 2011). Zudem können durch genaue Kenntnisse zur genetischen Verwandtschaft über Markerdaten Inzuchtdepressionen effektiver vermieden werden. Dies trägt zur Aufrechterhaltung der genetischen Vielfalt bei (Rodríguez-Ramilo et al., 2015).

Trotz kontinuierlichen Fortschritten im dynamischen Feld der Zuchtwertschätzung gibt es weiterhin Herausforderungen. Die Genotypisierung von Tieren ist mit Kosten verbunden, die je nach Umfang der genetischen Tests erheblich sein können. Zudem müssen die angewandten Phänotyp- und Genotypdaten von hoher Qualität sein, um genaue Schätzungen

zu erreichen. Zudem ist die Vorhersagegenauigkeit bei komplexen Merkmalen, die durch viele Gene mit kleinen Effekten beeinflusst werden, eingeschränkt (Dekkers, 2021).

Vom Aussterben bedrohte Rassen mit geringer Populationsgröße wie das DSN profitieren derzeit nur gering vom Entwicklungsfortschritt bei der genomischen Zuchtwertschätzung, die in den zahlenmäßig großen Populationen wie HF und FV in der Routine angewandt wird. Basis dafür sind kommerzielle SNP-Chips, die speziell auf die zahlenmäßig großen Populationen „zugeschnitten“ sind. Bei dieser kommerziellen SNP-Chip Entwicklung spielen die genomische Architektur der Hochleistungsrassen und die konventionellen Zuchtzielmerkmale (insbesondere Produktionsmerkmale) eine entscheidende Rolle. Während in den Populationen wie HF und FV durch genomische Selektionsschemata ein beschleunigter Zuchtfortschritt beobachtet werden kann, bleiben lokale Rassen mit geringer Populationsgröße im Hinblick auf ihre Wettbewerbsfähigkeit im Vergleich zu den größeren Rassen züchterisch zunehmend weiter zurück. Darüber hinaus sind für lokale Rassen das Inzuchtmanagement bzw. die Erhaltung der genetischen Diversität von großer Bedeutung. Züchterische Strategien müssen somit zwei zentrale Bedingungen gleichzeitig berücksichtigen: Die Maximierung von Zuchtfortschritt bei gleichzeitiger Restriktion für zukünftige verwandtschaftliche Beziehungen bzw. Maximierung von Kennzahlen der genetischen Diversität. Die genaueste Beurteilung der Diversität ist möglich (insbesondere bei unzureichender Tiefe der Abstammungsdateien), wenn hierzu genomische Markerdaten verwendet werden. Dabei erscheint der Heterozygotiegrad basierend auf SNP-Markerdaten die aktuell beste Maßzahl, um genetische Variabilität innerhalb von Populationen zu beschreiben. Allerdings kann auch dieser Heterozygotiegrad nur dann genau abgeleitet werden, wenn WGS Informationen der Tiere zur Verfügung stehen (siehe u.a. Malomane et al., 2018)

Die genomische Zuchtwertschätzung kann in vielerlei Hinsicht beim Erhalt der vom Aussterben bedrohter Rassen nützlich sein. Aufgrund der geringen Anzahl an Individuen in diesen Populationen sind die zur Verfügung stehenden phänotypischen Daten beschränkt. Durch das Einbeziehen von genomischen Informationen kann die Genauigkeit der Zuchtwertschätzungen verbessert werden. Des Weiteren ist Inzucht ein Problem in vielen lokalen Populationen. Genomische Informationen können dazu beitragen, genetisch ähnliche Anpaarungen zu vermeiden und so Inzucht zu minimieren. Durch gezielte Zuchtstrategien können wertvolle genetische Varianten, die möglicherweise mit Resistenz gegen Krankheiten oder besondere Anpassungen verbunden sind, innerhalb einer Rasse erhalten bleiben (Mészáros et al., 2015).

Für die erfolgreiche Anwendung genomischer Zuchtwertschätzung in Rassen mit geringer Populationsgröße in Europa gibt es unterschiedliche Beispiele. In Skandinavien beispielsweise wurden genomische Zuchtwertschätzungen für die finnische Ayrshire-Rasse

implementiert, um die Genauigkeit der Zuchtwertschätzungen zu verbessern und gleichzeitig die genetische Vielfalt zu erhalten. Diese Bemühungen führten dazu, den langfristigen Erhalt der lokalen Rassen zu sichern (Sarviaho et al., 2024). Für das Schweizer Original Braunvieh wurde die genomische Zuchtwertschätzung eingesetzt, um die genetische Basis dieser alten und numerisch kleinen Rasse zu verstärken. Genomische Daten helfen, wertvolle Zuchtlinien zu identifizieren und zielgerichtete Zuchtprogramme zu entwickeln, um wertvolle Eigenschaften wie Robustheit und Milchleistung zu verbessern (Häfliger et al., 2021). In Spanien wurden Projekte durchgeführt, um genomische Zuchtwertschätzungen bei lokalen Schaf- und Ziegenpopulationen zu testen. Diese Studien konzentrieren sich auf die Identifizierung von genetischen Markern für Krankheitsresistenz und andere wichtige Produktionsmerkmale, um die Attraktivität dieser Rassen für lokale Züchter zu erhöhen und ihren Erhalt zu unterstützen (Rupp et al., 2016).

1.4.3 *Optimum Genetic Contribution*

Die OGC Methode zielt darauf ab, die genetischen Fortschritte innerhalb einer Population zu maximieren und gleichzeitig die Zunahme der Inzucht zu minimieren (Meuwissen 1997). Für Zuchtprogramme, in denen die Erhaltung der genetischen Vielfalt eine hohe Priorität hat, ist diese Methode besonders relevant. Das Grundprinzip von OGC ist die Anzahl der Nachkommen, die jedes Zuchttier beisteuern sollte, zu optimieren, um langfristig sowohl den Zuchtfortschritt zu maximieren als auch das Inzuchtniveau zu kontrollieren. Hierbei werden die Zuchtwerte der zur Anpaarung vorausgewählten Tiere und ihre genetischen Beziehungen zueinander berücksichtigt, um anschließend die genetischen Beiträge jedes selektierten Elterntieres zur Erstellung der nächsten Generation zu berechnen. Die OGC-Methode nutzt oft lineare Programmierung oder Simulationsalgorithmen, um die optimalen Beiträge jedes Tieres zur nächsten Generation zu bestimmen. Diese Beiträge werden so festgelegt, dass der genetische Fortschritt gefördert wird, während gleichzeitig die Inzuchtakkumulation über Generationen hinweg kontrolliert wird (König & Simianer, 2006). Die Methode erweist sich als besonders vorteilhaft in Rassen mit geringer Populationsgröße, in denen die rasche Zunahme von Inzuchtkoeffizienten ein erhebliches Problem darstellen kann. In solchen Fällen hilft die OGC-Methode, wertvolle genetische Variation zu bewahren und die Gesundheit und Vitalität der Population langfristig zu sichern (Sonesson & Meuwissen, 2000).

König und Simianer (2006) zeigten, wie OGC in der HF-Population zur Selektion von Bullenmüttern und Bullenvätern in verschiedenen Zuchtszenarien eingesetzt werden kann. Die Untersuchung der Beziehungen zwischen den Tieren mit den höchsten Zuchtwerten wiesen Verwandtschaftskoeffizienten von bis zu 7% auf. Daher ist es auch in konventionellen Populationen sinnvoll, die OGC Methodik anzuwenden, um verwandtschaftliche Beziehungen zwischen Jungbullen zu kontrollieren und somit langfristig Inzuchtdepressionen in der Milchviehpopulation zu vermeiden. Für die gefährdete Schweinerasse „Bunte Bentheimer“

wurden verschiedene Paarungsdesigns mit OGC evaluiert, um eine geeignete Strategie zur Minimierung der Inzucht zu finden (Biermann et al., 2013). Dabei wurden die besten Ergebnisse durch den Einsatz künstlicher Besamung erzielt, welche eine wirklich flexible Nutzung der Vatertiere ohne räumliche Beschränkungen ermöglicht.

1.5 Ziele der Studie

Die in der Arbeit enthaltenden Analysen begleiten den Prozess im Rahmen der Entwicklung und Evaluierung eines DSN-spezifischen Genotypisierungs-Chips.

Die in Kapitel 2 durchgeführten genomweiten Assoziationsanalysen liefern wertvolle Hinweise auf genomische Marker, die mit Merkmalen der Fruchtbarkeit, Gesundheit und Resistenz gegen Endoparasiteninfektionen in Zusammenhang stehen. Diese teils neuen Merkmale reflektieren die robusten und fruchtbaren Rassecharakteristika des DSN.

Nach der erfolgreichen Entwicklung und bioinformatischen Erprobung des neuen DSN200K SNP-Chips (Neumann et al., 2021) wird in Kapitel 3 eine genetische Evaluation des neuen SNP-Chips vorgenommen. Dabei werden vergleichende genomische Zuchtwertschätzungen auf Basis unterschiedlicher Markerdichten durchgeführt (50K, 200K, WGS), um eine Bewertung des Chips vorzunehmen.

In Kapitel 3 werden auf Basis des neuen DSN200K-Chips Single-Step-Zuchtwertschätzungen durchgeführt. Die geschätzten GEBVs werden anschließend verwendet, um mithilfe der OGC-Methode DSN-Elitetiere zu selektieren. Dabei wird neben der Maximierung des Zuchtfortschritts auch ein Beitrag zum Diversitätsmanagement geleistet.

1.6 Literaturübersicht - Einleitung

Baker, L., Muir, P., & Sample, S. J. (2019). Genome-wide association studies and genetic testing: understanding the science, success, and future of a rapidly developing field. *Journal of the American Veterinary Medical Association*, 255(10), 1126–1136. https://doi.org/10.2460/javma.255.10.1126

Bengtsson, C., Stålhammar, H., Strandberg, E., Eriksson, S., & Fikse, W. F. (2020). Association of genomically enhanced and parent average breeding values with cow performance in Nordic dairy cattle. *Journal of Dairy Science*, 103(7), 6383-6391. https://doi.org/10.3168/jds.2019-17367

Biscarini, F., Nicolazzi, E. L., Stella, A., Boettcher, P. J., & Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics*, 6, 33. https://doi.org/10.3389/fgene.2015.00033

Boichard, D., & Brochard, M. (2012). New phenotypes for new breeding goals in dairy cattle. *Animal*, 6(4), 544-550. https://doi.org/10.1017/S1751731111002777

Brade, W., & Brade, E. (2013). Zuchtgeschichte der Deutschen Holsteinrinder. *Berichte über Landwirtschaft - Zeitschrift für Agrarpolitik und Landwirtschaft.*

Brem, G., & Brenig, B. (1992). Molekulare Genotyp-Diagnostik des Malignen Hyperthermie-Syndroms zur effizienten Zucht stressresistenter Schweine. *Wiener Tierärztliche Monatsschrift*, 79, 301-30.

Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews Genetics*, 10(11), 783-796. https://doi.org/10.1038/nrg2664

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., ... & Hayes, B. J. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46(8), 858-865. https://doi.org/10.1038/ng.3034

Dekkers, J. C., Su, H., & Cheng, J. (2021). Predicting the accuracy of genomic predictions. *Genetics Selection Evolution*, 53, 1-23. https://doi.org/10.1186/s12711-021-00581-6

Eggen, A. (2012). The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers*, 2(1), 10-15. https://doi.org/10.2527/af.2011-0032

FAO. (2015). The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome.

Goddard, M. E., & Hayes, B. J. (2007). Genomic selection. *Journal of Animal Breeding and Genetics*, 124(6), 323-330. https://doi.org/10.1111/j.1439-0388.2007.00702.x

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), 381-391. https://doi.org/10.1038/nrg2575

Häfliger, I. M., Seefried, F. R., Spengeler, M., & Drögemüller, C. (2021). Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. *Genetics Selection Evolution*, 53, 1-16. https://doi.org/10.1186/s12711-021-00628-6

Haller, M. (2015). Alte Haus- & Nutztierrassen neu entdeckt. Graz, Stuttgart: Leopold Stocker Verlag.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433-443. https://doi.org/10.3168/jds.2008-1646

Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J., & Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics*, 6(9), e1001139. https://doi.org/10.1371/journal.pgen.1001139

Hayes, B. J., & Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, 53(11), 876-883. https://doi.org/10.1139/g10-079

Hill, W. G. (2014). Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics*, 196(1), 1-16. https://doi.org/10.1534/genetics.112.147850

Jaeger, M. (2018). Phenotypic, quantitative genetic and genomic characterization of the German black and white dual-purpose cattle breed. (Dissertation, Agrarwissenschaften) Justus-Liebig-Universität Gießen. http://geb.uni-giessen.de/geb/volltexte/2019/14000.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. Trends in *Ecology & Evolution*, 17(5), 230-241. https://doi.org/10.1016/S0169-5347(02)02489-8

Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., ... & International Sheep Genomics Consortium Members. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology*, 10(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258

Knol, E. F., Nielsen, B., & Knap, P. W. (2016). Genomic selection in commercial pig breeding. *Animal Frontiers*, 6(1), 15-22. https://doi.org/10.2527/af.2016-0003

Korkuć, P., Arends, D., & Brockmann, G. A. (2019). Finding the optimal imputation strategy for small cattle populations. *Frontiers in Genetics*, 10, 430415. https://doi.org/10.3389/fgene.2019.00430

Korkuć, P., Arends, D., May, K., & König, S. (2021). Genomic loci affecting milk production in German Black Pied cattle (DSN). *Frontiers in Genetics*, 12, 640039. https://doi.org/10.3389/fgene.2021.640039

Korkuć, P., Neumann, G. B., Hesse, D., Arends, D., Reißmann, M., Rahmatalla, S., May, K., Wolf, M. J., König, S., & Brockmann, G. A. (2023). Whole-genome sequencing data reveal new loci affecting milk production in German Black Pied cattle (DSN). *Genes*, 14(3), 581. https://doi.org/10.3390/genes14030581

Lipkin, E., Mosig, M. O., Darvasi, A., Ezra, E., Shalom, A., Friedmann, A., ... & Soller, M. (1998). Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. *Genetics*, 149(4), 1557-1567.

Lindholm-Perry, A. K., Cunningham, H. C., Kuehn, L. A., Vallet, J. L., Keele, J. W., Foote, A. P., Cammack, K. M., & Freetly, H. C. (2017). Relationships between the genes expressed in the mesenteric adipose tissue of beef cattle and feed intake and gain. *Animal Genetics*, 48(4), 386–394. https://doi.org/10.1111/age.12565

Malomane, D. K., Simianer, H., Kadri, N. K., & Weigend, S. (2018). Genetic diversity in global chicken breeds as a function of genetic distance to wild populations. *Genetics Selection Evolution*, 50, 22. https://doi.org/10.1186/s12711-018-0390-5

Mészáros, G., Boison, S. A., Pérez O'Brien, A. M., Ferenčaković, M., Curik, I., Da Silva, M. V. B., ... & Sölkner, J. (2015). Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. *Frontiers in Genetics*, 6, 126535. https://doi.org/10.3389/fgene.2015.00013

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., ... & Van Tassell, C. P. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, 4(4), e5350. https://doi.org/10.1371/journal.pone.0005350

May, K., Scheper, C., Brügemann, K., Yin, T., Strube, C., Korkuć, P., Brockmann, G. A., & König, S. (2019). Genome-wide associations and functional gene analyses for endoparasite resistance in an endangered population of native German Black Pied cattle. *BMC Genomics*, 20, 1-15. https://doi.org/10.1186/s12864-019-6048-0

May, K., Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P., König, S., & Brockmann, G. A. (2023). Benefits of genotype and sequence data for genetic improvement in breeds with

a small population size with the example of the German Black Pied cattle. *Züchtungskunde* 95 (1), 18-34

Meier, S., Arends, D., Korkuć, P., Neumann, G. B., & Brockmann, G. A. (2020). A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. *Journal of Dairy Science*, 103(11), 10289-10298. https://doi.org/10.3168/jds.2019-17923

Meuwissen, T. H. E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science*, 75(4), 934-940. https://doi.org/10.2527/1997.754934x

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

Mügge, B. (1999). Deutsche Holsteins: Die Geschichte einer Zucht. Stuttgart: Ulmer.

Müller, R., Feiler, J., & Clare, A. (2022). A doomed technology? On gene editing in bavarian livestock agriculture, policy field conflicts and responsible research and innovation. *Frontiers in Political Science*, 4, 800211. https://doi.org/10.3389/fpos.2022.800211

Naderi, S., Moradi, M. H., Farhadian, M., Yin, T., Jaeger, M., Scheper, C., ... & May, K. (2020). Assessing selection signatures within and between selected lines of dual-purpose black and white and German Holstein cattle. *Animal Genetics*, 51(3), 391-408. https://doi.org/10.1111/age.12914

Neumann, G. B., Korkuć, P., Arends, D., Wolf, M. J., May, K., Reißmann, M., König, S., & Brockmann, G. A. (2021). Design and performance of a bovine 200k SNP chip developed for endangered German Black Pied cattle (DSN). *BMC genomics*, 22(1), 905 https://doi.org/10.18452/24547

Neumann, G. B., Korkuć, P., Arends, D., Wolf, M. J., May, K., König, S., & Brockmann, G. A. (2023a). Genomic diversity and relationship analyses of endangered German Black Pied cattle (DSN) to 68 other taurine breeds based on whole-genome sequencing. *Frontiers in Genetics,* 13, 993959. https://doi.org/10.3389/fgene.2022.993959

Neumann, G. B., Korkuć, P., Reißmann, M., Wolf, M. J., May, K., König, S., & Brockmann, G. A. (2023b). Unmapped short reads from whole-genome sequencing indicate potential infectious pathogens in German Black Pied cattle. *Veterinary Research*, 54(1), 95. https://doi.org/10.1186/s13567-023-01188-y

Notter, D. R. (1999). The importance of genetic diversity in livestock populations of the future. *Journal of Animal Science*, 77(1), 61–69. https://doi.org/10.2527/1999.77161x

Porto-Neto, L. R., Reverter, A., Prayaga, K. C., Chan, E. K., Johnston, D. J., Hawken, R. J., ... & Barendse, W. (2014). The genetic architecture of climatic adaptation of tropical cattle. *PLoS One*, 9(11), e113284. https://doi.org/10.1371/journal.pone.0113284

Purfield, D. C., Evans, R. D., & Berry, D. P. (2020). Breed-and trait-specific associations define the genetic architecture of calving performance traits in cattle. *Journal of Animal Science*, 98(5), skaa151. https://doi.org/10.1093/jas/skaa151

Rahmann, G. (2011). Modellvorhaben zum Einsatz verschiedener, gefährdeter Rinderrassen in extensiven Beweidungssystemen des Feuchtgrünlandes unter den Bedingungen der Agrarreform und einer innovativen, nachhaltigen Wirtschaftsweise im Biosphärenreservat Flusslandschaft Elbe. (Endbericht, wissenschaftliche Begleitung) Johann Heinrich von Thünen-Institut.

Rodríguez-Ramilo, S. T., García-Cortés, L. A., & de Cara, M. Á. R. (2015). Artificial selection with traditional or genomic relationships: consequences in coancestry and genetic diversity. *Frontiers in Genetics*, 6, 124866. https://doi.org/10.3389/fgene.2015.00013

Rupp, R., Mucha, S., Larroque, H., McEwan, J., & Conington, J. E. (2016). Genomic application in sheep and goat breeding. *Animal Frontiers*, 6(1), 39-44. https://doi.org/10.2527/af.2016-0005

Sahana, G., Cai, Z., Sanchez, M. P., Bouwman, A. C., & Boichard, D. (2023). Invited review: Good practices in genome-wide association studies to identify candidate sequence variants in dairy cattle. *Journal of Dairy Science*, 106(8), 5218–5241. https://doi.org/10.3168/jds.2022-22905

Sarviaho, K., Uimari, P., & Martikainen, K. (2024). Signatures of positive selection after the introduction of genomic selection in the Finnish Ayrshire population. *Journal of Dairy Science*, In press. https://doi.org/10.3168/jds.2024-24105

Schork, N. J., Fallin, D., & Lanchbury, J. S. (2000). Single nucleotide polymorphisms and the future of genetic epidemiology. *Clinical Genetics*, 58(4), 250-264. https://doi.org/10.1034/j.1399-0004.2000.580402.x

Sonesson, A. K., & Meuwissen, T. H. (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution*, 32(3), 231-248. https://doi.org/10.1186/1297-9686-32-3-231

Tabor, A. E., et al. (2021). Genotyping technologies for genetic improvement of livestock: A review. *Genetics Selection Evolution*, 53, 6. https://doi.org/10.1186/s12711-021-00608-y

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., ... & Zammit, G. (2023). How genomics can help biodiversity conservation. *Trends in Genetics*, 39(7), 545-559. https://doi.org/10.1016/j.tig.2023.03.001

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414-4423. https://doi.org/10.3168/jds.2007-0980

Van Tassell, C. P., Smith, T. P., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., ... & Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5(3), 247-252. https://doi.org/10.1038/nmeth.1185

Weber, J. L., & May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*, 44(3), 388-396.

Weller, J. I., Ezra, E., & Ron, M. (2017). Invited review: A perspective on the future of genomic selection in dairy cattle. *Journal of Dairy Science*, 100(11), 8633-8644. https://doi.org/10.3168/jds.2017-12879

Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., ... & Dekkers, J. C. (2011). Breeding value prediction for production traits in layer chickens using pedigree or genomic relationships in a reduced animal model. *Genetics Selection Evolution*, 43, 1-9. https://doi.org/10.1186/1297-9686-43-5

Wolc, A., Kranis, A., Settar, P., Fulton, J. E., O'Sullivan, N. P., Watson, K., ... & Dekkers, J. C. M. (2016). Implementation of genomic selection in the poultry industry. *Animal Frontiers*, 6(1), 23-31. https://doi.org/10.2527/af.2016-0004

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P., Brockmann, G. A., König, S., & May, K. (2021). Genome-wide association study using whole-genome sequence data for fertility, health indicator, and endoparasite infection traits in German black pied cattle. *Genes*, 12(8), 1163. https://doi.org/10.3390/genes12081163

Wolf, M. J., Neumann, G. B., Korkuć, P., Yin, T., Brockmann, G. A., König, S., & May, K. (2023). Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing. *Journal of Dairy Science*, 106(5), 3345-3358. https://doi.org/10.3168/jds.2022-22167

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P, Brockmann, G. A., & König, S. (2024). Single-step breeding value estimations and optimum contribution selection in endangered dual-purpose German Black and WhitePied cattle (DSN) using a breed specific SNP chip. *Journal of Animal Breeding and Genetics*. Eingereicht am 10.06.2024.

Zhang, Y., Fan, X., Sun, D., Wang, Y., Yu, Y., Xie, Y., ... & Zhang, Y. (2012). A novel method for rapid and reliable detection of complex vertebral malformation and bovine leukocyte adhesion deficiency in Holstein cattle. *Journal of Animal Science and Biotechnology*, 3, 1-6. https://doi.org/10.1186/2049-1891-3-6

Zucht- und Besamungsunion Hessen eG (ZBH) (2019). Zuchtprogramm der Zucht- und Besamungsunion Hessen eG - Deutsches Schwarzbuntes Niederungsrind. https://www.qnetics.de/fileadmin/downloads/Zuchtprogramme/Hessen_Milch_und_Zweinutzung/Zuchtprogramm_DSN_31.05.2019.pdf. Stand: 31.05.2019.

KAPITEL 2

Genome-wide association study using whole-genome sequence data for fertility, health indicator, and endoparasite infection traits in German Black Pied Cattle

Manuel J. Wolf[1], Tong Yin[1], Guilherme B. Neumann[2], Paula Korkuć[2], Gudrun A. Brockmann[2], Sven König[1] and Katharina May[1]

[1] Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany

[2] Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt-Universität of Berlin, 10115 Berlin, Germany

Genes 2021, 12(8), 1163; https://doi.org/10.3390/genes12081163

Accepted: 16 December 2022

Published: 28 July 2021

(This article belongs to the Special Issue “Genome-Wide Association Analysis of Cattle”)

genes

MDPI

Article

Genome-Wide Association Study Using Whole-Genome Sequence Data for Fertility, Health Indicator, and Endoparasite Infection Traits in German Black Pied Cattle

Manuel J. Wolf [1], Tong Yin [1], Guilherme B. Neumann [2], Paula Korkuć [2], Gudrun A. Brockmann [2], Sven König [1] and Katharina May [1,*]

[1] Institute of Animal Breeding and Genetics, Justus-Liebig—University of Gießen, 35390 Gießen, Germany; manuel.j.wolf@agrar.uni-giessen.de (M.J.W.); tong.yin@agrar.uni-giessen.de (T.Y.); sven.koenig@agrar.uni-giessen.de (S.K.)

[2] Albrecht Daniel Thaer—Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt-Universität of Berlin, 10115 Berlin, Germany; guilherme.neumann@hu-berlin.de (G.B.N.); paula.korkuc@hu-berlin.de (P.K.); gudrun.brockmann@hu-berlin.de (G.A.B.)

* Correspondence: katharina.may@agrar.uni-giessen.de

Abstract: This genome-wide association study (GWAS) aimed to identify sequence variants (SVs) and candidate genes associated with fertility and health in endangered German Black Pied cattle (DSN) based on whole-genome sequence (WGS) data. We used 304 sequenced DSN cattle for the imputation of 1797 genotyped DSN to WGS. The final dataset included 11,413,456 SVs of 1886 cows. Cow traits were calving-to-first service interval (CTFS), non-return after 56 days (NR56), somatic cell score (SCS), fat-to-protein ratio (FPR), and three pre-corrected endoparasite infection traits. We identified 40 SVs above the genome-wide significance and suggestive threshold associated with CTFS and NR56, and three important potential candidate genes (*ARHGAP21*, *MARCH11*, and *ZNF462*). For SCS, most associations were observed on BTA 25. The GWAS revealed 61 SVs, a cluster of 10 candidate genes on BTA 13, and 7 pathways for FPR, including key mediators involved in milk fat synthesis. The strongest associations for gastrointestinal nematode and *Dictyocaulus viviparus* infections were detected on BTA 8 and 24, respectively. For *Fasciola hepatica* infections, the strongest associated SVs were located on BTA 4 and 7. We detected 200 genes for endoparasite infection traits, related to 16 pathways involved in host immune response during infection.

Keywords: complex traits; candidate genes; dairy cows; dual-purpose cattle; endangered breed; endoparasite resistance; GWAS

Citation: Wolf, M.J.; Yin, T.; Neumann, G.B.; Korkuć, P.; Brockmann, G.A.; König, S.; May, K. Genome-Wide Association Study Using Whole-Genome Sequence Data for Fertility, Health Indicator, and Endoparasite Infection Traits in German Black Pied Cattle. *Genes* **2021**, *12*, 1163. https://doi.org/10.3390/genes12081163

Academic Editor: Emilia Bagnicka

Received: 7 July 2021
Accepted: 27 July 2021
Published: 28 July 2021

1. Introduction

Rapid advances in sequencing technology have opened new opportunities for dairy cattle breeding. Whole-genome sequencing (WGS) and genotype imputation to whole-genome sequence genotypes is an effective method to identify genes and causal mutations for single-gene traits, genetic defects, and complex polygenic traits in cattle [1]. Compared to commonly used single nucleotide polymorphism (SNP) chip arrays, WGS data enable discovering both common and rare variants affecting complex polygenic traits in cattle [2]. Consideration of complete linkage information in WGS data plus the possible consideration of the full genetic variance may contribute to the identification of variants, which do not exceed the significance threshold in genome-wide association studies (GWAS) using commercial SNP panels [2,3]. As an outstanding step forward, using WGS data in GWAS enhances the discovery of causative mutations [4,5] and increases the power to map genes for low heritable functional traits in cattle [4,6,7].

Powerful gene mapping approaches for functional traits is of special importance for the conservation of local and endangered breeds, since these are less competitive to large

cattle populations with regard to economically relevant traits [8,9]. Being predominantly kept in outdoor systems, local breeds are considered to be robust to harsh environmental conditions, displaying favorable adaptive trait characteristics such as heat stress or disease resistance [9–11]. The conservation of local endangered breeds is essential to offer solutions for future challenges in animal husbandry and to preserve genetic diversity [11]. In this regard, WGS data may include genomic markers which can be used to improve functional traits [12]. Moreover, compared with pedigree data, deep and dense genomic information as provided by WGS data improves inbreeding estimations and is powerful to identify deleterious variants in a population. Both aspects, i.e., controlling inbreeding and deleterious allele frequencies, are important mechanisms in genomic selection programs to maintain long-term selection response and genetic diversity [13].

With a small population size of ~2560 cattle, the German dual-purpose Black Pied cattle (DSN, German: Deutsches Schwarzbuntes Niederungsrind) is an endangered local breed in Germany [14]. The breed has a long breeding history in the German grassland region East Frisia, as well as in the German federal states of Lower Saxony and Brandenburg located in the eastern part of Germany [15,16]. DSN is considered the founder breed of the modern Holstein Friesian (HF) and characterized by higher fertility and milk protein and fat content compared to HF [10,17]. DSN cows are well adapted to grazing systems, and thus, 50% of all DSN cows in Germany are kept under organic pasture conditions [18]. However, grazing is associated with an increased risk for endoparasite infections [19].

Gastrointestinal nematodes (GIN), the bovine lungworm (*Dictyocaulus viviparus*), and the liver fluke (*Fasciola hepatica*) are the three most important helminthic species in pastured dairy cattle [20]. The annual estimated costs of these three helminth infections were 941 million euros in dairy cattle in Europe due to impairments in milk production and fertility [20]. Since anthelmintic treatment against endoparasite infections is restricted in organic dairy farming, there is a need to develop breeding strategies to improve cows' resistance to helminthic parasites. In this regard, research has aimed to investigate the genetic background of resistance to endoparasite infections in dairy cattle since 2017 [21–23]. Surprisingly, in a cross-classified research design for selection line comparisons, May et al. [21] identified higher GIN and *D. viviparus* infection rates for DSN compared to HF. An ongoing GWAS [22] detected the strongest associations for GIN and *D. viviparus* infections on BTA 2, 5, 8, 15, 17, 21, and 24 in a population of 148 DSN cows with 700K imputed genotypes. For *F. hepatica*, associated significant SNP markers were identified on BTA 1, 7, and 28. The genes *ALCAM*, *CDH2*, *EGFR*, and *PHLPP1* were annotated as main candidates involved in immunological functions during endoparasite infections in DSN. As an extension to commercial SNP chip applications, GWAS based on WGS data is expected to unravel the genetic architecture of endoparasite resistance and further important functional traits in DSN much deeper. For example, Twomey et al. [7] applied a GWAS for endoparasite infection traits in Irish dairy and beef cattle and demonstrated that only 0 to 11% of all quantitative trait loci (QTL) from imputed WGS data were detected when using 50 K genotypes.

Further GWAS in DSN were based on commercial 50 K SNP chip panels with focus on milk production [24] and udder health traits [25]. These studies revealed candidate genes previously described in HF and other cattle breeds, raising the interest in identifying species-specific variants for functional traits in DSN with WGS data. Clinical mastitis is the most common disease in DSN with incidences up to 26% [25], raising the interest in improving somatic cell counts (SCC) in milk as an indicator for udder health in DSN via genomic selection. Jaeger [10] studied udder health indicator traits in DSN phenotypically and identified a strong detrimental impact of increased SCC on feed intake and rumination. Although well adapted to grazing systems, DSN are exposed to metabolic stress due to heat stress impact on pastures and associated mobilization of body reserve in response to restricted feed intake. The fat-to-protein ratio (FPR) is a valuable indicator for metabolic health (e.g., ketosis) and for metabolic stability in this regard [26,27]. Moreover, FPR might be a novel indicator trait for robustness in DSN, especially in the context of challenging

heat stress environments in outdoor systems [28]. Due to their favorable grazing abilities, the conservation of DSN is financially supported by the German government. However, ongoing DSN breed competitiveness implies an optimization of the preventive health management in grazing and organic farming systems via genomic selection, especially to improve metabolic stability and resistance to infectious diseases.

Therefore, the present study aimed to identify genome-wide associations, potential candidate genes and pathways for female fertility, SCC as an indicator for udder health, FPR as an indicator for metabolic health and stability, and endoparasite infections in DSN based on WGS data. Providing genomic markers for fertility and functional health traits is important for the development of future genomic selection programs to preserve DSN and to improve the breed adaption towards pasture environments.

2. Materials and Methods

2.1. Cow Traits

Cow traits were available for 1886 DSN cows from eight dairy farms located in the German federal states of Brandenburg (five farms), Hesse (one farm), and Lower Saxony (two farms). The cows were born between 2005 and 2016. The first calving age ranged from 22 to 40 months (mean: 27.3 months). Fertility traits included calving-to-first service interval (CTFS) and the non-return after 56 days (NR56 = insemination success proved at day 56 after first insemination) in first parity cows. For the udder and metabolic health indicator traits SCC and FPR, we included the first test-day record between days in milk (DIM) 5 and 40 from the first parity. SCC was log-transformed into somatic cell score (=SCS =$\log_2$ (SCC/100,000) + 3) [29]. For fertility and health indicator traits, we excluded cows with a genetic breed (DSN) percentage lower than 90% according to our own developed algorithm to clearly differentiate between DSN and HF cows [30].

The endoparasite infection traits considered repeated measurements for fecal egg counts for GIN (FEC-GIN), fecal egg counts for *F. hepatica* (FEC-FH), and fecal larvae counts (FLC) for *D. viviparus* (FLC-DV). A modified McMaster technique [31] with a sensitivity of 25 eggs/g feces was used to determine FEC-GIN. For GIN, strongylid eggs were the predominant morphotype followed by *Strongyloides papillosus* and *Capillaria* spp. eggs. Fecal egg counts for *F. hepatica* were examined by the sedimentation technique using 10 g feces per sample. The fecal larvae count for *D. viviparus* was determined with the Baermann technique using 40 g feces per sample [32]. Endoparasite traits (FEC-GIN, FLC-DV, FEC-FH) were available for an initial dataset including 1166 untreated and pastured Black and White dairy cows (including DSN) examined in 2015 [21]. Using this initial dataset, endoparasite traits were pre-corrected for fixed effects via linear mixed models in the statistical software SAS (version 9.4 [33]) as described in May et al. [22]. In this regard, farm, parity, genetic line (DSN and other Holstein Friesian selection lines), season of parasitological examination, and lactation stage of cows were included as fixed effects. The pre-corrected phenotypes (residuals) for the three endoparasite traits (FEC-GIN, FEC-FH, and FLC-DV) from the linear mixed models are later denoted as RES-GIN, RES-FH, and RES-DV, respectively. Descriptive statistics for FEC-GIN, FLC-DV, and FEC-FH in the initial dataset of 1166 cows are presented in Table 1.

2.2. Genetic Architecture of Cow Traits

Genetic parameters were estimated using the software GCTA [34] for CTFS, NR56, SCS, and FPR considering the genomic relationship matrix (**G**) and applying linear mixed models as described in Section 2.5. for the GWAS. In this regard, the genomic relationship matrix was constructed in GCTA according to VanRaden [35]. The SNP-based heritabilities were 0.05 (±0.03) and 0.02 (±0.02) for CTFS and NR56, respectively. For SCS and FPR, the SNP-based heritabilities were 0.13 (±0.04) and 0.14 (±0.04), respectively. For endoparasite traits, we estimated heritabilities based on the pedigree matrix (**A**) using the initial dataset of 1166 cows [21]. The pedigree-based heritabilities were 0.06 (±0.04) for FEC-GIN, 0.05 (±0.04) for FEC-DV, and 0.33 (±0.06) for FEC-FH.

Table 1. Descriptive statistics for fertility traits, health and metabolic stability indicator traits, and endoparasite infection traits in DSN.

Trait [1]	No. of Records	No. of Cows	No. of Cows with Sequence Level Genotypes	Mean [2]	SD [2]	Min. [2]	Max. [2]
CTFS	1683	1683	1683	78.31	27.44	26.0	241.0
NR56	1683	1683	1683	0.64	0.48	0	1.0
SCS	1638	1638	1638	2.58	1.48	−1.06	8.88
FPR	1638	1638	1638	1.18	0.18	0.45	2.42
FEC-GIN	1997	1166	142	11.35	22.57	0	225.0
FEC-FH	2006	1166	142	0.61	3.64	0	89.0
FLC-DV	1988	1163	142	0.17	2.14	0	46.0

[1] CTFS = calving-to-first service interval; NR56 = non-return after day 56; SCS = somatic cell score (log-transformed somatic cell count: log2 (SCC/100,000) +3); FPR = fat-to-protein ratio; FEC-GIN = fecal egg count for gastrointestinal nematodes; FEC-FH = fecal egg count for *Fasciola hepatica*; FLC-DV = fecal larvae count for *Dictyocaulus viviparus*. [2] Mean, SD, Min. and Max. values are given for the number of records.

2.3. Whole-Genome Sequencing and Imputation of 50K Genotypes

Whole-genome sequencing data were available for 304 DSN cattle, of which 47 were bulls (all available DSN sires used for artificial insemination) and 257 were cows from eight different herds, reflecting the whole phenotypic range for milk yield, milk composition, and reproduction traits in the DSN population. Sequencing was performed on the Illumina NovaSeq platform (Novogene Bioinformatics Technology Co., Ltd., Beijing, China) with 150 paired end reads and 15× coverage. Sequence read mapping, variant calling, and recalibration were performed following the 1000 Bull Genome (http://www.1000bullgenomes.com; accessed on 21 February 2020 guidelines. Lower quality sequence variants (SVs) were discarded by applying the machine learning method "variant quality score recalibration" in the Genome Analysis Toolkit (GATK, version 4.1.3.0 [36]) and using training and truth set variants provided by the 1000 Bull Genomes database. After, the dataset included 20,567,619 SVs (including 18.5 million SNPs and 2.0 million indels). SVs with a minor-allele frequency (MAF) <1% and a call rate < 95% were discarded. SVs with 5% of Mendelian inconsistences (i.e., opposing homozygotes detected from 156 sire-offspring pairs) were removed. Moreover, SVs with a quality by depth (QD) <10 and coverage <3000 were removed in case they were not predicted as high or moderate impact by the variant effect predictor (VEP) [37]. After quality filtering, the dataset contained 16,175,216 SVs from 304 DSN. This dataset provided a reference panel for the imputation of 1797 DSN genotyped with the BovineSNP50 (50 K) Bead Chip V2 (Illumina Inc., San Diego, CA, USA). The average numerator relationship between the 304 cattle from the reference panel and the 1797 genotyped cows for imputation was 6.7%. The inbreeding coefficients were 1.9% for the reference group and 2.1% for the imputation group. The imputation was performed in the software BEAGLE (version 5.1 [38]) based on a one-step imputation accuracy approach [39]. The average imputation accuracy was 97.04%. SVs with MAF <5%, linkage disequilibrium (r^2) <0.5 (output from BEAGLE), and significant deviation from the Hardy–Weinberg equilibrium (HWE, $p < 10^{-6}$) were discarded from the imputed dataset. Then, the 1797 imputed DSN genotypes were merged with the 304 sequenced DSN, resulting in a total number of 12,164,173 SVs (including 11,043,497 SNPs and 1,120,676 indels) from 2101 sequence level DSN genotypes.

2.4. Quality Control of Whole-Genome Sequence Genotypes

Imputed sequence-level genotypes and cow traits were merged, resulting in a total number of 1683 cows with phenotypes for fertility traits, 1638 cows with phenotypes for health and metabolic stability indicator traits, and 142 cows with phenotypes for pre-corrected endoparasite infection traits. Quality control of imputed WGS genotypes was performed within the three datasets in three consecutive runs in PLINK (version 1.9 [40]). SVs with MAF <5% and significant deviation from HWE ($p < 10^{-6}$) or a call

rate <95% were discarded. After quality control, 11,391,082 SVs (10,343,725 SNPs and 1,047,357 indels) remained for cows with phenotypes for CTFS and NR56, 11,413,456 SVs (10,363,951 SNPs and 1,049,505 indels) remained for cows with phenotypes for SCS and FPR, and 10,595,540 SVs (9,624,332 SNPs and 971,208 indels) remained for cows with pre-corrected phenotypes (residuals) for endoparasite infection traits (RES-GIN, RES-FH and RES-DV). Descriptive statistics for all traits of cows with imputed sequence level genotypes after quality control is shown in Table 1.

2.5. Genome-Wide Association Analyses

We applied a single marker linear mixed model in the software package GCTA [34] to identify genome-wide associations for fertility traits (CTFS, NR56), health and metabolic stability indicator traits (SCS, FPR), and pre-corrected endoparasite infection traits (RES-GIN, RES-FH, RES-DV). All GWAS were performed as MLMA-LOCO "leaving one chromosome out" analysis for large datasets using the –mlma and –mlma-subtract-grm options in GCTA. The statistical model for testing single-locus effects was defined as follows:

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

where y = vector including records for CTFS, NR56, SCS, FPR, RES-GIN, RES-FH, and RES-DV; β = vector of fixed effects (farm, calving year, calving month, and a linear regression on age at first calving for CTFS and NR56; farm, test-day year-season, a linear regression on DIM, and a linear regression on fat percentage for SCS; farm, test-day year-season, and a linear regression on age at first calving for FPR); u = vector of polygenic effects with $u \sim N(0, G\sigma^2_u)$, with G denoting the genetic similarity matrix among individuals, and σ^2_u the polygenic variance; s = vector for marker effects; e = vector of random residuals; and X, Z, and S were incidence matrices for β, u, and s, respectively.

For model quality, we checked quantile-quantile (Q-Q) plots and calculated the genomic inflation factor λ for each association analysis. We applied a Bonferroni correction to account for multiple testing. Since traditional Bonferroni correction (i.e., relating the genome-wide significance threshold of 0.05 to the total number of markers) tends to produce many false-negative associations [41], we did not use the total number of SVs but estimated the effective number of independent SVs (n_{eff}). To calculate n_{eff}, one sequence variant (SV) of a SV pair in LD with $r^2 > 0.5$ was excluded using the –indep-pairwise option in PLINK [40]. A window size of 5000 SVs was chosen, which was shifted in an interval of 500 SVs. The adjusted Bonferroni-corrected genome-wide significance threshold based on n_{eff} ($p = 0.05/n_{eff}$) was pBonf = 1.60×10^{-7} (0.05/312,502) for CTFS and NR56, pBonf = 1.56×10^{-7} (0.05/319,686) for SCS and FPR, and pBonf = 1.22×10^{-7} (0.05/409,481) for RES-GIN, RES-FH, and RES-DV. In addition, we considered a less conservative suggestive significance threshold (pSug = $1/n_{eff}$). The suggestive significance threshold was pSug = 3.20×10^{-6} (1/312,502) for CTFS and NR56, pSug = 3.13×10^{-6} (1/319,686) for SCS and FPR, and pSug = 2.44×10^{-6} (1/409,481) for RES-GIN, RES-FH, and RES-DV, respectively.

2.6. Candidate Gene Annotations and Pathway Analyses

We applied the biomaRt R package [42,43] from Bioconductor to retrieve 'rs accession numbers' of associated SVs using the getBM() function. Potential candidate genes were queried and assigned to the associated SVs using the current gene annotations from ENSEMBL (release 104) [44] based on the *Bos taurus* ARS-UCD1.2 genome assembly [45]. A gene was considered as a candidate gene if at least one SV above pSug was positioned in the respective gene and/or within 100 kb up- and downstream of the respective candidate gene. Physiological functions and positions of potential candidate genes were manually reviewed in the ENSEMBL and KEGG [46] databases. In addition, the identified potential candidate genes were manually submitted to the DAVID (version 6.8 [47]) and KEGG pathway databases for pathway analysis.

3. Results

3.1. Female Fertility Traits

3.1.1. Calving-to-First Service Interval

The Manhattan plot from the GWAS with CTFS is given in Figure 1A. The genomic inflation factor λ was 1.041. We identified 33 SNPs above *p*Sug on BTA 5, 12, 13, 15, and 28 (Table S1). In addition, one SNP (rs380946888) on BTA 12 surpassed the genome-wide significance threshold *p*Bonf. The highest number of associations was identified on BTA 13 (n = 15) and on BTA 28 (n = 10). The associated markers were annotated to six potential candidate genes on BTA 12, 13, 15, and 28 (Table 2). On BTA 12, we identified the Kelch-like family member 1 (*KLHL1*) gene. The highest association ($p = 2.25 \times 10^{-7}$) on BTA 13 was found within the Rho GTPase activating protein 21 (*ARHGAP21*) gene with five SNPs located in the gene and nine SNPs and one indel close to the gene (Table 2). Further annotated genes were the lin-7 homolog C (*LIN7C*) gene, *ENSBTAG00000052005*, and the olfactory receptor family 5 subfamily member 5 (*OR5BE5*) gene on BTA 15. All associated variants on BTA 28 are located in the choline O-acetyltransferase (*CHAT*) gene.

3.1.2. Non-Return after Day 56

Figure 1B shows the Manhattan plot from the GWAS with NR56. The genomic inflation factor λ was 1.014. The GWAS revealed six SVs (five SNPs and one indel) above *p*Sug on BTA 8, 11, 13, and 20 for NR56 (Table S1). The associated markers were annotated to three potential candidate genes (Table 2). The SNPs rs110809463 and rs135364419 on BTA 8 are located in the zinc finger protein 462 (*ZNF462*) gene. The SNP rs383197946 on BTA 11 is located in the EFR3 homolog B (*EFR3B*) gene. The SNP rs207515592 on BTA 20 is closely related to the membrane associated ring-CH-type finger 11 (*MARCH11*) gene.

Table 2. Potential candidate genes related to the identified sequence variants (SVs) significantly associated with the female fertility traits calving-to first service interval (CTFS) and non-return 56 (NR56) in DSN.

BTA	Gene Position [1]	No. of SVs within/Close to Gene [2]	Position of Maximum Association (*p*-Value)	rs Number of Maximum Association	Gene Name
CTFS					
12	43,998,992–44,525,321	3/0	44,156,560 (2.93×10^{-6}) *	rs136060929	*KLHL1*
13	25,463,655–25,592,062	5/10	25,577,261 (2.25×10^{-7}) *	rs384930569	*ARHGAP21*
15	58,214,580–58,224,553	0/1	58,258,453 (1.32×10^{-6}) *	rs41777070	*LIN7C*
	78,914,154–78,915,587	0/1	78,929,182 (2.08×10^{-6}) *	rs379801720	*ENSBTAG00000052005*
	79,013,311–79,014,249	0/1	78,929,182 (2.08×10^{-6}) *	rs379801720	*OR5BE5*
28	43,760,392–43,804,579	10/0	43,796,638 (1.37×10^{-6}) *	rs42155599	*CHAT*
NR56					
8	96,644,114–96,759,752	2/0	96,724,190 (4.07×10^{-7}) *	rs110809463	*ZNF462*
11	74,126,883–74,224,519	1/0	74,190,369 (8.30×10^{-7}) *	rs383197946	*EFR3B*
20	56,953,217–57,071,806	0/1	56,867,118 (1.34×10^{-6}) *	rs207515592	*MARCH11*

[1] Gene position (start-end) in ENSEMBL build on *Bos taurus* genome assembly ARS-UCD1.2; [2] number of associations that reached the Bonferroni-corrected genome-wide significance threshold (*p*Bonf) or the suggestive significance threshold (*p*Sug) based on the position of the identified candidate gene ± 100 kb up- and downstream; * above *p*Sug; BTA = *Bos taurus* chromosome.

3.2. Health Indicator Traits

3.2.1. Somatic Cell Score

The Manhattan plot from the GWAS with SCS is presented in Figure 2A. The genomic inflation factor λ was 1.072. The GWAS revealed one SNP (rs137783421) above *p*Bonf on BTA 11 and 53 SNPs above *p*Sug on BTA 1, 9, 11, 13, 23, and 25 (Table S2). The majority of associations (n = 47) was found on BTA 25. We identified nine potential candidate genes for SCS (Table 3). The claudin 8 (*CLDN8*) gene on BTA 1 was annotated to the leukocyte transendothelial migration pathway and to the cell adhesion molecules pathway, playing a major role in gene expressions during *Escherichia coli*-induced mastitis in dairy cows (Table 4). One marker on BTA 9 was located in the SUMO specific peptidase 6 (*SENP6*) gene. The SNP rs137783421 (maximum association for SCS) on BTA 11 is located in the

Rap guanine nucleotide exchange factor 1 (*RAPGEF1*) gene. The RAN binding protein 9 (*RANBP9*) gene on BTA 23 was annotated to the SNP rs876215027. All genomic associations on BTA 25 (n = 47; bp: 6,629,679–6,656,058) are located in or close to the RNA binding fox-1 homolog 1 (*RBFOX1*) gene.

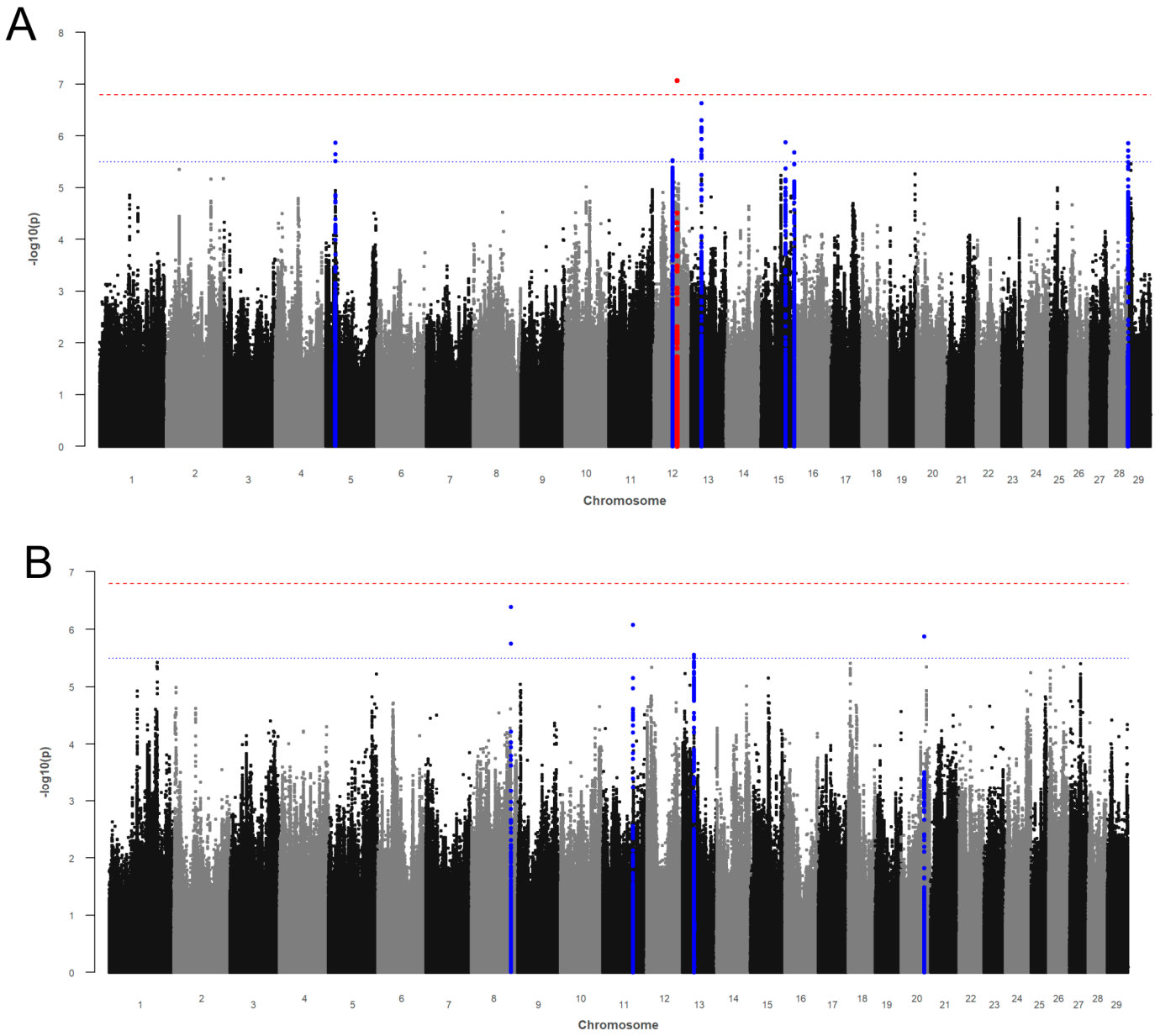

Figure 1. Manhattan plots for −log10 p-values of marker effects for (**A**) calving-to-first service interval (CTFS) and (**B**) non-return 56 (NR56). Markers above the genome-wide significance threshold pBonf (red dashed line) are highlighted in red and markers above the suggestive significance threshold pSug (blue dashed line) are highlighted in blue. Markers within a distance of 125 kb up- and downstream of the significantly or suggestively associated SVs are highlighted in red or blue, respectively.

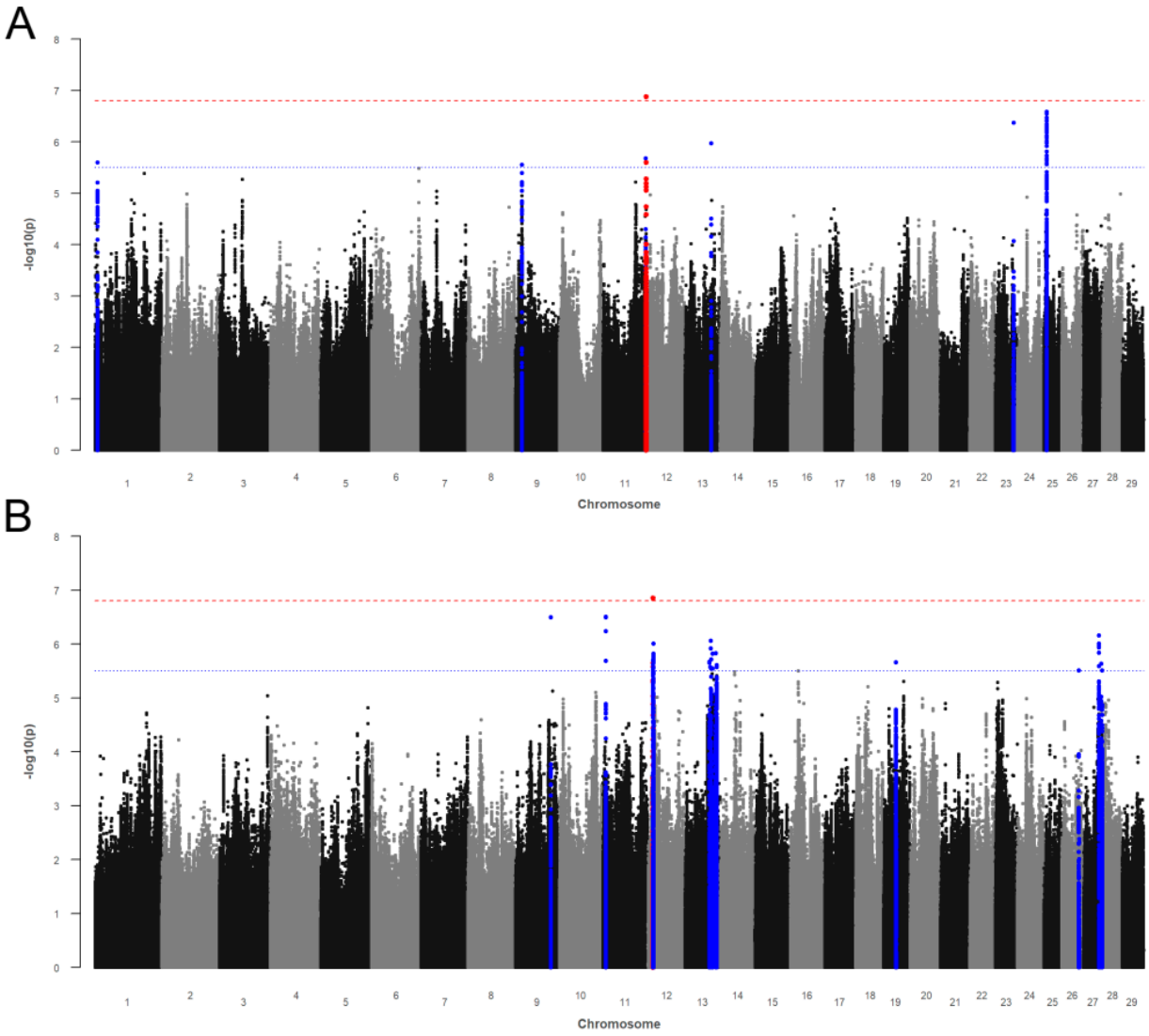

Figure 2. Manhattan plots for −log10 *p*-values of marker effects for (**A**) somatic cell score (SCS) and (**B**) fat-to-protein ratio (FPR). Markers above the genome-wide significance threshold *p*Bonf (red dashed line) are highlighted in red and markers above the suggestive significance threshold *p*Sug (blue dashed line) are highlighted in blue. Markers within a distance of 125 kb up- and downstream of the significantly or suggestively associated SVs are highlighted in red or blue, respectively.

3.2.2. Fat-to-Protein Ratio

The Manhattan plot from the GWAS for FPR is presented in Figure 2B. The genomic inflation factor λ was 1.088. The GWAS revealed one SNP (rs439994366) above *p*Bonf on BTA 12 and 60 SVs (54 SNPs and six indels) above *p*Sug on BTA 9, 11, 12, 13, 19, 26, and 27 (Table S2). The associated markers were annotated to 19 potential candidate genes (Table 3). The SNP rs439994366 is located in the Von Willebrand factor A domain containing 8 (*VWA8*) gene on BTA 12, with 15 SVs (thereof two indels) located in the gene and one further SNP at a close distance. The region between 58.4 and 75.3 Mb on BTA 13 harbored a cluster of ten potential candidate genes. On BTA 27, we identified the GINS complex subunit 4 (*GINS4*) gene, the glycerol-3-phosphatase acetyltransferase 4 (*GPAT4*) gene, and the thyroid hormone receptor β (*THRB*) gene.

The glutamate metabotropic receptor 1 (*GRM1*) gene is related to the calcium signaling, neuroactive ligand-receptor interaction, and phospholipase D signaling pathway, being associated with milk yield in dairy cattle (Table 4). Furthermore, the angiopoietin 4

(*ANGPT4*) gene and the thyroid hormone receptor β (*THRB*) gene are located in four pathways previously described for milk fat or protein synthesis in dairy cows (Table 4).

Table 3. Potential candidate genes related to the identified sequence variants (SVs) significantly associated with the udder health indicator trait somatic cell score (SCS) and with the metabolic health and stability indicator trait fat-to-protein ratio (FPR) in DSN.

BTA	Gene Position [1]	No. of SVs within/Close to Gene [2]	Position of Maximum Association (*p*-Value)	rs Number of Maximum Association	Gene Name
SCS					
1	5,689,903–5,690,709	0/1	5,693,321 (2.49×10^{-6}) *	-	*KRTAP24-1*
	5,770,225–5,772,218	0/1	5,693,321 (2.49×10^{-6}) *	-	*CLDN8*
9	15,242,718–15,373,861	1/0	15,301,365 (2.76×10^{-6}) *	-	*SENP6*
11	101,323,401–101,414,345	0/1	101,435,611 (2.06×10^{-6}) *	rs211669575	*NUP214*
	101,450,754–101,465,047	0/1	101,435,611 (2.06×10^{-6}) *	rs211669575	*FAM78A*
	101,696,993–101,834,040	2/0	101,761,967 (1.30×10^{-7}) **	rs137783421	*RAPGEF1*
13	61,874,277–61,955,381	0/1	61,959,686 (1.04×10^{-6}) *	rs211178277	*NOL4L*
23	42,945,450–43,017,711	1/0	42,9749,92 (4.20×10^{-7}) *	rs876215027	*RANBP9*
25	6,224,841–6,638,491	2/45	6,654,595 (2.58×10^{-7}) *	rs136166815	*RBFOX1*
FPR					
9	83,287,300–83,713,833	1/0	83,557,515 (3.16×10^{-7}) *	rs467161057	*GRM1*
11	6,481,883–6,555,244	0/4	6,464,004 (3.06×10^{-7}) *	rs134892674	*ENSBTAG00000054755*
12	11,675,264–12,071,822	15/1	11,736,920 (1.39×10^{-7}) **	rs439994366	*VWA8*
	12,256,326–12,294,096	0/1	12,195,520 (2.71×10^{-6}) *	rs209487147	*ENSBTAG00000053271*
	12,410,713–12,461,551	0/9	12,496,948 (1.49×10^{-6}) *	rs208793423	*AKAP11*
13	58,456,060–58,457,612	0/2	58,382,430 (2.12×10^{-6}) *	rs42020993	*ENSBTAG00000054668*
	60,191,141–60,234,885	1/0	60,228,471 (2.65×10^{-6}) *	-	*ANGPT4*
	61,107,684–61,147,486	2/0	61,145,994 (8.51×10^{-7}) *	-	*HM13*
	62,381,695–62,408,299	0/1	62,434,039 (1.91×10^{-6}) *	-	*BPIFB4*
	62,503,682–62,514,220	0/1	62,434,039 (1.91×10^{-6}) *	-	*BPIFA2A*
	62,645,078–62,655,447	0/1	62,659,594 (2.82×10^{-6}) *	-	*BPIFA2B*
	62,676,036–62,685,906	0/1	62,659,594 (2.82×10^{-6}) *	-	*ENSBTAG00000031373*
	65,100,388–65,148,653	2/0	65,123,915 (1.50×10^{-6}) *	rs109380861	*CNBD2*
	72,891,431–72,905,729	1/0	72,897,116 (1.46×10^{-6}) *	rs137243257	*TTPAL*
	75,255,202–75,339,020	2/0	75,261,912 (2.42×10^{-6}) *	rs137115876	*SLC13A3*
19	30,045,106–30,177,093	0/1	30,186,388 (2.15×10^{-6}) *	-	*ENSBTAG00000049618*
27	36,454,819–36,470,324	0/12	36,520,069 (6.89×10^{-7}) *	rs211250281	*GINS4*
	36,522,605–36,539,773	1/12	36,520,069 (6.89×10^{-7}) *	rs211250281	*GPAT4*
	41,461,703–41,896,044	1/0	41,614,645 (2.30×10^{-6}) *	rs135231909	*THRB*

[1] Gene position (start-end) in ENSEMBL build on *Bos taurus* genome assembly ARS-UCD1.2; [2] number of associations that reached the Bonferroni-corrected genome-wide significance threshold (*p*Bonf) or the suggestive significance threshold (*p*Sug) based on the position of the identified candidate gene ± 100 kb up- and downstream; * above *p*Sug; ** above *p*Bonf; BTA = *Bos taurus* chromosome.

3.3. Endoparasite Infection Traits

3.3.1. Gastrointestinal Nematode Infections

The associations between each SV and RES-GIN are shown in the Manhattan plot in Figure 3A. The genomic inflation factor λ was 1.019. We identified 47 SVs (thereof one indel) above *p*Bonf and 270 SNPs and 30 indels above *p*Sug on 14 chromosomes associated with RES-GIN, respectively (Table S3). Most significantly and suggestively associations were identified on BTA 2 (n = 165) and on BTA 24 (n = 89). The strongest associations were identified on BTA 8 with seven markers surpassing *p*Bonf. Additionally, 29 SNPs and one indel on BTA 24 and 10 SNPs on BTA 26 surpassed *p*Bonf. The majority of associations surpassing *p*Sug was found on BTA 2 (n = 165) and BTA 24 (n = 59) (Table S3). Significantly and suggestively associated SVs were annotated to 46 potential candidate genes (Table 5; Table S4). The most interesting candidate genes with at least 10 SVs located in the gene were the contactin-associated protein-like 5 (*CNTNAP5*) gene and the microtubule-associated protein 2 (*MAP2*) gene on BTA 2. On BTA 24, we identified 23 SVs (22 SNPs and one indel) located in the α kinase 2 (*ALPK2*) gene and 58 further SVs (thereof one indel) in close distance (<19.2 kb) to *ALPK2* (Figure 3B). Both genes NEDD4 like E3 ubiquitin protein ligase (*NEDD4L*) and MALT1 paracaspase (*MALT1*) flank *APLK2*, forming an association cluster (bp: 57,593,275–57,876,049) with a total number of 79 associated SNPs and three associated indels (Figure 3B). In addition, we detected ten SNPs in the cartilage acidic protein 1 (*CRTAC1*) gene on BTA 26.

As shown in Table 4, the gene activin receptor type 1C (*ACVR1C*) on BTA 2 is related to the cytokine–cytokine interaction pathway and to the TGF-ß signaling pathway, both involved in host immune response mechanisms. The adenylate cyclase 1 (*ADCY1*) on BTA 4 is part of the chemokine signaling and estrogen signaling pathways, previously identified for GIN infections in cattle (Table 4). Further pathways involved in host immune response were annotated to the *MALT1* and PH domain leucine rich repeat protein phosphatase 1 (*PHLPP1*) gene on BTA 24. These pathways include the B cell receptor signaling pathway, C-type lectin receptor signaling pathway, NF-kappa B signaling pathway, PI3K-Akt signaling pathway, and the T cell receptor signaling pathway (Table 4).

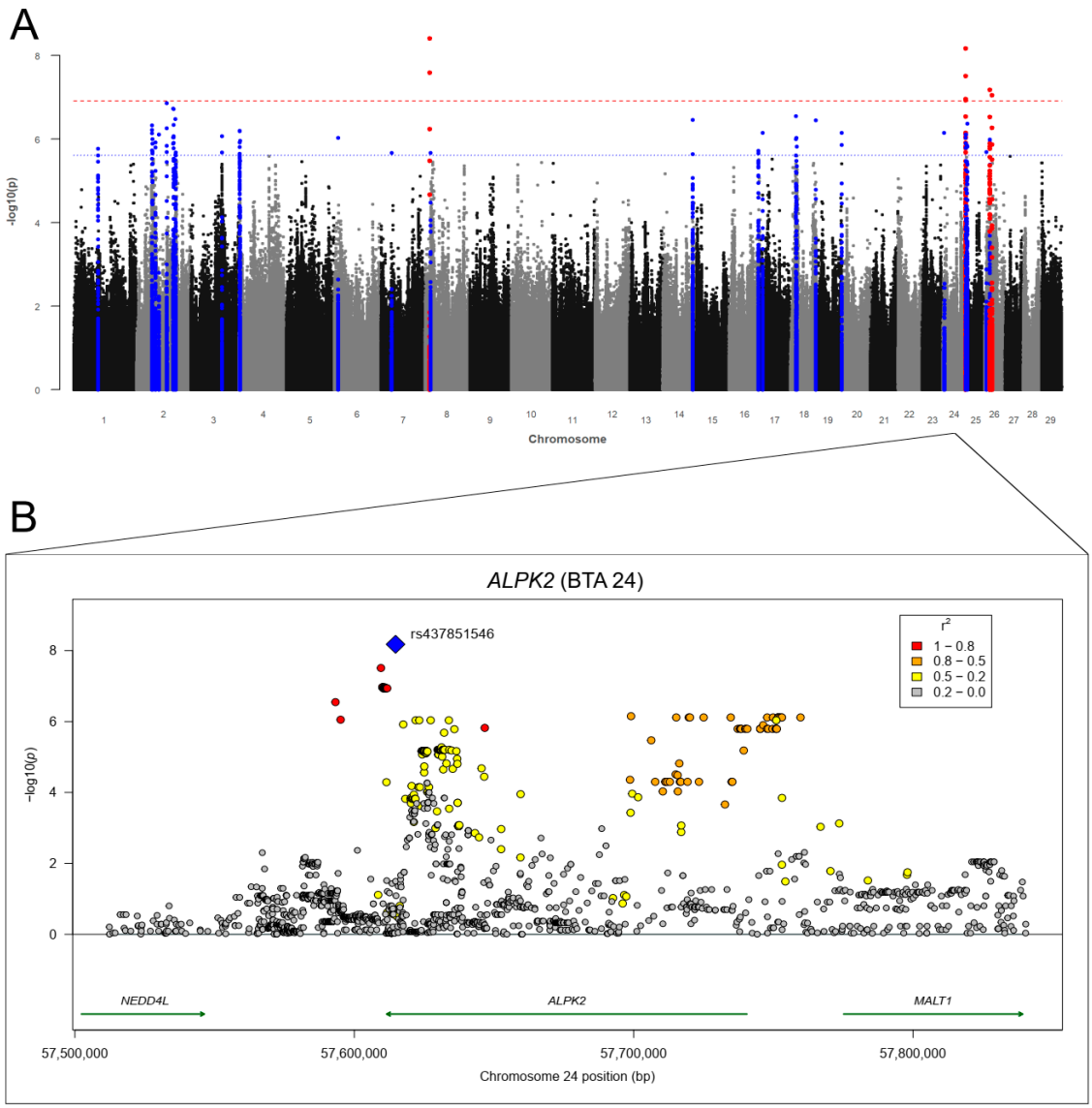

Figure 3. (**A**) Manhattan plot for −log10 *p*-values of marker effects for residuals of fecal egg counts of gastrointestinal nematodes (RES-GIN). Markers above the genome-wide significance threshold *p*Bonf (red dashed line) are highlighted in red and markers above the suggestive significance threshold *p*Sug (blue dashed line) are highlighted in blue. Markers within a distance of 125 kb up- and downstream of the significantly or suggestively associated SVs are highlighted in red or blue, respectively. (**B**) Regional association plot for the α kinase 2 (*ALPK2*) gene on BTA 24 with corresponding flanking regions (+/−100,000 bp). The SNP rs437851546 (blue square) was the highest associated SNP ($p = 6.63 \times 10^{-9}$) for RES-GIN on BTA 24. Circles show GWAS *p*-values, with different colors indicating linkage disequilibrium (LD): red: LD 0.8 to 1.0, orange: LD 0.5 to 0.8. yellow: LD 0.2 to 0.5, gray: LD 0.0 to 0.2. Genes with a green arrow pointing to the right are located on the forward strand, genes with an arrow pointing to the left are located on the reverse strand.

3.3.2. Liver Fluke (*Fasciola hepatica*) Infections

The Manhattan plot for the GWAS with RES-FH is shown in Figure 4A. The genomic inflation factor λ was 1.025. We identified 80 SVs (69 SNP and 11 indels) above *p*Bonf on BTA 3, 4, 7, 8, 9, 10, 11, 14, 15, 23, and 27 (Table S3). The strongest association ($p = 3.66 \times 10^{-12}$) was detected on BTA 7 (bp: 109,376,871). The GWAS revealed 224 further SVs (200 SNPs and 24 indels) above *p*Sug on all BTA despite of BTA 12, 18, 20, 22, and 25 (Table S3). The highest number of associations (above *p*Bonf and *p*Sug) were identified on BTA 4 ($n = 51$) and BTA 7 ($n = 64$). We identified 62 potential candidate genes for RES-FH (Table 5; Table S5). The region between 117.0 and 119.4 Mb on BTA 4 harbored a cluster of eight potential candidate genes (Table S5). The KIAA0825 (*KIAA0825*) gene and the mannosidase α class 2A member 1 (*MAN2A1*) gene were the most interesting candidate genes on BTA 7 with six SNPs in the *KIAA0825* gene and five SNPs in the *MAN2A1* gene plus five flanking SNPs, respectively. On BTA 27, 11 significantly associated SVs (thereof one indel) are located in the fibroblast growth factor receptor 1 (*FGFR1*) gene, and 12 significantly associated SVs (thereof on indel) are flanking around the gene (Figure 4B).

The identified candidate genes were annotated to 12 pathways potentially involved in host-parasite interactions (Table 4). The phospholipase C β 1 (*PLCB1*) gene on BTA 13 (three SNPs located in the gene) was assigned to four biological pathways potentially involved in host–*Fasciola hepatica* interactions (Table 4). We identified two suggestively associated SNPs in the protein tyrosine kinase 2 (*PTK2*) gene in BTA 14, which is part of the chemokine signaling, leukocyte transendothelial migration, and natural killer cell mediated cytotoxicity pathways (Table 4). The four genes interleukin 21 (*IL21*), prolactin-related protein VII (*PRP-VII*), prolactin-related protein IX (*PRP9*) and SMAD family member 4 (*SMAD4*) were related to the cytokine–cytokine receptor interaction, JAK-STAT signaling, PI3K-Akt signaling, TGF-β signaling, and Th17 cell differentiation pathways, respectively (Table 4).

3.3.3. Bovine Lungworm (*Dictyocaulus viviparus*) Infections

The Manhattan plot for the GWAS with RES-DV is shown in Figure 5A. The genomic inflation factor λ was 1.064. We identified 332 SVs (311 SNPs and 21 indels) above *p*Bonf on BTA 2, 5, 6, 9, 15, 20, and 24 (Table S3). The most (n = 208) and strongest associations above *p*Bonf were identified on BTA 24. The GWAS revealed 823 further SVs (761 SNPs and 62 indels) above *p*Sug on all BTA except for BTA 1, 12, 18, 19, 22, and 25. The highest number of associations above *p*Sug were detected on BTA 9 (n = 231) and on BTA 24 (n = 214). We identified 92 potential candidate genes for RES-DV (Table 5; Table S6). The region between 94.9 and 97.7 Mb on BTA 5 includes a cluster of seven potential candidate genes (Table S6). The genes including the largest number of associations within this cluster were the Rho GDP dissociation inhibitor β (*ARHGDIB*) gene, the glutamate ionotropic receptor NMDA type subunit 2B (*GRIN2B*) gene, and the epithelial membrane protein 1 (*EMP1*) gene. We identified the contactin 5 (*CNTN5*) gene on BTA 15 with 47 significantly associated SVs (thereof four indels) located in the gene and 56 associated SVs (thereof three indels) in the flanking region of the gene. The region between 68.2 and 68.4 Mb on BTA 21 includes a cluster of five potential candidate genes. The strongest association within this cluster ($p = 1.62 \times 10^{-7}$; bp 68,299,895) was located in the kinesin light chain 1 (*KLC1*) gene, with a total number of 59 SVs (49 SNPs and 10 indels) located in the gene (Table S6). On BTA 24, we identified a cluster of five genes including a large number of associations in the region between 7.0 and 8.4 Mb. These genes include the suppressor of cytokine signaling 6 (*SOCS6*) gene, the rotatin (*RTTN*) gene, the CD226 molecule (*CD226*) gene, the docking protein 6 (*DOK6*) gene, and the coiled-col domain containing 102B (*CCDC102B*) gene. Within this cluster, we identified the largest number of associations for *DOK6* (Figure 5B). A further potential candidate gene was laminin subunit α 3 (*LAMA3*) with 34 SVs (thereof two indels) located in the gene.

The genes *SOCS6*, *CD226*, and *LAMA3* were related to the JAK-STAT signaling, cell adhesion molecules, and PI3K-Akt signaling pathways, respectively (Table 4). The bone

morphogenetic protein receptor type 1B (*BMPR1*) gene on BTA 6 was annotated to the cytokine–cytokine receptor interaction and to the TGF-β signaling pathway, which are involved in the secretion and up- and downregulation of cytokines during helminth infections (Table 4). In addition, we identified the JAK-STAT signaling pathway and the PI3K-Akt signaling pathway for the cyclin D3 (*CCND3*) gene.

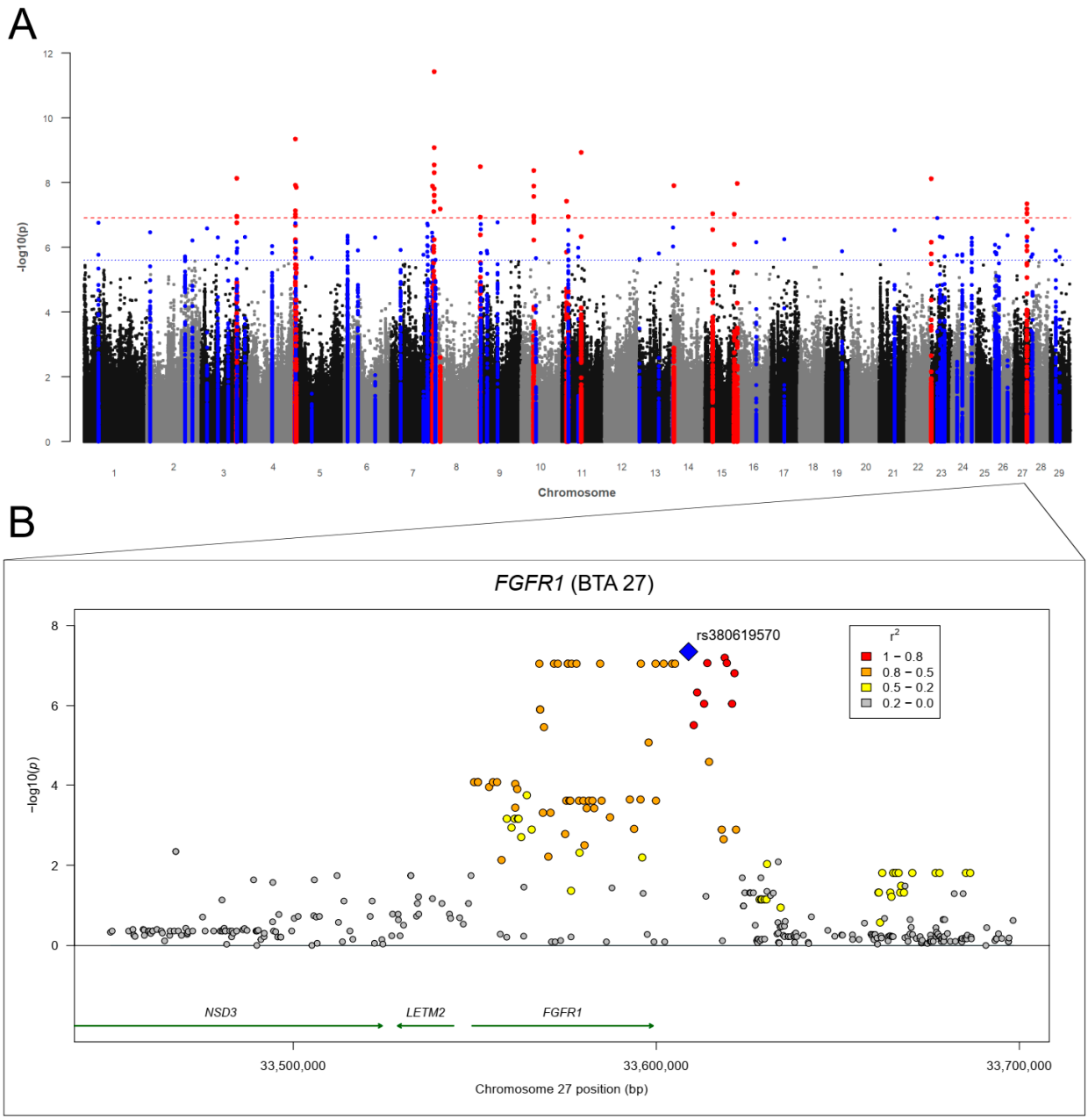

Figure 4. (**A**) Manhattan plot for −log10 *p*-values of marker effects for residuals of fecal egg counts of *Fasciola hepatica* (RES-FH); markers above the genome-wide significance threshold *p*Bonf (red dashed line) are highlighted in red and markers above the suggestive significance threshold *p*Sug (blue dashed line) are highlighted in blue. Markers within a distance of 125 kb up- and downstream of significantly or suggestively associated SVs are highlighted in red or blue, respectively. (**B**) Regional association plot for the fibroblast growth factor receptor 1 (*FGFR1*) gene on BTA 27 with corresponding flanking regions (+/−100,000 bp). The SNP rs380619570 (blue square) was the highest associated SNP ($p = 4.48 \times 10^{-8}$) for RES-FH on BTA 27. Circles show GWAS *p*-values, with different colors indicating linkage disequilibrium (LD): red: LD 0.8 to 1.0, orange: LD 0.5 to 0.8. yellow: LD 0.2 to 0.5, gray: LD 0.0 to 0.2. Genes with a green arrow pointing to the right are located on the forward strand, genes with an arrow pointing to the left are located on the reverse strand.

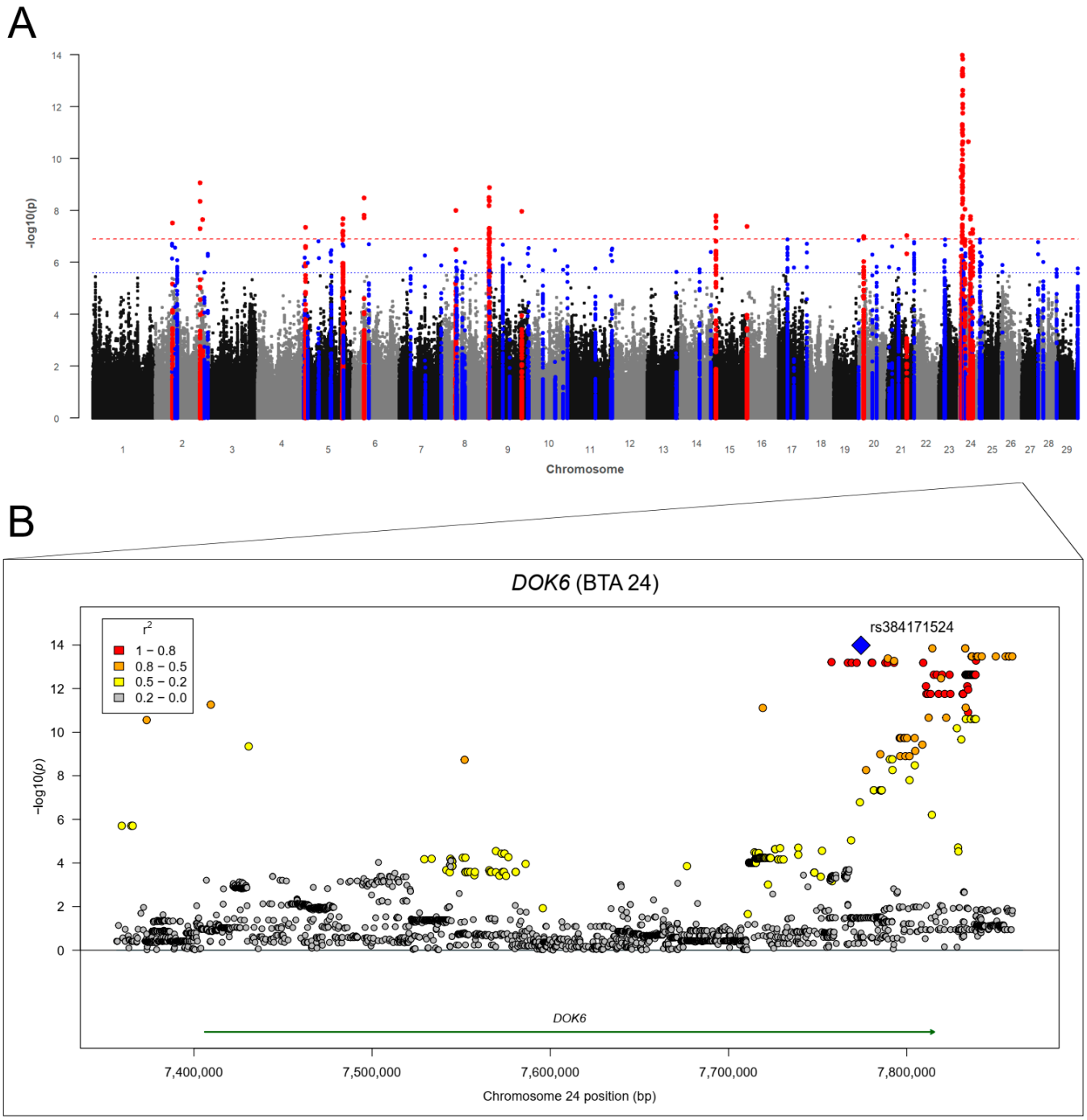

Figure 5. (**A**) Manhattan plot for −log10 *p*-values of marker effects for residuals for fecal larvae counts of *Dictyocaulus viviparus* (RES-DV); markers above the genome-wide significance threshold *p*Bonf (red dashed line) are highlighted in red and markers above the suggestive significance threshold *p*Sug (blue dashed line) are highlighted in blue. Markers within a distance of 125 kb up- and downstream of the significantly or suggestively associated SVs are highlighted in red or blue, respectively. (**B**) Regional association plot for the docking protein 6 (*DOK6*) gene on BTA 24 with corresponding flanking regions (+/−100,000 bp). The SNP rs384171524 (blue square) was the highest associated SNP ($p = 1.04 \times 10^{-14}$) for RES-DV. Circles show GWAS *p*-values, with different colors indicating linkage disequilibrium (LD): red: LD 0.8 to 1.0, orange: LD 0.5 to 0.8. yellow: LD 0.2 to 0.5, gray: LD 0.0 to 0.2.

Table 4. Pathways (sorted alphabetically) related to the identified candidate genes associated with fat-to-protein ratio (FPR), somatic cell score (SCS), and endoparasite infection traits (RES-GIN, RES-FH, RES-DV), selected from the DAVID and KEGG databases and previously described in literature to be associated with the corresponding trait.

Pathway	KEGG Entry	Trait [1]	Candidate Gene (BTA)	Possible Association of Pathway with Trait according to Literature
B cell receptor signaling pathway	bta04662	RES-GIN	*MALT1* (24)	Involvement of B cells in immune response to gastrointestinal nematodes in ruminants [48].
Calcium signaling pathway		FPR	*GRM1* (9)	Pathway associated with milk fat content in Holstein cattle [49].
Cell adhesion molecules	bta04514	SCS	*CLDN8* (1)	Identification of the cell adhesin molecules pathway for *Escherichia coli*-induced mastitis in RNAseq analysis [50].
		RES-DV	*CD226* (24)	Cell adhesion molecules pathway identified for *F. hepatica* infections in cattle [22].
cGMP-PKG signaling pathway	bta04022	RES-FH	*PLCB1* (13), *PRKG1* (26), *ADRB3* (27)	Cyclic GMP (cGMP) is an intracellular messenger that mediates the action of nitric oxide, which is increasingly produced by leukocytes during *F. hepatica* infections [51].
Chemokine signaling pathway	bta04062	RES-GIN	*ADCY1* (4)	Pathway was associated with resistance to gastrointestinal nematode infections in Angus cattle [52].
		RES-FH	*PLCB1* (13), *PTK2* (14)	Pathway was associated with *Fasciola gigantica* infections in Buffalo via proteomics analysis [53]; pathway associated with *F. hepatica* infections in mice [54].
C-type lectin receptor signaling pathway	bta04625	RES-GIN	*MALT1* (24)	C-type lectin receptors are involved in innate and adaptive immunity to pathogens [46]; helminth C-type lectins are involved in host–parasite interactions [55].
Cytokine–cytokine receptor interaction	bta04060	RES-GIN	*ACVR1C* (2)	Pathway was associated with resistance to GIN infections in Angus cattle [52] and in German Black Pied dairy cattle [22]; up- and downregulation of cytokines as immune mechanism in cattle in response to helminth infections [56,57].
		RES-DV	*BMPR1B* (6)	
		RES-FH	*IL21* (17), *PRP9* (23), *PRP-VII* (23)	
Estrogen signaling pathway	bta04915	RES-GIN	*ADCY1* (4)	Increase in reproduction rate of helminths as a result of increasing metabolism of 17-ß-estradiol in the host [58]
		RES-FH	*PLCB1* (13), *KRT31* (19), *KRT34* (19)	
JAK-STAT signaling pathway	bta04630	RES-FH	*IL21* (17), *PRP9* (23), *PRP-VII* (23)	*F. hepatica* excretory-secretory antigens suppress multiple proteins participating in the JAK-STAT signaling pathway in mice [59].
		RES-DV	*CCND3* (23), *SOCS6* (24)	JAK-STAT pathway is the principal signaling mechanism for cytokines involved in *D. viviparus* infections [46,60].
Leukocyte transendothelial migration	bta04670	RES-FH	*PTK2* (14)	Pathway was associated with *Fasciola gigantica* infections in buffalo via proteomics analysis [53].
		SCS	*CLDN8* (1)	Identification of the Leukocyte transendothelial migration pathway for *Escherichia coli*-induced mastitis in RNAseq analysis [50].
MAPK signaling pathway	bta04010	FPR	*ANGPT4* (13)	Pathway associated with milk fat traits in dairy cattle [61,62].
Natural killer cell mediated cytotoxicity	bta04650	RES-FH	*PTK2* (14)	Natural killer cells are lymphocytes of the innate immune response involved in host defense against infections with parasites [46]; cytotoxic natural killer cells were involved in early stage of infection by *F. hepatica* in rats [63]; pathway identified for *F. hepatica* infections in mice [54].
Neuroactive ligand-receptor interaction	bta04080	FPR	*GRM1* (9), *THRB* (27)	Pathway associated with milk fat traits in dairy cattle [49]

Table 4. *Cont.*

Pathway	KEGG Entry	Trait [1]	Candidate Gene (BTA)	Possible Association of Pathway with Trait according to Literature
NF-kappa B signaling pathway	bta04064	RES-GIN	*MALT1* (24)	Family of transcription factors regulating genes involved in immunity [46]; pathway associated with *F. hepatica* infections in sheep [64].
NOD-like receptor signaling pathway	bta04621	RES-FH	*PLCB1* (13)	Family of pattern recognition receptors responsible for various pathogens and generating innate immune response [46].
Phospholipase D signaling pathway	bta04072	FPR	*GRM1* (9)	Phospholipase D is an essential enzyme for the production of phosphatidic acid, a key intermediate in milk fat synthesis during lactation [61].
PI3K-Akt signaling pathway	bta04151	RES-GIN	*PHLPP1* (24)	Pathway has important functions in cellular immune response [46].
		RES-FH	*PTK2* (14), *PRP9* (23), *PRP-VII* (23), *EIF4EBP1* (27), *FGFR1* (27)	
		RES-DV	*CCND3* (23), *LAMA3* (24)	
		FPR	*ANGPT4* (13)	Pathway associated with milk fat traits in dairy cattle [62].
Rap 1 signaling pathway	bta04015	FPR	*ANGPT4* (13)	Pathway associated with milk fat traits in dairy cattle [61].
Ras signaling pathway	bta04014	FPR	*ANGPT4* (13)	Pathway associated with milk fat traits in dairy cattle [61].
T cell receptor signaling pathway	bta04660	RES-GIN	*MALT1* (24)	Involvement of B cells in immune response to gastrointestinal nematodes in ruminants [48].
TGF-β signaling pathway	bta04350	RES-GIN	*ACVR1C* (2)	TGF-ß involved in host immune response during *Ostertagia ostertagi* (GIN species) infections [65]; pathway is associated in host–*F. hepatica* interactions: binding of *F. hepatica* growth factors to host TGF- ß receptors and triggering SMAD (Sma-and Mad-related proteins) in host leukocytes [56].
		RES-FH	*SMAD4* (24)	
		RES-DV	*BMPR1B* (6)	
Th17 cell differentiation	bta04659	RES-FH	*IL21* (17), *SMAD4* (4)	*F. hepatica*-induced TGF-ß suppresses Th17 responses in infected mice [66].

[1] RES-GIN = residuals for fecal egg counts of gastrointestinal nematodes; RES-FH = residuals for fecal egg counts of liver flukes (*Fasciola hepatica*); RES-DV = residuals for fecal larvae counts of bovine lungworms (*Dictyocaulus viviparus*).

Table 5. Potential candidate genes (sorted alphabetically) related to the identified sequence variants (SVs) associated with residuals for fecal egg counts of gastrointestinal nematodes (RES-GIN), residuals for fecal egg counts of liver flukes (RES-FH), and residuals for fecal larvae counts of bovine lungworms (RES-DV) in DSN. Genes with at least one associated variant in the respective gene are marked in bold. A gene was considered a candidate gene if at least one SV above *p*Sug was positioned in the respective gene and/or within 100 kb up- and downstream of the gene.

BTA	RES-GIN	RES-FH	RES-DV
1	*ENSBTAG00000048985*	-	-
2	*ACVR1*, ***ACVR1C***, ***CNTNAP5***, *CPS1*, *ENSBTAG00000054211*, *ENSBTAG00000040367*, *ENSBTAG00000051630*, ***GPD2***, *KIF5C*, ***LANCL1***, ***LRP1B***, *LYPD6B*, ***MAP2***, *NR4A2*, ***UNC80***	***CRYGA***, *CRYGB*	*ENSBTAG00000033143 **, *ENSBTAG00000050185*, *ENSBTAG00000049959*, *ENSBTAG00000055116 FAM124B **, ***GALNT13***, *GPR55 **, *SPATA3 **
3	*PDE4B*	***CDC14A***, *ENSBTAG00000037539*, ***KCNN3***, ***TTC4***	-
4	*ADCY1*	***CNPY1 ****, *HTR5A **, ***LMBR1***, *MNX1*, *NCAPG2 **, *PTPRN2 **, *RBM33 **, *UBE3C*	***ESYT2***
5	-	-	*ARHGDIB **, *ASCL1*, *DRAM1*, ***EMP1 ****, ***FAM234B ****, *GNPTAB*, ***GRIN2B ****, ***LRP6***, ***NELL2***, *PDE6H **, *SLC38A1*, *TPH2 **, ***UTP20***, *YAF2*, *ZCRB1*
6	***NDST4***	*GPRIN3*	***BMPR1B ****, ***KCNIP4***, *PDLIM5 **
7	***MEGF10***	***ADAMTS19***, ***AADAT***, ***KIAA0825***, ***MAN2A1***, ***TMEM232 ****	*HAND1*
8	*ENSBTAG00000052065 **	-	*ALDH1A1*, *ENSBTAG00000052698*
9	-	***LCA5***, *SH3BGRL2*	*ENSBTAG00000055087 **, *SASH1 **, ***TBXT***, *UST **
10	-	*ENSBTAG00000050159 **, ***TMEM87A***	*ABCD4*, *DDHD1*, *FERMT2*, *VRTN*
11	-	***EXOC6B ****, *FAM98A **, ***REEP1****	*ENSBTAG00000009599*, ***LCN10***, ***NELFB***, ***RALGDS***, *TOR4A*
13	-	***PLCB1***, ***HAO1***	*SPINT3*
14	***CNBD1***	***PTK2***	***RALYL***
15	-	***ARHGAP20 ****, ***ALX4 ****, *FAM111B **, *GLYATL2 **	***CNTN5****
16	***SYT14***	*ENSBTAG00000053468*	*SOX13 **
17	-	*IL21*	*ENSBTAG00000033967*, *ENSBTAG00000055004*, *RAB36*, ***RSPH14***
18	***APRT***, ***CBFA2T3***, *DNAJA2*, ***GALNS***, *ENSBTAG00000048593*, *ENSBTAG00000048735*, *NETO2*, ***PHKB***	-	-
19	***HELZ***	*KRT31*, *KRT34*	
20	-	-	***DAB2***, *ENSBTAG00000047333*, *ENSBTAG00000000617*, ***ENSBTAG00000049964 ****
21	-	-	***APOPT1***, *ENSBTAG00000035184*, *ENSBTAG00000052298*, *ENSBTAG00000003957*, ***KLC1***, *MCTP2*, ***PPP1R13B***, *SNRPA1*, ***XRCC3***, ***ZFYVE21***

Table 5. *Cont.*

BTA	RES-GIN	RES-FH	RES-DV
23		*ENSBTAG00000048946*, *ENSBTAG00000053227*, *KHDRBS2* *, *PRP9*, *PRP-VII*, ***TFAP2D***, ***TRERF1***	*CCND3*, ***FOXP4***, *LRFN2*, *TAF8*, *TRERF1*, *UBR2*
24	***ALPK2*** *, *MALT1*, *NEDD4L* *, ***PHLPP1***, ***TSHZ1***, ***ZCCHC2***, *ZNF532*	***ELAC1***, *MEX3C*, *PIK3C3*, *SMAD4*	***CD226*** *, ***CDH19***, *C24H18orf63*, ***CYB5A*** *, *DIPK1C*, ***DOK6*** *, ***DTNA*** *, *ENSBTAG00000049503*, ***LAMA3*** *, *NPC1*, ***OSBPL1A***, ***PIK3C3*** *, ***RTTN*** *, ***RIOK3*** *, *SOCS6* *, ***TMEM241***, *WDR7*
26	*BCL2*, ***CRTAC1*** *, *ENSBTAG00000048707*, *IDE* *, ***KIF11***, ***PCDH15***, *TNKS2*	***ATRNL1***, *EXOC6*, *HHEX*, ***PCDH15***, ***PRKG1***	-
27	-	*ADRB3*, *EIF4EBP1*, ***FGFR1*** *, *ZNF385D*	***THRB***
28	-	*ACTA1*, *ENSBTAG00000048654*	***ARHGAP22***
29	-	*PRMT3*, *SLC6A5*	*ENSBTAG00000008274*, *ENSBTAG00000050398*

* Genes including at least one associated variant above the genome-wide significance threshold (pBonf = 1.22×10^{-7}).

4. Discussion

This is the first study investigating genome-wide associations and candidate genes for functional traits in the local endangered DSN breed based on imputed WGS data. In our study, we used a quite large reference population of 304 sequenced DSN for imputation, resulting in a high imputation accuracy of 97.04%. Brondum et al. [67] and Mao et al. [68] demonstrated higher imputation accuracies for WGS data when using multi-breed compared to single-breed reference populations. Lower imputation accuracies of up to 95% were achieved in cattle studies using multi-breed reference panels with 242 to 1577 sequenced cattle [5,67,69]. Korkuć et al. [39] compared different imputation strategies in DSN and suggested a large reference panel of the same breed instead of a multi-breed (DSN together with HF) approach. Achieving an imputation accuracy above 97%, we expected a high power in our GWAS to identify genomic loci associated with complex functional traits in DSN.

We focused on fertility, udder health, metabolic health and stability indicator traits, and endoparasite infection traits since these are highly relevant for the conservation of the breed [10,24]. For CTFS, we identified SVs on BTA 12, 13, 15, and 28 with the largest number of associations on BTA 13. In Holstein cows, genomic loci for CTFS were detected on BTA 13, too [70], but associated regions on BTA 13 did not overlap with our findings. We identified *ARHGAP21* on BTA 13 as one main candidate gene for CTFS with five SNPs located in the gene or within a region up to 42.3 kb downstream of the gene. The *ARHGAP21* gene was reported to be involved in post-partum anestrus in tropical beef cattle [71]. Rosa et al. [72] demonstrated in a mouse model that *ARHGAP21* is involved in insulin secretion. Insulin plays a major role in the regulation of energy balance and reproductive functions in dairy cattle [73]. Gong et al. [74] showed that diets with higher insulin levels significantly reduced the intervals from calving to first ovulation in dairy cows. Thus, *ARHGAP21* might impact CTFS by regulating important hormone functions in dairy cattle. For NR56, our GWAS revealed significant SVs on BTA 8, 11, 13, and 20. Holmberg and Andersson-Eklund et al. [75] identified SNPs on BTA 11 and on BTA 20 associated with NR56 in Swedish dairy cows. We identified three genes for NR56, of which two were previously reported to be involved in cattle female fertility. The *ZNF462* gene was identified as a transcription factor involved in fertility in Brangus heifers [76]. Moore et al. [77] showed reduced expressions of *ZNF462* in the corpus luteum of low fertility HF cows compared to the high fertility control group. Kiser et al. [78] described *MARCH11* on BTA 20 as a top locus for conception rate at first service and repeated artificial insemination in Holstein heifers. Furthermore, *MARCH11* was recently associated with endometriosis in HF cows in our studies [79]. Hence, our detected genes influencing fertility in DSN may have similar functions for fertility in other cattle breeds. This observation is in accordance with Korkuć et al. [24], who found a high overlap of genomic regions in DSN and HF for milk production traits, although they identified no common genes for DSN and HF. Although we identified genes for CTFS and NR56 in DSN which were previously described in other breeds, the same gene can be differentially expressed in different breeds, which might phenotypically explain improved fertility in DSN. For example, Timperio et al. [80] showed transcriptomic level disparities in liver tissues in two closely related *Bos taurus* breeds, which contribute to large physiological differences. Moreover, Lehnert et al. [81] demonstrated that gene expression levels for growth-related genes can differ in different breeds, although the genes are involved in growth in cattle generally.

For FPR, we detected a cluster of ten closely related genes on BTA 13. This finding is in contrast to the GWAS by Korkuć et al. [24], where no SNP reached the significance threshold on BTA 13 for milk fat and protein traits in a population of 1816 50K genotyped DSN. However, similar to our findings, the authors detected SVs on BTA 9, 11, 12, and 27 associated with milk fat content or milk fat yield in DSN. Interestingly, Klein et al. [27] identified similar associations for FPR in Holstein cows on BTA 9, 13, 14, and 27, indicating a strong genomic relatedness of HF and DSN. The *DGAT1* gene on BTA14 was highly associated with FPR in Holstein cows [27]. Furthermore, Jaeger [10] identified *DGAT1* for

fat content in a multi-breed GWAS including a quite large number of DSN cows. However, although genomic relatedness between DSN and HF is high, we detected no associations in the *DGAT1* gene for FPR in DSN, which coincides with the observation by Korkuć et al. [24] for other milk production traits in DSN. We identified the glycerol-3-phosphatase acyltranserase 4 (*GPAT4*) gene on BTA 27 as a potential candidate gene for FPR. The *GPAT4* gene is well described as a locus involved in lipid metabolism in HF cows [82]. The *THRB* gene on BTA 27 was associated with FPR and RES-DV, and might be an interesting candidate gene when aiming for improvements for both traits, i.e., less susceptibility to *D. viviparus* infections and stable FPR, which has been already proven by May et al. [21] quantitative-genetically. Furthermore, we found seven pathways for genes associated with FPR (e.g., MAPK signaling pathway), which were associated with milk production and fat yield in different cattle breeds [62,83]. The overlap of pathways identified for fat yield in previous studies and FPR in our study might be explained by the fact that fat yield is more variable than protein yield, and thus, contributes to a larger variation in FPR.

For the health and metabolic stability indicator traits SCS and FPR, Jaeger [10] estimated genetic correlations lower than 0.80 between two classes of average herd sizes and first calving ages in DSN based on pedigree data, indicating the presence of genotype-by-environment interactions (GxE). Hence, we assume that a large number of genetic loci contribute to GxE for SCS and FPR, which should be explored in ongoing studies. The knowledge of how GxE contributes to the phenotypic variance for adaptive and characteristic traits in DSN is of great importance for the conservation of the breed, since large GxE SNP effects in cattle indicate considerable opportunity to improve environmental resistance and health [84]. For SCS, our GWAS revealed SVs on BTA 1, 9, 11, 13, 23, and 25. Significant marker associations for SCS in other dairy cattle breeds were described on each chromosome [85], indicating the complex genetic architecture of the trait. In the present study, we estimated additive genetic effects by neglecting non-additive dominance effects, which could play a crucial role in the genomic architecture of complex functional traits [86]. As pointed out by Howard et al. [87], dominance decreases when alleles are almost fixed due to inbreeding. Interestingly, a large inbreeding depression was observed for SCS in Holstein cows [88], while inbreeding depression for SCS is currently not present in DSN, possibly due to different selection strategies with only a small number of DSN bulls with high SCS breeding values used for artificial insemination [30]. Substantial dominance effects in DSN for endoparasite infection traits were estimated for an immunological relevant gene [89]. The largest number of associations was detected on BTA 25 within and close to the *RBFOX1* gene, which was previously associated with subclinical ketosis in Jersey cattle [90]. An interesting candidate gene associated with SCS was the *CLDN8* gene on BTA 1, which was found to be increasingly expressed in the bovine mammary glands during reduced milking frequency [91]. Furthermore, *CLDN8* is part of the cell adhesion molecules and leukocyte transendothelial migration pathways, being involved in the recruitment and activation of macrophages during acute *Escherichia coli*-induced mastitis [50,92]. Due to the biological relevance of *CLDN8* in udder health, we recommend a more detailed investigation of polymorphisms and favorable alleles in *CLDN8*, which might contribute to improved udder health in DSN. No overlap was seen between the genomic loci affecting SCS and loci affecting mastitis, which have been identified by Meier et al. [24]. In contrast to our findings for SCS, Meier et al. [25] detected genomic loci on BTA 3, 6, 9, and 26 for mastitis in a population of 1026 50K genotyped DSN, and they identified *BMPR1B* on BTA 6 as the main candidate gene associated with mastitis. The *BMPR1B* gene was described to be involved in immune response to the mastitis pathogen *Staphylococcus aureus* and is related to the cytokine–cytokine interaction and TGF-ß signaling pathway. In the present study, *BMPR1B* was associated with RES-DV assuming that *BMPR1B* plays an important role in both bacterial and endoparasite infections in DSN.

The GWAS revealed a large number of associations and candidate genes for endoparasite infection traits. Interestingly, we identified three genes associated with two endoparasite traits (*TRERF1* on BTA 23, *PIK3C3* on BTA 24, and *PCDH15* on BTA 26). The

PCDH15 gene, identified for GIN and *F. hepatica* infections, was associated with *D. viviparus* infections in our preliminary GWAS in DSN based on imputed high-density (HD) 700K data [22]. Hence, *PCDH15* seems to be involved in overall resistance against endoparasite infections in DSN. Twomey et al. [7] applied a GWAS for antibody responses to endoparasite infections (*F. hepatica*, *Neospora caninum*, and *Ostertagia ostertagi*) in dairy and beef cattle and identified up to 248 SVs and up to 48 QTL for the different traits, which is similar to the number of SNPs and candidate genes for endoparasite infection traits detected in our study. However, the number of detected SVs for fertility and health indicator traits was smaller, which might be due to smaller heritabilities or a more complex genetic architecture for genes involved in host immune response to parasitic disease. Another explanation for the larger number of SVs detected for endoparasite traits addresses the chosen pre-correction approach for endoparasite infection traits in order to consider a larger number of cows examined for endoparasite infections. However, when applying the same approach for fertility and health indicator traits and using the model residuals from pre-corrected traits, the number of significantly and suggestively associated markers did not increase substantially for these traits (results not shown). Furthermore, the genomic inflation factor λ ranged from 1014 to 1088 for all traits and pre-corrected endoparasite traits, indicating that the number of false-positive associations for endoparasite infection traits does not differ to the other investigated traits.

For RES-GIN, a large number of associations surpassing *p*Bonf were detected on BTA 2, 8, 24, and 26. Using the same study design and statistical approach for endoparasite infection traits in DSN imputed to high-density (HD) 700K, we identified associations for RES-GIN on BTA 2, 4, 5, 8, 9, 18, 22, 24, and 26 in a previous GWAS [22]. Interestingly, a large proportion of SNPs detected with HD did not reach the significance threshold in the present study based on WGS data and only one candidate gene (*PHLPP1*) on BTA 24 overlapped in both datasets. The same observation was made by Wu et al. [6] when comparing GWAS results from different SNP panels and WGS data for the same trait. A possible explanation refers to differences in the genomic relationship matrix and LD between low- or high-density SNP panels and WGS data. Moreover, the significance thresholds in the present GWAS differs for different marker densities as a result of Bonferroni multiple testing correction and the calculated effective number of independent SVs. In consequence, markers detected with 700K HD data by May et al. [22] did not reach the suggestive and significance threshold in this study. However, for RES-DV, we detected linkage ($r^2 > 0.6$) between 21 significantly associated SNPs from the GWAS by May et al. [22] and 258 significantly or suggestively associated SVs in the present study. Furthermore, when marker density increases, LD within a region becomes stronger, implying a larger number of SVs around the identified genes affecting the trait of interest. For fecal egg excretion in Angus cattle, Kim et al. [52] detected significant associations on BTA 3, 5, 8, 15, and 27 and, similar to our GWAS, they found genes annotated to the chemokine signaling and cytokine–cytokine interaction pathway. In contrast, Twomey et al. [7] identified the strongest associations on BTA 3, 4, 12, 13, 14, 19, 21, and 23 for antibody response to *O. ostertagi* (most common GIN species in cattle) in a multi-breed GWAS. We identified *ALPK2* on BTA 24 as the gene with the largest number of associations within and in close distance to the gene. Surprisingly, *ALPK2* is not involved in immunological pathways and was not reported as a gene underlying parasite resistance in other animal species. In accordance to the study by May et al. [22], the largest number and strongest associations for RES-GIN and RES-DV were detected on BTA 2 and on BTA 24, which might be explained by the biological relatedness of both nematode traits RES-GIN and RES-DV. However, we detected no overlapping gene for RES-GIN and RES-DV. The gene *FAM234B* on BTA 5 and the gene *DOK6* on BTA 24 were previously reported for *D. viviparus* larvae counts [22]. This is the first GWAS for patent *D. viviparus* infections in cattle based on WGS data and suggesting that the cytokine–cytokine interaction, JAK-STAT signaling, PI3K-signaling, and TGF-β signaling pathways are associated with *D. viviparus* infections in cattle.

For RES-FH, WGS data revealed a large number of candidate genes and related pathways previously described for *F. hepatica* infections via transcriptomic studies [51,53,56,59,63], indicating the advantage of WGS data to identify causal mutations for complex health traits in dairy cattle. We identified the fibroblast growth factor receptor 1 (*FGFR1*) gene on BTA 27 associated with RES-FH. Yu et al. [93] showed that fibroblast growth factor polypeptides are associated with liver fibrosis in mice. Liver fibrosis plays a key role in host immune response to *F. hepatica* infections [64]. Interestingly, TGF-ß plays a crucial role in fibrogenic processes during *F. hepatica* infections and acts as a host receptor for a parasitic growth factor [64,94]. Fu et al. [64] described the *SMAD4* gene on BTA 24 as a key factor involved in fibrogenetic processes during *F. hepatica* infections. Here, we detected *SMAD4* associated with RES-FH and related to the TGF-ß signaling pathway, indicating the power of our GWAS approach to identify genes associated with complex traits in DSN. Together with *IL21*, *SMAD4* is related to the Th17 cell differentiation pathway. Walsh et al. [66] showed that TGF-ß suppresses Th17 immune response in the host via *F. hepatica* infections. Furthermore, we detected six genes for RES-FH related to the cGMP-PKG signaling pathway and the JAK-STAT signaling pathway, which are well known to be involved in *F. hepatica* infections [56,59]. Surprisingly, the genes prolactin-related protein IIV and IX (*PLR9* and *PRP-VII*) on BTA 23 were associated with RES-FH and annotated to the JAK-STAT signaling pathway, too. This finding addresses possible polymorphisms in the identified prolactin genes in DSN associated with both improved milk production and resistance to *F. hepatica* infections. This is of great importance, since genetic correlations between milk production and *F. hepatica* egg excretion was shown to be favorable in quantitative-genetic studies [21,22]. Hence, genetic correlations with milk production traits should be taken into consideration when selecting DSN being well adapted to pasture production systems with improved fertility and metabolic stability and enhanced resistance to endoparasite and udder bacterial infections.

5. Conclusions

Using WGS data, we identified a large number of SVs and potential candidate genes associated with fertility, udder and metabolic health indicator traits, metabolic stability, and endoparasite infection traits in the local DSN population. Such in-depth insights into the genomic particularities enable improved selection strategies in the local DSN breed, especially when defining criteria in the context of conservation of genetic resources. The investigated traits are known as specific adaptive traits in DSN. Genetic improvements of these traits contribute to overall robustness and to possible DSN breed advantages or DSN breed competitiveness over large commercial dairy cattle populations. Selection of DSN animals carrying favorable alleles for the identified SVs is an efficient breeding approach in this regard. For CTFS and NR56, three annotated genes have similar functions in other cattle breeds. *CLDN8* and *RBFOX1* were the most interesting potential candidate genes for SCS. The largest number of associations for FPR were detected on BTA 12 and 27. For endoparasite infection traits, we detected potential candidate genes and related biological pathways, which are involved in host immune response to endoparasite infections. In particular, the identified markers within immunological relevant genes should be used for future genomic selection strategies in DSN, aiming to improve health and adaption to pasture environments.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/genes12081163/s1, Table S1: List of all SVs associated with calving-to-first-service interval (CTFS) and non-return at day 56 (NR56), Table S2: List of all SVs associated with somatic cell score (SCS) and fat-to-protein ratio (FPR), Table S3: List of all SVs associated with residuals of gastrointestinal nematode infections (RES-GIN), *Dictyocaulus viviparus* infections (RES-DV), and *Fasciola hepatica* infections (RES-FH), Table S4: Potential candidate genes with corresponding number of SVs within and close to the gene related to the identified SVs associated with residuals of fecal egg counts for gastrointestinal nematodes, Table S5: Potential candidate genes with corresponding number of SVs within and close to the gene related to the identified SVs associated with residuals of

fecal egg counts for *Fasciola hepatica*, Table S6: Potential candidate genes with corresponding number of SVs within and close to the gene related to the identified SVs associated with residuals of fecal larvae counts for *Dictyocaulus viviparus.*

Author Contributions: Conceptualization of the study, S.K. and G.A.B.; preparation of sequence and SNP chip data, G.B.N. and P.K.; methodology and statistical analyses, M.J.W., K.M. and T.Y.; writing—original draft preparation, M.J.W. and K.M.; writing—review and editing, S.K. and K.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially funded by the German Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany, granted by the Federal Office for Agriculture and Food (BLE; grant number FKZ 2818BM090 and -091).

Institutional Review Board Statement: The whole genotyping process from tissue sampling up to the SNP database development was embedded in the logistics and infrastructure of the cow genotyping activities in Germany, organized by the German Holstein breeding organization and their participating regional breeding organizations and farmers. This activity is the basis to implement a national cow training set for genomic selection. Farmers agreed to participate in this study and to collect small fecal samples for endoparasite determinations. Such fecal collection does not influence the wellbeing of cows. All further data were provided by the respective breeding organizations and milk recording organizations provided (i.e., the traits from official milk recording schemes, genomic data, and pedigree data). Thus, no ethical approval was required for this study.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data supporting the results of this article are included within the article. The raw phenotypic and genotypic data are stored in the databases of the IT company vit Verden and the cluster justHPC from Giessen University (https://www.hkhlr.de/de/cluster/justhpc-giessen, accessed on 2 June 2021). All data can be provided on request.

Acknowledgments: The authors wish to thank the farmers for their cooperation in this study.

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

References

1. Daetwyler, H.D.; Capitan, A.; Pausch, H.; Stothard, P.; van Binsbergen, R.; Brondum, R.F.; Liao, X.; Djari, A.; Rodriguez, S.C.; Grohs, C.; et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Gen.* **2014**, *46*, 858–867. [CrossRef] [PubMed]
2. Das, A.; Panitz, F.; Gregersen, V.R.; Bendixen, C.; Holm, L.-E. Deep sequencing of Danish Holstein dairy cattle for variant detection and insight into potential loss-of-function variants in protein coding genes. *BMC Genom.* **2015**, *16*, 1043. [CrossRef] [PubMed]
3. Schork, A.J.; Thompson, W.K.; Pham, P.; Torkamani, A.; Roddey, J.C.; Sullivan, P.F.; Kelso, J.R.; O'Donovan, M.C.; Furberg, H.; Schork, N.J.; et al. All SNPs are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* **2013**, *9*, e1003449. [CrossRef]
4. Van Binsbergen, R.; Bink, M.C.A.M.; Calus, M.P.L.; van Eeuwijk, F.A.; Hayes, B.J.; Hulsegge, I.; Veerkamp, R.F. Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genet. Sel. Evol.* **2014**, *46*, 41. [CrossRef]
5. Pausch, H.; MacLeod, I.M.; Fries, R.; Emmerling, R.; Bowman, P.J.; Daetwyler, H.D.; Goddard, M.E. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. *Genet. Sel. Evol.* **2017**, *49*, 24. [CrossRef] [PubMed]
6. Wu, X.; Lund, M.S.; Sahana, G.; Guldbrandtsen, B.; Sun, D.; Zhang, Q.; Su, G. Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genet. Sel. Evol.* **2015**, *47*, 50. [CrossRef]
7. Twomey, A.J.; Berry, D.P.; Evans, R.D.; Doherty, M.L.; Graham, D.A.; Purfield, D.C. Genome-wide association study of endoparasite phenotypes using imputed whole-genome sequence data in dairy and beef cattle. *Genet. Sel. Evol.* **2019**, *51*, 15. [CrossRef] [PubMed]
8. Hartwig, S.; Bennewitz, J. Züchterische Aspekte zur Konsolidierung und Weiterentwicklung lokaler Rinderpopulationen. *Züchtungskunde* **2014**, *86*, 19–24.
9. Tijjani, A.; Utsunomiya, Y.T.; Ezekwe, A.G.; Nashiru, O.; Hannotee, O. Genome sequence analysis reveals selection signatures in endangered trypanotolerant West African Muturu cattle. *Front. Genet.* **2019**, *10*, 442. [CrossRef] [PubMed]
10. Jaeger, M.P.H. Phenotypic, Quantitative Genetic and Genomic Characterization of the German Black and White Dual-Purpose Cattle Breed. Ph.D. Thesis, Justus-Liebig University Gießen, Giessen, Germany, 5 October 2018.
11. Boettcher, P.J.; Hoffmann, I.; Baumung, R.; Drucker, A.G.; McManus, C.; Berg, P.; Stella, A.; Nilsen, L.B.; Moran, D.; Naves, M.; et al. Genetic ressources and genomics for adaptation of livestock to climate change. *Front. Genet.* **2015**, *5*, 461. [CrossRef]

12. Mei, C.; Wang, H.; Zhu, W.; Wang, H.; Cheng, G.; Qu, K.; Guang, X.; Li, A.; Zhao, C.; Yang, W. Whole-genome sequencing of the endangered bovine species Gayal (Bos frontalis) provides new insights into its genetic features. *Sci. Rep.* **2016**, *6*, 19787. [CrossRef]
13. Doublet, A.-C.; Restoux, G.; Fritz, S.; Balberini, L.; Fayolle, G.; Hozé, C.; Laloe, D.; Croiseau, P. Intensified use of reproductive technologies and reduced dimensions of breeding schemes put genetic diversity at risk in dairy cattle breeds. *Animals* **2020**, *10*, 1903. [CrossRef]
14. TGRDEU. Zentrale Dokumentation Tiergenetischer Ressourcen in Deutschland, Bundesanstalt für Landwirtschaft und Ernährung. 2019. Available online: https://tgrdeu.genres.de/ (accessed on 24 May 2021).
15. Mügge, B.; Lutz, W.E.; Südbeck, H.; Zelfel, S. *Deutsche Holsteins: Die Geschichte einer Zucht*; Verlag Eugen Ulmer: Stuttgart, Germany, 1999.
16. Grothe, P.O. *Holstein-Friesian: Eine Rasse Geht um die Welt*; Landwirtschaftsverlag: Münster, Germany, 1993.
17. May, K.; Brügemann, K.; König, S.; Strube, C. Patent gastrointestinal nematode infections in organically and conventionally pastured dairy cows and their impact on individual milk fertility parameters. *Vet. Parasitol.* **2017**, *245*, 119–127. [CrossRef]
18. Biedermann, G. Zuchtplanung für die Erhaltung des Alten Schwarzbunten Niederungsrindes. Bundesprogramm Ökologischer Landbau und Andere Formen Nachhaltiger Landwirtschaft. Available online: https://service.ble.de/ptdb/index2.php?detail_id=84928&site_key=141 (accessed on 17 May 2021).
19. Van Dijk, J.; Sargison, N.D.; Kenyon, F.; Skuce, P.J. Climate change and infectious disease: Helminthological challenges to farmed ruminants in temperate regions. *Animal* **2010**, *4*, 377–392. [CrossRef] [PubMed]
20. Charlier, J.; Rinaldi, L.; Musella, V.; Ploeger, H.W.; Chartier, C.; Vineer, H.R.; Hinney, B.; von Samson-Himmelstjerna, G.; Băcescu, B.; Mickiewicz, M.; et al. Initial assessment of the economic burden of major helminth infections to the ruminant livestock industry in Europe. *Prev. Vet. Med.* **2020**, *182*, 105103. [CrossRef]
21. May, K.; Brügemann, K.; Yin, T.; Scheper, C.; Strube, C.; König, S. Genetic line comparisons and genetic parameters for endoparasite infections and test-day milk production traits. *J. Dairy Sci.* **2017**, *100*, 7330–7344. [CrossRef] [PubMed]
22. May, K.; Scheper, S.; Brügemann, K.; Yin, T.; Strube, C.; Korkuc, P.; Brockmann, G.A.; König, S. Genome-wide associations and functional gene analyses for endoparasite resistance in an endangered population of native German Black Pied cattle. *BMC Genom.* **2019**, *20*, 277. [CrossRef] [PubMed]
23. Twomey, A.J.; Carroll, R.I.; Doherty, M.L.; Byrne, N.; Graham, D.A.; Sayers, R.G.; Blom, A.; Berry, D.P. Genetic correlations between endo-parasite phenotypes and economically important traits in dairy and beef cattle. *J. Anim. Sci.* **2018**, *96*, 407–421. [CrossRef] [PubMed]
24. Korkuć, P.; Arends, D.; May, K.; König, S.; Brockmann, G.A. Genomic loci affecting milk production in German Black Pied cattle (DSN). *Front. Genet.* **2021**, *12*, 640039. [CrossRef] [PubMed]
25. Meier, S.; Arends, D.; Korkuć, P.; Neumann, G.B.; Brockmann, G.A. A genome-wide association study for clinical mastitis in the dual-purpose German Black Pied cattle breed. *J. Dairy Sci.* **2020**, *103*, 10289–10298. [CrossRef]
26. Buttchereit, N.; Stamer, E.; Junge, W.; Thaller, G. Genetic parameters for energy balance, fat/protein ratio, body condition score and disease traits in German Holstein cows. *J. Anim. Breed. Genet.* **2012**, *129*, 280–288. [CrossRef] [PubMed]
27. Klein, S.-L.; Scheper, C.; Brügemann, K.; Swalve, H.H.; König, S. Phenotypic relationships, genetic parameters, genome-wide associations, and identification of potential candidate genes for ketosis and fat-to-protein ratio in German Holstein cows. *J. Dairy Sci.* **2019**, *102*, 6276–6287. [CrossRef] [PubMed]
28. Al-Kanaan, A. Heat Stress Response for Physiological Traits in Dairy and Dual Purpose Cattle Populations on Phenotypic and Genetic Scales. Ph.D. Thesis, Faculty of Organic Agriculture, University of Kassel, Kassel, Germany, January 2016.
29. Ali, A.K.A.; Shook, G.E. An optimum transformation for somatic cell concentration in milk. *J. Dairy Sci.* **1980**, *63*, 487–490. [CrossRef]
30. Jaeger, M.; Scheper, C.; König, S.; Brügemann, K. Inbreeding and genetic relationships of the endangered dual-purpose black and white cattle breed (DSN) based on own genetic breed percentage calculations. *Züchtungskunde* **2018**, *90*, 262–279.
31. Thienpont, D.; Rochette, F.; van Parijs, O.F.J. *Diagnosing Helminthiasis by Coproscopical Examination*; Janssen Research Foundation: Beerse, Belgium, 1979; p. 187.
32. Baermann, G. Eine einfache Methode zur Auffindung von Ankylostoma (Nematoden)-Larven in Erdproben. A simple method for isolation of Ancylostoma (nematode) larvae in soil samples. *Geneeskd. Tijdschr. Voor Ned. Indie* **1917**, *51*, 13–137.
33. *SAS Institute*, version 9.4; SAS Insitute, Inc.: Cary, NC, USA, 2013.
34. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef]
35. Van Raden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [CrossRef]
36. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]
37. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef]
38. Browning, B.L.; Zhou, Y.; Browning, S.R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **2018**, *103*, 338–348. [CrossRef]

39. Korkuć, P.; Arends, D.; Brockmann, G.A. Finding the optimal imputation strategy for small cattle populations. *Front. Genet.* **2019**, *10*, 52. [CrossRef] [PubMed]
40. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]
41. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3000 shared controls. *Nature* **2007**, *447*, 661–678. [CrossRef] [PubMed]
42. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [CrossRef] [PubMed]
43. Durinck, S.; Spellman, P.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [CrossRef]
44. ENSEMBL. Genome Browser. Available online: http://www.ensembl.org/index.html (accessed on 18 April 2021).
45. Yates, A.D.; Achuthan, P.; Akanni, W.; Allen, J.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Azov, A.G.; Bennett, R.; et al. Ensembl 2020. *Nucleic Acids Res.* **2020**, *48*, D682–D688. [CrossRef]
46. Kanehisa, M.; Goto, S.; Sato, Y.; Kawashima, M.; Furmichi, M.; Tanabe, M. Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* **2014**, *42*, 199–205. [CrossRef]
47. Hang, D.W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2008**, *4*, 44–57. [CrossRef] [PubMed]
48. McRae, K.M.; Stear, M.J.; Good, B.; Keane, O.M. The host immune response to gastrointestinal nematode infection in sheep. *Parasite Immunol.* **2015**, *37*, 605–613. [CrossRef]
49. Liu, L.; Zhou, J.; Chen, C.J.; Zhang, J.; Wen, W.; Tian, J.; Zhang, Z.; Gu, Y. GWAS-based identification of new loci for milk yield, fat, and protein in Holstein cattle. *Animals* **2020**, *10*, 2048. [CrossRef] [PubMed]
50. Luoreng, Z.-M.; Wang, X.-P.; Mei, C.-G.; Zan, L.-S. Expression profiling of peripheral blood miRNA using RNAseq technology in dairy cows with Escherichia coli-induced mastitis. *Sci. Rep.* **2018**, *8*, 12693. [CrossRef] [PubMed]
51. Kolodziejczyk, L.; Siemieniuk, E.; Skrzydlewska, E. Fasciola hepatica: Effects on the antioxidative properties and lipid peroxidation of rat serum. *Exp. Parasitol.* **2006**, *113*, 43–48. [CrossRef] [PubMed]
52. Kim, E.-S.; Sonstegard, T.S.; da Silva, M.V.G.B.; Gasbarre, L.C.; Van Tassel, C.P. Genome-wide scan of gastrointestinal nematode resistance in closed Angus population selected for minimized influence of MHC. *PLoS ONE* **2015**, *10*, e0119380. [CrossRef]
53. Hu, R.-S.; Zhang, F.-K.; Esheikkha, H.M.; Ma, Q.-N.; Ehsan, M.; Zhao, Q.; Zhu, X.-Q. Proteomic profiling of the liver, hepatic lymph nodes, and spleen of buffaloes infected with *Fasciola gigantica*. *Pathogens* **2020**, *9*, 982. [CrossRef]
54. Rojas-Caraballo, J.; López-Abán, J.; Moreno-Pérez, D.A.; Vicente, B.; Fernández-Soto, P.; del Olmo, E.; Patarroyo, M.A.; Muro, A. Transcriptomic profiling of gene expression during immunization trial against *Fasciola hepatica*: Identification of genes and pathways involved in conferring immunoprotection in a murine model. *BMC Infect. Dis.* **2017**, *17*, 94. [CrossRef] [PubMed]
55. Loukas, A.; Maizels, R.M. Helminth C-type lectins and host-parasite interactions. *Parasitol. Today* **2000**, *16*, 333–339. [CrossRef]
56. Zhang, X.-X.; Cong, W.; Elsheikha, H.M.; Li, G.-H.; Ma, J.-G.; Huang, W.-Y.; Zhao, Q.; Zhu, X.-Q. De novo transcriptome sequencing and analysis of the juvenile and adult sages of Fasciola gigantica. *Infect. Genet. Evol.* **2017**, *51*, 33–40. [CrossRef]
57. Hagberg, M. Immune Cell Responses to the Cattle Lungworm, Dictyocaulus viviparus. Ph.D. Thesis, University Uppsala, Uppsala, Sweden, 2008.
58. Escobedo, G.; Roberts, C.W.; Carrero, J.C.; Morales-Montor, J. Parasite regulation by host hormones: An old mechanism of host exploitation? *Trends Parasitol.* **2005**, *21*, 588–593. [CrossRef]
59. Aguayo, V.; Fernandez, B.N.V.; Rodríguez-Valentín, M.; Riz-Jiménez, C.; Ramos-Benítez, M.J.; Méndez, L.B.; Espino, A.M. Fasciola hepatica GST downregulates NF-κB pathway effectors and inflammatory cytokines while promoting survival in a mouse septic shock model. *Sci. Rep.* **2019**, *9*, 2275. [CrossRef] [PubMed]
60. Johnson, D.R.; Saes, J.; Matthews, J.B. Local cytokine responses in Dictyocaulus viviparus infection. *Vet. Parsitol.* **2005**, *128*, 309–318. [CrossRef]
61. Laodim, T.; Elzo, M.A.; Koonawootrittriron, S.; Suwanasopee, T.; Jattawa, D. Pathway enrichment and protein interaction network analysis for milk yield, fat yield and age at first calving in a Thai multibreed dairy population. *Asian Australas. J. Anim. Sci.* **2019**, *32*, 508–518. [CrossRef] [PubMed]
62. Zhou, C.; Shen, D.; Li, C.; Cai, W.; Liu, S.; Yin, H.; Shi, S.; Cao, M.; Zhang, S. Comparative transcriptomic and proteomic analyses identify key genes associated with milk fat traits in Chinese Holstein cows. *Front. Genet.* **2019**, *10*, 672. [CrossRef] [PubMed]
63. Jedlina, L.; Kozak-Ljunggren, M.; Wedrychowicz, H. In vivo studies of the early, peritoneal, cellular and free radical responses in rats infected with Fasciola hepatica by flow cytometric analysis. *Exp. Parasitol.* **2011**, *128*, 291–297. [CrossRef]
64. Fu, Y.; Browne, J.A.; Killick, K.; Mulcahy, G. Network analysis of the systemic response to *Fasciola hepatica* infection in sheep reveals changes in fibrosis, apoptosis, toll-like receptors 3/4 and B cell function. *Front. Immunol.* **2017**, *8*, 485. [CrossRef] [PubMed]
65. Foster, N.; Elsheikha, H.M. The immune response to parasitic helminths of veterinary importance and its potential manipulation for futre vaccine control strategies. *Parasitol. Res.* **2012**, *110*, 1587–1599. [CrossRef] [PubMed]
66. Walsh, K.P.; Brady, M.T.; Finlay, C.M.; Boon, L.; Mills, K.H.G. Infection with a helminth parasite attenuates autoimmunity through TGF-beta-mediated suppression of Th17 and Th1 responses. *J. Immunol.* **2009**, *183*, 1577–1586. [CrossRef]
67. Brondum, R.F.; Guldbrandtsen, B.; Sahana, G.; Lund, M.S.; Su, G. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. *BMC Genom.* **2014**, *15*, 728. [CrossRef] [PubMed]

68. Mao, X.; Sahana, G.; De Koning, D.-J.; Guldbrandtsen, B. Genome-wide association studies of growth traits in three dairy cattle breeds using whole-genome sequence data. *J. Anim. Sci.* **2016**, *94*, 1426–1437. [CrossRef]
69. Iso-Touru, T.; Sahana, G.; Guldbrandtsen, B.; Lund, M.S.; Vilkki, J. Genome-wide association analysis of milk yield traits in Nordic Red cattle using imputed whole genome sequence variants. *BMC Genet.* **2016**, *17*, 55. [CrossRef] [PubMed]
70. Hoglund, J.K.; Guldbrandtsen, B.; Su, G.; Thomsen, B.; Lund, M.S. Genome scan detects quantitative trait loci affecting female fertility traits in Danish and Swedish Holstein cattle. *J. Dairy Sci.* **2009**, *92*, 2136–2143. [CrossRef]
71. Fortes, M.R.S.; Suhaimi, A.H.M.S.; Porto-Neto, L.R.; McWilliam, S.M.; Flatscher-Bader, T.; Moore, S.S.; D'Occhio, M.J.; Meira, C.T.; Thomas, M.G.; Snelling, W.M. Post-partum anoestrus in tropical beef cattle: A systems approach combining gene expression and genome-wide association results. *Livest. Sci.* **2014**, *166*, 158–166. [CrossRef]
72. Rosa, L.R.O.; Soares, G.M.; Silveira, L.R.; Boschero, A.C.; Barbosa-Sampaio, H.C.L. ARHGAP21 as a regulator of multiple cellular processes. *J. Cell. Physiol.* **2018**, *233*, 8477–8481. [CrossRef]
73. Laskowski, D.; Sjunnesson, Y.; Humblot, P.; Andersson, G.; Gustafsson, H.; Båge, R. The functional role of insulin in fertility and embryonic development—What can we learn from bovine model? *Theriogenology* **2016**, *86*, 457–464. [CrossRef]
74. Gong, J.; Lee, W.; Garnswothy, P.; Webb, R. Effect of dietary-induced increases in circulating insulin concentrations during the early postpartum period on reproductive function in dairy cows. *Reproduction* **2002**, *123*, 419–427. [CrossRef] [PubMed]
75. Holmberg, M.; Andersson-Eklund, L. Quantitative trait loci affecting fertility and calving traits in Swedish dairy cattle. *J. Dairy Sci.* **2006**, *89*, 3664–3671. [CrossRef]
76. Fortes, M.R.S.; Sneling, W.M.; Reverter, A.; Nagaraj, S.H.; Lehnert, S.; Hawken, R.; DeAtley, K.L.; Peters, K.E.; Silver, G.A.; Rincon, G.; et al. Gene network analyses of first service conception in Brangus heifers: Use of genome and traits associations, hypothalamic-transcriptome information, and transcription factors. *J. Anim. Sci.* **2012**, *90*, 2894–2906. [CrossRef] [PubMed]
77. Moore, S.G.; Pryce, J.E.; Hayes, B.J.; Chamberlain, A.J.; Kemper, K.E.; Berry, D.P.; McCabe, M.; Cormican, P.; Lonergan, P.; Fair, T.; et al. Differentially Expressed Genes in Endometrium and Corpus Luteum of Holstein Cows Selected for High and Low Fertility Are Enriched for Sequence Variants Associated with Fertility. *Biol. Reprod.* **2016**, *94*, 1–11. [CrossRef]
78. Kiser, J.N.; Keuter, E.M.; Seabury, C.M.; Neupane, M.; Moraes, J.G.N.; Dalton, J.; Burns, G.W.; Spencer, T.E.; Neibergs, H.L. Validation of 46 loci associated with female fertility traits in cattle. *BMC Genom.* **2019**, *20*, 576. [CrossRef]
79. May, K.; Sames, L.; Scheper, C.; König, S. Genomic loci for uterine diseases in Holsten cows and impact on milk production and fertility. In Proceedings of the 72nd Annual Meeting of the European Association for Animal Production, Davos, Switzerland, 30 August–3 September 2021; Wageningen Academic Publishers: Wageningen, The Netherlands, 2021.
80. Timperio, A.M.; D'Alessandro, A.; Pariset, L.; D'Amici, G.M.; Valentini, A.; Zolla, L. Comparative proteomics and transcriptomics analyses of livers from two different *Bos taurus* breeds: "Chianina and Holstein Friesian". *J. Proteom.* **2009**, *73*, 309–322. [CrossRef] [PubMed]
81. Lehnert, S.A.; Reverter, A.; Byrne, K.A.; Wang, Y.; Nattrass, G.S.; Hudson, N.J.; Greenwood, P.L. Gene expression studies of developing bovine longissimus muscle from two different beef cattle breeds. *BMC Dev. Biol.* **2007**, *7*, 95. [CrossRef] [PubMed]
82. Wang, X.; Wurmser, C.; Pausch, H.; Jung, S.; Reinhardt, F.; Tetens, J.; Thaller, G.; Fries, R. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. *PLoS ONE* **2012**, *7*, e40711. [CrossRef]
83. Bhat, S.A.; Ahmad, S.M.; Ibeagha-Awemu, E.I.; Bhat, B.A.; Dar, M.A.; Mumtaz, P.T.; Shah, R.A.; Ganai, N.A. Comparative transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in gene expression and pathways regulating milk synthesis between Jersey and Kashmiri cattle. *PLoS ONE* **2019**, *14*, e011773. [CrossRef] [PubMed]
84. Braz, C.U.; Rowan, T.N.; Schnabel, R.D.; Decker, J.E. Extensive genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. *bioRxiv* **2020**. [CrossRef]
85. AnimalQTLdb. Available online: https://www.animalgenome.org/cgi-bin/QTLdb/index (accessed on 14 May 2021).
86. Liu, Y.; Xu, L.; Wang, Z.; Xu, L.; Chen, Y.; Zhang, L.; Xu, L.; Gao, X.; Zhu, B.; Li, J. Genomic prediction and association analysis with models including dominance effects for important traits in Chinese Simmental beef cattle. *Animal* **2019**, *9*, 1055. [CrossRef]
87. Howard, J.T.; Pryce, J.E.; Baes, C.; Maltecca, C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *J. Dairy Sci.* **2017**, *100*, 6009–6024. [CrossRef] [PubMed]
88. Sorensen, A.C.; Madsen, P.; Sorenzen, M.K.; Berg, P. Udder health shows inbreeding depression in Danish Holsteins. *J. Dairy Sci.* **2006**, *89*, 4077–4082. [CrossRef]
89. May, K.; Weimann, C.; Scheper, C.; Strube, S.; König, S. Allele substitution and dominance effects of *CD166/ALCAM* gene polymorphisms for endoparasite resistance and test-day traits in a small cattle population using logistic regression analyses. Mamm. *Genome* **2019**, *30*, 301–317. [CrossRef] [PubMed]
90. Fugate, R.T.; Dauten, L.H.; Wiggans, G.R.; White, H.M. Determination of Single Nucleotide Polymorphisms Associated with Subclinical Ketosis in Jersey Cattle. Available online: https://heatherwhite.dysci.wisc.edu/wp-content/uploads/sites/164/2014/07/ADSA-Jersey-SNP-poster-FINAL.pdf (accessed on 19 May 2021).
91. Littlejohn, M.D.; Walker, C.G.; Ward, H.E.; Lehnert, K.B.; Snell, R.G.; Verkerk, G.A.; Spelman, R.J.; Clark, D.A.; Davis, S.R. Effects of reduced frequency of milk removal on gene expression in the bovine mammary gland. *Physiol. Genom.* **2010**, *41*, 21–32. [CrossRef]
92. Buitenhuis, B.; Rontved, C.M.; Edwards, S.M.; Ingvartsen, K.; Sorensen, P. In depth analysis of genes and pathways of the mammary gland involve in the pathogenesis of bovine Escherichia coli-mastitis. *BMC Genom.* **2011**, *12*, 130. [CrossRef]

93. Yu, C.; Wang, F.; Jin, C.; Huang, X.; Miller, D.L.; Basilico, C.; McKeehan, W.L. Role of fibroblast growth factor type 1 and 2 in carbon tetrachloride-induced hepatic injury and fibrogenesis. *Am. J. Pathol.* **2003**, *163*, 1653–1662. [CrossRef]
94. Sulaiman, A.A.; Zolnierczyk, K.; Japa, O.; Owen, J.P.; Maddison, B.C.; Emes, R.D.; Hodgkinson, J.E.; Gough, K.C.; Flynn, R.J. A trematode parasite derived growth factor binds and exerts influences on host immune functions via host cytokine receptor complexes. *PLoS Pathog.* **2016**, *12*, e1005991. [CrossRef] [PubMed]

KAPITEL 3

Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing

Manuel J. Wolf[1]*, Guilherme B. Neumann[2], Paula Korkuć[2], Tong Yin[1], Gudrun A. Brockmann[2], Sven König[1] and Katharina May[1]*

[1] Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany

[2] Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt-Universität of Berlin, 10115 Berlin, Germany

Journal of Dairy Science, 106(5), 3345-3358, https://doi.org/10.3168/jds.2022-22665

Accepted: 16 December 2022

Available online: 5 April 2023

Version of Record: 23 April 2023

J. Dairy Sci. 106:3345–3358
https://doi.org/10.3168/jds.2022-22665

Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing

Manuel J. Wolf,[1]* Guilherme B. Neumann,[2] Paula Kokuć,[2] Tong Yin,[1] Gudrun A. Brockmann,[2] Sven König,[1]† and Katharina May[1]*
[1]Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany
[2]Animal Breeding Biology and Molecular Genetics, Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Humboldt Universität zu Berlin, 10115 Berlin, Germany

ABSTRACT

Genetic evaluations of local cattle breeds are hampered due to small reference groups or biased due to the utilization of SNP effects estimated in other large populations. Against this background, there is a lack of studies addressing the possible advantage of whole-genome sequences (WGS) or consideration of specific variants from WGS data in genomic predictions for local breeds with small population size. Consequently, the aim of this study was to compare genetic parameters and accuracies of genomic estimated breeding values (GEBV) for 305-d production traits, fat-to protein ratio (FPR), and somatic cell score (SCS) at the first test date after calving and confirmation traits of the endangered German Black Pied cattle (DSN) breed using 4 different marker panels: (1) the commercial 50K Illumina BovineSNP50 BeadChip, (2) a customized 200K chip designed for DSN (DSN200K) which considers the most important variants for DSN from WGS, (3) randomly generated 200K chips based on WGS data, and (4) a WGS panel. The same number of animals was considered for all marker panel analyses (i.e., 1,811 genotyped or sequenced cows for conformation traits, 2,383 cows for lactation production traits, and 2,420 cows for FPR and SCS). Mixed models for the estimation of genetic parameters directly included the respective genomic relationship matrix from the different marker panels plus the trait-specific fixed effects. For the calculation of GEBV accuracies, we applied repeated random subsampling validation. In the process of separate cross-validations per trait, we created a validation set including 20% of cows with masked phenotypes, and a training set comprising 80% of the cows. The cows were selected randomly in a procedure with 10 replicates considering replacements in the different scenarios. The accuracy was defined as the correlation between the direct GEBV and the phenotypes with subtracted corresponding fixed effects for the cows in the validation set. For FPR and SCS, as well as for lactation production traits, heritabilities were largest based on WGS data, but the increase compared with the 50K or DSN200K applications was quite small in the range from 0.01 to 0.03. Also, for most of the conformation traits, heritabilities were largest based on WGS and DSN200K data, but the increase was in the range of the corresponding standard error. Accordingly, GEBV accuracies for most of the studied traits were highest based on WGS data or when utilizing the DSN200K chip, but the accuracy differences across the marker panels were quite small and nonsignificant. In conclusion, WGS data and the DSN200K chip only contributed to minor improvements in genomic predictions, still justifying the use of the commercial 50K chip. Nevertheless, WGS and the 200KDSN chip harbor breed-specific variants, which are valuable for studying causal genetic mechanisms in the endangered DSN population.

Key words: local cattle, whole-genome sequences, genomic predictions, genetic parameters

INTRODUCTION

The Illumina BovineSNP50 BeadChip with ~54,000 markers (hereafter referred as 50K chip) is commonly used for genetic evaluations in various cattle breeds worldwide. The 50K chip is designed based on a reference panel of large commercial populations such as Holstein Friesian (**HF**) and Simmental, and works particularly well in breeds closely related to these reference populations (Matukumalli et al., 2009). With regard to genetically different endangered cattle breeds, Hozé et al. (2013) reported a loss of up to 20,000 markers after filtering the 50K data, and indicated possible effects on accuracies of genomic estimated breeding values (**GEBV**). In this regard, van den Berg et al. (2017)

Received August 17, 2022.
Accepted December 16, 2022.
*These authors contributed equally to this work.
†Corresponding author: sven.koenig@agrar.uni-giessen.de

and Zhang et al. (2019) showed an accuracy decline when including a larger number of low minor allele frequency (**MAF**) markers into genomic prediction data sets. Theoretically, marker density or utilization of sequence data can affect the power of genetic evaluations (Zhang et al., 2019). Generally, local cattle breeders have a strong interest in implementing genomic selection schemes, because breeding goals of these breeds reflect a large pattern of new functional traits. Also, for numerically small populations, Buch et al. (2012) showed that the establishment of cow training sets and relating cow genotypes directly to novel cow phenotypes is more efficient than basing selection on inaccurate sire breeding values.

The German Black Pied cattle (**DSN**, German: Deutsches Schwarzbuntes Niederungsrind) is an endangered local dual-purpose breed with a small population size of ∼2,500 cattle, and is the founder breed of the modern HF (Brade and Brade, 2013). In pedigree-based analyses, the average inbreeding coefficient for DSN cows in different regions varied from 2.9 to 4.1% (Jaeger et al., 2019), and the effective population size of the active DSN population of the birth year 2015 comprised 85 animals (Jaeger et al., 2018). The DSN is the founder breed of the modern HF, but the average genetic relationship between the current DSN and HF populations was quite low with only 0.1% (Naderi et al., 2020). In DSN, only ∼37,000 markers of the 50K chip are informative after filtering for a SNP call rate >95% and a MAF >5% (Korkuć et al., 2021). Divergent selection strategies in DSN and HF during the past decades contributed to the genomic differentiation between both breeds. For example, Naderi et al. (2020) identified different selection signature segments in both breeds and detected *CLU* as a main candidate gene affecting milk protein content in DSN, but not in HF. Korkuć et al. (2021) showed that the well described *DGAT1* gene with major effects on milk production traits in HF is not associated with milk yield or fat percentage in DSN. Therefore, a large proportion of breed-specific variants in DSN might not be covered by the 50K chip, implying GEBV with smaller accuracy. Because of the breed-specific signals in genome-wide associations, we recently designed a customized 200K chip for DSN (**DSN200K**) including breed-specific variants (Neumann et al., 2021). We assume favorable effects on GEBV accuracies compared with the 50K chip application when considering the more specific DSN genome characteristics on the denser DSN200K.

The effect of marker density on prediction accuracy or reliability was evaluated in several studies and in different breeds. Su et al. (2012) reported an improved genomic prediction reliability of 0.5% in Nordic Holstein and of 1.0% in Red Dairy cattle for high density 777K marker panels compared with the medium-density 50K chip for protein yield, female fertility, and udder health. Similarly, Erbe et al. (2012) indicated a small increase in GEBV accuracies of 0.01 in HF and 0.03 in Jerseys for milk, protein, and fat yield (**FY**), when using imputed 777K high-density SNP panels compared with the 50K chip. With regard to whole-genome sequence (**WGS**) data analyses, Meuwissen and Goddard (2010) applied stochastic simulations and compared genomic predictions from WGS data with those from SNP chips with ∼30,000 markers. The prediction accuracies increased by ∼40% when using WGS, due to the direct consideration of the functional mutations. In simulations, Druet et al. (2014) observed a plus of 1.4% in GEBV accuracy of WGS data compared with commercial SNP panels. Results from these simulation studies (Meuwissen and Goddard, 2010; Druet et al., 2014) indicated an increase of prediction accuracies from WGS data with increasing effective population size and decreasing MAF for the causative mutation. However, the advantage of WGS data over dense marker panels in genomic predictions was not confirmed in real data sets of large commercial HF populations (van Binsbergen et al., 2015; Raymond et al., 2018). In genetic evaluations in the large HF population, test runs based on sequence variants increased the prediction reliability only marginally by 0.6 percentage points compared with high-density SNP genotypes (VanRaden et al., 2017). The inclusion of imputed WGS data did not significantly contribute to improved GEBV accuracies for 33 traits in HF (VanRaden et al., 2017), and for conformation, fertility and udder health traits in Brown Swiss cattle (Frischknecht et al., 2018).

Lund et al. (2016) recommended WGS data especially for genetic evaluations in breeds with small reference populations. However, there is a lack of studies addressing the possible advantage of WGS data in genomic predictions for small-sized cattle populations. To fill this gap, the objective of this study was to evaluate and to compare genetic parameter estimates and GEBV accuracies for 305-d production, test-day and conformation traits in real data sets of the endangered DSN breed for 4 different marker types: (1) the Illumina 50K SNP chip, (2) our novel designed breed-specific DSN200K including a large number of informative variants specifically for DSN, (3) randomly generated 200K chips based on WGS data, and (4) WGS data (imputed).

MATERIALS AND METHODS

Because no human or animal subjects were used, this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board.

Table 1. Descriptive statistics for 305-d production and traits from the first test-day after calving

Trait[1]	N	Mean	SD	Minimum	Maximum
305-d production					
MY	2,383	6,162.11	1,246.86	2,248.00	10,574.00
ECM	2,383	6,398.44	1,222.40	2,093.07	9,728.75
FY	2,383	256.71	48.61	76.00	412.00
Fat%	2,383	4.19	0.37	2.16	5.60
PY	2,383	221.16	44.72	77.00	346.00
Pro%	2,383	3.59	0.22	2.85	4.31
First test-day after calving					
FPR	2,420	1.18	0.19	0.45	2.52
SCS	2,420	2.61	1.51	−1.06	8.88

[1]MY = milk yield; PY = protein yield; FY = fat yield; Fat% = fat percentage; Pro% = protein percentage; FPR = fat-to-protein ratio.

Cow Traits

We considered 2,420 first-parity DSN cows kept in 10 different herds. The herds were located in the German federal states Hesse, Berlin-Brandenburg and Mecklenburg-West Pomerania. The cows were born between 2006 and 2018. The age at first calving ranged from 22 to 46 mo (mean: 27.93 mo). For 305-d production traits, we included only cows with a lactation length of more than 250 d in milk. Hence, 2,383 cows had records for lactation milk yield (**MY**), ECM, FY, fat percentage (**Fat%**), protein yield (**PY**) and protein percentage (**Pro%**). The ECM was calculated according to the following formula by Zumwald et al. (2018): $MY/3.14 \times (0.38 \times Fat\% + 0.24 \times Pro\% + 0.816)$. Test-day records from the first official test-day after calving included fat-to-protein ratio (**FPR**) and milk somatic cells from all 2,420 cows. Somatic cell count was transformed into SCS applying the formula by Ali and Shook (1980): $\log_2 (SCC/100{,}000) + 3$. For FPR and SCS, we focused on the challenging early lactation period, because the early lactation measurements were reliable indicators for herd life especially in the large-scale herds from the eastern part of Germany (Bergk and Swalve, 2011). Consequently, we included only the first official test-day after calving between 5 and 40 DIM. Descriptive statistics for 305-d milk production traits, FPR and SCS are given in Table 1.

Conformation traits were recorded in a subset comprising the 5 largest herds located in the eastern part of Germany including 1,811 records for 18 linear traits and 4 type trait composites. The linear traits included dairy character, stature, body depth, chest width, rump angle, rump width, BCS, rear leg set side view, foot angle, hock quality, rear leg set rear view, rear udder height, central ligament, teat placement front, teat placement rear, fore udder attachment, udder depth, and teat length and were scored on a scale from 1 to 9 in increments of 1. Stature was measured in centimeters. The conformation composites for dairy type, body type, feet and legs, and udder were recorded on a scale from 65 to 88 points. Conformation trait scoring was conducted by trained classifiers from German breeding organizations. Descriptive statistics for the type trait composites and linear conformation traits are presented in Table 2.

SNP Chip Design, Imputation of Genotypes, and Quality Control

The design of the customized DSN-specific 200K SNP chip (Axiom myDesign TG Array; Thermo Fisher Scientific) is described in detail by Neumann et al. (2021). Variants for this DSN200K chip construction were selected from 20,587,181 sequence variants consisting of SNPs, insertions and deletions detected from WGS of 304 DSN animals. In a first step of the SNP chip construction, variants were filtered for technical suitability (e.g., low read coverage, high specificity of allele calls). In total, 3,069,815 SVs passed the technical suitability check. Afterward, we selected variants being informative in DSN. In detail, the chip contains DSN-unique variants, variants associated with important traits of interest in DSN (e.g., disease resistance, MY, fertility), variants with low, high, or moderate effects on gene transcripts, and DSN informative variants from the 50K chip (criteria in this regard are outlined by Neumann et al., 2021). These sum up to 175,537 SNPs and 8,618 insertions and deletions corresponding to 103,801 haplotype blocks in DSN.

For the evaluation of the 3 different marker densities in the context of genomic predictions, 3 genotype panels (50K, DSN200K and WGS) of the same animals were constructed using different steps of imputation and downscaling. Details for imputation procedures are given by Korkuć et al. (2019). Briefly, genotyping was performed for 1,797 animals using the 50K chip, for 1,595 animals using the DSN200K chip, and for 304 animals using WGS data. The 200K data (167,271 variants with MAF > 0.1) of 1,579 animals were used

Table 2. Descriptive statistics for type trait composites and linear conformation traits

Trait	N	Mean	SD	Minimum	Maximum
Type trait composite					
Dairy type	1,811	79.01	3.02	68.0	86.0
Body type	1,811	79.21	3.42	65.0	87.0
Feet and legs	1,811	81.82	3.08	65.0	87.0
Udder	1,811	78.24	3.66	65.0	86.0
Linear trait					
Dairy character	1,811	3.38	1.35	1.0	9.0
Stature (in cm)	1,811	136.83	2.99	126.0	152.0
Body depth	1,811	5.34	1.29	1.0	9.0
Chest width	1,811	6.28	1.13	2.0	9.0
Rump angle	1,811	4.79	0.97	1.0	9.0
Rump width	1,811	5.74	1.23	2.0	9.0
BCS	1,811	6.06	1.16	2.0	9.0
Rear leg set side view	1,811	5.23	0.98	1.0	9.0
Foot angle	1,811	5.04	1.00	1.0	9.0
Hock quality	1,811	5.95	1.38	1.0	9.0
Rear leg set rear view	1,811	5.86	1.52	1.0	9.0
Rear udder height	1,811	4.25	1.36	1.0	9.0
Central ligament	1,811	4.34	1.37	1.0	8.0
Teat placement front	1,811	4.00	1.16	1.0	8.0
Teat placement rear	1,811	4.96	1.06	1.0	8.0
Fore udder attachment	1,811	5.32	1.32	1.0	9.0
Udder depth	1,811	4.39	1.23	1.0	9.0
Teat length	1,811	4.95	1.00	2.0	9.0

as a reference panel for the imputation of 1,797 DSN 50K genotypes to 200K genotypes. The WGS data (16,175,216 variants) was used as a reference panel for the imputation of either the 50K data or the DSN200K data to WGS genotypes. The reference panel for imputation to WGS genotypes was filtered beforehand to increase the imputation accuracy. Only variants with MAF >0.01 and call rate >95% were retained. Variants with more than 5% Mendelian error rate (based on opposing homozygotes for the available 156 parent-offspring pairs) were removed. In addition, variants with quality by depth <10 and depth <3,000 (as predicted with the Ensembl Variant Effect Predictor (McLaren et al., 2016), were discarded due to low transition or transversion ratio. Phasing and imputation of variants was performed using Beagle v.5.1 (Browning et al., 2018). For the calculation of imputation accuracy, we considered the variants of 15 sequenced DSN animals genotyped with both the 50K chip and the DSN200K chip. For that reason, these 15 animals were not included in the reference panels. The imputation accuracy from 50K to DSN200K was on average 99.99% (0.07% SD). The imputation accuracy to sequence level was on average 96.50% (0.58% SD) for the 50K chip and 98.58% (0.21% SD) for the DSN200K chip.

We performed quality control of the 3 genotype panels (50K, 200K, and WGS) in PLINK (Purcell et al., 2007) in 2 steps. In a first step, we filtered genotypes and imputed WGS data by including only variants and individuals with a 90% genotyping rate, variants with MAF >0.05 and variants without significant deviation ($P < 10^{-6}$) from Hardy-Weinberg equilibrium (**HWE**). Moreover, we considered only SNPs on autosomes. This step resulted in several 31,539 variants for 50K, 125,678 variants for 200K, and 11,827,995 variants for WGS. Afterward, those data sets were restricted to the animals with available phenotypes for the investigated traits (305-d production, FPR, SCS, conformation traits), and the same filtering steps for MAF, call rates, and HWE, were applied. The final number of variants after filtering depending on the respective traits were on average ~31,400 for 50K, ~124,300 for the DSN200K, and ~11.6 million for WGS data. The 304 sequenced cows were considered in genomic studies for the DSN200K development, and further on in the present study in genomic predictions, implying possible biased estimates or some extent of overfitting

Generation of Random 200K SNP Chips from WGS Data

From the 3,069,815 technically suitable variants, 2,833,148 variants were available from WGS data after previous filtering steps. We used the sample() function in R (R Core Team, 2022) to generate 20 random chips from WGS data including the same number of 125,678 variants as remained after the first filtering step for the DSN200K. The same number of markers guaranteed the comparison of genetic parameters and GEBV accuracies between the breed-specific DSN200K with the same marker densities, but without considering breed-specific variants (hereafter referred as random 200K chips).

Genetic Parameter and Genomic Breeding Value Estimations

We created genomic relationship matrices $\mathbf{G}$ in GCTA (Yang et al., 2011) considering all cows with genotypes for each trait. Afterward, for BLUPF90 applications (Aguilar, 2018), we computed the inverse of $\mathbf{G} + 0.02\mathbf{I}$, with $\mathbf{I}$ being an identity matrix. We used single-trait animal models and the AI-REML procedure to estimate variance components and heritabilities by applying the AIREMLF90 package from the BLUPF90 software package (Misztal et al., 2014). The statistical model considering the whole data set for each trait was:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e},$$

where $\mathbf{y}$ = vector of phenotypes for 305-d production, FPR, SCS, and conformation traits; $\mathbf{b}$ = vector of fixed effects including the following fixed effects for the 305-d production traits, FRP, and SCS: herd, calving year, calving season, age at first calving, and lactation length as linear regression on lactation length for incomplete 305-d records and on DIM after calving for the test-day records; and for conformation traits: herd, classifier, year of conformation trait scoring, lactation stage, and age at first calving; $\mathbf{g}$ = vector of additive-genetic effects with $\sim N\left(\mathbf{0}, \mathbf{G}\sigma_g^2\right)$, with $\mathbf{G}$ denoting the genomic relationship matrix, and σ_g^2 the genetic variance; $\mathbf{e}$ = residual effects with $\sim N\left(\mathbf{0}, \mathbf{I}\sigma_e^2\right)$, and σ_e^2 residual variance; $\mathbf{X}$ and $\mathbf{Z}$ = design matrices for $\mathbf{b}$ and $\mathbf{g}$, and $\mathbf{I}$ = identity matrix.

Cross-Validation and Calculation of GEBV Accuracies

For the calculation of GEBV accuracies, we applied repeated random subsampling validation (**RSV**; i.e., Monte Carlo cross-validation; Picard and Cook, 1984). We performed cross-validation for each trait separately. In this regard, we divided the data set into a validation set (20% of cows) with masked phenotypes and a training set (80% of cows). Number of animals in the validation set were 477 for 305-d production traits, 484 for FPR and SCS, and 363 for conformation traits. We selected the cows randomly in a procedure with 10 replicates, allowing replacements in consecutive replicates. In each validation step, the breeding values from the cows in the masked group (validation set) were estimated based on the genomic relationship information and phenotypes for the cows in the training set. The accuracy was defined as the correlation between the direct GEBV and the phenotypes with subtracted corresponding fixed effects (i.e., $\mathbf{y} - \mathbf{Xb}$) for the cows with masked phenotypes (cows in the validation set). Afterward, the mean accuracy was calculated considering the 10 replicates for each trait for 50K, DSN200K, random 200K (20 random subsets), and WGS data sets.

RESULTS

Genetic Parameter Estimates

The heritabilities and variance components with corresponding SE for 305-d production traits, FPR, and SCS from the whole data set based on different marker densities are presented in Table 3. For 305-d production traits, the heritabilities ranged from 0.355 (WGS) for FY to 0.646 (WGS) for Pro%. For all 305-d production traits, highest heritabilities were estimated based on WGS data. However, the heritabilities only increased by 0.011 to 0.028 when using WGS data compared with 50K or 200K marker densities. For SCS and FPR from the first test-day, the additional genetic variance captured by WGS data were quite small. In most cases, the increase of marker densities was associated with increasing heritabilities, but heritability differences from the 4 chips were in a narrow range of corresponding SE from 0 to 0.036 indicating nonsignificant effects. The heritabilities for FPR based on the different chips were 0.145 (50K), 0.152 (DSN200K), 0.143 (random 200K), and 0.148 (WGS). For SCS, the heritabilities were very similar from the different genotype data sets (i.e., 0.097 based on WGS data, 0.095 based on the random 200K chip data, 0.088 based on the 50K chip, and 0.085 based on the DSN200K). Similarly, the genetic variance for SCS was larger based on WGS data (0.202) compared with the other 3 chips (i.e., 0.198 for the random 200K, 0.183 for 50K, and 0.179 for the DSN200K).

The heritabilities and variances with corresponding SE for conformation traits based on different marker densities are presented in Table 4. For the 4 type trait composites, heritabilities ranged from 0.276 (50K) to 0.292 (WGS) for dairy type, from 0.317 (50K) to 0.353 (WGS) for body type, from 0.080 (50K) to 0.089 (DSN200K) for feet and legs, and from 0.215 (50K) to 0.230 (WGS) for udder. The heritabilities for the 18 linear traits ranged from 0.010 for foot angle with the 200K chips to 0.450 for stature with WGS. For teat placement front and central ligament, heritabilities were largest with the 50K chip. Generally, heritability differences were quite small for the same trait based on the different marker panels, but WGS data contributed to the largest heritabilities for 15 of the 22 conformation traits. This finding again indicates that denser marker data (i.e., WGS compared with the 50K or the 200K density) captures the additive-genetic variance for conformation traits in DSN more

Table 3. Heritabilities, additive-genetic variances $\left(\sigma_g^2\right)$, and residual variances $\left(\sigma_e^2\right)$ with corresponding SE of estimates for 305-d production, and fat-to-protein ratio (FPR) and SCS from the first test-day after calving based on different marker densities[1]

	Marker density[3]							
	50K		DSN 200K		Random 200K (1–20)		WGS	
Trait[2]	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
MY								
h^2	0.410	0.036	0.412	0.036	0.414	0.037	*0.433	0.037
σ_g^2	354,020.000	38,857.000	354,370.000	38,593.000	357,023.500	39,674.600	374,540.000	40,714.000
σ_e^2	507,870.000	25,385.000	504,540.000	25,263.000	503,745.000	26,368.600	488,960.000	26,632.000
ECM								
h^2	0.348	0.036	0.349	0.036	0.350	0.037	*0.367	0.038
σ_g^2	260,210.000	32,101.000	260,680.000	31,968.000	262,169.000	32,824.250	275,160.000	33,784.000
σ_e^2	486,380.000	22,974.000	484,230.000	22,909.000	483,819.500	23,785.100	472,840.000	24,075.000
FY								
h^2	0.338	0.036	0.342	0.036	0.341	0.037	*0.355	0.038
σ_g^2	423.420	52.902	428.030	53.093	426.869	54.161	445.460	55.673
σ_e^2	827.410	38.545	821.870	38.527	822.737	39.882	806.860	40.405
Fat%								
h^2	0.543	0.033	0.558	0.033	0.546	0.034	*0.562	0.035
σ_g^2	0.060	0.005	0.062	0.005	0.060	0.005	0.062	0.005
σ_e^2	0.050	0.003	0.049	0.003	0.050	0.003	0.048	0.003
PY								
h^2	0.349	0.036	0.347	0.036	0.350	0.037	*0.367	0.038
σ_g^2	317.710	39.305	315.500	38.927	318.838	40.020	334.740	41.189
σ_e^2	591.060	28.037	590.000	27.930	588.329	28.975	575.030	29.330
Pro%								
h^2	0.618	0.030	0.640	0.030	0.635	0.032	*0.646	0.032
σ_g^2	0.021	0.002	0.022	0.002	0.021	0.002	0.022	0.002
σ_e^2	0.013	0.001	0.012	0.001	0.012	0.001	0.012	0.001
FPR								
h^2	0.145	0.031	*0.152	0.032	0.143	0.032	0.148	0.033
σ_g^2	0.005	0.001	0.005	0.001	0.005	0.001	0.005	0.001
σ_e^2	0.028	0.001	0.028	0.001	0.028	0.001	0.028	0.001
SCS								
h^2	0.088	0.025	0.086	0.025	0.095	0.026	*0.097	0.027
σ_g^2	0.183	0.052	0.179	0.052	0.198	0.056	0.202	0.057
σ_e^2	1.893	0.068	1.895	0.067	1.877	0.069	1.874	0.069

[1]The highest heritability per trait is indicated with an asterisk.
[2]MY = 305-d milk yield; PY = 305-d protein yield; FY = 305-d fat yield; Fat% = 305-d fat percentage; Pro% = 305-d protein percentage.
[3]DSN = German Black Pied cattle; mean heritabilities based on the 20 random 200K chips and respective mean SE; WGS = whole-genome sequence.

precisely. For the feet and leg composite and for the linear type traits dairy character, rear udder height and teat length, heritabilities were largest with the DSN200K. Heritabilities increased up to 2.2% points with the DSN200K compared with the 50K chip for most of the 22 conformation traits, except for chest width, rump angle, BCS, foot angle, central ligament, and teat placement front. Only for rump angle, the heritability from the random 200K chips was the same as estimated based on WGS data.

GEBV Accuracies

The mean GEBV accuracies from RSV for 305-d production traits, FPR and SCS are presented in Table 5. For the moderately to highly heritable 305-d production traits, we achieved the highest GEBV accuracies based on our novel designed DSN200K chip. In this regard, accuracies increased by 0.002 to 0.008 compared with the 50K SNP chip, and by 0.0002 to 0.011 compared with WGS data (Figure 1). For 305-d production

Table 4. Heritabilities, additive-genetic variances $\left(\sigma_g^2\right)$, and residual variances $\left(\sigma_r^2\right)$ with corresponding SE of estimates for conformation traits for different marker densities[1]

	Marker density[2]							
	50K		DSN 200K		Random 200K (1–20)		WGS	
Trait	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Dairy type								
h^2	0.276	0.039	0.291	0.040	0.286	0.041	*0.292	0.041
σ_g^2	2.358	0.373	2.490	0.386	2.439	0.387	2.499	0.396
σ_e^2	6.141	0.312	6.040	0.314	6.070	0.322	6.018	0.326
Body type								
h^2	0.317	0.042	0.328	0.043	0.346	0.044	*0.353	0.045
σ_g^2	3.327	0.506	3.451	0.518	3.641	0.538	3.712	0.548
σ_e^2	7.124	0.389	7.035	0.392	6.840	0.402	6.772	0.408
Feet and legs								
h^2	0.080	0.030	*0.089	0.031	0.086	0.031	0.088	0.032
σ_g^2	0.694	0.259	0.779	0.273	0.750	0.272	0.763	0.276
σ_e^2	7.970	0.336	7.894	0.339	7.913	0.342	7.901	0.344
Udder								
h^2	0.215	0.039	0.221	0.040	0.228	0.040	*0.230	0.040
σ_g^2	2.817	0.551	2.905	0.559	2.980	0.564	3.016	0.569
σ_e^2	10.260	0.504	10.179	0.505	10.069	0.510	10.022	0.514
Stature								
h^2	0.416	0.043	0.428	0.043	0.437	0.045	*0.450	0.046
σ_g^2	3.404	0.439	3.512	0.445	3.583	0.460	3.704	0.469
σ_e^2	4.757	0.292	4.664	0.292	4.597	0.305	4.492	0.309
Dairy character								
h^2	0.227	0.039	*0.242	0.041	0.229	0.041	0.234	0.042
σ_g^2	0.360	0.068	0.387	0.071	0.363	0.069	0.372	0.071
σ_e^2	1.221	0.061	1.202	0.061	1.218	0.062	1.211	0.063
Body depth								
h^2	0.332	0.043	0.351	0.044	0.354	0.045	*0.360	0.046
σ_g^2	0.521	0.079	0.553	0.081	0.555	0.082	0.565	0.083
σ_e^2	1.041	0.059	1.017	0.059	1.009	0.061	0.999	0.061
Chest width								
h^2	0.180	0.037	0.183	0.038	0.195	0.039	*0.199	0.040
σ_g^2	0.212	0.046	0.216	0.047	0.231	0.049	0.235	0.050
σ_e^2	0.963	0.045	0.961	0.046	0.945	0.047	0.941	0.047
Rump angle								
h^2	0.315	0.043	0.305	0.043	*0.331	0.045	*0.331	0.045
σ_g^2	0.294	0.046	0.284	0.045	0.309	0.048	0.308	0.048
σ_e^2	0.635	0.035	0.642	0.035	0.620	0.036	0.620	0.037
Rump width								
h^2	0.303	0.043	0.309	0.043	0.316	0.044	*0.325	0.045
σ_g^2	0.419	0.067	0.427	0.068	0.436	0.070	0.449	0.071
σ_e^2	0.957	0.052	0.950	0.053	0.940	0.054	0.928	0.055
Fore udder attachment								
h^2	0.177	0.037	0.191	0.038	0.189	0.038	*0.196	0.039
σ_g^2	0.296	0.065	0.320	0.068	0.317	0.067	0.327	0.069
σ_e^2	1.371	0.065	1.350	0.065	1.349	0.065	1.339	0.066
Rear udder height								
h^2	0.221	0.042	0.243	0.043	0.237	0.043	*0.246	0.044
σ_g^2	0.364	0.074	0.402	0.077	0.391	0.077	0.406	0.079

Continued

Table 4 (Continued). Heritabilities, additive-genetic variances $\left(\sigma_g^2\right)$, and residual variances $\left(\sigma_r^2\right)$ with corresponding SE of estimates for conformation traits for different marker densities[1]

	Marker density[2]							
	50K		DSN 200K		Random 200K (1–20)		WGS	
Trait	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
σ_e^2	1.277	0.065	1.245	0.065	1.250	0.067	1.238	0.068
Udder depth								
h^2	0.268	0.042	0.270	0.042	0.285	0.044	*0.287	0.044
σ_g^2	0.387	0.067	0.389	0.067	0.411	0.070	0.414	0.071
σ_e^2	1.049	0.055	1.046	0.055	1.025	0.057	1.022	0.057
Teat length								
h^2	0.335	0.043	0.343	0.044	0.337	0.045	*0.344	0.045
σ_g^2	0.336	0.051	0.346	0.051	0.338	0.051	0.345	0.053
σ_e^2	0.664	0.038	0.658	0.038	0.660	0.039	0.654	0.039
Teat placement front								
h^2	*0.312	0.041	0.306	0.041	0.307	0.041	0.307	0.042
σ_g^2	0.380	0.057	0.371	0.056	0.372	0.057	0.372	0.057
σ_e^2	0.834	0.045	0.839	0.045	0.837	0.046	0.836	0.046
Teat placement rear								
h^2	0.199	0.038	0.206	0.038	0.199	0.039	*0.207	0.040
σ_g^2	0.213	0.043	0.220	0.044	0.213	0.044	0.221	0.045
σ_e^2	0.854	0.041	0.846	0.041	0.852	0.042	0.844	0.042
Rear leg set side view								
h^2	0.122	0.033	0.140	0.036	0.148	0.037	*0.153	0.038
σ_g^2	0.112	0.031	0.128	0.034	0.136	0.035	0.140	0.036
σ_e^2	0.800	0.035	0.786	0.036	0.778	0.037	0.774	0.037
Rear leg set rear view								
h^2	0.097	0.032	0.111	0.034	0.112	0.035	*0.116	0.036
σ_g^2	0.195	0.066	0.223	0.070	0.226	0.072	0.233	0.073
σ_e^2	1.807	0.079	1.782	0.080	1.777	0.081	1.770	0.082
Central ligament								
h^2	*0.166	0.039	0.157	0.038	0.162	0.039	0.159	0.039
σ_g^2	0.298	0.073	0.282	0.072	0.290	0.073	0.284	0.073
σ_e^2	1.490	0.072	1.502	0.072	1.493	0.073	1.497	0.073
Hock quality								
h^2	0.150	0.036	0.158	0.037	0.159	0.038	*0.161	0.038
σ_g^2	0.279	0.069	0.293	0.071	0.295	0.073	0.299	0.074
σ_e^2	1.569	0.072	1.558	0.073	1.554	0.074	1.550	0.075
Foot angle								
h^2	0.011	0.017	0.010	0.017	0.011	0.018	*0.012	0.019
σ_g^2	0.009	0.015	0.009	0.015	0.010	0.016	0.010	0.016
σ_e^2	0.869	0.032	0.870	0.032	0.869	0.033	0.868	0.033
BCS								
h^2	0.260	0.042	0.260	0.042	0.265	0.043	*0.266	0.044
σ_g^2	0.307	0.054	0.308	0.055	0.313	0.056	0.315	0.057
σ_e^2	0.871	0.045	0.871	0.046	0.864	0.047	0.862	0.047

[1]The highest heritability per trait is indicated with an asterisk.

[2]DSN = German Black Pied cattle; mean heritabilities based on the 20 random 200K chips; WGS = whole-genome sequence.

traits, the accuracies from the 20 random 200K chips were consistently lower compared with the DSN200K. Hence, the increase in GEBV accuracies is not only due to increased marker density from 50K to 200K, but additionally caused by the inclusion of DSN-specific variants on the DSN200K chip. For SCS, we achieved

Table 5. Accuracy of genomic predictions for 305-d production and fat-to-protein ratio (FPR) and SCS from the first test-day after calving using different marker densities[1]

	Marker density[3]			
Trait[2]	50K	DSN 200K	Random 200K (1–20)	WGS
305-d production trait				
MY	0.453	*0.455	0.449	0.452
ECM	0.391	*0.392	0.385	0.387
FY	0.375	*0.378	0.370	0.371
Fat%	0.558	*0.567	0.557	0.558
PY	*0.394	*0.394	0.387	0.390
Pro%	0.605	*0.614	0.605	0.605
First test-day after calving				
FPR	0.178	*0.185	0.172	0.173
SCS	0.156	0.157	0.160	*0.161

[1]The accuracy presents the mean of all 10 runs from cross-validation per trait and marker density. The highest accuracy per trait is indicated with an asterisk. Standard deviations of mean accuracies ranged from 0.02 to 0.05.

[2]MY = milk yield; PY = protein yield; FY = fat yield; Fat% = fat percentage; Pro% = protein percentage.

[3]DSN = German Black Pied cattle; mean accuracies based on the 200 calculated accuracies for the 20 random 200K chips (20 random chips × 10 runs); WGS = whole-genome sequence.

the highest GEBV accuracy with WGS data. Again, as explained above, we strongly considered markers from GWAS being significantly associated with MY, PY, FY, Pro%, and Fat% in DSN when designing the novel DSN200K chip, while neglecting such variants being associated with SCS. Generally, for the low heritability test-day traits FPR and SCS, GEBV accuracies were much smaller compared with the lactation traits and ranged from 0.172 to 0.185 for FPR, and from 0.156 to 0.161 for SCS. For both test-day traits, we used the first record from official test-day recording after calving, representing a quite large phenotypic and residual variance (Table 1), possibly explaining the low GEBV accuracies.

Table 6 presents the mean GEBV accuracies from RSV for conformation traits. The number of records for conformation traits in our genomic prediction models was limited to 1,811 cows. Nevertheless, we achieved moderate GEBV accuracies for the moderate heritability type traits up to 0.41 for stature based on WGS and the DSN200K. For 10 of the 22 conformation traits, GEBV accuracies were highest with WGS data. For feet and legs, the GEBV accuracies were identical with the DSN200K chip, the random 200K chip and WGS. The accuracy was also identical with the random 200K chip and WGS data for the genetic evaluation of body depth. Genomic estimated breeding value accuracies were improved up to 0.012 when using the novel designed DSN200K chip compared with the WGS data for 11 of the 22 conformation traits (Figure 1). As shown in Figure 1, improvements in GEBV accuracies from the 50K chip to the DSN200K chip were achieved for the 3 trait categories 305-d lactation traits, test-day FPR and SCS, and conformation composite traits. For the linear type traits rump angle, udder depth, teat placement front, central ligament and BCS, the DSN200K implied slightly lower accuracies of 0.001 to 0.003 compared with the 50K chip (Figure 1). For conformation traits, even the maximal difference in accuracies (0.012) was smaller than the minimal SD across replicates of 0.03, again indicating nonsignificant effects of the chosen marker panel.

DISCUSSION

Genetic Parameter Estimates

In accordance with our results, Veerkamp et al. (2016) reported a small increase in genomic heritabilities of less than 0.04 for PY and SCS when using WGS data instead of the 50K chip. Gonzalez-Recio et al. (2015) showed that sequence variants captured 2–5% additional genetic variance for milk, fat, and PY compared with 600K SNP chip data. In Duroc pigs, Zhang et al. (2018) observed a significant increase of 8–13% in genomic heritabilities for moderate heritable traits (average daily gain, backfat depth, feed intake, loin muscle depth) when using imputed WGS data compared with 80K or 650K marker densities. For PY of HF dairy cows and applying GBLUP, van Binsbergen et al. (2015) estimated heritabilities of 0.94 based on BovineHD SNP chip data, of 0.97 based on imputed HD data (50K to ∼777K) and of 0.98 based on imputed WGS data. As pointed out by van Binsbergen et al. (2015), heritabilities were close to 1 when using deregressed proofs for PY due to small residual variances. We expected a larger heritability for SCS with the DSN200K compared with the 50K chip because we

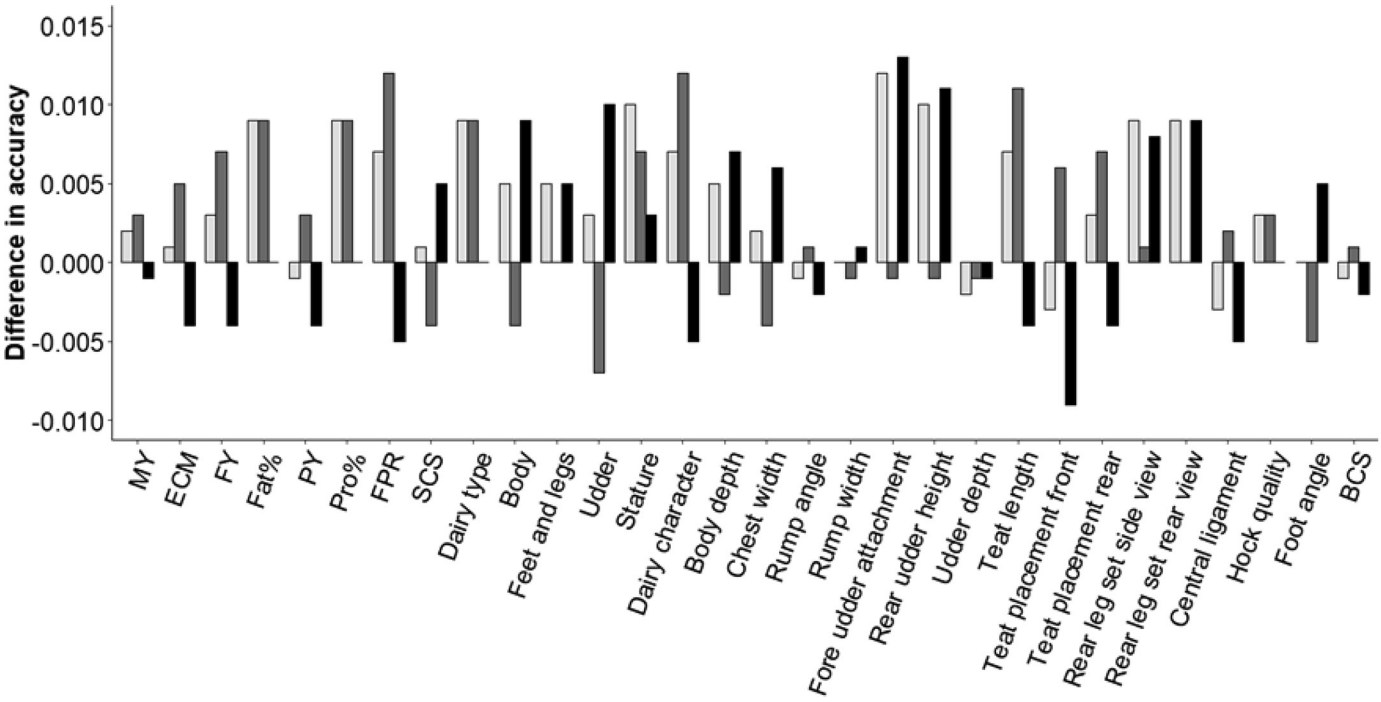

Figure 1. Differences in prediction accuracy (mean of all 10 runs from cross-validation) for 305-d milk production traits, test-day traits, and conformation traits. Light gray bars = accuracy of the German Black Pied cattle 200K chip (DSN200K) compared with the 50K Illumina chip; dark gray bars = accuracy of the DSN200K chip compared with the imputed whole-genome sequence (WGS) data; black bars = accuracy of the WGS data compared with the 50K Illumina chip. Milk yield (MY) = 305-d milk yield; PY = 305-d protein yield; FY = 305-d fat yield; Fat% = 305-d fat percentage; Pro% = 305-d protein percentage; ECM = 305-d energy-corrected milk; FPR = fat-to-protein ratio at first test-day after calving; SCS = somatic cell score at first test-day after calving.

considered variants associated with mastitis resistance in DSN when designing the breed-specific 200K SNP chip (Neumann et al., 2021). In contrast, there is only a moderate genetic correlation between SCS and clinical mastitis of ~0.75, explaining the differences in selection signals in GWAS for both traits (Naderi et al., 2020). In our study, we analyzed first test-day SCS, but some other studies indicated only accurate predictions for clinical mastitis when focusing on a longitudinal somatic cell data structure, in the ideal case with repeated measurements from the different udder quarters (Wagner et al., 2021). Nevertheless, Sender et al. (2013) and Welderufael et al. (2018) identified different QTL on the same chromosome for SCS and clinical mastitis, indicating that genetic variance explained by SNPs for clinical mastitis may not fully explains the genetic variance for SCS. When designing our novel DSN200K, we included only 34,039 SNPs from the commercial 50K chip (Neumann et al., 2021). Hence, it might be possible that some of the excluded ~20,000 SNPs located on the 50K chip but not included in the DSN200K, explain parts of the genetic variance for SCS (because SCS was not considered in the cascade of the SNP chip development). Overall, WGS data only contributed to minor changes in genetic parameter estimates (also for the conformation traits in the present study), but has potential to identify breed-specific variants and genes for low heritable and complex traits. Veerkamp et al. (2016) showed that the heritability for PY and SCS significantly increased when considering variants selected from prior GWAS.

Genomic Predictions

Mean GEBV accuracies for 305-d production traits in the small DSN population were slightly lower compared with the accuracies for MY, PY, and FY from larger training sets of HF cows (Habier et al., 2010; Erbe et al., 2012; van den Berg et al., 2017). In our study, we used a GBLUP model to estimate GEBV, but Bayesian methods might be more appropriate to analyze sequence data for some of the studied traits. In this regard, Van Binsbergen et al. (2015) reported an increase in prediction reliability of 0.01 when applying Bayesian stochastic search variable selection models compared with GBLUP. Chiaia et al. (2017) and Bhuiyan et al. (2018) indicated that the BayesR method performed substantially better compared with

Table 6. Accuracy of genomic prediction for type trait composites and linear conformation traits using different marker densities[1]

	Marker density[2]			
Trait	50K	DSN 200K	Random 200K (1–20)	WGS
Type trait composite				
Dairy type	0.339	*0.348	0.338	0.339
Body type	0.323	0.328	0.331	*0.332
Feet and legs	0.129	*0.134	*0.134	*0.134
Udder	0.262	0.265	0.270	*0.272
Linear trait				
Stature	0.398	*0.408	0.398	0.401
Dairy character	0.264	*0.271	0.260	0.259
Body depth	0.319	0.324	0.325	*0.326
Chest width	0.232	0.234	*0.239	0.238
Rump angle	0.316	0.315	*0.317	0.314
Rump width	0.274	0.274	0.274	*0.275
Fore udder attachment	0.225	0.237	0.236	*0.238
Rear udder height	0.225	0.235	0.234	*0.236
Udder depth	0.277	0.275	0.276	*0.278
Teat length	0.340	*0.347	0.335	0.336
Teat placement front	*0.385	0.382	0.377	0.376
Teat placement rear	0.292	*0.295	0.285	0.288
Rear leg set side view	0.197	*0.206	0.205	*0.206
Rear leg set rear view	0.148	*0.157	0.156	*0.157
Central ligament	*0.193	0.190	0.190	0.188
Hock quality	0.213	*0.216	0.214	0.213
Foot angle	0.039	0.039	0.042	*0.044
BCS	*0.299	*0.299	*0.299	*0.299

[1]The accuracy presents the mean of all 10 runs from cross-validation per each trait and marker density. The highest accuracy per trait is indicated with an asterisk. Standard deviations of mean accuracies ranged from 0.03 to 0.06.

[2]DSN = German Black Pied cattle; mean accuracies based on the 200 calculated accuracies for the 20 random 200K chips (20 random chips × 10 runs); WGS = whole-genome sequence.

GBLUP for traits that are controlled by SNPs or few major genes with large effects.

Moreover, GEBV accuracies might be influenced by the chosen cross-validation strategy. We used the RSV method with random selection and replacements, which was superior over classical cross-validations (e.g., in vegetation mappings; Lyons et al., 2018). Especially for small-sized data sets with up to 1,000 records, Berrar et al. (2006) and Baba et al. (2021) suggested RSV for genomic predictions. Thus, we assume that RSV has an advantage over other classical cross-validation methods for cattle breeds with a small population size such as DSN. Most commonly, RSV partitions the data set into training sets with 90% of animals and validation sets with 10% of animals (González-Camacho et al., 2012; Morota and Gianola, 2014). In our study, we adapted RSV by portioning the data set in 80% of cows in the training set and 20% of cows in the validation set. With decreased proportion of animals in the training set, the accuracies are expected to be lower as previously shown by Luan et al. (2009) for milk production traits. Accordingly, Zhu et al. (2017), calculated larger GEBV accuracies in 10-fold cross-validation with 90% of cows in the training set compared with 5-fold cross-validation with 80% of animals in the training set. However, with 10% animals in the validation set, standard deviations across replicates tend to be larger compared with 20% allocation strategy.

One explanation for the moderate accuracies from the smaller conformation data set might be the genetic relatedness between the cows in the training and validation set (e.g., Pérez-Cabal et al., 2012), which was higher for the conformation trait data set (average genetic relationship of 3.40% between training and validation sets in the WGS scenario) compared with the production traits (average genetic relationship of 2.04% between training and validation sets in the WGS scenarios for FPR and SCS, and of 2.80% for lactation traits). A further point explaining the acceptable accuracies for conformation traits with moderate heritability in a comparable small cow training set addresses the herd structure. German Black Pied cattle conformation trait scoring considered a subsample including only modern large-scale herds. Herold et al. (2021) identified herd size with associated feeding and management strategies in the large-scale herds from the eastern part of Germany as major components affecting genetic differentiation and accuracy of genetic evaluations. For all conformation traits with

estimated heritabilities close to zero (i.e., feet and legs, foot angle, rear leg set side view, and rear leg set rear view) GEBV accuracies were largest with WGS data. Accordingly, Druet et al. (2014) indicated the strongest increase in GEBV accuracies when using WGS data for low heritability traits. For simple traits which are controlled by one or only a few genes with major effects, a high marker density implied a smaller GEBV accuracy if the quantitative trait nucleotide is directly included in the model (Zhang et al., 2019).

General Aspects

Regarding genetic evaluations in purebred populations, only minor advantages in prediction accuracies were identified when using WGS data instead of dense SNP marker panels (e.g., VanRaden et al., 2017), which was supported in the present study for the DSN breed. Possible benefits of WGS data, mostly based on results from simulations, address multibreed predictions (Iheshiulor et al., 2016). Accordingly, Toosi et al. (2010) and Su et al. (2012) reported that high-density markers are more beneficial for genomic predictions across populations than within populations. Druet et al. (2014) pointed out that the advantage of WGS data in genomic predictions is greater for populations with a large effective population and with a large number of phenotyped and genotyped individuals. Otherwise, the small effects of causative mutations will be estimated with too much error. In addition, it is possible to preselect breed-specific variants from sequence data for genomic selection. Liu et al. (2020) integrated SNPs from imputed WGS data (e.g., regulatory regions or genes, peaks of QTLs) into the 50K chip, contributing to improved genomic predictions in Danish Jersey cattle. Raymond et al. (2018) showed that GEBV accuracies were significantly higher when considering pre-selected markers from GWAS based on WGS data than using WGS alone.

Overall, the benefits of WGS with regard to improved prediction accuracies were quite small. Nevertheless, WGS data are imperative to identify causal variants for functional mutations (Hozé et al., 2013), and to improve LD between markers and causal mutations (Lund et al., 2016). Finally, WGS data contributes to the understanding of the mechanisms of selection and genetic drift in small populations (Hulsegge et al., 2022), and to an improved management of genetic diversity (Neumann et al., 2023).

As indicated above, it is important to note that 304 sequenced cows were considered when developing the breed-specific DSN200K chip, implying possible biased estimates in the subsequent genetic evaluations. Weigel (2017) described the overfitting of data points in the training sets as general challenge in genomic predictions. Consequently, Jia (2017) strongly emphasized the importance of sophisticated cross-validations. In the present study, we considered 10 replicates per SNP chip and trait, which is a common number in several other genomic prediction studies, also for methodological evaluations based on simulated data (e.g., Naderi et al., 2016). Nevertheless, it is really imperative to have independent subsets when pre-selecting markers for the design of SNP chips in genome-wide association analyses and in ongoing genomic predictions, but the practical feasibility is a great challenge in breeds with small population size (May et al., 2022). In this context, Schöpke and Swalve (2016) emphasized the "limited information" in endangered breeds, from a phenotype as well as from a genotype perspective.

CONCLUSIONS

Estimated genetic parameters and GEBV accuracies were very similar from the different marker densities for 305-d lactation production, test-day FPR and SCS, and conformation traits. Nevertheless, especially for production traits, consideration of WGS data or the DSN200K contributed to increased heritabilities and GEBV accuracies, but increases were mostly in the range of estimated standard errors and nonsignificant. The improved prediction accuracies based on WGS or DSN200K are not sufficient to cover the extra costs for animal sequencing or using the breed-specific SNP chip in DSN. For conformation traits with moderate heritability (stature, teat length), acceptable prediction accuracies were realized even for a smaller sample size of genotyped or sequenced cows, indicating the positive effects of accurate trait recording in specific large-scale contract herds. Overall, WGS and the DSN200K chip harbor breed-specific variants, which are valuable for studying causal variants in the endangered DSN population, but the extra benefit for genetic evaluations is too small.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the German Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany, granted by the Federal Office for Agriculture and Food (BLE; grant number FKZ 2818BM090 and −091; Bonn, Germany). The authors have not stated any conflicts of interest.

REFERENCES

Aguilar, I. 2018. Creation of genomic relationship matrices and BLUPF90 "user file." Accessed Jun. 15, 2022. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=user_file.pdf.

Ali, A. K. A., and G. E. Shook. 1980. An optimum transformation for somatic cell concentration in milk. J. Dairy Sci. 63:487–490. https://doi.org/10.3168/jds.S0022-0302(80)82959-6.

Baba, T., S. Pegolo, L. F. M. Mota, F. Peñagaricano, G. Bittante, A. Cecchinato, and G. Morota. 2021. Integrating genomic and infrared spectral data improves the prediction of milk protein composition in dairy cattle. Genet. Sel. Evol. 53:29. https://doi.org/10.1186/s12711-021-00620-7.

Bergk, N., and H. H. Swalve. 2011. Fat-to-protein-ratio in early lactation as an indicator of herdlife for first lactation dairy cows. Zuchtungskunde 83:89–103.

Berrar, D., I. Bradbury, and W. Dubitzky. 2006. Avoiding model selection bias in small-sample genomic datasets. Bioinformatics 22:1245–1250. https://doi.org/10.1093/bioinformatics/btl066.

Bhuiyan, M. S. A., Y. K. Kim, H. J. Kim, D. H. Lee, S. H. Lee, H. B. Yoon, and S. H. Lee. 2018. Genome-wide association study and prediction of genomic breeding values for fatty-acid composition in Koran Hanwoo cattle using a high-density single-nucelotide polymorphism array. J. Anim. Sci. 96:4063–4075. https://doi.org/10.1093/jas/sky280.

Brade, W., and E. Brade. 2013. Zuchtgeschichte der Deutschen Holsteinrinder. Berichte über Landwirtschaft-Zeitschrift für Agrarpolitik und Landwirtschaft 91:1–41. https://doi.org/10.12767/buel.v91i2.25. [In German]

Browning, B. L., Y. Zhou, and S. R. Browning. 2018. A one-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. 103:338–348. https://doi.org/10.1016/j.ajhg.2018.07.015.

Buch, L. H., M. Kargo, P. Berg, J. Lassen, and A. C. Sørensen. 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal 6:880–886. https://doi.org/10.1017/S1751731111002205.

Chiaia, H. L. J., E. Peripoli, R. M. O. Silva, C. Aboujaoude, F. L. B. Feitosa, M. V. A. Lemos, M. P. Berton, B. F. Olivieri, R. Espigolan, R. L. Tonussi, D. G. M. Gordo, T. Bresolin, A. F. B. Magalhães, G. A. F. Júnior, L. G. Albuquerque, H. N. Oliveira, J. J. M. Furlan, A. M. Ferrinho, L. F. Mueller, H. Tonhati, A. S. C. Pereira, and F. Baldi. 2017. Genomic prediction for beef fatty acid profile in Nellore cattle. Meat Sci. 128:60–67. https://doi.org/10.1016/j.meatsci.2017.02.007.

Druet, T., I. M. Macleod, and B. J. Hayes. 2014. Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. Heredity 112:39–47. https://doi.org/10.1038/hdy.2013.13.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114–4129. https://doi.org/10.3168/jds.2011-5019.

Frischknecht, M., T. H. E. Meuwissen, B. Bapst, F. R. Seefried, C. Flury, D. Garrick, H. Signer-Hasler, C. Stricker, A. Bieber, R. Fries, I. Russ, J. Sölkner, A. Bagnato, and B. Gredler-Grandl. 2018. Short communication: Genomic prediction using imputed whole-genome sequence variants in Brown Swiss cattle. J. Dairy Sci. 101:1292–1296. https://doi.org/10.3168/jds.2017-12890.

González-Camacho, J. M., G. de los Campos, P. Pérez, D. Gianola, J. E. Cairns, G. Mahuku, R. Babu, and J. Crossa. 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. Theor. Appl. Genet. 125:759–771. https://doi.org/10.1007/s00122-012-1868-9.

Gonzalez-Recio, O., H. D. Daetwyler, I. M. MacLeod, J. E. Pryce, P. J. Bowman, B. J. Hayes, and M. E. Goddard. 2015. Rare variants in transcript and potential regulatory regions explain a small percentage of the missing heritability of complex traits. PLoS One 10:e0143945. https://doi.org/10.1371/journal.pone.0143945.

Habier, D., J. Tetens, F.-R. Seefried, P. Lichtner, and G. Thaller. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet. Sel. Evol. 42:5. https://doi.org/10.1186/1297-9686-42-5.

Herold, J., K. Brügemann, and S. König. 2021. Herd clustering strategies and corresponding genetic evaluations based on social-ecological characteristics for a local endangered cattle breed. Arch. Tierzucht 64:187–198. https://doi.org/10.5194/aab-64-187-2021.

Hozé, C., M.-N. Fouilloux, E. Venot, F. Guillaume, R. Dassonneville, S. Fritz, V. Ducrocq, F. Phocas, D. Boichard, and P. Croiseau. 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genet. Sel. Evol. 45:33. https://doi.org/10.1186/1297-9686-45-33.

Hulsegge, I., K. Oldenbroek, A. Bouwman, R. Veerkamp, and J. Windig. 2022. Selection and drift: A comparison between historic and recent Dutch Friesian cattle and recent Holstein Friesian using WGS data. Animals (Basel) 12:329. https://doi.org/10.3390/ani12030329.

Iheshiulor, O. O., J. A. Woolliams, X. Yu, R. Wellmann, and T. H. E. Meuwissen. 2016. Within- and across-breed genomic prediction using whole-genome sequence and single-nucleotide polymorphism panels. Genet. Sel. Evol. 48:15. https://doi.org/10.1186/s12711-016-0193-1.

Jaeger, M., K. Brügemann, S. Naderi, H. Brandt, and S. König. 2019. Variance heterogeneity and genotype by environment interactions in native Black and White dual-purpose cattle for different herd allocation schemes. Animal 13:2146–2155. https://doi.org/10.1017/S1751731119000144.

Jaeger, M., C. Scheper, S. König, and K. Brügemann. 2018. Inbreeding and genetic relationships of the endangered dual-purpose Black and White cattle breed based on own genetic breed percentage calculations. Zuchtungskunde 90:262–279.

Jia, Z. 2017. Controlling the overfitting of heritability in genomic selection through cross validation. Sci. Rep. 7:13678. https://doi.org/10.1038/s41598-017-14070-z.

Korkuć, P., D. Arends, and G. A. Brockmann. 2019. Finding the optimal imputation strategy for small cattle populations. Front. Genet. 10:52. https://doi.org/10.3389/fgene.2019.00052.

Korkuć, P., D. Arends, K. May, S. König, and G. A. Brockmann. 2021. Genomic loci affecting milk production in German Black Pied cattle (DSN). Front. Genet. 12:640039. https://doi.org/10.3389/fgene.2021.640039.

Liu, A., M. S. Lund, D. Boichard, E. Karaman, S. Fritz, G. P. Aamand, U. S. Nielsen, Y. Wang, and G. Su. 2020. Improvement of genomic prediction by integrating additional single nucleotide polymorphisms selected from imputed whole genome sequencing data. Heredity 124:37–49. https://doi.org/10.1038/s41437-019-0246-7.

Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126. https://doi.org/10.1534/genetics.109.107391.

Lund, M. S., I. van den Berg, P. Ma, R. F. Brøndum, and G. Su. 2016. Review: How to improve genomic predictions in small dairy cattle populations. Animal 10:1042–1049. https://doi.org/10.1017/S1751731115003031.

Lyons, M. B., D. A. Keith, S. R. Phinn, T. J. Mason, and J. Elith. 2018. A comparison of resampling methods for remote sensing classification and accuracy assessment. Remote Sens. Environ. 208:145–153. https://doi.org/10.1016/j.rse.2018.02.026.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4:e5350. https://doi.org/10.1371/journal.pone.0005350.

May, K., M. J. Wolf, G. B. Neumann, P. Korkuć, G. A. Brockmann, and S. König. 2022. Genetic evaluations and genomic characteristics for local cattle using genome sequences, 50K, and a specific SNP chip. 12th World Congress on Genetics Applied to Livestock Production, Rotterdam, Netherlands.

McLaren, W., L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. 2016. The Ensembl variant effect predictor. Genome Biol. 17:122. https://doi.org/10.1186/s13059-016-0974-4.

Meuwissen, T., and M. E. Goddard. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–631. https://doi.org/10.1534/genetics.110.116590.

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. Manual for BLUPF90 family of programs. Accessed Jun. 21, 2021. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf.

Morota, G., and D. Gianola. 2014. Kernel-based whole-genome prediction of complex traits: A review. Front. Genet. 5:363. https://doi.org/10.3389/fgene.2014.00363.

Naderi, S., M. H. Moradi, M. Farhadian, T. Yin, M. Jaeger, C. Scheper, P. Korkuc, G. A. Brockmann, S. König, and K. May. 2020. Assessing selection signatures within and between selected lines of dual-purpose black and white and German Holstein cattle. Anim. Genet. 51:391–408. https://doi.org/10.1111/age.12925.

Naderi, S., T. Yin, and S. König. 2016. Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. J. Dairy Sci. 99:7261–7273. https://doi.org/10.3168/jds.2016-10887.

Neumann, G., P. Korkuć, D. Arends, M. J. Wolf, K. May, S. König, and G. A. Brockmann. 2023. Genomic diversity and relationship analyses of endangered German Black Pied cattle (DSN) to 68 other taurine breeds based on whole-genome sequencing. Front. Genet. 13:993959.

Neumann, G. B., P. Korkuć, D. Arends, M. J. Wolf, K. May, M. Reißmann, S. Elzaki, S. König, and G. A. Brockmann. 2021. Design and performance of a bovine 200 k SNP chip developed for endangered German Black Pied cattle (DSN). BMC Genomics 22:905. https://doi.org/10.1186/s12864-021-08237-2.

Pérez-Cabal, M. A., A. I. Vazquez, D. Gianola, G. J. M. Rosa, and K. A. Weigel. 2012. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. Front. Genet. 3:27. https://doi.org/10.3389/fgene.2012.00027.

Picard, R. R., and R. D. Cook. 1984. Cross-validation of regression models. J. Am. Stat. Assoc. 79:575–583. https://doi.org/10.1080/01621459.1984.10478083.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575. https://doi.org/10.1086/519795.

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Raymond, B., A. C. Bouwman, C. Schrooten, J. Houwing-Duistermaat, and R. F. Veerkamp. 2018. Utility of whole-genome sequence data for across-breed genomic prediction. Genet. Sel. Evol. 50:27. https://doi.org/10.1186/s12711-018-0396-8.

Schöpke, K., and H. H. Swalve. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal 10:1050–1060. https://doi.org/10.1017/S1751731116000410.

Sender, G., A. Korwin-Kossakowska, A. Pawlik, K. G. A. Hameed, and J. Oprządek. 2013. Genetic basis of mastitis resistance in dairy cattle – A review. Ann. Anim. Sci. 13:663–673. https://doi.org/10.2478/aoas-2013-0043.

Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (∼54,000) and high-density (∼777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy cattle populations. J. Dairy Sci. 95:4657–4665. https://doi.org/10.3168/jds.2012-5379.

Toosi, A., R. L. Fernando, and J. C. M. Dekkers. 2010. Genomic selection in admixed and crossbred populations. J. Anim. Sci. 88:32–46. https://doi.org/10.2527/jas.2009-1975.

van Binsbergen, R., M. P. L. Calus, M. C. A. M. Bink, F. A. van Eeuwijk, C. Schrooten, and R. F. Veerkamp. 2015. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. Genet. Sel. Evol. 47:71. https://doi.org/10.1186/s12711-015-0149-x.

van den Berg, I., P. J. Bowman, I. M. MacLeod, B. J. Hayes, T. Wang, S. Bolormaa, and M. E. Goddard. 2017. Multi-breed genomic prediction using Bayes R with sequence data and dropping variants with a small effect. Genet. Sel. Evol. 49:70. https://doi.org/10.1186/s12711-017-0347-9.

VanRaden, P. M., M. E. Tooker, J. R. O'Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. Genet. Sel. Evol. 49:32. https://doi.org/10.1186/s12711-017-0307-4.

Veerkamp, R. F., A. C. Bouwman, S. Schrooten, and P. L. Calus. 2016. Genomic prediction using preselected variants from a GWAS with whole-genome sequence data in Holstein-Friesian cattle. Genet. Sel. Evol. 48:95. https://doi.org/10.1186/s12711-016-0274-1.

Wagner, P., K. Brügemann, T. Yin, P. Engel, C. Weimann, K. Schlez, and S. König. 2021. Microscopic differential cell count and specific mastitis pathogens in cow milk from compost-bedded pack barns and cubicle barns. J. Dairy Res. 88:413–419. https://doi.org/10.1017/S0022029921000844.

Weigel, K. A. 2017. Genomic selection of dairy cattle: A review of methods, strategies, and impact. J. Anim. Breed. Genom. 1:1–15. https://doi.org/10.12972/jabng.20170001.

Welderufael, B. G., P. Løvendahl, D.-J. de Koning, L. L. G. Janss, and W. F. Fikse. 2018. Genome-wide association study for susceptibility to and recoverability from mastitis in Danish Holstein cows. Front. Genet. 9:141. https://doi.org/10.3389/fgene.2018.00141.

Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher. 2011. GCTA: A tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Zhang, C., R. A. Kemp, P. Stothard, Z. Wang, N. Boddicker, K. Krivushin, J. Dekkers, and G. Plastow. 2018. Genomic evaluation of feed efficiency component traits in Duroc pigs using 80K, 650K, and whole-genome sequence variants. Genet. Sel. Evol. 50:14. https://doi.org/10.1186/s12711-018-0387-9.

Zhang, H., L. Yin, M. Wang, X. Yuan, and X. Liu. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. Front. Genet. 10:189. https://doi.org/10.3389/fgene.2019.00189.

Zhu, B., J. Zhang, H. Niu, L. Guan, P. Guo, L. Xu, Y. Chen, L. Zhang, H. Gao, X. Gao, and J. Li. 2017. Effects of marker density and minor allele frequency on genomic prediction for growth traits in Chinese Simmental beef cattle. J. Integr. Agric. 16:911–920. https://doi.org/10.1016/S2095-3119(16)61474-0.

Zumwald, J., M. Braunschweig, T. Nemecek, B. Schüpbach, P. Jeanneret, A. Roesch, P. Hofstetter, and B. Reidy. 2018. Life cycle assessment of grassland-based dairy production systems in Switzerland. Sustainable Meat and Milk Production from Grasslands: Proceedings of the 27th General Meeting of the European Grassland Federation. Grassland Science in Europe 23:541–543.

KAPTIEL 4

Single-step breeding value estimations and optimum contribution selection in endangered dual-purpose German Black Pied cattle (DSN) using a breed specific SNP chip

Manuel J. Wolf[1], Tong Yin[1], Guilherme B. Neumann[2,3], Paula Korkuć[2], Gudrun A. Brockmann[2] and Sven König[1]

[1] Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany

[2] Albrecht Daniel Thaer-Institute for Agricultural and Horticultural Sciences, Animal Breeding Biology and Molecular Genetics, Humboldt-Universität of Berlin, 10115 Berlin, Germany

[3] Leibnitz Institute for Zoo and Wildlife Research,10315 Berlin, Germany

Journal of Animal Breeding and Genetics

Submitted: 24.06.2024

Summary

The aims of the present study were to perform single-step genomic predictions in the dual-purpose *German Black Pied cattle* (DSN) breed considering a DSN specific SNP chip (DSN_200K), and to use the corresponding estimated breeding values (EBV) in ongoing optimum genetic contribution (OGC) selection. All results were compared with the application of the commercial *Illumina*® *BovineSNP50 BeadChip* (50K). The traits of interest in the present study (due to the differing breeding history of these traits in the past) included 305-d lactation protein percentage (Pro%) of 9029 DSN cows, fat-to-protein ratio (FPR) from the first test-day of 8773 DSN cows, and stature (STAT) measured in cm of 4409 DSN cows. The DSN cows represented the calving years 2008 to 2019. Genotyping of 2797 DSN animals was conducted using both the DSN_200K and the 50K. From the genotyped animals, a subset of 1800 cows had phenotypic records for all three traits FPR, Pro% and STAT. Heritabilities from the single-step genetic parameter estimations were quite large for Pro% (0.69) and STAT (0.78), but small for FPR (0.11). The choice of the SNP chip only had minor effects on variance components, heritabilities and EBVs. Furthermore, genetic parameters were very similar from genetic statistical models additionally considering a linear regression on pedigree-based inbreeding coefficients. OGC selection was applied to a pool of 1125 pre-selected bull sires (BS) and bull dams (BD). A more relaxed genetic relationship constraint was associated with favorable effects on the average EBVs for Pro%, FPR and STAT, and a declining number of selected BS. The gains in genetic merit were marginal when relaxing the constraint at 0.06 for the genetic relationships or higher. The same associations were found for an overall breeding index (I-DSN), considering the three traits with equal weights. Consequently, we suggested OGC applications with a genetic relationship constraint of 0.06, which contributed to genetic gain in I-DSN of 17.9%, and to increased diversity due to an increased number of BS, when compared to the current practical elite animal selection scheme. A large number of finally selected BS and BD was identical when either using EBV from the DSN_200K or from the 50K. From such perspective, we only see marginal extra value for the specific DSN SNP-chip application.

Key words: Local cattle, single-step genetic evaluation, optimum genetic contributions

INTRODUCTION

Local dual-purpose *German Black Pied cattle* (**DSN**, German: Deutsches Schwarzbuntes Niederungsrind) with a population size of ~2500 cows is the founder breed of modern Holstein Friesian (**HF**) in Germany (Jaeger et al., 2018a). Due to the breeding history and still existing genetic relationships with HF (Naderi et al., 2020), official genetic evaluations of DSN are embedded in the infrastructure as developed for HF. Nevertheless, divergent breeding strategies in DSN and HF over the past decades with focus on different trait categories contributed to a genetic differentiation between both selection lines. The genetic differentiation resulted in linkage disequilibrium pattern (Naderi et al., 2020), in allele frequencies for important genes such as *DGAT1* (Korkuć et al., 2021), and in allele substitution and dominance effects for selected SNP (May et al., 2019), which differed in the HF and the DSN population.

The breeding goal definition for DSN during the past decades consistently emphasized lactation protein percentage (**Pro%**). A small body size reflected via measurements in cm for stature (**STAT**) was a breeding objective from the late 1970s until 2015, but in commercial farms, intra-herd selection strategies recently favored taller cows. The idea behind a breeding focus on small STAT was to be complementary with the conformation characteristics of modern HF cows, and to achieve low body weights, i.e., following the cattle breeding approach as implemented for pasture based systems in New Zealand (Alemu et al., 2023). Breeding on small-sized cattle might favor energy efficiency aspects and grassland suitability, but for the dual-purpose DSN, also carcass traits and body weight play an economic role (Meier et al., 2021). Consequently, current DSN breeding efforts favor an increase of body measurements and weights. The fat-to-protein ratio (**FPR**) in early lactation has been suggested as a reliable indicator for metabolic stability (Jaeger et al., 2018b; Klein et al., 2019). Furthermore, Bergk and Swalve (2011) identified strong associations between FPR from test-days in early lactations with cow longevity. Consequently, practical DSN breeding aims on the consideration of FPR, but the intermediate trait optimum (a low FPR indicates acidosis, a high FPR indicates ketosis), implies some difficulties in this regard. Overall, different selection strategies have been applied for Pro%, STAT and FPR in the DSN breed in the past. Hence, these traits are interesting for investigations addressing selection and mating designs.

The commercial *Illumina® BovineSNP50 BeadChip* (hereafter referred to as **50K**) as designed for the large dairy populations does not fully reflect the genetic characteristics and the importance of specific traits in DSN. Consequently, Neumann et al. (2021) developed a breed specific *DSN 200K SNP chip* (**DSN_200K**), which especially considered variants and genomic regions with major

effects on Pro%, STAT and FPR (May et al., 2022). The DSN_200K was used for the estimation of genetic parameters for production and conformation traits based on the genomic relationship matrix (Wolf et al., 2023), and for genome wide association studies (Korkuć et al., 2023). Estimated heritabilities were slightly larger and association signals were clearer compared to 50K SNP chip applications. Not all DSN cows are genotyped, but trait recording comprises the whole population. Hence, accuracies of genomic predictions might increase when combining genomic and pedigree-based approaches in single-step genetic evaluations. Such effects were clearly elaborated for genetic evaluations in organic populations with scarcely genotyped cows (Shabalina et al., 2020). The methodology to combine pedigree and genomic relationships through the so-called **H**-matrix is described, by, e.g., Legarra et al. (2014).

An issue following the genetic evaluations addresses the selection and mating scheme. Especially for small-sized populations, it is imperative to monitor and to manage inbreeding and genetic relationships in a long-term perspective. In the DSN population, selection for females and natural service sires is mostly based on phenotype characteristics, and on EBVs for the three major traits for cow and bull sires (Jaeger et al., 2018a). Some large scale DSN herds utilize mating software for the selection of cow sires for intra-herd replacements, but already existing close relationships among sires from artificial insemination (**AI**) programs narrow the availability of alternative or outcross genetics. Based on historical datasets spanning a 40 years period from 1975 to 2005, trait responses on inbreeding alterations were quite stable, disproving any significant inbreeding depressions for primary as well as for functional traits (Jaeger et al., 2018a). Nevertheless, high relationships of some recent influential sires with the current cow population point to detrimental inbreeding effects in the near future. Also in an international study simultaneously considering DSN populations from Germany, Poland and The Netherlands, high genetic relationships among the most popular bulls used for AI, were identified (Jaeger et al., 2018b). Close genetic relationships among influential sires imply a time-lagged increase of genetic connectedness, a loss of genetic diversity and associated possible future inbreeding depressions in the DSN cow population. Hence, it is imperative that selection strategies aiming on the conservation of genetic variability in the long term perspective regulate genetic relationships among bull sires and bull dams. In this regard, the optimum genetic contribution (**OGC**) concept as developed by Meuwissen (1997) successfully has been applied in dairy cattle breeding schemes, i.e., to determine optimal mating frequencies for bull sires and bull dams (König and Simianer, 2006).

Following all the constraints, challenges and background information as stated above, the aims of the present study were a) to perform single-step genetic evaluations for DSN on the basis of the

newly developed DSN_200K for the traits FPR, Pro% and STAT, and b) to use the estimated breeding values (**EBV**) and an overall index for an OGC approach to select DSN bull sires and bull dams. Genetic parameters from the single-step genomic predictions, and genetic merit and the number of selected bull sires from the different OGC scenarios, were compared with commercial 50K chip applications.

MATERIALS AND METHODS

Phenotypes

Phenotypic records of first parity DSN cows comprised the calving years 2008 to 2019. In this study, we used FPR from the first official test-day after calving from 9029 cows, 305-d lactation records for Pro% from 8373 cows, and STAT measurements from 4409 cows. For FPR, previous data editing excluded first test-days outside the period between 5 and 40 days after calving. Only completed 305-d lactations were considered for Pro%. In a subset of the 8 largest herds, STAT in cm was measured in the process of official type trait classification. The descriptive trait statistics are given in Table 1. The DSN cows recorded for FPR and Pro% were kept in 13 herds, with an average herd size of 63 cows and a minimum of 23 cows per herd and a maximum of 473 cows per herd. However, apart from 2 large-scale farms located in former East Germany (average herd size: 378 cows), the DSN cows were mostly kept in small-sized family farms with focus on organic or grazing systems.

SNP chips and genotyping

Genotyping considered 2797 DSN animals. From the genotyped animals, a subset of 1800 cows had phenotypic records for all three traits FPR, Pro% and STAT. Further 997 genotyped animals were sires, paternal or maternal grandsires of the cows with phenotypes. Genotyping was performed using two different SNP chips. First, the standard 50K chip with ~54,000 markers as commonly used for genetic evaluations of the commercial breeds HF and Simmental, was used. The second SNP chip was the DSN specific DSN_200K. Our approach in developing the DSN_200K based on whole-genome sequence data with in total 20,587,181 sequence variants is outlined by Neumann et al. (2021). For SNP quality control for the genotyped DSN cattle in the present study, we used the software package PLINK (Purcell et al., 2007). Filtering of SNP genotypes comprised the standard criteria, i.e., exclusion of SNPs with a call rate smaller than 0.01, a minor allele frequency smaller than 0.05 and significant deviation ($P < 1 \times 10^{-8}$) from Hardy-Weinberg equilibrium, and locations on a sex chromosome. After the filtering steps, 31,248 SNPs

(50K chip) and 152,493 SNPs (DSN_200K chip) from the 2797 animals, were available for the ongoing genomic analyses.

Inbreeding and genetic structure

Pedigrees were corrected and some gaps were filled using our own algorithm specifically developed for the DSN breed (Jaeger et al., 2018a). The DSN pedigree (only considering DSN animals with a DSN breed percentage larger than 90%) comprised in total 19,593 animals, traced back to the oldest founder animal born in 1906. The second oldest animal in the dataset was born in 1922. All birth years from 1952 to 2019 were represented in the pedigree.

For the genotyped animals, we computed the pedigree-based relationship matrix A using the CFC computer package (Sargolzaei et al., 2006). A pedigree-based inbreeding coefficient larger than 0 was calculated for 12,203 animals. The average inbreeding coefficient considering all animals from birth years 1952 to 2019 was 0.013, with a current value of 0.038 for animals born in 2019. For the cows with phenotypic records, the average inbreeding coefficient was 0.034. The average generation interval on the cow-dam pathway of selection for the cows with phenotypes comprised 5.9 years. Criteria reflecting the depth of the pedigree were calculated using CFC. The values for the longest ancestral path and the average number of discrete generation equivalents were 22 and 6.45, respectively. The pedigree completeness index according to MacCluer et al. (1983) for the cows with phenotypic records was larger than 96% for all birth years.

Single-step genetic evaluations

Genetic parameters and breeding values for FPR, Pro% and STAT were estimated applying single-step GBLUP methodology, as implemented in the packages PREGSF90 and POSTGSF90 from the BLUPF90 program family (Aguilar et al., 2014). In this regard, we defined the following single-trait animal model 1:

$$y = X\beta + Za + e \quad (1)$$

where y is a vector of observations for FPR, Pro% or STAT; β is a vector of fixed effects including herd, calving year, calving season (spring, summer, autumn, winter) and age at first calving for FPR and Pro%, and additionally the classifier for STAT; a is a vector of random additive genetic effects, with $a \sim N(0, H\sigma^2_a)$, and σ^2_a denoting the additive genetic variance and H denoting the combined (pedigree and genomics) relationship matrix constructed according to Legarra et al. (2009); e is a vector of random residual effects with $e \sim N(0, I\sigma^2_e)$ with σ^2_e denoting the residual variance; and X and Z are the incidence matrices for fixed and additive genetic effects, respectively. The single-step genetic parameter estimations were conducted in consecutive runs

based on the 50K and the DSN_200K genotypes using the same pedigree, the same data, and the same base populations. We run two sub-models in this regard, with or without a linear regression on the pedigree-based inbreeding coefficient.

Optimum genetic contribution selection

The OGC application aims to maximize genetic gain, assuming that higher EBVs are also favorable in the sense of breeding, which implies challenges for traits with an intermediate optimum. Consequently, for FPR, we transformed the original EBVs to reflect the physiological and economic background in ongoing OGC applications. A high original FPR EBV indicates ketosis, and a low original FPR EBV indicates acidosis. According to Bergk and Swalve (2011), an intermediate original EBV for FPR of exactly 0 or an average phenotypic value of 1.1 was related with smallest frequencies for involuntary cow disposals. Consequently, for all animals, the difference of the maximal EBV for FPR (1.05 from the 50K genetic evaluation, 1.09 for the DSN_200K genetic evaluation) minus the absolute value of the original EBV for FPR, was calculated. Such procedure implied highest transformed FPR EBVs for animals with original EBV of 0 (= favorable in the sense of breeding), and smallest transformed FPR (transformed EBV = 0) for extreme animals.

For the OGC applications, we additionally constructed an overall index (I-DSN), considering the three EBVs for FPR, Pro% and STAT with equal weights (33.3% each). The I-DSN was standardized to a mean of 100 and a SD of 12 points.

The OGC approach focused on the selection of elite animals, i.e., bull sires (BS) and bull dams (BD), for a genomic breeding program as outlined by Schaeffer (2006). In this regard, we aimed on determining the optimal genetic contributions of BS and BD to generate the next cohort of young bulls (YB). The pool of possible selection candidates as pre-selected by the DSN breeding organization for genotyping comprised 1125 active animals (48 genotyped BS and 1077 genotyped BD from the 13 herds). The generated YB are used for AI and / or as natural service sires in a rotational system, implying that natural service sires are exchanged across herds in distinct intervals. Such strategy in utilizing natural service sires enables gene flow from specific sires in all DSN herds. The DSN breeding organization generates 15 YB per year. The guidelines of organic farming in Germany strongly restrict the utilization of reproduction biotechnologies, especially of embryo transfer. Hence, the application of selection schemes without embryo transfer and semen sexing requires the selection of 30 BD to produce 15 YB. Theoretically, in an AI program, one specific outstanding BS could be mated with all 30 BD. Nevertheless, we defined a maximal constraint in this regard, i.e., fixing the maximal genetic contribution of a male selection

candidate to 0.02 = 20%. The minimal genetic contribution of an individual selected BS is a single mating with one BD, i.e., 1/30 = 0.033 = 3.3%.

The applied OGC algorithm to determine optimal mating frequencies (= genetic contributions) for BS and BD followed the equations as developed by Meuwissen and Sonesson (1998), and as implemented in the respective software package GENCONT. Input parameters were the single-step EBVs for FPR, Pro% and STAT, and I-DSN of the 1125 potential selection candidates (stored in vector u in consecutive runs for the different traits) and the respective sex information. The genetic contributions of the selected elite animals are included in vector c. In brief, the maximization of selection response in each single trait implies to maximize c´u, by constraining the average genetic relationships through c'Ac in the range from 0.02 to 0.10 in consecutive runs. Major output criteria were the average EBV of selected elite animals, and the number of selected BS at the constraint for genetic relationships.

RESULTS

Single-step genetic parameters and breeding values

The variance components and the respective heritabilities for FPR, Pro% and STAT are given in Table 2. For the two different SNP chips (50K or DSN_200K) and from the two different sub-models (with or without a linear regression on the pedigree-based inbreeding coefficient), the genetic variances and residual variances were very similar for the same trait, implying heritabilities in the narrow range from 0.691 to 0.697 for Pro%, from 0.106 to 0.108 for FPR, and from 0.777 to 0.780 for STAT. Consideration of the pedigree-based inbreeding coefficient in the statistical model had no effect on variance components and heritabilities. All heritabilities were associated with small SE in the range from 0.016 (FPR) to 0.022 (Pro%).

For the same trait but different chip applications (50K versus DSN_200K), the mean, minimum and maximum EBVs as well as I-DSN were very similar (Table 3) for the whole dataset (including the cows with phenotypic records as specified in Table 1 and their sires), as well as for the group including the 1125 pre-selected animals for OGC applications. Interestingly, the selection candidates had lower EBV for STAT compared to all animals, reflecting the selection strategy on small-sized DSN cows in the breeding herds, but on larger animals in commercial herds.

Inbreeding effects and optimum genetic contribution selection

From model 1 additionally considering a linear regression on the pedigree-based inbreeding coefficient, respective results depict the change in the unit of the trait per 0.01 increase in

inbreeding (Table 4). The effects of pedigree inbreeding coefficients on Pro%, FPR and STAT were quite small. For Pro%, 0.01 pedigree based inbreeding increase was associated with a reduction of 0.003% (50K) and 0.002% (DSN_200K) in milk protein content. Increased inbreeding was associated with smaller body sizes, i.e., a reduction of 0.062 cm per 0.01 pedigree inbreeding coefficient (50K), and of 0.059 cm per 0.01 pedigree inbreeding coefficient (DSN_200K). The effects on FPR were negligible with regression coefficients very close to zero. However, for FPR, as outlined before, smallest and highest values reflecting acidosis and ketosis, respectively, are unfavorable. Consequently, a linear regression on inbreeding cannot fully depict the underlying physiological mechanisms. Consequently, in pure phenotypic association analyses, we classified the cows according to FPR as follows: i) FPR < 1.0 indicating acidosis, ii) FPR > 1.6 indicating ketosis, and iii) the intermediate class with FPR between ≥ 1.0 and ≤ 1.6. The average pedigree-based inbreeding coefficient was smallest for the intermediate cow group (0.031), but higher in the acidosis (0.0350) and ketosis group (0.034).

Figure 1 displays the main output parameters for the OGC selection for Pro%, i.e., the average genetic merit of the selected parents and the number of selected BS at different constraint for genetic relationships. Genetic gain in terms of average EBVs for Pro% is increasing with a relaxing constraint in genetic relationships. The average EBV of selected animals was 0.083 (50K) and 0.087 (DSN_200K) at a strict constraint of 0.02 for the genetic relationships, but 0.347 at a relaxed genetic relationship constraint of 0.10. However, the increase in genetic merit per additional 0.01 relaxed relationship constraint was larger at strict restriction levels compared to already relaxed restriction levels. At an already relaxed constraint of 0.07 genetic relationships, further relaxations were associated with only minor positive effects on genetic merit. As expected, a strict constraint in genetic relationships was related with more diversity in terms of selected BS, i.e., an increased number of different BS. For the strict constraint at 0.02 genetic relationships, 25 (DSN_200K) and 26 (50K) BS were selected, but at a 0.10 genetic relationship constraint, only 8 influential BS heavily contributed with gene flow to the next generation. At most of the genetic relationship constraint, the slightly higher genetic merit based on the DSN_200K EBV was associated with a slightly lower number of BS compared to the selection based on the 50K EBV. The effects of the genetic relationship constraint on the number of selected BS were very similar based on the “50K EBV” or on the “DSN_200K EBV”. The similarity is reflected through the large overlap of same BS which have been selected in both scenarios. For example, at 0.10 genetic relationship constraint, from the 8 selected BS, 7 BS were identical.

Figure 2 displays genetic merit and the number of selected BS at different constraint for genetic

relationships for the OGC selection for STAT. The current selection change towards larger cows implies that higher STAT EBV are favorable. Consequently, a more relaxed constraint for genetic relationships was associated with an increase of the average EBV for STAT, due to a lower number of selected BS with favorable (higher) STAT EBVs. However, at all genetic relationship constraints, the average EBV were negative, indicating the generally low STAT EBV among the current pool of selection candidates, and the above mentioned strategy on small-sized cattle in the past two decades. Nevertheless, the strict OGC application contributed to the desired breeding objective with focus on a larger body size towards the population average. Again, the increase in genetic merit was very obvious when relaxing the constraint at a generally low level for genetic relationships. The effect on genetic gain based on EBVs from either the 50K or the DSN_200K was very similar, because the breeding value correlation between both EBVs was 0.99. Consequently, identical BS were selected in both scenarios across all constraints for genetic relationships. As expected from the curve pattern for genetic merit, the reduction in the number of selected BS was strongest when relaxing the genetic relationship constraint at already low levels for genetic relationships.

Figure 3 displays genetic merit and the number of selected BS at different constraint for genetic relationships for the OGC selection for FPR. The shape of curves for genetic merit and number of selected BS follows the pattern as described for Pro% and STAT. A more relaxed genetic relationship constraint was associated with favorable effects on the average EBV for FPR, contributing to prevent both diseases acidosis and ketosis. In contrast to STAT, the selected BS were not identical based on 50K EBV or on DSN_200K EBV, albeit the differences were quite small. With regard to the intermediate optimum trait FPR and the applied EBV transformation, there is no guarantee that for each BD with a value above the optimum there is an appropriate mating partner with a value below the optimum, and vice versa. Nevertheless, the small number of finally selected BD and BS for a genetic relationship constraint larger than 0.04 had favorable EBVs in the sense of metabolic stability. These BS and BD had original FPR EBV very close to zero, implying large transformed FPR EBVs. The large transformed FPR EBVs disprove any concerns for either ketosis or acidosis, ultimately contributing to stabilized selection for FPR.

The OGC results for the selection based on I-DSN are depicted in Figure 4, indicating the same curve pattern for the genetic merit and the number of selected BS as observed for single traits. Again, both EBV sources (DSN_200K versus 50K) only had very minor effects on the selection of elite animals.

Overall, the OGC application is in agreement with the DSN breeding objectives, i.e., to maintain a

balance between long term genetic gain and genetic diversity in terms of genetic relationships. For all three traits Pro%, STAT and FPR, as well as for I-DSN, a genetic relationship constraint larger than 0.06 was associated with only minor additional genetic merit. Consequently, the 0.06 genetic relationship constraint is suggested for OGC applications in the current DSN breeding program. The average pedigree relationship coefficient among the elite animals used for matings in practice (5 selected BS and 30 selected BD) is 0.058, which yielded average EBVs of 0.25 for Pro%, of -4.10 for STAT and of 0.42 for FPR. The respective average EBVs of selected BS and BD based on OGC (50K EBV) were 0.31 for Pro%, -2.96 for STAT and 0.49 for FPR, indicating additional genetic merit in the range from 17% to 27% for the different traits compared to the reference scenario (current DSN selection scheme as practiced by the breeding organization). With regard to the constructed index, OGC selection at a 0.06 constraint for genetic relationships resulted in a genetic merit of 139.4 (DSN_200K). The average I-DSN of BD and BS selected by the breeding organization was 118.1, indicating a genetic gain of 17.9%.

DISCUSSION

Single-step genetic parameters and breeding values

Utilization of the DSN specific SNP chip (DSN_200K) or of the commercial 50K resulted in almost identical results in single-step genetic parameter estimations for variance components and heritabilities for all three traits Pro%, STAT and FPR. In genomic evaluations for production traits in DSN, Wolf et al. (2023) modelled pure genomic relationship matrices (**G**-matrices) considering different marker densities. Similarly to our study, they found that genetic parameters based on the specific DSN_200K or on whole genome sequence data did not differ significantly from the results based on the 50K or a randomly constructed 200 SNP chip. As outlined in previous genomic evaluations with **G**-matrices, the chosen genotype platform or even utilization of whole genome sequences, had minor impact on genetic parameter estimates (e.g., van Binsbergen et al., 2015). Instead, the size and the genetic composition of the training set was the crucial parameter influencing the genetic parameter estimates (e.g., Naderi et al., 2016). In contrast to genomic evaluations only considering the **G**-matrix, pedigree relationships additionally determine the outcome in single-step approaches, and pedigree relationships are completely independent from the marker panel or the marker density. Hence, the even higher similarity of genetic parameter estimates from different SNP chips in the present single-step study compared to "**G**-approaches" could be expected. The correlation coefficients larger than 0.99 between EBVs from the DSN_200K and the 50K in same traits are a further sign in this regard. A breed specific SNP chip

might have greater application potential in breeds with larger genetic distances to HF than the DSN. The still existing genetic relationship between DSN and HF (Naderi et al., 2020) might explain the only minor differences in the present study.

In the present study, all variance components and heritabilities had quite small standard errors. Major reasons for the reliable genetic parameter estimates and EBVs are the large datasets of phenotyped cows for all three traits Pro%, FPR and STAT, and the high quality of the pedigree relationships, because several gaps in the pedigree could be filled by applying the DSN specific pedigree imputation algorithm (Jaeger et al., 2018a). The importance of the high quantity and quality of phenotyped cows in genomic predictions based on so-called cow training sets was especially highlighted for local breeds, which are characterized by a small number of AI sires with highly accurate conventional EBV (Schöpke and Swalve, 2016). The heritabilities in the DSN population for all three traits are at the upper range compared to estimates in commercial breeds with large population size, especially for STAT. An explanation might be the accurate and objective STAT measurements in cm in the current study compared to subjective classifier scores (e.g., Schierenbeck et al., 2009). Similarly to our study, large heritabilities for Pro% have been reported in other cattle populations, and a large heritability especially for DSN was expected, because of the identified major candidate genes with large effects (Korkuć et al., 2023).

The genetic-statistical modelling approach, i.e., with or without regressions on the pedigree-based inbreeding coefficient, had no effect on the genetic parameter estimates. Similarly, in HF cows from Iran, Rokouei et al. (2011) applied genetic statistical models including the pedigree relationship matrix, and additionally considering or ignoring a linear regression on the pedigree-based inbreeding coefficients. Variance components and genetic parameters for production and reproduction traits from both models were very similar, and the rank correlations between EBVs for male animals were throughout larger than 0.95. Fioretti et al. (2001) enhanced genetic models by additionally considering the inbreeding effect of the sire and the dam as linear regressions, but the maximum change in heritabilities compared to the baseline model was only 0.01.

Inbreeding effects

In the present study, the negligible effects of inbreeding on genetic parameter estimates might be due the generally small effects of inbreeding on the cow traits Pro%, STAT and FPR. In Australian HF, Pryce et al. (2014) indicated generally stronger inbreeding depressions in genomic models compared to pedigree-based analyses. However, the negligible detrimental effects of inbreeding on production traits in the present DSN study are in agreement with results from pure genomic

approaches in Canadian and Dutch HF populations (Makanjuola et al., 2020; Doekes et al., 2019, respectively). In their studies, genomic inbreeding measures such as runs of homozygosity or coefficients from the genomic relationship matrix, outlined inbreeding depressions only for a very few low heritability functional traits. Results from simulation studies indicated over-estimated inbreeding depressions based on inbreeding coefficients from the genomic relationship matrix in populations with small effective population size (Caballero et al., 2020).

Optimum genetic contribution selection

The current pedigree inbreeding coefficient for DSN bulls and cows from the most recent birth year of 0.038 is on a generally moderate level compared to other cattle populations (e.g., Hinrichs et al., 2015). However, the quite close genetic relationships of 0.079 among the current available AI sires and the obvious increase of inbreeding in the most recent birth years suggest to implement a proper inbreeding management in DSN breeding approaches. The reason for the accumulation of inbreeding and genetic relationships in DSN in the recent years might be the shift from the strong utilization of natural service sires towards AI. In this regard, Jaeger et al. (2018a) recently identified close genetic relationships among intensively used AI sires, and they suggested to control genetic relationships in a long-term perspective through the determination of optimum genetic contributions for BD and BS. Application of mating programs within cow herds, i.e., to mate a specific cow sire with a specific cow dam, only supports to minimize inbreeding in a short-term, but not in long-term perspective (König and Simianer, 2006).

Generally, the applied OGC approach in the DSN breed based on genomic breeding values from single-step genomic evaluations are in agreement with previous pure pedigree-based approaches, or with enhanced genomic applications. The general pattern of associations found in the present study, i.e., a quite strong accumulation of genetic merit with a relaxing constraint at low genetic relationship levels, but only minor effects on EBVs at already high levels, have been displayed by Kearny et al. (2004) in the UK HF population, and by König and Simianer (2006) in German HF. Both latter studies used pedigree EBV and constraints for the pedigree relationships. For local breeds, Biermann et al. (2014) emphasized the importance of natural matings, implying difficulties to mate elite animals according to their suggested genetic contributions. Consequently, they enhanced the specific mating algorithm by Sonesson and Meuwissen (2000) through a second constraint, which considered the regional availability of sires. However, in DSN, the percentage of AI is larger than 80% and almost 100% in the BD herds. Additionally, an exchange of natural service sires has been observed across producer herds (Jaeger et al., 2018a). In some local cattle breeds with small population size, crossbreeding with related breeds is

common practice, e.g., between Angler and dual-purpose Red and White cattle (Addo et al., 2017). Consequently, to account for the breed origin of alleles, Wang et al. (2017) additionally considered migration, genetic uniqueness and native allele diversity in optimum contribution selection for Vorderwald and Angler cattle. In this regard, for the handling of genetic introgression with other breeds, the software package Optisel (Wellmann, 2019) can consider native contributions from pedigree data, or segment-based kinships from genomic marker data. Nevertheless, such aspects might be important for some local cattle breeds with historical or recent migration, but not for DSN. DSN has a strong purebred history, implying that migration of other breeds was only practiced up to the end of the 19th century (Mügge et al., 1999). To enlarge the DSN gene pool, Jaeger et al. (2018b) suggested across-country genetic evaluations considering the purebred DSN from Poland and The Netherlands. For such an international purebreed approach, the OGC concept from this study can be directly applied. An increase of BS for elite matings (being a major result from the current OGC applications) is a very efficient tool to maintain long-term genetic diversity in AI programs for dairy cattle. It was the widespread use of frozen semen of influential sires in the late 1960s and early 1970s, which mainly contributed to a fast increase of genetic relationships in German HF, with long-term impact on the current population structure (König and Simianer, 2006). Accordingly, in all 3 continents Europe, North America and Oceania, Miglior (2000) identified the same 5 influential BS contributing to more than 50% to the young bull generations. The 14 (for Pro%), the 9 (for FPR), the 7 (for STAT) and the 10 (for I-DSN) selected BS in the present OGC approach at a 0.06 constraint for the genetic relationship exceeds the number of only 5 bull sires, which are currently used in practical DSN breeding schemes. In practice, the DSN breeding organization suggests the same BS to improve the overall breeding goal, which is mainly determined by Pro%, FPR and STAT. In the current OGC approach, 3 BS were in common between all three traits and for I-DSN. Brito et al. (2021) associated loss of diversity with intensive selection on only a few traits reflecting a specific trait category, or on a specific breeding index. König et al. (2013) draw similar conclusions, especially if the few breeding traits are genetically closely correlated. In contrast, the three traits Pro%, FPR and STAT reflect different trait categories with different underlying physiological and genetic mechanisms. However, from a practical perspective, it might be difficult to implement specific BS and BD selection strategies for different traits. Consequently, genetic improvements based on optimum genetic contributions for I-DSN, is the most realistic approach.

CONCLUSION

Heritabilities for the major breeding goal traits from the single-step genetic parameter estimations in the local DSN breed were quite large for Pro% (0.69) and STAT (0.78), but quite small for FPR (0.11). The choice of the SNP chip (50K versus DSN_200K) as well as of the statistical modelling approach (with or without a linear regression on pedigree-based inbreeding coefficients), implied negligible effects on all variance components. An increase of inbreeding was associated with only small unfavorable or neutral effects on Pro%, STAT and FPR. Optimum genetic contribution selection based on single-step EBVs for BS and BD led to an increase in genetic merit and a decline in the number of selected BS with a relaxed constraint for the genetic relationship. However, the respective effects were marginal when relaxing the constraint for the genetic relationship at 0.06 or higher. Consequently, we suggest OGC applications with a constraint of genetic relationships at 0.06, which contributed to a genetic gain of 17.9% in the overall index I-DSN and more diversity due to an increased number of BS when compared to the current practical DSN elite animal selection scheme.

Acknowledgement

The authors gratefully acknowledge funding from the German Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany, granted by the Federal Office for Agriculture and Food (BLE; grant numbers 2818BM090 2818BM091).

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

Conflict of interest statement

The authors declare that they have no conflicts of interest.

REFERENCES

Addo, S., J. Schäler, D. Hinrichs, and G. Thaller. 2017. Genetic diversity and ancestral history of the German Angler and the Red-and-White dual-purpose cattle breeds assessed through pedigree analysis. Agricultural Sciences 8, 1033-1047. doi:10.4236/as.2017.89075.

Aguilar, I., Misztal, I., Tsuruta, S., and A. Legarra. 2014. PREGSF90–POSTGSF90: Computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, BC, Canada, 2014; p. 3. doi:10.13140/2.1.4801.5045

Alemu, S.W., Handcock, R.C., and D.C. Garrick. 2023. Estimated genetic parameters for mature cow live weights during lactation in a multi-breed dairy cattle population. New Zealand J. of Agricultural Research, https://doi.org/10.1080/00288233.2023.2261875

Bergk, N., and H.H. Swalve. 2011. Fat-to-protein-ratio in early lactation as an indicator of herdlife for first lactation dairy cows. Züchtungskunde 83 (2), 89–103.

Biermann, A.D.M., Pimentel, E.C.G., Tietze, M., Pinent, T., and S. König. 2014. Implementation of genetic evaluation and mating designs for the endangered local pig breed 'Bunte Bentheimer'. J. Anim. Breed. Genet. 131, 36-45.

Brito, L.F., N. Bedere, F. Douhard, H.R. Oliveira, M. Arnal, F. Peñagaricano, A.P. Schinckel,

C.F. Baes, and F. Miglior. 2021. Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world. Animals 15, https://doi.org/10.1016/j.animal.2021.100292

Caballero, A., B. Villanueva, and T. Druet. 2020. On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evolutionary Applications 14(2), 416-428.

Doekes, H.P., R.F. Veerkamp, P. Bijma, G. de Jong, S.J. Hiemstra, and J.J. Windig. 2019. Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. Genetics Selection Evolution 51, 54.

Fioretti, M., Rosati, A., Pieramati, C., and L.D. Van Vleck. 2001. Effect of including inbreeding coefficients for animal and dam on estimates of genetic parameters and prediction of breeding values for reproductive and growth traits of Piedmontese cattle. Livestock Prod. Sci. 74, 137–145.

Hinrichs, D., J. Bennewitz, R. Wellmann und G. Thaller. 2015. Estimation of ancestral inbreeding effects on stillbirth, calving ease and birthweight in German Holstein dairy cattle. J. Anim. Breed. Genet. 132, 59-67.

Jaeger, M., Scheper, C., König, S., and K. Brügemann. 2018a. Inbreeding and genetic relationships of the endangered dual-purpose Black and White cattle breed based on own genetic breed percentage calculations. Züchtungskunde 90, 262-279.

Jaeger, M., Brügemann, K., Naderi, S., Brandt, H., and S. König. 2018b. Variance heterogeneity and genotype by environment interactions in native Black and White dual-purpose cattle for different herd allocation schemes. Animal 13(10), 2146-2155, https://doi.org/10.1017/s1751731119000144

Kearney, J.F., E. Wall, B. Villanueva , and M.P. Coffey. 2004. Inbreeding trends and application of optimized selection in the UK Holstein Population. J. Dairy Sci. 2004 87, 3503-3509.

Klein, S.L., Scheper, C., Brügemann, K., Swalve, H.H., and S. König. 2019. Phenotypic relationships, genetic parameters, genome-wide associations and identification of potential candidate genes for ketosis and fat-to-protein ratio in German Holstein cows. J. Dairy Sci. 102, 6276-6287, https://doi.org/10.3168/jds.2019-16237

Korkuć, P., Neumann, G.B., Hesse, D., Arends, D., Reißmann, M., Rahmtalla, S., May, K., Wolf, M.J., König, S., and G.A. Brockmann. 2023. Whole-genome sequencing data reveal new loci affecting milk production in German Black Pied cattle (DSN). Genes14(3):581. doi:10.3390/genes14030581

König, S., and H. Simianer. 2006. Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. Livest. Sci. 103, 40-53.

König, S., K. Brügemann, and E.C.G. Pimentel. 2013. Züchterische Strategien für Tier- und Klimaschutz: Was ist möglich und was brauchen wir? Züchtungskunde 85 (1):22-33

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi .org/10.3168/jds.2009 -2061.

Legarra, A., Christensen, O.F., Aguilar, I., and I. Misztal. 2014. Single Step, a general approach for genomic selection. Livest. Sci. 106, 54-65.

MacCluer, J. W., A.J. Boyce, B. Dyke, L.R. Weitkamp, D.W. Pfennig, and C.J. Parson, C. J. 1983. Inbreeding and pedigree structure in Standardbred horses. J. of Heredity 74, 394-399.

Makanjuola, B.O., C. Maltecca, F. Miglior, F.S. Schenkel, and C.F. Baes. 2020. Effect of recent and ancient inbreeding on production and fertility traits in Canadian Holsteins. BMC genomics 21, 1-15.

May, K., Weimann, C., Scheper, C., Strube, C., and S. König, S. 2019. Allele substitution and dominance effects of CD166/ALCAM gene polymorphisms for endoparasite resistance and test-day traits in a small cattle population using logistic regression analyses.

Mammalian Genome 30, 301-317, https://doi.org/10.1007/s00335-019-09818-z

May, K., Wolf, M.J., Neumann, G.B., Korkuć, P., Brockmann, G.A., and S. König. 2022.

Genetic evaluations and genomic characteristics for local cattle using genome sequences, 50K and a specific SNP chip. Proceedings of the World Congress on Genetics Applied to Livestock Production, 3.-8. Juli 2022, Rotterdam, Niederlande. https://www.wageningenacademic.com/pb-assets/wagen/WCGALP2022/54_011.pdf

Meier, S., Korkuć, P., Arends, D., and G.A. Brockmann. 2019. DNA sequence variants and protein haplotypes of casein genes in German Black Pied Cattle (DSN). Front. Genet. 10, 1129. doi: 10.3389/fgene.2019.01129

Meier, S., Arends, D., Korkuć, P., Kipp, S., Segelke, D., Filler, G., and G.A. Brockmann. 2021. Implementation of an economic lifetime Net Merit for the dual-purpose German Black Pied cattle breed. Agriculture 11(1), 41, doi: 10.3390/agriculture11010041

Meuwissen, T.H.E. 1997. Maximizing the response of selection with a predefined rate of inbreeding. J. Anim. Sci. 75, 934-940.

Meuwissen, T.H.E., and A. Sonesson. 1998. Maximizing the response of selection with a predefined rate of inbreeding: overlapping generations. J. Anim. Sci. 76, 2575–2583.

Miglior, F. 2000. Impact of inbreeding – Managing a declining Holstein pool. Proc. 10th World Holstein Friesian conference, Sydney, Australia: 108-113.

Mügge, B., W.E. Lutz, H. Südbeck, and S. Zelfel. 1999. Deutsche Holsteins: Die Geschichte einer Zucht. Verlag Eugen Ulmer. ISBN-10. 3800145383.

Naderi, S., Yin, T., and S. König. 2016. Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. J. Dairy Sci. 99, 7261-7273.

Naderi, S., Moradi, M.H., Farhadian, M., Yin, T., Jaeger, M., Scheper, C., Korkuc, P., Brockmann, G.A., König, S., and K. May. 2020. Assessing selection signatures within and between selected lines of dual-purpose black and white and German Holstein cattle. Anim. Genet. 51, 391-408. https://doi.org/10.1111/age.12925.

Neumann, G.B., Korkuć, P., Arends, D., Wolf, M.J., May, K., Reißmann, M., Elzaki, S., König, S., and G.A. Brockmann. 2021. Design and performance of a bovine 200 k SNP chip developed for endangered German Black Pied cattle (DSN). BMC Genomics 22, 905. https://doi.org/10.1186/s12864-021-08237-2.

Pryce, J.E., M. Haile-Mariam, M.E. Goddard, and B. J. Hayes. 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genetics Selection Evolution 46, 71.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., and M.A. Ferreira. 2007. PLINK: A tool for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575. https://doi.org/10.1086%2F519795.

Rokouei, M., R.V. Torshizi, M. Moradi Sharbabak, M. Sargolzai, and A. Sorencen. 2011. Effect of inbreeding on genetic parameters and breeding values for production and reproduction traits in Holstein cattle of Iran. Iranian J. of Anim. Sci., 42(1), https://ijas.ut.ac.ir/article_22711.html

Schaeffer, L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123, 218-223.

Sargolzaei, M., Iwaisaki, H., and J. J. Colleau. 2006. CFC: a tool for monitoring genetic diversity. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 27-28 ref. 13, ISBN: 85-60088-01-6

Schierenbeck, S., König, S., and H. Simianer. 2009. Genetic and environmental impact on auction prices for Holstein cows. Livestock Sci. 121, 327-334.

Schöpke K., and H.H. Swalve. 2016. Review: Opportunities and challenges for small populations of dairy cattle in the era of genomics. Animal 10,1050–1060.

Shabalina, T., Yin, T., May, K., and S. König. 2020. Proofs for genotype by environment interactions considering pedigree and genomic data from organic and conventional cow reference populations. J. Dairy Sci. 104, 4452-4466. https://doi.org/10.3168/jds.2020-19384

Sonesson, A.K., and T.H.E. Meuwissen. 2000. Mating schemes for optimum contribution selection with constrained rates of inbreeding. Genet. Sel. Evol. 32, 231-248.

Van Binsbergen, R., Calus, M.P.L., Bink, M.C.A.M., van Eeuwijk, F.A., Schrooten, C., and R.F. Veerkamp. 2015. Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. Genet. Sel. Evol. 47, 71. https://doi.org/10.1186/s12711-015-0149-x.

Wang, Y., J. Bennewitz, and R. Wellmann. 2017. Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. Genet. Sel. Evol. 49, 45. doi:10.1186/s12711-017-0320-7

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

Wolf, J.M., Neumann, G.B., Korkuć, P., Yin, T., Brockmann, G. A., König, S., and K. May. 2023. Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing. J. Dairy Sci. 106, 3345-3358. https://doi: 10.3168/jds.2022-22665

Wang, Y., J. Bennewitz, and R. Wellmann. 2017. Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. Genet. Sel. Evol. 49, 45. doi:10.1186/s12711-017-0320-7

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

Wolf, J.M., Neumann, G.B., Korkuć, P., Yin, T., Brockmann, G. A., König, S., and K. May. 2023. Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing. J. Dairy Sci. 106, 3345-3358. https://doi: 10.3168/jds.2022-22665

Table 1. Descriptive statistics for the cow traits 305-d lactation protein percentage (Pro%), fat-to-protein ratio (FPR) at the first test-day and stature (STAT).

Trait	Mean	SD	Min.	Max.
Pro% (in %)	3.48	0.24	2.45	4.31
FPR (a ratio)	1.22	0.21	0.41	3.03
STAT (in cm)	135.38	5.61	125.00	143.00

Table 2. Single-step variance components and heritabilities with respective SE for 305-d protein percentage (Pro%), first test-day fat-to-protein ratio (FPR), and stature (STAT) based on genotypes from the commercial 50K Chip (50K) and the specific DSN chip (DSN_200K) from model 1 with and without linear regression of pedigree based inbreeding

Trait	SNP chip	Model	Genetic variance	Residual variance	h^2	SE
Pro%	50K	Without inbreeding	0.0276	0.1233	0.6912	0.022
		With inbreeding	0.0276	0.0123	0.6913	0.022
	DSN_200K	Without inbreeding	0.0279	0.0121	0.6967	0.022
		With inbreeding	0.0278	0.0121	0.6968	0.022
FPR	50K	Without inbreeding	0.2608	2.1915	0.1061	0.020
		With inbreeding	0.2612	2.1914	0.1062	0.016
	DSN_200K	Without inbreeding	0.2659	2.1871	0.1081	0.020
		With inbreeding	0.2663	2.1871	0.1083	0.020
STAT	50K	Without inbreeding	9.6249	2.7515	0.7771	0.017
		With inbreeding	9.6755	2.7563	0.7777	0.017
	DSN_200K	Without inbreeding	9.6626	2.7106	0.7803	0.017
		With inbreeding	9.5528	2.7123	0.7788	0.017

Table 3: Descriptive statistics for estimated breeding values for protein percentage (Pro%), fat-to-protein ratio (FPR), transformed FPR (FPR-Trans)[1], stature (STAT) and the overall index (I-DSN)[1] for all animals (ALL; including the cows with phenotypic record as specified in Table 1 and their sires), and for the 1125 pre-selected elite animals for optimum genetic contribution selection (ELITE) from the different SNP chips 50K and DSN_200K.

Trait	Group	Chip	Mean	SD	Min.	Max.
Pro% (in %)	ALL	50K	0.029	0.125	-0.479	0.604
	ALL	DSN_200K	0.028	0.126	-0.483	0.609
	ELITE	50K	0.138	0.131	0.320	0.686
	ELITE	DSN_200K	0.140	0.130	0.339	0.689
FPR (a ratio)	ALL	50K	-0.027	0.212	-1.073	1.052
	ALL	DSN_200K	-0.029	0.213	-1.058	1.087
	ELITE	50K	-0.137	0.262	-1.073	0.878
	ELITE	DSN_200K	-0.133	0.263	-1.058	0.864
FPR-Trans	ALL	50K	0.355	0.115	0	1.052
	ALL	DSN_200K	0.360	0.120	0	1.087
	ELITE	50K	0.242	0.154	0	0.878
	ELITE	DSN_200K	0.248	0.157	0	0.864
STAT (in cm)	ALL	50K	-4.847	5.386	-18.313	11.576
	ALL	DSN_200K	-4.829	5.379	-18.309	11.642
	ELITE	50K	-9.662	2.017	-15.762	-0.325
	ELITE	DSN_200K	-9.631	2.026	-15.723	-0.355
I-DSN	ALL	50K	101.48	11.01	74	136
	ALL	DSN_200K	101.93	11.43	72	137
	ELITE	50K	105.80	9.49	96	152
	ELITE	DSN_200K	106.02	9.64	96	153

[1]Transformations of EBV and calculations of the index are explained in the materials and methods

Table 4. Regression coefficients with respective standard errors (SE) and test for significance[1] for the change in protein percentage (Pro%), fat-to-protein ratio (FPR) and stature (STAT) per 0.01 increase of the inbreeding coefficient from the genetic-statistical model using the 50K or the DSN-200K genotype data

Trait	Chip	Regression coefficient	SE	Significance
Pro% (in %)	50K	-0.003	0.001	n.s
	DSN_200K	-0.002	0.001	n.s.
FPR (a ratio)	50K	0.001	< 0.000	n.s.
	DSN_200K	0.001	< 0.000	n.s.
STAT (in cm)	50K	-0.062	0.008	n.s.
	DSN_200K	-0.059	0.008	n.s.

[1]n.s.: Regression coefficient is not significantly different from zero ($P > 0.05$)

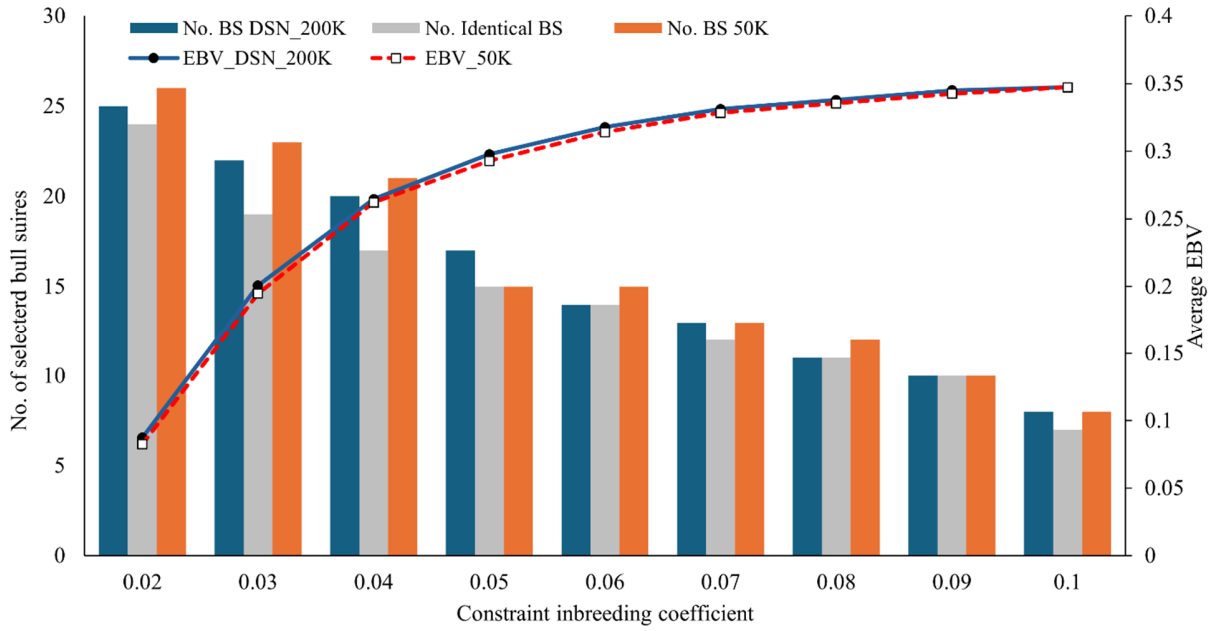

Figure 1. Average estimated breeding values (EBV) for protein percentage of selected bull sires (BS) and bull dams and the number of selected BS at constraint for average genetic relationships. (DSN_200K = the DSN specific SNP chip was used for the genetic evaluation; 50K = the standard 50K chip was used for the genetic evaluations).

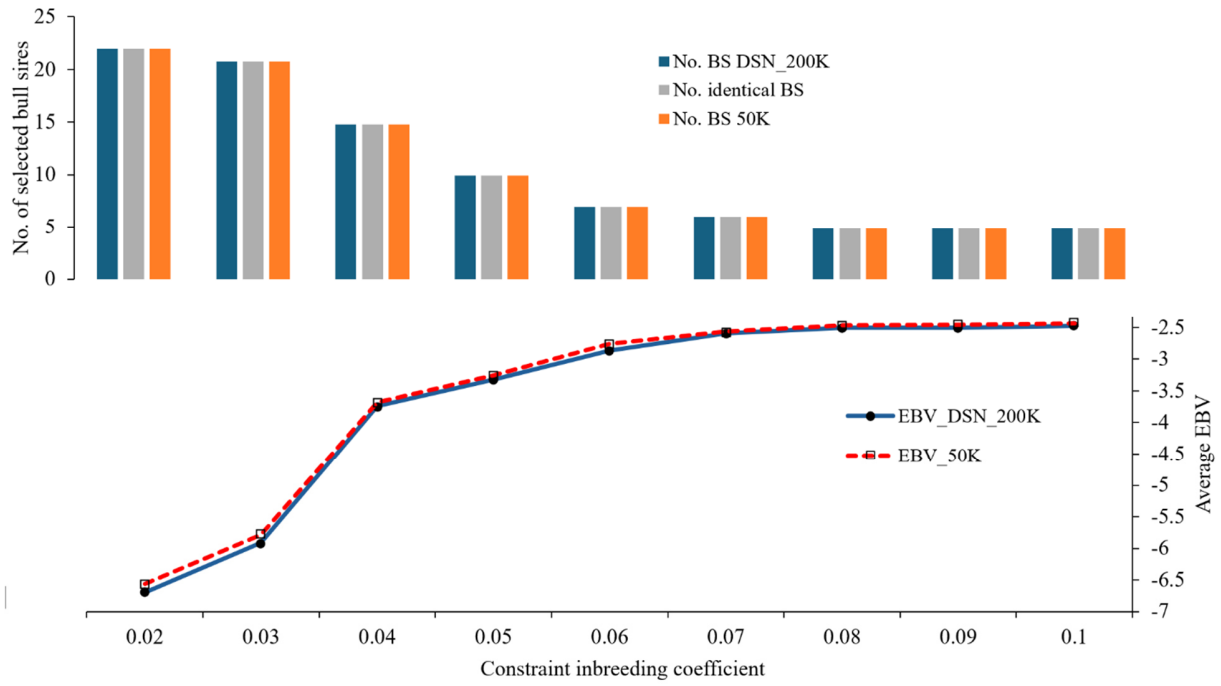

Figure 2. Average estimated breeding values (EBV) for stature of selected bull sires (BS) and bull dams and the number of selected BS at constraint for average genetic relationships. (DSN_200K = the DSN specific SNP chip was used for the genetic evaluation; 50K = the standard 50K chip was used for the genetic evaluations).

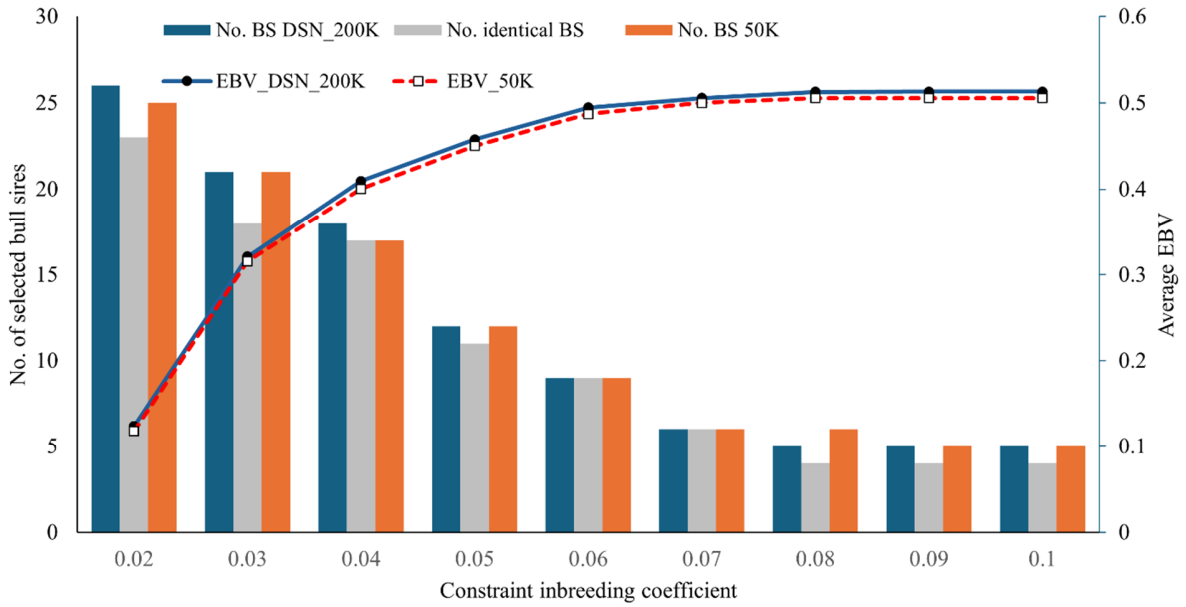

Figure 3. Average estimated breeding values (EBV) for the fat-to-protein ratio of selected bull sires (BS) and bull dams and the number of selected BS at constraint for average genetic relationships. (DSN_200K = the DSN specific SNP chip was used for the genetic evaluation; 50K = the standard 50K chip was used for the genetic evaluations).

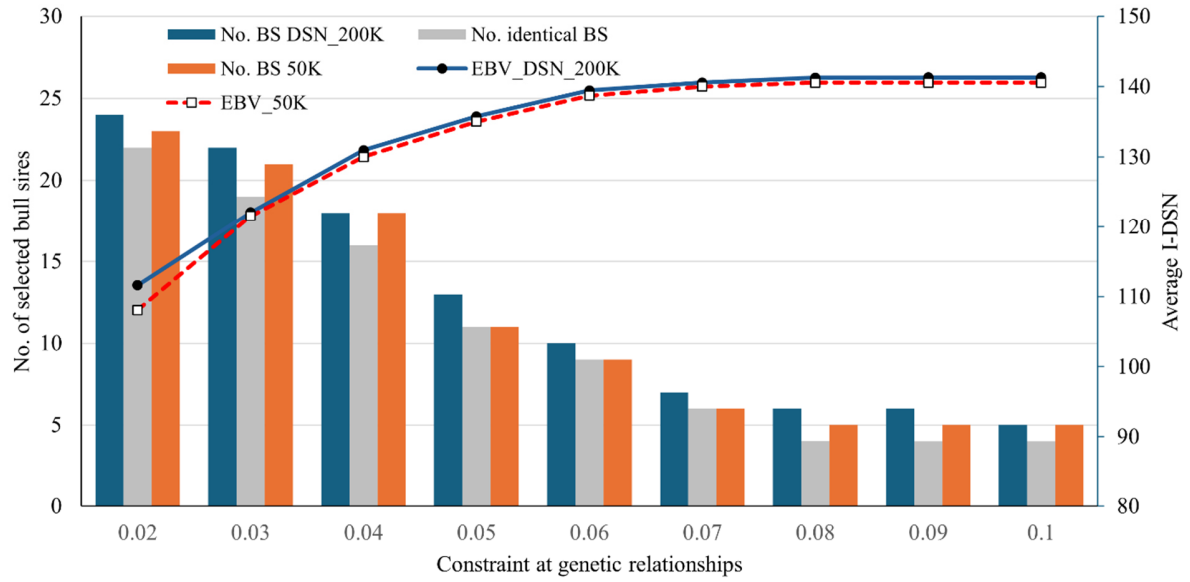

Figure 4. Average index (I-DSN) of selected bull sires (BS) and bull dams and the number of selected BS at constraint for average genetic relationships. (DSN_200K = the DSN specific SNP chip was used for the genetic evaluation; 50K = the standard 50K chip was used for the genetic evaluations).

KAPITEL 5

DISKUSSION

Die in den Kapiteln 2-4 angewandten Methoden wurden im Rahmen eines Projektes zur Entwicklung und Evaluation eines rassespezifischen SNP-Chips für die Rasse DSN veröffentlicht.

Eine Stichprobe mit 304 sequenzierten DSN wurde zu Beginn des Projektes aufgebaut. Durch Anwendung bioinformatischer Tools und genetisch-statistischer Analysen auf den WGS-Datensatz können Erkenntnisse über die genetische Architektur der bedrohte Rinderrasse DSN erlangt werden. WGS ermöglicht die Erfassung einer viel größeren Anzahl von genetischen Varianten als traditionelle SNP-Chips. Dies führt zu einer präziseren Identifizierung von Varianten, die mit spezifischen Merkmalen oder in bestimmten Populationen bzw. unter bestimmten Umweltbedingungen bedeutend sein können. Das kann besonders bei Rassen mit kleineren Populationsgrößen nützlich sein, da es die Möglichkeit bietet, auch seltene genetische Variationen zu erkennen (Daetwyler et al., 2014). Durch die erhöhte Informationsdichte wird eine verbesserte Zuchtauswahl erwartet (Iheshiulor et al., 2016). Die WGS-Datenbasis bietet, im Gegensatz zu Daten die durch SNP-Chip Genotypisierungen erhoben werden, eine viel detailliertere Überwachung von genetischen Ressourcen, insbesondere bei lokalen Rassen, die oft von Inzucht bedroht sind. Dies ist entscheidend, um die langfristige Vitalität und Anpassungsfähigkeit der Populationen zu sichern. Die Nachteile von WGS sind zum einen die hohen Kosten im Vergleich zur Verwendung von beispielsweise SNP-Chips. Zudem benötigt die anfallende Datenmenge den Umgang mit fortgeschrittenen bioinformatischen Werkzeugen und eine geeignete technische Infrastruktur. Trotz der Herausforderungen, die WGS Daten mit sich bringen, sind ihr Potenzial für die Züchtung und genetische Forschung enorm.

5.1 Imputation

Die Stichprobengröße an Sequenzdaten einer bedrohten Rasse von insgesamt 304 DSN ist einzigartig. In vielen Studien wird nur eine geringe Anzahl an Tieren sequenziert, z.B. 16 Rinder von zwei chinesischen Rinderrassen (Jiang et al., 2021), 30 Rinder von sechs schwedischen Rassen (Harish et al., 2024) oder 32 Kühe des rumänischen grauen Steppenrinds (Davidescu et al., 2022). Hierbei handelt es sich um Studien, die die genetische Diversität und Herkunft der Rassen analysierten. Für Methoden wie GWAS oder für die Schätzung von GEBVs wird eine größere Anzahl an Tieren benötigt, um eine hohe statistische Aussagekraft zu erzielen. Aufgrund der hohen Kosten der Sequenzierung wird hier meist eine Imputation der verfügbaren Genotypen (z.B. 50K) auf WGS-Level angewandt. Dafür wird eine ausreichend große Anzahl an sequenzierten Tieren benötigt, um eine Referenzpopulation zu bilden. Hier sind single- und multi-breed Verfahren möglich. Bei Rassen mit geringen Populationsgrößen ist es sinnvoll eine multi-breed Methode anzuwenden, um die Größe der Referenzpopulationen für das Imputationsverfahren zu erhöhen. Das Ziel besteht stets darin, eine möglichst hohe Imputationsgenauigkeit zu erreichen. Bei Wang et al. (2022) wurden mit

einem Referenzdatensatz von 30 Large White Schweinen für Genomsequenzen insgesamt 2655 50K genotypisierte Schweine mit einer Imputationsgenauigkeit von ca. 0,94 auf WGS-Niveau imputiert. In einer Studie bei Hunden von Hayward et al. (2019) stand ein multi-breed Datensatz mit WGS-Daten von 365 Tieren zur Verfügung. In dem Datensatz waren reinrassige Hunde, Kreuzungsrassen und Wölfe enthalten. Anhand dieser Referenzdaten wurden die Genotypen von 6112 Hunden mit einer Genauigkeit von etwa 0,88 auf WGS-Level imputiert. Die Autoren merken an, dass die Genauigkeit auf durchschnittlich 92,7% steigt, wenn man die Tiere aus dem Zieldatensatz betrachtet, deren Rasse vermehrt im Referenzdatensatz vorhanden ist. Dies stimmt mit der Beobachtung überein, dass in einem single-breed Imputationsverfahren, wie es in Kapitel 2 eingesetzt wurde, höhere Genauigkeiten erzielt werden können. Die große Anzahl an DSN mit WGS-Daten (n = 304) ermöglichte es, insgesamt 1797 genotypisierte DSN (50K) mit einer Genauigkeit von 97,04% auf WGS-Level zu imputieren. Somit standen WGS Informationen mit insgesamt 11.413.456 SVs, also SNPs und INDELS, von 1886 Tieren für weitere Analysen zur Verfügung (Wolf et al., 2021). Für die Imputation wurde eine 1-step Imputationsmethode mit einem reinen DSN Referenzpanel, anstatt eines multi-breed Ansatzes mit DSN und HF, gewählt. Obwohl in der Literatur 2-step Verfahren eine größere Genauigkeit versprechen (VanRaden et al., 2013; van Binsbergen et al., 2014), ist die Beobachtung in kleinen Populationen eine andere, da diese keine ausreichend große Anzahl an Tieren für die Zwischen-Imputationsstufe zur Verfügung stellen können (Korkuć et al., 2019).

Insgesamt ist die Imputation von Genotypdaten ein wichtiger Forschungsbereich, der erhebliche Vorteile für die Tierzucht bietet. Die Technik erlaubt unvollständige Genotypdaten zu ergänzen, indem sie Informationen aus Referenzgenomsequenzen verwenden, um fehlende genetische Variationen in den Daten vorherzusagen. Die Anwendung dieser Methode ist besonders nützlich, wenn eine vollständige Genomsequenzierung aus Kostengründen nicht möglich ist. Durch die Methode kann die Qualität und Vollständigkeit genetischer Daten verbessert werden. Durch die höhere Dichte an Informationen erhöht sich zudem auch die analytische Leistung, sodass eine präzisere genetische Bewertung vorgenommen werden kann und die Selektionsentscheidung verbessert wird (Druet et al., 2014).

Die Genauigkeit der Imputation hängt stark von der Verfügbarkeit und Qualität des Referenzdatensatzes ab. Dies führt insbesondere bei bedrohten Rassen zu dem Problem, dass nur wenige Referenzdaten verfügbar sind. Hier kann mit multi-breed Ansätzen Abhilfe geleistet werden. Dabei ist allerdings zu bedenken, dass die einbezogenen Rassen eine starke Verwandtschaft mit der zu untersuchenden Rasse aufweisen sollten. Da selbst bei nahverwandten Rassen eine sehr differierende genetische Architektur zugrunde liegen kann, ist bei diesem methodischen Ansatz von einer geringeren Genauigkeit auszugehen, als bei einem single-breed Verfahren. Zu große Unterschiede zwischen der Studienpopulation mit der

Referenzpopulation können zu einer inkorrekten Imputation führen (Marchini & Howie, 2010). Die Imputation von Genotypdaten zu WGS kann deutliche Vorteile bieten, muss jedoch auch sorgfältig unter Berücksichtigung der jeweiligen genetischen und technischen Gegebenheiten durchgeführt werden.

5.2 Genomweite Assoziationsstudien

Unter Beachtung spezifischer Herausforderung bei Rassen mit geringen Populationsgrößen, bieten GWAS bedeutende Möglichkeiten zur Verbesserung und Erhaltung bedrohter Nutztierrassen. Mit GWAS können wichtige genomische Regionen bedrohter Rassen identifiziert werden, die mit genetischen Variationen und wichtigen Merkmalen verbunden sind (Goddard &Hayes, 2009). Diese Informationen können dabei helfen einzigartige genetische Ressourcen zu detektieren und langfristig zu erhalten.

Wie aufschlussreich GWAS mit WGS in bedrohten Rassen sein können, zeigen die Studien von Korkuć et al. (2021, 2022, 2023). In den Studien wurden GWAS für Milchproduktionsmerkmale mit einer großen Übereinstimmung an genotypisierten Individuen, jedoch mit unterschiedlichen Markerdichten (50K, DSN200K und WGS), durchgeführt. Aufgrund der nahen Verwandtschaft zur Rasse HF wurde davon ausgegangen, dass ähnliche Marker bzw. mit diesen Markern annotierte Gene, identifiziert werden. Eines der wesentlichen Gene in der Milchproduktion bzw. Milchzusammensetzung bei Rindern ist *DGAT1* (Diacylglycerol O-Acyltransferase 1) auf BTA 14. Es ist verantwortlich für den letzten Schritt der Triglyceridsynthese im Fettstoffwechsel und beeinflusst somit direkt den Fettgehalt der Milch. Dafür wird insbesondere der K232A-Polymorphismus im *DGAT1*-Gen verantwortlich gemacht. Tiere mit der K-Allel Variante zeigen tendenziell einen höheren Milchfettgehalt (Thaller et al., 2003). In den oben aufgeführten Studien von Korkuć et al. konnten weder auf 50K, 200K noch auf WGS Datenbasis *DGAT1* in den GWAS für Milchproduktionsmerkmale identifiziert werden. Auch in der Studie aus Kapitel 2 (Wolf et al., 2021), in der FPR als Merkmal in einer GWAS mit WGS Daten beim DSN untersucht wurde, konnte *DGAT1* nicht identifiziert werden. Da WGS-Daten vorlagen, konnten Korkuć et al. (2023) feststellen, dass *DGAT1* in der untersuchten DSN-Population nahezu fixiert ist für die Alanin-Protein-Variante des K232A-Polymorphismus (Häufigkeit von 0,97) und daher mit keinem Merkmal assoziiert werden kann.

Die angegebenen Studien zeigen, dass auf Basis des rassespezifischer DSN200K SNP-Chips und WGS-Daten Loci identifiziert werden können, die mit dem konventionellen 50K Chip der Firma Illumina unentdeckt blieben. Darüber hinaus konnte der WGS- Datensatz direkt die Erklärung zu liefern, dass aufgrund einer Fixierung des K232A-Polymorpismus das *DGAT1* Gen trotz seines großen Einflusses auf den Milchfettgehalt und der starken Verwandtschaft mit HF, nicht identifiziert werden konnte.

Der o.g. Datensatz mit insgesamt 1886 DSN auf WGS-Level wurde in Kapitel 2 in GWA-Studien angewandt. Angesichts der hohen Imputationsgenauigkeiten von über 97% wurde eine große Power in den GWAS erwartet um Loci zu identifizieren, die mit komplexen funktionellen Merkmalen (Fruchtbarkeit, Eutergesundheit, metabolische Gesundheit, Endoparasitenresistenzen) beim DSN assoziiert sind.

Neben den additiv genetischen Effekten, die in den GWAS berücksichtigt wurden, sollten zukünftig auch Genotyp-Umweltinteraktionen (GxE) für die komplexen Merkmale beim DSN untersucht werden. Jäger et al. (2018a) schätzten basierend auf einem großen DSN-Pedigreedatensatz Korrelationen von unter 0,80 zwischen zwei unterschiedlichen Klassen an Herdengrößen und Erstkalbealtern für die Merkmale SCS und FPR. Dies weist darauf hin, dass ggf. relevante GxE Effekte vorliegen. Je nach Einfluss von GxE auf die phänotypische Varianz für adaptive und charakteristische DSN Merkmale kann diese Information bedeutend sein für die Konservierung der Rasse und ihren adaptiven Eigenschaften. Die Berücksichtigung von GxE bei der Festlegung der Zuchtstrategie kann das Wohlergehen und die Leistung der Tiere verbessern (Hayes et al., 2009). Die Genauigkeiten in den Zuchtwertschätzungen können bei der Berücksichtigung von GxE Interaktionen in Bezug zu unterschiedlichen Umweltbedingungen, verbessert werden (Maltecca et al., 2020; O'Hara et al. 2020). Neben GxE blieben in den GWAS in Kapitel 2 auch nicht-additive Dominanzeffekte unberücksichtigt. Doch auch diese können eine entscheidende Rolle in der genomischen Architektur von komplexen funktionalen Merkmalen spielen und sollten zukünftig berücksichtigt werden (Liu et al., 2019).

In der GWAS mit WGS von Wolf et al. (2020) wurde mehrfach beobachtet, dass viele signifikante SNPs aus einer vorangegangenen Studie von May et al. (2019) mit gleichem Studiendesign und statistischem Modell, nicht mit entsprechenden Signifikanzen bestätigt werden konnten. Bei dem Merkmal RES-GIN überlappte nur ein Kandidatengen (*PHLPP1*) auf BTA 24. Derartige Sachverhalte wurden auch in anderen Studien beobachtet (s.o. Korkuć et al. 2021, 2022, 2023; Wu et al. 2017) und sind möglicherweise auf die Unterschiede in den genomischen Verwandtschaftsmatrizen und Kopplungsungleichgewichten zwischen unterschiedlich dichten SNP-Paneelen bzw. WGS-Daten zurückzuführen. Darüber hinaus muss die Korrekturmethode zur Signifikanzschwelle bei WGS-Datensätzen angepasst werden. Diesbezüglich wurde bei Wolf et al. (2020) die Bonferroni-Korrektur anhand der effektiven Anzahl an SNPs durchgeführt. Durch die große Anzahl an SNPs in dem WGS-Datensatz liegt die Signifikanzschwelle niedriger als in der Studie von May et al. (2019) mit HD-Daten (600K). Der p-Wert eines SNPs muss unterhalb der bestimmten Signifikanzschwelle liegen, diese somit „unterschreiten", damit die Nullhypothese verworfen wird und der SNP als signifikant gewertet wird. Daher müssen die p-Werte der einzelnen SNPs bei den GWAS mit WGS im Vergleich zu den Studien mit z.B. HD-Daten (600K) kleiner sein, um die

Signifikanzschwellen zu unterschreiten und somit als signifikant eingestuft zu werden. In der Literatur werden verschiedene Ansätze vorgeschlagen, um die konservative Bonferroni-Korrektur etwas zu verfeinern. Sveinbjornsson et al. (2016) schlagen z.B. eine Methode der gewichteten Bonferroni-Anpassung vor, bei der die Anreicherungen von Sequenzannotationen unter den Assoziationssignalen verwendet wird und somit die Power Assoziationen zu detektieren nachweislich gesteigert wird. Ein positiver Effekt von der großen Markerdichte in dem WGS Datensatz ist, dass in der Nähe von Kandidatengenen durch Kopplungsungleichgewicht nicht nur einzelne, sondern oftmals viele Marker, niedrigere p-Werte aufweisen und ggf. die gesetzten Schwellenwerte unterschreiten. Die Interpretation von GWAS-Ergebnissen, insbesondere das Verständnis der funktionalen Bedeutung identifizierter genetischer Varianten, bleibt komplex. Es besteht unabhängig vom zur Verfügung stehenden Datensatz die Gefahr von falsch-positiven Ergebnissen, besonders bei kleinen Stichprobengrößen, was in Rassen mit geringer Populationsgröße häufig der Fall ist (Lund et al., 2011).

Besonders positiv werden bei Wolf et al. (2021) mögliche Zusammenhänge zwischen Genorten beobachtet, die sowohl mit Endoparasitenresistenzen als auch mit verbesserter Milchproduktion in Verbindung stehen. Konkret wird ein Gen für das Prolaktin-verwandte Protein IIV und IX auf BTA 23 auch mit RES-FH assoziiert. Daher sollten zukünftig genetische Korrelationen von Merkmalen der Endoparasitenresistenzen mit Merkmalen der Milchproduktion berücksichtigt werden, wenn es darum geht, DSN Tiere zu selektieren, die an Weidesysteme angepasst sind und eine verbesserte Fruchtbarkeit und metabolische Stabilität aufweisen.

GWAS mit WGS sind ressourcenintensiv in Bezug auf Genotypisierung und Datenanalyse. Für lokale Rassen, die möglicherweise nicht im Fokus großer kommerzieller Interessen stehen, kann es schwieriger sein, die notwendigen finanziellen Mittel zu beschaffen. Umso wichtiger ist, dass auch in Zukunft Fördermittel bereitgestellt werden, um vom Aussterben bedrohte Rassen gezielt genomisch testen und letztlich erhalten zu können. Eindeutige GWAS-Ergebnisse sollten weiter untersucht werden, bevor diese für Zuchtszenarien bedacht werden. Die Power der durchgeführten GWAS ist teilweise fraglich. Das liegt vor allem an den vergleichsweise geringen Tierzahlen in den Studien. Dieses Dilemma ist bei Rassen mit geringer Populationsgröße nicht zu korrigieren. Daher sollten auch hier Ansätze einer multi-breed GWAS in Betracht gezogen werden. Jedoch müssen mögliche Verzerrungen durch die erhöhte genetische Heterogenität im multi-breed Datensatz bedacht werden. Der multi-breed Ansatz kann zu populationsspezifischen Effekten und deren Fehlinterpretation führen (VanRaden et al., 2011). Insgesamt wird die Komplexität durch aufwändigere Modelle und höheren technischen Anforderungen erhöht.

Um zu testen, ob die identifizierten Gene für die neuen Merkmale der Endoparasitenresistenzen oder auch der Fruchtbarkeit tatsächlich wirksam sind sollten weitere Untersuchungen erfolgen. Zum einem sollte die biologische Funktionsweise der identifizierten Gene näher betrachtet werden. Weitere Studien in anderen Populationen und unter verschiedenen Umweltbedingungen können zeigen, ob die Assoziationen konsistent sind. Es ist auch denkbar einige Marker in praktischen Zuchtszenarien zu integrieren, um nach mehreren Zuchtdurchgängen und nach erfolgreicher Anreicherung der als vorteilhaft eingestuften Allele, eine langfristige Wirksamkeit auf das gewünschte Merkmal zu testen.

5.3 Zuchtwertschätzung

Aus der Studie Wolf et al. (2022) in Kapitel 3 geht hervor, dass die Anwendung von rassespezifischen SNP-Chips und die Anwendung von WGS Daten bei der Schätzung von genetischen Parametern und in der Zuchtwertschätzung in kleinen Populationen nur marginale Vorteile mit sich bringen. Routinemäßig werden Zuchtwertschätzungen in Deutschland zentral von der Vereinigte Informationssysteme Tierhaltung w.V. (vit) durchgeführt. Dabei bilden in den meisten Schätzungen Holstein Schwarzbunt und Rotbunt, Rotbunt-Doppelnutzung, Rotvieh-Angler, Jersey und DSN eine gemeinsame Datengrundlage. Für die ZWS der Milchleistungsmerkmale stehen somit insgesamt Daten von ca. 14 Millionen Tiere zur Verfügung (vit, 2024). In der o.g. Studie sollte zunächst davon ausgegangen werden, dass eine rassespezifische Zuchtwertschätzung genauere Zuchtwerte impliziert, da die spezifische genetische Struktur der Rassen genauer berücksichtigt werden kann. Dies wird zusätzlich durch die Anwendung rassespezifischer Varianten (DSN200K) bzw. WGS-Daten untermauert. Die Zuchtziele können auf die speziellen Bedürfnisse oder Vorzüge der Rasse angepasst werden. Dazu zählen Anpassungsfähigkeit an lokale Umweltbedingungen, Krankheitsmerkmale oder auch der Doppelnutzungscharakter für Milch und Fleisch. Der größte Nachteil bei kleinen Populationen ist jedoch der geringe Umfang der zu Verfügung stehenden Datensätze, was die statistische Aussagekraft der Analysen, sowie die Zuverlässigkeit der ZWS, einschränkt (Hayes et al., 2009). Für zukünftige Anwendungen könnte es jedoch interessant sein, kausale genetische Mechanismen, die durch WGS oder rassespezifische Chips, wie dem DSN200K aufgedeckt wurden und in einem direkten Zusammenhang mit rassespezifischen Vorzügen, wie z.B. Krankheitsresistenzen stehen, in Zuchtprogrammen dieser Rassen gezielt zu berücksichtigten.

5.4 Genetische Parameter und Genauigkeit der Zuchtwertschätzung

Für die Schätzung genetischer Parameter, wie Heritabilität, nutzten Wolf et al. (2023) unter Einbeziehung der rein genomischen Verwandtschaftsmatrix **G** verschiedene Markerdichten (50K, DSN200K und WGS) und zeigten, dass diese nur geringfügige Auswirkungen auf die Schätzwerte haben. Diese Beobachtungen werden durch Binsbergen et al. (2015) bestätigt. Bei zweistufigen Verfahren sind die Größe und die genetische Zusammensetzung des

Trainingssatzes die entschiedenen Einflüsse auf die Schätzungen der genetischen Parameter (Naderi et al., 2016). In Single-Step Ansätzen sollten daher möglichst viele genotypisierte Tiere mit bekannten Phänotypen in den Datensatz einbezogen werden (König, 2023). Werden im Rahmen eines Single-Step-Verfahrens neben Genotypen auch Stammbauminformationen miteinbezogen, haben diese zusätzliche Einflüsse auf die genetischen Parameter. Somit wurde in dem Manuskript in Kapitel 4 davon ausgegangen, dass die Übereinstimmungen der Zuchtwerte, basierend auf 50K, 200K und WGS höhere Ähnlichkeiten aufweisen, als bei den Schätzungen in denen nur rein genomische Schätzungen berücksichtigt wurden. Die in Kapitel 4 berechneten Korrelationskoeffizienten von größer 0,99 zwischen den Zuchtwerten, basierend auf 50K- und 200K-SNP-Chip-Informationen, bestätigen diese These. Aus den Zuchtwertschätzungen in Wolf et al. (2023) und Wolf et al. (2024) geht hervor, dass die Anwendungen von unterschiedlichen SNP-Chips mit verschiedenen Markerdichten nur geringe Auswirkungen auf die Zuchtwerte und die Genauigkeiten der Zuchtwerte haben. In Wolf et al. (2024) werden neben der genomischen Informationen zusätzlich Stammbauminformationen miteinbezogen, wodurch der Einfluss der gewählten SNPs Chips noch geringer wird. Der DSN200K Chip schneidet in den Schätzungen geringfügig besser ab als der konventionelle 50K Chip. Neben der höheren Markerdichte, die nur einen geringen Einfluss hat, ist eventuell der rassespezifische Ansatz bei dem DSN200K ebenfalls von Vorteil. Man könnte annehmen, dass durch die enge Verwandtschaft zwischen DSN und HF (Naderi et al., 2020) ein multi-breed Ansatz und die dadurch höheren Anzahl an verfügbaren Genotyp- und Phänotypinformationen eine bessere Vorhersagegenauigkeit erzielt werden könnte. Dies wurde in den Studien für DSN nicht weiter untersucht. Jedoch zeigt eine Studie von Erbe et al. (2012), dass die Nutzung des kommerziellen 50K SNP-Chips in Kombination mit einer HF-Referenzpopulation zu einem marginalen Verlust an Vorhersagegenauigkeiten für Jersey-Rinder im Vergleich zur Implementierung rassespezifischer Jersey-Referenzpopulationen oder der Nutzung dichterer SNP-Chips führt.

In der Studie in Kapitel 4 waren die geschätzten genetischen Parameter mit sehr geringen Standardfehlern assoziiert, was auf den großen Datensatz phänotypisierter Kühe und die hohe Qualität der Stammbaumbeziehungen zurückzuführen ist. Die Lücken im Stammbaum wurden durch einen speziellen Algorithmus für DSN geschlossen (Jäger et al., 2018b). Die geschätzten Heritabilitäten sind im Vergleich zu Schätzungen in kommerziellen Rassen hoch. Hierfür werden etwa objektive Messungen, anstatt subjektive Klassifizierungen (z.B. Schierenbeck et al., 2009), beispielsweise für das Merkmal Größe verantwortlich gemacht. Eine hohe Erblichkeit für den Proteingehalt der Milch war zu erwarten, da Korkuć et al. (2021) beim DSN Hauptkandidatengene mit großen Effekten auf das Merkmal identifizierten. Die in Kapitel 4 berechneten Inzuchteffekte wurden teilweise in den angewandten genetisch-statistischen Modellen als Regressionen berücksichtigten, zeigten jedoch vernachlässigbare

Effekte auf die genetischen Parameter. Inzuchtdepressionen zeigten sich in anderen Studien insbesondere bei funktionalen Merkmalen mit niedrigen Erblichkeiten (Makanjuola et al., 2020; Doekes et al., 2019). Da die gewählten Kuhmerkmale (Pro%, STAT, FPR) relativ hohe Heritabilitäten aufweisen, ist nicht von einer vorhandenen Inzuchtdepression auszugehen.

5.5 Optimum Genetic Contribution

Es ist notwendig, die Entwicklung von genetischen Beziehungen und Zuchtfortschritt langfristig durch die Bestimmung der OGC von Bullenvätern und -müttern zu regulieren. Besonders herauszustellen ist hierbei eine langfristige Betrachtung der Inzucht in der Zuchtplanung (König und Simianer, 2006). Insgesamt zeigt der aktuelle, stammbaumbasierte Inzuchtkoeffizient von DSN-Bullen und Kühen aus dem letzten Geburtsjahr ein moderates Niveau von 3,78%. Andere Rinderpopulationen weisen höhere Koeffizienten auf (Hinrichs et al., 2015). Die aktuell verfügbaren DSN Bullen für die künstliche Besamung weisen eine enge genetische Beziehung untereinander von durchschnittlich 4,9% auf. Durch den Wandel von dem Einsatz von Bullen mit Natursprung hin zur künstlichen Befruchtung ist mit einem vermehrten Anstieg von Inzucht zu rechnen. Daher ist es wichtig, ein Inzuchtmanagement in den Zuchtstrategien beim DSN zu implementieren.

Die Ergebnisse der OGC-Studie mit genomischen Daten in Kapitel 4 stimmen mit den Beobachtungen rein Stammbaum basierter Studien überein (Kearny et al., 2004; König und Simianer, 2006). Bei niedriger Inzucht ist ein starker Zuchtfortschritt bei entspannten Beschränkungen bei der Auswahl von Zuchttieren möglich. Ist bereits eine hohe Stufe an Inzucht erreicht, so stagniert der Zuchtfortschritt, einhergehend mit einer geringen Anzahl an selektieren Bullenvätern. In Rassen mit geringer Populationsgröße hat der Einsatz von Deckbullen noch eine große Relevanz, was zu Schwierigkeiten bei der gezielten und geplanten r Verpaarung von Elitetieren führen kann. Daher muss bei diesen Rassen zusätzlich die regionale Verfügbarkeit berücksichtigt werden. In der DSN Population ist das Problem nicht so stark ausgeprägt. Hier wird über 80% künstliche Besamung eingesetzt und Natursprungbullen werden zwischen Herden teilweise getauscht (Jäger et al., 2018a). Ein weiterer Aspekt bei Rassen mit kleiner Populationsgröße ist die Kreuzung mit verwandten Rassen, wie z.B. zwischen Angler und Rotbunt Doppelnutzung (Addo et al., 2017). Dabei müssen neben Zuchtwert, Inzucht und regionaler Verfügbarkeit auch Migration, genetische Einzigartigkeit und native Allelvielfalt bei der Auswahl berücksichtigt werden (Wang et al., 2017). Eine Kreuzung mit verwandten Rassen ist beim DSN nicht notwendig, da die Rasse aufgrund ihrer großen Bedeutung in der Vergangenheit eine lange reinrassige Zuchtgeschichte aufweisen kann (Mügge et al., 1999). Für die Erweiterung des Genpools sind länderübergreifende Aktivitäten wie beispielsweise die Berücksichtigung von DSN Populationen aus Polen und den Niederlanden, angeraten. Dafür könnte die in Kapitel 4 angewandte OGC-Methode direkt im internationalen Kontext übertragen werden. Die

Ergebnisse von Kapitel 4 zeigen, dass die Steigerung der Anzahl an Bullen für Elitepaarungen ein effizientes Vorgehen ist, um langfristig die genetische Diversität in Zuchtprogrammen mit künstlicher Besamung aufrechtzuerhalten. Des Weiteren sollten in den Programmen unterschiedliche Merkmale und Zuchtindizes betrachtet werden, um möglichst unterschiedliche genetische und physiologische Mechanismen abzudecken. Sonesson und Meuwissen (2000) entwickelten einen Algorithmus für die Zuweisung spezifischer Paarungspartner, der darauf abzielt, die Inzucht auf Basis von OGC in der nächsten Generation zu minimieren. Die Anwendung von OGC zur gezielten Anpaarung eines spezifischen Bullenvaters mit einer spezifische Bullenmutter hilft nur kurzfristig, Inzucht zu minimieren, dafür folgende Generationen durch diese Maßnahme weniger Spielraum zur Maximierung von genetischer Diversität vorhanden ist. Um eine langfristige Kontrolle der Inzucht zu erlangen ist es daher notwendig, OGC in Anpaarungsprogrammen über mehrere Genrationen zu betrachten (König und Simianer, 2006). Abgesehen von der Frage, wie die Inzucht kontrolliert werden kann, bieten genomische Herdenpaarungswerkzeuge weitere Vorteile, insbesondere um letale rezessive Allele in einem homozygoten Zustand in Embryonen oder Nachkommen zu vermeiden (Uppermann et al., 2019).

Die OGC-Methode ist ein wichtiger Ansatz, um langfristig genetische Vielfalt zu erhalten, insbesondere in Populationen mit kleiner effektiver Populationsgröße wie dem DSN. Dieses Konzept wurde unter anderem von König und Simianer (2006) ausführlich beschrieben und findet Anwendung in der Praxis, um die Inzucht zu kontrollieren und eine ausgewogene genetische Vielfalt zu fördern. In lokalen Rassen mit geringer Populationsgröße ist es wichtig, Inzuchtstrategien sorgfältig zu planen und genetische Beiträge optimal zu bestimmen, um genetische Erosion zu vermeiden.

5.6 Notwendigkeit der Entwicklung eines spezifischen Genotypisierungschips

Die Verwendung von spezifischen SNP-Chips für bestimmte Rassen könnte in Populationen mit größeren genetischen Distanzen zu dominierenden Rassen von größerem Nutzen sein. Bei der Rasse DSN ist immer noch eine Verwandtschaft zur großen HF Population gegeben. In Fällen starker Eigenständigkeit einer Rasse, könnte die Verwendung von rassespezifischen SNP-Chips helfen, Kandidatengene und funktionelle Mutationen zu identifizieren, die für die Erhaltung und Verbesserung der Rasse wichtig sind. Dabei spielt nicht nur die Abdeckung der genetischen Variabilität eine Rolle, sondern auch die genetische Architektur und die spezifische Selektionsgeschichte der Rassen. Obwohl die Ergebnisse vielversprechend sind, gibt es Einschränkungen hinsichtlich der Stichprobengröße und der Repräsentativität der genotypisierten Tiere, die die Generalisierbarkeit der Ergebnisse beeinträchtigen könnten. Weiterführende Studien mit umfangreicheren Datensätzen sind notwendig, um die Robustheit und Effektivität des DSN200K-Chips weiter zu validieren. Es wird empfohlen, weitere Forschungen zur Interaktion zwischen genetischen Merkmalen und Umweltbedingungen

durchzuführen, um die Anpassungsfähigkeit des DSN weiter zu verbessern. Zudem sollten Langzeitstudien zur Überwachung der genetischen Diversität und der Inzuchteffekte initiiert werden, um die nachhaltige Entwicklung der Population zu gewährleisten. Die Ergebnisse der Single-Step-Zuchtwertschätzung in Kombination mit der OGC-Analyse zeigten, dass eine sorgfältige Balance zwischen genetischem Fortschritt und Diversitätserhalt möglich ist. Diese Balance ist essenziell, um langfristige genetische Erosion zu vermeiden und die Anpassungsfähigkeit der Population zu erhalten.

5.7 Schlussfolgerungen und Empfehlungen

Für die Entschlüsselung der genetischen Struktur von quantitativen Merkmalen bieten insbesondere WGS-Daten Vorteile, da die Abdeckung des gesamten Genoms mit Markern gegeben ist und somit auch seltene Allele in lokalen Rassen identifiziert werden können, die mit kommerziellen Arrays möglicherweise nicht erfasst werden können. Dies kann für die Erhaltung der genetischen Vielfalt und die Anpassungsfähigkeit von Rassen von Bedeutung sein. Die Zusammenstellung eines aussagekräftigen Datensatzes ist jedoch mit hohen Kosten verbunden. Daher sollte nach Aufbau eines Referenzdatensatzes mithilfe von standardisierten Genotypisierungsarrays und durch Anwendung von Imputationsmethoden eine aussagekräftige Testpopulation auf WGS-Niveau erstellt werden. Dieses Vorgehen ist im Vergleich zur Anwendung von kommerziellen SNP-Chips mit einem hohen Rechenaufwand verbunden. Trotz höherer Kosten kann WGS wesentliche Einblicke in die genetische Struktur einer Rasse liefern und sollte regelmäßig eingesetzt werden, um genetische Veränderungen über die Zeit zu überwachen. Die identifizierten Marker innerhalb relevanter Gene sollten für künftige genomische Selektionsstrategien bei verwendet werden. Beim DSN sollten die identifizierten Marker innerhalb immunologisch relevanter Gene genutzt werden, um die Adaption an die Weidehaltung weiter zu verbessern. Rassespezifische Genotypisierungsarrays sind verhältnismäßig günstiger als WGS und aufgrund der geringeren Datendichte, leichter auszuwerten. Vorausgesetzt für die zu untersuchende Rasse ist bereits ein rassespezifischer SNP-Chip verfügbar, der die Bereiche der relevanten Merkmale der Rasse gut abdeckt, dann sind diese für Assoziationsstudien zu empfehlen. Für Rassen mit geringer Populationsdichte können WGS daten und rassespezifische Genotypisierungsarrays für Untersuchungen kausaler Varianten wertvoll sein, um die genetischen Besonderheiten und Bedürfnisse dieser Rassen abzudecken.

Für weitere Methoden, wie Zuchtwertschätzungen und den Berechnungen von genetischen Parametern, waren die Vorteile von rassespezifischen SNP-Chips und WGS-Daten, nur marginal gegenüber den Anwendungen mit dem kommerziellen Illumina BovineSNP50 Chip. Die verbesserten Vorhersagegenauigkeiten und höhere Heritabilitäten auf der Grundlage von WGS- oder DSN200K-Daten reichen nicht aus, um die zusätzlichen Kosten für die

Sequenzierung der Tiere oder die Verwendung des rassespezifischen SNP-Chips beim DSN zu rechtfertigen.

Auch bei der Anwendung von OGC, zeigte der rassespezifische DSN200K Chip kaum Vorteile gegenüber der Datenbasis aus den Genotypisierungen mit dem BovineSNP50. Die OGC Ergebnisse zeigen, dass für eine langfristige Zuchtplanung in der bedrohten Rasse DSN die Inzuchtbeschränkungen auf 6% festgelegt werden sollten, um ein ausgewogenes Verhältnis zwischen Inzucht und Leistungssteigerung (17-27% in den untersuchten Merkmalen) zu erhalten. Zudem könne durch den Einsatz einer höheren Anzahl von Bullenvätern im Vergleich zur derzeitigen Selektion von wenigen Elitetieren eine größere genetische Diversität erzielt werden.

5.8 Literaturübersicht - Diskussion

Addo, S., Schäler, J., Hinrichs, D., & Thaller, G. (2017). Genetic diversity and ancestral history of the German Angler and the Red-and-White dual-purpose cattle breeds assessed through pedigree analysis. *Agricultural Sciences*, 8(9), 1033-1047. https://doi.org/10.4236/as.2017.89075

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., ... & Hayes, B. J. (2014). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, 46(8), 858-865. https://doi.org/10.1038/ng.3034

Davidescu, M.-A., Simeanu, D., Gorgan, D.-L., Ciorpac, M., & Creanga, S. (2022). Analysis of Phylogeny and Genetic Diversity of Endangered Romanian Grey Steppe Cattle Breed, a Reservoir of Valuable Genes to Preserve Biodiversity. *Agriculture*, 12, 2059. https://doi.org/10.3390/agriculture12122059

Doekes, H. P., Veerkamp, R. F., Bijma, P., de Jong, G., Hiemstra, S. J., & Windig, J. J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. *Genetics Selection Evolution*, 51, 1-16. https://doi.org/10.1186/s12711-019-0465-2

Druet, T., Macleod, I. M., & Hayes, B. J. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, 112(1), 39-47. https://doi.org/10.1038/hdy.2013.13

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., ... & Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of dairy science*, 95(7), 4114-4129. https://doi.org/10.3168/jds.2011-5019

Flury, C., Tapio, M., Sonstegard, T., Drögemüller, C., Leeb, T., Simianer, H., ... & Rieder, S. (2010). Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. *Journal of Animal Breeding and Genetics*, 127(5), 339-347. https://doi.org/10.1111/j.1439-0388.2010.00861.x

Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics, 10(6), 381-391.

Harish, A., Lopes Pinto, F. A., Eriksson, S., et al. (2024). Genetic diversity and recent ancestry based on whole-genome sequencing of endangered Swedish cattle breeds. *BMC Genomics*, 25, 89. https://doi.org/10.1186/s12864-024-05666-8

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433-443. https://doi.org/10.3168/jds.2008-1646

Hayward, J. J., White, M. E., Boyle, M., Shannon, L. M., Casal, M. L., Castelhano, M. G., et al. (2019). Imputation of canine genotype array data using 365 whole-genome sequences improves power of genome-wide association studies. *PLoS Genet*, 15(9), e1008003. https://doi.org/10.1371/journal.pgen.1008003

Hinrichs, D., Bennewitz, J., Wellmann, R., & Thaller, G. (2015). Estimation of ancestral inbreeding effects on stillbirth, calving ease and birthweight in German Holstein dairy cattle. *Journal of Animal Breeding and Genetics*, 132(1), 59-67. https://doi.org/10.1111/jbg.12110

Iheshiulor, O. O., Woolliams, J. A., Yu, X., Wellmann, R., & Meuwissen, T. H. (2016). Within- and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genetics Selection Evolution*, 48, 1-15. https://doi.org/10.1186/s12711-016-0210-2

Jaeger, M., Brügemann, K., Naderi, S., Brandt, H., & König, S. (2018a). Variance heterogeneity and genotype by environment interactions in native Black and White dual-purpose cattle for different herd allocation schemes. *Animal*, 13(10), 2146-2155. https://doi.org/10.1017/S1751731119000144

Jaeger, M., Scheper, C., König, S., & Brügemann, K. (2018b). Inbreeding and genetic relationships of the endangered dual-purpose Black and White cattle breed based on own genetic breed percentage calculations. *Züchtungskunde*, 90, 262-279.

Jiang, L., Kon, T., Chen, C., et al. (2021). Whole-genome sequencing of endangered Zhoushan cattle suggests its origin and the association of MC1R with black coat colour. *Scientific Reports*, 11, 17359. https://doi.org/10.1038/s41598-021-96880-8

Kearney, J. F., Wall, E., Villanueva, B., & Coffey, M. P. (2004). Inbreeding trends and application of optimized selection in the UK Holstein population. *Journal of Dairy Science*, 87(10), 3503-3509. https://doi.org/10.3168/jds.S0022-0302(04)73485-3

König, S. (2023). Single-step-Zuchtwertschätzung: Allgemeine Anmerkungen, Hintergrundinformation und Notwendigkeit. Züchtungskunde, 95(2).

Koenig, S., & Simianer, H. (2006). Approaches to the management of inbreeding and relationship in the German Holstein dairy cattle population. *Livestock Science*, 103(1-2), 40-53. https://doi.org/10.1016/j.livsci.2006.01.006

Korkuć, P., Arends, D., & Brockmann, G. A. (2019). Finding the optimal imputation strategy for small cattle populations. *Frontiers in genetics*, 10, 430415. https://doi.org/10.3389/fgene.2019.00430

Korkuć, P., Arends, D., May, K., & König, S. (2021). Genomic loci affecting milk production in German Black Pied cattle (DSN). *Frontiers in Genetics*, 12, 640039. https://doi.org/10.3389/fgene.2021.640039

Korkuć, P., Neumann, G. B., Arends, D., Wolf, M. J., May, K., & König, S., & Brockmann, G. A. (2022). Verbesserte genomweite Assoziationen bei Verwendung eines rassespezifischen 200K SNP-Chips für Deutsche Schwarzbunte Niederungsrinder (DSN). Vortrag: Vortragstagung der DGfZ und GfT, Kiel, Deutschland, 21.-22.09.2022.

Korkuć, P., Neumann, G. B., Hesse, D., Arends, D., Reißmann, M., Rahmatalla, S., ... & Brockmann, G. A. (2023). Whole-Genome Sequencing Data Reveal New Loci Affecting Milk Production in German Black Pied Cattle (DSN). *Genes*, 14(3), 581. https://doi.org/10.3390/genes14030581

Liu, Y., Xu, L., Wang, Z., Xu, L., Chen, Y., Zhang, L., ... & Li, J. (2019). Genomic prediction and association analysis with models including dominance effects for important traits in Chinese Simmental beef cattle. *Animals*, 9(12), 1055. https://doi.org/10.3390/ani9121055

Lund, M. S., de Roos, A. P., de Vries, A. G., Druet, T., Ducrocq, V., Fritz, S., ... & Su, G. (2011). A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genetics Selection Evolution*, 43, 1-8. https://doi.org/10.1186/1297-9686-43-43

Makanjuola, B. O., Miglior, F., Abdalla, E. A., Maltecca, C., Schenkel, F. S., & Baes, C. F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of Dairy Science*, 103(6), 5183-5199. https://doi.org/10.3168/jds.2019-17926

Maltecca, C., Tiezzi, F., Cole, J. B., & Baes, C. (2020). Symposium review: Exploiting homozygosity in the era of genomics—Selection, inbreeding, and mating programs. *Journal of Dairy Science*, 103(6), 5302-5313. https://doi.org/10.3168/jds.2019-17925

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, 11(7), 499-511. https://doi.org/10.1038/nrg2796

May, K., Scheper, C., Brügemann, K., Yin, T., Strube, C., Korkuć, P., ... & König, S. (2019). Genome-wide associations and functional gene analyses for endoparasite resistance in an endangered population of native German Black Pied cattle. *BMC Genomics*, 20, 1-15. https://doi.org/10.1186/s12864-019-6233-3

Mügge, B., Lutz, W. E., Südbeck, H., & Zelfel, S. (1999). Deutsche Holsteins: Die Geschichte einer Zucht. Verlag Eugen Ulmer. ISBN-10. 3800145383.

Naderi, S., Yin, T., & König, S. (2016). Random forest estimation of genomic breeding values for disease susceptibility over different disease incidences and genomic architectures in simulated cow calibration groups. *Journal of Dairy Science*, 99(9), 7261-7273. https://doi.org/10.3168/jds.2016-11109

Naderi, S., Moradi, M. H., Farhadian, M., Yin, T., Jaeger, M., Scheper, C., ... & May, K. (2020). Assessing selection signatures within and between selected lines of dual-purpose black and white and German Holstein cattle. *Animal Genetics*, 51(3), 391-408. https://doi.org/10.1111/age.12931

O'Hara, E., Neves, A. L. A., Song, Y., & Guan, L. L. (2020). The Role of the Gut Microbiome in Cattle Production and Health: Driver or Passenger?. *Annual Review of Animal Biosciences*, 8, 199–220. https://doi.org/10.1146/annurev-animal-021419-083952

Schierenbeck, S., König, S., & Simianer, H. (2009). Genetic and environmental impact on auction prices for Holstein cows. *Livestock Science*, 121(2-3), 327-334. https://doi.org/10.1016/j.livsci.2008.07.025

Sonesson, A. K., & Meuwissen, T. H. (2000). Mating schemes for optimum contribution selection with constrained rates of inbreeding. *Genetics Selection Evolution*, 32(3), 231-248. https://doi.org/10.1186/1297-9686-32-231

Sveinbjornsson, G., Albrechtsen, A., Zink, F., Gudjonsson, S. A., Oddson, A., Másson, G., Holm, H., Kong, A., Thorsteinsdottir, U., Sulem, P., Gudbjartsson, D. F., & Stefansson, K. (2016). Weighting sequence variants based on their annotation increases power of whole-genome association studies. Nature genetics, 48(3), 314–317. https://doi.org/10.1038/ng.3507

Thaller, G., Krämer, W., Winter, A., Kaupe, B., Erhardt, G., & Fries, R. (2003). Effects of DGAT1 variants on milk production traits in German cattle breeds. *Journal of Animal Science*, 81(8), 1911–1918. https://doi.org/10.2527/2003.8181911x

Upperman, L. R., Kinghorn, B. P., MacNeil, M. D., & Van Eenennaam, A. L. (2016). Management of lethal recessive alleles in beef cattle through the use of mate selection software. *Genetics Selection Evolution*, 51(1), 36. https://doi.org/10.1186/s12711-016-0229-4

van Binsbergen, R., Bink, M. C., Calus, M. P., van Eeuwijk, F. A., Hayes, B. J., Hulsegge, I., & Veerkamp, R. F. (2014). Accuracy of imputation to whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution*, 46, 1-13. https://doi.org/10.1186/1297-9686-46-41

van Binsbergen, R., Calus, M. P., Bink, M. C., van Eeuwijk, F. A., Schrooten, C., & Veerkamp, R. F. (2015). Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genetics Selection Evolution*, 47, 1-13. https://doi.org/10.1186/s12711-015-0103-7

VanRaden, P. M., Olson, K. M., Null, D. J., & Hutchison, J. L. (2011). Harmful inbreeding effects of recessive mutations discovered by genomic testing. Journal of Dairy Science, 94(12), 6153-6161.

VanRaden, P. M., Null, D. J., Sargolzaei, M., Wiggans, G. R., Tooker, M. E., Cole, J. B., ... & Doak, G. A. (2013). Genomic imputation and evaluation using high-density Holstein genotypes. *Journal of Dairy Science*, 96(1), 668-678. https://doi.org/10.3168/jds.2012-5702

Vereinigte Informationssysteme Tierhaltung (VIT). (2024). Beschreibung der Zucht eschreibung der Zuchtwertschätzung für alle Schätzmerkmale bei den Milchrinderrassen für die vit mit der Zuchtwertschätzung beauftragt ist. Zuchtwertschätzung. VIT. https://www.vit.de/fileadmin/DE/Zuchtwertschaetzung/2404_Zws_Bes_deu.pdf. Stand April 2024

Wang, Y., Bennewitz, J., & Wellmann, R. (2017). Novel optimum contribution selection methods accounting for conflicting objectives in breeding programs for livestock breeds with historical migration. *Genetics Selection Evolution*, 49, 45. https://doi.org/10.1186/s12711-017-0320-5

Wang, X., Wang, L., Shi, L., Zhang, P., Li, Y., Li, M., Tian, J., Wang, L., & Zhao, F. (2022). GWAS of Reproductive Traits in Large White Pigs on Chip and Imputed Whole-Genome Sequencing Data. *International Journal of Molecular Sciences*, 23, 13338. https://doi.org/10.3390/ijms232113338

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P., Brockmann, G. A., König, S., & May, K. (2021). Genome-wide association study using whole-genome sequence data for fertility, health indicator, and endoparasite infection traits in German black pied cattle. *Genes*, 12(8), 1163. https://doi.org/10.3390/genes12081163

Wolf, M. J., Neumann, G. B., Korkuć, P., Yin, T., Brockmann, G. A., König, S., & May, K. (2023). Genetic evaluations for endangered dual-purpose German Black Pied cattle using 50K SNPs, a breed-specific 200K chip, and whole-genome sequencing. *Journal of Dairy Science*, 106(5), 3345-3358. https://doi.org/10.3168/jds.2022-22721

Wolf, M. J., Yin, T., Neumann, G. B., Korkuć, P, Brockmann, G. A., & König, S. (2024). Single-step breeding value estimations and optimum contribution selection in endangered dual-

purpose German Black and WhitePied cattle (DSN) using a breed specific SNP chip. *Journal of Animal Breeding and Genetics*. Eingereicht am 10.06.2024.

Wu, Y., Zheng, Z., Visscher, P. M., & Yang, J. (2017). Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biology*, 18, 1-10. https://doi.org/10.1186/s13059-017-1285-0

Yin, H., Zhou, C., Shi, S., Fang, L., Liu, J., Sun, D., ... & Zhang, S. (2019). Weighted single-step genome-wide association study of semen traits in Holstein bulls of China. *Frontiers in Genetics*, 10, 1053. https://doi.org/10.3389/fgene.2019.01053

DANKSAGUNG

Es heißt, dass es ein ganzes Dorf braucht, um ein Kind großzuziehen – und ebenso brauchte es ein ganzes Netzwerk von wunderbaren Menschen, um diese Dissertation zu vollenden.

Ein großer Dank geht an meinen Doktorvater, Prof. Dr. Sven König, der mich mit großer Unterstützung auf meiner Reise durch die Wissenschaft begleitet hat. Für die herausragende Zusammenarbeit und großartige Betreuung möchte ich mich bei Frau Dr. Dr. Katharina May bedanken. Danke an PD Dr. Tong Yin für den intensiven Austausch über die Möglichkeiten und Grenzen von genetisch-statistischen Analysen. Ich danke auch Prof. Dr. Gudrun Brockmann und ihrer Arbeitsgruppe von der Humboldt-Universität zu Berlin, insbesondere Frau Dr. Paula Korkuć und Dr. Guilherme Neumann, für die wertvolle Zusammenarbeit während des DSN-SNP-Projekts.

Ich möchte allen meinen Kolleginnen und Kollegen am Institut für Tierzucht und Haustiergenetik danken. Die vielen inspirierenden Gespräche und freundschaftlichen Begegnungen lassen mich mit einem Lächeln auf die drei Jahre am Institut blicken. Insbesondere danke ich Patricia, Sina und Larissa für die schöne gemeinsame Zeit im Büro.

Meinen Eltern und Geschwistern gilt mein besonderer Dank. Sie haben mir stets den Rücken gestärkt und mich ermutigt, das Ziel nicht aus den Augen zu verlieren. Danke für all eure Unterstützung und lieben Worte.

Meine Freunde waren stets eine Quelle der Ermutigung und haben in den passenden Momenten für den nötigen Ausgleich gesorgt. Danke, dass ihr immer für mich da wart und mir geholfen habt, die nötige Balance zu finden.

Meiner Freundin Lena gilt an dieser Stelle ein ganz besonderer Dank. Trotz der Distanz, die während der Arbeit am Institut unumgänglich war, hast du mich auch in den aufreibenden Zeiten begleitet und gestärkt. Deine Liebe und Unterstützung in dieser Zeit und darüber hinaus waren keine Selbstverständlichkeit. Danke, Lena!

Diese Dissertation ist euch allen gewidmet. Danke, dass ihr das Dorf wart, dass nötig war um dieses Kind großzuziehen!

Manuel Wolf