

Universität für Bodenkultur
Department für Nachhaltige Agrarsysteme
Institut für Nutztierwissenschaften
Arbeitsgruppe Tierzucht



Genomische Selektion bei Fleckvieh mit niedriger Markerdichte

Diplomarbeit

vorgelegt von

Elisabeth Kirnbauer

Beurteiler:

Univ. Prof. DI Dr. Johann Sölkner

Betreuer:

Univ. Prof. DI Dr. Johann Sölkner

Dr. Gábor Mészáros

Dr. Gregor Gorjanc (Universität Ljubljana)

Wien, im Oktober 2012

Widmungen

Als erstes möchte ich meinen Betreuern Dr. Johann Sölkner und Dr. Gabor Meszaros meinen Dank aussprechen. Ich bedanke mich für die Unterstützung beim Verfassen meiner Diplomarbeit und die freundschaftliche Atmosphäre währenddessen. Besonders möchte ich Gabor für das Lösen meiner „kleinen“ Computerprobleme danken.

Ich bedanke mich auch bei meinen Freunden und meiner Familie für die Unterstützung und Aufmunterung zur richtigen Zeit, DANKE.

Danke, auch an meinen Freund Michael, dass er mich und meine Launen während des Studiums ertragen hat und mir beigestanden ist.

Außerdem möchte ich meinen Großvater, Gerhard Raganitsch, hervorheben für die unzähligen Male in denen er mich beraten, mir zugehört und geholfen hat. Ihm gebührt ein besonderer Dank.

Abschließend möchte ich mich noch bei der Zuchtdata für das Bereitstellen des Datenmaterials bedanken.

Inhaltsverzeichnis

Widmungen	Fehler! Es wurde keine Folge festgelegt.
Abbildungsverzeichnis	III
Tabellenverzeichnis	III
Abkürzungen	VIII
1 Einleitung.....	1
2 Literaturübersicht.....	2
2.1 Definition genomische Selektion	2
2.2 Stellungnahmen zur genomische Selektion	3
2.3 Molekulare Marker	4
2.4 SNP Verwendung.....	6
2.4 SNP Chips	7
2.5.1 BovineSNP50 Bead Chip.....	8
2.5.2 BovineHD Bead Chip.....	9
2.5.3 BovineLD Bead Chips	9
2.5.3.1 Bovine3K Bead Chip	9
2.5.3.2 Bovine7K Bead Chip	10
2.6 Sicherheit der Zuchtwertschätzung und Zuchtfortschritt	10
2.7 Linkage Disequilibrium und Effektive Populationsgröße	11
2.8 Phänotypische Daten	13
2.9 Imputation	13
2.10 Studien zur Imputation	15
2.11 Verlauf der genomischen Zuchtwertschätzung in Österreich	16
3 Material	18
3.1 Beschreibung der Rasse Fleckvieh	18
3.2 Genotypen	19
3.3 Phänotypen	20
3.3.1 Ausschlachtung	20
3.3.2 Eiweißprozent, Fettprozent und Fettkilogramm	21

3.3.3 Fruchtbarkeit.....	21
3.3.4 Nutzungsdauer	22
3.3.5 Zellzahl	23
3.4 Referenz- und Validierungsgruppe.....	23
4 Methoden.....	27
4.1 SNP Selektion	27
4.2 Methoden zur genomischen Zuchtwertschätzung	30
4.2.1 BayesB	30
4.2.2 GBLUP	31
4.3 Genauigkeit und Kalkulation der genomisch geschätzten Zuchtwerte	32
4.4 Imputationsmethode.....	33
5 Ergebnisse und Diskussion	34
5.1 Ergebnisse der genomischen Zuchtwertschätzung mit direkten Zuchtwerten unter Verwendung verschiedener SNP-Chips	34
5.2 Ergebnisse der genomischen Zuchtwertschätzung mit deregressierten Zuchtwerten unter Verwendung verschiedener SNP-Chips	38
5.3 Vergleich von direkten und deregressierten Zuchtwerten als Phänotypen	41
5.4 Ergebnisse der Imputation	44
5.5 Ergebnisse der imputierten SNP-Chips und Vergleich mit den originalen Chips	45
6 Conclusio.....	52
7 Zusammenfassung.....	54
8 Summary	56
9 Literaturverzeichnis	58
10 Tabellenanhang.....	65

Abbildungsverzeichnis

Abbildung 1: Verteilung der Fleckviehstiere in Referenz-und Testgruppe.....	24
---	----

Abbildung 2: Verteilung der Fleckviehstiere nach Geburtsjahrgängen	25
Abbildung 3: Verteilung der SNPs (nach Anzahl) auf den Chromosomen je nach verwendeten SNP-Chip	28
Abbildung 4: Verteilung der SNPs (in Prozenten) auf den Chromosomen nach verwendeten SNP-Chips	29
Abbildung 5: Korrelationen von EBVs und GEBVs von verschiedenen SNP Panelen berechnet mit BayesB	36
Abbildung 6: Korrelationen der EBVs und GEBVs mit BayesB und GBLUP berechnet für den 50K Chip	37
Abbildung 7: Korrelationen von deregressierten EBVs und GEBVs mit BayesB und GBLUP berechnet für den 50K Chip	41
Abbildung 8: Unterschiede zwischen mit BayesB berechneten Korrelationen mit direkten oder deregressierten Zuchtwerten als Phänotypen	42
Abbildung 9: Unterschiede zwischen mit GBLUP berechneten Korrelationen mit direkten oder deregressierten Zuchtwerten als Phänotypen	44
Abbildung 10: Unterschiede im Test-Set bei imputierten und originalen 3K Chip berechnet mit GBLUP	46
Abbildung 11: Korrelationen von ausgewählten Merkmalen	47

Tabellenverzeichnis

Tabelle 1: Anzahl (N), Mittelwert (Mean), Standardabweichung (Std.dev.), Minimum (Min) und Maximum (Max) von ausgewählten Zuchtwerten 26

Tabelle 2: Anzahl (N), Mittelwert (Mean), Standardabweichung (Std.dev.), Minimum (Min) und Maximum (Max) von ausgewählten deregressierten Zuchtwerten 26

Tabelle 3: SNP Unterschiede in den verschiedenen SNP-Chips 27

Tabelle 4: Korrelationen zwischen konventionellen EBVs und GEBVs für das Test-Set mit verschiedenen SNP-Panelen..... 35

Tabelle 5: Korrelationen zwischen deregressierten EBVs und GEBVs für das Test-Set mit verschiedenen SNP-Panelen..... 40

Tabelle 6: Unterschiede der Korrelationen von direkten und deregressierten Zuchtwerten geschätzt mittels BayesB 41

Tabelle 7: Unterschiede der Korrelationen von direkten und deregressierten Zuchtwerten geschätzt mittels GBLUP 43

Tabelle 8: Korrelationen zwischen EBVs und GEBVs für den 3K Chip 65

Tabelle 9: Korrelationen zwischen EBVs und GEBVs für den 7K Chip 65

Tabelle 10: Korrelationen zwischen EBVs und GEBVs für den 50K Chip 66

Tabelle 11: Korrelationen zwischen deregressierten EBVs und GEBVs für den 3K Chip	66
Tabelle 12: Korrelationen zwischen deregressierten EBVs und GEBVs für den 7K Chip	66
Tabelle 13: Korrelationen zwischen deregressierten EBVs und GEBVs für den 50K Chip	67
Tabelle 14: Korrelationen zwischen direkten EBVs und GEBVs von den originalen und imputierten Chips berechnet mit GBLUP	45
Tabelle 15: Korrelationen zwischen deregressierten EBVs und GEBVs von originalen und imputierten Chips berechnet mit GBLUP	48
Tabelle 16: Korrelationen zwischen direkten EBVs und GEBVs von den imputierten Chips	67
Tabelle 17: Korrelationen zwischen deregressierte EBs und GEBVs von den imputierten Chips	68

Abkürzungen

	Englisch	Deutsch
cM	Centi-Morgan	Centi-Morgan (genetisches Distanzmaß)
DGV	Direct Genomic Value	direkt genomischer Zuchtwert
EBV	Estimated Breeding Value	geschätzter Zuchtwert (phänotypisch)
GEBV	Genomic Estimated Breeding Value	genomisch geschätzter Zuchtwert
HD	High-Density	hohe Dichte
LD	Low-Density	niedrige Dichte
MAF	Minor Allel Frequency	Mindest- Allelfrequenz
Ne	Effective Population Size	Effektive Populationsgröße
QTL	Quantitative Trait Locus	quantitativer Merkmalsgenort
SNP	Single Nucleotide Polymorphism	Single Nucleotide Polymorphismus

1 Einleitung

Die Domestikation der Rinder (*Bos taurus*) fand vor circa 8.000 bis 10.000 Jahren statt. Die Art *Bos taurus* hat sich im Laufe der Geschichte zu einer der ökonomisch und kulturell wichtigsten Spezies der Erde entwickelt. Die vielen verschiedenen Rinder wurden auf Grund ihrer unterschiedlichen Eigenschaften wie Milchmenge oder Fleischqualität selektiert, sodass sich bis heute mehr als 800 Rassen entwickelt haben (Lewis et al. 2011).

Die traditionelle Selektion von quantitativen Merkmalen mit ökonomischer Bedeutung basierte auf den eigenen phänotypischen Leistungen und auf den erbrachten Verwandtschaftsleistungen des Tieres (Meuwissen et al. 2001).

Die neue Technik der genomischen Selektion revolutioniert zurzeit die Milchrinderzucht. Bei der genomischen Selektion werden mit Hilfe von sogenannten genetischen Markern direkt genomische Zuchtwerte (DGV) auf DNA-Ebene geschätzt. Der genomischen Zuchtwertschätzung gingen zwei wichtige Entwicklungen voraus. Die erste war die Sequenzierung des Rindergenoms, wofür man tausende SNPs (Single Nukleotide Polymorphismen), eine Form von DNA-Markern, entdecken musste. Die zweite wichtige Entwicklung war die Erkenntnis, dass man mit alleiniger Verwendung von genomischer Selektion relativ hohe Sicherheiten für die Zuchtwertschätzung erzielt (Hayes et al. 2009).

Durch die Nutzung von genomischen Markern ist es möglich, sicher geschätzte Zuchtwerte für junge Tiere beider Geschlechter zu erhalten, dies wirkt sich positiv auf die Erhöhung des Zuchtfortschritts durch Verringerung des Generationsintervalls aus. Außerdem kann der Kostenaufwand für Prüftiere gesenkt werden (Lillehammer et al. 2011, Moser et al. 2009).

Momentan wird die genomische Selektion beim Fleckvieh mit SNP-Daten des Illumina Bovine SNP50 Bead Chip mit rund 54.000 SNPs durchgeführt. Seit September 2011 bietet die Firma Illumina den neuen Bovine LD (low density) BeadChip mit 6.909 SNPs zu einem deutlich niedrigeren Preis an (Illumina 2011).

Ziel dieser Arbeit ist es zu untersuchen, wie sicher genomische Zuchtwertschätzung mit einem solchen Bovine LD Bead Chip durchgeführt werden kann.

2 Literaturübersicht

Die Literaturstudie dieser Arbeit wird in diesem Kapitel wiedergegeben. Es werden für diese Arbeit relevante Begriffe und interessante Resultate von anderen Studien, sowie der Verlauf der genomischen Zuchtwertschätzung erläutert.

2.1 Definition Genomische Selektion

Bisher konnte das genetische Potential eines Tieres nur über dessen Leistung oder die Leistung von verwandten Tieren geschätzt werden. Bei der genomischen Zuchtwertschätzung beziehungsweise genomischen Selektion wird der Zuchtwert nicht, wie häufig angenommen, direkt vom Erbgut bestimmt, sondern er wird über sogenannte genetische Marker (SNPs) geschätzt. Grundlage dafür sind über das gesamte Genom verteilte Variationen der DNA (Fürst 2011).

„Die grundlegende Theorie der genomischen Selektion ist einfach: der Zuchtwert eines Tieres soll direkt aus seinen Erbanlagen (dem genetischen Code) abgeleitet werden. Dazu müssen genetische Marker in ausreichender Zahl über das gesamte Genom vorhanden sein. Genetische Marker sind erkennbare Markierungen im Genom, welche oft sehr nahe bei Genen mit Einfluss auf ein Leistungsmerkmal liegen. Diese genetischen Marker werden SNP-Marker genannt. In einer aufwendigen statistischen Analyse wird für jeden dieser SNPs seine Beziehung bzw. sein Anteil am Zuchtwert geschätzt. Der genomische Zuchtwert eines Tieres wird dann einfach aus der Summe aller SNP-Effekte berechnet. So kann für Tiere schon sofort nach der Geburt ein genomischer Zuchtwert geschätzt werden. Die Sicherheiten des genomischen Zuchtwertes sind allerdings nicht 100 %, sondern liegen derzeit je nach Schätzverfahren größtenteils bei 40 bis 60 %“ (verändert nach Gredler o.J.).

Die Grundlage einer aussagekräftigen genomischen Zuchtwertschätzung ist eine solide Datenqualität, deshalb werden sämtliche Genotypdaten, die in die

Zuchtwertschätzung eingehen, genauestens geprüft. Im Verlauf dieser Überprüfung werden unzuverlässige Marker verworfen bzw. Tiere ausgeschlossen, deren genomisch bestimmte Verwandtschaft einen Konflikt mit der abstammungsbasierten Verwandtschaft aufweist.

Ein eindeutiger Vorteil der genomischen Zuchtwertschätzung ist, dass durch die Marker die Verwandtschaftsverhältnisse genauer bestimmt werden können. In der konventionellen Zuchtwertschätzung hatten Großväter und Enkel immer ein Verwandtschaftsverhältnis von 0,25. In der genomischen Zuchtwertschätzung werden durch Marker Verhältnisse zwischen 0,15 – 0,35 beobachtet (Fürst 2011).

Man unterscheidet in der Zuchtwertschätzung zwischen genomischem und genomisch optimiertem Zuchtwert:

- **„Genomisch direkter Zuchtwert (DGV):** Der genomisch direkte Zuchtwert, ist jener Zuchtwert, der aus den Informationen aus dem Erbgut (Genom, SNP-Marker) geschätzt wird.
- **Genomisch optimierter Zuchtwert:** Beim genomisch optimierten Zuchtwert wird der genomisch direkte Zuchtwert mit dem Ahnenindex bzw. dem konventionellen ZW über eine Gewichtung kombiniert. Er ist dem genomisch direkten Zuchtwert von der Sicherheit und Aussagekraft überlegen“ (Fürst 2011).

Die aktuell verwendete Methode zur Schätzung des Zuchtwertes ist GBLUP (genomic BLUP), wobei international auch die Methode BayesB zum Einsatz kommt (Fürst 2011).

2.2 Stellungnahmen zur genomischen Selektion

Die Verwendung von molekularen Markern zur Vorhersage von genetischen Werten führt zu einem schnelleren Zuchtfortschritt als er im Vergleich mit dem traditionellen Selektionschema erreichbar ist (Meuwissen et al. 2001).

Genomische Selektion ermöglicht eine sehr präzise Aussage über Zuchtwerte sehr junger Tiere (Hayes et al. 2008).

Die Berechnung und der Einsatz von genomischen Werten werden allgemein als genomische Selektion bezeichnet. Die Genomische Selektion wird in der Zukunft einen Haupteinfluss auf die Milchviehzucht haben (Van der Beek o.J.).

Genomische Selektion verspricht die Technologie zu werden, mit der man den höchsten Zuchtfortschritt der letzten 20 Jahre erreichen kann (Hayes et al. 2008).

Der hohe Anteil an genotypisierten Bullen und die hohe Zuverlässigkeit der Genotypisierung lässt eine umfangreiche Abstammungsuntersuchung zu. Dies ermöglicht ein großes Vertrauen in die genomische Selektion. (Wiggans et al. 2009).

Genomische Daten erhöhen stark die Sicherheit der Zuchtwertschätzung, wenn die Genotypen auf die Phänotypen für eine große Anzahl an Tieren abgestimmt werden (Van Raden et al. 2008a).

Genomische Selektion ist die Methode, die in den meisten Ländern für die Zuchtwertschätzung von Milchkühen verwendet wird, wobei die Kosten für die Erfassung eines Genotyps mit einem HD-Panel für den einzelnen Landwirt oft zu hoch sind. Mittels Imputierung kann von einem günstigeren LD-Panel auf ein HD-Panel geschlossen werden und damit werden auch die Kosten für die genomische Selektion reduziert (Berry und Kearney 2011).

2.3 Molekulare Marker

In der klassischen Tierzucht wird der Genotyp über das Infinitesimalmodell dargestellt. Dieses Modell geht davon aus, dass die Ausprägung eines Merkmals von einer unendlichen Anzahl von Loci bestimmt wird, jedoch hat jeder einzelne Locus einen sehr kleinen Effekt auf das Merkmal (Hayes und Goddard 2001). Die Identifizierung von Quantitative Trait Loci (QTL's), welche einen deutlich messbaren Effekt auf ein Merkmal besitzen, wird durch die moderne Molekulargenetik ermöglicht

(Neuner 2009). Ob die Molekulargenetik angewendet werden kann um Populationen genetisch zu verbessern, hängt davon ab ob es möglich ist, Tiere für bestimmte Genorte zu genotypisieren (Dekkers 2004). Um die Beziehung zwischen Marker und QTL aufzudecken, wurden in der Vergangenheit hauptsächlich Mikrosatellitenmarker verwendet. Heutzutage kommen zum größten Teil SNP-Marker zum Einsatz. Bei einem SNP handelt es sich um einen Austausch einer einzelnen Base im Genom (Neuner 2009). Aufgrund dessen wie Marker und QTL in Beziehung stehen, können drei Arten von genetischen Markern unterschieden werden:

- Direkte Marker: Orte, die für eine funktionelle Mutation codieren
- LD-Marker: Orte, die sich in einem populationsweiten Linkage Disequilibrium (Kopplungsungleichgewicht) mit einer funktionellen Mutation befinden
- LE-Marker: Orte, die sich in einer ausgezüchteten Population einem populationsweiten Linkage Equilibrium (Kopplungsgleichgewicht) mit einer funktionellen Mutation befinden (Dekkers 2004)

Die Verwendung der drei verschiedenen Marker in Selektionsprogrammen differiert sehr stark. Die direkten Marker und zu einem geringeren Grad die LD-Marker ermöglichen, wegen der konsistenten Bindung von Phänotyp und Genotyp über die Population hinweg, die Selektion. Die LE-Marker müssen aufgrund der verschiedenen Kopplungsphasen zwischen Marker und QTL innerhalb einzelner Familien bestimmt und berücksichtigt werden (Dekkers 2004). Molekulare Marker wurden benutzt um Genorte und Chromosomenregionen zu identifizieren, welche ein monogenes Merkmal oder ein quantitatives Merkmal beeinflussen. Zu dem Zweck der QTL- Erkennung und Verwendung können quantitative Merkmale eingeteilt werden in:

- routinemäßig aufgezeichnete Merkmale (Milchmenge, Fettprozent)
- schwierig aufzuzeichnende Merkmale (Futterraufnahme, Produktqualität)
- unaufgezeichnete Merkmale (Krankheitsresistenzen)

Diese Merkmale können noch unterteilt werden, je nachdem man Werte von beiden Geschlechtern erhält, ob es sich um geschlechtsspezifische Merkmale handelt oder ob man die Merkmale erst spät im Leben erfassen kann (Dekker 2004, Meuwissen und Goddard 1996). Die Anzahl der Marker in einem Marker Panel kann von einem einzelnen Marker bis zu tausenden von Markern reichen. Potenziell wäre sogar die

komplette DNA-Sequenz eines Tieres denkbar. Die Ergebnisse eines DNA-Test für ein Merkmal werden normaler Weise zusammengefasst zu einem Single-Marker-Score (Einzel-Merkmal-Punktzahl) oder einem Molekularen Zuchtwert. In den meisten Fällen ist der Marker-Score eine gewichtete Summe der Anzahl an Kopien von verschiedenen gewichteten Allelen, welche von der Referenzpopulation geschätzt wurden (Kachman 2008).

Die Exaktheit mit der Zuchtwerte über die Marker geschätzt werden können, ist von zwei Faktoren limitiert.

Erstens ist unter Umständen das Linkage Disequilibrium zwischen Marker und QTL unvollständig, sodass der Marker Haplotyp nicht die gesamte Varianz auf dem QTL erklären kann (Meuwissen et al. 2001). Ein Haplotyp beschreibt die Erbinformation, die in einem Chromosom vorliegt. Sie zeigen an, welche Allele auf jedem Chromosom sind und können zwischen dem mütterlichen Chromosom aus der Eizelle aus dem väterlichen Chromosom aus dem Spermium erbracht unterscheiden. Zum Beispiel ist ATCCC ein Haplotyp, der einen Chromosomenabschnitt beschreibt, welcher aus den Basen Adenin, Thymin, und dreimal Cytosin besteht. Ein Genotyp beschreibt die Erbinformation, die in einem Chromosomenpaar vorliegt, durch die Kombination zweier Haplotypen. Genotypen zeigen nur, wie viele Kopien jedes Allels ein Individuum von seinen beiden Eltern geerbt hat (Van Raden et al. 2011, Elberfeld 2007). Eine Anzahl von Markern, welche rund um einen QTL verstreut ist, bildet zusammen einen Markerhaplotyp (Dekkers 2004).

Der zweite limitierende Faktor ist der Stichprobenfehler bei der Schätzung von Haplotypen, wobei es zu einer Erhöhung des Fehlers führt, wenn sich die Umweltvarianz zwischen den genotypisierten und den leistungsgeprüften Tieren vergrößert (Meuwissen et al. 2001).

2.4 SNP Verwendung

Genotypen von circa 40.000 SNPs fungieren jetzt als dritte Quelle von Daten für die genetische Evaluierung für die Milchwirtschaft, zusätzlich zu den Phänotypen und Verwandtschaftsdaten, welche die Grundlage der Selektion der letzten 100 Jahre bildeten (Van Raden et al. 2008a). SNPs sind für die Verwendung als Marker sehr geeignet, da sie sehr häufig vorkommen. Die klassische Annahme ist, dass in 1.000

Basenpaaren mehr als 1 SNP vorkommt, beziehungsweise mehr als 3 Millionen SNPs im Genom existieren. Außerdem haben SNPs nur 2 Allele, weshalb die Genotypisierung nur eine Plus-Minus Analysierung verlangt. Wegen der 2 Allele wird ein SNP-Marker auch als Biallelic-Marker bezeichnet (Kryglyak 1997). Durch die Entwicklung von Hochdurchsatzverfahren wurde die Genotypisierung von Proben mittels SNPs leistbar und führte dazu, dass Tausende von Tieren aus der Milchwirtschaft genotypisiert wurden, wobei hauptsächlich leistungsgeprüfte Tiere in Besamungsprogrammen beziehungsweise Selektionskandidaten für diese Besamungsprogramme analysiert wurden. Die Genotypisierung wird vor allem mit dem BovineSNP50 Bead Chip der Firma Illumina durchgeführt (Weigel et al. 2010). Bei diesem Verfahren wird der Chip mit der DNA des Tieres versetzt. Danach werden mittels Computerprogrammen die Reaktionen zwischen Chip und DNA ausgewertet. Das Ergebnis besteht für jedes Tier aus den Genotypen, welche sich an über 50.000 verschiedenen Genorten befinden (Gredler o.J.). Die Qualität von genomischen Daten wird von Wiggans et al. (2009) als sehr hoch beschrieben. Die meisten SNPs haben weniger als 1% fehlende Genotypen und die Fehlerquoten liegen bei weniger als 0,05%. Durch den SNP Vergleich eng verwandter Tiere können fehlende Genotypen eingetragen und Fehler korrigiert werden (Van Raden 2008a). Wiggans et al. (2009) kamen zu der Erkenntnis, dass viele polymorph ausgeprägte SNPs bei der Rasse Holstein, bei den Rassen Jersey und Brown Swiss monomorph ausgeprägt sind. Dies ergab, dass entweder rassenspezifische SNP-Sets für die Genotypisierung notwendig sind, oder dass alle Rassen in dem SNP Auswahlprozess berücksichtigt werden müssen.

Goddard et al. (2009) haben in einer Studie beschrieben, dass eine Steigerung der Anzahl an SNPs über 50.000 nicht notwendig ist, solange man innerhalb einer Rasse arbeitet.

2.5 SNP Chips

Habier et al. (2009) beschrieben in ihrer Studie, dass die Genotypisierung von HD-SNP Chips nicht zu kostenintensiv sein darf, wenn in einem Zuchtprogramm sehr viele Selektionskandidaten enthalten sind oder wenn der ökonomische Nutzen im

Vergleich zu den Kosten der Genotypisierung klein ist. Dies ist zum Beispiel in Rinderzuchtprogrammen bei Kühen der Fall, da diese sehr geringe Reproduktionsraten haben. Aktuell sind die Kosten für die Genotypisierung mit LD-SNP Panelen wesentlich geringer als jene mit HD-SNP Panelen, deshalb besteht momentan ein großes Interesse daran, Möglichkeiten zu finden um die genomische Selektion mit LD-SNP Panelen realisieren zu können (Habier et al. 2009).

In Europa betragen die Kosten für einen Nachkommen geprüften Stier betragen \$500.000. Die Kosten für die Genotypisierung mittels 50K Chip betragen inklusive Labor \$500. Die Genotypisierung mittels 50K Chip ist auf jedenfalls eine Alternative um Kosten zu reduzieren, allerdings kommt sie trotzdem reduzierten Kosten nur für Elitestiere zum Einsatz. Für die weibliche Population sind die Kosten immer noch zu hoch, deshalb wurden LD-Chips entwickelt, welche deutlich billiger sind als der 50K Chip und mit Hilfe der Imputation gute Ergebnisse erzielen (Dassoneville 2012).

2.5.1 BovineSNP50 Bead Chip

Der BovineSNP50 Bead Chip ist der am häufigsten verwendete Chip zur Genotypisierung von Rindern. Er enthält 54.001 informative SNPs, welche gleichmäßig über das Rindergenom verteilt sind. Die Anwendung eines solchen SNP Chips ermöglicht die Identifikation von QTL's, die Bewertung der genetischen Veranlagung eines Tieres und damit die genomische Selektion, da die Leistung von interessanten Phänotypen vorausgesagt werden kann. Mittlerweile hat die Firma Illumina bereits die zweite Generation dieses Chips herausgebracht, den BovineSNP50 v2 Bead Chip mit 54.609 SNPs (Illumina 2011). Der Illumina BovineSNP50 Bead Chip bietet ein exaktes und zuverlässiges SNP-Set zur genomischen Evaluierung (Wiggans et al. 2008). Allerdings können für die Berechnungen nie alle SNPs verwendet werden, da sie bestimmte Anforderungen erfüllen müssen. Zum Beispiel wurde in der Studie von Wiggans et al. (2008) die Anzahl der SNPs vom ursprünglich verwendeten Bovine SNP50 BeadChip auf 40.874 SNP für die Rasse Holstein reduziert. Hier mussten die SNPs eine MAF (Mindest-Allelfrequenz) von größer 2 Prozent haben, durften nicht hochgradig miteinander korrelieren und durften keine Wertungsprobleme haben, ansonsten wurden sie eliminiert (Wiggans et al. 2008).

Die Kosten für einen 50K Chip betragen 2009 ungefähr \$209 und dieser Preis ist inzwischen deutlich gesunken. Im Moment beträgt er circa zwischen \$125 und \$135 pro Chip (Dassonville 2012, Schefers und Weigel 2012). Die gesamten Kosten für eine Genotypisierung, was sowohl den Chippreis, als auch den Laborpreis beinhaltet, kommen auf \$500 pro Tier (Dassoneville 2012).

2.5.2 BovineHD Bead Chip

Die im Moment umfassendste genomweite Genotypisierung kann mittels des BovineHD Bead Chip der Firma Illumina durchgeführt werden. Der Chip enthält mehr als 700.000 SNPs, mit denen man die genetischen Unterschiede zwischen Fleisch- und Milchrinderrassen untersuchen kann. Außerdem ermöglicht der BovineHD Bead Chip eine breite Palette von Anwendungen wie zum Beispiel genomweite Selektion, Identifikation von quantitativen Merkmalsgenorten, Evaluierung der genetischen Leistungen, Linkage Disequilibrium Studien, vergleichende genetische Studien und Rassencharakterisierungen für die Evaluierungen von Biodiversität (Illumina 2010). Da dieser SNP Chip für diese Arbeit keine Rolle spielt, ist er hier nur vollständigheitshalber angeführt.

2.5.3 BovineLD Bead Chips

Die BovineLD Bead Chips stellen eine kostengünstige Alternative zum üblichen in der Milchwirtschaft verwendeten Bovine50SNP Bead Chip dar. Die LD Chips sind dafür gedacht, dass wegen der verringerten Kosten auch Tiere mit einem geringeren Zuchtwert genotypisiert werden (Illumina 2011).

2.5.3.1 Bovine3K Bead Chip

Seit dem Jahr 2009 ist der Golden Gate Bovine3K Bead Chip in dem Sortiment der Firma Illumina vorhanden und wird als LD Chip bezeichnet. Er enthält 2.900 SNPs

und wurde entwickelt um die Kosten der genomischen Selektion zu reduzieren. Deshalb wurde auch die günstigere Golden Gate Technologie verwendet (Illumina 2010, Dassonneville et al. o.J.). Der Chip enthält eine Auswahl von wichtigen SNPs um für etliche interessante Merkmale eine Leistung der Tiere vorhersagen zu können. Wie auch im Folgenden erwähnten 7K Chip werden hier die auf den 50K Chip fehlenden SNPs imputiert. Die Imputationseffizienz (Anteil an korrekt imputierten SNPs) liegt hier zwischen 80-90%. (Illumina 2010).

2.5.3.2 Bovine7K Bead Chip

Seit September 2011 bietet die Firma Illumina einen 7K Chip an, welcher 6.909 SNPs enthält. Der 7K Chip ist eine neuere Variante der LD Chips und sollte den 3K Chip durch eine bessere Imputationseffizienz ersetzen.

Die SNPs sind gleichmäßig über das Rindergenom verteilt, jedoch mit einer höheren Dichte an den Chromosomenenden um die Imputationseffizienz zu steigern. Die Firma Illumina gibt diese 6.909 enthaltenen SNPs als perfekte Basis an um mit einer hohen Exaktheit auf den Bovine50SNP Bead Chip zu imputieren. Die Imputation nutzt den Vorteil des Linkage Disequilibrium, dem Phänomen, dass Gruppen von Allelen in Blöcken entlang des Chromosoms vererbt werden.

Die demonstrierte Imputationseffizienz der Firma Illumina mit dem BovineLD Bead Chip lag bei über 98%. Damit ermöglicht dieser Chip ebenfalls die Erkenntnis von den genetischen Auswirkungen auf die Milchproduktion, die Gesundheit und die Reproduktion (Illumina 2011). Die Kosten betragen zwischen \$43 und \$55 (Scheffers und Weigel 2012).

2.6 Sicherheit der Zuchtwertschätzung und Zuchtfortschritt

Milchviehzüchter bezeichnen die quadrierte Korrelation des geschätzten Zuchtwertes zum wahren Zuchtwert als Sicherheit der Zuchtwertschätzung (Van Raden et al. 2008a). Traditionell wurden die Sicherheiten der Zuchtwertschätzung durch Benutzung der Verwandtschaftsmatrix aus dem elterlichen Mittel errechnet. Die zukünftigen Leistungen der Jungtiere konnten allerdings nur mit einer Sicherheit von

32% vorhergesagt werden. Van Raden fand in einer Studie (2008) heraus, dass die mittels genomischer Daten erlangten Zuchtwerte mit Sicherheiten von über 60% geschätzt werden können. Dies bedeutet, genomische Daten entsprechen dem Wert von 20 Töchterleistungen und jener Wert sollte sich umso mehr erhöhen je mehr verwandte Tiere genotypisiert werden (Van Raden 2008). Die Sicherheit der genomischen Zuchtwertschätzung hängt von 4 Parametern ab:

- der Vollständigkeit zwischen Linkage Disequilibrium, Marker und QTL
- der Anzahl der geno- und phänotypisierten Tiere der Referenzpopulation, von der die SNP-Effekte geschätzt werden
- der Verteilung der QTL-Effekte
- der Heritabilität des Merkmals (Hayes et al. 2009)

Die Genauigkeit der Selektion durch die genomische Zuchtwertschätzung ist vergleichbar mit der nach der Leistungsprüfung durchgeführten Selektion. Der Einsatz der genomischen Zuchtwertschätzung wird den Zuchtfortschritt drastisch erhöhen, besonders bei Merkmalen, die sich phänotypisch nicht einfach erfassen lassen wie zum Beispiel: Merkmale mit niedriger Heritabilität, geschlechtsspezifische Merkmale, Krankheitsresistenzen oder Schlachtmerkmale (Meuwissen et al. 2001).

Zuchtunternehmen nutzen die höheren Sicherheiten der Zuchtwerte, welche durch die genomische Zuchtwertschätzung erreicht werden auf 2 verschiedene Arten. Eine Strategie ist die Genotypisierung aller potentiellen Zuchtstierkandidaten um nur für genomisch geprüfte Topstiere eine Nachkommensprüfung durchzuführen. Dies reduziert die Kosten des Zuchtprogramms und erhöht den Zuchtfortschritt. Eine andere Strategie ist, nur auf die genomischen Zuchtwerte zu vertrauen, dies reduziert das Generationsintervall und erhöht damit den Zuchtfortschritt (Hayes 2008).

2.7 Linkage Disequilibrium und Effektive Populationsgröße

Die klassische Definition des Linkage Disequilibrium (Kopplungsungleichgewicht) bezieht sich auf die nicht zufällige Verbindung von Allelen zwischen zwei Loci (Hayes 2008). Ein Linkage Disequilibrium tritt auf, wenn Genotypen auf zwei Loci nicht unabhängig voneinander sind (De Roos et al. 2008).

Der Einsatz von direkten Markern für QTL-Mapping und für die Selektion basiert auf dem Linkage Disequilibrium zwischen Marker und QTL, wobei das Linkage Disequilibrium zwischen Marker und QTL immer innerhalb einer Familie existent ist. Manchmal kann es jedoch auch auf dem Populationslevel existieren. LD-Marker werden immer in einer Referenzpopulation entdeckt, in der das ursprüngliche Experiment durchgeführt wird, wie zum Beispiel einer genomweiten Assoziationsstudie. Der Wert der Marker hängt außer in der Referenzpopulation von der Persistenz der Linkage Disequilibrium Phase zwischen Referenzpopulation und der zweiten Population ab (Dekkers und Hospital 2002). Es ist wichtig die Persistenz einer Linkage Disequilibrium Phase zu kennen, da das Linkage Disequilibrium zwischen Marker und QTL, welches in einer Rasse entdeckt wurde, in einer anderen Rasse nicht dasselbe sein muss.

De Roos et al. (2008) erwähnten in ihrer Studie, dass mit zunehmender genomischer Distanz die Linkage Disequilibrien schneller zerfallen. In ihrer Studie verglichen sie die Ausdehnung des Linkage Disequilibrium über verschiedenen Rinderrassen und kamen zu dem Ergebnis, dass das durchschnittliche Linkage Disequilibrium größer ist, wenn die Rassen näher verwandt sind.

Die Linkage Disequilibrien sind außerdem von der effektiven Populationsgröße (N_e) abhängig. Die effektive Populationsgröße ist definiert als die Anzahl der Individuen in einer idealisierten Population mit zufälliger Paarung und ohne Selektion, die auf die gleiche Rate der Inzucht führen würde wie in der realen Bevölkerung beobachtet wird. Die effektive Populationsgröße kann wegen des ungleichen genetischen Beitrags des einzelnen Individuums an die nächste Generation sehr viel kleiner sein als die tatsächliche Populationsgröße (Goddard und Hayes 2009). Linkage Disequilibrien auf kurzen Distanzen auf dem Genom sind eine Funktion von N_e viele Generationen zuvor, während die Linkage Disequilibrien auf langen Distanzen eine jüngere Populationsgeschichte widerspiegeln. Das bedeutet, dass nicht nur die aktuelle N_e wichtig ist, sondern auch die vergangene N_e . N_e für Tierarten mögen in der Vergangenheit viel größer gewesen sein, als sie es heute sind. In der jüngsten Vergangenheit haben die verbreitete Verwendung der künstlichen Besamung und der starke Einsatz von ein paar Elite-Vererbern die N_e stark reduziert (Hayes 2008). Der Zerfall von Linkage Disequilibrien gibt an, dass sich die effektive Populationsgröße bei Rindern vor circa 10.000 Generationen verringerte. Von ungefähr 10.000 Tieren ausgehend, reduzierte sich die N_e nach der Domestikation

ein weiteres Mal auf ein paar Tausend Tiere und für die modernen Rinderpopulationen fand eine Reduktion auf ~100 Tiere statt (Roos et al. 2008).

2.8 Phänotypische Daten

Genotypen müssen auf Phänotypen abgestimmt werden um SNP-Effekte schätzen zu können. Milchviehzüchter besitzen ein hochentwickeltes System um weltweite Zuchtwerte miteinander vergleichen zu können. Nationale Rechenzentren nutzen größtenteils BLUP-Modelle zur Berechnung von geschätzten Zuchtwerten für Merkmale. Die national geschätzten Zuchtwerte werden dann an Interbull Center mit Sitz in Uppsala (Schweden) gesendet. Mit Hilfe dieser Daten schätzt Interbull die genetische Korrelation der Leistungen für dasselbe Merkmal über verschiedene Länder hinweg und benutzt dafür die MACE (Multi-trait across country Evaluierung), um einen geschätzten Zuchtwert für jeden Stier für die erwartete Leistung in jeder Umwelt zu erhalten. Etwa 25 Merkmale werden zurzeit evaluiert. Die genomische Evaluation wird mittels Gebrauch von geschätzten Zuchtdaten anstatt von Rohdaten als Inputs berechnet. Internationale Kooperation ist für Rassen mit einer geringeren Anzahl an Tieren wie zum Beispiel Jersey oder Brown Swiss viel bedeutender als für Rassen mit einer hohen Anzahl an Individuen wie zum Beispiel Holstein (Van Raden et al. 2008a).

2.9 Imputation

Unbekannte Genotypen können aus beobachteten Genotypen an den gleichen oder in der Nähe von verwandten Loci „bekannt gemacht“ (imputiert) werden (Van Raden et al. 2011).

Imputationsmethoden arbeiten durch die Kombination eines Referenzpanels bestehend aus Individuen, die mittels SNPs genotypisiert werden und einer Testgruppe von einer genetisch ähnlichen Population, welche nur für eine Teilmenge dieser SNPs genotypisiert wird (Howie et al. 2009).

Manche SNPs sind in beiden Gruppen (Referenzgruppe und Testgruppe) inkludiert, allerdings hat die Testgruppe eine große Anzahl an SNPs mit fehlenden Genotypdaten. Die verschiedenen Imputationsmethoden versuchen die fehlenden Genotypen mit Hilfe von Algorithmen vorherzusagen, wobei die in beiden Gruppen vorkommenden SNPs eine Verbindung zwischen den Gruppen schaffen (Dassonneville 2012).

Das bedeutet, dass die Haplotypen der Untersuchungsprobe mit den „dichteren“ Haplotypen des Referenzpanels verglichen werden um Übereinstimmungen zu finden. Fehlende Haplotypen der Untersuchungsprobe werden dann durch die Benutzung von übereinstimmenden Haplotypen aus dem Referenzpanel imputiert. In realen Beispielen werden die Genotypen mit Unsicherheit unterstellt und es wird eine Wahrscheinlichkeitsverteilung über alle drei möglichen Genotypen erzeugt (Marchini and Howie 2010).

Stammbauminformationen (Beziehungen zwischen den Individuen) beziehungsweise die Informationen über Position der Marker auf dem Genom können zusätzliche Informationen zur Verbesserung der Imputation enthalten. Der wesentliche Unterschied zwischen den humangenetischen Datensätzen und jenen in der Nutztierwissenschaft liegt in der "Tiefe" der Ahnentafel. Humangenetischen Datensätze werden in der Regel auf kleine Familien oder sogar nicht verwandten Individuen aufgebaut. Auf der anderen Seite werden die genetischen Daten von Nutztierpopulationen üblicherweise von verwandten Individuen über mehrere Generationen hinweg gebildet. Zum Beispiel werden im Milchvieh Halbgeschwisterfamilien über mehrere Generationen hinweg untersucht und bilden riesige Datensätze. Angesichts dieser Unterschiede entwickelten einige auf Tierpopulationen spezialisierte Genetiker gezielt spezielle Software, die die Vorteile der sehr spezifischen Familienstruktur der Tierbestände zu verwerthen im Stande sind. In der Tat, wenn ein oder beide Elternteile genotypisiert werden, können einige spezifische Regeln angewendet werden um teilweise mit Sicherheit den Genotyp der Nachkommen zu bestimmen. Diese Regeln basieren auf der Mendelschen Spaltungsregel. Das einfachste Beispiel ist, wenn der Vater an einem Locus homozygot ist, kennt man mit Sicherheit das väterliche Allel der Nachkommen. (Dassonneville 2012).

Bis heute haben die meisten genomweiten Imputationsanalysen Referenzdaten vom internationalen HapMap Projekt verwendet. Während diese Strategie erfolgreich war, können Studien in naher Zukunft auf zusätzliche Referenzinformationen zugreifen, wie zum Beispiel auf Kontrollgruppen von dem 1.000-Genom-Projekt, welche für mehrere SNP Chips genotypisiert werden. Die neuen Referenzpanele sollten die Qualität und den Umfang der Anrechnung verbessern, aber sie stellen auch eine neue methodische Herausforderung dar. In ihrer Studie haben sich Howie et al. mit einer neuen Genotyp Imputationsmethode beschäftigt, IMPUTE version2, die den neuen Herausforderungen gewachsen ist. Ein Referenzpanel kann Tausende von Chromosomen enthalten und damit höhere Genauigkeiten erreichen als mit HapMap allein (Howie et al. 2009).

2.10 Studien zur Imputation

Bisherige Ergebnisse zeigen, dass die Benutzung von HD SNP Genotypen den genetischen Prozess in der Milchviehzucht bedeutend verbessert (Weigel et al. 2010). Weigel et al. (2010) sind jedoch der Meinung, dass dies zu den derzeitigen Preisen nur für Elitetiere in Frage kommt. Laut ihrer Studie könnte die Entwicklung der Imputation von LD auf HD Genotypen zu einer weit verbreiteten Akzeptanz der Genetik auf kommerziellen Betrieben führen.

Die Kosten der genomischen Selektion können durch die Genotypisierung des Bovine3K Chips und der anschließenden Verwendung von Algorithmen zur Imputation auf den Bovine50K Chip erheblich gesenkt werden, wobei der Schlüssel zu einer genauen Imputation in der Kenntnis der HD Genotypen im Stammbaum des zu imputierenden Tieres liegt. Die Senkung der Kosten der genomischen Selektion führt zu einem höheren Nutzen für die Landwirte. Dadurch vergrößert sich die Referenzpopulation für die Schätzung der SNP-Effekte und auch die Genauigkeit der Vorhersage des Zuchtwerts. Anschließend erhöht sich auch der Zuchtfortschritt (Berry und Kearney 2011).

Habier et al. (2009) untersuchten in ihrer Studie den Verlust an Genauigkeit von genomisch geschätzten Zuchtwerten, wenn zur Genotypisierung die LD SNPs

eingesetzt werden und kamen zu dem Ergebnis, dass in der ersten Generation mit weniger als 5% der Verlust an Genauigkeit sehr klein ist (bei einer LD SNP Dichte von 10 cM), sich aber im Laufe von Generationen vergrößert.

Informationen über die umliegenden Marker verbessern die Genauigkeit der Imputation nur, wenn die Tiere deren Genotypen imputiert werden, auch für die umliegenden Marker genotypisiert werden. Die Ergebnisse liefern eine geringere Steigerung, wenn das Kopplungsungleichgewicht (Linkage Disequilibrium) zwischen den SNPs niedrig war, aber es kommt zu einer wesentlichen Steigerung der Genauigkeit der Imputation, wenn das Kopplungsungleichgewicht hoch ist (Calus et al. 2011).

Berry und Kearney (2011) zeigten, dass es bei genomischen Vorhersagen für Zuchtwerte keine, beziehungsweise nur eine kleine Differenz gibt, vergleicht man direkte genomische Zuchtwerte von realen oder imputierten Genotypen. Die Differenzen variierten von 0.13 (Fettklasse des Schlachtkörpers der Nachkommen) bis 0.47 (Fortbewegung). Die durchschnittliche Korrelation der direkt genomisch geschätzten Zuchtwerte zwischen den realen oder imputierten Genotypen, für die 15 geschätzten Merkmale, betrug 0.97. Deshalb kann man auf eine gute Übereinstimmung zwischen den realen und imputierten Genotypen schließen. Ergebnisse der Studie zeigen, dass verfügbare HD Markerpanele von einem im Handel verfügbaren LD Markerpanel imputiert werden können. Dies führt zu einer Kostenreduktion in der genomischen Selektion (Berry und Kearney 2011).

2.11 Verlauf der genomischen Zuchtwertschätzung in Österreich

Seit August 2010 wurde die genomische Zuchtwertschätzung routinemäßig bei der Rasse Holstein eingeführt. Im Jahr 2011 kam die genomische Zuchtwertschätzung für Braunvieh und Fleckvieh in den Praxiseinsatz. Durch die offizielle Einführung des genomischen Zuchtwertes, liegen die Zuchtwerte von Jungstieren vor, deren Sicherheiten bereits einen breiten Besamungseinsatz von ungeprüften Vererbern rechtfertigen. Die vom EU-Recht definierte Mindestsicherheit von 50% beim Milchwert um mehr als 2.000 Samenportionen eines Stieres vermarkten zu dürfen,

wird beim Fleckvieh und bei Holstein mittels genomisch optimierten Zuchtwerts ohne Probleme überschritten.

Derzeit wird die genomische Zuchtwertschätzung parallel zur herkömmlichen Zuchtwertschätzung durchgeführt. Zur Kalibrierung der genomischen Zuchtwertschätzung ist eine große Anzahl von konventionell sicher geprüften Altstieren notwendig. Aktuell tragen beim Fleckvieh je nach Merkmal zwischen 5.000 und 8.000 Tieren zur Ableitung der Schätzformel bei.

Die Teilnahme am genomischen Zuchtwert steht grundsätzlich allen Züchtern offen. Der Zuchtverband des jeweiligen Bundeslandes ist für die Abwicklung der genomischen Selektion in seinem Gebiet verantwortlich. Der Züchter bekommt das Antragsformular für die genomische Zuchtwertschätzung vom Zuchtverband, sowie das Röhrchen mit Barcode für die Probenahme. Jedes Antragsformular ist vollständig (mit Tag der Probenahme, Unterschriften von Tierarzt und Züchter) auszufüllen und der Barcode des Antragsformulars nach der Blutentnahme auf das Proberöhrchen zu kleben. Der gesamte Ablauf dauert circa 6 Wochen und besteht aus DNA-Extraktion, Genotypisierung, Validierung, genomischer Zuchtwertschätzung und Ergebnisbereitstellung:

- **DNA-Extraktion:** Für die Bestimmung der DNA eines Tieres ist eine Gewebeprobe notwendig, wobei sich Blut bislang am besten bewährt hat. Allerdings ist Blut bei Zwillingen wegen des Blut-Chimärismus nicht geeignet. Als Alternative werden derzeit Ohrstanzproben verwendet.
- **Genotypisierung:** Die Blutprobe wird an das Austrian Institute of Technology (AIT) in Tulln gesendet. Die erfolgreich extrahierte DNA wird zur Genotypisierung zur GeneControl nach München gesendet, welche die Informationen der Genomdatenbank im Rahmen des Rinderdatenverbundes übermittelt (RDV).
- **Validierung:** Die Zuchtwertrechenstelle in Grub vergleicht die Genotypen mit denen verwandter Tiere um Abstammungsfehler oder Vertauschungen der Proben zu ermitteln.

- **Genomische Zuchtwertschätzung:** Für die validierten Genotypen werden von den Rechenstellen genomische Zuchtwerte geschätzt.
- **Ergebnisbereitstellung und Kosten:** In der inoffiziellen Phase werden die Ergebnisse dem Zuchtverband vorgelegt, wobei diese Informationen den Ahnenindex, den direkt genomischen Zuchtwert und den genomisch optimierten Zuchtwert enthalten. In der offiziellen Phase sind sie dann auch in der offiziellen Zuchtwertschätzung enthalten. Die Abrechnung erfolgt über den Zuchtverband, wobei sich die Nettokosten ohne Blutabnahme auf circa € 140 belaufen.

Der genomische Zuchtwert wird monatlich geschätzt, wobei festzuhalten ist, dass die genomische Zuchtwertschätzung weder die bisherige Zuchtwertschätzung noch die Leistungsprüfung ersetzt, sondern nur ergänzt (Fürst 2011).

3 Material

Das Kapitel Material beschreibt die Daten, welche zur Berechnung und Bewertung von den verschiedenen SNP-Chips zur Verfügung standen. Die Herkunft der Tiere und die Verteilung der Tiere in Referenz- und Testgruppe werden hier erläutert.

3.1 Beschreibung der Rasse Fleckvieh

Die Rasse Fleckvieh, auch Simmentaler oder Simmentaler Fleckvieh genannt, ist auf allen fünf Kontinenten vertreten und zählt weltweit zu den bedeutendsten Rinderpopulationen.

Um etwa 1890 begannen in Österreich die ersten zielstrebigen züchterischen Tätigkeiten mit Hilfe von importierten Zuchttieren aus der Schweiz. Der erste Fleckviehzuchtverband wurde 1894 gegründet, und 1895 wurde ein weiterer Meilenstein mit der Einführung einer offiziellen Milchleistungskontrolle gelegt

(Raganitsch 2001). Mittlerweile ist in Österreich das Fleckvieh mit einem Anteil von 77,6% (1,6 Millionen Tiere) am Gesamtrinderbestand die dominierende Rinderrasse (ZAR 2011).

Die Färbung variiert von hellgelb bis dunkelbraun gedeckt, gescheckt, gefleckt oder gesprenkelt, wobei in Klimaten mit hoher Sonneneinstrahlung eine dunklere Färbung günstiger ist. Die Züchtungsschwerpunkte liegen in der Milch- und Fleischleistung, wobei die Gewichtung dieser Eigenschaften je nach Nutzungsschwerpunkt schwankt (Raganitsch 2001).

Im Jahr 2011 waren in Österreich 277.579 Herdebuchkühe registriert. Die Milchleistungen aller Laktationen lagen im Durchschnitt bei 6.840 kg Milch mit 4,15% Fett und 3,41% Eiweiß.

Fleckvieh aus Österreich ist auch international sehr beliebt. Im Jahr 2010 wurden über 27.835 Fleckviehrinder exportiert (ZAR 2011).

3.2 Genotypen

Die Zuchtdata EDV-Dienstleistungs Ges.m.b.H, welche für die genetische Evaluierung von österreichischen Rinderrassen verantwortlich ist, stellte für diese Arbeit Daten von 2.123 Fleckviehtieren zur Verfügung. Es wurden ausschließlich Daten von Stieren in österreichischen Besitz und mit Zuchtwertsicherheiten von über 40% verwendet. Diese Stiere wurden mit dem Illumina Bovine SNP50 Bead Chip für 54.001 SNPs genotypisiert. Der Qualitätscheck wurde bereits von der Zuchtdata durchgeführt, dabei wurden SNPs ausgeschieden, die nicht wenigstens eines von den folgenden Kriterien erfüllten: Minimale Call Rate für die Marker musste 0,95 sein, kleinste Allelfrequenz von 0,02 durfte nicht unterschritten werden, außerdem wurde das Vorhandensein des Hardy-Weinberg-Equilibrium kontrolliert. Nach dem Qualitätscheck blieben 41.082 SNPs über, die für weitere Berechnungen verwendbar waren.

3.3 Phänotypen

Von der Zuchtdata wurden geschätzte Zuchtwerte für die Merkmale Ausschlachtung, Eiweißprozent, Fettprozent, Fettkilogramm, Nutzungsdauer, weibliche Fruchtbarkeit und somatische Zellzahl als Phänotypen zur Verfügung gestellt. Bei der Auswahl der Merkmale wurde darauf geachtet, dass die Möglichkeit besteht unterschiedliche Heritabilitäten zu untersuchen.

Die Merkmale Fettprozent (45% Heritabilität), Fettkilogramm (29% Heritabilität), Eiweißprozent (55% Heritabilität), Ausschlachtung (45% Heritabilität) fallen unter die Kategorie der hoch heritablen Merkmale und Nutzungsdauer (12% Heritabilität), Zellzahl (12% Heritabilität) und weibliche Fruchtbarkeit (2% Heritabilität) sind nieder heritabel (Fürst 2011).

Diese Daten stammen aus der Routine-Zuchtwertschätzung von August 2011, wobei zu den direkt geschätzten Zuchtwerten immer die entsprechenden Zuverlässigkeiten (R^2) angegeben wurden und nur jene Stiere die eine Mindestsicherheit von 40% aufweisen, verwendet wurden. Außerdem wurden deregressierte Zuchtwerte und die korrespondierenden Zuverlässigkeiten zur Verfügung gestellt. Deregressierte Zuchtwerte sind nur von den entsprechenden Töchterleistungen abgeleitet.

3.3.1 Ausschlachtung

Die Ausschlachtung hat eine Heritabilität von 45% und ist somit ein hoch heritables Merkmal. Die Fleischmerkmale werden in Österreich seit 1995 mit einer multivariaten BLUP-Tiermodell Zuchtwertschätzung geschätzt. Die Ausschlachtung berechnet sich indem man das Schlachtkörpergewicht (warm) durch das Lebendgewicht dividiert. Die Daten für die Zuchtwertschätzung werden durch eine un gelenkte Feldprüfung (Schlachthofdaten) ermittelt. Die zu prüfenden Teststiere werden mehr oder weniger nach dem Zufallsprinzip in Zuchtbetrieben für Testpaarungen eingesetzt. Dadurch fallen ausreichend Schlachtdaten von Maststieren, Ochsen (nur Grauvieh) bzw. Mastkälbern (nur Pinzgauer und Grauvieh) an. Die erhobenen Daten werden von der ÖFK (Österreichische Fleischkontrolle GmbH) an die ZuchtData zur Auswertung weitergeleitet. Bei dieser Prüfmethode sind bei geringem Aufwand (Organisation der Sammlung der Schlachthofdaten) hohe Genauigkeiten der Zuchtwertschätzung zu

erzielen. Der Nachteil, der infolge hoher Umweltvarianzen verminderten Heritabilitäten der Merkmale entsteht, kann durch große Nachkommengruppen ausgeglichen werden (Fürst 2011).

3.3.2 Eiweißprozent, Fettprozent, Fettkilogramm

Das Merkmal Fettprozent hat eine Heritabilität von ungefähr 45%, das Merkmal Fettkilogramm von ungefähr 29 % und das Merkmal Eiweißprozent ist zu circa 55% heritabel. Die Fettmenge und Eiweißmenge fließen bei der Zuchtwertschätzung in den Milchwert hinein, das bedeutet der Milchwert errechnet sich aus den Zuchtwerten für Fett-kg und Eiweiß-kg. Seit 2002 werden die Zuchtwerte gemeinsam mit Deutschland geschätzt. Die Umsetzung der gemeinsamen Zuchtwertschätzung mit der Einführung des Testtagsmodells in der Milch-Zuchtwertschätzung ist als Meilenstein in der Geschichte der Zuchtwertschätzung anzusehen. Als wirtschaftliche Gewichte wird ein Verhältnis von 1 : 10 (Fleckvieh, Braunvieh) bzw. 1 : 4 (alle anderen Rassen) für Fettprozent : Eiweißprozent verwendet. Die ökonomischen Gewichte entsprechen dem Grenznutzen für ein zusätzlich erzeugtes Kilogramm Fett bzw. Eiweiß (Fürst 2011).

3.3.3 Fruchtbarkeit

Die Fruchtbarkeit wird seit 1995 in Österreich geschätzt. Die Zuchtwertschätzung wird mit einem Mehrmerkmals-BLUP-Tiermodell durchgeführt. Das bedeutet, dass die 5 Fruchtbarkeitsmerkmale (Non-Return-Rate - Kalbin, Non-Return-Rate-Kuh, Rastzeit, Verzögerungszeit - Kalbin, Verzögerungszeit - Kuh) unter Berücksichtigung der genetischen Beziehungen gemeinsam (multivariat) geschätzt werden. Die Bedeutung der Fruchtbarkeit wird dann sehr stark ins Bewusstsein gerückt, wenn einerseits die Besamungskosten dadurch stark ansteigen und andererseits die Zwischenkalbezeiten so hoch werden, dass die geregelte Remontierung der weiblichen Nachzucht in Frage gestellt wird. Allerdings ist bei Fruchtbarkeit mit einer Heritabilität von etwa 2% die Bedeutung der Veranlagung gering. Anders ausgedrückt heißt das, dass die züchterischen Möglichkeiten zur Verbesserung der

Fruchtbarkeit begrenzt sind, hier ist die Optimierung der Umwelt wesentlich erfolgversprechender. Generell gilt, dass bei einer sehr hohen Heritabilität wenige Tiere bzw. Leistungsinformationen genügen, um ausreichend zuverlässige Zuchtwerte schätzen zu können. So erzielt man zum Beispiel bei der Milchleistung mit ungefähr 20 Töchtern schon eine Sicherheit von 60%, wogegen bei der Fruchtbarkeit ca. 300 Belegungen benötigt, um auf dieselbe Zuverlässigkeit zu kommen. Obwohl Fruchtbarkeitsmerkmale eine niedrige Heritabilität aufweisen, kann durch eine lang andauernde Selektion auf ein negativ korreliertes Merkmal (z. B. die Milchleistung) ein entsprechend unerwünschter Effekt eintreten. Die Behebung dieses unerwünschten Effektes auf genetischem Weg dauert wegen der niedrigen Heritabilität allerdings ebenfalls sehr lange (Fürst 2011).

3.3.4 Nutzungsdauer

Die Nutzungsdauer wird in Österreich seit 1995 geschätzt und ist ungefähr zu 12% heritabel. Für die Zuchtwertschätzung der Nutzungsdauer wird das international bestmögliche Programmpaket verwendet, das sogenannte ‚Survival Kit‘ von Dr. Vincent Ducrocq (INRA, Frankreich) und Univ.-Prof. Dr. Hans Sölkner (BOKU Wien). Das Merkmal Nutzungsdauer wird in der Zuchtwertschätzung durch das sogenannte Abgangsrisiko ausgedrückt. Dieses stellt das Risiko dar, dass eine Kuh vom Betrieb abgeht bzw. ausfällt. Das Abgangsrisiko verhält sich zur Nutzungsdauer gegenläufig. Das heißt, ein höheres Abgangsrisiko bedeutet eine niedrigere Nutzungsdauer. Die Nutzungsdauer ist nach der Milchleistung das wohl wirtschaftlich bedeutsamste Merkmal in der Milchviehhaltung. Kühe mit langer Nutzungsdauer müssen sowohl fruchtbar als auch von guter Konstitution (z.B. Fundament, Eutersitz) sein. Die tatsächliche Nutzungsdauer hängt allerdings ganz entscheidend von der Milchleistung einer Kuh ab, weil einerseits Kühe mit schlechter Leistung früher gemerzt werden, andererseits der Bauer Kühen mit besonders hoher Milchleistung eine Sonderbehandlung zukommen lässt (z.B. bei der Anzahl der Besamungen). Bei langer Nutzungsdauer wünschenswerte Nebeneffekte sind die volle Ausnützung des altersbedingten Leistungsmaximums, die Reduzierung der anteiligen Aufzuchtkosten und die höhere innerbetriebliche Selektionsschärfe (Fürst 2011).

3.3.5 Zellzahl

Die Zellzahl hat eine Heritabilität von 12% und wird seit 1998 in Österreich geschätzt. In der Zuchtwertschätzung werden die seit 1990 angefallenen Zellzahlergebnisse vom 8. bis zum 312. Laktationstag der Laktationen 1 bis 3 berücksichtigt und mittels eines Mehrmerkmals-Testtags-Tiermodell (Fixed Regression) durchgeführt. Die Verteilung der Zellzahlen in der Praxis entspricht nicht einer Normalverteilung, wie sie bei der Zuchtwertschätzung angenommen wird. Um die gewünschte Form der Verteilung zumindest näherungsweise zu erreichen, erfolgt vor der Zuchtwertschätzung eine logarithmische Transformation der Zellzahlen zum Linear Somatic Cell Score (SCS). Um dem Anwender der Zuchtwerte übersichtliche und verständliche Maßzahlen für die Selektion der Tiere in die Hand zu geben, werden die Zuchtwerte über die Laktationstage 8 bis 312 zum vertrauten 305-Tage-Zuchtwert aufsummiert. Bei dem für die Zellzahl veröffentlichten Zuchtwert handelt es sich um den Durchschnittswert der für die drei Laktationen geschätzten Einzelzuchtwerte. (Fürst 2011).

3.4 Referenz- und Validierungsgruppe

Um die Genauigkeit von genetischen Werten zu analysieren, wurden die Fleckviehstiere in zwei Gruppen geteilt und entweder der Referenzgruppe (Trainingsgruppe) oder der Testgruppe (Validierungsgruppe) zugewiesen. Die Tiere wurden nach Geburtsjahrgängen getrennt, wobei die Referenzgruppe alle Stiere enthält, die vor und im Jahr 2003 geboren wurden. Die Testgruppe umfasst alle Stiere, die nach dem Jahr 2003 geboren wurden. In Abbildung 1 ist zu erkennen, dass die Verteilung der Stiere der Referenzgruppe 1.637 Stiere und der Validierungsgruppe 486 Stiere zuspricht. Die Strategie, die jüngsten Tiere in das Test-Set zu selektieren, war die gemeinsame Schätzung von genomisch geschätzten Zuchtwerten (GEBVs) von Söhnen mit deren Vätern zu verhindern.

Dies würde zu einer Erhöhung der Genauigkeit führen, jedoch wäre die Erhöhung nur ein Effekt der Pedigreestruktur.

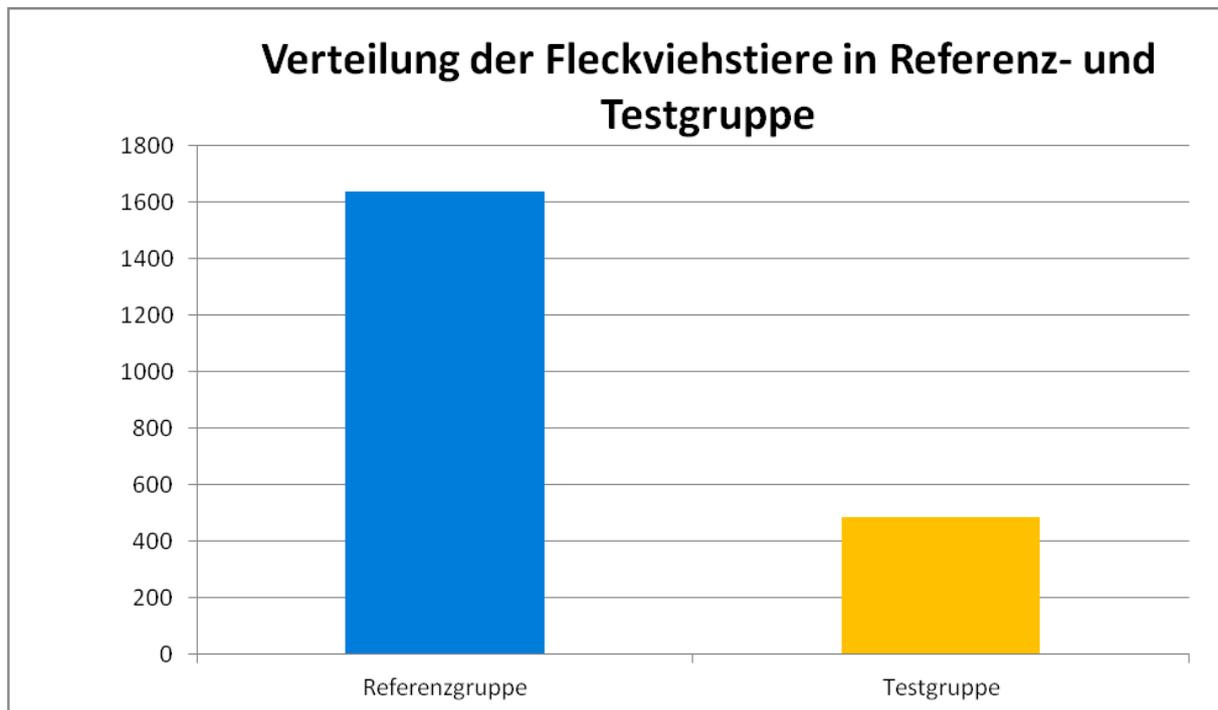


Abbildung 1: Verteilung der Fleckviehstiere in Referenz- und Testgruppe

Abbildung 2 zeigt die Verteilung der Fleckviehstiere nach Geburtsjahrgängen. Die verwendeten Stiere sind innerhalb der Jahre 1969 und 2006 geboren. Hier ist ersichtlich, dass zwischen 1969 und 1989 nur eine kleine Anzahl an Tieren im Verhältnis zu der Geburtengruppe zwischen 1990 und 2002 geboren ist. Außerdem ist nochmals das Verhältnis von der Referenzgruppe zur Testgruppe ersichtlich.

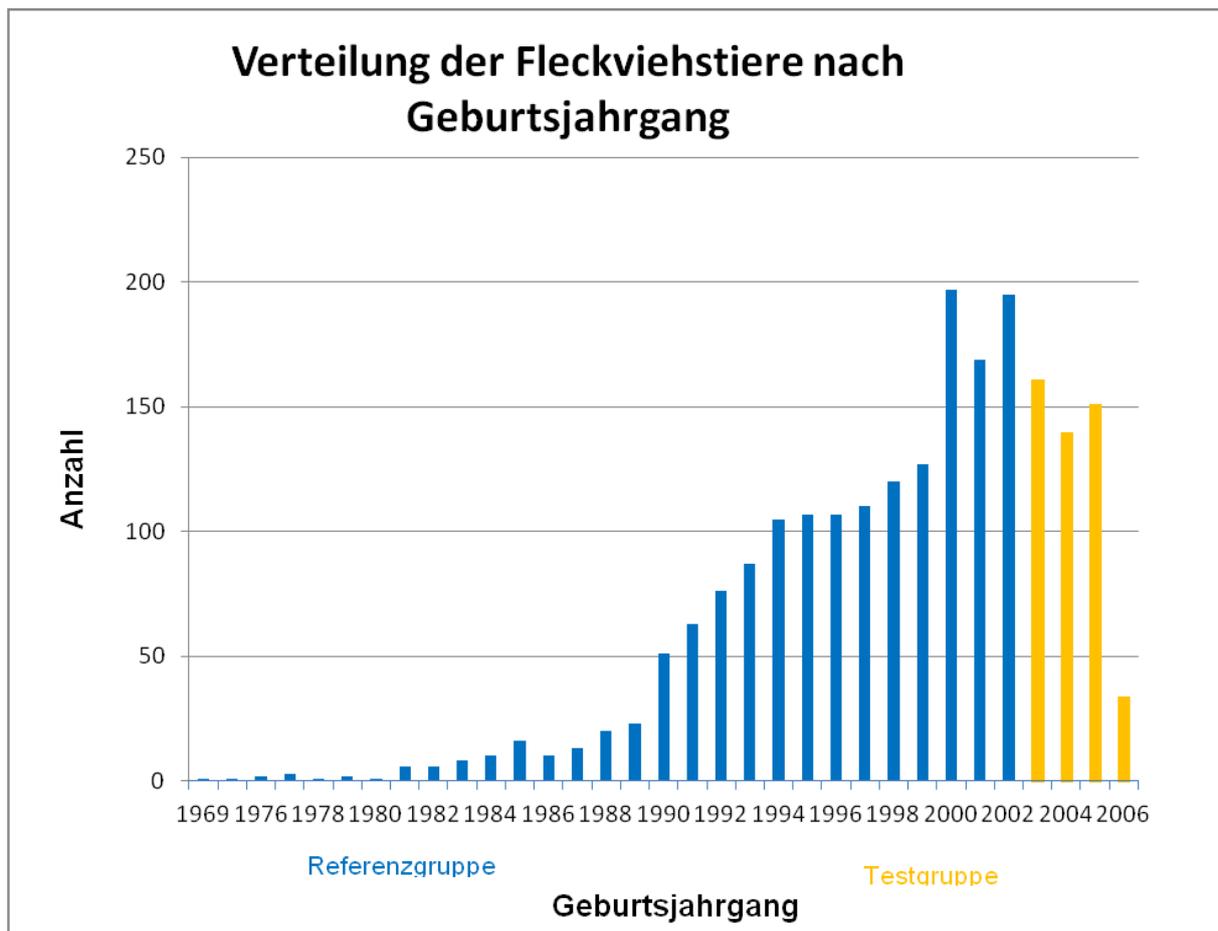


Abbildung 2: Verteilung der Fleckviehstiere nach Geburtsjahrgängen

Tabelle 1 stellt dar, wie viele Zuchtwerte für das jeweilige Merkmal zur Verfügung standen und differenziert außerdem zwischen Referenz- und Testgruppe. Zusätzlich gibt die Tabelle noch Mittelwert, Standardabweichung, Minimum und Maximum für das jeweilige Merkmal an.

In Tabelle 2 ist dasselbe dargestellt wie in Tabelle 1, allerdings für die deregressierten Zuchtwerte.

Tabelle 1. Anzahl (N), Mittelwert (mean), Standardabweichung (Std.dev.), Minimum (Min), und Maximum (Max) von ausgewählten Zuchtwerten

Merkmal	Referenz-Datenset					Test-Datenset				
	N	Mean	Std.dev.	Min	Max	N	Mean	Std.dev.	Min	Max
Ausschlachtung	1637	101,00	9,11	70,00	128,00	486	101,63	9,14	73,00	130,00
Eiweißprozent	1637	0,04	0,13	-0,38	0,52	479	0,00	0,11	-0,35	0,32
Fettkilogramm	1637	-14,08	21,40	-96,00	53,00	479	7,23	15,19	-35,00	55,00
Fettprozent	1637	0,44	0,22	-0,62	0,88	479	-0,03	0,20	-0,56	0,57
Nutzungsdauer	1637	98,37	10,93	62,00	133,00	485	104,06	8,30	71,00	129,00
weibl.Fruchtbarkeit	1637	100,78	6,92	77,00	122,00	485	99,42	6,08	79,00	117,00
Zellzahl	1637	99,97	10,90	59,00	131,00	484	101,64	9,69	69,00	129,00

Tabelle 2. Anzahl (N), Mittelwert (mean), Standardabweichung (Std.dev.), Minimum (Min), Maximum (Max) von ausgewählten deregressierten Zuchtwerten

Merkmal	Referenz-Datenset					Test-Datenset				
	N	Mean	Std.dev.	Min	Max	N	Mean	Std.dev.	Min	Max
Ausschlachtung	1629	101,01	11,39	41,47	141,19	480	101,62	11,60	68,40	134,16
Eiweißprozent	1623	0,04	0,14	-0,41	0,59	479	0,01	0,13	-0,40	0,38
Fettprozent	1623	-14,05	22,05	81,97	53,34	479	6,67	17,57	-44,93	58,12
Fettkilogramm	1623	0,05	0,24	-0,67	0,95	479	-0,02	0,24	-0,63	0,72
Nutzungsdauer	1616	98,59	15,57	42,07	169,66	348	107,32	12,27	38,58	138,00
weibl.Fruchtbarkeit	1636	101,59	10,72	64,06	142,88	373	100,80	15,36	43,98	142,40
Zellzahl	1636	99,84	12,11	56,07	135,97	482	101,74	12,62	47,16	136,56

4 Methoden

Das folgende Kapitel beschreibt die verwendeten Methoden zur Analysierung der Daten und Berechnung der Korrelationen zwischen EBVs (geschätzter Zuchtwert, aufgrund von Phänotypen) und GEBVs (genomisch geschätzter Zuchtwert).

4.1 SNP Selektion

Die Zuchtdata GMBH stellte die Daten von 2.123 Stieren zur Verfügung, welche mit dem 50K Chip genotypisiert wurden. Nach dem von der Zuchtdata durchgeführten Qualitätscheck blieben von den ursprünglich enthaltenen 54.001 SNPs 41.082 über. Die Tiere wurden nicht mit dem 3K und 7K Chip genotypisiert und deshalb wurden für die Auswertungen, die in den BovineLD Chips enthaltenen SNPs von dem 50K Chip selektiert. Der 50K Chip enthielt nach dem Qualitätscheck 6.565 SNPs, welche im 7K Chip vorhanden sind und für die Berechnungen zur Verfügung standen. Ursprünglich enthält der 7K Chip 6.909 SNPs, für die fehlenden 344 SNPs gab es keinen Daten, und somit schieden diese aus den Berechnungen aus. Der 3K Chip enthält ursprünglich 2.900 SNPs, und für die Berechnungen in dieser Arbeit konnten Daten von 2.890 SNPs aus dem 50K Chip selektiert werden.

	3K Chip 2.890 SNPs	7K Chip 6.565 SNPs	50K Chip 41.082 SNPs
3K Chip	-	800	0
7K Chip	4.476	-	0
50K Chip	38.193	34.517	-

Tabelle 3 zeigt, wie viele SNPs die verschiedenen Chips enthalten und die Unterschiede zwischen den Chips. Ersichtlich ist zum Beispiel, dass der in der Arbeit verwendete 3K Chip 800 SNPs enthält, die in dem 7K Chip nicht vorkommen. Der 7K Chip hingegen enthält 4.476 SNPs, die im 3K Chip nicht enthalten sind.

Da aus dem 50K Chip die SNPs für den 3K und 7K Chip extrahiert wurden, enthalten die LD Chips keine SNPs, die nicht im 50K Chip vorkommen.

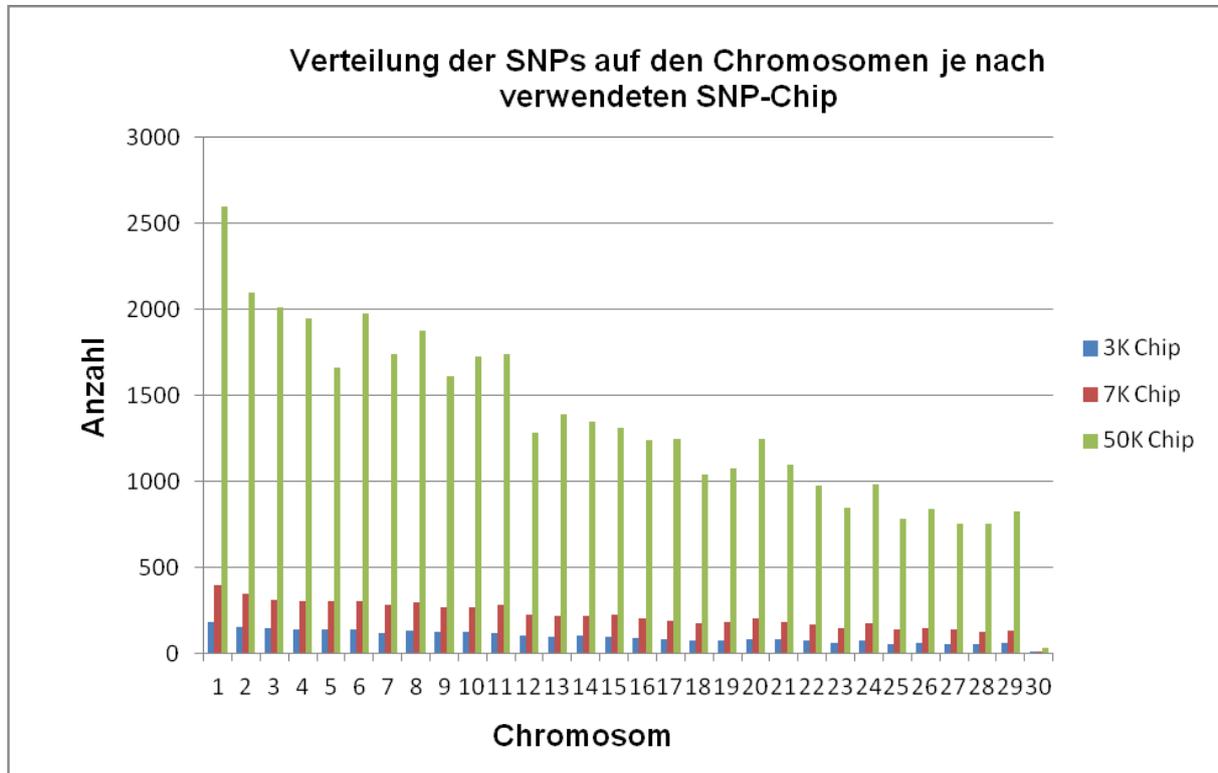


Abbildung 3: Verteilung der SNPs (nach Anzahl) auf den Chromosomen je nach verwendetem SNP-Chip

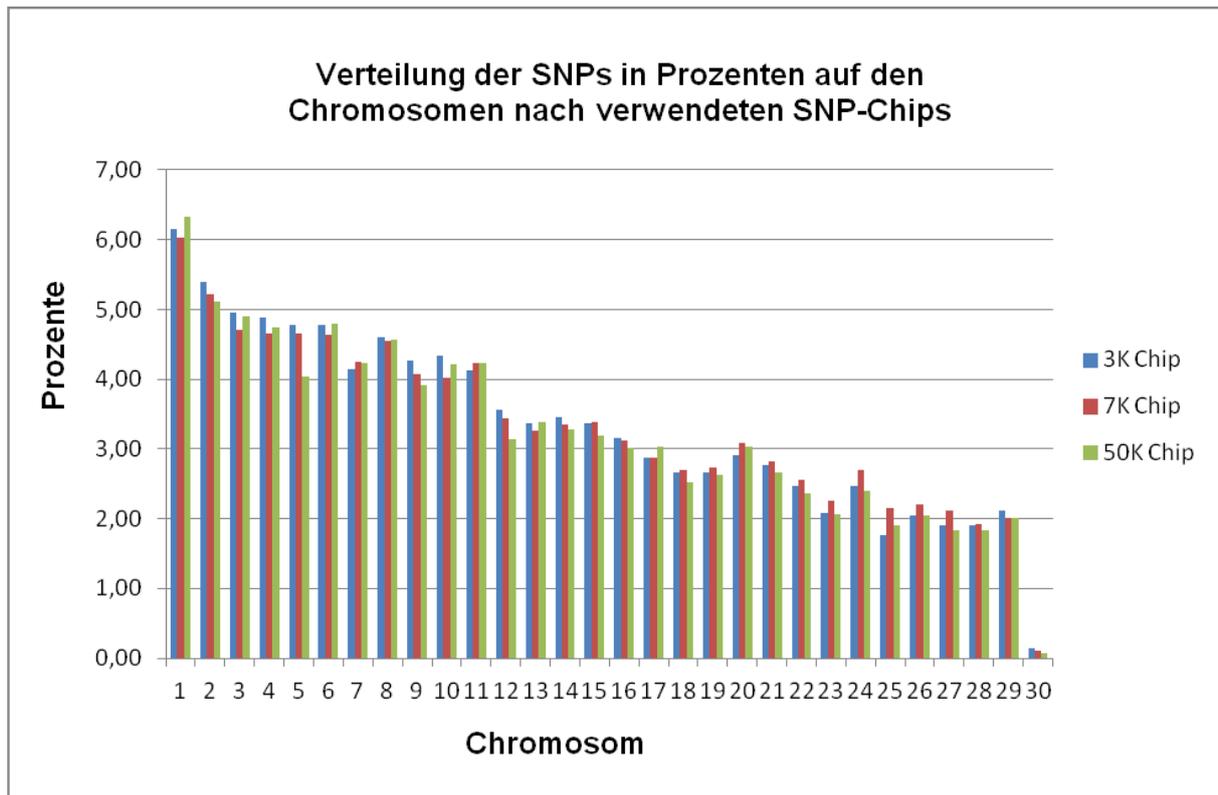


Abbildung 4: Verteilung der SNPs (in Prozenten) auf den Chromosomen nach verwendeten SNP-Chips

Abbildung 3 zeigt die Verteilung der SNPs nach Anzahl auf den jeweiligen Chromosomen. Wobei man in Abbildung 3 den Mengenunterschied zwischen den verschiedenen Chips eindeutig erkennen kann. In den verwendeten Chips sind die SNPs zwischen den Chromosomen 1 und 30 verteilt.

In Abbildung 4 sieht man, die prozentmäßig gleiche Verteilung der SNPs von den verschiedenen Chips auf den Chromosomen. Trotz der unterschiedlichen Chipgrößen ist die Verteilung in Prozenten annähernd ident.

4.2 Methoden zur genomischen Zuchtwertschätzung

Um die genomischen Zuchtwerte zu schätzen wurden zwei verschiedene Ansätze verwendet: BayesB und GBLUP. Die verschiedenen Methoden wurden verwendet um die berechneten Werte vergleichen und kontrollieren zu können.

4.2.1 BayesB

Die BayesB (Bayes-BLUP) Methode zur Schätzung von Zuchtwerten und SNP Effekten wurde von Meuwissen et al. (2009) entwickelt. Das statistische Model mit dem die Zuchtwerte und SNP-Effekte in dieser Arbeit geschätzt wurden, folgt dem Ansatz nach Meuwissen (2009):

$$y = \mu + \sum_{j=1}^{NM} X_j a_j + e$$

- y ist ein $N \times 1$ Vektor von Phänotypen
- μ ist der Mittelwert
- a_j ist der Effekt von Marker j
- X_j ist ein $N \times 1$ Vektor, welcher den Genotyp der Individuen von Marker j angibt, wobei 0 bedeutet, dass das erste Allele homozygot ist
 - $1/\sqrt{H_j}$ bedeutet heterozygot
 - $2/\sqrt{H_j}$ bedeutet homozygot für das zweite Allel (wobei H_j der Marker für Heterozygotie ist)
Der $1/\sqrt{H_j}$ Term in X_j standardisiert die Varianz der Markergenotypdaten auf 1.
- e ist die Standardabweichung

Die Varianz von a_j wurde mittels BayesB geschätzt (Meuwissen 2009). BayesB ist ein genomisch nichtlineares Variablen Auswahlverfahren (Daetwyler et al. 2010).

Bayesgg ist eine Geschwindigkeit verbesserte Version des BayesB. Wobei im Bayesgg angenommen wird, dass die SNPs, welche in BayesB ohne Wirkung sind hier eine geringe Wirkung haben. Die Varianz dieser kleinen Effekte wird abgeschätzt, und folglich werden die kleinen Effekte wie in einem BLUP-Modell modelliert. Einigen SNPs wird wie im richtigen BayesB erlaubt eine große Wirkung zu haben. Diese Software benötigt das Linux-System (Meuwissen 2009a).

4.2.2 GBLUP

Die grundlegende GBLUP Methode, welche benutzt wird um direkt genomische Zuchtwerte zu schätzen ist:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

- \mathbf{y} ist der Datenvektor der analysierten Leistungen (bzw SNPs) der genotypisierten Bullen
- μ ist der Mittelwert
- $\mathbf{1}$ ist der Vektor von Einsen
- \mathbf{Z} ist die Designmatrix, die die Leistungen (SNPs) den Zuchtwerten zuordnet
- \mathbf{g} ist der Vektor der DGV, welche geschätzt werden
- \mathbf{e} ist die Standardabweichung (Gao et al. 2012)

GBLUP ist eine genomisch lineare Methode um Zuchtwerte zu schätzen (Daetwyler et al. 2010). Bei der BayesB Methode wird ein Bayesian Mixture Modell verwendet, bei dem nur eine gewisse Anzahl an SNPs als wichtig angesehen wird, hingegen werden bei der GBLUP- Methode alle SNPs als gleich wichtig angesehen. In dieser Arbeit wurde mit unterschiedlichen Anteilen an wichtigen SNPs bei BayesB experimentiert, aber die Ergebnisse waren nahezu ident ungeachtet des Phänotyps. Aufgrund dessen werden nur die Ergebnisse mit 10% wichtigen SNPs in dieser Arbeit dargestellt.

Die Analyse für BayesB und GBLUP wurde mittels Verwendung des Bayesgg Programms durchgeführt, welches von Prof. Theo Meuwissen entwickelt wurde. Die Analyse mit Bayesgg rannte über das Linux-System auf den 4 NONA-Servern der BOKU. Analysiert wurden der 3K, der 7K und der 50K Chip mit den Werten der direkten und deregressierten Zuchtwerte. Die imputierten Werte wurden nur mit der GBLUP Methode berechnet.

4.3 Genauigkeit und Kalkulation der genomisch geschätzten Zuchtwerte

In Abbildung 1 ist die Verteilung der Stiere in eine Referenz- und Testgruppe ersichtlich. Die Referenzgruppe enthält 77,11% und die Validierungsgruppe enthält 22,89% der Stiere. Der Korrelationskoeffizient zwischen genomisch geschätztem Zuchtwert und realisierten Zuchtwert beziehungsweise deregressiertem Zuchtwert, wurde als Maß für die Genauigkeit der Vorhersage von genomisch geschätzten Zuchtwerten verwendet.

Die Vorhersagegleichungen für jedes Merkmal wurden aus dem Referenzset abgeleitet und dann mit den Genotypdaten kombiniert um die DGV für die Testtiere vorherzusagen, wobei dies auf der folgenden Formel basiert:

$$\mathbf{DGV} = \mathbf{X}\hat{\mathbf{b}}$$

- DGV ist der Vektor der direkt genomischen Werte, welche mit den Markergenotypen geschätzt wurden
- X ist eine Designmatrix, die den Genotyp des SNP auf alle Individuen bezieht
- $\hat{\mathbf{b}}$ ist der Vektor der geschätzten SNP Effekte (Moser et al. 2010)

4.4 Imputationsmethode

In diesem Projekt wurden AlphaImpute als Imputationsansatz verwendet.

AlphaImpute ist ein von John Hickey geschriebenes Softwarepaket, welches der Imputation und der Synchronisierung von Genotypdaten in Populationen mit Stammbaum Informationen dient (Hickey et al. 2011). AlphaImpute verwendet sowohl das Linkage Disequilibrium, als auch die Stammbauminformationen um Allele und Genotypen richtig zu imputieren und sollte gut geeignet sein um Imputationen in Milchviehpopulationen durchzuführen (Gredler et al. 2011).

AlphaImpute besteht aus einem einzelnen Programm, welches entweder AlphaPhase1.1 oder GeneProbforAlphaImpute genannt wird. Alle Informationen angefangen vom Analysenmodell zu den Inputdateien bzw. deren Layout, sind in einer einzigen Parameterdatei spezifiziert. AlphaImpute ist in Linux und Mac OSX ausführbar (Hickey et al. 2011).

5 Ergebnisse und Diskussion

Ziel dieser Arbeit war es zu untersuchen, wie sicher genomische Zuchtwertschätzung mit einem Bovine LD Bead Chip (7K) durchgeführt werden kann, wenn die fehlenden rund 47.000 SNPs imputiert werden. Dafür sind die Korrelationen zwischen EBVs und GEBVs mit BayesB und GBLUP für den 50K, 7K und 3K Chip berechnet worden. Die Korrelationswerte der imputierten Chips wurden für Analysen mittels GBLUP berechnet. Als Phänotypen standen direkte und deregressierte Zuchtwerte zur Verfügung. Es wurde untersucht, wie groß der Unterschied in der Sicherheit der Zuchtwertschätzung zwischen dem imputierten 3K Chip (3Ki), dem imputierten 7K (7Ki) Chip und dem originalen 50K Chip ist um die durch Imputation erreichte Verbesserung beurteilen zu können.

5.1 Ergebnisse der genomischen Zuchtwertschätzung mit direkten Zuchtwerten unter Verwendung verschiedener SNP-Chips

In Tabelle 4 sind die unterschiedlichen Korrelationen der drei SNP-Chips für die direkt geschätzten Zuchtwerte ersichtlich, wobei hier die Ergebnisse für die verschiedenen Merkmale für die Testgruppe zusehen sind. Mit steigender SNP-Dichte steigen die Korrelationen zwischen genomisch geschätzten und direkt beobachteten Zuchtwerten, was man in Tabelle 4 eindeutig erkennen kann.

Die Korrelationen, welche durch die Genotypisierung mittels 3K Chip für die Testgruppe erreicht wurden, variierten zwischen 0,3 – 0,55 für BayesB. Im Vergleich dazu variierten die Korrelationen des 50K Chips zwischen 0,45 – 0,71. Die Werte des 7K Chips (0,44 – 0,67) sind im Vergleich zum 3K Chip besser und liegen nahe an den Werten des 50K Chips. Das bedeutet, dass im Durchschnitt über die ermittelten Merkmale der 3K Chip 75% und der 7K Chip 90% der Genauigkeit des 50K Chips bei Berechnung mit BayesB erzielt haben.

Mit GBLUP variieren die Werte zwischen 0,25 – 0,49 für den 3K Chip und zwischen 0,44 – 0,58 beim 50K Chip. Die Werte des 7K Chips (0,39 – 0,55) liegen auch bei der

Berechnung mit GBLUP sehr nahe an jenen des 50K Chips. GBLUP erzielte im Durchschnitt über die Merkmale gesehen für den 3K Chip 71% für den 7K Chip 88% der Genauigkeit des 50K Chips.

Tabelle 4. Korrelationen zwischen konventionellen EBVs und GEBVs für das Test-Set mit verschiedenen SNP-Panelen

	3K		7K		50K	
	BayesB	GBLUP	BayesB	GBLUP	BayesB	GBLUP
Ausschlachtung	0,36	0,34	0,43	0,43	0,52	0,52
Eiweißprozent	0,40	0,35	0,46	0,41	0,45	0,48
Fettkilogramm	0,30	0,25	0,45	0,39	0,49	0,44
Fettprozent	0,55	0,49	0,67	0,55	0,71	0,56
Nutzungsdauer	0,32	0,34	0,44	0,45	0,50	0,51
weibliche Fruchtbarkeit	0,49	0,43	0,53	0,52	0,58	0,58
Zellzahl	0,43	0,38	0,45	0,44	0,51	0,51

Eine Abstufung der Korrelation zwischen 50K Chip, 7K Chip und 3K Chip ist bei allen ausgewählten Merkmalen der Testgruppe zu erkennen, wobei der 50K Chip die besten Ergebnisse erzielt. Zum Beispiel hat das Merkmal Ausschlachtung (berechnet mit BayesB) den Korrelationswert 0,36 bei einer Genotypisierung mittels 3K Chip, 0,43 bei einer Genotypisierung mit dem 7K Chip und für den 50K Chip wird ein Korrelationswert von 0,52 erreicht. Bei dem Merkmal Eiweißprozent erzielte der 7K Chip berechnet mit BayesB eine höhere Korrelation als der 50K Chip, dies ist allerdings nur bei BayesB der Fall. Bei GBLUP hat der 50K Chip eine bessere Performanz.

Die höchste Korrelation hat das Merkmal Fettprozent mit einem Wert von 0,71 durch Genotypisierung mit dem 50K Chip erreicht. Den niedrigsten Wert, der durch Genotypisierung mit 50K Chip ermittelt wurde, war 0,45 für das Merkmal Eiweißprozent. Die niedrigste Korrelation wurde mit einem Wert von 0,30 durch Genotypisierung mit dem 3K Chip für das Merkmal Fettkilogramm erreicht und der höchste Wert für den 3K Chip ist 0,55 für das Merkmal Fettprozent. Die Werte des 7K Chips liegen denen des 50K Chips sehr nahe (Abbildung 5). Die Abweichungen des 7K Chips für die EBVs liegen zwischen 0,04 – 0,09. Der Unterschied der Referenzgruppe für BayesB zwischen den SNP-Chips ist in Abbildung 5 eindeutig

ersichtlich. Die Werte des 50K Chip sind sehr nahe an 1 (0,99), die des 7K Chip (0,95 – 1,00) liegen knapp darunter und für den 3K Chip variieren die Werte zwischen 0,83 – 0,95.

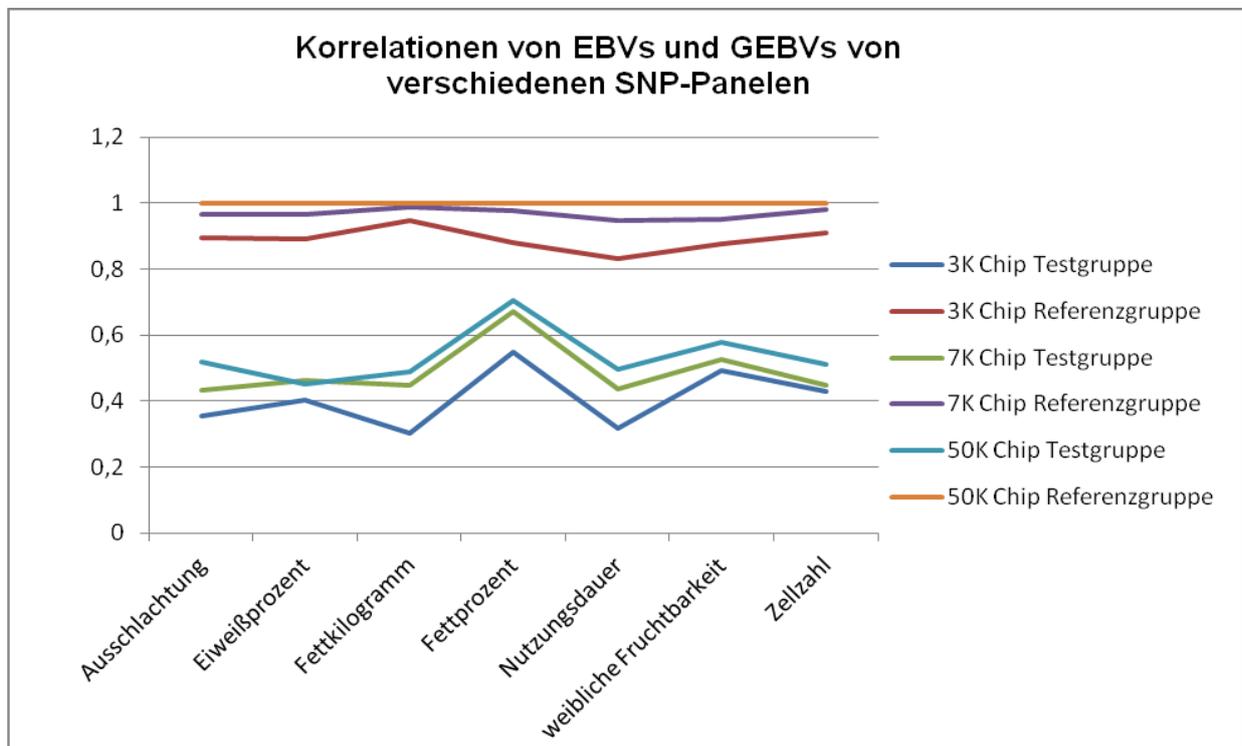


Abbildung 5: Korrelationen von EBVs und GEBVs von verschiedenen SNP Panelen berechnet mit BayesB

Die Werte für die Referenzgruppe berechnet mit GBLUP liegen für den 3K Chip zwischen 0,98 – 1,00, für den 7K Chip und für den 50K Chip nahe 1 (0,99). Die Werte sind in den Tabellen 8-10 dargestellt.

Die Abbildung 6 zeigt, dass die Referenzgruppe für alle Merkmale bei einer Genotypisierung mittels 50K Chip einen Wert nahe 1 erreicht, außerdem stellt die Abbildung graphisch den Unterschied zwischen den mit BayesB und GBLUP berechneten Werten dar. Beim 50K Chip sind die Ergebnisse der Referenzgruppe für BayesB und GBLUP nahezu ident mit Ausnahme des Merkmals Fettprozent und auch bei der Testgruppe sind nur kleine Unterschiede zu erkennen, bei welchem BayesB besser abschneidet. Dies stimmt mit der Studie von Daetwyler et al. (2010) gut überein, in welcher sie zeigten, dass mit zunehmender SNP-Dichte der

Unterschied zwischen den zwei Methoden niedriger wird und eventuell dieselben Werte erzielt werden können.

Bei allen anderen SNP Panelen ist es im Durchschnitt so, dass BayesB die besseren Werte für die Testgruppe aufweist, dies wurde auch in diversen anderen Studien beobachtet (Meuwissen et al. 2001, Hayes et al. 2010, Zhan et al. 2011).

Interessant in dieser Arbeit ist allerdings, dass BayesB eindeutig die schlechteren Korrelationen für die mit 3K genotypisierte Referenzgruppe erzielt.

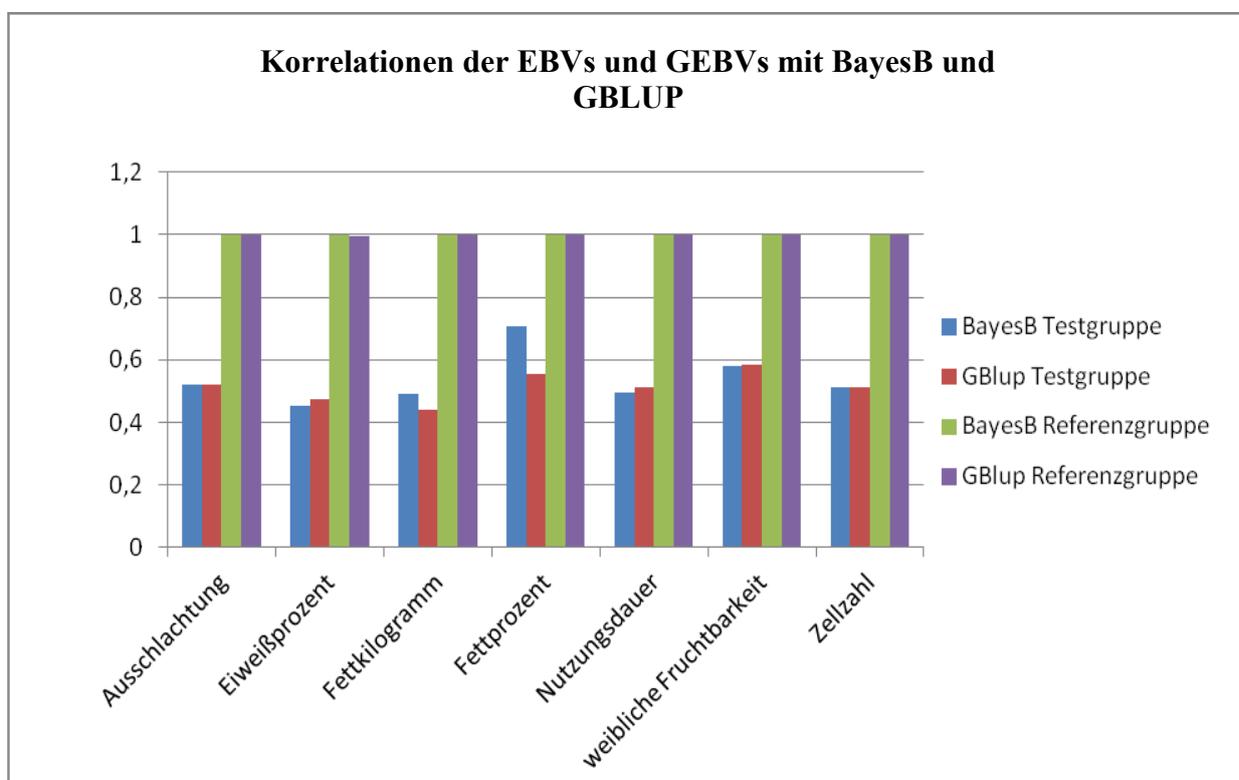


Abbildung 6: Korrelationen der EBVs und GEBVs mit BayesB und GBLUP berechnet für den 50K Chip

Verschiedenste Studien zeigten, dass hohe Genauigkeiten erzielt werden können, wenn SNP Daten verwendet werden um genomische Zuchtwerte zu schätzen. In der Studie von Meuwissen et al. (2001) wurden unter Verwendung von circa 50.000 SNPs, Korrelationen von 0,73 mit BLUP und Genauigkeiten von 0,85 mit BayesB kalkuliert. Calus et al. (2008) beschrieben Genauigkeiten von 0,83 für Merkmale mit

einer Heritabilität von 0,5 und Genauigkeiten von 0,66 für Merkmale mit einer Heritabilität von 0,1.

Korrelationen von Studien, die mit realen anstatt simulierten Daten rechneten, waren niedriger. Zum Beispiel berichteten De Roos et al. (2011) von Korrelationen zwischen EBVs und GEBVs mit Werten um die 0,63 für niedrig heritable Merkmale. Merkmale mit einer höheren Heritabilität kamen auf eine Genauigkeit von 0,76.

Die Korrelationen des 50K Chips sind in dieser Arbeit im Vergleich zu den in der Literatur beschriebenen Werten recht niedrig. Ein Grund könnte die höhere Anzahl an verwendeten Genotypen in den zitierten Studien sein. Van Raden et al. (2008) kamen zu dem Schluss, dass eine Verdopplung der genotypisierten Tiere die Zuverlässigkeit der geschätzten Zuchtwerte mehr erhöht als eine Verdopplung der Marker. Gredler et al. (2009) führte eine Studie mit Fleckvieh in einer Größenordnung vergleichbar mit jener in unserer Arbeit durch. Ihre Ergebnisse für GBLUP waren für die Merkmale Fettprozent, Eiweißprozent, weiblichen Fruchtbarkeit und Zellzahl zwischen 0,42 – 0,47. Die von Gredler et al. (2009) benutzte BayesC Methode erzielte Resultate zwischen 0,43 – 0,60, welche gut mit jenen von in dieser Arbeit berechneten Werte für BayesB (0,45 – 0,58) zusammen passen, abgesehen von einem Ausreißer von 0,71 berechnet für das Merkmal Fettprozent.

5.2 Ergebnisse der genomischen Zuchtwertschätzung mit deregressierten Zuchtwerten mit Verwendung verschiedener SNP-Chips

Tabelle 5 stellt Ergebnisse bei Verwendung von deregressierten Zuchtwerten als Phänotypen. Der Aufbau entspricht Tabelle 4, in der geschätzte Zuchtwerte als Phänotypen verwendet wurden. Bei den deregressierten Zuchtwerten steigt die Korrelation ebenfalls mit zunehmender SNP-Dichte bei der Referenzgruppe. Der mittels BayesB berechnete Korrelationswert (Testgruppe) für das Merkmal Eiweißprozent ist für den 3K Chip 0,36, den 7K Chip 0,41 und für den 50K Chip 0,47.

Die Ergebnisse für GBLUP und BayesB liegen beim 3K Chip zwischen 0,11 – 0,45, beim 7K Chip zwischen 0,18 – 0,63 und beim 50K Chip zwischen 0,15 – 0,64. Die

Werte des 50K Chips entsprechen sehr gut jenen von Van Raden et al. (2008b). Sie berechneten Genauigkeiten zwischen 0,19 – 0,63 für Töchterleistungen mit linearen und nicht linearen Verfahren.

Die Werte des 7K Chips liegen wieder sehr nahe an denen des 50K Chips sehr, wobei die berechneten Abweichungen hier zwischen 0 – 0,6 liegen. Die deregressierten Zuchtwerte haben sehr niedrige Ergebnisse für die Merkmale Nutzungsdauer und weibliche Fruchtbarkeit unabhängig von der verwendeten SNP-Dichte. Dies kann durch die niedrige Heritabilität der Merkmale erklärt werden. Außerdem spielt auch die niedrige Anzahl an Phänotypen die zur Berechnung der Nutzungsdauer zur Verfügung standen bei diesem Merkmal eine Rolle.

Moser et al. (2010) kamen bei ihrer Studie, für welche sie unterschiedliche SNP Subsets verwendeten um Korrelationen zwischen GEBVs und Phänotypen (deregressierte Zuchtwerte) zu kalkulieren, auf Ergebnisse, die mit dieser Studie übereinstimmen. Die Korrelationen für das 50K SNP Panel waren für höher heritable Merkmale genauer mit Werten zwischen 0,52 – 0,64 als für nieder heritable mit Werten zwischen 0,19 – 0,20. Moser et al. (2010) gelangten zu dem Schluss, dass die Heritabilität Einfluss auf die Genauigkeit hat. Außerdem berichteten sie in ihrer Studie, dass man mit der Genotypisierung durch den 3K Chip im Durchschnitt 90% der Genauigkeiten des 50K Chips erlangt, was mit den Daten unserer Studie nicht übereinstimmt. Da der 3K Chip durchschnittlich um die 75% der Genauigkeit des 50K Chips eindeutig niedrigere Werte als in Mosers Studie (2010) angegeben erzielt. Der 7K Chip erlangt allerdings im Durchschnitt über 90% der Genauigkeit des 50K Chips für die deregressierten Zuchtwerte. Dies sind für Korrelationen (EBV-GEBV), die mit deregressierten Phänotypen berechnet wurden sehr gute Resultate. Ein Grund für die niedrigeren Ergebnisse in dieser Arbeit könnte die niedrigere Anzahl an Phänotypen sein, die verwendet wurden.

Tabelle 5. Korrelationen zwischen deregressierten EBVs und GEBVs für das Test-Set mit verschiedenen SNP-Panelen

	3K		7K		50K	
	BayesB	GBLUP	BayesB	GBLUP	BayesB	GBLUP
Ausschlachtung	0,28	0,28	0,43	0,28	0,44	0,25
Eiweißprozent	0,36	0,21	0,41	0,31	0,47	0,40
Fettkilogramm	0,29	0,27	0,44	0,40	0,46	0,43
Fettprozent	0,51	0,45	0,63	0,50	0,64	0,50
Nutzungsdauer	0,11	0,14	0,19	0,18	0,15	0,18
weibliche Fruchtbarkeit	0,18	0,14	0,23	0,23	0,25	0,26
Zellzahl	0,37	0,38	0,39	0,40	0,42	0,45

Die Werte der Referenzgruppe für den 3K Chip sind zwischen 0,79 – 0,89 für BayesB und zwischen 0,93 – 1,00 für GBLUP. Der 7K Chip erzielte Werte zwischen 0,92 – 0,98 für BayesB und zwischen 0,96 – 1,00 für GBLUP (Tabellen 11-13). Die Werte des 50K Chips sind für BayesB als auch für GBLUP nahe 1 und sind in Abbildung 9 graphisch dargestellt. Abbildung 9 zeigt den Unterschied zwischen den mittels BayesB und GBLUP berechneten Korrelationen der deregressierten Zuchtwerte. Der Korrelationswert der Testgruppe ist im Durchschnitt nicht so hoch wie jener Wert der Testgruppe der direkten Zuchtwerte, außerdem ist der Korrelationswert der Referenzgruppe für die Merkmale Zellzahl und weibliche Nutzungsdauer mit einem Wert von 0,97 berechnet mit GBLUP eindeutig niedriger als für die anderen Merkmale. Dieses Ergebnis könnte durch die niedrigere Anzahl an Phänotypen, die für die beiden Merkmale zur Berechnung zur Verfügung standen und durch die niedrige Heritabilität der beiden Merkmale zu Stande gekommen sein.

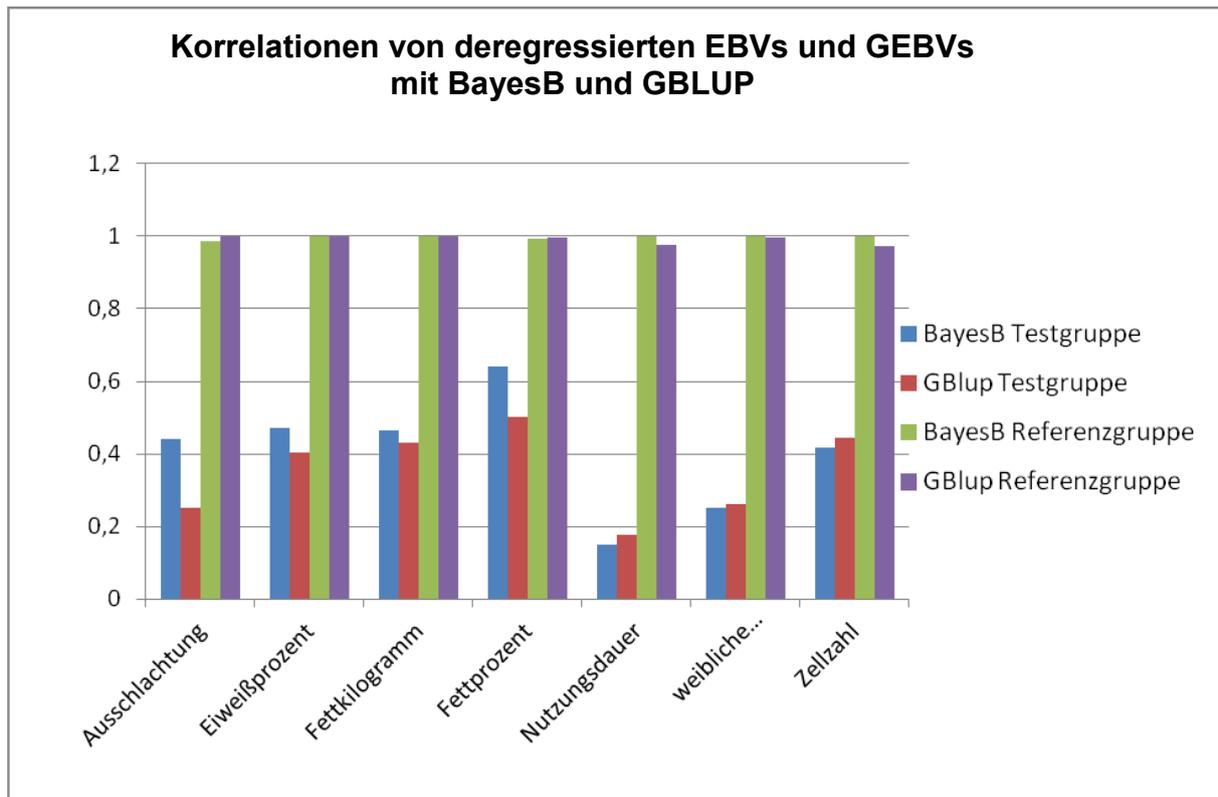


Abbildung 7: Korrelationen von deregressierten EBVs und GEBVs mit BayesB und GBLUP berechnet für den 50K Chip

5.3 Vergleich von direkten und deregressierten Zuchtwerten als Phänotypen

Tabelle 6. Unterschiede der Korrelationen von direkt und deregressierten Zuchtwerten geschätzt mittels BayesB

	3K		7K		50K	
	direkt ZW	deregr. ZW	direkt ZW	deregr. ZW	direkt ZW	deregr. ZW
Ausschlachtung	0,36	0,28	0,43	0,43	0,52	0,44
Eiweißprozent	0,40	0,35	0,46	0,41	0,45	0,47
Fettkilogramm	0,30	0,25	0,45	0,44	0,49	0,46
Fettprozent	0,55	0,49	0,67	0,63	0,71	0,64
Nutzungsdauer	0,32	0,34	0,44	0,19	0,50	0,15
weibliche Fruchtbarkeit	0,49	0,43	0,53	0,23	0,58	0,25
Zellzahl	0,43	0,38	0,45	0,39	0,51	0,42

Vergleicht man die Ergebnisse der direkten Zuchtwerte mit jenen der deregressierten Zuchtwerte, stellt man fest, dass größtenteils die Korrelationen der direkt geschätzten Zuchtwerte höher sind. Das Merkmal Ausschachtung hat bei einer Genotypisierung mit dem 50K Chip den Wert 0,52 bei einer Schätzung mittels direkter Zuchtwerte den Wert 0,44, wenn mit deregressierten Zuchtwerten geschätzt wurde. Ein anderes Beispiel ist das Merkmal Zellzahl, hier liegen die Werte bei 0,51 für die direkten Zuchtwerte und bei 0,41 für die deregressierten Zuchtwerte. Wenn die direkten Zuchtwerte als Phänotypen verwendet werden sind die Ergebnisse im Durchschnitt für den 3K Chip um 0,05 höher, beim 7K um 0,10 und beim 50K um 0,13 wenn die deregressierten Zuchtwerte als Phänotypen zum Einsatz kommen und BayesB verwendet wird. In Abbildung 8 wird graphisch der Unterschied zwischen den mit BayesB berechneten Korrelationen für die Testgruppe dargestellt, je nachdem ob als Phänotypen direkte oder deregressierte Zuchtwerte verwendet wurden. Als Beispiel wurde das Merkmal Ausschachtung gewählt.

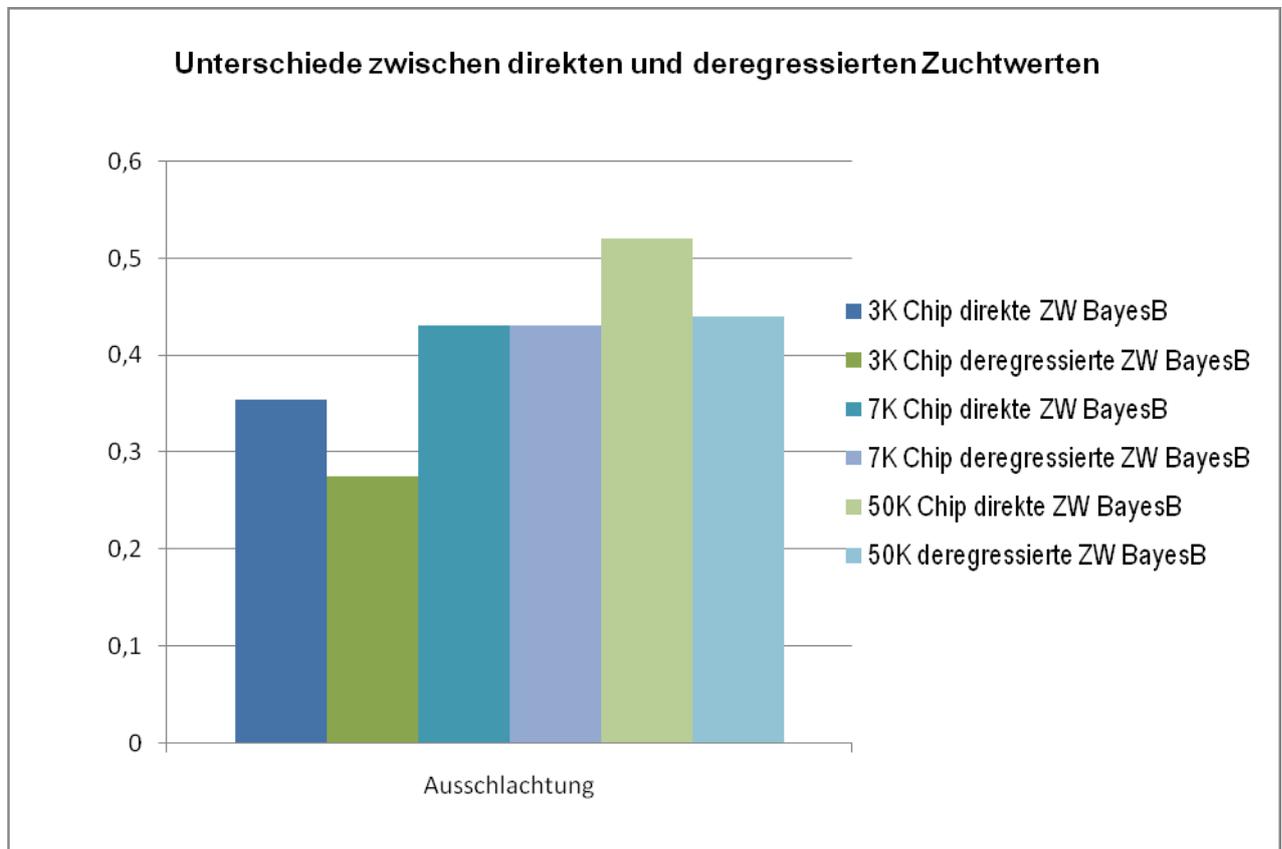


Abbildung 8: Unterschiede zwischen mit BayesB berechneten Korrelationen mit direkten oder deregressierten Zuchtwerten als Phänotypen

Tabelle 7. Unterschiede der Korrelationen von direkten und deregressierten Zuchtwerten geschätzt mittels GBLUP

	3K		7K		50K	
	direkt ZW	deregr. ZW	direktZW	deregr. ZW	direktZW	deregr. ZW
Ausschlachtung	0,34	0,28	0,43	0,28	0,52	0,25
Eiweißprozent	0,35	0,21	0,41	0,31	0,48	0,40
Fettkilogramm	0,25	0,27	0,39	0,40	0,44	0,43
Fettprozent	0,49	0,45	0,55	0,50	0,56	0,50
Nutzungsdauer	0,34	0,14	0,45	0,18	0,51	0,18
weibliche Fruchtbarkeit	0,43	0,14	0,52	0,23	0,58	0,26
Zellzahl	0,38	0,38	0,44	0,40	0,51	0,45

Tabelle 7. zeigt die Unterschiede der Korrelationen von direkten und deregressierten Zuchtwerten, welche mit GBLUP geschätzt wurden. Die Ergebnisse, mit den direkten Zuchtwerten als Phänotypen sind ebenfalls für fast alle Merkmale höher. Dies ist unabhängig des untersuchten SNP Panels ersichtlich. Wenn die direkten Zuchtwerte als Phänotypen verwendet werden, sind die Ergebnisse im Durchschnitt für den 3K Chip um 0,10 höher, beim 7K um 0,13 und beim 50K um 0,15 wenn die deregressierten Zuchtwerte als Phänotypen zum Einsatz kommen und GBLUP verwendet wird.

Abbildung 9 zeigt den Unterschied der mit GBLUP berechneten Korrelationen (EBV-GEV) zwischen direkten und deregressierten Zuchtwerten für das Test-Set, wobei hier das Merkmal Ausschlachtung als Beispiel gewählt wurde. Sowohl bei den durchschnittlichen Unterschieden, als auch aus den Abbildungen 8 und 9 ist ersichtlich, dass bei der Verwendung von GBLUP die Differenz zwischen direkten und deregressierten Zuchtwerten größer ist.

Man kann also daraus schließen, dass sowohl bei der Verwendung von BayesB als auch von GBLUP die Unterschiede zwischen den Korrelationen der direkten und deregressierten Zuchtwerte umso höher sind je mehr SNPs im Panel vorhanden sind. In dieser Arbeit wurden bis auf ein paar Ausnahmen größtenteils höhere Korrelationen mit der Verwendung von direkten Zuchtwerten als Phänotypen erzielt. Zu diesem Ergebnis kamen auch Gredler et. al (2009) in ihrer Studie.

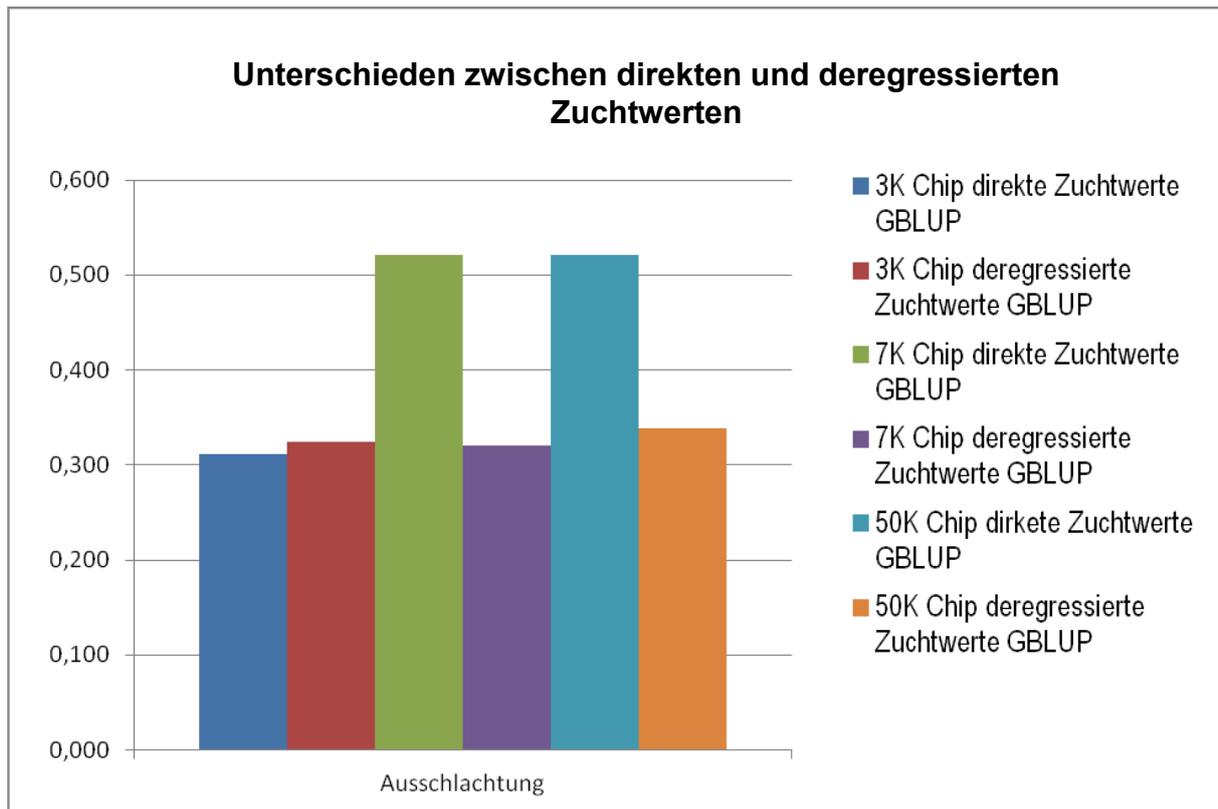


Abbildung 9: Unterschiede zwischen mit GBLUP berechneten Korrelationen mit direkten oder deregressierten Zuchtwerten als Phänotypen

5.4 Ergebnisse der Imputation

Imputiert wurde wie im Kapitel Methoden schon erwähnt wurde mit dem Programm AlphaImpute von John Hickey. Die Imputationseffizienz betrug für den 3Ki 92% und für den 7Ki 94%. In dem Kapitel werden die Daten zur Imputation angegeben.

3K --> 50 K

- Richtig imputierte Genotypen: 74.71%
- Falsch imputierte Genotypen: 5.30%
- Nicht als ganze Zahl imputiert: 19.99%
- Korrelation: 0.90

7K --> 50 K

- Richtig imputierte Genotypen: 80.08%
- Falsch imputierte Genotypen: 2.90%
- Nicht als ganze Zahl imputiert: 17.01%

- Korrelation: 0.93

5.5 Ergebnisse der imputierten SNP-Chips und Vergleich mit den originalen Chips

Nach der Imputation mit AlphaImpute des 3K auf den 50K Chip und des 7K auf den 50K Chip blieben jeweils 40.008 SNPs für die weiteren Berechnungen erhalten.

Tabelle 14: Korrelationen zwischen direkten EBVs und GEBVs von den originalen und imputierten Chips berechnet mit GBLUP					
	3K-50K	3K	7K-50K	7K	50K
Ausschlachtung	0,47	0,34	0,46	0,43	0,52
Eiweißprozent	0,44	0,35	0,44	0,41	0,48
Fettprozent	0,48	0,49	0,48	0,55	0,56
Fettkilogramm	0,39	0,25	0,39	0,39	0,44
Nutzungsdauer	0,52	0,34	0,51	0,45	0,51
weibliche Fruchtbarkeit	0,57	0,43	0,57	0,52	0,58
Zellzahl	0,48	0,38	0,49	0,44	0,51

Tabelle 14 zeigt den Vergleich der imputierten SNP-Chips mit den originalen Chips, berechnet mit GBLUP. Die Werte des originalen 7K Chips sind im Vergleich zum originalen 3K Chip besser, aber für die Zuchtwertschätzung führen die durch Imputation ermittelten Werte zu vielversprechenderen Ergebnissen.

Die berechneten Korrelationen für den imputierten 3K (3Ki) Chip erreichten Werte zwischen 0,39 - 0,57. Im Vergleich zum originalen 3K Chip (Korrelationen zwischen 0,25 – 0,49) schnitt der 3Ki deutlich besser ab, was die Bedeutung der Imputation für die Zuchtwertschätzung klar hervorhebt. In Abbildung 8 wird der Unterschied zwischen 3K und 3Ki Chip nochmals graphisch dargestellt. Für den 7Ki wurden mit GBLUP Korrelationen zwischen 0,39 – 0,57 berechnet und die Ergebnisse sind im

Vergleich zum originalen 7K Chip mit den Ausnahmen Fettprozent und Fettkilogramm ebenfalls deutlich besser.

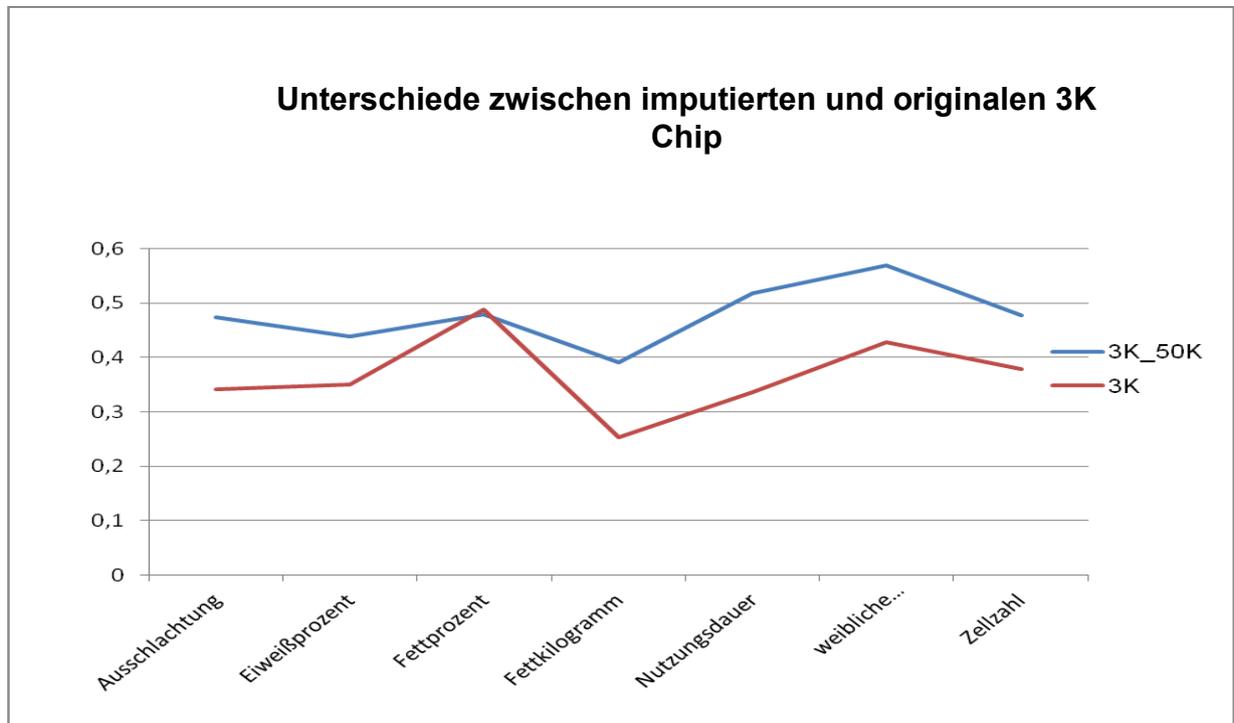


Abbildung 10: Unterschiede im Test-Set bei imputierten und originalen 3K Chip berechnet mit GBLUP

Die Ergebnisse der imputierten Chips liegen in dieser Arbeit unter jenen des 50K Chips (Korrelationen zwischen 0,44 – 0,58), was Abbildung 11 graphisch dargestellt wird. Die einzige Ausnahme ist der 3Ki Chip, welcher beim Merkmal Nutzungsdauer besser performt als der originale 50K Chip. Chen et al. (2011) beschrieben in ihrer Studie, dass die Sicherheit der GEBVs bei der Imputation abhängig von den imputierten Genotypen für eine Gruppe von Merkmalen im Durchschnitt um circa 2,6% sinkt. In dieser Arbeit sind die besseren Resultate des 50K Chips allerdings möglicherweise auch auf die höhere Anzahl an verwendeten SNPs zurückzuführen, da beim 3Ki und 7Ki Chip um 1.670 SNPs weniger in die Berechnung einfließen als beim 50K Chip, sodass ungefähr 93% der Genauigkeit des 50K Chips erzielt wurde.

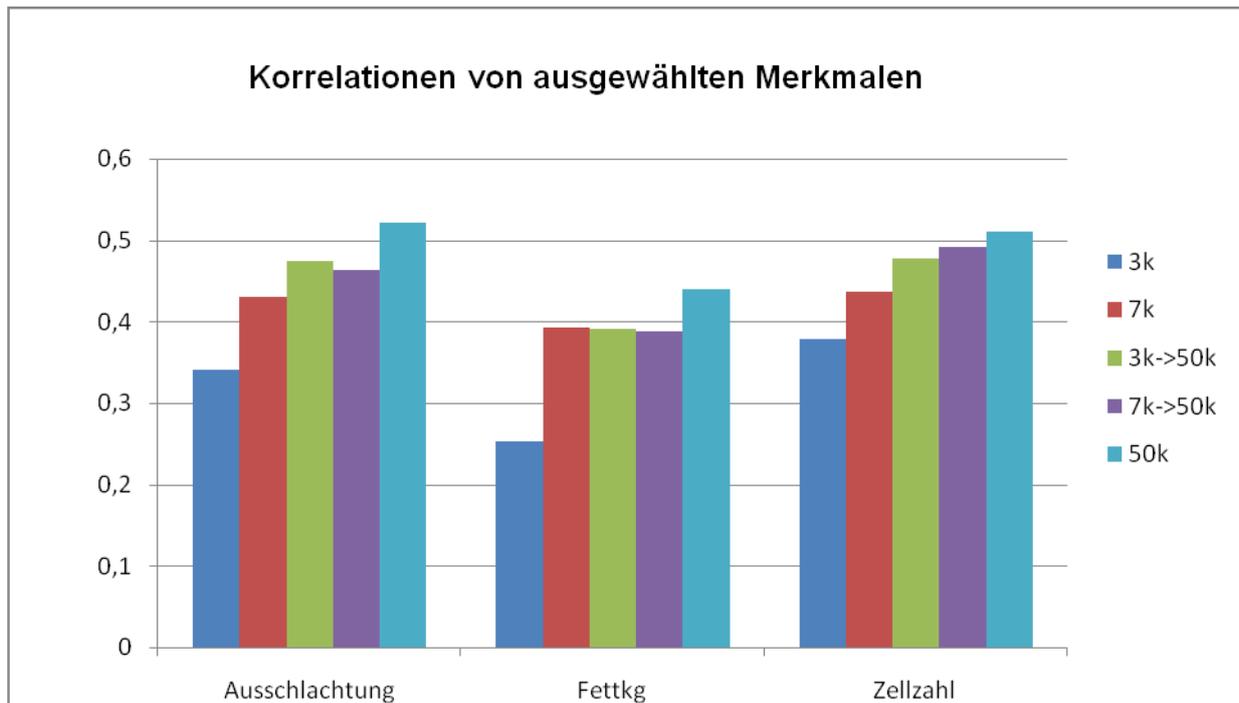


Abbildung 11: Korrelationen von ausgewählten Merkmalen

Derzeit wird der 50K Chip als Industrie-Standard verwendet, jedoch hat sich der 3K Chip in Kombination mit der Imputation als ein nützliches Hilfsmittel erwiesen, welcher eine geringfügig niedrigere Zuverlässigkeit der Ergebnisse aufweist, aber signifikant niedrigeren Kosten. Im September 2011, verbesserte Illumina Inc. seine Low-Density-Chip-Angebot durch den Austausch des 3K mit dem 7K Chip (Wiggans et al. 2012), welcher im Durchschnitt über verschiedene Rassen gesehen um 3% genauer ist, als der 3K Chip. In der Literatur wird fast immer beschrieben, dass der 7Ki Chip besser performt als der 3Ki, wobei ein Grund dafür die unterschiedliche Technologie in den Chips ist. Im 3K Chip wird die GoldenGate und im 7K Chip wird die Infinium Technologie verwendet, welche jener Technologie des 50K Chips entspricht (Boichard et al. 2012). Der wahrscheinlichste Grund für die besseren Ergebnisse ist die höhere Imputationsgenauigkeit des 7K Chips, welche aufgrund der Verdopplung der SNPs im Vergleich zum 3K Chip erreicht wird. Bei einer Studie von Boichard et. al (2012), erreichte der 7Ki Chip eine Imputationsgenauigkeit für die verwendete Holsteinpopulation von 98%, wobei der 3Ki nur auf 94% kam. Ein ähnliches Ergebnis erlangten auch Gredler et al. (2011). Bei einer Imputation mit AlphaImpute von dem 3K auf den 50K Chip wurde eine Imputationsgenauigkeit von 97,74% für die Rasse Brown Swiss erzielt. In dieser Arbeit erreichte der 3Ki Chip allerdings nur eine Imputationsgenauigkeit von 92% und der 7Ki Chip von 94%, auf

ein ähnliches Ergebnis kamen Daetwyler et al. (2011). Sie imputierten mit dem Programm ChromoPhase von dem 3K auf den 50K Chip für eine Holsteinpopulation (1.183 Individuen und Stammbauminformationen von 3.674 Tieren wurde verwendet) und erhielten eine unserer Arbeit ähnliche Imputationseffizienz von 92%. Die in unserer Arbeit erhaltene Imputationseffizienz ist im Vergleich zu den angeführten Studien, die mit AlphaImpute sehr niedrig, allerdings kann der Grund dafür sein, dass wir in dieser Arbeit mit Fleckviehdaten gerechnet haben. In Fleckviehpopulationen existieren mehr Variationen als in den zuvor erwähnten Rasse, zum Beispiel berechneten Boichard et al. (2012), dass im Fleckvieh 99,5% der untersuchten Loci polymorph sind, während im Brown Swiss 96,2% nur alle polymorph sind und in der Nordamerikanischen Holsteinpopulation 98,5% aller Loci polymorph sind. Aufgrund dessen könnte es sein, dass dadurch verminderte Linkage Disequilibrien existieren und deshalb die Imputationseffizienz geschmälert wird. Durch die sehr ähnliche Imputationseffizienz und die Tatsache, dass aufgrund des Extrahierens der SNPs für den 3Ki und 7Ki aus dem originalen 50K Chips, die unterschiedliche Technologie keinen Einfluss auf die Resultate hat, haben der 3Ki und der 7Ki Chip sehr ähnliche Ergebnisse geliefert.

Tabelle 15: Korrelationen zwischen deregressierten EBVs und GEBVs von originalen und imputierten Chips berechnet mit GBLUP					
	3K-50K	3K Chip	7K-50K	7K Chip	50K Chip
Ausschlachtung	0,37	0,28	0,36	0,43	0,44
Eiweißprozent	0,33	0,36	0,34	0,41	0,47
Fettprozent	0,44	0,29	0,44	0,44	0,46
Fettkilogramm	0,38	0,51	0,38	0,63	0,64
Nutzungsdauer	0,21	0,11	0,20	0,19	0,15
weibliche Fruchtbarkeit	0,25	0,18	0,24	0,23	0,25
Zellzahl	0,39	0,37	0,41	0,39	0,42

Tabelle 15 stellt die Korrelationen zwischen deregressierten und genomisch geschätzten Zuchtwerten von originalen und imputierten SNPs dar. Die Werte wurden mit GBLUP berechnet und liegen für den 3Ki zwischen 0,21 – 0,44 und für

den 7Ki ebenfalls zwischen 0,20 – 0,44. Die niedrigsten Korrelationen wurden für die Nutzungsdauer und die weibliche Fruchtbarkeit berechnet, dies kann durch niedrigen Heritabilitäten der beiden Merkmale als auch durch die niedrige Anzahl der Phänotypen, die zur Schätzung der beiden Merkmale zur Verfügung standen, erklärt werden. Die sehr ähnlichen Ergebnisse des 3Ki und 7Ki Chips sind bei den deregressierten Werten gleich zu erklären wie bei den direkten Zuchtwerten. Die Imputationseffizienz des 7Ki Chips ist nicht entscheidend besser als jene des 3Ki Chips, außerdem hat durch das Extrahieren der SNPs des 3Ki und 7Ki Chips aus dem 50K Chip die unterschiedliche Technologie der Chips keinen Einfluss auf die Resultate. Die Imputation des 3Ki und 7Ki Chips bringt für fast alle Merkmale bessere Resultate als die originalen Chips, auffällig sind allerdings die Merkmale Eiweißprozent und Fettkilogramm. Diese Merkmale hatten unerklärlicher Weise nach der Imputation der Chips (3K und 7K) schlechtere Ergebnisse als vorher. Für die Referenzgruppe wurden Korrelationen zwischen 0,98 – 1,00, sowohl für den 3Ki Chip als auch für den 7Ki Chip.

Bisherige Ergebnisse zeigen, dass die Benutzung von HD-SNP Genotypen den genetischen Prozess in der Milchviehzucht bedeutend verbessert, jedoch ist dies zu derzeitigen Preisen den Elitetieren vorbehalten (Berry und Kearny 2011, Weigel et al. 2010). Um die Akzeptanz der Genetik auf kommerziellen Betrieben zu erweitern, müsste eine Kostenreduktion der Genotypisierung erfolgen (Weigel et al. 2010). Laut einer Studie von Weigel et. al. (2010) könnte die Entwicklung der Imputation von LD auf HD Genotypen diese Akzeptanz fördern, wobei die Genotypisierung mittels der 3K Chips weniger als die Hälfte als jene mit dem 50K Chip durchgeführte Genotypisierung kostet. Die Senkung der Kosten der genomischen Selektion führt zu einem höheren Nutzen durch die Landwirte, wodurch sich auch die Referenzpopulation für die Schätzung der SNP-Effekte und auch die Genauigkeit der Zuchtwertschätzung vergrößert, womit sich auch der Zuchtfortschritt erhöht (Berry und Kearney 2011).

Die mittels Imputation berechneten genomischen Zuchtwerte sind vielversprechend. Der Sicherheitsverlust ist je nach Imputationsmethode unterschiedlich groß und für die Zuchtwertschätzung für die breite Masse vertretbar.

In der Studie von Chen et al. (2011) wurden ebenfalls der 3K Chip auf den 50K Chip imputiert und dann die Werte für verschiedene Merkmale getestet und verglichen, und es wurde ein Sicherheitsverlust im Durchschnitt von 2,6% ermittelt bei einer Imputation mit Beagle. Unter anderen wurden die Merkmale Fettkilogramm, Eiweißkilogramm, Nutzungsdauer und Zellzahl kalkuliert, wobei der Genauigkeitsverlust für diese Merkmale zwischen 1,3 – 3,8% schwankt. Den niedrigsten Verlust für jene Merkmale wurde für Eiweißkilogramm, der höchste für Fettkilogramm berechnet. Nimmt man diesen Durchschnittswert plus die etwas geringere SNP Anzahl im 3Ki und 7Ki Chip, entspricht das ungefähr den Werten, die in dieser Studie berechnet wurden. Chen et al. (2011) kamen außerdem zu der Erkenntnis, dass die Benutzung der imputierten Genotypen die Ergebnisse der genomischen Zuchtwerte genauso wenig verzerrt wie die Benutzung des 50K Chips. Weigel et al. (2011) benutzten eine Referenzpopulation von 1446 Jersey Stiere, die mit dem 50K Chip für die Merkmale Milchkilogramm, Eiweißprozent und weibliche Fruchtbarkeit genotypisiert waren. Sie berichteten, dass mit 2.942 SNPs und Imputation auf den 50K Chip DGVs mit einer Genauigkeit von rund 95% im Vergleich zur Verwendung von echten Genotypen vorhergesagt werden können, wenn eine geeignete Referenzpopulation mit Genotypen vorhanden ist, die mit hoher SNP-Dichte genotypisiert wurde. Dieses Ergebnis kommt jenen in unserer Arbeit ermittelten Ergebnis nahe, wenn man bedenkt, dass die amerikanische Jerseypopulation mit 94% polymorpher Loci deutlich weniger Variation aufweist als die Fleckviehpopulation mit 99,5% polymorpher Loci. Im Durchschnitt wurde für unsere Fleckviehpopulation eine Genauigkeit von rund 93% für den 3Ki und 7Ki Chip im Vergleich zum 50K Chip erzielt.

Eine andere Studie von Berry und Kearney (2011), in welcher der 3K auf den 50K Chip imputiert wurde, zeigte für die genomischen Zuchtwertschätzung nur eine kleine mittlere Differenz, vergleicht man direkte genomische Zuchtwerte von realen oder imputierten Genotypen. Die Korrelation zwischen der DGVs der Referenzgruppe, wenn sie ihre wirklichen oder imputierten Genotypen für die genomischen Vorhersage benutzt wurden, variierten zwischen 0,92 - 0,99, was den in dieser Arbeit ermittelten Werten entspricht.

Schefers und Weigel erwähnten in ihrem Programm (2012), dass ein großer Nachteil der LD-Chips die Imputationsfehler sind, die auftreten können und damit die Genauigkeit der anschließend geschätzten Zuchtwerte negativ beeinflussen.

Aufgrund der Werte, welche in dieser und den angeführten anderen Studien ermittelt wurden, kann man auf eine relativ gute Übereinstimmung zwischen den realen und imputierten Genotypen schließen.

6 Conclusio

- Die Panelgröße hat einen erheblichen Einfluss auf die berechneten Genauigkeiten der genomischen Zuchtwerte, das bedeutet die Korrelationen zwischen EBVs und GEBVs steigen mit der Panelgröße. Allerdings sind die berechneten Genauigkeiten für den 7K Chip erheblich besser als jene des 3K Chips. Aufgrund dieser Ergebnisse kann man darauf schließen, dass eine Zuchtwertschätzung unter alleiniger Verwendung von Tieren, die mit dem 3K Chip genotypisiert werden, ausgeschlossen ist, da er im Durchschnitt nur 75% der Genauigkeit des 50K Chips erzielte. Für den 7K Chip wurde in unserer Studie im Durchschnitt 90% der Genauigkeit des 50K Chip ermittelt, weshalb dieser Chip eher zur alleinigen Verwendung in der Zuchtwertschätzung in Frage käme. Eine echte Alternative um die Kosten der genomischen Selektion zu reduzieren ist jedoch die Nutzung eines LD-Chips mit anschließender Imputation, durch welche Genauigkeiten von ungefähr 93% des 50K Chips erreicht wurden. Daraus kann man schließen, dass die Verwendung der Imputation eines LD-Chips gut geeignet ist um weibliche Populationen zu genotypisieren bzw. für die Vorselektion junger Stiere.
- Unsere Arbeit zeigte, dass die Verwendung von deregressierten Zuchtwerten als Phänotypen für die Zuchtwertschätzung im Vergleich zu den konventionellen Zuchtwerten nicht sinnvoll ist. Bei nahezu allen Werten schnitten die mit deregressierten Zuchtwerten berechneten Korrelationen schlechter ab, als die mit konventionellen Zuchtwerten berechneten Korrelationen.
- Der imputierte 3K und der imputierte 7K Chip liefern annähernd dieselben Ergebnisse, wobei man erwarten würde, dass der imputierte 7K Chip besser abschneidet aufgrund der mehr als verdoppelten SNP Menge, die in diesem Chip inkludiert ist. Da, dies ist unserer Studie nicht der Fall ist, legt es die Vermutung nahe, dass die normaler Weise besseren Ergebnisse des imputierten 7K Chips hauptsächlich ein Resultat der besseren

Imputationseffizienz sind. Die beiden Chips hatten in unserem Fall bei der Imputationseffizienz mit 92% für den 3K Chip und 94% für den 7K Chip keine großen Unterschiede, somit konnte es auch zu den beinahe identen Resultaten kommen.

7 Zusammenfassung

Die genomische Selektion ist eine neue Technologie um Zuchtwerte auf DNA Ebene zu schätzen. Die Analyse erfordert über das gesamte Genom verteilte molekulare Marker. Die am häufigsten verwendeten Marker sind sogenannten Single Nucleotide Polymorphismen (SNP), wobei es sich bei einem SNP um eine Punktmutation handelt. Eine Punktmutation bezeichnet einen Austausch eines einzigen Nukleotiden im Genom.

Mittels genomischer Selektionstechniken ist es möglich direkt nach der Geburt Zuchtwerte für die Kälber zu erhalten. Dadurch kann man die Kälber früher zu selektieren und wodurch das Generationsintervall reduziert und der Zuchtfortschritt erhöht wird.

Im Rinder-Genom existieren mehrere Millionen SNPs, aber in der Tierzucht wird nur ein Bruchteil davon verwendet. Die Genotypisierung von Rindern wurde mit dem Bovine SNP50 Bead Chip (enthält 54.001 SNPs) durchgeführt. Seit September 2011 bietet die Firma Illumina Chips die 2.900 und 6.909 SNPs enthalten. Das Ziel dieser Arbeit war es zu untersuchen, wie sicher genomische Zuchtwertschätzung mit einem solchen Bovine LD Bead Chip durchgeführt werden kann, wenn die fehlenden SNPs imputiert, d.h. mit vorhandenen Informationen aus der Population und von engen Verwandten ergänzt werden. Außerdem wurde die Genauigkeit der genomischen Zuchtwerte, die mit verschiedenen originalen imputierten SNP Chips erzielt werden können, untersucht.

Genotypen wurden von 2.123 Simmental Stiere zur Verfügung gestellt, klassifiziert nach Geburtsjahr in Test- und Referenzgruppen. Konventionell geschätzte Zuchtwerte (EBV) und deregressierte Zuchtwerte mit hohen Sicherheiten für Fettkilogramm, Fettprozent, Eiweißprozent, Nutzungsdauer, weibliche Fruchtbarkeit, Zellzahl und Ausschlagung wurden als Phänotypen verwendet. Die im 3K und 7K Chip enthaltenen SNPs, wurden aus dem 50K Chip extrahiert und nicht separat mit den kleineren Chips genotypisiert. Nach der Qualitätskontrolle des 50K Chips blieben 41.082 SNPs für die Analyse übrig. Der 3K Chip enthielt 2.890 SNPs und die 7K Chip

6.565 SNPs für weiter Berechnungen. Imputiert wurde mit der Software AlphaImpute entwickelt von John Hickey. Der 3K und 7K Chip wurden auf die 50K Chip imputiert. Nach der Imputation enthielten der 3K und der 7K Chip 40.008 SNPs (nach dem Entfernen von SNPs mit unbekannter Position), die für die Schätzung verwendet werden konnten. BayesB und GBLUP wurde verwendet um genomische Zuchtwerte für die nicht-imputierten Chips zu schätzen. Für die imputierten SNP Sets wurde nur GBLUP verwendet und die Ergebnisse wurden mit denen der nicht-imputierten Chips verglichen.

Für die mit 50K Chip berechneten Korrelationen zwischen genomischen und konventionellen Zuchtwerten wurden Korrelationen zwischen 0,44 – 0,71 mit GBLUP und BayesB ermittelt. Die Ergebnisse, die mit deregressierten Zuchtwerten erzielt wurden, sind in nahezu allen SNP-Panelen für alle Merkmale niedriger als jene mit direkten Zuchtwerten berechneten Korrelationen und lagen für den 50K Chip zwischen 0,15 – 0,64.

Die Ergebnisse des 3K Chips waren zwischen 0,25 – 0,49 für die direkten Zuchtwerte. Die Werte des 7K Chips kommen jenen des 50K Chips sehr nahe und differierten für die direkten Zuchtwerte nur um 0,04 – 0,09 und für die deregressierten Zuchtwerte um 0 - 0,06. Dies bedeutet, dass der 3K Chip eine Genauigkeit von um die 75% und der 7K Chip von um die 90% im Vergleich zum 50K Chip erreicht hat. Durch eine Imputation der LD-Chips auf den 50K Chip konnten die erzielten Werte noch verbessert werden. Die imputierten Ergebnisse brachten sowohl für den 3Ki als auch für den 7Ki (Korrelationen zwischen 0,39 – 0,57) im Durchschnitt 93% der Genauigkeit des 50K Chips für die direkten Zuchtwerte.

Die LD-Chips mit anschließender Imputation auf den 50K Chip könnten eine mögliche Alternative für die Vorselektierung junger Tiere, beziehungsweise für die Genotypisierung einer großen weiblichen Population sein, wie auch in der Literatur berichtet wird (Dassonville et al. 2011).

Schlüsselwörter: Genomische Selektion, SNP Subsets, Rinder, Genauigkeiten, Fleckvieh, genomische Zuchtwerte.

8 Summary

Genomic selection is a new technique to estimate breeding values on the DNA level. The analysis requires molecular markers distributed throughout the genome. The most widely used markers are the so called single nucleotide polymorphisms (SNP), which are point mutations replacing a single nucleotide in the genome.

With genomic selection techniques it is possible to get breeding values of calves right after birth, which gives the possibility to select them earlier and therefore reduce the generation interval and increase genetic gain.

There are several million SNPs in the cattle genome, but only a fraction of these is used in animal breeding. The genotyping of animals was done with the Bovine SNP50 BeadChip (contains 54.001 SNPs). Since September 2011, the company Illumina provides chips that contain 2.900 and 6.909 SNPs. The aim of this study was to investigate the accuracies of genomic breeding values with different SNP chips and in case of SNP imputation.

Genotypes were from 2123 Simmental bulls, classified according to year of birth into test and reference groups. Highly reliable estimated breeding values (EBV) and deregressed breeding values for fat content, fat yield, protein content, longevity, non-return rate after 56 days, somatic cell score and dressing percentage were used as phenotypes. The information contained in the 3K and 7K SNPs were extracted from the 50K chip, and not separately genotyped with the smaller chips. After the quality check of the 50K chip 41.082 SNPs remained for analysis. The 3K chip contained 2.890 SNPs and the 7K chip 6.565 SNPs. With AlphaImpute, developed by John Hickey the 3K and 7K chip was imputed to the 50K chip. After the imputation the 3K and the 7K Chip contained 40.008 SNPs (due to deletion of SNPs with unknown position), which could be used for estimation. BayesB and GBLUP were used to estimate genomic breeding values for the non-imputed chips. Only the GBLUP was used for the imputed SNP sets and the results were compared to those from non-imputed sets. For the 50K chip the correlations between genomic and conventional EBVs were between 0,44 – 0,71 with using GBLUP and BayesB. The results from

deregressed breeding values were in almost all SNP panels for all traits lower than those calculated with direct breeding values. The calculated correlations for the 50K chip were between 0,15 to 0,64.

The results from the 3K chips were between 0,25 to 0,49 computed for the direct breeding values. The results from the 7K chips are very close to those of the 50K chips and differ for direct breeding values by only 0,04 to 0,09 and for the deregressed breeding values by 0 – 0,06. This means that the 3K chip achieves an accuracy of around 75%, and the chip 7K of around 90% in comparison to the 50K chip. By imputation of lower density chips to 50K chip these values could be further improved. The results for imputed sets of the 3Ki and 7Ki (correlations between 0,39 to 0,57) were in average 93% of the accuracy of the 50K chip calculated for the direct breeding values. The lower density chip with subsequent imputation on the 50K chip could be a possible alternative for the preselection of young animals, or for genotyping a large female population, as reported in the literature (e.g. Dassonville et al. 2011).

Keywords: Genomic selection, SNP subsets, accuracies, cattle, Fleckvieh, GEBV.

9 Literaturverzeichnis

Berry, D.P. und Kearney, J.F. (2011). Imputation of genotypes from low- to high-density genotyping platforms and implications for genomic selection. *Animal*(2011) 5:8: 1162-1169.

Boichard, D., Hoyoung Chung, H., Dassonneville, R., David, X., Eggen, A., Fritz, S., Kimberly J. Gietzen, K.J., Hayes, B.J., Lawley, C.T., Sonstegard, T.S., Van Tassell, C.P., Van Raden, P.M., Viaud-Martin, K.A. und Wiggans, G.R. (2012). Design of a Bovine Low-Density SNP Array Optimized for Imputation. *PLoS one*. Volume 7/ Issue 3/ e34130.

BovineHD Genotyping BeadChip. Illumina 2010.

BovineLD Genotyping BeadChip. Illumina 2011.

BovineSNP50 Genotyping BeadChip. Illumina. Inc. 2011.

Browning, B.L. und Browning, S.R. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81:1084–1097.

Browning, B. L. und Browning, S.R. (2009). A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics* 84: 210–223.

Calus, M. P. L., Veerkamp, R. F. und Mulder, H. A. (2011). Imputation of missing single nucleotide polymorphism genotypes using a multivariate mixed model framework. *J. Anim. Sci.* 89: 2042-2049.

Chen, J., Liu, Z., Reinhardt, F. und Reents, R. (2011). Reliability of Genomic Prediction Using Imputed Genotypes for German Holsteins: Illumina 3K to 54K Bovine Chip. *INTERBULL BULLETIN NO.* 44:51-54.

Daetwyler, H.D., Pong-Wong, R., Villanueva, B., Woolliams, J.A., 2010. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* 185:1021-1031.

Daetwyler, H.D., Wiggans, G.R., Hayes, B.J., Woolliams, J.A. und Goddard, M.E. (2011). Imputation of Missing Genotypes From Sparse to High Density Using Long-Range Phasing. *Genetics* 189:317-327.

Dassonneville, R., Brøndum, R.F, Druet, T., Fritz, S., Guillaume, F., Guldbrandtsen, B., Lund, M.S., Ducrocq, V., und Su, G. (2011). Effect of imputing markers from a low-density chip on the reliability of genomic breeding values in Holstein populations. *J. Dairy Sci.* 94:3679–3686.

Dassonneville, R. (2012). Dissertation: Genomic selection of diary cows. AgroParisTech. S 1-163.

Dassonneville R., Fritz S., Boichard D. and Ducrocq V. (o.J.). Imputation efficiency with different low density chips in French diary and beef breeds.

Dekkers, J. C. M. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *J. Anim. Sci.* 82:313–328.

Dekkers, J.C.M. und Hospital, F.(2002). Multifactorial genetics: The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev.Genet.* 3:22–32.

De Roos, A. P. W., Hayes, B.J., Spelman, R. J. und Goddard, M. E, (2008). Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle. *Genetics* 179:1503–1512.

Elberfeld, M. (2007). Diplomarbeit: Komplexität von Haplotypisierung mittels perfekten Phylogenien und kleinsten Haplotypmengen. Lübeck: Universität zu Lübeck. S.1-77.

De Roos, A.P.W., Schrooten C., Druet T. 2011. Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. *Journal of Dairy Science*, 94, S. 4708–4714.

Fürst, C. (2011). Zuchtwertschätzung beim Rind: Grundlagen, Methoden und Interpretationen. Wien: Zuchtdata EDV-Dienstleistungs GmbH. S1-109.

Gredler, B., Nirea, K.G., Solberg, T.R, Egger-Danner, C., Meuwissen, T. und Sölkner, J. (2009). A comparison of methods for genomic selection in Austrian dual purpose Simmental cattle Association for Advancement of Animal Breeding and Genetics. Proceedings of the 18th Conference Barossa Valley, South Australia 28 September – 1 Oktober 2009, S.568.

Gredler, B., Seefried, F.R., Schuler, U., Bapst, B., Schnyder, U. und Hickey, M.J. (2011). INTERBULL BULLETIN NO 44: 8-11.

Gredler, B., Schwarzenbacher, H., Egger-Danner, C., Fürst, C., Emmerling, R. und Sölkner, J. (o.J.). Predictive ability of genomic selection in dual purpose Fleckvieh cattle using different types of dependent variables.

Howie, B.N., Donnelly, P. und Marchini, J.(2009). A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genetics*. 5:1-15.

Gao, H., Christensen, F.C., Madsen, P. , Nielsen, S.N., Zhang, Y., Mogens S Lund, M.S. und Su, G. (2012). Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet. Sel. Evol.* 44:8.

Golden Gate Bovine3K Genotyping BeadChip. Illumina Inc. 2010.

Goodard, M.E. und Hayes, B.J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381-391.

Gredler, B.() Was ist genomische Selektion. <http://www.fleckvieh.at/news-ticker/bericht/details/was-ist-die-genomische-selektion-1735.html> besucht am 24.03.2012.

Habier, D., Fernando, R.L. und Dekkers, J.C.M. (2009). Genomic Selection Using Low-Density Marker Panels. *Genetics* 182:343-353.

Hayes, B.J und Goddard, M. E.(2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Gen. Sel. Evol.*33:209-229.

Hayes, B.J. (2008). Whole Genome Association and Genomic Selection. Salzburg: Course Notes.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. und Goddard, M.E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433-443.

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-FatPercentage, and Type in Holstein Cattle as Contrasting Model Traits. *PLoS Genet* 6, e1001139.

Hickey, J., Kinghorn, B., Tier, B., Van der Werf, J. und Cleveland, M. (2011). AlphaImpute. HYPERLINK "<https://sites.google.com/site/hickeyjohn/alphaimpute>" <https://sites.google.com/site/hickeyjohn/alphaimpute> aufgerufen am 08.09.2012.

Kachman, S.L. (2008). Incorporation of Marker Scores into National Genetic Evaluations. University of Nebraska-Lincoln: Department of Statistics.

Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature Genet.* 17: 21-24.

Lillehammer, M., Meuwissen T.H.E. und Sonesson, A.K. (2011). A comparison of dairy cattle breeding designs that use genomic selection. *J. Anim.Sci.* 94:493-500.

Lewis J., Abas Z., Dadousis C., Lykidis D., Paschou P und Drineas P. (2011). Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs. PLOS one, S.1-8.

Marchini J. and Howie B. (2010). Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11:499-511.

Meuwissen, T.H.E. und Goddard M.E. (1996). The use of marker haplotypes in animal breeding schemes. Genet. Sel. Evol. 28:161-176.

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

Meuwissen, T.H.E. (2009). Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:1-9.

Meuwissen, T.H.E. (2009a). Manual for program bayesgg.

Moser G., Tier, B., Crump, R.E., Khatkar M.S. und Raadsma H. W. (2009). A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. Genet. Sel. Evol. 41:1-56.

Moser G., Khatkar M.S., Hayes B.J. und Raadsma H.W. (2010). Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. Genet. Sel. Evol. 42:37.

Neuner S.J. (2009). Dissertation: Untersuchung zur Optimierung von markerunterstützten Zuchtwertschätzverfahren in der Rinderzucht. Bayern:Universität Kiel. S1-91.

Raganitsch, G. (2001). Das Österreichische Fleckvieh und seine Genetik (Bd. 1). Zwettl: Eigenverlag der AGÖF. S 1-15.

Schefers, J.M. und Weigel, K.A. (2012). Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Department of Dairy Science, University of Wisconsin, Madison 537061-9.

Van der Beek, S. (o.J.). Effect of Genomic Selection on National on International Genetic Evaluations.

Van Raden, P.M. (2008). Efficient Methods to Compute Genomic Predictions. J.Dairy Sci. 91:4414-4423.

Van Raden, P.M., Wiggans, G.R., Sonstegard, T.S. und Van Tassell C.P. (2008a). Using Genomic Data to Improve Dairy Cattle Genetic Evaluations. Beltsville: USDA Animal Improvement Programs and Bovine Functional Genomics Laboratories.

Van Raden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2008b). Invited review: reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 92:16-24.

Van Raden, P.M., O'Connell J.R., Wiggans G.R. und Weigel K.A. (2011). Genomic evaluations with many more genotypes. 43:10.

Weigel, K.A., De los Campos, G., Vazques, A.I., Van Tassell, C.P., Rosa, G.J.M., Gianola D., O'Connell, J.R., Van Raden, P.M. und Wiggans, G.R. (2010). Genomic Selection and its Effects on Dairy Cattle Breeding Programs.

Weigel, K.A., de los Campos G., Vazques, A.I., Rosa, G.J.M., Gianola D. and Van Tassel, D.P. (2010b). Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. J. Dairy Sci. 93: 5423-5435.

Wiggans, G.R., Cooper, T.A., Van Raden P.M., Olson, K.M. und Tooker, M.E. (2012). Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation¹. J. Dairy Sci. 95 :1552–1558.

Wiggans, G.R., Sonstegard, T.S., Van Raden, P.M., Matukumalli, L.K., Schnabel, R.D., Taylor, J.F., Schenkel F.S. und Van Tassell C.P. (2008). Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J. Dairy Sci.* 92 :3431–3436.

Zhang, Z., Ding, X., Liu, J., Zhang, Q. und de Koning, D.J. (2011). Accuracy of genomic prediction using low-density marker panels. *J. Dairy Sci.* 94:3642-3650.

ZAR. (2011). Jahresbericht. Wien: HAMMERER GmbH & CoKG.

10 Tabellenanhang

In den folgenden Tabellen 8-13,16 und 17 steht T die für Testgruppe und R für die Referenzgruppe.

Tabelle 8: Korrelationen zwischen EBVs und GEBVs für den 3K Chip				
	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,36	0,34	0,89	1,00
Eiweißprozent	0,40	0,35	0,89	0,98
Fettkilogramm	0,30	0,25	0,95	1,00
Fettprozent	0,55	0,49	0,88	0,99
Nutzungsdauer	0,32	0,34	0,83	0,99
weibliche Fruchtbarkeit	0,49	0,43	0,88	0,99
Zellzahl	0,43	0,38	0,91	0,99

Tabelle 9: Korrelationen zwischen EBVs und GEBVs für den 7K Chip				
	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,43	0,43	1,00	1,00
Eiweißprozent	0,46	0,41	1,00	0,99
Fettkilogramm	0,45	0,39	1,00	1,00
Fettprozent	0,67	0,55	0,98	1,00
Nutzungsdauer	0,44	0,45	0,95	1,00
weibliche Fruchtbarkeit	0,53	0,52	0,95	1,00
Zellzahl	0,45	0,44	0,98	1,00

	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,52	0,52	1,00	1,00
Eiweißprozent	0,45	0,48	1,00	1,00
Fettkilogramm	0,49	0,44	1,00	1,00
Fettprozent	0,71	0,56	1,00	1,00
Nutzungsdauer	0,50	0,51	1,00	1,00
weibliche Fruchtbarkeit	0,58	0,58	1,00	1,00
Zellzahl	0,51	0,51	1,00	1,00

	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,28	0,28	0,86	1,00
Eiweißprozent	0,36	0,21	0,88	1,00
Fettkilogramm	0,29	0,27	0,94	0,99
Fettprozent	0,51	0,45	0,88	0,98
Nutzungsdauer	0,11	0,14	0,79	0,93
weibliche Fruchtbarkeit	0,18	0,14	0,79	0,98
Zellzahl	0,37	0,38	0,90	0,94

	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,43	0,28	0,97	1,00
Eiweißprozent	0,41	0,31	0,96	1,00
Fettkilogramm	0,44	0,40	0,98	1,00
Fettprozent	0,63	0,50	0,97	0,99
Nutzungsdauer	0,19	0,18	0,94	0,97
weibliche Fruchtbarkeit	0,23	0,23	0,92	0,99
Zellzahl	0,39	0,40	0,97	0,96

Tabelle 13: Korrelationen zwischen deregressierten EBVs und GEBVs für den 50K Chip

	BayesB T	GBLUP T	BayesB R	GBLUP R
Ausschlachtung	0,44	0,25	0,99	1,00
Eiweißprozent	0,47	0,40	1,00	1,00
Fettkilogramm	0,46	0,43	1,00	1,00
Fettprozent	0,64	0,50	1,00	1,00
Nutzungsdauer	0,15	0,18	1,00	0,98
weibliche Fruchtbarkeit	0,25	0,26	1,00	1,00
Zellzahl	0,42	0,45	1,00	0,97

Tabelle 16: Korrelationen zwischen direkten EBVs und GEBVs von den imputierten Chips

	GBLUP T	GBLUP R	GBLUP T	GBLUP R
	3K-50K	3K-50K	7K-50K	7K-50K
Ausschlachtung	0,47	1,00	0,46	1,00
Eiweißprozent	0,44	1,00	0,44	1,00
Fettprozent	0,48	1,00	0,48	1,00
Fettkilogramm	0,39	1,00	0,39	1,00
Nutzungsdauer	0,52	1,00	0,51	1,00
weibliche Fruchtbarkeit	0,57	1,00	0,57	1,00
Zellzahl	0,48	1,00	0,49	1,00

Tabelle 17: Korrelationen zwischen deregressierten EBVs und GEBVs von den imputierten Chips				
	GBLUP T	GBLUP R	GBLUP T	GBLUP R
	3K-50K	3K-50K	7K-50K	7K-50K
Ausschlachtung	0,37	1,000	0,36	1,000
Eiweißprozent	0,33	1,000	0,34	1,000
Fettprozent	0,44	0,99	0,44	0,99
Fettkilogramm	0,38	1,00	0,38	1,00
Nutzungsdauer	0,21	0,98	0,20	0,98
weibliche Fruchtbarkeit	0,25	1,00	0,24	1,00
Zellzahl	0,39	1,00	0,41	1,00