

**Aus dem Institut für Tierzucht und Haustiergenetik
der Georg-August-Universität Göttingen**

**Untersuchungen zur Genauigkeit einer
genomgestützten Zuchtwertschätzung**

**Dissertation
zur Erlangung des Doktorgrades
der Fakultät für Agrarwissenschaften
der Georg-August-Universität Göttingen**

**vorgelegt von
Dörthe Agena
geboren in Göttingen**

Göttingen, Dezember 2008

D7

1. Referent: Prof. Dr. Henner Simianer
2. Korreferent: Prof. Dr. Christoph Knorr

Tag der mündlichen Prüfung: 29. Januar 2009

Inhaltsverzeichnis

1 EINLEITUNG	1
2 LITERATURÜBERSICHT	3
2.1 ENTWICKLUNG DER TIERZUCHT.....	3
2.2 GENETISCHE MARKER.....	6
2.3 MARKERGESTÜTZTE SELEKTION	9
2.4 GENOMIC SELECTION – GRUNDLEGENDE METHODEN	12
2.4.1 Least squares	15
2.4.2 BLUP	16
2.4.2.1 Die Modelle der BLUP-Zuchtwertschätzung	18
2.4.2.2 BLUP in der genomic selection	21
2.4.3 Bayes Methode.....	22
2.5 GENOMIC SELECTION – ANGEWANDTE METHODEN.....	23
3 MATERIAL UND METHODEN	39
3.1 DIE SIMULATIONSPROGRAMME	39
3.2 DIE AUSWERTUNG	44
4 ERGEBNISSE UND DISKUSSION.....	48
4.1 HAPLOTYPEN.....	48
4.2 EINFLUSS DER PARAMETER AUF DEN GENOMISCHEN ZUCHTWERT.....	50
4.2.1 Einfluss der Größe der Analysegeneration	50
4.2.2 Einfluss der Anzahl der Generationen.....	56
4.2.3 Einfluss der Populationsgröße.....	58
4.2.4 Einfluss der Segmentgröße.....	60
4.2.5 Einfluss der Mutationsrate	66
4.2.6 Einfluss der SNP-Anzahl	67
4.2.7 Einfluss der Heritabilität.....	73
4.2.8 Genauigkeit der genomischen Zuchtwerte in den Folgegenerationen.....	76
4.2.8.1 Einfluss der Größe der Analysegeneration	76
4.2.8.2 Einfluss der Segmentgröße.....	80
4.2.8.3 Einfluss der Heritabilität.....	81
4.2.9 Die Selektion.....	84
4.2.9.1 Die zufällige Selektion	84
4.2.9.2 Die gerichtete Selektion	86
4.2.9.3 Vergleich von zufälliger und gerichteter Selektion	88
4.3 RESAMPLING	89

5 SCHLUSSFOLGERUNG	92
6 ZUSAMMENFASSUNG.....	95
7 SUMMARY	97
LITERATURVERZEICHNIS.....	99

Tabellenverzeichnis

Tabelle 1:	Übersicht über die Veröffentlichungen mit Angabe von Verfahren und Ergebnissen bei simulierten Daten.....	24
Tabelle 2:	Korrelation zwischen dem wahren und genomischen Zuchtwert für vier verschiedene Methoden und eine unterschiedliche Anzahl von Tieren (verändert nach Meuwissen et al., 2001)	25
Tabelle 3:	Korrelationen zwischen dem wahren und genomischen Zuchtwert für verschiedene Heritabilitäten, QTL-Gewichtungen und Positionen (Kolbehdari et al., 2007).....	30
Tabelle 4:	Korrelationen zwischen dem wahren und genomischen Zuchtwert für Großeltern, Eltern und Nachkommen bei $h^2 = 0,1$ und $0,5$ und polygenen Effekten (verändert nach Calus und Veerkamp, 2007)	32
Tabelle 5:	Korrelation zwischen dem wahren und genomischen Zuchtwert für unterschiedliche Markerdichten bei Markergenotypen und -haplotypen sowie SNP-Genotypen und Haplotypen (verändert nach Solberg et al., 2008)	37
Tabelle 6:	Übersicht über die veränderbaren Parameter	39
Tabelle 7:	Anzahl selektierter Tiere und Selektionsintensität für drei verschiedene Szenarien	43
Tabelle 8:	Signifikanzen (t-Test) zwischen Nachkommenanzahlen für alle Elternpaaranzahlen (EP) für MSE- und Korrelationswerte (r).....	52
Tabelle 9:	Signifikanzen (t-Test) zwischen Elternpaaranzahlen für alle Nachkommenanzahlen (NK) für MSE- und Korrelationswerte (r).....	53
Tabelle 10:	Signifikanzen (t-Test) zwischen den Generationenanzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen	57
Tabelle 11:	Signifikanzen (t-Test) zwischen den Gründertieranzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen	59
Tabelle 12:	Signifikanzen (t-Test) zwischen den Segmentanzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen.....	62
Tabelle 13:	Signifikanzen (t-Test) zwischen den Mutationsraten für den MSE-Wert.....	67
Tabelle 14:	Signifikanzen (t-Test) zwischen den einzelnen SNP-Anzahlen für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen	71

Tabelle 15:	Signifikanzen (t-Test) zwischen den Heritabilitäten für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen.....	74
Tabelle 16:	Darstellung Anzahl selektierter Tiere und Selektionsintensität für drei verschiedene Szenarien	84

Abbildungsverzeichnis

Abbildung 1:	Polymorphe Stellen in der DNA-Sequenz im Vergleich zweier Chromosomen für einen SNP, einen Mikro- und einen Minisatelliten (Wiederholungseinheiten sind farbig markiert) aus: Cichon et al. (2002) 7
Abbildung 2:	Genetische Variabilität in einem 100000 Basenpaare langen DNA-Sequenzabschnitt auf Chromosom 8. Die Striche unter dem DNA-Strang zeigen die Lokalisation von bekannten Mikrosatellitenmarkern (rot) und SNPs (schwarz) an, aus: Cichon et al. (2002) 8
Abbildung 3:	Darstellung der Genauigkeit der genomischen Zuchtwerte bei verschiedenen Anzahlen von Trainingsgenerationen für eine Population im Hardy-Weinberg-Gleichgewicht (a) und im Mutations-Drift-Gleichgewicht (b) für eine Heritabilität von 0,1 (Abbildungen auf der linken Seite) und 0,5 (Abbildungen auf der rechten Seite), aus: Muir (2007)..... 26
Abbildung 4:	Genauigkeit des genomischen Zuchtwertes bei 512 genotypisierten Tieren über vier Trainingsgenerationen und bei 1024 genotypisierten Tieren über zwei Generationen, aus: Muir (2007) 27
Abbildung 5:	Genauigkeiten des genomischen Zuchtwertes bei verschiedenen Genomgrößen bei 100 Markern für vier Trainingsgenerationen, $h^2 = 0,1$ und Populationen im Hardy-Weinberg- und Mutations-Drift-Gleichgewicht 27
Abbildung 6:	Genauigkeiten des genomischen Zuchtwertes für eine unterschiedliche Anzahl von Markern bei gleich bleibender Markerdichte von 10 Markern/cM für unterschiedliche Anzahlen an Trainingsgenerationen bei $h^2 = 0,1$ und einer Population im Hardy-Weinberg-Gleichgewicht (a) und einer im Mutations-Drift-Gleichgewicht (b), aus: Muir (2007) 28
Abbildung 7:	Genauigkeit für zufällige oder gerichtete Selektion auf den genomischen Zuchtwert nach vier Trainingsgenerationen für $h^2 = 0,1$ für eine Population im Hardy-Weinberg-Gleichgewicht (a) und eine im Mutations-Drift-Gleichgewicht (b), aus: Muir (2007) 29
Abbildung 8:	Genauigkeiten des genomischen Zuchtwertes für Eltern und Nachkommen für $h^2 = 0,5$ bei verschiedener Markerdichte und verschiedenen Haplotypenzusammenstellungen, aus Calus et al. (2008) .. 35
Abbildung 9:	Genauigkeiten des genomischen Zuchtwertes für Eltern und Nachkommen für $h^2 = 0,1$ bei verschiedener Markerdichte und verschiedenen Haplotypenzusammenstellungen, aus Calus et al. (2008) .. 36
Abbildung 10:	Generelle Struktur der Simulations- und Auswertungsprogramme..... 42

Abbildung 11:	Anzahl unterschiedlicher Haplotypen je Segment.....	49
Abbildung 12:	Durchschnittliche Anzahl des Vorkommens je Haplotyp über alle zehn Segmente.....	49
Abbildung 13:	Einfluss der Anzahl Elternpaare (EP) und Nachkommen auf die MSE-Werte.....	51
Abbildung 14:	Einfluss der Anzahl der Elternpaare (EP) und der Nachkommen auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert.....	54
Abbildung 15:	Einfluss der Anzahl der Generationen auf die MSE-Werte (mit SE) und der Korrelation zwischen dem wahren und dem genomischen Zuchtwert...	56
Abbildung 16:	Einfluss der Generationsgröße auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert.....	59
Abbildung 17:	Einfluss der Segmentanzahl auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und genomischen Zuchtwert.....	61
Abbildung 18:	Haplotypeneffekte für 50 Segmente (QTL-Positionen rot).....	63
Abbildung 19:	SNP-Effekte für Einzel-SNP-Analyse (QTL Positionen rot).....	64
Abbildung 20:	Einfluss der Mutationsrate auf den MSE-Wert mit SE.....	66
Abbildung 21:	Einfluss der SNP-Anzahl auf den MSE-Wert bei 100 und 200 SNPs für zwei verschiedene Segmentgrößen und Tierzahlen.....	68
Abbildung 22:	Einfluss der SNP-Anzahl auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert bei 100 und 200 SNPs für zwei verschiedene Segmentgrößen und Tierzahlen.....	70
Abbildung 23:	Einfluss der Heritabilität auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert.....	73
Abbildung 24:	Korrelation zwischen dem wahren und dem genomischen Zuchtwert für verschiedene Größen der Analysegeneration für $h^2 = 0,1$	75
Abbildung 25:	Einfluss der Anzahl der Elternpaare auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und die Folgegeneration.....	77
Abbildung 26:	Einfluss der Größe der Analysegeneration auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und vier Folgegenerationen.....	78
Abbildung 27:	Einfluss der Anzahl der Segmente auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und die Folgegeneration.....	80

Abbildung 28:	Einfluss der Heritabilität und der Anzahl der Nachkommen auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und Folgegeneration.....	82
Abbildung 29:	Einfluss der Größe der Analysegeneration auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und vier Folgegenerationen bei $h^2 = 0,1$	83
Abbildung 30:	Einfluss der zufälligen Selektion auf die Korrelation zwischen wahren und dem genomischem Zuchtwert für vier verschiedene Szenarien für die Analysegeneration und vier Folgegenerationen	85
Abbildung 31:	Einfluss der gerichteten Selektion auf die Korrelation zwischen wahren und dem genomischen Zuchtwert für vier verschiedene Szenarien für die Analysegeneration und vier Folgegenerationen.....	86
Abbildung 32:	Mittlere Varianz an den Genorten für $h^2 = 0,5$ (QTL-Positionen rot)	87
Abbildung 33:	Mittlere Varianz an den Genorten für $h^2 = 0,1$ (QTL-Positionen rot)	88
Abbildung 34:	Einfluss der gerichteten und zufälligen Selektion auf die Korrelation zwischen wahren und dem genomischen Zuchtwert für die Analysegeneration und vier verschiedene Szenarien	89
Abbildung 35:	Vergleich der Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die BLUP-Methode und das Resampling für zwei verschiedene Anzahlen an SNPs und Segmenten.....	90

1 Einleitung

Mit der Domestikation der einzelnen Nutztierarten wie Rind, Schwein, Schaf und Ziege vor ungefähr 10000 Jahren sowie des Huhnes vor ca. 4500 Jahren beginnt auch die Geschichte der Tierzucht. Seit dieser Zeit ist es ihr Ziel, die Leistungen der Tiere so anzupassen, dass sie den menschlichen Anforderungen entsprechen (Andersson, 2001). Die phänotypische Selektion durch Beobachtung und Erfahrung der Züchter war dabei über den längsten Zeitraum, neben der natürlich Selektion, das einzige Selektionskriterium (Andersson, 2001; Kräußlich, 1997).

Eine starke Entwicklung fand erst ab dem 20. Jahrhundert statt, was auf die Wiederentdeckung der Mendelschen Gesetze zurückzuführen ist. Mit Hilfe ihrer Anwendung wurde die Tierzucht zum ersten Mal auf ein naturwissenschaftlich begründetes Fundament gestellt. Durch fortschreitendes Verständnis der Genetik und Weiterentwicklung der Methoden konnten in den letzten Jahrzehnten große Fortschritte in der Tierzucht erzielt werden (Kräußlich, 1997). Bei vielen Leistungsmerkmalen, besonders Merkmalen mit mittlerer Heritabilität, wie z.B. Milch- und Fleischleistung, konnte ein bemerkenswerter Zuchtfortschritt erzielt werden.

Aber obwohl die Anwendung der klassischen Selektionsverfahren bei vielen Merkmalen zu großen Erfolgen führte, haben sie auch ihre Grenzen, da die Verfahren hauptsächlich additiv-genetische Effekte nutzen und die Selektionskriterien auf phänotypischen Eigen- oder Verwandtschaftsleistungen beruhen. Bei Merkmalen mit geringer Heritabilität, wenigen Beobachtungen, später Ausprägung, Schlachtmerkmalen, Krankheitsresistenzen sowie negativ korrelierten Merkmalen zeigen sich die Schwierigkeiten der bisherigen Selektionsverfahren. Eine züchterische Verbesserung solcher Merkmale ist mit den klassischen Methoden nur schwer realisierbar (Schwerin, 2003; Wolf, 2008).

Aufgrund der Globalisierung und des damit verbundenen Preisdrucks, aber auch durch sich ändernde Marktanforderungen wird ein weiterer Zuchtfortschritt in allen Bereichen der Tierzucht unausweichlich sein. Ansätze zur Lösung der aufgetretenen Probleme wurden durch die Entwicklung in der Molekulargenetik ermöglicht. Durch sie ist man heute in der Lage, das Genom zu analysieren, wodurch Einblicke in das Erbgeschehen und die Genstrukturen von Erbanlagen ermöglicht werden. Die Suche nach Bereichen des Genoms, die quantitative Merkmale wie z.B. Leistungsmerkmale beeinflussen, so genannte *Quantitative Trait Loci* (QTL), ist viel versprechend und führte in den letzten Jahrzehnten zur Lokalisation etlicher QTL (Schwerin, 2003; Wolf, 2008).

Genetische Marker, die in QTL-Nähe liegen, können somit als Selektionshilfsmittel in die Zucht einbezogen werden. Die Vorteile einer solchen markergestützten Selektion (MAS) liegen in der genauen Typisierungsmöglichkeit für jedes Tier unabhängig von der Merkmalsausprägung sowie Alter und Geschlecht. Mögliche Zuchttiere können so schon kurz nach ihrer Geburt bestimmt werden, ohne auf langwierige Nachkommentests zurückgreifen zu müssen (Mäki-Tanila, 2006; Schwerin, 2003). Aber trotz umfangreicher Forschungen auf diesem Gebiet waren die Zuchterfolge bisher eher gering. Verschiedene Faktoren wie z.B., Genauigkeit des bisherigen Zuchtwertes, Anteil der genetischen Varianz, die durch die DNA-Marker erklärt werden, oder Genauigkeit der geschätzten QTL-Effekte, können den Erfolg der MAS beeinflussen (Goddard & Hayes, 2007).

Ein weiterer Ansatz, Marker in die Zucht einzubeziehen, wurde erstmals von Meuwissen et al. (2001) vorgestellt. Bei der Methode der genomic selection werden eine sehr große Anzahl von Single Nucleotide Polymorphismen (SNP), die dicht über das Genom verteilt sind, gleichzeitig zur Selektion herangezogen. Man geht dabei von der Annahme aus, dass die gesamte genetische Varianz durch die SNPs erklärt wird und außerdem die QTL im Kopplungsungleichgewicht (LD) mit wenigstens einigen Markern sind (Goddard & Hayes, 2007).

Da für die meisten Tierarten mittlerweile eine sehr große Anzahl an SNPs zur Verfügung steht und die Kosten der Typisierung gering sind, sind die Hauptprobleme der genomic selection behoben. Vor einer Einführung der Methode sollte allerdings eine sorgfältige Bewertung der möglichen Vorteile erfolgen sowie die besten Einbindungsmöglichkeiten in Zuchtprogramme besprochen werden, um den größtmöglichen Nutzen aus der genomic selection zu erhalten.

In dieser Arbeit sollten für eine simulierte Schweinepopulation unter verschiedensten Bedingungen, wie z.B. unterschiedlicher Populationsgrößen, Heritabilitäten, Selektionsszenarien oder Haplotypengrößen mit der BLUP-Methode die Marker/Haplotypen-Effekte geschätzt und daraus genomische Zuchtwerte für jedes Tier berechnet werden. Weiterhin sollte die Genauigkeit der genomgestützten Zuchtwertschätzung bestimmt werden. Dadurch sollten die Vorteile, die die genomic selection bietet, aufgezeigt und die besten Einsatzmöglichkeiten dieser relativ neuen Methode herausgearbeitet werden.

2 Literaturübersicht

2.1 Entwicklung der Tierzucht

Die Entstehung der Haustiere aus den einzelnen Wildtierarten war eine rein menschliche Leistung. Durch die starke Verbesserung von Jagdwaffen, Jagdmethoden und Werkzeugen wurde die Nahrungsbeschaffung immer einfacher, was einerseits zu einem Anstieg der Bevölkerungszahlen, andererseits zu einer deutlichen Abnahme der natürlichen Nahrungsquellen führte. Die Menschen waren daraufhin gezwungen, sich neue Methoden des Nahrungserwerbs anzueignen und begannen Wildpflanzen zu kultivieren und Wildtiere als Haustiere zu halten. Diese Domestikation begann vor etwa 12000 Jahren, und seither ist es das Ziel, die verschiedenen Haustierarten an die menschlichen Anforderungen und Bedürfnisse anzupassen (Kräußlich, 1997).

Dabei kam der Tierhaltung im deutschsprachigen Raum über einen langen Zeitraum nur wenig Interesse entgegen. Züchtung, Fütterung und Haltung der Tiere spielten nur eine untergeordnete Rolle, wohingegen die Vermehrung im Vordergrund stand. Ein wachsendes Interesse an der Landwirtschaft und somit an der Tierhaltung trat im 18. Jahrhundert ein. Mit den wachsenden Bevölkerungszahlen nach dem Dreißigjährigen Krieg im 17. Jahrhundert ergab sich eine erhöhte und differenzierte Nachfrage nach landwirtschaftlichen Erzeugnissen. Eine zusätzliche Umstrukturierung der Gesellschaft durch die zunehmende Industrialisierung zu Anfang des 19. Jahrhunderts und der damit einhergehenden Verstädterung und somit Lösung von der landwirtschaftlichen Selbstversorgung taten ihr Übriges. Eine züchterische Entwicklung der teils wenig leistungsfähigen Tierarten war unumgänglich. Aber trotz eines starken Befassens mit genetischen Vorgängen in der folgenden Zeit trat ein großer Entwicklungsschub in der Haustierzucht erst im 20. Jahrhundert ein (Comberg, 1984).

Erste Leistungserfassungen um 1900 steigerten das Interesse am Vererbungsgeschehen der Leistungsmerkmale, und mit der Anwendung der Mendelschen Gesetze war es möglich das Erbgeschehen der qualitativen Merkmale zu erfassen, zu erklären und der Zucht ein weites Betätigungsfeld zu eröffnen. Bei quantitativen Merkmalen zeigte sich jedoch sehr bald, dass mit dieser Methode der erhoffte Erfolg nicht realisiert werden konnte (Comberg, 1984).

Hier stellten die ermittelten Leistungsdaten, die für die Selektion verwendet werden sollten, den Ausgangspunkt zur Erfassung des Zuchtwertes dar. Für die Anwendung in der Tierzucht eigneten sich Modelle aus der Populationsgenetik am besten. Bald gab es erste Versuche, die Leistung von Tieren miteinander zu vergleichen. Verschiedene Verfahren wurden entwickelt, erweitert und verbessert, um die Leistung der einzelnen Tiere zu bewerten und den Zuchtwert zu schätzen (Comberg, 1984). Eine viel versprechende Methode war der Selektionsindex, der lange in der Tierzucht eingesetzt wurde. Er führte erstmals zu einer Gewichtung verschiedener Informationsquellen nach Verwandtschaft und Beziehung des gemessenen Merkmals zum Zuchtwert. Somit konnten erstmalig nicht nur Eigenleistungen verwendet werden. Außerdem war die Berücksichtigung mehrerer Merkmale in der Selektion möglich. Nachteilig war, dass die Methode normalverteilte Beobachtungen voraussetzte und die Einbeziehung von großen Mengen von Verwandteninformationen sehr aufwendig war. Des Weiteren war die Korrektur systematischer Umwelteinflüsse nicht gegeben, wodurch die Ergebnisse nur innerhalb einer Gruppe aussagekräftig waren, die unter den gleichen Umweltbedingungen gehalten wurde (Littmann et al., 2006).

Eine Weiterentwicklung des Selektionsindex stellt die BLUP-Methode dar, welche schon 1949 von Henderson entwickelt wurde, aber erst in den 70er Jahren publiziert (Henderson, 1975) und deren Einführung aufgrund des großen Rechenaufwandes bis in die 80er Jahre dauerte. Heute wird BLUP weltweit bei allen Nutztierspezies mit großem Erfolg eingesetzt und ist die Methode der Wahl für die Zuchtwertschätzung. Bei dieser Methode werden fixe und zufällige Effekte gleichzeitig geschätzt, und die Einbeziehung aller Verwandteninformationen stellt kein Problem mehr dar, sodass alle Tiere einer Population miteinander verglichen werden können.

Die bisher angewendeten Methoden der quantitativen Tierzucht haben zu einer extremen Verbesserung der Leistungsmerkmale geführt. Gerade bei Merkmalen mit mittlerer und hoher Heritabilität konnten sehr gute Fortschritte erzielt werden. Die Grenzen dieser Verfahren zeigen sich bei Merkmalen mit niedriger Heritabilität, wie z.B. Konstitutionsparametern, Langlebigkeit und Fruchtbarkeit, die in den letzten Jahren aus wirtschaftlichen Gründen vermehrt in den Vordergrund getreten sind, bei Merkmalen mit nur wenigen Beobachtungen (teure, aufwendige Erfassung), bei Merkmalen mit später Ausprägung, bei Schlachtmerkmalen (Erfassung der Merkmale erst nach der Schlachtung möglich), bei Krankheitsresistenzen sowie bei negativ korrelierten Merkmalen. Eine züchterische Verbesserung solcher Merkmale ist mit den klassischen Methoden nur schwer realisierbar (Muir, 2007; Schwerin, 2003; Wolf, 2008). Auch in Zukunft wird genetischer Fortschritt unumgänglich sein. Dieser wird durch verschiedene Faktoren wie

Selektionsintensität, Genauigkeit der Zuchtwertschätzung, genetische Varianz der Zuchtpopulation sowie das Generationsintervall bestimmt (Wolf, 2008).

Die Entwicklung in der Molekulargenetik mit ihrer Identifizierung von Genen und genetischen Markern schafft die Möglichkeit, komplexe Mechanismen, die Leistungs- und Gesundheitsmerkmale beeinflussen, zu analysieren und damit neue Merkmale für die Auswahl von Zuchttieren zu definieren (Wolf, 2008). Dadurch, dass Marker vererbbar sind und gegeben durch ihre mögliche unmittelbare Lage zu Genen oder QTLs gekoppelt mit diesen vererbt werden können, bietet eine Selektion unter Einbeziehung genetischer Marker eine Reihe von Vorteilen (Meuwissen, 2003; Wolf, 2008):

- Verkürzung des Generationsintervalls: Durch Berücksichtigung von Markerinformationen kann eine Selektion von Tieren vor der Merkmalsausprägung oder bereits von Embryonen durchgeführt werden. Nachkommentests müssen nicht mehr abgewartet werden. Das Generationsintervall kann somit beträchtlich verkürzt werden.
- Erhöhung der Selektionsgenauigkeit: Bei Merkmalen, bei denen die Selektionsgenauigkeit relativ niedrig ausfällt, wie Merkmale mit niedriger Heritabilität, wenigen Messungen, später Erfassung und Schlachtmerkmalen führt die Einbeziehung von Markerinformationen zu einer höheren Genauigkeit.
- Erhöhung der Selektionsintensität: Die Kosten zur Markeridentifikation sind in den letzten Jahren immer weiter gesunken und eine immer größere Anzahl an Markern kann gleichzeitig analysiert werden. Eine größere Anzahl von Tieren kann so analysiert werden und steht für einen Selektionsprozess zur Verfügung.

Außerdem kann mit den Techniken der Molekulargenetik die genetische Vielfalt in Nutztierpopulationen erfasst und eine sichere Identifizierung von Individuen ermöglicht werden.

2.2 Genetische Marker

Als genetische Marker, oder auch nur Marker, werden in der Molekularbiologie eindeutig identifizierbare, vererbare, kurze DNA-Abschnitte, deren Ort im Genom bekannt ist, bezeichnet. Grundsätzlich gibt es Marker in kodierenden und nichtkodierenden Regionen. Marker in kodierenden Regionen, also Genen, werden als „direkte“ Marker bezeichnet und sind von großer Bedeutung, da die Polymorphismen des Markers die Variation des Merkmals bestimmen. Marker in nichtkodierenden Regionen können allenfalls sehr nah an Genen liegen, weshalb sie gekoppelt mit den Genen vererbt werden können. Nur solche mit dem Gen gekoppelten Marker sind für die Einbeziehung in die Selektion geeignet. Je enger die Marker mit dem Gen gekoppelt vorkommen, desto besser wird ihre Aussagekraft in Bezug auf den Phänotyp des Tieres, da man über die verschiedenen Allele des Markers auf die Allele des Genes schließen kann (Williams, 2005). Es werden verschiedene Typen von Markern unterschieden:

Restriktionsfragment-Längen-Polymorphismen

Restriktionsfragment-Längen-Polymorphismen (RFLPs) waren die ersten Marker, die in der Genetik benutzt wurden. Dabei werden Restriktionsenzyme verwendet, die jeweils auf eine spezifische Erkennungssequenz in der DNA reagieren und eine Schnittstelle für die Restriktionsenzyme darstellen. Die Länge der Restriktionsfragmente wird durch Variationen in der DNA-Sequenz und durch Mutationen bestimmt, bei denen eine Erkennungssequenz für ein Restriktionsenzym entsteht oder verloren geht. Auf diese Weise entstehen unterschiedliche Längen an Restriktionsfragmenten, die verglichen werden können (Montaldo & Meza-Herrera, 1998; Williams, 2005).

Minisatelliten

Minisatelliten, auch „variable number of tandem repeats“ (VNTRs), werden erstmals Mitte der 80er Jahre beschrieben. Sie zählen zu den repetitiven Sequenzen, zeichnen sich also durch sich mehrfach wiederholende Sequenzmotive aus, die in der Regel in nichtkodierenden Sequenzen liegen. Sie haben eine Länge von 12 bis 500 Basenpaaren und zeigen zusammen mit den Mikrosatelliten eine große Variabilität. In Abbildung 1 sind polymorphe Stellen für einen Minisatelliten dargestellt. Der 13 Basenpaare lange Minisatellit kommt auf dem oberen Chromosom nur einmal, auf dem unteren Chromosom

zweimal vor. Der Großteil der Populationen ist für die meisten der Minisatelliten heterozygot. Der Nachteil ist die relative Größe, die die Amplifikation der DNA-Abschnitte schwierig gestaltet sowie die ungleichmäßige Verteilung. Die höchste Dichte findet man in den telomerischen Bereichen (Cichon et al., 2002; Montaldo & Meza-Herrera, 1998).

Mikrosatelliten

Mikrosatelliten oder „short tandem repeats“ (STRs) sind DNA-Sequenzen von tandemartig wiederholten einfachen Basenmotiven mit 10-50 Kopien. Sie bestehen aus zwei bis elf Basenpaaren von variabler Basenzusammensetzung. In Abbildung 1 kann man erkennen, dass der Mikrosatellit aus CA-Wiederholungen im oberen Chromosom aus nur fünf Wiederholungen besteht, während im unteren sieben Wiederholungen vorliegen. Mikrosatelliten bieten den Vorteil, gleichmäßig und in großer Zahl, ca. 50 000 bis 100 000, über das Genom verteilt zu sein (sie machen etwa 0,5 % des Gesamtgenoms aus), liegen aber wie die Minisatelliten in der Regel in nichtkodierenden Bereichen. Purin/Pyrimidinbasen Wiederholungen wie CA, AT, GC sind häufig. Besonders CA-Wiederholungen, welche mit 80 000 die häufigste Form darstellen, haben eine sehr hohe Frequenz und werden als Marker verwendet (Knippers, 2001; Montaldo & Meza-Herrera, 1998; Wolz, 1993).

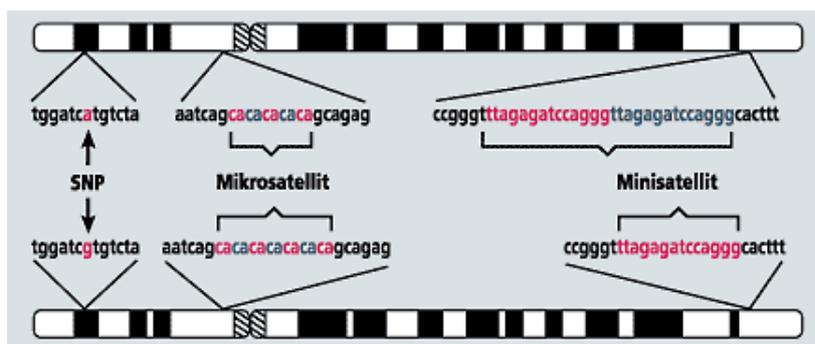


Abbildung 1: Polymorphe Stellen in der DNA-Sequenz im Vergleich zweier Chromosomen für einen SNP, einen Mikro- und einen Minisatelliten (Wiederholungseinheiten sind farblich markiert) aus: Cichon et al. (2002)

SNPs

Mit SNP „single nucleotide polymorphisms“ werden Variationen von einzelnen Basenpaaren im DNA-Strang bezeichnet. In Abbildung 1 ist ein Austausch einer Base dargestellt. Im oberen Chromosom befindet sich an der rot markierten Stelle die Base Adenin, während auf dem unteren Chromosom Guanin zu finden ist.

Der Austausch einer Base gilt erst als polymorph, wenn die Frequenz innerhalb der Population bei 1 Prozent oder mehr liegt. Darunter spricht man von einer Mutation. Man geht davon aus, dass es im humanen Genom ca. 11 Millionen SNPs gibt, deren minores Allel eine Frequenz von wenigstens einem Prozent in der Weltbevölkerung hat. Solche SNPs kommen durchschnittlich wahrscheinlich alle 290 Basenpaare vor (Bertsch & Neumaier, 2003; Cichon et al., 2002).

SNPs sind mit Abstand der häufigste Typ genetischer Variabilität und machen etwa 90 Prozent der interindividuellen genetischen Variabilität aus und treten im Durchschnitt etwa alle 200 bis 300 Basenpaare auf. In Abbildung 2 ist die genetische Variabilität auf einem 100 000 Basenpaare langen DNA-Abschnitt dargestellt. Die Striche unterhalb des DNA-Stranges zeigen die bekannten Mikrosatelliten (rot) und SNPs (schwarz). Deutlich ist die sehr viel größere Häufigkeit und regelmäßigere Verteilung der SNPs (40 in diesem Bereich) im Vergleich zu lediglich 3 bekannten Mikrosatelliten (Aitken et al., 2004; Cichon et al., 2002; Williams, 2005).

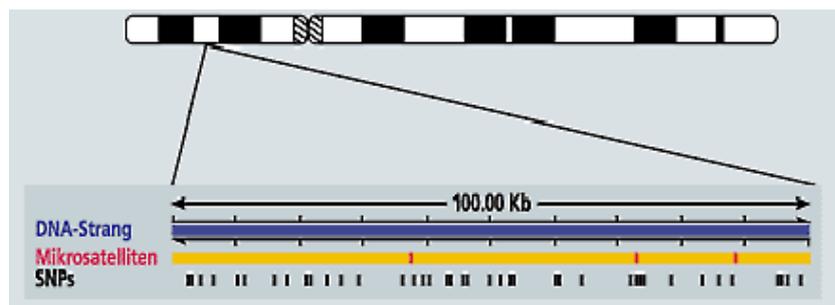


Abbildung 2: Genetische Variabilität in einem 100000 Basenpaare langen DNA-Sequenzabschnitt auf Chromosom 8. Die Striche unter dem DNA-Strang zeigen die Lokalisation von bekannten Mikrosatellitenmarkern (rot) und SNPs (schwarz) an, aus: Cichon et al. (2002)

SNPs können in kodierenden und nichtkodierenden Bereichen des DNA-Stranges auftreten. Dabei zeigt sich eine geringere Nukleotid-Diversität in den kodierenden Bereichen im Vergleich zu den nichtkodierenden, was den Erwartungen entspricht, dass auf der Variabilität in kodierenden Bereichen ein größerer negativer Selektionsdruck lastet. Innerhalb der kodierenden Bereiche zeigt sich für die einzelnen Basen eine unterschiedliche Variabilität. So weiß man seit längerem, dass der durch Basentriplets definierte genetische Code degeneriert ist, da bekannt ist, dass mehrere Triplets ein und dieselbe Aminosäure kodieren. So kodieren sehr häufig die ersten beiden Basen in einem Codon sehr spezifisch für eine Aminosäure (nichtdegenerierte Positionen); Mutationen an diesen Positionen bewirken einen Aminosäureaustausch. An der dritten Stelle im Codon sind dagegen häufig mehrere Basen möglich, um die gleiche Aminosäure zu erhalten (degenerierte Position). Mutationen an diesen Stellen ändern die Aminosäure dann nicht oder nur selten. Aus Untersuchungen am humanen Genom geht hervor, dass die Variabilität an nichtdegenerierten Positionen geringer ausfällt als an degenerierten Positionen. Mutationen an Stellen, die einen Aminosäurenwechsel nach sich ziehen führen wohl häufig zu einem Selektionsnachteil, weil das resultierende Protein seine normale Funktion ganz oder zum Teil verliert. Die Mutation kann sich somit in der Population nicht festigen und geht verloren.

Für die Tierzucht relevant sind solche SNPs in kodierenden Regionen, welche einen Einfluss auf den Phänotyp des zu selektierenden Merkmals haben.

Der Großteil der SNPs in nichtkodierenden Bereichen wird mit großer Wahrscheinlichkeit keine phänotypischen Auswirkungen haben. Für die Tierzucht wären hier wieder nur solche SNPs von Bedeutung, welche eng gekoppelt mit Genen oder QTL-Regionen sind, sodass Aussagen über den Einfluss auf den Phänotyp mit Hilfe des Markers zu treffen sind (Cichon et al., 2002).

2.3 Markergestützte Selektion

Selektions- und Anpaarungsstrategien in Zuchtprogrammen beruhten bis in die Anfänge der 90er Jahre auf genetischen Schätzern (Zuchtwerten), die anhand der phänotypischen Leistung ermittelt wurden. Diese phänotypischen Merkmale werden von genetischen und umweltbedingten Faktoren beeinflusst. Durch die Entdeckung der DNA (Watson & Crick, 1953) und Techniken zum einfachen Lesen der Basenabfolge, mittels der Polymerase Kettenreaktion (Saiki et al., 1988) konnten bestimmte Bereiche als genetische Marker identifiziert werden. Durch Restriktionsfragment-Längen-Polymorphismen (RFLPs)

(Botstein et al., 1980) und Mikrosatelliten konnten genetische Unterschiede zwischen Tieren direkt auf dem Level der DNA gemessen werden (Meuwissen & van Arendonk, 1992). Diese Unterschiede stellen nur in seltenen Fällen QTLs dar. In der Regel liegen sie in deren Nähe, sodass sie einen QTL markieren. Man spricht in solchen Fällen von einem gekoppelten Marker (Soller, 1978).

Ist eine Kopplung zwischen einem Marker und einem QTL bekannt, kann man die Information über den Marker in das Zuchtprogramm einbinden. Die Selektion beruht dann nicht nur auf phänotypischen Leistungen und quantitativ genetischen Schätzwerten, sondern zusätzlich auf Informationen von Markern, die an QTLs gekoppelt sind. In diesem Fall spricht man von markergestützter Selektion (MAS) (Lande & Thompson, 1990). Die Markerinformationen können im Gegensatz zu vielen phänotypischen Leistungen bereits beim Jungtier gemessen werden. Das bietet einen Vorteil der Selektionsentscheidung am juvenilen Tier (Woolliams & Smith, 1988).

Der Einsatz genetischer Marker erfordert vier Schritte (Meuwissen & Van Arendonk, 1992):

1. Suche nach genetischen Markern innerhalb Tierart und Rasse
2. Erfassung der Kopplung zwischen Markern
3. Bestimmung der Assoziation zwischen Markern und QTL
4. Einbindung der Markerinformation in das Zuchtprogramm

Die genetischen Marker können von drei unterschiedlichen Arten sein (Dekkers, 2004): Direkte Marker, linkage disequilibrium (LD) Marker und linkage equilibrium (LE) Marker. Die Bestimmung der Marker wurde bei Andersson (2001) beschrieben.

Direkte Marker liegen innerhalb eines Gens und kommen somit selten vor. LE Marker sind Marker, die innerhalb der Population nicht in LD vorliegen. Sie werden auf genomweiter Ebene ermittelt, anhand von Kreuzungen oder großen Halbgeschwisterfamilien innerhalb einer Rasse. Der genomweite Scan erfordert mittlere bis geringe Markerdichten (15 – 50 cM Abstand) (Darvasi & Soller, 1994).

LD Marker werden für Feinkartierungen genutzt, um die funktionelle Mutation zu lokalisieren (Andersson, 2001; Rothschild & Soller, 1997) bzw. ein Gen zu identifizieren, und liegen in den untersuchten Bereichen 1-5 cM auseinander (Dekkers, 2004).

Diese genetischen Marker haben alle gemeinsam, dass sie im Gegensatz zu heutigen SNPs, weit auseinander liegen, relativ selten vorkommen und unterschiedlich polymorph sind. Der Informationsgehalt der Marker kann stark schwanken, und die Entwicklung der einzelnen Marker ist aufwendig und teuer. Der große Aufwand lohnt sich jedoch bei der Einbindung der Markerinformationen in Zuchtprogramme (Dekkers, 2004).

Die zusätzliche Effizienz der MAS beruht auf der Kombination von molekularer und phänotypischer Information verglichen mit den Standardmethoden der Selektion, welche auf dem Phänotyp beruhen und stark von der Heritabilität des Merkmals, dem Anteil durch genetische Marker erklärter additiv genetischer Varianz und dem Selektionsschema abhängig sind (Lande & Thompson, 1990). Am größten ist die Effizienz unter Individualelektion, wenn die Heritabilität des Merkmals gering und der durch Marker erklärte Teil der genetischen Varianz mittel bis hoch ist (Lande & Thompson, 1990).

Molekulare Marker haben in den seltensten Fällen größere Effekte auf ökonomisch wichtige Merkmale. Die MAS beruht allein auf Kopplungseffekten zwischen Marker und QTL, welche die Effizienz je nach Kopplungsgrad verringern. Der Erfolg von MAS-Programmen ist von vielen verschiedenen Faktoren abhängig. Eine ausführliche Übersicht ist bei Davis und DeNise (1998) aufgeführt. Dazu gehören Populationsgröße, Ausmaß des LD, Anzahl selektierter Generationen, Heritabilität, Anzahl an Markern, Höhe der QTL-Effekte uvm. (Goddard & Hayes, 2007).

Die Berücksichtigung der Markerinformation erfolgt über die Einbindung in einen Selektionsindex. Dekkers (2007) beschreibt die Erweiterung des Standardmodells

$$P=G+E$$

auf

$$P=Q+R+E$$

mit

P= Phänotyp

G= genetische Komponente

E= umweltbedingte Komponente

Q= Anteil der genetischen Komponente, der durch Marker erklärt wird

R= polygener Effekt

Dadurch lässt sich der herkömmliche Selektionsindex (Hazel, 1943) auf Markerinformationen ausweiten. Der von Lande und Thompson (1990) vorgeschlagene Selektionsindex ist bei Dekkers (2007) so erweitert worden, dass er in Standardverfahren und –programmen (z.B. SelAction, Rutten et al., 2002) verwendet werden kann.

Trotz der erhöhten Informationsmenge der MAS mit ihrer höheren Sicherheit und möglicher Selektion von Jungtieren hat dieses Verfahren auch seine Nachteile. Die Marker für MAS können mit dem QTL gekoppelt, jedoch im Kopplungsgleichgewicht sein, wodurch die Marker weniger informativ werden. Das kann dazu führen, dass die Anzahl der zu typisierenden Marker stark steigt, da man gezwungen ist, noch zusätzliche Marker

zur Analyse hinzuzunehmen, ohne dass der Selektionserfolg in gleichem Maße steigt (Boichard et al., 2006).

In Simulationsstudien wurde der Vorteil der MAS gegenüber klassischer Selektion über mehrere Generationen berechnet (Schulman & Dentine, 2005). Dort wurde dargestellt, dass die MAS einen größeren Selektionserfolg aufweist, aber nach mehreren Generationen wieder zurückgeht. Besonders bemerkenswert war der Rückgang an genetischer Varianz mit MAS, der so nicht zu erwarten war. Der QTL muss aber einen hohen Effekt haben, um den Selektionserfolg geringfügig zu erhöhen, was in der Praxis nicht zu erwarten sei. MAS hat sich bis heute nur in geringen Maße durchgesetzt (Schulman & Dentine, 2005).

2.4 Genomic selection – Grundlegende Methoden

Die genomic selection hat in den letzten Jahren extrem an Bedeutung gewonnen und wurde zu einem intensiven Betätigungsfeld. Die Methode wurde erstmals von Meuwissen et al. (2001) präsentiert. Der Ausdruck „genomic selection“ wurde allerdings schon 1998 von Haley und Visscher auf dem World Congress on Genetics Applied to Livestock (WCGALP) in Armidale eingeführt, die ihn auf die gleiche Art und Weise wie Meuwissen et al. (2001) in ihrem Paper verwendet haben (Meuwissen, 2007).

Die Methode der genomic selection ist in gewisser Hinsicht eine markergestützte Selektion auf genomweiter Ebene. Sie wurde entwickelt, um die Nachteile der markergestützten Selektion zu umgehen und weiteren genetischen Fortschritt zu erzielen (Goddard & Hayes, 2007; Meuwissen, 2007). Ausgehend von der Annahme, dass Zuchtwerte mit Hilfe von Informationen aus DNA-Sequenzvariationen zwischen einzelnen Tieren mit größerer Genauigkeit geschätzt werden können, wurden umfangreiche Untersuchungen bezüglich der markergestützten Selektion unternommen. Die Anwendung dieser Methode war beschränkt und der daraus gezogene Nutzen sehr gering. Verschiedene Faktoren beeinflussen den zusätzlichen Nutzen der markergestützten Selektion (Goddard & Hayes, 2002; 2007):

- Genauigkeit der existierenden Zuchtwerte. Sind die Genauigkeiten schon sehr hoch, kann nur wenig zusätzliche Genauigkeit hinzugewonnen werden.
- Anteil der genetischen Varianz, der durch die DNA-Marker erklärt wird.
- Genauigkeit, mit der die Effekte der QTL-Allele geschätzt werden.
- Möglichkeit, das Generationsintervall durch markergestützte Selektion zu verkürzen.

Da die meisten Merkmale durch viele Gene beeinflusst werden, wird bei Auffinden von nur einigen Genen durch Marker nur ein kleiner Teil der genetischen Varianz erklärt. Außerdem haben einzelne Gene nur einen geringen Effekt, wodurch eine große Menge an Daten benötigt wird, um ihre Effekte genau zu bestimmen. Verschlimmert wird dieses Problem noch bei Verwendung von Markerhaplotypen, um QTL aufzufinden, da hier viele Haplotypeneffekte bestimmt werden müssen (Goddard & Hayes, 2007; Meuwissen, 2003).

Die Marker, welche für die markergestützte Selektion benutzt werden, können mit dem QTL gekoppelt sein, aber trotzdem im Kopplungsgleichgewicht mit ihm sein. Wenn das der Fall ist, gelten alle QTL-Allele in Gründertieren als unterschiedlich und die Anzahl an QTL-Allelen, deren Effekte geschätzt werden müssen, ist nochmals erhöht (Goddard & Hayes, 2007).

Um diese Schwierigkeiten zu lösen, wurden von Meuwissen et al. (2001) eine Variante der markergestützten Selektion vorgestellt, welche sie genomic selection nannten. Bei dieser Methode ist eine Vielzahl an Markern dicht über das gesamte Genom verteilt. Hier wird insgesamt von mehreren Zehn- oder Hunderttausenden von Markern ausgegangen, sodass die Annahme besteht, dass alle Marker zusammen die gesamte genetische Varianz erklären und alle Gene, die das zu untersuchende Merkmal beeinflussen, mit zumindest einigen Markern im LD stehen. Alle Markereffekte werden simultan geschätzt. Für Marker bzw. Genomregionen mit hohem bzw. kleinem Einfluss erwartet man hohe bzw. kleine Schätzwerte, sodass direkt auf diese selektiert werden kann (Goddard & Hayes, 2007; Meuwissen, 2003; Meuwissen et al., 2001).

Die größten Hindernisse bei der Einführung der genomic selection lagen zu Anfang in der immensen Anzahl von benötigten Markern sowie ihren Typisierungskosten. Durch die enorme Entwicklung in der Genotypisierungs-Technologie sind die Kosten stark gesunken, und große Sätze an SNPs können routinemäßig erfasst werden (de Koning et al., 2007; Goddard & Hayes, 2007; Meuwissen, 2007).

Die Berechnung der genomischen Zuchtwerte läuft in drei Schritten ab. Die Marker werden dazu benutzt, um den Genotyp für jedes Tier an jedem QTL zu bestimmen. Der einfachste Weg hierzu ist, die Marker als QTL zu behandeln. Daraufhin werden die Effekte der Chromosomensegmente bzw. der einzelnen Marker anhand einer Referenzpopulation geschätzt. In einem letzten Schritt wird der genomische Zuchtwert für die Tiere nach der folgenden Formel berechnet (Goddard & Hayes, 2007; Hayes, 2008):

$$gZW = \sum_i^n x_i \hat{g}_i$$

n = Anzahl der Chromosomsegmente/Marker über das Genom

x_i = Vektor, welcher die Tiere mit den Chromosom-/Markereffekten verknüpft

\hat{g}_i = Vektor der Haplotyp-/Markereffekte innerhalb des Chromosomensegments

Für die Berechnung des Zuchtwertes werden die einzelnen Chromosomensegmente bzw. Marker, die jedes Tier trägt, bestimmt und der jeweilige Effekt des Chromosomensegments bzw. des Markers aufsummiert.

Der entscheidende Parameter bei der Schätzung der Markereffekte ist der Anteil der QTL-Varianz, der durch die Marker erklärt wird. Dieser Anteil ist abhängig von der Kopplung zwischen einem QTL und einem Marker, der hoch variabel ausfällt. Mit steigendem Abstand zwischen Marker und QTL sinkt die Kopplung und somit der Anteil an erklärter QTL-Varianz. Die Verwendung vieler Marker, um hohe Markerdichten zu erzeugen, ist daher notwendig (Goddard & Hayes, 2007).

Das Problem im ersten Schritt der Methode stellt die extrem große Anzahl von Markern dar. Bei einer Verwendung von mehreren Zehn- oder Hunderttausenden von Markern und nur einigen tausend typisierten Tieren soll eine Vielzahl von Markereffekten mit Hilfe einer geringen Menge von Leistungsdaten geschätzt werden, was eine starke Überparametrisierung zur Folge hat (Meuwissen, 2003).

Insgesamt werden verschiedene Verfahren vorgeschlagen, um Marker- oder Haplotypeneffekte zu berechnen (Hayes, 2008; Meuwissen, 2003). Ein Hauptunterschied der Methoden ist die Annahme über die Varianz der Haplotyp- oder Markereffekte über das Genom. Die einfachere Annahme ist, dass die Varianz der Haplotyp-/Markereffekte über alle Segmente gleich ausfällt. Da bekannt ist, dass in einigen wenigen Segmenten QTLs mit großen Effekten vorkommen, während in anderen QTLs mit geringem Einfluss oder gar keine QTLs vorhanden sind, gehen andere Methoden davon aus, dass die Varianz der einzelnen Segmente unterschiedlich ausfällt.

2.4.1 Least squares

Die "Methode der kleinsten Quadrate" (Least Squares) geht auf Legendre und Gauss zurück und ist der älteste Ansatz zur Schätzung von Populationsparametern. Das Ziel dieser Methodik ist, die Summe der quadrierten, nicht erklärbaren Resteffekte zu minimieren und diejenigen Schätzwerte der gesuchten Parameter auszuwählen, die diese Vorgabe am besten erfüllen (Kräußlich, 1997).

Beim Least Squares Ansatz gibt es keine Annahme hinsichtlich unterschiedlicher Chromosomensegment-Effekte, da hierbei alle Effekte als fix betrachtet werden. In einem ersten Schritt wird hier für jedes Segment eine Regressionsanalyse durchgeführt (Hayes, 2008):

$$y = \mu 1_n + X_i g_i + e$$

y = Datenvektor

μ = Mittel

1_n = Vektor von n Einsen (n = Anzahl Daten)

g_i = genetische Effekte der Haplotypen am i -ten Chromosomensegment

X_i = Designmatrix für das i -te Chromosomensegment

e = Resteffekt

Anschließend wird eine bestimmte Anzahl (m) der signifikanten Segmente ausgewählt und ihre Effekte simultan über eine multiple Regressionsanalyse geschätzt (Hayes, 2008):

$$y = \mu 1_n + \sum_m X_i g_i + e$$

Die Summe bezieht sich hier auf alle signifikanten QTL-Positionen. Sämtliche andere Haplotypeneffekte werden auf null gesetzt.

Der Least Squares Ansatz hat zwei wesentliche Nachteile. Zum einen muss ein Signifikanzniveau gewählt werden, welches nicht zu gering ausfällt, da andererseits die Anzahl an signifikanten Segmenten zu hoch ausfällt und die Anzahl an Phänotypdaten übersteigt. Andererseits führt die Auswahl der entsprechenden Segmente und ihrer

erneuten Effektschätzung zu einer Überschätzung der Effekte (Goddard & Hayes, 2007; Hayes, 2008).

2.4.2 BLUP

Die „best linear unbiased prediction“ – Methodik (BLUP) wurde im Wesentlichen 1949 von Henderson entwickelt. Sie beinhaltet die gemeinsame Schätzung von fixen Effekten und Zuchtwerten in so genannten gemischten Modellen (Kräußlich, 1997). Die Eigenschaften werden schon zum Teil durch den Namen ausgedrückt:

- B: Best** die Korrelation zwischen dem wahren und dem geschätzten Zuchtwert wird maximiert, bzw. die Varianz des Schätzfehlers minimiert
- L: Linear** die Schätzformel ist eine lineare Funktion der Beobachtungen
- U: Unbiased** die Schätzwerte, Zuchtwerte wie auch fixe Effekte sind unverzerrt
- P: Prediction** beinhaltet die Schätzung der Zuchtwerte

Die BLUP-Zuchtwertschätzung wird heute weltweit bei allen Nutztierspezies wegen ihrer umfassenden statistischen Möglichkeiten eingesetzt. Durch fortwährende Entwicklung der Computerkapazitäten war man in der Lage, von anfangs sehr einfachen Einmerkmalsmodellen mit Vätern auf komplexe Tiermodelle zu erweitern (Mrode, 1996).

Für die Berechnungen wird ein lineares gemischtes Modell benötigt:

$$y = Xb + Zu + e$$

y = Vektor der Beobachtungswerte

b = Vektor der unbekannt, fixen, zu schätzenden Parameter

u = Vektor der unbekannt, zufälligen, zu schätzenden Parameter

e = Vektor der Resteffekte

X = Designmatrix für fixe Effekte

Z = Designmatrix für zufällige Effekte

Die Erwartungswerte der Variablen sind:

$$E(y) = Xb; \quad E(u) = E(e) = 0$$

Die Varianzen sind:

$\text{var}(e) = I\sigma_e^2 = R$, $\text{var}(u) = A\sigma_g^2 = G$ und $\text{cov}(u,e) = 0$, wobei A die Verwandtschaftsmatrix darstellt

Die geschätzten Lösungen für \hat{u} und \hat{b} ergeben sich aus den Mischmodellgleichungen (mixed model equations = MME) (Henderson, 1950)

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

Die Voraussetzung ist, dass weder R noch G singulär sind und $R = I\sigma_e^2$. Dadurch, dass die zufälligen Effekte und die Resteffekte unkorreliert sind, vereinfacht sich die MME zu:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

mit $\lambda = \sigma_e^2/\sigma_g^2$, bzw. $(1-h^2)/h^2$

Zur Lösung des Gleichungssystems werden die Inverse der Verwandtschaftsmatrix und später die Inverse der Koeffizientenmatrix benötigt. Da bedingt durch die Größe der Matrizen eine direkte Invertierung nicht möglich ist, muss auf andere Methoden zurückgegriffen werden. Da bei der Verwandtschaftsmatrix immer gleiche Werte vorkommen, wurde schon 1976 von Henderson ein Algorithmus zur direkten Aufstellung der inversen Verwandtschaftsmatrix entwickelt (Henderson, 1976). Dieses Verfahren ist für die Koeffizientenmatrix nicht realisierbar. Eine geeignete Möglichkeit zur Lösung stellt die Iteration, hier Gauss-Seidel-Iteration, dar (Mrode, 1996).

2.4.2.1 Die Modelle der BLUP-Zuchtwertschätzung

Die BLUP-Zuchtwertschätzung ermöglicht durch den relativ einfachen Aufbau ihrer Mischmodellgleichungen die Berechnung von Zuchtwerten für die unterschiedlichsten Datenstrukturen. Die vorkommenden Leistungsdaten und Merkmale können aufgrund des unterschiedlichen Ausgangsmaterials wie Tierart, Verwandtschaftsstruktur, Zuchtprogramm, Produktionssystem, Leistungsprüfung usw. sehr unterschiedlich sein. Der Vorteil der BLUP-Zuchtwertschätzung liegt darin, dass sie ein universelles Werkzeug für alle Bereiche der Tierzucht darstellt. Im Folgenden sollen kurz die gängigsten Modelle vorgestellt werden, wobei man grundsätzlich von zwei Unterschieden ausgehen kann:

- Unterschiede im genetischen Aufbau (Tiermodell – Vatermodell)
- Unterschiede in der Datenstruktur (Einmerkmals-, Mehrmerkmals-, Wiederholbarkeitsmodell)

Das Tiermodell

Das Tiermodell ist das am meisten genutzte Modell in der Zuchtwertschätzung. Es ermöglicht die Schätzung der Zuchtwerte für alle im Pedigree vorhandenen Tiere, wobei auch Tiere ohne eigene Leistung aufgrund ihrer Verwandtschaft zu anderen Tieren einen Zuchtwert erhalten. In diesem Modell wird für jedes Tier und jeden Effekt eine Gleichung aufgestellt, wodurch bei großen Populationen das System sehr schnell sehr groß werden kann.

Das Vatermodell

Das Vatermodell wurde schon sehr früh in der Praxis, besonders in der Rinderzucht, eingesetzt. Der große Vorteil liegt in der reduzierten Anzahl von Gleichungen im Vergleich zu dem Tiermodell, da nur für die Väter ein Zuchtwert geschätzt wird. Kühe werden nur als Halbgeschwister in Nachkommengruppen berücksichtigt und ihre Leistungen über die genetische Kovarianz zwischen Eltern und Nachkommen dem Vater zugewiesen. Der Nachteil liegt in dem unbekanntem Zuchtwert des Anpaarungspartners, was zu Verzerrungen in den Schätzwerten führen kann (Mrode, 1996).

Das Mehrmerkmalsmodell

In der Tierzucht werden häufig mehrere Merkmale mit unterschiedlicher ökonomischer Bedeutung zur Selektion herangezogen. Diese Merkmale sind untereinander korreliert, sodass eine Schätzung der einzelnen Merkmalszuchtwerte gleichzeitig unter Berücksichtigung der genetischen und phänotypischen Korrelation erfolgen muss. Das Modell muss dabei folgendermaßen um das zweite Merkmal angepasst werden (2 Merkmale) (Mrode, 1996):

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

y_i = Vektor der Beobachtungen im Merkmal i

b_i = Vektor der fixen Effekte im Merkmal i

a_i = Vektor der zufälligen Tiereffekte im Merkmal i

e_i = Vektor der zufälligen Resteffekte im Merkmal i

X_i und Z_i sind Designmatrizen, die die Beobachtungen im jeweiligen Merkmal zu den Effekten zuordnen

Dabei wird angenommen, dass:

$$\text{var} \begin{bmatrix} a_1 \\ a_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} g_{11}A & g_{12}A & 0 & 0 \\ g_{21}A & g_{22}A & 0 & 0 \\ 0 & 0 & r_{11} & r_{12} \\ 0 & 0 & r_{21} & r_{22} \end{bmatrix}$$

Dadurch, dass die Merkmale untereinander korreliert sind, ist die Matrix R , die Varianz-Kovarianzmatrix der Resteffekte, keine Einheitsmatrix mehr. Daher muss das komplette Gleichungssystem aufgestellt werden.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

$$G = A \otimes G_0$$

$$R = I \otimes R_0$$

G_0 = additiv genetische Kovarianzmatrix

R_0 = Kovarianzmatrix der Resteffekte

Das Zeichen \otimes bezeichnet das Kronecker- oder direkte Produkt zweier Matrizen, bei dem jedes Element der links stehenden Matrix mit der gesamten rechten Matrix multipliziert wird.

Das komplett ausgeschriebene Gleichungssystem für zwei Merkmale sieht folgendermaßen aus:

$$\begin{bmatrix} X_1' R^{11} X_1 & X_1' R^{12} X_2 & X_1' R^{11} Z_1 & X_1' R^{12} Z_2 \\ X_2' R^{21} X_1 & X_2' R^{22} X_2 & X_2' R^{21} Z_1 & X_2' R^{22} Z_2 \\ Z_1' R^{11} X_1 & Z_1' R^{12} X_2 & Z_1' R^{11} Z_1 + A^{-1} G^{11} & Z_1' R^{12} Z_2 + A^{-1} G^{12} \\ Z_2' R^{21} X_1 & Z_2' R^{22} X_2 & Z_2' R^{21} Z_1 + A^{-1} G^{21} & Z_2' R^{22} Z_2 + A^{-1} G^{22} \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} X_1' R^{11} y_1 + X_1' R^{12} y_2 \\ X_2' R^{21} y_1 + X_2' R^{22} y_2 \\ Z_1' R^{11} y_1 + Z_1' R^{12} y_2 \\ Z_2' R^{21} y_1 + Z_2' R^{22} y_2 \end{bmatrix}$$

G^{ij} bzw. R^{ij} sind Elemente der G^{-1} bzw. R^{-1}

In den meisten Zuchtprogrammen wird ein Mehrmerkmalsmodell angewandt, weil unterschiedliche, nicht sehr hoch korrelierte Merkmale zu einem Gesamtzuchtwert kombiniert werden sollen. Hier wäre ein einfacheres Modell keine Alternative.

Das Wiederholbarkeitsmodell

Es kommt in der Tierzucht häufig vor, dass nicht nur eine Messung von einem Tier, sondern mehrere Messungen eines Merkmals vorliegen. Dadurch kann die Umweltvarianz in eine spezielle Umweltvarianz σ_{es}^2 und in eine permanente Umweltvarianz σ_{pe}^2 aufgeteilt werden. Dieses Modell berücksichtigt die Wiederholbarkeit und ermöglicht es, neben den Zuchtwerten auch den Effekt der generellen Umwelt zu schätzen (Mrode, 1996).

Das Modell hat folgende Form (Mrode, 1996):

$$y = Xb + Zu + Wpe + e$$

- y = Vektor der Beobachtungen
 b = Vektor der fixen Effekte
 u = Vektor der zufälligen Tiereffekte
 p_e = Vektor der zufälligen Effekte der permanenten Umwelt und der nicht additiv genetischen Effekte
 e = Vektor der zufälligen Resteffekte
 X, Z und W sind Designmatrizen, die die Beobachtungen jeweils den Effekten zuordnen

Die Matrizen W und Z sind identisch, wenn die Beobachtungen pro Tier gleich ausfallen, die Matrizen also gleich aufgebaut sind.

Bei der Mischmodellgleichung handelt es sich um eine Erweiterung des Einmerkmalsmodells um den Effekt der permanenten Umwelt:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\lambda_1 & Z'W \\ W'X & W'Z & W'W + I\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \\ p\hat{e} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}$$

$$\text{mit } \lambda_1 = \frac{\sigma_e^2}{\sigma_g^2} \text{ und } \lambda_2 = \frac{\sigma_e^2}{\sigma_{pe}^2}$$

2.4.2.2 BLUP in der genomic selection

Bei Anwendung der BLUP-Methode zur Berechnung der genomischen Zuchtwerte bleibt das verwendete Modell gleich. Einzig in der Z -Matrix werden nun die Haplotypen bzw. Marker mit den jeweiligen Phänotypen verknüpft, anstatt die Leistungen den Tieren zuzuordnen. Ein entsprechendes Modell wird in Kapitel 3.2.2 dargestellt. Geschätzt werden die Effekte für jeden einzelnen Haplotypen/Marker, worüber dann letztendlich für jedes Tier entsprechend seiner Haplotypen/Marker ein genomischer Zuchtwert berechnet werden kann.

Beim BLUP-Verfahren können zwar alle Marker simultan geschätzt werden, allerdings wird hierfür die Varianz benötigt. Man geht dabei davon aus, dass alle Marker zu gleichen Teilen die genetische Varianz erklären, was eine unrealistische Annahme darstellt, da es viele Gene geben wird, die nur einen geringen Einfluss auf das Merkmal haben werden, während einige einen großen Einfluss haben. Diese Annahme wird von dem Bayes-Verfahren aufgegriffen, da hier die Varianz eines jeden Locus variieren kann (Meuwissen et al., 2001).

2.4.3 Bayes Methode

Das Bayes Theorem, welches nach dem englischen Mathematiker und Pfarrer Thomas Bayes benannt wurde, ist ein Ergebnis der Wahrscheinlichkeitstheorie. Mit dem Satz von Bayes kann die Wahrscheinlichkeit für das Eintreten eines Ereignisses A berechnet werden unter der Voraussetzung, dass ein Ereignis B vorher eingetreten ist (Hayes, 2008).

Die Anwendung der Bayes-Methode ermöglicht es, das Wissen um die Verteilung der QTL-Effekte zu berücksichtigen. Nur wenige Chromosomensegmente beinhalten QTLs mit hohen Effekten, einige haben QTLs mit mittleren bis kleinen Effekten und einige Segmente beinhalten gar keine. Bei der Methode BayesA nach Meuwissen et al. (2001) werden die Daten auf zwei Ebenen untersucht. Die erste ist das Modell auf Ebene der Daten und die zweite die der Varianzen der Chromosomensegmente. Das Modell auf Ebene der Daten ist mit dem von BLUP gleichzusetzen, außer dass die Varianzen der Segmente hier $V(\mathbf{g}_{ij}) = \sigma_{gi}^2$ sind, welche für jedes Segment variieren und über das Modell der Varianzen der Segmente geschätzt werden. Die apriori Verteilung der Varianzen der Segmente ist eine skalierte invertierte Chi-Quadrat Verteilung, $X^{-2}(\nu, S)$, wobei S der Skalierungsparameter und ν die Anzahl der Freiheitsgrade darstellt. Die aposteriori-Verteilung der Varianzen der Segmente kombiniert die Informationen der apriori mit denen der Daten (Hayes, 2008):

$$P(\sigma_{gi}^2 | \mathbf{g}_i) = X^{-2}(\nu + n_i, S + \mathbf{g}_i' \mathbf{g}_i)$$

n_i = Anzahl der Haplotypen-Effekte in Segment i

Die aposteriori-Verteilung kann nicht direkt dazu verwendet werden, um σ_{gi}^2 zu schätzen, da sie abhängig von den unbekanntem \mathbf{g}_i Effekten ist. Meuwissen et al. (2001) benutzten daher das Gibbs-sampling Verfahren zur Schätzung der Effekte und Varianzen.

Die Methode BayesA stellt die Verteilung der genetischen Varianz über das Genom nicht vollständig dar. Meuwissen et al. (2001) zeigten, dass die inverse X^2 - Verteilung eine Verteilung der genetischen Varianz voraussetzt, bei der sehr viele SNPs einen sehr kleinen Effekt haben und sehr wenige einen großen Effekt. Die Möglichkeit, dass der Effekt gleich null ist, wird jedoch durch diese Verteilung nicht modelliert, weil die X^2 - Verteilung für den Wert Null eine unendliche Wahrscheinlichkeit hat. Es sollte daher die Verteilung so geändert werden, dass nicht nur der Wert Null eine realistische Wahrscheinlichkeit hat, sondern die Gesamtverteilung bei dem Wert Null einen Gipfel

bildet, da die Wahrscheinlichkeit bei sehr vielen SNPs am höchsten ist, dass dort kein genetischer Effekt liegt. Diese Methode mit der veränderten Verteilung wird bei Meuwissen et al. (2001) BayesB genannt. Hier wird eine apriori-Verteilung verwendet, die eine hohe Dichte bei $\sigma_{gi}^2 = 0$ und eine invertierte X^2 -Verteilung für alle Werte von $\sigma_{gi}^2 > 0$ hat. Die apriori-Verteilung gliedert sich in zwei Verteilungen, $\sigma_{gi}^2 = 0$ mit der Wahrscheinlichkeit π und $\sigma_{gi}^2 \sim X^{-2}(\nu, S)$ mit der Wahrscheinlichkeit von $1-\pi$.

Der Gibbs Sampler kann in diesem Fall zur Schätzung der Effekte und Varianzen nicht genutzt werden, weil der komplette Parameterraum nicht erfasst wird. Das Verfahren lässt keine Verteilungen mit $\sigma_{gi}^2 = 0$ zu, wenn die a-priori-Verteilung größer null war. Dieses Problem wird gelöst, indem σ_{gi}^2 und g_i simultan geschätzt mittels eines Metropolis-Hastings Algorithmus bestimmt wurden (Hayes, 2008).

2.5 Genomic selection – Angewandte Methoden

Die Methode der genomic selection ist in den letzten sieben Jahren seit ihrer Einführung durch Meuwissen et al. (2001) zu einem intensiven Betätigungsfeld geworden, wodurch vermehrt Studien durchgeführt wurden. Eine Übersicht über die verschiedenen Veröffentlichungen mit Angaben zu den verwendeten Methoden und Korrelationen zwischen dem wahren und genomischen Zuchtwert befindet sich in Tabelle 1.

Die erste Veröffentlichung zu diesem neuen Verfahren kam von Meuwissen et al. (2001). In der von ihnen durchgeführten Simulationsstudie wurden die Genauigkeiten der berechneten genomischen Zuchtwerte der drei angewandten Verfahren (Least Squares, BLUP, Bayes) miteinander verglichen. Dazu wurden 10 Chromosomen mit je 100 cM Länge simuliert. Auf jedem Chromosom befanden sich in der Mitte eines jeden Centimorgans ein QTL, welcher zu beiden Seiten im Abstand von 0,5 cM von einem Marker flankiert wurde. Die Mutationsrate für QTLs betrug $2,5 \times 10^{-5}$, die für Marker $2,5 \times 10^{-3}$. Es wurde eine effektive Populationsgröße von 100 erstellt und über 1000 Generationen erhalten. In den Generationen 1001 bis 1003 wurde die Anzahl von Tieren auf 200 bzw. 2000 (Generation 1002 und 1003) erhöht. Die Tiere in den Generationen 1001 und 1002 wurden genotypisiert, und der Phänotyp wurde ermittelt. Für die Tiere der letzten Generation sollte ein Zuchtwert mittels der Markerinformationen geschätzt werden. Die Markereffekte wurden an den Tieren aus Generation 1001 und 1002 ermittelt. Dafür wurden die Marker, welche einen QTL flankieren, zu einem Haplotyp zusammengefasst.

Tabelle 1: Übersicht über die Veröffentlichungen mit Angabe von Verfahren und Ergebnissen bei simulierten Daten

Autor	Verfahren	Korrelation
	Least Squares	0,12 – 0,36
Meuwissen et al. (2001)	BLUP	0,58 – 0,73
	Bayes B	0,71 – 0,85
Muir (2007)	BLUP	0,00 – 0,87
Kolbehdari et al. (2007)	BLUP	0,60 – 0,79
Calus and Veerkamp (2007)	Gibbs sampler	0,80 – 0,36
Dekkers (2007)	Selektionsindex	-
	Least Squares	0,49 – 0,79
Habier et al. (2007)	RR-Blup	0,42 – 0,87
	Bayes B	0,54 – 0,88
Calus et al. (2008)	Gibbs Sampler	0,47 – 0,90
Solberg et al. (2008)	BayesB	0,63 – 0,86

Die Effekte wurden über das Least-Squares-, BLUP- und Bayes-Verfahren geschätzt und die Korrelationen zwischen dem wahren und geschätzten Zuchtwert, welcher der Genauigkeit der Zuchtwerte entspricht und damit die Genauigkeit der Selektion angibt, miteinander verglichen. In Tabelle 2 werden die Ergebnisse der Arbeit für die vier Methoden für eine unterschiedliche Anzahl von Tieren (2200, 1000 und 500 Tiere) bei einem Marker Abstand von 1cM aufgezeigt. Weiterhin sind die Markerabstände bei einer gleich bleibenden Anzahl von 2200 Tieren von 1 cM auf 2 cM und 4 cM dargestellt.

Eine Abnahme der Genauigkeit durch eine Verringerung der Tier- sowie der Markeranzahl konnte somit aufgezeigt werden. Einzig die Werte beim Least-Squares-Verfahren sind bei fallender Markeranzahl angestiegen.

Tabelle 2: Korrelation zwischen dem wahren und genomischen Zuchtwert für vier verschiedene Methoden und eine unterschiedliche Anzahl von Tieren (verändert nach Meuwissen et al., 2001)

		Methoden			
		Least Squares	BLUP	BayesA	BayesB
Anzahl Tiere	2200	0,318	0,732	0,798	0,848
	1000	0,204	0,659	-	0,787
	500	0,124	0,579	-	0,708
Abstand zwischen zwei Markern	1 cM	0,318	0,732	0,798	0,848
	2 cM	0,354	0,708	-	0,810
	4 cM	0,363	0,668	-	0,737

Muir (2007) untersuchte SNP-Effekte von simulierten Daten und berechnete genomische Zuchtwerte mit Hilfe des BLUP-Verfahrens. Die erstellte Populationsgröße variierte dabei von 128 bis 1024 Tieren bei ausgeglichenem Geschlechterverhältnis. Die simulierte Genomgröße lag zufällig zwischen 1 und 10000 cM, welche auf drei Chromosomen aufgeteilt wurde. Die Anzahl an QTL-Loci lag bei 100. Sie wurden zufällig über das gesamte Genom verteilt. SNPs wurden in zufälligem, 0,1 bis 100 cM, aber gleichmäßigem Abstand über das Genom gelegt. QTLs und SNPs waren biallelisch. Insgesamt wurden zwei Variationen simuliert. Zum einen wurde eine Population im Hardy-Weinberg-Gleichgewicht erzeugt und zum anderen eine im Mutations-Drift-Gleichgewicht, wofür 5000 Generationen erzeugt wurden, um das Gleichgewicht herzustellen. Zur Analyse folgten 1 bis 5 Trainingsgenerationen, an denen die SNP-Effekte geschätzt wurden. Hier erfolgte die Auswahl zur Zucht zufällig. Für die folgenden 5 – 9 Generationen wurden jeweils die genomischen Zuchtwerte berechnet. Die Tiere wurden hier zufällig oder direkt (über traditionelle BLUP-Zuchtwertschätzung oder genomische Zuchtwerte) zur Zucht ausgewählt. Für die Vergleiche der Ergebnisse wurden die Korrelationen zwischen den genomischen und den wahren Zuchtwerten berechnet.

Die Ergebnisse für eine Populationsgröße von 128 Tieren im Hardy-Weinberg-Gleichgewicht, bei einer Heritabilität von 0,1 (Abbildung 3, linke Seite) und einer Markeranzahl von 100 bei 100 cM Genomgröße zeigten auf, dass die Genauigkeit je nach Anzahl der Trainingsgenerationen und nach Anzahl nachfolgender Generationen stark schwanken. Je höher die Anzahl von Trainingsgenerationen war, desto höher fiel die Genauigkeit aus. So lag die Genauigkeit bei einer Trainingsgeneration zwischen 0,0 und

0,2 und stieg bis auf 0,65 bei vier Trainingsgenerationen an. Für die Population im Mutations-Drift-Gleichgewicht fielen die Genauigkeiten leicht höher aus. Aber auch hier wurden mit steigender Anzahl an Trainingsgenerationen höhere Genauigkeiten erreicht. Bei einer Heritabilität von 0,5 (Abbildung 3, rechte Seite) lagen die Genauigkeiten je nach Anzahl an Trainingsgenerationen in einem Bereich von 0,1 bis 0,8 für eine Population im Hardy-Weinberg-Gleichgewicht. Für eine Population im Mutations-Drift-Gleichgewicht konnten Genauigkeiten von 0,15 bis 0,90 erzielt werden. Auch hier wurde deutlich, dass die Genauigkeiten mit steigender Anzahl an Trainingsgenerationen zunahmen.

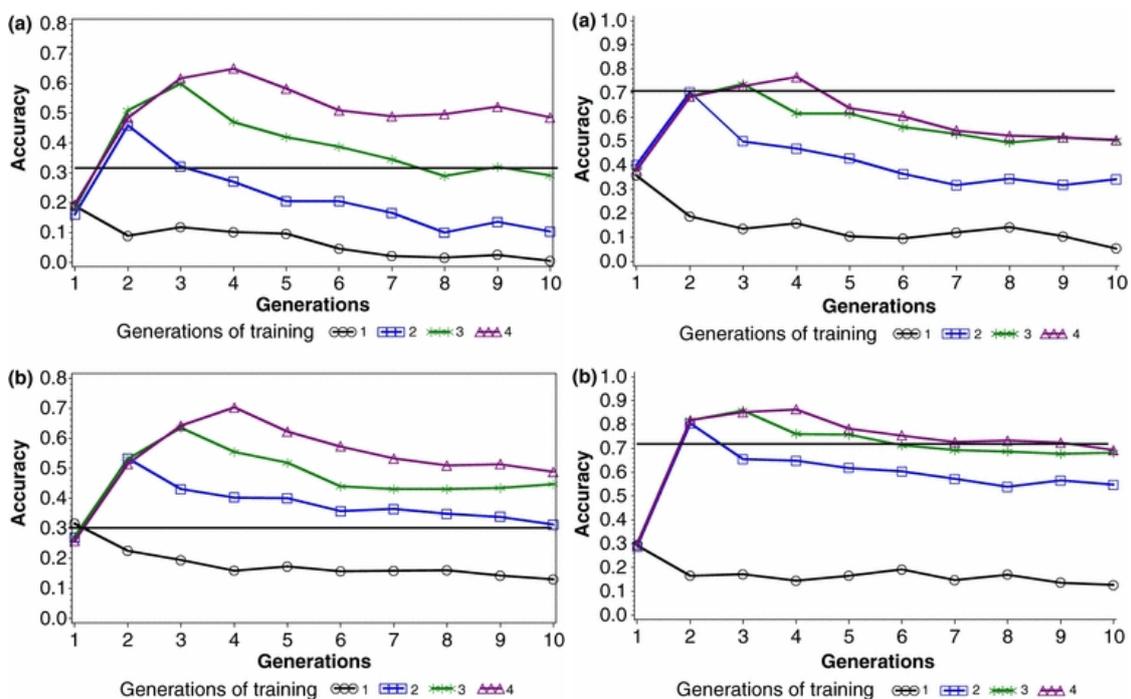


Abbildung 3: Darstellung der Genauigkeit der genomischen Zuchtwerte bei verschiedenen Anzahlen von Trainingsgenerationen für eine Population im Hardy-Weinberg-Gleichgewicht (a) und im Mutations-Drift-Gleichgewicht (b) für eine Heritabilität von 0,1 (Abbildungen auf der linken Seite) und 0,5 (Abbildungen auf der rechten Seite), aus: Muir (2007)

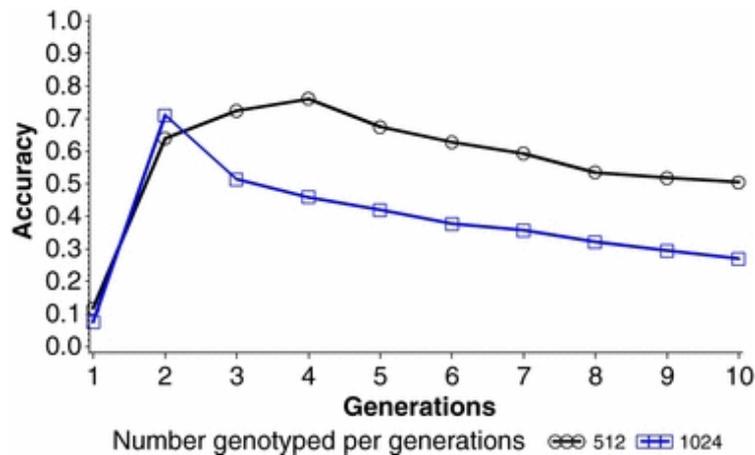


Abbildung 4: Genauigkeit des genomischen Zuchtwertes bei 512 genotypisierten Tieren über vier Trainingsgenerationen und bei 1024 genotypisierten Tieren über zwei Generationen, aus: Muir (2007)

Bei einem Vergleich der Anzahl von genotypisierten Tieren pro Trainingsgeneration für eine Population im Hardy-Weinberg-Gleichgewicht bei einer Heritabilität von 0,1 wurde gezeigt, dass eine geringere Zahl von 512 Tieren über 4 Trainingsgenerationen gegenüber einer hohen Anzahl von 1024 Tieren über nur 2 Trainingsgenerationen zu höheren Genauigkeiten führte (Abbildung 4).

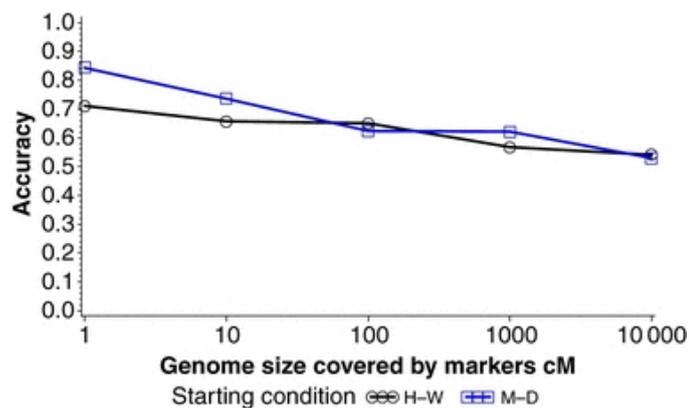


Abbildung 5: Genauigkeiten des genomischen Zuchtwertes bei verschiedenen Genomgrößen bei 100 Markern für vier Trainingsgenerationen, $h^2 = 0,1$ und Populationen im Hardy-Weinberg- und Mutations-Drift-Gleichgewicht

In Abbildung 5 ist der Einfluss der Änderung der Markerdichte durch Veränderung der Genomgröße bei gleich bleibender Anzahl von 100 Markern, einer Heritabilität von 0,1 und einer Populationsgröße von 128 aufgezeigt. Die Genauigkeit des genomischen Zuchtwertes fiel bei sinkender Markerdichte von 0,84 auf 0,60 (Mutations-Drift-Gleichgewicht) und von 0,71 auf 0,60 (Hardy-Weinberg-Gleichgewicht). Geringere Markerdichten bzw. größere Abstände zwischen einzelnen Markern führten somit zu einer höheren Rekombinationsrate. Die Kopplung fällt niedriger aus und die Genauigkeiten nehmen ab (Muir, 2007).

Die Veränderung der Markeranzahl bei gleich bleibender Markerdichte von 10 Markern/cM, einer Heritabilität von 0,1 und einer Tieranzahl von 128 je Trainingsgeneration für eine Population im Hardy-Weinberg- und eine im Mutations-Drift-Gleichgewicht ist in Abbildung 6 dargestellt. Die Werte ergaben einen Höhepunkt der Genauigkeit von 0,69 bei 1000 Markern für eine Population im Hardy-Weinberg-Gleichgewicht. Bei der Population im Mutations-Drift-Gleichgewicht lag die höchste Genauigkeit mit 0,78 bei 10 Markern. Muir (2007) geht von einem optimalen Marker:QTL-Verhältnis von 10:1 für eine Population im Hardy-Weinberg-Gleichgewicht aus, während das optimale Verhältnis für eine Population im Mutations-Drift-Gleichgewicht bei 2,5:1 liegt.

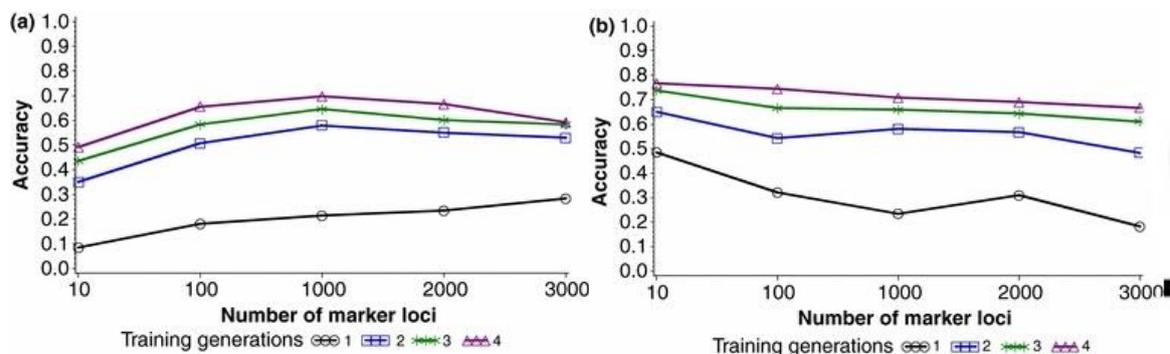


Abbildung 6: Genauigkeiten des genomischen Zuchtwertes für eine unterschiedliche Anzahl von Markern bei gleich bleibender Markerdichte von 10 Markern/cM für unterschiedliche Anzahlen an Trainingsgenerationen bei $h^2 = 0,1$ und einer Population im Hardy-Weinberg-Gleichgewicht (a) und einer im Mutations-Drift-Gleichgewicht (b), aus: Muir (2007)

Weiterhin wurden die Genauigkeiten der genomischen Zuchtwerte unter Selektion berechnet (Abbildung 7). Hierzu wurde eine 256 Individuen starke Population im Hardy-Weinberg- oder Mutations-Drift-Gleichgewicht mit einer Heritabilität von 0,1 und 100 Markern erstellt. 32 Tiere wurden in den 4 Trainingsgenerationen zufällig für die Zucht verwendet. In den folgenden sechs Generationen wurden die Tiere zufällig oder direkt auf den genomischen Zuchtwert selektiert. Die zufällige Selektion resultierte in hohen Korrelationen besonders für die Population im Mutations-Drift-Gleichgewicht. Dagegen nahm die Korrelation für direkte Selektion nach den Trainingsgenerationen schnell ab, was sich besonders bei der Population im Mutations-Drift-Gleichgewicht bemerkbar machte.

Muir (2007) nahm daher an, dass man Ergebnisse aus Zufallspaarung nicht ohne weiteres auf Selektion übertragen kann.

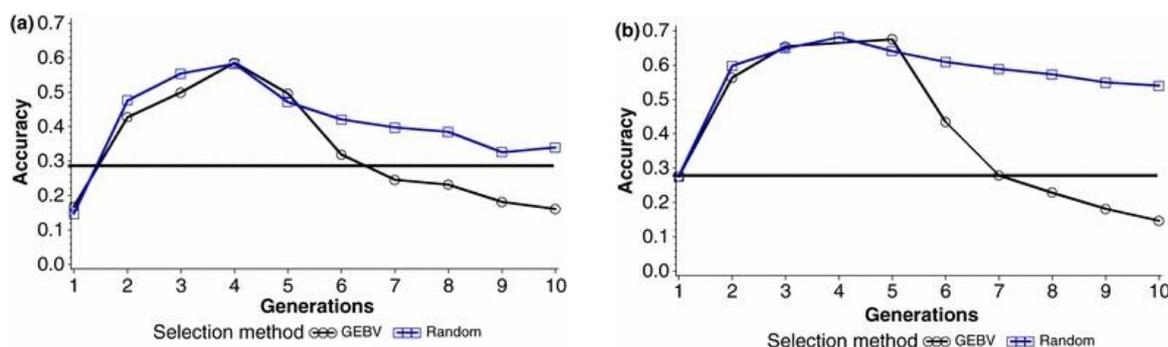


Abbildung 7: Genauigkeit für zufällige oder gerichtete Selektion auf den genomischen Zuchtwert nach vier Trainingsgenerationen für $h^2 = 0,1$ für eine Population im Hardy-Weinberg-Gleichgewicht (a) und eine im Mutations-Drift-Gleichgewicht (b), aus: Muir (2007)

In der Veröffentlichung von Kolbehdari et al. (2007) wurden genomische Zuchtwerte für eine Gruppe von Söhnen berechnet. Für die Simulationsstudie wurde ein Chromosom von 100 cM simuliert. Auf diesem Chromosom lagen 10 QTL und 100 gleichmäßig verteilte SNPs. QTLs und SNPs waren biallelisch mit gleichen Allelfrequenzen, die in der Basispopulation zufällig vergeben wurden. Die Basispopulation bestand aus 30 Bullen und 400 Kühen, auf die 25 Generationen Zufallspaarung folgten. In der letzten Generation wurden aus der Vorhergehenden zufällig 20 Bullen ausgewählt, jeder mit 100 Söhnen, die zufällig in zwei Gruppen aufgeteilt wurden. Eine Gruppe mit 10 Bullen und deren Söhnen bildete die Testgruppe, an welcher mit Hilfe der daughter yield deviation (DYD) der jeweils 100 Töchter die Markereffekte über das BLUP-Verfahren geschätzt wurden. Dazu wurden die Marker in Intervalle von je zwei SNPs eingeteilt, so dass insgesamt 99 Intervalle

vorlagen. Für alle Tiere der beiden Gruppen wurden über die SNP-Intervall-Effekte die genomischen Zuchtwerte berechnet. Verschiedene Einstellungen konnten dabei verändert werden. Zum einen wurden unterschiedliche Heritabilitäten von 0,05, 0,3 und 0,5 verwendet. Außerdem wurden die QTL-Effekt-Gewichtungen von gleichmäßig (alle 10 QTL haben einen gleich großen Einfluss) auf ungleichmäßig (ein QTL hat einen großen Einfluss, zwei einen mittleren und sieben einen kleinen Einfluss) variiert, und die QTL-Positionen konnten von gleichmäßig auf zufällige Verteilung geändert werden. Berechnet wurde auch hier wieder die Korrelation zwischen dem wahren und genomischen Zuchtwert. Die Ergebnisse wurden in Tabelle 3 dargestellt.

Tabelle 3: Korrelationen zwischen dem wahren und genomischen Zuchtwert für verschiedene Heritabilitäten, QTL-Gewichtungen und Positionen (Kolbehari et al., 2007)

h ²	QTL-Gewichtungen	QTL-Positionen	Korrelation
0,5	gleichmäßig	gleichmäßig	0,63
0,5	gleichmäßig	zufällig	0,77
0,5	ungleichmäßig	gleichmäßig	0,62
0,5	ungleichmäßig	zufällig	0,79
0,3	gleichmäßig	gleichmäßig	0,60
0,3	gleichmäßig	zufällig	0,76
0,3	ungleichmäßig	gleichmäßig	0,61
0,3	ungleichmäßig	zufällig	0,77
0,05	gleichmäßig	gleichmäßig	0,60
0,05	gleichmäßig	zufällig	0,74
0,05	ungleichmäßig	gleichmäßig	0,60
0,05	ungleichmäßig	zufällig	0,75

Die gesamte Spannweite bei diesen Daten lag von $r = 0,60$ bis $0,79$, wobei Korrelationen über $r = 0,70$ bei Simulationen mit zufälliger QTL-Verteilung erreicht wurden. Mit steigender Heritabilität von 0,05 auf 0,5 stieg die Korrelation leicht von 0,75 auf 0,79 bei ungleichmäßigem QTL-Einfluss und zufälliger QTL-Position. Bei gleichem QTL-Einfluss und zufälliger QTL-Position lagen die Werte mit 0,74 – 0,77 nahezu im gleichen Bereich. Bei gleichmäßigem QTL-Einfluss lagen die Korrelationen um 0,01 höher als bei ungleichmäßigem Einfluss. Dieser Unterschied ist aber nicht signifikant. Die Korrelationen bei gleichmäßiger QTL-Position lagen generell unter 0,65.

Den Einfluss eines polygenen Effektes in die genomische Zuchtwertschätzung und ihre Auswirkung auf die Genauigkeit der genomischen Zuchtwerte wurde von Calus und Veerkamp (2007) untersucht. In ihrer Arbeit gingen sie davon aus, dass in der Praxis das Genom der verschiedenen Tierarten noch nicht vollständig und regelmäßig mit SNPs bedeckt ist. Die vorhandenen SNPs sind so nicht in der Lage, die gesamte QTL-Varianz zu erklären. Die genetische Varianz, welche nicht von den SNPs erklärt wird, zu ignorieren, könnte die Genauigkeit des genomischen Zuchtwertes und damit der Selektion reduzieren. Aus diesem Grund wurden verschiedene Datensätze simuliert, um Modelle mit und ohne polygene Effekte zu testen. Dafür wurde eine Population von 100 Tieren mit gleicher Anzahl männlicher und weiblicher Tiere erstellt, deren Größe über 1000 Generationen konstant gehalten wurde. Jedes Tier pro Generation verpaarte sich zufällig und erzeugte zwei Nachkommen. Das simulierte Genom umfasste drei Chromosomen von je einem Morgan Länge. 50000 SNPs und 10000 QTLs wurden zufällig über das gesamte Genom verteilt. Alle enthielten nur eines von zwei möglichen Allelen, welches innerhalb der 1000 Generationen nur ein einziges Mal mutieren konnte. Alle Marker mit Allelfrequenzen kleiner 0,02 in den Generationen 1002 und 1003 wurden für die Analyse verworfen. Allen mutierten QTLs wurde ein Effekt zugeteilt, wobei eine große Anzahl einen kleinen Einfluss zugesprochen bekam und nur wenige einen großen Effekt hatten. In Generation 1001 hatten alle Tiere je 10 Nachkommen, sodass 1000 Tiere in Generation 1002 vorlagen, welche dann für die Generation 1003 wieder nur zwei Nachkommen produzierten. Die Tiere in den Generationen 1001 und 1002 bekamen einen Phänotypwert zugeteilt, und alle Tiere der Generation 1001 bis 1003 wurden genotypisiert.

Die Heritabilität des Merkmals wurde auf 0,5 und 0,1 gesetzt. Fünf verschiedene Szenarien wurden getestet, in denen verschiedene Kombinationen genotypisierter Tiere zur Analyse verwendet wurden. Im 1. Szenario wurden alle Informationen der Eltern und Großeltern analysiert, im 2. Szenario die der Eltern und Großväter. Im 3. Szenario wurden nur die Geno- und Phänotypwerte der Eltern verwendet. Im 4. Modell wurden nur die Werte der Väter und Großväter und als letzte Wahl nur noch Informationen der Väter für die Analyse verwendet. Für die Analyse wurden die SNPs in 2 SNP-Haplotypen zusammengefasst und die Markereffekte mit einem Gibbs sampler über 30000 Runden geschätzt und anschließend genomische Zuchtwerte berechnet. Die Korrelation zwischen dem wahren und genomischen Zuchtwert wurde für alle Tiere in den Generationen 1001 – 1003, die in die Analyse einbezogen waren, berechnet. Die Korrelationen zwischen dem wahren und dem genomischen Zuchtwert der Elterntiere waren stark von der Anzahl der vorhergehenden Tieren in der Analyse abhängig. Sie fiel mit sinkender Anzahl an Tieren ab (Tabelle 4).

Die Korrelationen zwischen dem wahren und dem genomischen Zuchtwert der Selektionskandidaten aus Generation 1003 bei $h^2 = 0,1$ fielen von $r = 0,55$ bei Einbezug sämtlicher Eltern und Großeltern-Informationen auf $r = 0,38$, wenn nur noch die Väter zur Analyse verwendet wurden. Für die hohe Heritabilität rangierten die Werte zwischen $r = 0,79$ und $0,73$. Der Ausschluss der Großeltern (Generation 1001) aus der Analyse führte zu einer geringen Abänderung von $r = 0,02$ bei $h^2 = 0,1$ (Tabelle 4). Insgesamt konnten höhere Genauigkeiten der Zuchtwerte durch steigende Anzahl von Tieren in der Analyse besonders bei geringen Heritabilitäten erzielt werden. Der Einbezug polygener Effekte bewirkte in allen Szenarien nur eine geringfügige Veränderung. Die Korrelationswerte schwankten hier für $h^2 = 0,1$ und in Generation 1003 zwischen $r = 0,55$ und $0,36$ und für $h^2 = 0,5$ zwischen $r = 0,8$ und $0,74$.

Tabelle 4: Korrelationen zwischen dem wahren und genomischen Zuchtwert für Großeltern, Eltern und Nachkommen bei $h^2 = 0,1$ und $0,5$ und polygenen Effekten (verändert nach Calus und Veerkamp, 2007)

h ²	polygener Effekt	Generation/Tiere				
		1001 Großväter	1001 Großmütter	1002 Väter	1002 Mütter	1003 Nachkommen
0,1	nicht einbezogen	0,65	0,65	0,60	0,60	0,55
	einbezogen	-	-	0,59	0,60	0,54
		-	-	0,58	0,59	0,53
		0,54	-	0,48	-	0,40
		-	-	0,47	-	0,38
	einbezogen	0,68	0,67	0,61	0,61	0,55
		0,66	-	0,60	0,60	0,53
		-	-	0,59	0,60	0,52
		0,55	-	0,49	-	0,39
		-	-	0,47	-	0,36
-		-	-	-	-	
0,5	nicht einbezogen	0,91	0,91	0,86	0,86	0,79
	einbezogen	0,91	-	0,86	0,86	0,79
		-	-	0,86	0,85	0,79
		0,87	-	0,83	-	0,74
		-	-	0,82	-	0,73
	einbezogen	0,93	0,93	0,88	0,88	0,80
		0,93	-	0,88	0,87	0,80
		-	-	0,87	0,87	0,80
		0,88	-	0,84	-	0,75
		-	-	0,84	-	0,74
-		-	-	-	-	

Um die Auswirkung der polygenen Effekte weiter zu untersuchen, wurden alle Tiere aus den Generationen 1001 und 1002 analysiert, aber nur 50 bzw. 20 % der Marker verwendet und somit die Markerdichte verringert. Die Ergebnisse zeigten hier, dass bei einer geringen Heritabilität von 0,1 ein Gewinn an Genauigkeit durch die Einbindung von polygenen Effekten erst bei geringen Markerdichten erzielt werden konnte. Bei höheren Heritabilitäten fielen die Korrelationen mit sinkender Markerdichte unter Einbeziehung der polygenen Effekte höher aus als ohne. Allerdings fielen die Korrelationen mit sinkender Markerdichte generell niedriger aus, so dass insgesamt nur ein sehr geringer Nutzen aus der Einbindung der polygenen Effekte gezogen werden konnte und diese nur sinnvoll bei geringen Markerdichten erscheint.

Dekkers (2007) schätzt den Selektionsfortschritt, der durch den Einbezug von Markerdaten in die Zuchtwertschätzung entsteht. Es wurden Daten für ein vereinfachtes Schweinezuchtprogramm mit Heritabilitäten für das Merkmal von 0,4 und 0,1 simuliert. In jeder Generation wurden 20 männliche Tiere selektiert und jedes Tier an drei selektierte weibliche Tiere angepaart, die je acht Nachkommen erzeugten. Die Selektion geschah mit Hilfe von BLUP-Zuchtwerten, welche auf Phänotyp- und/oder Markerdaten basierten. Die Daten werden mit dem traditionellen Selektionsindex analysiert, der um den Markereffekt erweitert wurde:

$$I = b'X = [b'_Q, b'_P] \begin{bmatrix} X_Q \\ X_P \end{bmatrix}$$

X_Q = Vektor der genomischen Zuchtwerte der Individuen selbst und/oder ihrer Verwandten

X_P = Vektor der Phänotypen der Individuen selbst und/oder ihrer Verwandten

b'_Q, b'_P = Vektoren der Index-Gewichtungen

Die Ergebnisse zeigten, dass die genomischen Zuchtwerte, die allein auf Markerdaten basierten, eine Genauigkeit von 0,75 für $h^2 = 0,4$ und 0,55 für $h^2 = 0,1$ erforderten, um den gleichen Zuchtfortschritt zu erlangen, der nur auf Phänotypdaten basierte. Erst bei höheren Genauigkeiten der genomischen Zuchtwerte konnte ein höherer Zuchtfortschritt bis zu 35 % für $h^2 = 0,4$ und bis über 80 % für $h^2 = 0,1$ erzielt werden. Die kombinierte Selektion auf Marker- und Phänotypdaten resultierte dagegen in einem erheblichen Gewinn an Zuchtfortschritt gegenüber der phänotypbasierten Selektion. Hier konnte schon bei relativ geringen Genauigkeiten des genomischen Zuchtwertes ein erhöhter Zuchtfortschritt erzielt werden, besonders bei niedrigen Heritabilitäten und bei

Verwendung der Phänotypen ausschließlich der weiblichen Tiere. Weiterhin wurde der Einfluss der Einbeziehung von Markerdaten auf die Inzuchtrate dargestellt. Hier war zu erkennen, dass die Verwendung von Markerdaten allein die Rate erheblich senken konnte, von 2 bis über 3 % für die Selektion auf den Phänotyp und auf weniger als 1 % bei alleiniger Verwendung von Markern.

In der Arbeit von Habier et al. (2007) wird der Einfluss der Verwandtschaft zwischen und innerhalb einzelner Generationen auf LD und die Genauigkeit der genomischen Zuchtwerte anhand simulierter Daten geschätzt. Dazu wurde eine Population mit 100 Tieren simuliert, die über 1000 Generationen einer Zufallspaarung unterlagen. Die Population wurde in zwei Linien aufgeteilt und über fünf Generationen auf je 1000 Tiere in Generation 1005 erhöht. In den folgenden fünf Generationen wurden jeweils nur 50 männliche Tiere an alle 500 weiblichen Tiere angepaart. Anhand der Linie 1 wurden die Markereffekte mittels Least Squares, BLUP, BayesA und BayesB geschätzt. Für die Tiere der Linie 2 wurden die genomischen Zuchtwerte berechnet. Insgesamt wurde ein Szenario mit 1000 Markern im LD zu 50 QTL simuliert. Die SNPs hatten einen Abstand von 1 cM. 1000 Tiere standen hier zur Berechnung der Effekte zur Verfügung. Weiterhin wurde ein Szenario mit 100, 1000 und 2000 Markern in LE mit 10 QTL simuliert. Die Effekte wurden an 500 männlichen Tieren geschätzt.

Selbst bei fehlenden LD wurden signifikante Genauigkeiten der genomischen Zuchtwerte gefunden. Bei vorhandenen LD stiegen die Genauigkeiten um fast 50 % an, sanken jedoch in den Folgegenerationen stark ab. Dies wurde in der Arbeit mit dem Absinken der LD aufgrund geringerer Verwandtschaftsbeziehungen erklärt. Um diesen Effekt aufzufangen, wurde in dem statistischen Modell die Verwandtschaft zwischen den Tieren berücksichtigt und das LD daraufhin korrigiert. Eine über Generationen persistente Genauigkeit der genomischen Zuchtwerte konnte mit dieser Korrekturmethode erreicht werden.

In der Arbeit von Calus et al. (2008) wurden die Genauigkeiten der genomischen Zuchtwerte für drei verschiedene Haplotypengrößen und verschiedene Markerdichten berechnet. Die Populationsgröße lag für 1000 Generationen bei 100 Tieren, die einer Zufallspaarung unterlagen. Das simulierte Genom umfasste drei Chromosomen von je 1 Morgan Länge. Über das gesamte Genom waren 300000 SNPs und 50000 QTLs zufällig verteilt und erhielten alle das Allel 1. Pro Generation traten durchschnittlich 300 Mutationen bei SNPs und 50 bei QTLs auf. Die Mutationsrate der SNPs und QTLs wurde anhand vorhergehender Simulationen in der Generation 1000 auf der Basis der polymorphen SNPs und QTLs bestimmt, ca. 25000 SNPs und 75 QTL. Alle SNPs mit

Frequenzen $< 0,02$ in den Generationen 100-1003 wurden verworfen. Aus den verbleibenden Markern wurden fünf verschiedene Markerdichten simuliert, indem zufällig 100, 50, 20, 10 oder 5 % der polymorphen SNPs ausgewählt wurden. Mutierte QTL-Allele (im Durchschnitt 78,3) erhielten einen positiven oder negativen Wert über eine Gamma-Verteilung. Nach 1000 Generationen wurden weitere drei Generationen simuliert. Ab Generation 1002 wurde die Populationsgröße auf 1000 Tiere angehoben. Alle Tiere der Generationen 1001 und 1002 erhielten einen Phänotypwert. Die Heritabilität lag bei 0,5 oder 0,1. Anhand dieser Tiere wurden die Gewichtungsfaktoren für die Haplotypen/SNPs über den Gibbs Sampler geschätzt und die genomischen Zuchtwerte für die Generationen 1001-1003 berechnet. Vier Annahmen über die Haplotyp-Konstruktion wurden gemacht:

- Einzel-SNP-Analyse
- Haplotyp aus zwei benachbarten SNPs
- Haplotyp aus zwei SNPs mit Verwendung von identical by descent (IBD) - Wahrscheinlichkeiten
- Haplotyp aus zehn SNPs mit Verwendung von IBD-Wahrscheinlichkeiten

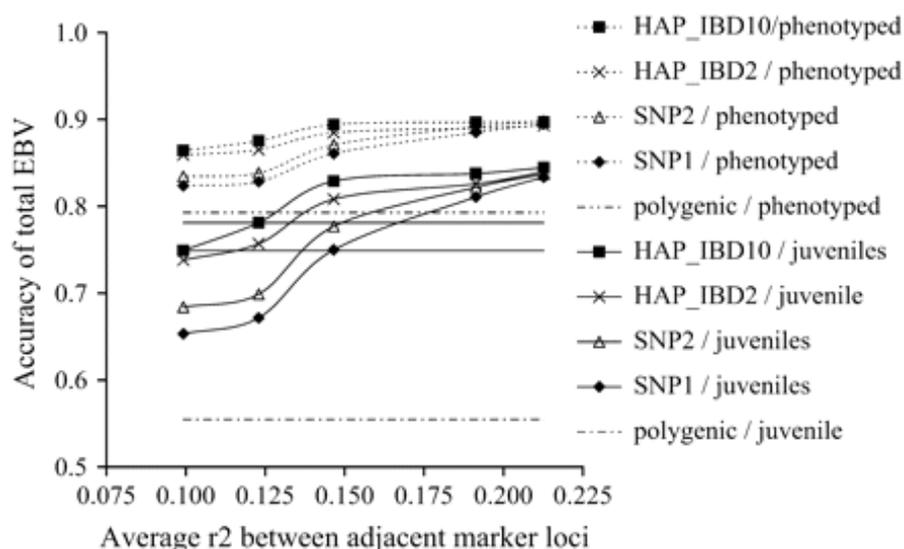


Abbildung 8: Genauigkeiten des genomischen Zuchtwertes für Eltern und Nachkommen für $h^2 = 0,5$ bei verschiedener Markerdichte und verschiedenen Haplotypenzusammenstellungen, aus Calus et al. (2008)

Die Genauigkeiten der genomischen Zuchtwerte für $h^2 = 0,5$ lagen für die verschiedenen Haplotyp-Modelle zwischen 0,65 und 0,9 (Abbildung 8), wobei die Genauigkeiten mit steigender Markerdichte zunahmen. Das Modell für Haplotypen aus 10 SNPs erhielt die höchsten Genauigkeiten, das für die Einzel-SNP-Analyse die niedrigsten. Die Unterschiede zwischen den einzelnen Haplotyp-Modellen nahmen mit steigender Markerdichte ab und lagen zwischen 0,0 und 0,11.

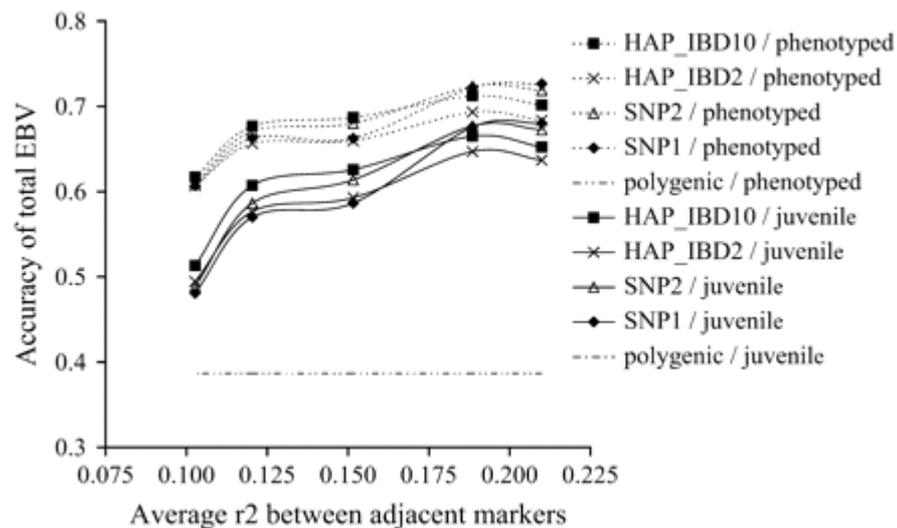


Abbildung 9: Genauigkeiten des genomischen Zuchtwertes für Eltern und Nachkommen für $h^2 = 0,1$ bei verschiedener Markerdichte und verschiedenen Haplotypenzusammenstellungen, aus Calus et al. (2008)

Für die niedrige Heritabilität (Abbildung 9) lag die Genauigkeit zwischen 0,47 und 0,72. Auch hier nahm die Genauigkeit mit steigender Markerdichte zu. Für geringe Markerdichten bekam der 10er Haplotyp die höchsten Genauigkeiten, bei hohen Markerdichten konnte die Einzel-SNP-Analyse leicht höhere Genauigkeiten erzielen. Die Unterschiede zwischen den einzelnen Haplotyp-Modellen fielen allerdings für alle Markerdichten sehr gering aus. Höhere Genauigkeiten konnten in dieser Arbeit somit durch den Einbezug von Informationen durch Kopplungsanalysen bei hohen Heritabilitäten erzielt werden. Bei niedrigen Heritabilitäten fielen die Unterschiede zwischen den Modellen dagegen sehr gering aus, und kein Modell war dem anderen signifikant überlegen.

Die Auswirkungen von Markerdichte und Größe der Haplotypen anhand von SNPs und Mikrosatelliten auf die Genauigkeit der genomischen Zuchtwerte wurden in der Arbeit von Solberg et al. (2008) behandelt. Hierfür wurde eine Population von 100 Tieren mit gleichem Geschlechterverhältnis simuliert, die für 1000 Generationen der Zufallspaarung unterlag. In Generation 1001 wurde die Populationsgröße auf 1000 Tiere erhöht, indem die männlichen Tiere an je 20 weibliche angepaart wurden. Diese Größe wurde auch in der folgenden Generation 1002 beibehalten. Für jedes Tier wurden 10 Chromosomen mit einer Gesamtgenomgröße von 10 Morgan simuliert. Die beiden Markertypen, SNPs und Mikrosatelliten, wurden gleichmäßig, aber in verschiedenen Dichten, über die Chromosomen verteilt. Für beide Markertypen gab es je vier simulierte Dichten. Für die Mikrosatelliten waren die Dichten: 2/cM, 1/cM, 0,5/cM und 0,25/cM, während sie für die SNPs bei 8/cM, 4/cM, 2/cM und 1/cM lagen. Die möglichen QTL lagen gleichmäßig über das Chromosom verteilt, ihre Anzahl bei 100/Chromosom. In der 1. Generation gab es keinerlei Polymorphismen bei Markern oder QTL. Mutationen ließen Änderungen mit einer Rate von $2,5 \times 10^{-3}$ für Marker und $2,5 \times 10^{-5}$ für QTL zu. Mutierte QTL-Allele bekamen einen Effekt zugeteilt. Die Markereffekte für die Einzelmarker bzw. Haplotypen (zwei benachbarte Marker) wurden anhand der 1000 Tiere der 1001ten Generation über BayesB geschätzt.

Tabelle 5: Korrelation zwischen dem wahren und genomischen Zuchtwert für unterschiedliche Markerdichten bei Markergenotypen und -haplotypen sowie SNP-Genotypen und Haplotypen (verändert nach Solberg et al., 2008)

Marker pro Morgan	Korrelation	
	Genotypen	Haplotypen
25	0,626	
50	0,723	
100	0,770	0,764
200	0,827	0,798
SNP pro Morgan		
100	0,690	
200	0,790	
400	0,841	0,802
800	0,860	0,821

Die Korrelationen der genomischen Zuchtwerte (Tabelle 5) für Mikrosatellitengenotypen lagen zwischen $r = 0,626$ und $0,827$ für die verschiedenen Markerdichten, wobei die Genauigkeit mit steigender Markerdichte zunahm. Auch für die SNPs-Genotypen stieg die Korrelation von $r = 0,690$ auf $0,860$ mit steigender Markerdichte an. Bei der Verwendung von Mikrosatelliten- und SNP-Haplotypen konnte mit steigender Markerdichte ein leichter Anstieg der Genauigkeit von $0,764$ auf $0,798$ bei den Mikrosatelliten verzeichnet werden. Bei den SNPs stieg die Genauigkeit von $0,802$ auf $0,821$ an. Tendenziell konnte beobachtet werden, dass die Genauigkeiten für die Haplotypen geringer ausfielen und die Einzelmarker somit eine höhere Genauigkeit lieferten.

3 Material und Methoden

3.1 Die Simulationsprogramme

Die Ansätze und Grundzüge dieser Arbeit basieren auf den Simulationen von Meuwissen et al. (2001), die in Kapitel 2.4 näher beschrieben wurden. Für die vorliegende Arbeit wurde eine Schweinepopulation erstellt. Alle benötigten Daten, wie das Pedigree, SNPs sowie Geno- und Phänotypwerte wurden mit Hilfe eines Fortran Programmes simuliert. Sämtliche genauere Beschreibungen, wie Zahlen- und Größenangaben des Simulationsprogrammes basieren auf den Grundeinstellungen, welche alle variabel gestaltet wurden, so dass alle Parameter verändert werden konnten. Im Folgenden hier eine Liste der veränderbaren Parameter in ihren Grundeinstellungen:

Tabelle 6: Übersicht über die veränderbaren Parameter

Parameter	Grundeinstellung
Anzahl Tiere pro Generation und Geschlecht	150
Anzahl Generationen bis zur Analyse	15
Anzahl Generationen ab der Analyse	5
Anzahl ausgewählter Elternpaare für die Analyse	30
Anzahl Nachkommen pro Elternpaar	2
Anzahl Nachkommen pro Elternpaar für Analysegeneration	10
Chromosomenlänge in Morgan	1
SNP-Anzahl	100
QTL-Anzahl	12
Allele pro Locus	2
Segmentanzahl	100
Mutationsrate	$2,5 \times e^{-3}$
Heritabilität	0,5
Anzahl Wiederholungen je Lauf	100

Für die Gründergeneration wurden insgesamt 300 Tiere, je 150 pro Geschlecht, simuliert. Für jedes Tier wurde ein Chromosomenpaar mit je 1 Morgan Länge erzeugt. Auf jedem Chromosom lagen in gleichmäßigem Abstand von 1 cM 100 SNP mit zwei möglichen Allelen, die in der Gründergeneration zufällig vergeben wurden. Die Allelfrequenzen waren mit $P(\text{Allel1}) = P(\text{Allel2}) = 0,5$ identisch. Die Chromosomen wurden für die spätere Analyse der SNP-Effekte in eine bestimmte Anzahl gleich großer Segmente mit je gleicher Anzahl an SNPs unterteilt. Die Kombination der SNPs innerhalb eines homologen Segments entsprach hierbei jeweils einem Haplotyp.

Aus den 100 SNPs wurden nach dem Zufallsprinzip 12 Loci ausgewählt und als QTL ausgewiesen. Hierdurch wurde jeder QTL in gleichmäßigem Abstand von 1cM auf beiden Seiten von einem SNP flankiert. Die gesamte genetische Varianz wurde von den 12 QTLs erklärt. Die Allele der QTLs bekamen Werte zugeteilt; Allel 1 = 1 für einen positiven und Allel 2 = -1 für einen negativen Einfluss. Der Einfluss der einzelnen QTLs auf die Leistung der Tiere war für alle gleich groß. Die insgesamt 24 Allelwerte, für jeden Locus über beide Chromosomen, wurden für jedes Tier aufaddiert und als Genotypwert ausgegeben. Dieser Genotypwert wurde gleichzeitig auch als der wahre Zuchtwert des Tieres angenommen. Der Phänotypwert für jedes Tier ergab sich aus dem Genotypwert und einer Umweltkomponente.

$$P = G + U$$

P = Phänotyp

G = Genotyp

U = Umweltabweichung

Die Ausprägung des Phänotyps wird durch viele verschiedene Umweltkomponenten (Klima, Fütterung, Haltung, Hygiene) beeinflusst. Die Varianz der Umweltabweichung wurde anhand der Heritabilität, die bei 0,5 lag, und der berechneten genetischen Varianz geschätzt (Mrode, 1996):

$$\sigma^2_U = \sigma^2_g \times ((1-h^2)/h^2)$$

σ^2_g = genetische Varianz

h^2 = Heritabilität

Die Folgegeneration wurde erzeugt, indem alle 300 Tiere der Gründergeneration zufällig zu insgesamt 150 Paaren zusammengestellt und angepaart wurden. Für jedes Paar wurden jeweils zwei Nachkommen – ein männlicher und ein weiblicher – simuliert. Die Nachkommen ersetzen ihre Eltern in der F1-Generation. Der Markergenotyp der Eltern wurde zufällig auf die Nachkommen übertragen. Die ausgewählten QTL-Orte und ihre Allelwerte blieben über alle Generationen gleich. Die Genotyp- und Phänotypwerte wurden für alle Nachkommen berechnet.

Rekombinationen wurden simuliert, indem für jedes Segment über eine Poissonverteilung die Anzahl der Rekombinationen bestimmt wurde. Im Falle einer Rekombination wurde der Ort und der Strang definiert, wo die Rekombination stattfinden sollte und die jeweiligen „Stücke“ auf den Strängen innerhalb des Segments gegeneinander ausgetauscht. Zusätzlich wurden Mutationen mit einer Rate von $2,5 \times 10^{-3}$ unterstellt. Auch hier wurde in jedem Segment die Anzahl der stattfindenden Mutationen ausgewählt und daraufhin der Ort und der Strang festgelegt, auf dem der Austausch des Allels gegen das andere mögliche erfolgte. Es wurde bei der Mutation nicht zwischen SNP oder QTL unterschieden, womit es dem Zufall überlassen wurde, ob ein QTL mutierte.

Die auf diese Weise generierten Nachkommen bildeten die nächste Generation. Mit ihnen wurde auf gleiche Weise die nächste Generation erzeugt, bis insgesamt 15 Generationen entstanden waren.

In Generation 15 wurden für jedes der 150 Elternpaare zehn statt zwei Nachkommen erzeugt, sodass insgesamt 1500 Nachkommen in Generation 16 vorlagen. Aus den 150 Elternpaaren in Generation 15 wurden zufällig 30 Elternpaare ausgewählt. Deren 300 Nachkommen aus Generation 16 bildeten die Analysegeneration und wurden für die Berechnungen der SNP-Effekte herangezogen und in eine Ausgabedatei geschrieben.

Nach Generation 16, an der die Untersuchungen der SNP-Effekte stattfanden, wurde noch eine bestimmte Anzahl von Folgegenerationen erzeugt. Dies geschah wiederum nach dem gleichen Prinzip wie zuvor, nur dass nun wieder nur zwei Nachkommen pro Elternpaar erzeugt wurden. Des Weiteren wurden nicht nur die 300 zufällig ausgewählten Tiere aus Generation 16, die zur Analyse verwendet wurden, in der Zucht eingesetzt, sondern alle vorhandenen 1500 Tiere aus dieser Generation. Somit enthielten alle Generationen ab Generation 17 jeweils 1500 Tiere. Die grafische Darstellung der Simulations- und Auswertungsprogramme ist in Abbildung 10.

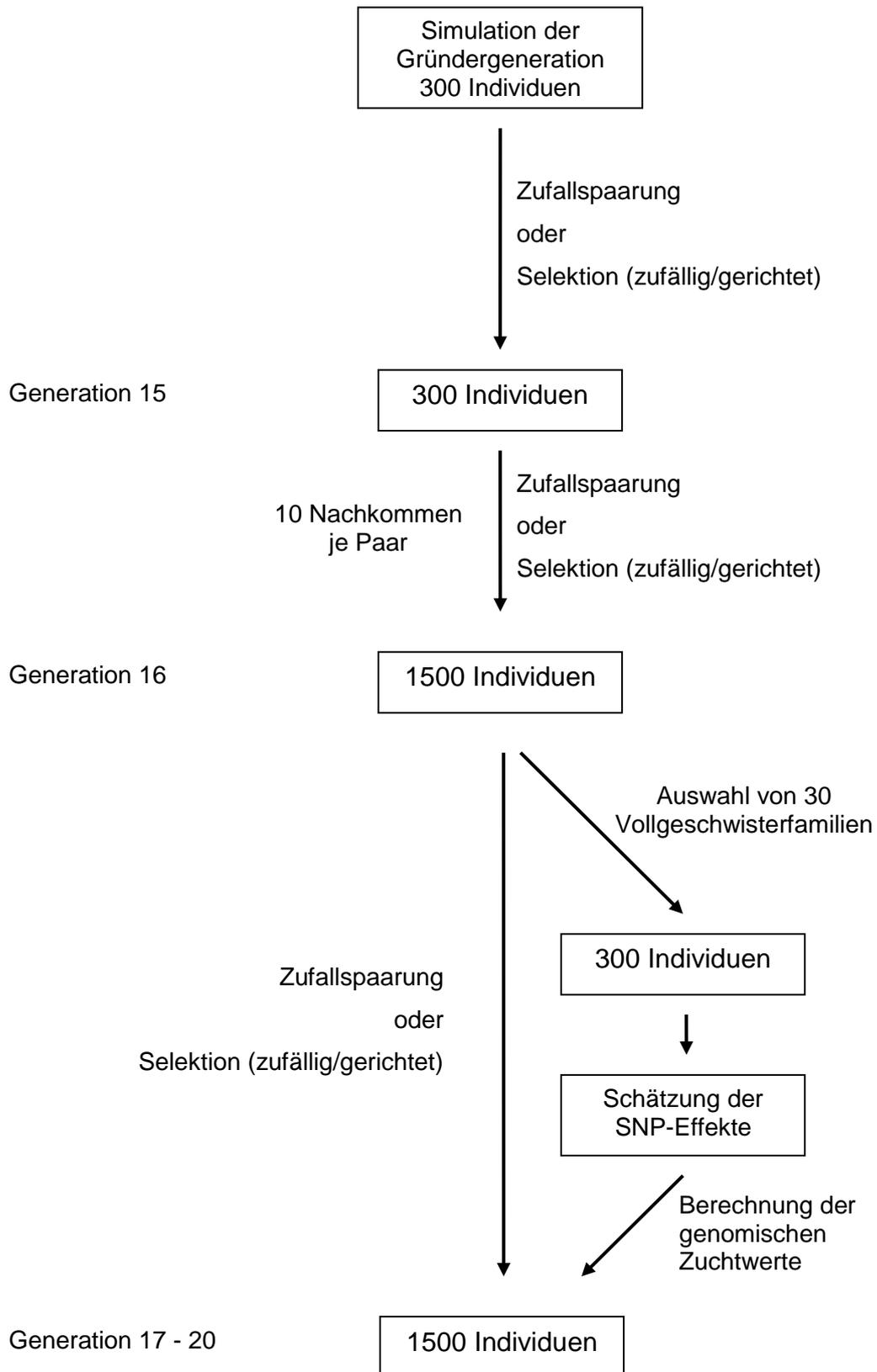


Abbildung 10: Generelle Struktur der Simulations- und Auswertungsprogramme

Sonderformen des Simulationsprogramms

Abweichung vom bisher beschriebenen Simulationsprogramm stellt die Selektion dar. In dieser Arbeit wurden zusätzlich zu der zufälligen Paarung eine zufällige und eine gerichtete Selektion nach dem Phänotyp durchgeführt. In beiden Fällen startete die Selektion in der ersten Generation, und die Anzahl der selektierten Tiere konnte für jedes Geschlecht unabhängig voneinander definiert werden. Drei verschiedene Selektionsszenarien wurden angewendet, welche in Tabelle 7 aufgezeigt werden.

Tabelle 7: Anzahl selektierter Tiere und Selektionsintensität für drei verschiedene Szenarien

	Anzahl selektierter Tiere		Selektionsintensität	
	männliche	weibliche	männliche	weibliche
Szenario 1	50	125	1,10	0,30
Szenario 2	25	100	1,52	0,56
Szenario 3	15	75	1,76	0,80

Auf der männlichen Seite wurden im ersten Szenario mit 50 Tieren nur noch 1/3 aller verfügbaren Tiere zur Zucht ausgewählt, was einer Selektionsintensität von 1,10 entsprach. Im zweiten Szenario stieg die Intensität auf 1,52 an. Es wurden nur 25 von 150 männlichen Tieren (1/6) in der Zucht eingesetzt. In Szenario 3 wurden noch 10% der Tiere ausgewählt. Die Selektionsintensität lag mit 1,76 am höchsten.

Auf der weiblichen Seite wurde in allen Szenarien eine höhere Anzahl von Tieren in der Zucht eingesetzt. Im ersten Szenario lag die Selektionsintensität mit 0,30 am niedrigsten, und es wurden 125 Tiere für die Zucht ausgewählt. Im zweiten Szenario wurde mit insgesamt zwei Dritteln (100 Tiere) der 150 möglichen weiblichen Tiere gezüchtet. Die Selektionsintensität stieg hier auf 0,56. Im dritten Szenario stieg die Intensität auf 0,80 an, und es wurde mit der Hälfte der verfügbaren Tiere (75 Tiere) gezüchtet.

Bei der zufälligen Selektion wurde in der Gründergeneration (300 Tiere) je nach Szenario die entsprechende Anzahl von männlichen und weiblichen Tieren zufällig ausgewählt. Diese ausgewählten Tiere wurden so lange zufällig zu Elternpaaren zusammengestellt und erzeugten je zwei Nachkommen bis die Anzahl von 300 Tieren der vorangegangenen Generation wieder erreicht wurde. Diese Nachkommen ersetzten dann ihre Eltern in der F1-Generation. Aus ihnen wurde wieder die entsprechende Anzahl von Tieren zufällig

ausgewählt, um die nächste Generation zu erzeugen. Die folgenden Generationen wurden auf die gleiche Weise simuliert. In Generation 15 wurden auch hier zehn statt zwei Nachkommen gebildet und 30 Elternpaare ausgewählt, deren Nachkommen zur SNP-Effekt-Analyse herangezogen wurden. Die folgenden Generationen wurden auf die gleiche Weise wie bisher erzeugt, nur dass nun mit je 750 Tieren pro Geschlecht eine größere Anzahl von möglichen Zuchttieren zur Verfügung stand, aus denen die entsprechende Anzahl von Tieren ausgewählt wurde. Damit änderten sich auch die Selektionsintensitäten für diese Generationen. Auf der männlichen Seite lagen die Intensitäten nun bei 1,99 (Szenario 1), 2,27 (Szenario 2) und 2,42 (Szenario 3). Die Selektionsintensitäten auf der weiblichen Seite fielen mit 1,49 (Szenario 1), 1,62 (Szenario 2) und 1,76 (Szenario 3) etwas niedriger aus.

Die gerichtete Selektion lief grundsätzlich wie die zufällige Selektion ab, nur dass die Tiere in jeder Generation innerhalb der Geschlechter nach dem Phänotyp rangiert wurden. Die entsprechende Anzahl der Tiere mit dem höchsten Phänotypwert pro Geschlecht wurde in der Zucht eingesetzt. Der weitere Ablauf der gerichteten Selektion war der gleiche wie bei der zufälligen Selektion bzw. wie im Kapitel zuvor (3.1) beschrieben.

3.2 Die Auswertung

Die Auswertung der Daten fand mittels verschiedener Perl und Fortran Programme statt. In einem ersten Schritt wurden sämtliche Tiere sowie ihre SNP- und Leistungsdaten aus der Datei, die für die Analyse der SNP-Daten angelegt wurde, ausgelesen. Alle Haplotypen pro Segment wurden registriert und erhielten eine Nummer. Gleiche Haplotypen pro Segment erhielten dabei die gleiche Nummer. Pro Segment und Tier kamen zwei Haplotypen vor, bei denen aufgrund der Simulation bekannt war, ob der Haplotyp vom Vater oder von der Mutter stammte. Die Nummern wurden dann anstelle der Haplotypen bzw. SNPs eingesetzt. Für die Analyse wurde eine temporäre Datei erstellt, in der die Tiernummer, die Haplotypennummer, der Geno- sowie der Phänotyp eingetragen wurden.

Für die Untersuchung der einzelnen Haplotypen und der Berechnung ihres Einflusses auf den Phänotyp wurde als Grundlage das BLUP-Verfahren benutzt. Der Vorteil seiner Anwendung lag zum einen in dem leicht umzusetzenden und zu programmierenden Modell. Außerdem war es möglich, in einer relativ kurzen Zeiteinheit sichere und eindeutige Ergebnisse zu liefern. Zum anderen war durch die Behandlung der

Haplotypeneffekte als zufällige Effekte die Möglichkeit gegeben, alle Effekte simultan zu schätzen, da dadurch keine Probleme mit zu wenigen Freiheitsgraden entstanden (Meuwissen, 2003).

Die angewendeten Programme

In dieser Arbeit wurde ein Einmerkmals-Tiermodell mit folgendem linearen Modell angewendet:

$$y = 1\mu + Zu + e$$

y = Vektor der Phänotypwerte

μ = Vektor der unbekannt, fixen, zu schätzenden Parameter; da in dieser Arbeit keine fixen Effekte vorlagen, wurde hier das Leistungsmittel μ eingesetzt

u = Vektor der unbekannt, zufälligen, zu schätzenden Parameter; hier die einzelnen Haplotypeneffekte

e = Vektor der zufälligen Resteffekte

1 = Designmatrix für fixe Matrizen; da in dieser Arbeit als fixer Effekt nur das Leistungsmittel eingesetzt wurde, stellt die Matrix einen Vektor aus Einsen dar

Z = Designmatrix der zufälligen Matrizen, welche die Leistung mit den einzelnen Haplotypen verknüpft

Die Lösungen für \hat{u} und \hat{b} ergeben sich aus folgender Mischmodellgleichung (Henderson, 1950)

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I \times \lambda \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

Der Faktor λ ist ein Schätzwert für die angenommene genetische Varianz der einzelnen Haplotypen, der an die Anzahl der Segmente und somit an die Größe der Haplotypen

angepasst wird. Für den Faktor wurde der Wert $\lambda = 400/\text{Anzahl der Segmente}$ angenommen. Die Matrix I stellt eine Einheitsmatrix dar.

Zur Lösung des Gleichungssystems wurde eine Gauss-Seidel-Iteration angewendet. Das Konvergenzkriterium lag bei 10^{-5} .

Am Ende der Analyse erhielt man für jeden Haplotypen einen Gewichtungsfaktor. In einem weiteren Schritt wurde der genomische Zuchtwert für jedes Tier über diese Gewichtungsfaktoren berechnet. Dazu wurden die Tiere der Analysegeneration sowie der Folgegenerationen eingelesen. Für jedes Tier wurden die Gewichtungsfaktoren der bei ihnen vorkommenden Haplotypen über alle Chromosomen aufaddiert und als genomischer Zuchtwert ausgegeben. Haplotypen, die in Folgegenerationen erstmalig vorkamen und für die keine Gewichtungsfaktoren vorlagen, erhielten den Faktor null.

Zur weiteren Analyse wurde aus der Abweichung zwischen dem wahren Zuchtwert und dem genomischen Zuchtwert über alle Tiere der mean squared error (MSE) sowie der Standardfehler (SE) berechnet. Des Weiteren wurde die Korrelation zwischen dem wahren Zuchtwert und dem genomischen Zuchtwert ermittelt. Für jeden Lauf wurden 100 Wiederholungen durchgeführt, sodass man einen gemittelten Wert für die Korrelation, den MSE- und SE-Wert, über alle 100 Wiederholungen erhielt.

Im weiteren Verlauf der Arbeit wurde das zuvor beschriebene BLUP-Verfahren um ein Resampling erweitert. Dabei blieben das BLUP-Verfahren sowie das verwendete Modell unverändert. Die Berechnung und Schätzung der Gewichtungsfaktoren aller Haplotypen fand allerdings nicht mehr gleichzeitig statt. Pro Runde, insgesamt 10000, wurden zufällig nur so viele Haplotypen ausgewählt, wie auch Tiere mit Leistungen zur Verfügung standen. Ausgehend von den Grundeinstellungen mit 300 Tieren in der Analysegeneration wurden so 300 Haplotypen ausgewählt. Nach der Ziehung der Haplotypen, bei der Doppelte vermieden wurden, wurde die Z-Matrix neu aufgebaut. Hierbei wurden für jedes Tier nur die Haplotypen mit in die Berechnungen einbezogen, die vorher ausgewählt wurden. Alle weiteren wurden auf „Null“ gesetzt. Sollte ein Tier die gewählten Haplotypen nicht führen, so wurde auch in diesem Fall eine Null eingetragen.

Weiterhin wurde bei diesem Resampling die Berechnung des Faktors λ , der ein Schätzwert für die angenommene genetische Varianz der Haplotypen darstellt, variabel gestaltet. Anstelle einer fixen und für alle Haplotypen gleich großen genetischen Varianz wie beim zuvor beschriebenen BLUP-Verfahren, wurde aus den unterschiedlichen Gewichtungs-Schätzwerten der Haplotypen eine genetische Varianz für jeden Haplotypen berechnet und in jeder Runde neu angepasst. Zu Beginn der Berechnungen wurde die Varianz für jeden Haplotypen auf den gleichen Wert wie beim BLUP-Verfahren gesetzt. Sobald für einen Haplotypen mehrere Schätzwerte vorlagen, wurde die Varianz neu berechnet und dieser Wert anstelle des alten eingesetzt. Bei erneutem Hinzukommen

eines weiteren Schätzwertes für einen Haplotypen wurde die Varianz für diesen Haplotypen neu berechnet und dieser Wert übernommen.

Die ersten 2000 Runden beim Resampling galten als Burn-in-Runden. Die errechneten Gewichtungsschätzwerte für die Haplotypen wurden verworfen. Erst die Ergebnisse aller späteren Runden wurden in die Berechnung der Gesamt-Gewichtungsfaktoren einbezogen. Nach Abschluss der 10000 Runden wurde die Summe der Gewichtungsfaktoren je Haplotyp durch die jeweilige Anzahl an durchgeführten Schätzungen dividiert, und man erhielt einen gemittelten Gewichtungsfaktor für jeden Haplotypen. Der genomische Zuchtwert für jedes Tier sowie die Korrelation zwischen dem wahren und genomischen Zuchtwert wurde auf die schon beschriebene Weise ermittelt.

Der Vorteil des Resamplings liegt in der angepassten genetischen Varianz für jeden Haplotypen. Da in der Wirklichkeit viele Haplotypen keinen oder einen sehr geringen Einfluss auf die genetische Varianz haben und nur einige wenige einen größeren Einfluss vorweisen können, entspricht diese Annahme eher der Realität, als feste gleich große Varianzen für alle Haplotypen anzunehmen (Meuwissen et al., 2001). Der große Nachteil dieses Verfahrens liegt in der sehr langen Laufzeit des Programms, welches schon bei diesen geringen Datenmengen zwei Wochen in Anspruch nahm.

Um die Signifikanz der Unterschiede zwischen den einzelnen Korrelationswerten sowie den MSE-Werten zu testen, wurde ein t-Test angewendet.

4 Ergebnisse und Diskussion

Für alle in diesem Kapitel beschriebenen Simulationsläufe wurden die Grundeinstellungen für die Parameter, die im Kapitel 3.1 angegeben sind, angewendet. Veränderungen dieser Einstellungen werden jeweils angegeben.

4.1 Haplotypen

Die Ergebnisse der Haplotypenregistrierung, die zu Beginn der Auswertung erfolgt, soll in den folgenden Abbildungen an einem Simulationslauf näher dargestellt werden. Über die Darstellungen soll ein besseres Verständnis für den Aufbau der einzelnen Chromosomensegmente (Anzahl Haplotypen/Segment, Häufigkeit jeden Haplotyps) erreicht werden. Für den Simulationslauf wurden die Chromosomen in zehn Segmente unterteilt, da bei dieser Segmentgröße eine relativ große Anzahl an unterschiedlichen Haplotypen zu erwarten ist (2^{10} mögliche Haplotypen). Bei einer Einstellung von 100 Segmenten wären dagegen nur noch zwei Haplotypen pro Segment möglich. Bei einer Größe von 300 Tieren in der Analysegeneration lagen pro Segment insgesamt 600 Haplotypen vor. In der nachfolgenden Abbildung 11 ist die Anzahl von unterschiedlichen Haplotypen für jedes der zehn Segmente abgebildet.

Es ist zu erkennen, dass für die meisten Segmente zwischen 90 und 100 verschiedene Haplotypen vorliegen. Einzige Ausnahme bilden die Segmente fünf und sechs. Dort sind es nur 88 Segmente. In Segment acht liegen sogar 109 unterschiedliche Haplotypen vor. Über alle Segmente gerechnet kommen in der Analysegeneration 957 verschiedene Haplotypen vor. Das ergibt einen Durchschnitt von 96 Haplotypen pro Segment, was bedeutet, dass in der Simulation pro Segment nur 9,38 % der möglichen 1024 Haplotypen erzeugt wurden.

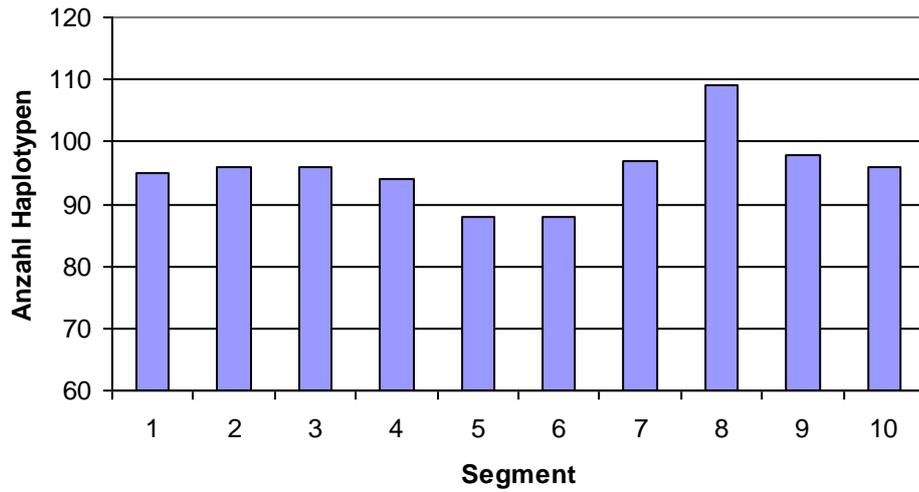


Abbildung 11: Anzahl unterschiedlicher Haplotypen je Segment

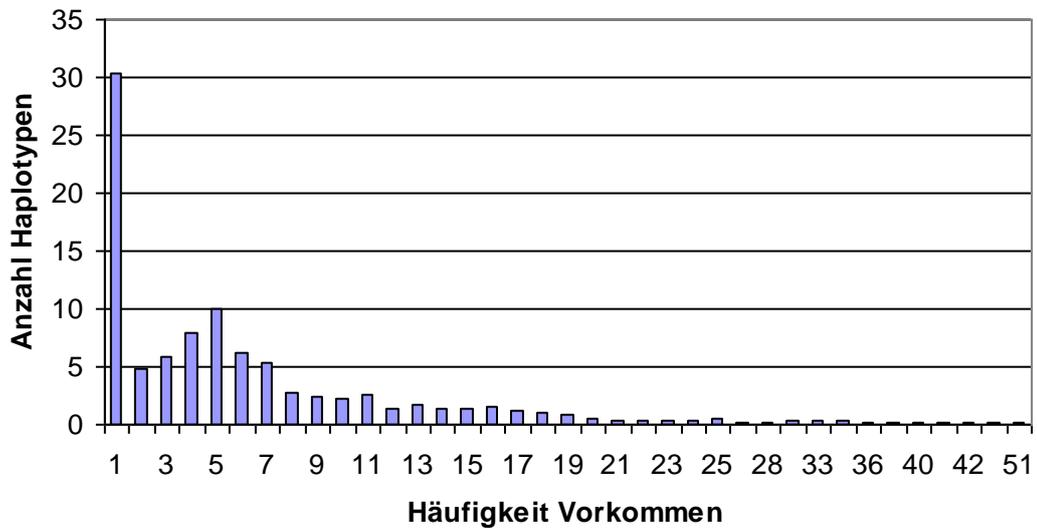


Abbildung 12: Durchschnittliche Anzahl des Vorkommens je Haplotyp über alle zehn Segmente

Abbildung 12 soll die Häufigkeit des Vorkommens der einzelnen Haplotypen über alle zehn Segmente näher darstellen. Auf der y-Achse kann man die Anzahl der Haplotypen ablesen, auf der x-Achse ist die Häufigkeit des Vorkommens für jeden der angegebenen Haplotypen aufgetragen. Man kann sehen, dass die meisten Haplotypen, über 30, nur einmal vorkommen. Die Anzahl der Haplotypen, die häufiger vorkommen nimmt dagegen stark ab, so kommen zum Beispiel nur noch knapp fünf Haplotypen je zwei Mal vor. Die Anzahl an Haplotypen die dreimal, viermal oder fünfmal vorkommen, steigt dann leicht auf sechs, acht und zehn Haplotypen an. Danach fällt die Anzahl an mehrfach vertretenen Haplotypen jedoch stark ab. Haplotypen, die öfter als zwanzigmal je Segment vertreten sind, sind eher die Ausnahme. Die Anzahl der Haplotypen liegt hier durchgängig unter eins. In diesen Fällen handelt sich um einzelne Haplotypen, die in einem Segment sehr häufig (42-, 47- oder 51-Mal) vorkommen. Allgemein ist zu sagen, dass ein großer Teil der Haplotypen, hier knapp 32%, nur ein einziges Mal in der Generation vorkommt. Den größten Anteil (knapp 50%) haben die Haplotypen, die nur wenige Male (2- bis 10-Mal) wiederholt in der Generation vorkommen. Einen geringeren Anteil mit 14% haben Haplotypen die 11- bis 20-Mal in der Generation vorkommen. Haplotypen, die mehr als 20-Mal in der Generation vertreten sind, sind selten (4%).

4.2 Einfluss der Parameter auf den genomischen Zuchtwert

4.2.1 Einfluss der Größe der Analysegeneration

In den folgenden Abbildungen wird der Einfluss der Größe sowie des Aufbaus der Analysegeneration, die zur Berechnung der Gewichtungsfaktoren benutzt wird, auf den MSE-Wert und die Korrelation zwischen dem wahren und genomischen Zuchtwert dargestellt. Hierzu wurde jeweils die Größe der Stichprobe verändert, indem unterschiedliche Anzahlen von 10, 20, 30 oder 50 Elternpaaren aus Generation 15 ausgewählt wurden. Zusätzlich wurde auch die Anzahl der Nachkommen pro Elternpaar in Generation 15 verändert. Hier wurden 10, 20 oder 30 Nachkommen gewählt. Durch Kombination der beiden Parameter konnte so eine unterschiedlich große Generation 16 erzeugt werden. Die geringste Anzahl von Tieren (100) wurde mit 10 Elternpaaren mit je 10 Nachkommen erzielt. Mit 1500 Tieren bei 50 Elternpaaren und je 30 Nachkommen war die höchste Anzahl erreicht. Die Anzahl der Segmente lag bei 10.

In Abbildung 13 wird der Einfluss der beiden Parameter auf die MSE-Werte aufgezeigt, wobei je eine Linie für eine Anzahl von Elternpaaren steht.

Betrachtet man alle vier Linien für die verschiedenen Anzahlen von Elternpaaren gemeinsam, so fällt auf, dass die Werte mit steigender Nachkommenanzahl sinken. Je nach Anzahl der Elternpaare ist die Veränderung unterschiedlich stark ausgeprägt. Der größte Abfall der MSE-Werte ist bei 10 Elternpaaren zu erkennen. Hier sinkt der Wert von 12,19 bei 10 Nachkommen auf 9,12 bei 30 Nachkommen. Bei den anderen Elternpaaren ist der Abfall geringer. Die Werte mit 30 Nachkommen sind aber immer die niedrigsten Werte. Generell fällt die Abnahme der MSE-Werte von 10 auf 20 Nachkommen über alle Elternpaaranzahlen größer aus als von 20 auf 30 Nachkommen. Das bestätigen auch die Signifikanzen zwischen den Nachkommenanzahlen (Tabelle 8). Zwischen 10 und 20 Nachkommen fallen die Unterschiede der MSE-Werte für alle Elternpaaranzahlen signifikant, teilweise hochsignifikant aus. Dagegen sind die Unterschiede zwischen 20 und 30 Nachkommen nur für 30 und 50 Elternpaare signifikant und auch hier nur für 30 Elternpaare hochsignifikant.

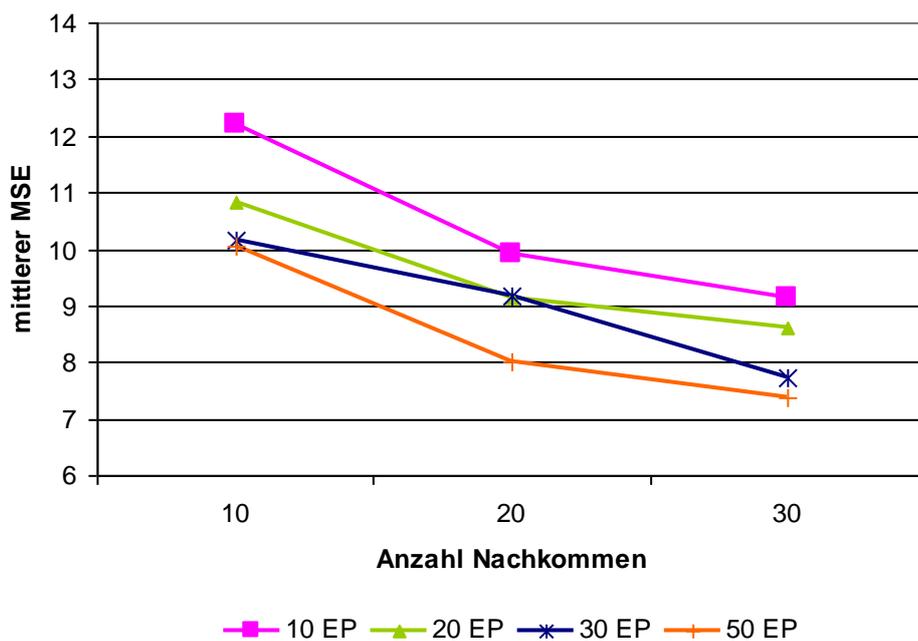


Abbildung 13: Einfluss der Anzahl Elternpaare (EP) und Nachkommen auf die MSE-Werte

Tabelle 8: Signifikanzen (t-Test) zwischen Nachkommenanzahlen für alle Elternpaaranzahlen (EP) für MSE- und Korrelationswerte (r)

		zwischen Nachkommenanzahlen					
		10 und 20		10 und 30		20 und 30	
		MSE	r	MSE	r	MSE	r
Anzahl EP	10	***	***	***	***	ns	**
	20	**	***	***	***	ns	***
	30	*	***	***	***	***	***
	50	***	***	***	***	*	***
nicht signifikant (ns): $p > 0,05$;		*: $p \leq 0,05$		**: $p \leq 0,01$		***: $p \leq 0,001$	

Ein nahezu kontinuierlicher Abfall der Werte ist auch mit steigender Anzahl von Elternpaaren zu erkennen. Die Werte für 10 Elternpaare fallen am höchsten aus, die für 50 Elternpaare am niedrigsten. Die Werte von 20 und 30 Elternpaaren liegen zwischen den beiden anderen Elternpaarwerten. Einzige Ausnahme bildet hier der Wert von 30 Elternpaaren mit 20 Nachkommen, der sogar leicht über dem von 20 Elternpaaren und 20 Nachkommen liegt. Insgesamt fallen die Unterschiede zwischen den Elternpaaren bei 10 Nachkommen am größten aus und verringern sich dann mit steigender Nachkommenanzahl.

Die Unterschiede zwischen direkt benachbarten Elternpaaranzahlen fallen nur selten signifikant aus (Tabelle 9). Gerade bei 10 Nachkommen sind fast alle Unterschiede zwischen den Elternpaaranzahlen nicht signifikant, nur bei größeren Elternpaarunterschieden, 10 und 30 sowie 10 und 50. Erst bei höheren Nachkommen- und Elternpaaranzahlen fallen die Unterschiede signifikant bis hochsignifikant aus.

Betrachtet man Werte mit gleicher Anzahl von Tieren, wie 10 Elternpaare mit 20 Nachkommen und 20 Elternpaare mit 10 Nachkommen, so fällt deutlich auf, dass der Wert für 10 Elternpaare mit 9,91 niedriger ausfällt als der für 20 Elternpaare (10,84). Das gleiche Ergebnis bekommt man auch für 600 Tiere. Der Wert von 20 Elternpaaren mit 30 Nachkommen (8,61) fällt niedriger aus als der von 30 Elternpaaren mit 20 Nachkommen (9,17). Allerdings sind die Unterschiede für beide Vergleiche nicht signifikant.

In Abbildung 14 ist der Einfluss auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert dargestellt. Es ist zu sehen, dass die Korrelation mit steigender Nachkommenanzahl für alle Elternpaaranzahlen zunimmt. Bei 10 Nachkommen liegen die Werte im Bereich von $r = 0,80 - 0,81$ und steigen bei 30 Nachkommen bis auf $r = 0,84 -$

0,87 an. Man kann erkennen, dass der Anstieg der Korrelationen von 10 auf 20 Nachkommen leicht höher ausfällt als für 20 auf 30 Nachkommen. Alle Unterschiede zwischen den Nachkommenanzahlen sind für alle Elternpaaranzahlen hochsignifikant.

Tabelle 9: Signifikanzen (t-Test) zwischen Elternpaaranzahlen für alle Nachkommenanzahlen (NK) für MSE- und Korrelationswerte (r)

		zwischen Elternpaaranzahlen											
		10 und 20		10 und 30		10 und 50		20 und 30		20 und 50		30 und 50	
		MSE	r	MSE	r	MSE	r	MSE	r	MSE	r	MSE	r
Anzahl NK	10	ns	ns	**	ns	**	***	ns	ns	ns	***	ns	***
	20	ns	ns	ns	***	***	***	***	**	*	***	*	***
	30	ns	**	**	***	***	***	*	*	***	***	ns	***

nicht signifikant (ns): $p > 0,05$; *: $p \leq 0,05$ **: $p \leq 0,01$ ***: $p \leq 0,001$

Betrachtet man die Unterschiede zwischen den einzelnen Anzahlen der Elternpaare, ist zu erkennen, dass die Unterschiede bei 10 Nachkommen wesentlich geringer ausfallen als bei 30 Nachkommen. Bei 10 Nachkommen liegen alle Werte der Elternpaare nahe zusammen und unterscheiden sich nur um $r = 0,01$. Hier fallen die Unterschiede auch nur zwischen 10 und 50 sowie 20 und 50 Elternpaaren signifikant aus. Zwischen allen anderen Elternpaaranzahlen sind die Unterschiede nicht signifikant. Erst bei 20 und 30 Nachkommen sind deutlichere Unterschiede zu erkennen. Hier fallen die Unterschiede dann auch signifikant aus. Die Unterschiede zwischen den Elternpaaranzahlen nehmen also mit steigender Nachkommenanzahl zu. Generell kann man sagen, je höher die Anzahl der Elternpaare, desto höher auch die Korrelationen.

Betrachtet man hier Werte mit gleicher Tieranzahl - 10 Elternpaare mit 20 Nachkommen und 20 Elternpaare mit 10 Nachkommen - so fällt auf, dass der Korrelationswert von 10 Elternpaaren sehr viel höher ausfällt als der für 20 Elternpaare. Auch bei einem Vergleich von 600 Tieren sind die Werte von 20 Elternpaaren mit 30 Nachkommen höher als die von 30 Elternpaaren mit 20 Nachkommen. Die Unterschiede fallen für beide Tieranzahlen hochsignifikant aus. Diese Ergebnisse lassen darauf schließen, dass die Anzahl der Nachkommen eine höhere Bedeutung für die Schätzung der Gewichtungsfaktoren hat als die Anzahl der Elternpaare. Je mehr Informationen von verwandten Tieren vorliegen, desto genauer kann der Einfluss eines Haplotyps auf die Leistung geschätzt werden.

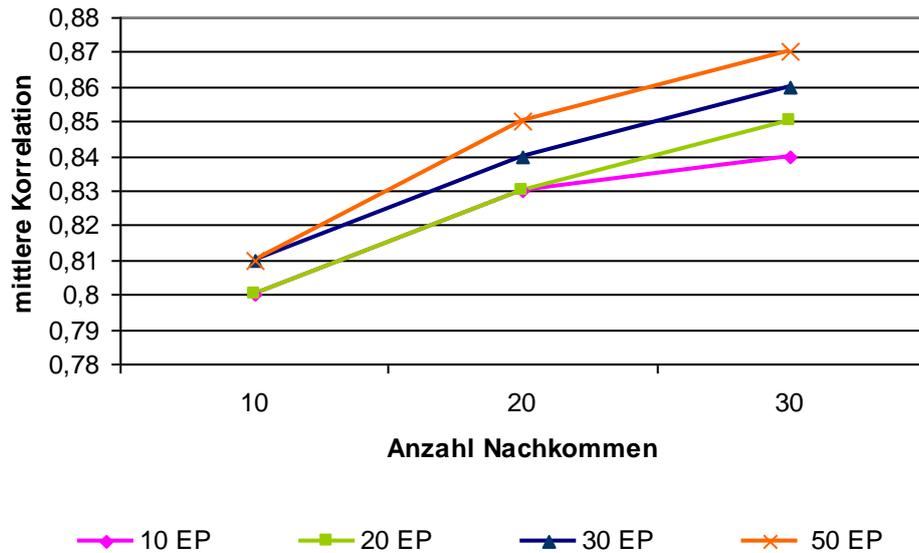


Abbildung 14: Einfluss der Anzahl der Elternpaare (EP) und der Nachkommen auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert

Aus den Ergebnissen über den Einfluss der Größe und Struktur der Analysegeneration ist festzuhalten, dass der höchste MSE-Wert und die niedrigste Korrelation bei 10 Elternpaaren und 10 Nachkommen zu finden sind, während der niedrigste MSE-Wert und die höchste Korrelation bei 50 Elternpaaren und 30 Nachkommen liegen. Man kann also davon ausgehen, dass die Gewichtungsfaktoren für die einzelnen Haplotypen umso genauer geschätzt werden, je mehr Tiere in der Analysegeneration vorhanden sind. Durch die erhöhte Anzahl von Tieren sind mehr Informationen für jeden Haplotypen verfügbar, und der Einfluss des Haplotyps auf die Leistung kann besser geschätzt werden. Mit steigender Anzahl von Tieren fällt dieser Einfluss nicht mehr so hoch aus. Die MSE-Werte bei 30 Nachkommen (9,12 für 10 Elternpaare zu 7,39 bei 50 Elternpaaren) unterscheiden sich nicht mehr so stark wie die bei 10 Nachkommen (12,19 bei 10 Elternpaaren zu 10,05 bei 50 Elternpaaren). Die Werte für Korrelation liegen alle über 0,8 und fallen damit sehr hoch aus. Der Unterschied in der Genauigkeit zwischen 30 Elternpaaren mit 30 Nachkommen ($r = 0,86$) und 50 Elternpaaren mit 30 Nachkommen ($r = 0,87$) fällt trotz hoher Anstiege an Tierzahlen (1500 gegenüber 900) sehr gering aus. Die gleiche Zunahme der Korrelation konnte schon bei 30 Nachkommen zwischen 20 und 30 Elternpaaren erreicht werden, wobei die Anzahl der Tiere hier nur um 300 zunimmt. Aus diesem Grund ist zu überlegen, ob eine Typisierung aller Tiere notwendig ist, da schon mit weniger Tieren sehr gute Ergebnisse erzielt werden können. Generell kann die

Genauigkeit durch mehr Nachkommen pro Elternpaar effektiver erhöht werden als durch eine Erhöhung der Elternpaaranzahlen.

Meuwissen et al. (2001) schätzten genomische Zuchtwerte an drei verschiedenen Tieranzahlen (500, 1000 und 2200) für eine Heritabilität von 0,3. Hier nahmen die Genauigkeiten für die genomischen Zuchtwerte von 0,579 über 0,659 auf 0,732 für die BLUP-Methode zu. Auch für die beiden anderen Methoden (Least Squares und BayesB) konnte ein Anstieg der Werte verzeichnet werden. Calus und Veerkamp (2007) kamen zu ähnlichen Ergebnissen. Die Anzahl von Tieren anhand derer die Haplotypeneffekte geschätzt wurden, wurde von 500 in mehreren Schritten auf 1100 erhöht. Für eine Heritabilität von 0,5 lag die Korrelation zwischen dem wahren und genomischen Zuchtwert bei 0,73. Sie stieg um 0,01 durch Hinzunahme von 50 Tieren. Wurden insgesamt 500 Tiere hinzugenommen, so konnte die Korrelation auf $r = 0,79$ ansteigen. Eine weitere Steigerung der Tierzahlen führte zu keiner höheren Genauigkeit. Für eine niedrige Heritabilität von 0,1 waren die Anstiege der Korrelation sehr viel deutlicher. Mit steigender Tierzahl (von 500 auf 1000 Tiere) stieg die Korrelation von 0,38 auf 0,53. Durch 50 bzw. 100 weitere Tiere stieg sie nochmals um je 0,01. Ein Einfluss der Tierzahlen auf die Genauigkeit des genomischen Zuchtwertes konnte somit gezeigt werden. Besonders für niedrige Heritabilitäten sind hohe Tierzahlen wichtig. Verglichen mit den Korrelationen aus dieser Arbeit fallen die Werte bei Meuwissen et al. (2001) und bei Calus und Veerkamp (2007) niedriger aus. Alle Werte lagen in dieser Arbeit bei größer oder gleich 0,80. Allerdings wurde von einer hohen Heritabilität von 0,5 ausgegangen und die genomischen Zuchtwerte beziehen sich auf die Analysegeneration und nicht auf die nachfolgenden Tiere ohne Phänotyp. Bei Calus und Veerkamp (2007) wurden für die Tiere der Analysegenerationen höhere Genauigkeiten von $r = 0,84 - 0,93$ für $h^2 = 0,5$ und $r = 0,47 - 0,65$ für $h^2 = 0,1$ berechnet. Verglichen mit dem Wert aus dieser Arbeit für 1000 Tiere ($r = 0,85$), liegen die Korrelationen von Calus und Veerkamp (2007) bei 1000 Tieren mit $r = 0,87$ nur leicht höher. Bei 550 Tieren ($r = 0,84$) zu 600 Tieren ($r = 0,84-0,85$) in dieser Arbeit liegt die Korrelation sogar gleich bis etwas höher.

Auch Muir (2007) hat seine Tierzahlen durch Erhöhung der Trainingsgenerationen vergrößert. Dabei lag seine Generationsgröße mit 128 Tieren relativ niedrig. Durch Verwendung von vier statt einer Trainingsgeneration konnte die Genauigkeit von $r = 0,2$ auf $r = 0,7 - 0,8$ angehoben werden. Somit konnte Muir (2007) mit erheblich weniger Tieren gleiche Genauigkeiten wie bei Meuwissen et al. (2001) erreichen. Die Korrelationswerte von Muir (2007) in der 1. Analysegeneration lagen zwischen $r = 0,3$ und $r = 0,4$, was verglichen mit den Werten für 100 Tiere ($r = 0,8$) aus dieser Arbeit sehr viel niedriger ausfiel. Weiterhin konnte Muir (2007) zeigen, dass es effektiver war, eine höhere

Anzahl von Trainingsgenerationen zu verwenden als mehr Tiere je Generation zu analysieren.

Bei Solberg et al. (2008) wurde die Analysegeneration von 1000 auf 2000 Tiere erweitert. Auch hier konnte eine höhere Genauigkeit (von $r = 0,77$ auf $0,84$) durch die größere Anzahl von Tieren erzielt werden. Die Untersuchung fand allerdings nur anhand von Mikrosatelliten statt.

4.2.2 Einfluss der Anzahl der Generationen

Der Einfluss der Anzahl der Generationen auf den MSE-Wert mit SE-Wert und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert ist in der folgenden Abbildung 15 dargestellt. Verändert wurde die Anzahl der Generationen bis zur Analysegeneration. Sechs verschiedene Szenarien wurden simuliert: 16, 30, 50, 100, 250 sowie 500 Generationen. Die Segmentanzahl betrug 10.

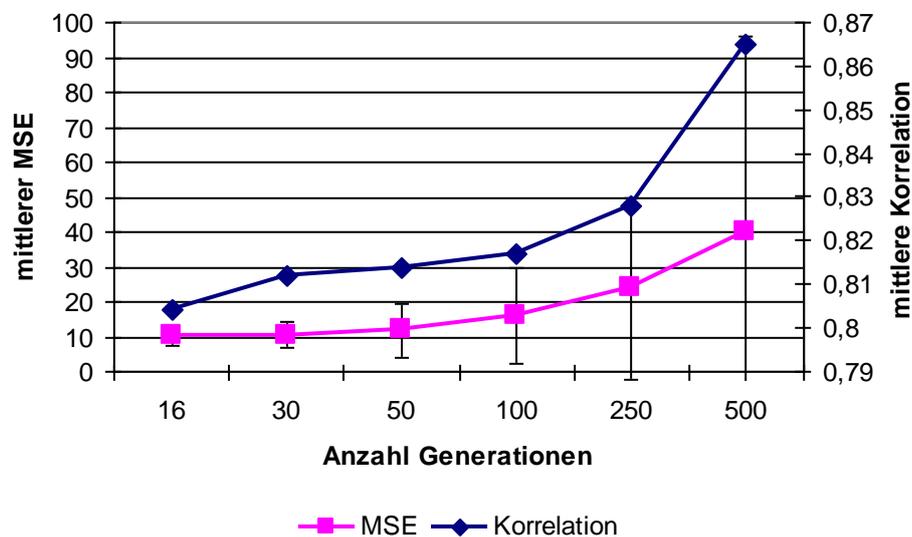


Abbildung 15: Einfluss der Anzahl der Generationen auf die MSE-Werte (mit SE) und der Korrelation zwischen dem wahren und dem genomischen Zuchtwert

Man kann sehen, dass die MSE-Werte mit steigender Anzahl der Generationen zunehmen. Der Wert steigt von 10,28 bei 16 Generationen auf 39,73 bei 500 Generationen. Auch der SE-Wert nimmt mit steigender Generationenanzahl extrem stark zu. Liegt der Wert bei 16 Generationen noch bei 2,83, so vergrößert er sich auf 56,03 bei 500 Generationen. Die Unterschiede zwischen den MSE-Werten fallen signifikant bis hochsignifikant aus (Tabelle 10). Nur zwischen 16 und 30 sowie zwischen 30 und 50 Generationen sind die Unterschiede nicht signifikant.

Tabelle 10: Signifikanzen (t-Test) zwischen den Generationenanzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen

	Anzahl Generationen					
	16	30	50	100	250	500
16		ns	*	***	***	***
30	*		ns	***	***	***
50	*	ns		*	***	***
100	**	ns	ns		**	***
250	***	***	***	**		**
500	***	***	***	***	***	

nicht signifikant (ns): $p > 0,05$; *: $p \leq 0,05$ **: $p \leq 0,01$ ***: $p \leq 0,001$

Auch die Korrelation zwischen dem wahren und dem genomischen Zuchtwert nimmt mit steigender Generationenanzahl zu. Bei 16 Generationen liegt sie bei 0,80 und steigt bis auf 0,86 bei 500 Generationen an. Dabei ist der Anstieg zwischen 30 und 100 Generationen weniger ausgeprägt als von 16 auf 30 und von 100 bis 500 Generationen. Dieser Eindruck wird auch durch die Signifikanzen bestärkt (Tabelle 10). Sämtliche Unterschiede zwischen den einzelnen Generationenanzahlen sind signifikant bis hochsignifikant. Nur zwischen 30 und 50, 30 und 100 sowie 50 und 100 Generationen sind die Unterschiede nicht signifikant. Ein Einfluss der Generationenanzahl auf die Genauigkeit des genomischen Zuchtwertes ist somit gegeben. Allerdings fällt die Erhöhung der Genauigkeit bei Generationenanzahlen im mittleren Bereich (30 – 100 Generationen) sehr gering aus. Eine höhere Zunahme der Genauigkeit ist erst ab hohen Generationenanzahlen zu realisieren.

Außer bei Solberg et al. (2008) wurden in keiner anderen wissenschaftlichen Arbeit Simulationen mit unterschiedlichen Generationenanzahlen gemacht. Meuwissen et al.

(2001) und Calus und Veerkamp (2007) simulierten 1000 Generationen vor den Analysegenerationen. Muir (2007) arbeitete mit 5000 Generationen, während Kolbehdari et al. (2007) nur 25 Generationen simulierten. Bei Kolbehdari et al. (2007) lagen die Korrelationen zwischen dem wahren und dem genomischen Zuchtwert bei $r = 0,77$. Verglichen mit dem Wert aus dieser Arbeit, der für 16 und 30 Generationen um die $r = 0,80$ schwankte, fiel er somit kleiner aus. Allerdings beschränkten sich deren Daten ausschließlich auf 10000 Töchterinformationen, womit sie schlecht mit den Daten aus dieser Arbeit zu vergleichen waren. Bei Muir (2007) fielen die Genauigkeiten mit $r = 0,3 - 0,4$ extrem gering aus und lagen unterhalb der Werte aus dieser Arbeit für alle Generationenanzahlen. Allerdings wurden die Haplotypeneffekte an 128 gegenüber 300 Tieren in dieser Arbeit geschätzt, was den Unterschied zwischen den Werten ausmachen kann. Meuwissen et al. (2001) erhielten eine Genauigkeit von $r = 0,579$ bei 1000 Generationen und 500 Tieren. Auch dieser Wert lag weit unter dem Wert aus dieser Arbeit für 500 Generationen und 300 Tiere ($r = 0,865$). Allerdings berechneten Meuwissen et al. (2001) die Werte schon für eine Folgegeneration und für eine niedrigere Heritabilität ($h^2 = 0,3$), was in dieser Arbeit nicht der Fall ist und was den Unterschied zwischen den Werten erklärt. Calus und Veerkamp (2007) schätzten die Genauigkeit auf $r = 0,82$ bei 500 Tieren. Auch dieser Wert lag damit unter dem in dieser Arbeit. Erst bei der Analyse von 1000 Tieren lagen die Werte mit $r = 0,85-0,86$ auf einem nahezu gleichen Level mit denen in dieser Arbeit.

Solberg et al. (2008) testeten verschiedene Anzahlen von Generationen von 1 bis 1000. Sie konnten zeigen, dass mit zunehmender Anzahl von Generationen die Kopplung zwischen Markern und QTL von fast null auf 0,2 stieg und damit verbunden die Genauigkeiten der genomischen Zuchtwerte von $r = 0,4$ auf 0,69 zunahmen. Dieses Ergebnis stimmt damit völlig mit dem gezeigten Ergebnis in dieser Arbeit überein.

4.2.3 Einfluss der Populationsgröße

Den Einfluss der Größe der Population auf den MSE-Wert mit SE-Wert und die Korrelation zwischen dem wahren und genomischen Zuchtwert ist in Abbildung 16 dargestellt. Verändert wurde die Anzahl an männlichen und weiblichen Tieren pro Generation bis zur Analysegeneration. Neben den Simulationen für je 150 Tiere pro Geschlecht wurde die Anzahl auf je 250, 500 sowie 1000 Tiere erhöht.

Aus der Abbildung ist ersichtlich, dass der MSE-Wert von 10,28 bei 150 Tieren auf 9,77 bei 1000 Tieren abfällt. Die Veränderung fällt mit insgesamt 0,51 sehr gering aus.

Sämtliche Unterschiede zwischen den MSE-Werten waren nicht signifikant (Tabelle 11). Die SE-Werte fallen mit steigender Populationsgröße viel geringer aus und sind von 2,83 bei 150 Tieren auf 1,23 bei 1000 Tieren abgefallen.

Die Korrelation zwischen dem wahren und dem genomischen Zuchtwert fällt von 0,804 (150 Tiere) auf 0,8 und 0,798 bei 250 und 500 Tieren ab. Bei 1000 Tieren steigt sie dann wieder auf $r = 0,805$ an. Die Änderungen sind alle sehr gering und bis auf den Unterschied zwischen 500 und 1000 Tieren nicht signifikant (Tabelle 11).

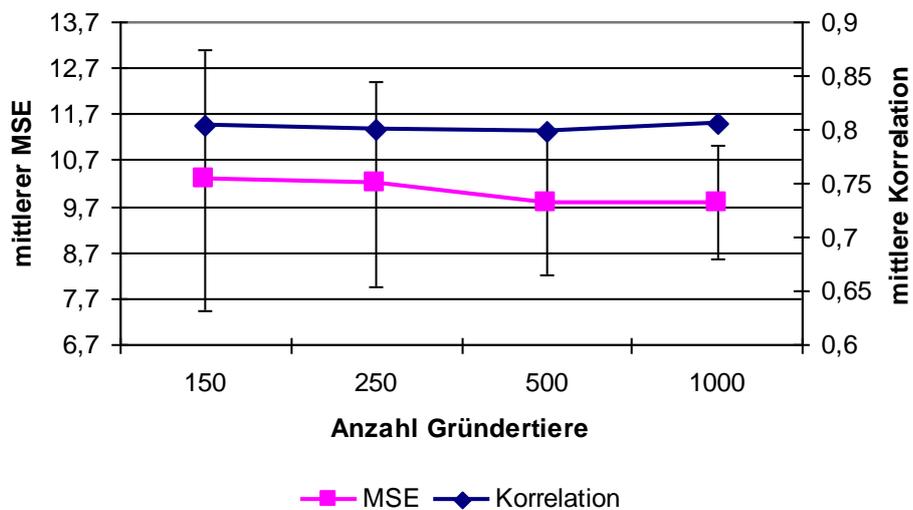


Abbildung 16: Einfluss der Generationsgröße auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert

Tabelle 11: Signifikanzen (t-Test) zwischen den Gründertieranzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen

		Anzahl Gründertiere			
		150	250	500	1000
Anzahl Gründertiere	150		ns	ns	ns
	250	ns		ns	ns
	500	ns	ns		ns
	1000	ns	ns	*	

nicht signifikant (ns): $p > 0,05$; *: $p \leq 0,05$ **: $p \leq 0,01$ ***: $p \leq 0,001$

Muir (2007) simulierte vier unterschiedliche Populationsgrößen. In den Generationen vor den vier Trainingsgenerationen wurde mit 16, 32, 64 oder 128 Tieren je Geschlecht gezüchtet. Die Korrelationen lagen zwischen 0,6 für Populationen im Hardy-Weinberg-Gleichgewicht und 0,7 für Populationen im Mutations-Drift-Gleichgewicht bei 16 Zuchttieren und fielen bis auf 0,3 – 0,5 für 128 Zuchttiere ab. Somit fiel die Genauigkeit bei Muir (2007) erheblich stärker ab als es in dieser Arbeit der Fall war, allerdings wurden bei Muir (2007) auch sehr viel weniger Tiere in der Zucht eingesetzt. Der höchste Wert von 128 Tieren kommt dem niedrigsten in dieser Arbeit mit 150 Tieren am nächsten. Hier lag der Wert mit $r = 0,804$ sehr viel höher als der von Muir (2007), allerdings sind die Werte der beiden Arbeiten sehr schwer miteinander vergleichbar, da bei Muir (2007) eine geringe Heritabilität von $h^2 = 0,1$ verwendet wurde. Muir (2007) sieht in der Populationsgröße einen wichtigen Parameter, der besonders für Populationen im Hardy-Weinberg-Gleichgewicht von Bedeutung ist, da hier die Genauigkeit mit steigender Populationsgröße stark abfällt. Calus und Veerkamp (2007) gaben an, dass höhere effektive Populationsgrößen zu einem geringeren LD zwischen SNP und QTL führen, während kleinere effektive Populationsgrößen zu einem höheren LD führen (Zhao et al., 2005). Sie gehen daher davon aus, dass bei höheren Populationsgrößen ein geringerer Teil der genetischen Varianz von den Haplotypen erklärt wird und somit die Genauigkeit der genomischen Zuchtwerte abfällt.

Auch bei Solberg et al. (2008) wurde eine erhöhte Populationsgröße getestet. Sie konnten zeigen, dass die Genauigkeit der genomischen Zuchtwerte mit steigender Populationsgröße von 100 auf 200 Tiere um 0,013 abnahm und führten das auf eine sinkende Kopplung zwischen Markern und QTL zurück.

Ausgehend von den Ergebnissen in dieser Arbeit, bei denen der Einfluss der Populationsgröße auf den MSE-Wert und die Korrelation nur sehr gering und nicht signifikant war, können die Ergebnisse von Muir (2007) nicht in dem Umfang bestätigt werden.

4.2.4 Einfluss der Segmentgröße

In der folgenden Abbildung 17 wird der Einfluss der Segmentanzahl auf den MSE-Wert mit SE und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert gezeigt. Insgesamt wurden fünf verschiedene Segmentgrößen untersucht. Vier Segmente mit je 25 SNPs/Segment, zehn Segmente mit je 10 SNPs/Segment, zwanzig Segmente mit je 5 SNPs/Segment, fünfzig Segmente mit je 2 SNPs/Segment und 100 Segmente, was einer Einzel-SNP-Analyse entspricht.

Beim MSE-Wert kann ein deutlicher Abfall von vier (18,24) auf einhundert (8,9) Segmente verzeichnet werden. Der stärkste Abfall des Wertes findet von vier (18,24) auf zehn (10,15) Segmente statt. Die Veränderungen des Wertes fallen mit größerer Segmentanzahl dann sehr viel geringer aus. Von zehn auf zwanzig Segmente liegt der Abfall bei 0,79. Der Wert von fünfzig Segmenten steigt auf 10,42 wieder deutlich an, liegt sogar über dem von zehn Segmenten. Der MSE-Wert für 100 Segmente nimmt mit 8,9 den niedrigsten Wert ein. Die Unterschiede zwischen den Segmenten sind häufig hochsignifikant. Nur zwischen 10 und 20 sowie 10 und 50 Segmenten, aber auch zwischen 20 und 50 sowie 20 und 100 Segmenten sind die Unterschiede nicht signifikant (Tabelle 12).

Der Standardfehler ist bei vier Segmenten mit 4,24 am größten und fällt auf 2,93 bei zehn Segmenten ab. Bei zwanzig Segmenten steigt er um 0,37 an, fällt bei fünfzig Segmenten wieder auf den früheren Wert (2,95) und nimmt mit 2,44 bei 100 Segmenten den niedrigsten Wert ein.

Die Korrelation zwischen dem wahren und dem genomischen Zuchtwert hat den niedrigsten Wert bei vier Segmenten mit 0,79. Die Werte steigen mit größerer Segmentanzahl kontinuierlich an und nehmen bei 100 Segmenten ($r = 0,87$) den höchsten Wert an. Hier fallen sämtliche Unterschiede zwischen den Segmentanzahlen hochsignifikant aus (Tabelle 12).

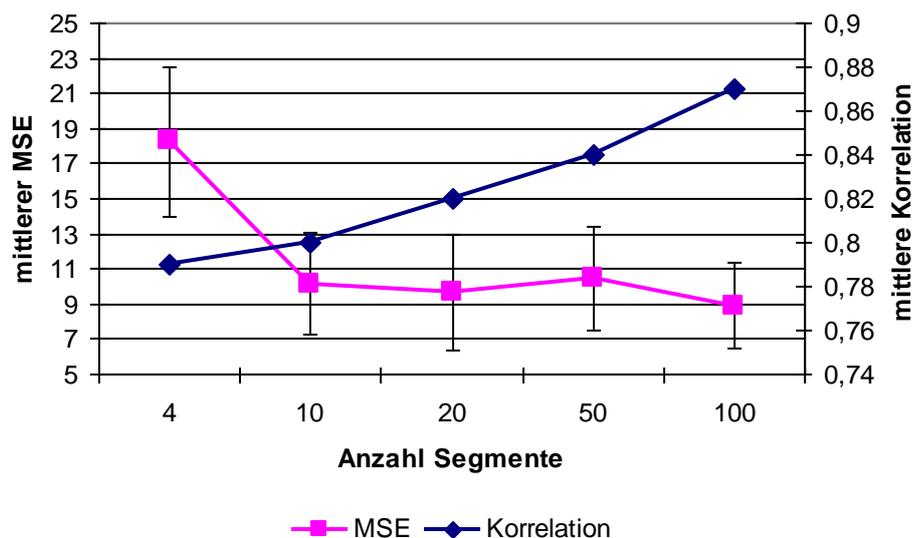


Abbildung 17: Einfluss der Segmentanzahl auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und genomischen Zuchtwert

Der hohe MSE-Wert von 18,24 bei vier Segmenten lässt sich durch die relativ großen Segmente von 25 SNPs/Segment erklären. Bei nur vier Segmenten und zwölf vorliegenden QTL können durchschnittlich drei QTL pro Segment vorkommen. Da die QTL über den Genotypwert die Leistung eines Tieres mitbestimmen, ist auch ein Einfluss auf den Effekt der Haplotypen gegeben. Bei Vorhandensein mehrerer QTL pro Segment werden ihre Einflüsse zusammen geschätzt, was zu Ungenauigkeiten in der Bewertung der Haplotypeneffekte führen kann. Aus diesem Grund fällt der MSE-Wert bei vier Segmenten viel höher aus als bei den höheren Segmentanzahlen. Bei zehn Segmenten ist die Wahrscheinlichkeit größer, dass in jedem Segment nur ein QTL liegt. Die Effekte des Haplotyps können so genauer geschätzt werden.

Tabelle 12: Signifikanzen (t-Test) zwischen den Segmentanzahlen – für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen

		Anzahl Segmente				
		4	10	20	50	100
Anzahl Segmente	4		***	***	***	***
	10	***		ns	ns	**
	20	***	***		ns	ns
	50	***	***	***		***
	100	***	***	***	***	

Weiterhin ist bei 4 Segmenten mit je 25 SNPs/Segment die Anzahl an möglichen Haplotypen je Segment mit über 33 Millionen extrem hoch. Viele der Haplotypen werden nur ein einziges Mal in der Population vertreten sein, und nur wenige kommen mehrfach vor (siehe Abbildung 12). Eine sichere Schätzung der Haplotypeneffekte ist somit kaum gegeben.

Zusätzlich nimmt die Wahrscheinlichkeit einer Kopplung der SNPs an einen QTL mit Größe der Segmente bzw. Entfernung zum QTL ab. Bei 25 SNPs in einem Segment liegt die Größe des Segments bei 25 cM. Der Wert der Kopplung r^2 für ein SNP am Anfang des Segments mit einem SNP am Ende des Segments liegt hier nur noch bei einem sehr niedrigen Wert von $< 0,018$. Bei kleineren Segmenten und somit kurzen Entfernungen steigt der Wert an und liegt für 1 cM bei $r^2 = 0,206$ (Du et al., 2007).

Die sichersten Schätzwerte mit dem geringsten MSE-Wert und der höchsten Korrelation erreicht man somit bei 100 Segmenten. Mit einer Korrelation von 0,87 zwischen dem wahren und dem genomischen Zuchtwert, bei einer Größe von 300 Tieren in der Analysegeneration erhält man äußerst genaue Schätzwerte für die einzelnen SNPs. Zur Verdeutlichung sind in den folgenden beiden Abbildungen 18 und 19 die Haplotypen- bzw. SNP-Effekte für 50 und 100 Segmente dargestellt.

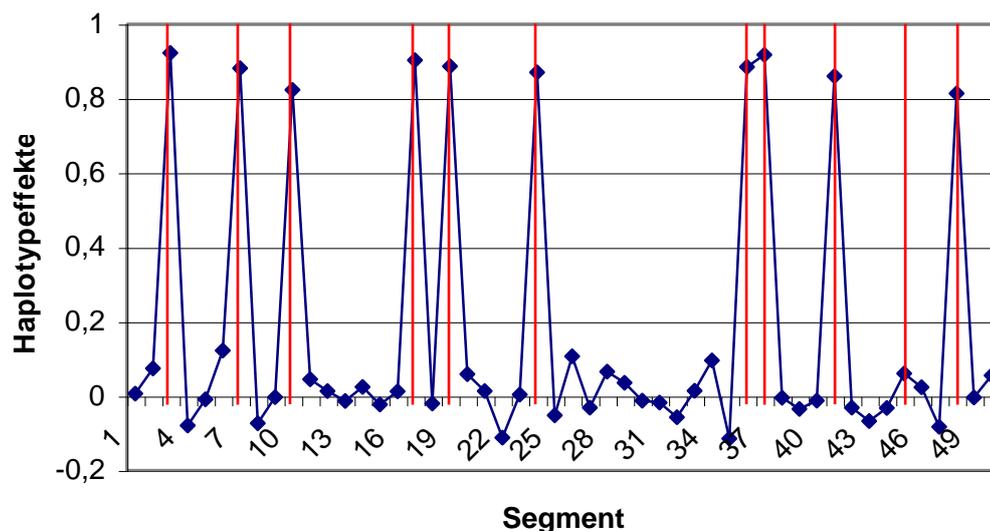


Abbildung 18: Haplotypeneffekte für 50 Segmente (QTL-Positionen rot)

Für den Vergleich wurden für beide Simulationen die QTL-Positionen für alle 100 Wiederholungen an den gleichen Positionen gewählt. In Abbildung 18 wurde die Analyse für 50 Segmente, also zwei SNPs pro Segment durchgeführt. Die höchsten Effekte sind an den QTL-Positionen zu finden. Einzige Ausnahme bildet das Segment 45. Hier sind zwei QTLs in einem Segment enthalten. Der Effekt des Segments fällt sehr niedrig aus und liegt im Bereich der Effekte von anderen SNPs. Da die SNP-Allele bei jedem Tier anders ausfallen, kann nicht eindeutig geklärt werden, warum der Effekt so niedrig ausfällt. In Abbildung 19 ist dagegen die Analyse für 100 Segmente durchgeführt worden. Auch hier finden sich die höchsten Effekte an den QTL-Positionen. Die beiden QTLs an den Positionen 89 und 90, die bei der Berechnung zuvor noch in einem Segment waren, sind nun in zwei verschiedenen Segmenten und bewirken beide hohe Effekte. Es ist davon auszugehen, dass die Berechnung des Haplotypeneffekts an Position 45 bei 50

Segmenten ungenauer verlief, da beide Effekte der QTLs in einem Segment berücksichtigt wurden.

Zu gleichen Ergebnissen bei der Darstellung der Effekte kamen Kolbehdari et al. (2007). Die QTL Positionen konnten anhand der Haplotypeneffekte sehr gut ausgemacht werden.

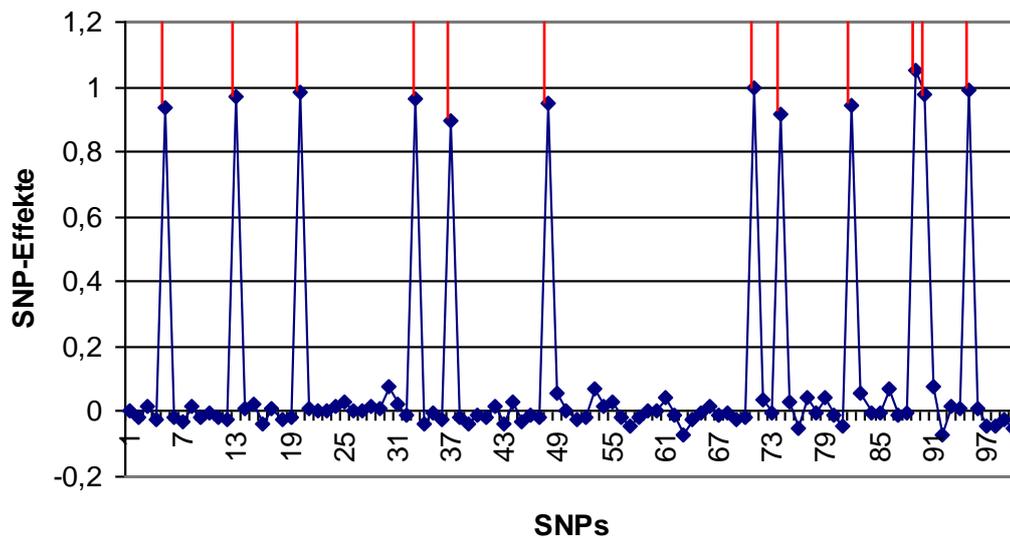


Abbildung 19: SNP-Effekte für Einzel-SNP-Analyse (QTL Positionen rot)

Eine Änderung der Haplotypengröße wurde nur bei Solberg et al. (2008) und Calus et al. (2008) vorgenommen. Beide Arbeiten kamen zu unterschiedlichen Ergebnissen. Calus et al. (2008) kamen zu dem Ergebnis, dass ein Haplotyp, bestehend aus 10 SNPs, die höchsten Genauigkeiten für eine hohe Heritabilität liefert. Bei niedrigen Heritabilitäten fielen die Unterschiede zwischen den Modellen sehr gering aus. Die Einzel-SNP-Analyse konnte hier bei hohen Markerdichten aber eine hohe Genauigkeit liefern. Allerdings wurden IBD-Wahrscheinlichkeiten und Kopplungsphasen zwischen den Haplotypen als Informationen mit in die Analyse einbezogen. Solberg et al. (2008) verglichen die Einzel-Marker-Analyse mit einer Haplotypenanalyse, bestehend aus zwei benachbarten SNPs sowie Mikrosatelliten. Hier konnte die Einzel-Marker-Analyse tendenziell höhere Genauigkeiten liefern, wodurch Solberg et al. (2008) zu einer Einzel-Marker-Analyse raten, da auf diese Weise die Markerphase nicht geschätzt werden muss.

In dieser Arbeit konnte für die Einzel-SNP-Analyse die höchste Genauigkeit erzielt werden, was sich mit den Ergebnissen von Solberg et al. (2008) deckt. Der Wert fiel mit $r = 0,87$ leicht höher aus als die Werte bei Solberg et al. (2008), die von $r = 0,69$ bis $0,86$ je

nach Markerdichte schwankten. Da in dieser Arbeit keine Informationen zur IBD-Wahrscheinlichkeit und Kopplungsphase zwischen den Haplotypen einbezogen wurden, ist ein Vergleich der Ergebnisse mit denen von Calus et al. (2008) für die entsprechenden Segmentgrößen sehr schwer. Die Werte für die Einzel-SNP-Analyse und die für die benachbarten SNPs fielen in dieser Arbeit etwas niedriger aus, wobei nur ein leichter Unterschied von ungefähr 0,01- 0,02 bei der Einzel-SNP-Analyse zu dem Wert von Calus et al. (2008) auftrat.

Hayes et al. (2007) fanden heraus, dass die Anteile an erklärter QTL-Varianz durch die Verwendung von Markerhaplotypen größer ausfielen als bei Gebrauch von Einzel-Markern. So stieg der Anteil an der erklärten QTL-Varianz von 0,1 für den nächstgelegenen Marker auf 0,55 für einen Sechs-Marker-Haplotyp an. Auch die Genauigkeit, mit der die QTL-Alleleffekte geschätzt werden konnten, fiel für den Sechs-Markerhaplotyp am höchsten aus und für den am nächsten gelegenen Einzelmarker am niedrigsten. Gute Ergebnisse konnte der beste Marker erzielen. Er lag mit der Genauigkeit im Bereich des Vier-Marker-Haplotypen, was Hayes et al. (2007) zu dem Schluss kommen ließ, dass die QTLs nicht so genau lokalisiert werden können wie angenommen, da der am nächsten zum QTL gelegene Marker sehr viel geringere Genauigkeiten erzielte. Die Ergebnisse von Hayes et al. (2007) stimmen somit nicht mit den Ergebnissen dieser Arbeit überein. Allerdings muss man dabei beachten, dass bei Hayes et al. (2007) reale Daten für die Analyse verwendet wurden, wobei die Markerverteilung unregelmäßig ausfiel und es große Lücken ohne und Bereiche mit vielen Markern gab.

Mit steigender Tieranzahl - und somit steigender Anzahl an Phänotypen und Markergenotypen - wird die Genauigkeit, mit der Markergenotypeffekte geschätzt werden, sich Eins annähern, wie auch die Genauigkeit, mit der Haplotypeneffekte geschätzt werden. Dies wird langsamer geschehen, da die Anzahl zu schätzender Haplotypen größer ist als die der SNPs, die nur bei zwei je Genort liegen. Daher wird angenommen, dass der Vorteil von Haplotypen über SNP-Genotypen zunehmen wird, wenn die Menge an Daten für die Effektschätzung zunimmt, besonders bei geringen Markerdichten, da nur hier ein Vorteil der Haplotypen über die Einzel-SNP-Analyse aufgezeigt werden konnte (Goddard & Hayes, 2007, Zhao et al., 2007). Zhao et al. (2007) merkten weiter an, dass eine Einzel-SNP-Analyse höhere Genauigkeiten als eine Analyse von Haplotypen liefern wird, sofern eine begrenzte Anzahl an Phänotypen für die Effektschätzung zur Verfügung steht und die Kopplung zwischen SNPs und QTL sehr hoch ist.

Die Einzel-SNP-Analyse stellt somit unter den gegebenen Umständen dieser Arbeit die effektivste Methode für eine Schätzung der SNP-Effekte dar. Die zukünftigen Berechnungen werden sich somit nur noch auf viele Segmente mit wenigen SNPs/Segment oder die Einzel-SNP-Analyse konzentrieren.

4.2.5 Einfluss der Mutationsrate

Abbildung 20 zeigt den Einfluss der Mutationsrate für 100 Segmente auf den MSE-Wert. Neben dem Wert von $2,5 \times 10^{-3}$ (Mutationsrate 1) wurde ein Wert $2,5 \times 10^{-2}$ (Mutationsrate 2) und $2,5 \times 10^{-1}$ (Mutationsrate 3) eingesetzt. Man kann erkennen, dass der MSE-Wert über die verschiedenen Mutationsraten abnimmt. Mit 8,9 bei Mutationsrate 1 hat er den höchsten Wert. Hier liegt auch der höchste Wert für den Standardfehler vor (2,44). Mutationsrate 3 erreicht den niedrigsten MSE-Wert (8,05) und den geringsten Standardfehler (1,68). Die Veränderungen zwischen den verschiedenen Mutationsraten fallen mit 0,43 zwischen Mutationsrate 1 und 2 und mit 0,42 zwischen der 2. und 3. Mutationsrate nahezu gleich groß aus. Die Unterschiede zwischen den Mutationsraten fallen bis auf den Unterschied zwischen Mutationsrate 1 und 3 nicht signifikant aus (Tabelle 13).

Die Korrelation zwischen dem wahren und dem genomischen Zuchtwert bleibt bei allen drei Mutationsraten konstant bei 0,87. Damit hat die Mutationsrate zwar einen leichten Einfluss auf den MSE-Wert, trägt aber zu keiner Veränderung der Genauigkeit des genomischen Zuchtwertes bei.

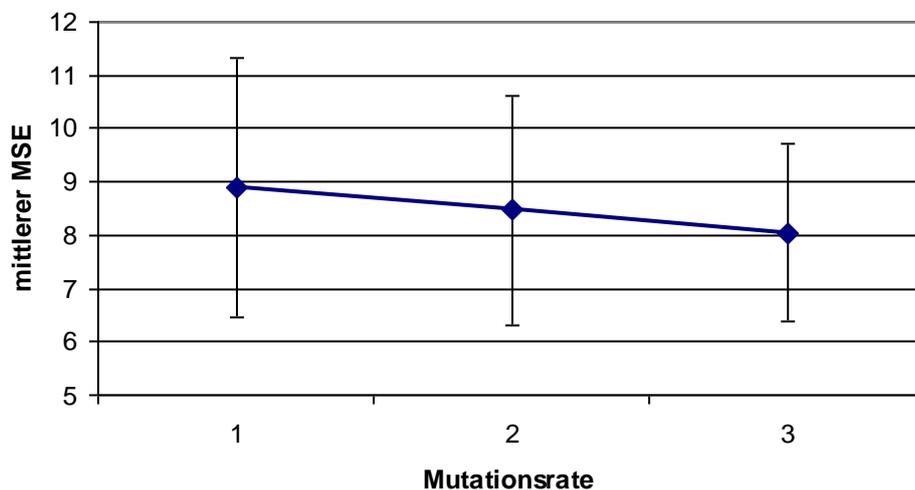


Abbildung 20: Einfluss der Mutationsrate auf den MSE-Wert mit SE

Tabelle 13: Signifikanzen (t-Test) zwischen den Mutationsraten für den MSE-Wert

Mutationsrate 1 mit 2	Mutationsrate 1 mit 3	Mutationsrate 2 mit 3	
ns	**	ns	
nicht signifikant (ns): $p > 0,05$;	*: $p \leq 0,05$	**: $p \leq 0,01$	***: $p \leq 0,001$

Eine Veränderung der Mutationsraten wurde in keiner anderen bekannten Arbeit getestet. Die Mutationsraten der SNPs lagen bei $2,5 \times 10^{-3}$ (Meuwissen et al., 2001; Muir, 2007; Solberg et al., 2008). Damit entsprachen sie dem hier eingestellten Wert von Mutationsrate 1, welcher gleichzeitig auch als Grundeinstellung angenommen wurde. Kumar und Subramanian (2002) wiesen allerdings darauf hin, dass die Mutationsrate für SNPs in ihrer Studie bei einem viel geringeren Wert von $2,2 \times 10^{-9}$ lag, wodurch die angenommene Rate in dieser und anderen Arbeiten zu hoch ausfallen würde. Auch bei Aitken et al. (2004) wurde ein Wert von ungefähr 10^{-8} angegeben. Da die Änderungen der Mutationsrate zu einem höheren Wert allerdings keinerlei Auswirkung auf die Korrelation zwischen dem wahren und genomischen Zuchtwert hatte, sollte eine Veränderung der Mutationsrate auf andere Werte zu keinen Auswirkungen in der Korrelation führen. Weiterhin konnte die Beobachtung gemacht werden, dass die Mutationsraten abhängig von der Genomregion sind (Ellegren et al., 2003; Lercher & Hurst, 2002). Eine Anpassung der Mutationsraten aufgrund dieser Beobachtungen wäre sicherlich für die Zukunft wünschenswert, um die Simulationen realistischer zu gestalten.

4.2.6 Einfluss der SNP-Anzahl

Der Einfluss einer erhöhten SNP-Anzahl auf den MSE-Wert mit SE und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert wird in den beiden nächsten Abbildungen gezeigt. Verglichen werden hierbei die Werte für 100 und 200 SNPs für zwei verschiedene Segmentgrößen. Zum einen wurde die Segmentgröße von zwei SNPs pro Segment gewählt, wodurch bei 100 SNPs insgesamt 50 Segmente gebildet werden, bei 200 SNPs 100 Segmente. Zum Vergleich wurde eine Einzel-SNP-Analyse durchgeführt, wobei die Anzahl an SNPs der Anzahl an Segmenten entspricht. Zwei verschiedene Tieranzahlen in der Analysegeneration wurden gewählt: Zum einen 300 Tiere aus 30 Elternpaaren und zum anderen 1000 Tiere aus 100 Elternpaaren.

In Abbildung 21 ist der Einfluss auf den MSE-Wert dargestellt. Für die Segmentgröße von zwei SNPs pro Segment ist zu sehen, dass der MSE-Wert für 200 SNPs/300 Tiere mit

22,34 extrem hoch ausfällt und deutlich über dem von 100 SNPs/300 Tiere (10,15) liegt. Der MSE-Wert für 200 SNPs/1000 Tiere fällt mit 8,81 sehr viel niedriger aus als der für 200 SNPs/300 Tiere. Am niedrigsten fällt der Wert von 100 SNPs/1000 Tieren aus (3,97). Der Unterschied zwischen 100 SNPs/300 Tieren und 100 SNPs/1000 Tieren fällt mit 6,18 weniger als halb so groß aus wie der Unterschied zwischen 200 SNPs/300 und 200 SNPs/1000 Tieren (13,53). Der SE-Wert für 200 SNPs/300 Tiere fällt mit 13,35 sehr hoch aus und liegt deutlich über dem von 100 SNPs/300 Tiere (2,95). Der Wert für 200 SNPs/1000 Tiere fällt mit 3,04 nur etwas größer aus als der Wert für 100 SNPs/300 Tiere. Mit 1,00 fällt der SE-Wert bei 100 SNPs/1000 Tieren am niedrigsten aus.

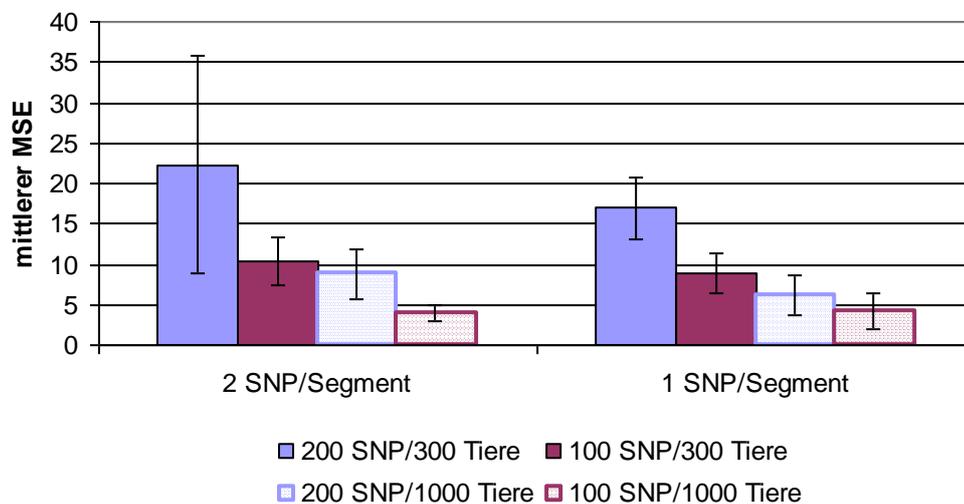


Abbildung 21: Einfluss der SNP-Anzahl auf den MSE-Wert bei 100 und 200 SNPs für zwei verschiedene Segmentgrößen und Tierzahlen

Bei der Einzel-SNP-Analyse fällt der Wert von 200 SNPs/300 Tiere mit 16,94 auch ziemlich hoch aus und liegt auch hier wieder über dem von 100 SNPs/300 Tiere (8,9). Hier ist noch ein Unterschied von 8,04 zu verzeichnen. Der Wert für 200 SNPs/1000 Tiere fällt mit 6,19 niedriger aus als bei 200 SNPs/300 Tiere. Der Unterschied zwischen diesen liegt hier bei 10,75. Der MSE-Wert für 100 SNPs/1000 Tiere fällt mit 4,23 am niedrigsten aus. Der Unterschied zwischen 100 SNPs/300 Tieren und 100 SNPs/1000 Tieren fällt mit 4,67 sehr viel niedriger aus als der bei 200 SNPs. Der SE-Wert fällt mit 3,75 für 200 SNPs/300 Tiere sehr viel geringer aus als bei der größeren Segmentgröße von 2 SNPs/Segment. Der Wert für 100 SNPs/300 Tiere ändert sich dagegen mit 2,44 sehr geringfügig gegenüber dem größeren Segment. Bei 200 SNPs/1000 Tiere liegt der Wert bei 2,41 und bei 2,28 für 100 SNPs/1000 Tiere.

Die MSE-Werte für die Einzel-SNP-Analyse liegen somit leicht niedriger als die Werte für die zwei SNP-Segmente. Einzige Ausnahme bilden hier die 100 SNPs bei 1000 Tieren. Hier fällt der MSE-Wert bei zwei SNPs je Segment (3,97) etwas geringer aus als der für die Einzel-SNP-Analyse (4,23).

Sämtliche Unterschiede zwischen den MSE-Werten sind in Tabelle 14 dargestellt. Es ist zu sehen, dass alle Unterschiede hochsignifikant ausfallen, bis auf den Unterschied 100 SNPs/1000 Tiere zwischen den beiden Segmentgrößen und den Unterschied von 100 SNP/300 Tieren bei einem SNP/Segment mit 200 SNPs/1000 Tieren bei zwei SNPs/Segment. In beiden Fällen fallen die Werte nahezu gleich aus.

In der folgenden Abbildung 22 ist der Einfluss auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert aufgezeigt. Der Unterschied zwischen den Werten für 100 und 200 SNPs ist für beide Segmentgrößen und Tieranzahlen deutlich zu erkennen. Bei zwei SNPs pro Segment liegt der Wert für die Korrelation mit 0,75 für 200 SNPs/300 Tiere deutlich unter dem von 100 SNPs/300 Tiere ($r = 0,84$). Das entspricht einem Unterschied von 0,09. Die Korrelation für 200 SNPs/1000 Tiere erreicht mit 0,88 einen um 0,13 höheren Wert als der für 300 Tiere. Den höchsten Wert ($r = 0,93$) erreichen 100 SNPs/1000 Tiere. Hier liegt der Wert um 0,09 höher als der für 300 Tiere.

Für die Einzel-SNP-Analyse erhält man das gleiche unterschiedliche Niveau zwischen den beiden Korrelationswerten, nur liegt hier der Wert für 200 SNPs/300 Tiere bei 0,78 und bei 0,87 für 100 SNPs/300 Tiere. Die Werte für 1000 Tiere liegen auch hier wieder höher ($r = 0,91$ bei 200 SNPs/1000 Tiere und $r = 0,95$ 100 SNPs/1000 Tiere).

Somit liegen die Korrelationen für die Einzel-SNP-Analyse leicht höher als für die zwei SNP-Segmente. Insgesamt fallen alle Unterschiede zwischen den Werten hochsignifikant aus (Tabelle 14).

Ein Einfluss der SNP-Anzahl auf den MSE-Wert und die Korrelation ist somit definitiv gegeben. Eine erhöhte SNP-Anzahl wirkt sich in diesem Fall allerdings nicht positiv auf den MSE-Wert und die Genauigkeit der genomischen Zuchtwerte aus. Die MSE-Werte bei 300 Tieren in der Analysegeneration für 200 SNPs fallen sehr viel höher aus als die für 100 SNPs, und auch der sehr hohe SE-Wert für zwei SNPs pro Segment bei 200 SNPs spricht für eine hohe Ungenauigkeit bei der Bewertung der Haplotypeneffekte. Die Werte bei 1000 Tieren in der Analysegeneration fallen generell viel niedriger aus. Aber auch hier liegen die MSE-Werte für 200 SNPs über denen für 100 SNPs. Allerdings fallen die Unterschiede zwischen 200 und 100 SNPs nicht mehr so hoch aus: 4,84 bei zwei SNPs/Segment (dagegen 11,92 für 300 Tiere) und nur 1,96 bei einem SNP/Segment (dagegen 8,04 für 300 Tiere).

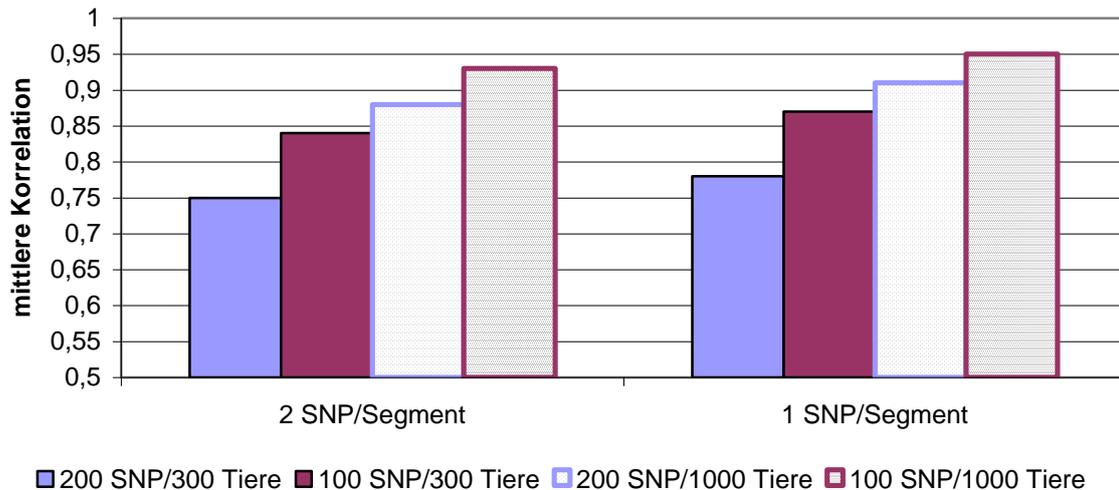


Abbildung 22: Einfluss der SNP-Anzahl auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert bei 100 und 200 SNPs für zwei verschiedene Segmentgrößen und Tierzahlen

Auch die Werte für die Korrelation zwischen dem wahren und genomischen Zuchtwert fallen bei 300 Tieren für 200 SNPs um 0,09 für zwei SNPs/Segment und 0,10 für einem SNP/Segment geringer aus als die Werte für 100 SNPs. Auch hier konnte eine Erhöhung der Anzahl Tiere in der Analysegeneration keine Veränderung herbeiführen. Die Korrelationswerte für 1000 Tiere liegen auch hier für 200 SNPs unter denen für 100 SNPs, auch wenn die Unterschiede nicht mehr ganz so groß ausfallen (0,05 für zwei SNPs/Segment und 0,04 für ein SNP/Segment). Eine genauere Schätzung der Gewichtungsfaktoren konnte somit mit einer erhöhten SNP-Anzahl nicht erreicht werden. Man muss allerdings bedenken, dass die gewählte Stichprobe von 300 Tieren relativ klein ausfällt. In dem Fall von 100 SNPs stellt das kein Problem dar, da hier bei beiden Segmentgrößen eine maximale Haplotypenanzahl von 200 erreicht werden kann. Somit liegen mehr Tiere mit Leistungsinformationen vor als Haplotypen vorkommen. Bei 200 SNPs kommen aber jeweils maximal 400 Haplotypen vor. Hier stehen also weniger Tiere als Haplotypen zur Verfügung, wodurch die Berechnungen ungenauer ausfallen, was sich auf den MSE-Wert und die Korrelation auswirkt. Um einen passenderen Vergleich zu erhalten, wurde die Stichprobe auf 1000 Tiere erhöht. Allgemein konnten damit wie aufgezeigt, höhere Korrelationen und niedrigere MSE-Werte erzielt werden. Allerdings fielen die Werte für 100 SNPs noch immer vorteilhafter aus als die für 200 SNPs. Durch die Erhöhung der Anzahl der SNPs von 100 auf 200 bei gleich bleibender Genomgröße von einem Morgan je Chromosom, wird die Markerdichte erhöht und der

Abstand der einzelnen Marker auf 0,5 cM verkleinert. Dabei bleibt die Anzahl an QTLs gleich.

Tabelle 14: Signifikanzen (t-Test) zwischen den einzelnen SNP-Anzahlen für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen

		Anzahl		2 SNPs/Segment				1 SNP/Segment			
		Tiere	SNPs	300		1000		300		1000	
				100	200	100	200	100	200	100	200
Anzahl		Tiere	SNPs								
2 SNPs/ Segment	300	100		***	***	***	***	***	***	***	***
		200		***		***	***	***	***	***	***
	1000	100		***	***		***	***	***	ns	***
		200		***	***	***		ns	***	***	***
1 SNP/ Segment	300	100		***	***	***	***		***	***	***
		200		***	***	***	***	***		***	***
	1000	100		***	***	***	***	***	***		***
		200		***	***	***	***	***	***	***	

nicht signifikant (ns): $p > 0,05$; *: $p \leq 0,05$ **: $p \leq 0,01$ ***: $p \leq 0,001$

Meuwissen et al. (2001) verwendeten unterschiedliche Markerdichten. Der Abstand zwischen zwei Markern war dabei 1, 2 oder 4 cM groß. Mit steigendem Abstand zwischen den Markern fiel die Genauigkeit dabei von $r = 0,732$ auf $r = 0,668$ ab.

Bei Muir (2007) wurde eine unterschiedliche Markerdichte getestet, indem zum einen die Anzahl an Markern konstant bei 100 gehalten wurde und die Größe des Genoms von 1 cM bis hin zu 10000 cM variierte. Hier fiel die Genauigkeit mit steigender Genomgröße und somit sinkender Markerdichte aufgrund von höheren Rekombinationsraten deutlich ab. Zum anderen wurde die Markeranzahl von 10 bis 3000 Marker variiert bei gleich bleibender Markerdichte von 10/cM. Hier konnten die Ergebnisse ein Optimum der Genauigkeit bei 1000 Markern für eine Population im Hardy-Weinberg-Gleichgewicht finden, was einem Verhältnis von Markern zu QTL von 10:1 entsprach und dem Verhältnis in dieser Arbeit sehr nahe kam. Bei einer Population im Mutations-Drift-Gleichgewicht, lag dieses Optimum bei einem Marker:QTL-Verhältnis von 2,5:1. Bei einer höheren Anzahl

QTL als Marker können die Marker nicht die gesamte genetische Varianz erklären. Andererseits ist eine zu hohe Anzahl an Markern, welche die Anzahl an QTL weit übertrifft, nicht von Vorteil, da Confounding-Effekte zwischen den Markern so nicht getrennt werden können. Ein angemessenes Marker:QTL-Verhältnis ist daher laut Muir (2007) notwendig.

Bei Calus und Veerkamp (2007) wurden insgesamt drei Markerdichten getestet. Die Ergebnisse zeigten, dass mit steigender Markerdichte die Genauigkeiten der genomischen Zuchtwerte zunahmen.

In der Arbeit von Calus et al. (2008) wurden unterschiedliche Markerdichten getestet, indem die Anzahl an Markern bei gleichbleibender Genomgröße variiert wurde. Dabei wurden im Durchschnitt zwischen 120 und 2350 Markern über das Genom (3 Morgan). Die Ergebnisse zeigten, dass die Genauigkeit der genomischen Zuchtwerte bei einer hohen Heritabilität mit steigender Markerdichte zunahm, allerdings fielen die Zunahmen bei höheren Markerdichten gering aus. Für eine geringe Heritabilität fiel der Anstieg der Genauigkeit insgesamt höher aus, aber auch hier fiel der Anstieg mit höheren Markerdichten geringer aus. Somit fiel bei Calus et al. (2008) das Marker:QTL-Verhältnis mit 30:1 viel höher aus. Allerdings ist ein Vergleich mit den Ergebnissen von Muir (2007) schwierig, da bei ihm die Markerdichte konstant gehalten wurde, während bei Calus et al. (2008) die Dichte verändert wurde. Der Wert für die Genauigkeit bei Calus et al. (2008) ($r = 0,83-0,84$) lag für die gleiche Markerdichte (ca. 1 cM zwischen den Markern) und die gleiche Heritabilität ($h^2 = 0,5$) in der nahezu gleichen Größenordnung wie die Genauigkeit in dieser Arbeit ($r = 0,87$).

Auch in der Arbeit von Solberg et al. (2008) wurde die Markerdichte durch Anheben der Markeranzahl von einem bis acht pro cM verändert bei gleichbleibender Anzahl von 100 möglichen QTL. Die Genauigkeiten der genomischen Zuchtwerte stiegen hier mit steigender Markeranzahl von $r = 0,69$ auf $0,86$. Der Anstieg von vier auf acht Marker/cM fiel mit $0,019$ allerdings nur noch gering aus. Insgesamt gesehen, fielen die Genauigkeiten geringer aus als in dieser Arbeit. Bei einer Anzahl von einem Marker/cM lag die Genauigkeit bei $r = 0,69$ und fiel somit $0,19$ niedriger aus als in dieser Arbeit. Erst bei acht Markern/cM erreichte die Genauigkeit den gleichen Wert wie in dieser Arbeit.

Eine Erklärung für die abweichenden Ergebnisse in dieser Arbeit stellen die unterschiedlichen Bedingungen dar. So wurden nur 300 Tiere in der Analysegeneration verwendet, während in anderen Arbeiten höhere Anzahlen von mindestens 1000 Tieren eingesetzt wurden. Weiterhin wurden insgesamt 15 Generationen erzeugt, wohingegen bei anderen Arbeiten in der Regel 1000 Generationen simuliert wurden, um eine LD zwischen Markern und QTL zu erzielen. Aufgrund der niedrigen Anzahl an Generationen liegt die Vermutung nahe, dass kaum LD vorhanden war und die Genauigkeiten daher

niedriger ausfielen, wofür auch die Ergebnisse in Kapitel 4.2.2 sprechen. Auch Solberg et al. (2008) konnten für eine geringere Anzahl an Generationen ein geringeres LD nachweisen, was in geringeren Genauigkeiten resultierte. Weiterhin wurden in dieser Arbeit die QTL als Marker mit in die Analyse einbezogen, was in anderen Arbeiten nicht der Fall war. Da nach so wenigen Generationen kaum LD vorkommt, ist ein positiver Effekt auf die Genauigkeit durch zusätzliche Marker nicht gegeben. Die Marker, die direkt QTLs waren, hatten maximale LD, alle anderen kein LD, wodurch eine Erhöhung der Anzahl der Marker keine höhere Genauigkeit zur Folge hatte.

4.2.7 Einfluss der Heritabilität

In der folgenden Abbildung 23 ist der Einfluss der Heritabilität auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für 100 Segmente aufgezeigt. Ausgewählt wurden drei verschiedene Heritabilitätswerte von 0,1, 0,3 und 0,5.

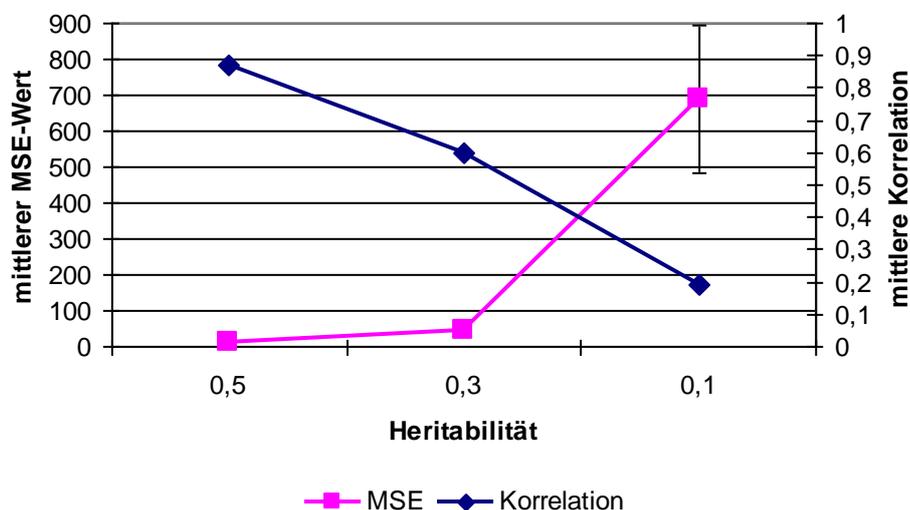


Abbildung 23: Einfluss der Heritabilität auf den MSE-Wert (mit SE) und die Korrelation zwischen dem wahren und dem genomischen Zuchtwert

Man sieht deutlich, dass der MSE-Wert für $h^2 = 0,5$ mit 8,6 am kleinsten ist. Auch der Standardfehler ist sehr gering ($SE = 2,62$). Der Wert für $h^2 = 0,3$ fällt mit 43,66 schon sehr viel höher aus, aber auch hier fällt der Standardfehler (10,76) sehr klein aus. Der Wert für $h^2 = 0,5$ ist extrem groß (686,75), und auch der Standardfehler fällt hier mit 205,94 sehr viel größer aus. Der Wert für die Heritabilität von 0,1 sticht damit hervor und unterscheidet sich deutlich von den anderen beiden Werten. Die Unterschiede zwischen den Werten fallen alle hochsignifikant aus (Tabelle 15).

Die Werte für die Korrelation zwischen dem wahren und dem genomischen Zuchtwert geben ein ähnliches, nur nicht ganz so ausgeprägtes Bild, wieder. Der Korrelationswert für $h^2 = 0,5$ ist mit 0,87 am höchsten und fällt auf 0,6 für eine Heritabilität von 0,3 ab. Der Wert für $h^2 = 0,1$ liegt mit 0,19 am niedrigsten und weist einen sehr großen Unterschied zu den anderen beiden Werten auf. Auch hier fallen alle Unterschiede hochsignifikant aus (Tabelle 15).

Tabelle 15: Signifikanzen (t-Test) zwischen den Heritabilitäten für MSE-Werte oberhalb der Diagonalen, für Korrelationen unterhalb der Diagonalen

		Heritabilität		
		0,5	0,3	0,1
Heritabilität	0,5		***	***
	0,3	***		***
	0,1	***	***	

nicht signifikant (ns): $p > 0,05$; *: $p \leq 0,05$ **: $p \leq 0,01$ ***: $p \leq 0,001$

In Abbildung 24 sind weitere Untersuchungen mit einer niedrigen Heritabilität von 0,1 aufgezeigt. Dargestellt ist die Korrelation zwischen dem wahren und genomischen Zuchtwert. Verändert wurde hier die Größe der Analysegeneration in vier Schritten von 300 auf 3000 Tiere. Es ist deutlich zu sehen, dass die Korrelation zwischen dem wahren und dem genomischen Zuchtwert mit steigender Anzahl von Tieren in der Analysegeneration kontinuierlich zunimmt. Der Wert steigt von $r = 0,19$ bei 300 Tieren auf $r = 0,51$ bei 3000 Tieren. Damit konnte gezeigt werden, dass auch bei geringer Heritabilität eine hohe Genauigkeit erreicht werden kann, indem die Anzahl der Tiere, die zur Analyse verwendet wurden, gesteigert wurde.

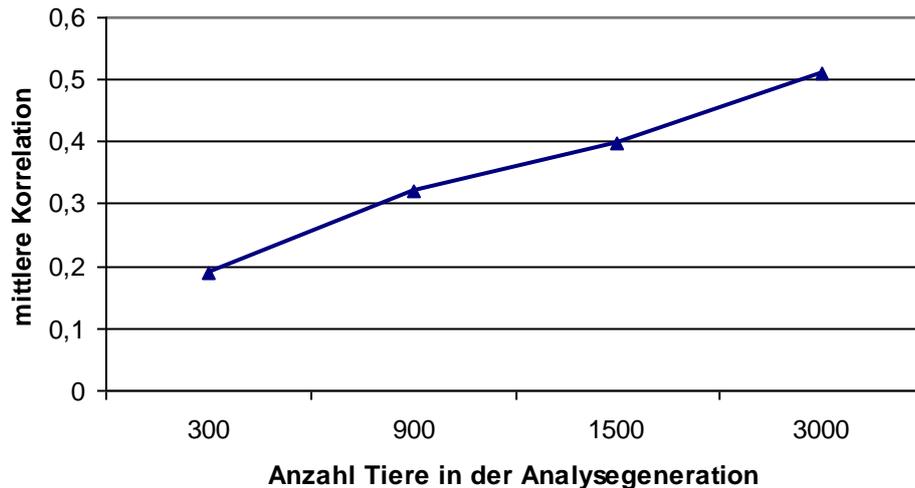


Abbildung 24: Korrelation zwischen dem wahren und dem genomischen Zuchtwert für verschiedene Größen der Analysegeneration für $h^2 = 0,1$

In verschiedenen Arbeiten von Calus und Veerkamp (2007), Kolbehdari et al. (2007), Muir (2007) und Calus et al. (2008) wurden verschiedene Heritabilitäten getestet. Calus und Veerkamp (2007) verwendeten Heritabilitäten von 0,5 und 0,1. Für $h^2 = 0,5$ errechneten sie Werte zwischen $r = 0,86$ und $0,91$ für 1100 Tiere in der Analysegeneration. Bei nur 500 Tieren lag der Wert bei $r = 0,82$. Bei einer Heritabilität von 0,1 fiel die Genauigkeit für 1100 Tiere auf $r = 0,6$ bis $0,65$ und bei 500 Tieren auf $r = 0,47$ ab. Verglichen mit den Werten in dieser Arbeit fallen die Genauigkeiten für $h^2 = 0,5$ mit $r = 0,87$ bei 300 Tieren in der Analysegeneration höher aus. Die Genauigkeiten für $h^2 = 0,1$ liegen mit $r = 0,19$ für 300 Tiere und $r = 0,32$ für 900 Tiere deutlich niedriger, was durch die Anwendung unterschiedlicher Methoden zur Effektschätzung zu erklären ist.

Kolbehdari et al. (2007) verwendeten drei verschiedene Heritabilitäten von 0,5, 0,3 und 0,05. Die Genauigkeiten fielen von $r = 0,77$ bei $h^2=0,5$ über $r = 0,76$ ($h^2 = 0,3$) auf $r = 0,74$ bei $h^2 = 0,05$. Somit war auch hier ein Abfall der Genauigkeiten zu verzeichnen. Allerdings lagen die Werte außer für $h^2 = 0,5$ deutlich höher als in dieser Arbeit, was darauf zurückzuführen sein kann, dass als Phänotyp nicht die Leistungen der Tiere sondern der DYD eingesetzt wurden. Weiterhin wurden die Effekte anhand einer sehr großen Nachkommengruppe von 10000 Töchtern aus 100 Bullen getestet.

Bei Muir (2007) und Calus et al. (2008) wurden Heritabilitäten von 0,5 und 0,1 verwendet. Die Werte fielen hier von $r = 0,7 - 0,8$ für $h^2 = 0,5$ auf $r = 0,44 - 0,5$ für $h^2 = 0,1$ bei Muir (2007) und von $r = 0,82$ für $h^2 = 0,5$ auf $r = 0,68$ bei $h^2 = 0,1$ bei Calus et al. (2008) ab. Auch hier fielen die Werte für $h^2 = 0,1$ in dieser Arbeit niedriger aus, was an den unterschiedlichen Zusammensetzungen der Analysegenerationen liegen kann. Bei Calus

et al. (2008) wurden insgesamt mehr Tiere und eine Bayes-Methode zur Schätzung der Effekte verwendet. Bei Muir (2007) wurden Tiere aus verschiedenen Generationen herangezogen. Bei Verwendung von nur einer Generation und weniger Tieren lagen die Genauigkeiten bei $r = 0,2 - 0,3$, was der Genauigkeit in dieser Arbeit entspricht.

Die Heritabilität hat somit einen deutlichen Einfluss auf den genomischen Zuchtwert. Die genauesten Schätzungen für die Gewichtungsfaktoren der Haplotypen und somit für den genomischen Zuchtwert bekommt man bei einer hohen Heritabilität. Hier ist die Korrelation am höchsten und der MSE-Wert am niedrigsten. Bei einer hohen Heritabilität, hier 0,5, bestimmt die Umwelt zur Hälfte die Leistung des Tieres. Der andere Teil der Leistung ist auf die Genetik zurückzuführen. Bei einer niedrigen Heritabilität ist dagegen der Umwelteffekt sehr viel größer. Der Großteil der Leistung wird durch die Umwelt beeinflusst, was die Berechnung der Gewichtungsfaktoren erschwert und ungenau werden lässt. Aus diesem Grund fallen die MSE-Werte bei $h^2 = 0,3$ und $0,1$ höher aus und die Korrelationen niedriger. Mit einer Steigerung der Tierzahlen in der Analysegeneration konnten aber höhere Genauigkeiten erzielt werden.

Bei geringen Heritabilitäten sollte daher die Analysegeneration eine hohe Anzahl an Tieren aufweisen, da gezeigt werden konnte, dass die Genauigkeit mit steigender Tierzahl zunimmt. Andere Autoren konnten höhere Genauigkeiten bei niedrigen Heritabilitäten mit höheren Tierzahlen, verglichen mit dieser Arbeit, erzielen. Weiterhin könnte eine Anwendung von DYDs bzw. YDs (yield deviation) anstelle der Phänotypen, wie bei Kolbehdari et al. (2007) angesprochen, einen positiven Effekt auf die Genauigkeit haben, da Kolbehdari et al. (2007) auch bei geringen Heritabilitäten sehr hohe Genauigkeiten angegeben haben.

4.2.8 Genauigkeit der genomischen Zuchtwerte in den Folgegenerationen

4.2.8.1 Einfluss der Größe der Analysegeneration

In den folgenden Abbildungen soll gezeigt werden, wie sich die Korrelation zwischen dem wahren und dem genomischen Zuchtwert in der 1. Folgegeneration oder in der 1. bis 4. Folgegeneration im Vergleich zur Analysegeneration verändert. Hierzu wurde jeweils für verschiedene Parameterveränderungen die Korrelation zwischen dem wahren und genomischen Zuchtwert der Analysegeneration mit der der Folgegeneration/en verglichen.

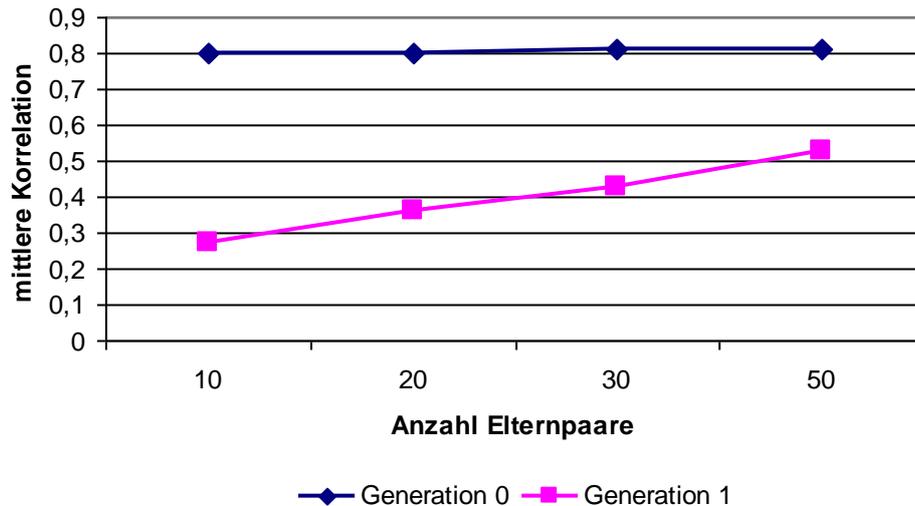


Abbildung 25: Einfluss der Anzahl der Elternpaare auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und die Folgegeneration

Abbildung 25 zeigt die Entwicklung der Korrelationen bei verschiedenen Anzahlen von Elternpaaren (10, 20, 30 und 50) für 10 Segmente. Es fällt auf, dass die Korrelationswerte für die Folgegeneration deutlich unter denen der Analysegeneration liegen. Während die Werte der Analysegeneration von $r = 0,80$ auf $0,81$ nur geringfügig ansteigen, steigt die Korrelation der Folgegeneration von $r = 0,27$ bei 10 Elternpaaren kontinuierlich auf $0,53$ bei 50 Elternpaaren an. Somit ist bei 10 Elternpaaren mit $0,53$ der größte Unterschied zwischen den beiden Generationen zu verzeichnen. Dieser extreme Unterschied verringert sich mit steigender Anzahl von Elternpaaren und ist bei 50 Elternpaaren mit $0,28$ am geringsten.

Ein Einfluss der Anzahl der Elternpaare auf die Genauigkeit der genomischen Zuchtwerte für die Folgegeneration ist damit gegeben. Er fällt hier sogar wesentlich größer aus als bei der Analysegeneration, was damit zu erklären ist, dass die Gewichtungsfaktoren für die Haplotypen nur in der Analysegeneration geschätzt werden. Somit liegen keine Gewichtungen für neu entstandene Haplotypen durch Mutation oder Rekombination vor. Außerdem werden die Faktoren nur anhand einer kleinen Stichprobe von 300 Tieren errechnet. Bei 10 Segmenten können allerdings 1024 mögliche Haplotypen pro Segment vorkommen. Somit ist eine Erfassung aller Haplotypen durch die 300 Tiere nicht gegeben. In der Folgegeneration treten dadurch eine gewisse Menge an neu hinzugekommenen Haplotypen auf, für die keine Gewichtungsfaktoren geschätzt wurden.

Man kann deutlich erkennen, dass in der Analysegeneration die Erhöhung der Anzahl der Elternpaare nur einen verhältnismäßig geringen Zuwachs der Korrelation einbringt,

während der Anstieg in der Folgegeneration dagegen sehr hoch ist. Je mehr Tiere für die Stichprobe ausgewählt werden, desto mehr unterschiedliche Haplotypen werden erfasst und desto genauer kann der genomische Zuchtwert in der Folgegeneration bestimmt werden. Weiterhin stehen durch eine höhere Anzahl von Tieren mehr Phänotypen für die Analyse der Effekte zur Verfügung, was die Schätzungen genauer ausfallen lässt.

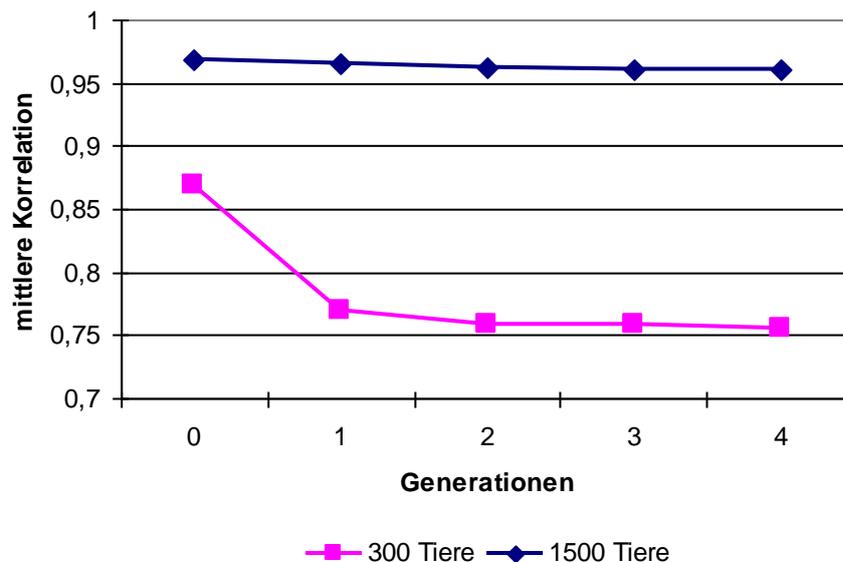


Abbildung 26: Einfluss der Größe der Analysegeneration auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und vier Folgegenerationen

Sollen die Gewichtungsfaktoren also nicht für jede Generation neu geschätzt werden, so ist die Berechnung der Faktoren an einer großen Stichprobe mit einer hohen Anzahl von Elternpaaren und Nachkommen durchaus sinnvoll, da zwar die Genauigkeit in der Analysegeneration kaum zunimmt, die Berechnung des genomischen Zuchtwertes in der Folgegeneration allerdings so viel genauer wird. Zur Verdeutlichung soll in der folgenden Abbildung 26 der Einfluss der Größe der Analysegeneration für 300 und 1500 Tiere dargestellt werden. Abgebildet ist die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration (Generation 0) und vier Folgegenerationen für 300 und 1500 Tiere in der Analysegeneration bei einer Einzel-SNP-Analyse. Man kann erkennen, dass für 300 Tiere die Korrelation von $r = 0,869$ in der Analysegeneration auf $r = 0,769$ in der 1. Folgegeneration abfällt und dann weiter leicht bis auf $r = 0,756$ in der vierten Folgegeneration abfällt. Für 1500 Tiere liegt die Korrelation schon in der Analysegeneration deutlich höher ($r = 0,969$) und fällt dann viel langsamer

bis auf $r = 0,960$ in der vierten Folgegeneration ab. Der Abfall der Genauigkeit fällt somit bei 300 Tieren mit 0,113 sehr viel größer aus als bei 1500 Tieren. Das kann man damit erklären, dass bei 1500 Tieren alle in Generation 16 zur Verfügung stehenden Tiere zur Analyse herangezogen werden. Somit stehen sehr viel mehr Phänotypleistungen zur Verfügung, und die SNP-Effekte können genauer geschätzt werden.

Der direkte Vergleich der Genauigkeiten der Analysegeneration mit Folgegenerationen wurde nur in wenigen Arbeiten von Calus und Veerkamp (2007), Muir (2007), Habier et al. (2007) und Calus et al. (2008) gezogen. In anderen Arbeiten werden nur die Genauigkeiten der folgenden Generationen bzw. für Tiere ohne Phänotyp angegeben.

Bei Calus und Veerkamp (2007) fiel die Genauigkeit, gemessen an 1100 Tieren von $r = 0,91-0,86$ in der Analysegeneration auf $r = 0,79$ für die Folgegeneration ab. Für 500 Tiere fiel sie von $r = 0,82$ auf $r = 0,73$. Die Werte der Genauigkeiten lagen damit ungefähr im Bereich der Werte in dieser Arbeit. Für 300 Tiere fielen sie von $r = 0,869$ auf $r = 0,769$ in der ersten Folgegeneration ab. Der Abfall der Genauigkeit entsprach dem bei Calus und Veerkamp (2007). Für 1500 Tiere lag der Wert in dieser Arbeit mit $r = 0,969$ höher und fiel auf $r = 0,965$. Damit fielen die Genauigkeiten hier höher aus, allerdings wurden auch 400 Tiere mehr analysiert.

Bei Muir (2007) wurden für insgesamt zehn Generationen die Genauigkeiten der genomischen Zuchtwerte berechnet, dabei waren je ein bis vier Generationen Analysegenerationen, wobei je Generation nur 128 Tiere analysiert wurden. Die Genauigkeiten für die Analysegenerationen lagen im Bereich von $r = 0,7 - 0,8$ und fielen in den folgenden Generationen auf $r = 0,4 - 0,6$ ab. Damit lagen die Genauigkeiten von Muir (2007) etwas unterhalb der Genauigkeiten in dieser Arbeit. Allgemein war aber ein Abfall der Genauigkeiten in den folgenden Generationen zu erkennen, der weniger stark ausgeprägt ist, je mehr Tiere zur Analyse verwendet wurden.

Habier et al. (2007) schätzten für die Analysegeneration sowie zehn weitere Generationen die genomischen Zuchtwerte. Die Genauigkeiten fielen dabei von $r=0,78$ in der Analysegeneration auf $r=0,42$ in der 10. Generation ab. Damit lagen die Werte deutlich unter denen in dieser Arbeit, obwohl 1000 Tiere für die Analyse der Effekte zu Verfügung standen.

Calus et al. (2008) schätzten die Genauigkeiten der genomischen Zuchtwerte für die Analysegeneration und eine Folgegeneration. Die Ergebnisse zeigten, dass die Genauigkeit für die Folgegeneration abnahm von $r = 0,83$ auf $r = 0,7$, womit sie unterhalb der Genauigkeit in dieser Arbeit lagen, da insgesamt 1100 Tiere zur Analyse der Effekte verwendet wurden.

4.2.8.2 Einfluss der Segmentgröße

In Abbildung 27 wird der Einfluss der Segmentgröße auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analyse- sowie Folgegeneration gezeigt. Dargestellt werden fünf verschiedene Segmentgrößen: 4, 10, 20, 30 und 100 Segmente.

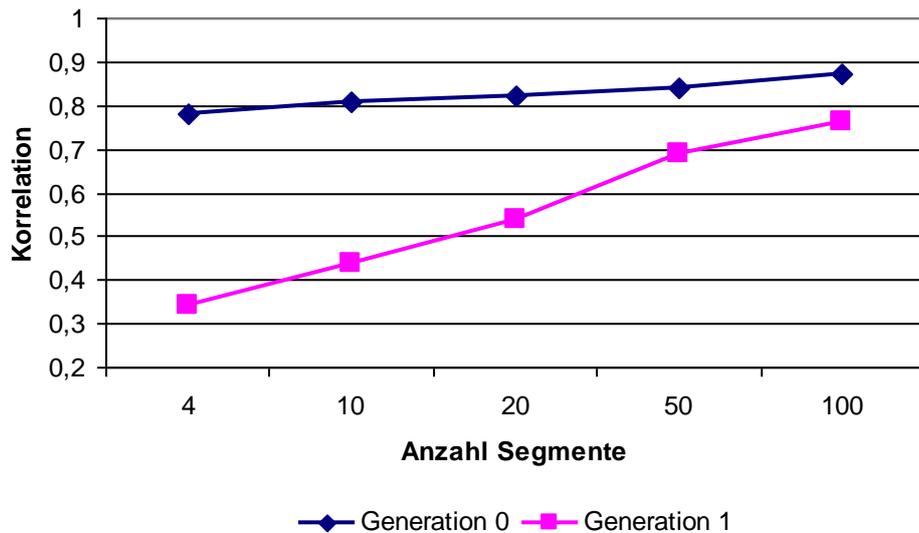


Abbildung 27: Einfluss der Anzahl der Segmente auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und die Folgegeneration

Die Korrelation der Analysegeneration steigt, wie in Kapitel 4.2.4 beschrieben, von 0,78 bei vier Segmenten auf 0,87 bei 100 Segmenten. Die Korrelationswerte der Folgegeneration unterscheiden sich davon deutlich und fallen teilweise sehr viel niedriger aus. Bei vier Segmenten liegt der Wert bei $r = 0,34$, womit er deutlich unter dem der Analysegeneration ($r = 0,78$) liegt. Die Werte nehmen mit steigender Segmentanzahl stark zu und liegen schließlich mit $r = 0,69$ (50 Segmente) und $r = 0,76$ (100 Segmente) in einem hohen Bereich. Hier fällt der Unterschied der Korrelationswerte zwischen der Analyse- und Folgegeneration relativ gering aus.

Der große Unterschied zwischen den Korrelationswerten der beiden Generationen bei vier, zehn und 20 Segmenten lässt sich darauf zurückführen, dass zum einen wieder nur 300 von 1500 Tieren zur Analyse ausgewählt wurden. Zum anderen ist bei so großen Segmenten eine große Anzahl von Haplotypen möglich. Bei vier Segmenten mit je 25 SNPs pro Segment sind je Segment mehr als 33 Millionen verschiedene Kombinationen möglich. Aus den 300 Tieren könnten aber nur maximal 600 verschiedene Haplotypen

erfasst werden. In Kapitel 4.1 (Abbildung 11) wurde schon dargestellt, dass nur ein geringer Teil der möglichen Haplotypen erfasst werden kann. Von den 1024 möglichen Haplotypen wurden im Schnitt nur 96 und somit weniger als 10% registriert und analysiert. Aus diesem Grund fallen die Korrelationen der Folgegeneration für vier, zehn und 20 Segmente so niedrig aus, da für einen Großteil der Haplotypen in der Folgegeneration keine Gewichtungsfaktoren geschätzt wurden. Bei 50 und 100 Segmenten fallen die Unterschiede geringer aus, weil hier die Anzahl an möglichen Haplotypen mit 200 sehr viel geringer ausfällt. Schon in der Analysegeneration werden alle möglichen Haplotypen erfasst und es können Gewichtungsfaktoren berechnet werden. In der Folgegeneration treten somit keine neuen Haplotypen mehr auf. Dass trotz Bekanntheit aller Haplotypen Unterschiede zwischen den Korrelationswerten der beiden Generationen auftreten, liegt an der Berechnung der Gewichtungsfaktoren. Diese fand nur einmalig an der Analysegeneration statt und es werden die gleichen Faktoren auch auf die Folgegeneration angewandt, was zu den Ungenauigkeiten führt.

4.2.8.3 Einfluss der Heritabilität

Der Einfluss der Heritabilität auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analyse- und Folgegeneration bei 30 Elternpaaren mit unterschiedlicher Nachkommenanzahl wird in Abbildung 28 gezeigt. Dargestellt werden drei Heritabilitäten von 0,1, 0,3 und 0,5.

Es ist zu erkennen, dass die Korrelationswerte der Analysegeneration bei allen drei Heritabilitäten und bei allen Nachkommenanzahlen über den Werten der Folgegeneration liegen.

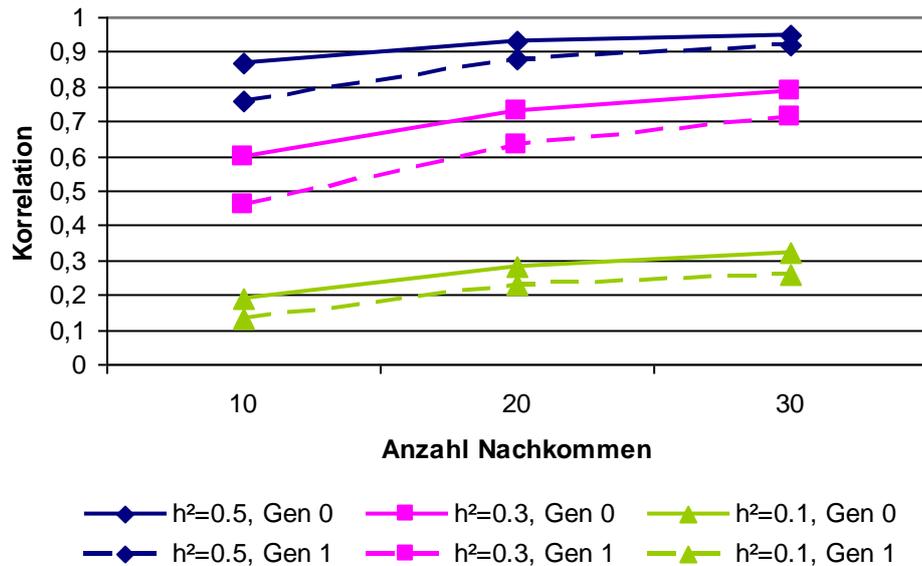


Abbildung 28: Einfluss der Heritabilität und der Anzahl der Nachkommen auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und Folgegeneration

Die Korrelationswerte der Folgegeneration für $h^2 = 0,1$ liegen in einem sehr niedrigen Bereich von 0,13 (10 Nachkommen) bis zu 0,26 (30 Nachkommen). Der Unterschied zu den Werten der Analysegeneration ist dabei über die Nachkommenanzahl gleich bleibend gering bei 0,05 – 0,06. Für die mittlere Heritabilität von 0,3 liegt die Korrelation bei 0,46 für 10 Nachkommen und steigt auf 0,71 bei 30 Nachkommen an. Hier fällt auf, dass der Unterschied zwischen den Generationen bei 10 Nachkommen mit 0,14 größer ausfällt und dann mit steigender Nachkommenanzahl auf 0,08 bei 30 Nachkommen abnimmt. Die Korrelationswerte der Folgegeneration für eine hohe Heritabilität von 0,5 nehmen mit 0,76 (10 Nachkommen) bis hin zu 0,92 (30 Nachkommen) sehr hohe Werte an. Aber auch hier sieht man, dass der Unterschied zwischen den Generationen bei niedrigen Nachkommenzahlen größer ist und geringer bei höheren Nachkommenzahlen ausfällt. Somit ist auch in diesem Beispiel wieder deutlich der Einfluss der Größe der Analysegeneration zu erkennen. Bei höheren Nachkommenzahlen können mehr Haplotypeneffekte sehr viel genauer geschätzt und der genomische Zuchtwert für die Folgegeneration effektiver bestimmt werden. Gerade bei niedrigen Heritabilitäten kann durch eine größere Stichprobe eine höhere Korrelation erreicht werden (siehe Abbildungen 25 und 26).

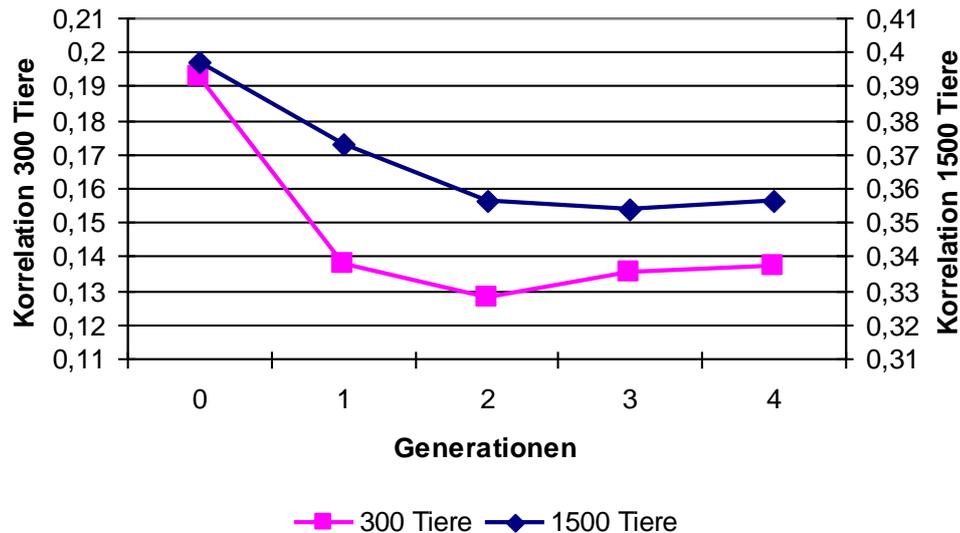


Abbildung 29: Einfluss der Größe der Analysegeneration auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und vier Folgegenerationen bei $h^2 = 0,1$

Deutlich wird dies auch anhand der Abbildung 29. Hier wird die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die Analysegeneration und vier Folgegenerationen für eine niedrige Heritabilität von 0,1 gezeigt. Dargestellt sind zwei verschiedene Größen der Analysegeneration von 300 Tieren verglichen mit 1500 Tieren. Die Korrelation bei 300 Tieren fällt von $r = 0,193$ auf $r = 0,137$ in der vierten Folgegeneration ab. Besonders stark ausgeprägt ist hier der Abfall von der Analysegeneration auf die erste Folgegeneration. Bei 1500 Tieren ist der Abfall nicht so stark. Hier fällt die Korrelation von $r = 0,397$ auf $r = 0,356$ in der vierten Folgegeneration. Auch bei Calus und Veerkamp (2007), Muir (2007) und Calus et al. (2008) fielen die Genauigkeiten in den Folgegenerationen für alle Heritabilitäten niedriger aus. Die Unterschiede zu den Ergebnissen in dieser Arbeit sind in den Kapiteln 4.2.7 und 4.2.8.1 beschrieben.

4.2.9 Die Selektion

4.2.9.1 Die zufällige Selektion

In Abbildung 30 wird die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die zufällige Selektion für drei verschiedene Selektionsszenarien und ohne Selektion bei einer Einzel-SNP-Analyse gezeigt.

Tabelle 16: Darstellung der Anzahl selektierter Tiere und der Selektionsintensität für drei verschiedene Szenarien

	Anzahl selektierter Tiere		Selektionsintensität	
	männliche	weibliche	männliche	weibliche
Szenario 1	50	125	1,10	0,30
Szenario 2	25	100	1,52	0,56
Szenario 3	15	75	1,76	0,80

Es ist zu sehen, dass die Korrelation für die Analysegeneration (Generation 0) kaum Veränderungen durch die unterschiedliche Selektionsintensität der drei verschiedenen Szenarien erfährt und bei $r = 0,866$ bis $0,869$ liegt. Die Korrelation für die Folgegenerationen fällt deutlich niedriger aus als die der Analysegeneration und liegt bei $r = 0,764$ bis $0,774$ in der ersten Folgegeneration und fällt auf $r = 0,741$ bis $0,756$ in der vierten Folgegeneration ab.

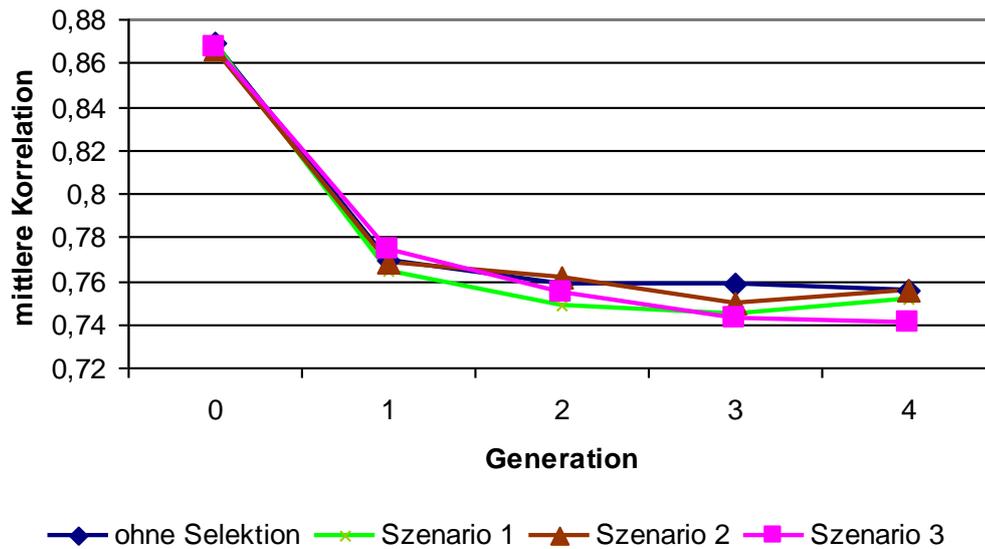


Abbildung 30: Einfluss der zufälligen Selektion auf die Korrelation zwischen wahren und dem genomischem Zuchtwert für vier verschiedene Szenarien für die Analysegeneration und vier Folgegenerationen

Damit unterscheiden sich die drei Selektionsszenarien und das Szenario ohne Selektion nur sehr geringfügig. Besonders für die Analysegeneration und die erste Folgegeneration fallen die Unterschiede extrem gering aus. Erst mit steigender Generationenanzahl nehmen die Unterschiede zwischen den Szenarien leicht zu. „Ohne Selektion“ weist dabei die höchsten Korrelationen auf. Mit steigender Selektionsintensität fallen die Genauigkeiten geringer aus, Szenario 3 hat die niedrigsten Korrelationswerte in Generationen 3 und 4.

Dieses Ergebnis entspricht den Ergebnissen von Muir (2007). Er testete eine zufällige Selektion, bei der aus 256 Tieren je 16 männliche und weibliche Tiere für die Zucht ausgewählt wurden. Die Genauigkeiten fielen in der Analysegeneration durch die Selektion um $r = 0,02-0,05$ niedriger aus. Auch in den Folgegenerationen fielen die Genauigkeiten durch die Selektion bis zu $r = 0,15$ geringer aus. Ein Vergleich der absoluten Werte mit den Werten in dieser Arbeit ist durch die unterschiedlichen Heritabilitäten nicht angezeigt.

Sämtliche genomweite Assoziationen zwischen Markern und QTL basieren auf Unterschieden in den Frequenzen von LD gegenüber nicht LD. Durch Selektion werden genetische Varianzen und Allelfrequenzen verändert und unvorteilhafte LD erzeugt (Bulmer, 1971), welche für die Berechnung von genomischen Zuchtwerten nachteilig ist (Muir, 2007).

4.2.9.2 Die gerichtete Selektion

In Abbildung 31 wird die Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die zufällige Selektion für drei verschiedene Selektionsszenarien und ohne Selektion bei einer Einzel-SNP-Analyse gezeigt.

Die Abbildung zeigt, dass die Korrelation für „ohne Selektion“ am höchsten ausfällt $r = 0,869$ (Generation 0) bis $r = 0,756$ (Generation 4) und deutlich von den anderen Selektionsszenarien zu unterscheiden ist. Die Selektionsszenarien zeigen alle deutlich niedrigere Korrelationen. Nur in der Analysegeneration liegen die Werte noch etwas höher und näher an dem für „ohne Selektion“. Die Korrelationen für die Szenarien 1-3 liegen in der Analysegeneration bei $r = 0,793$ bis $0,631$ und fallen in der ersten Folgegeneration stark auf $r = 0,256$ bis $0,137$ ab. Der Abfall von Generation 1 bis 4 ist danach nicht mehr so ausgeprägt. Die Korrelation in Generation 4 liegt bei $r = 0,112$ bis $0,118$. In der vierten Generation liegen die Werte damit sehr nah beieinander.

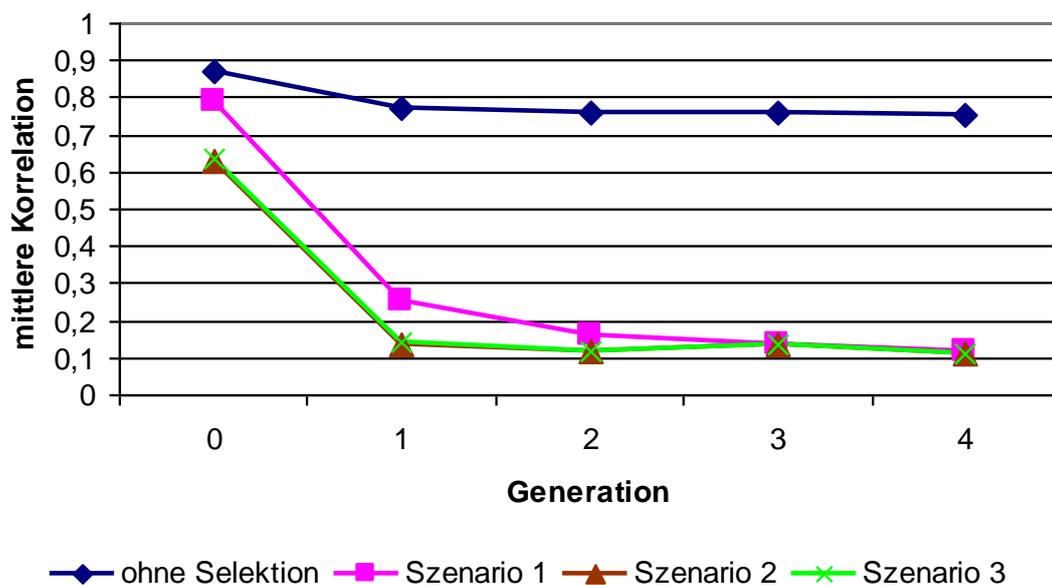


Abbildung 31: Einfluss der gerichteten Selektion auf die Korrelation zwischen wahren und dem genomischen Zuchtwert für vier verschiedene Szenarien für die Analysegeneration und vier Folgegenerationen

Dieses Ergebnis entspricht den Ergebnissen von Muir (2007). Er testete neben einer zufälligen Selektion auch eine gerichtete, bei der aus 256 Tieren je 16 männliche und weibliche Tiere für die Zucht ausgewählt wurden. Allerdings wurde die gerichtete Selektion erst nach den Analysegenerationen gestartet. Bis dahin wurden die Tiere zufällig selektiert. Die Selektion fand nach dem genomischen Zuchtwert statt. Die Genauigkeiten der genomischen Zuchtwerte fielen in den Analysegenerationen durch die Selektion ungefähr um 0,1 niedriger aus. In den Folgegenerationen fielen die Genauigkeiten sehr schnell sehr stark ab und lagen bis zu 0,35 unterhalb der Genauigkeiten ohne Selektion. Diskutiert wurde dieses Ergebnis schon in Kapitel 4.2.9.1. Als Verdeutlichung ist die genetische Varianz an allen Genorten für die beiden Heritabilitäten 0,5 und 0,1 in den folgenden Abbildungen 32 und 33 dargestellt.

Die Abbildung 32 zeigt für einen Lauf die genetische Varianz an allen Genorten für eine Heritabilität von 0,5. Die QTL-Positionen sind rot markiert. Man kann deutlich erkennen, dass die Varianz an allen 12 QTL-Orten auf null abfällt, während die Varianz an allen anderen Orten von 0,52 bis 0,58 schwankt. Durch die Selektion auf den Phänotyp wurden auf die QTL mit den positiven Werten selektiert und dieses Allel fixiert. Dadurch fallen die Genotypwerte aller Tiere gleich aus und auch die Phänotypwerte gleichen sich an, da die genetische Varianz null beträgt und der Resteffekt auch um einen Mittelwert schwankt. Daher ist eine sichere Schätzung der einzelnen Markereffekte nicht mehr gegeben und die Genauigkeiten nehmen einen geringen Wert an.

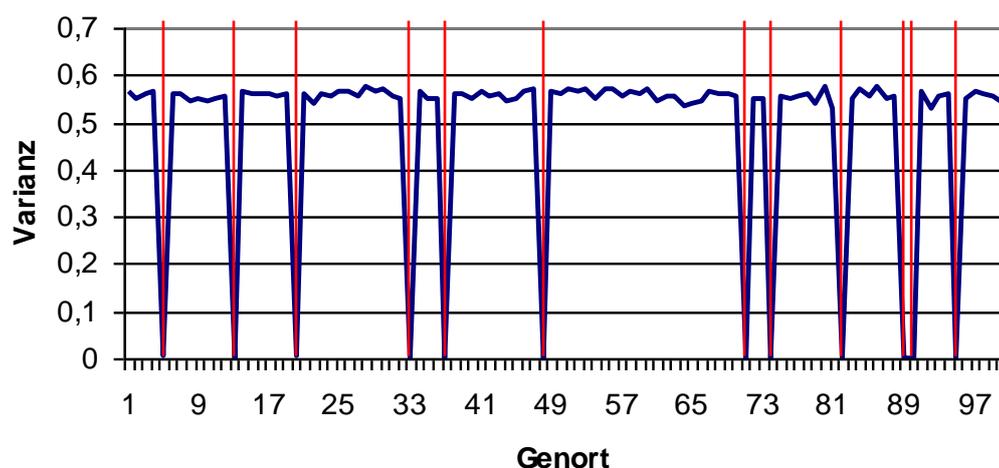


Abbildung 32: Mittlere Varianz an den Genorten für $h^2 = 0,5$ (QTL-Positionen rot)

Abbildung 33 zeigt die genetische Varianz an allen Genorten für eine Heritabilität von 0,1. Hier ist zu sehen, dass die Varianz an den QTL-Orten nicht ganz so stark abfällt wie bei $h^2 = 0,5$. Die geringste Varianz weist der QTL an Position 95 mit 0,39 auf, die höchste der an Position 74 mit 0,42. Für alle anderen Genorte schwankt die Varianz zwischen 0,54 und 0,58. Da hier die QTL-Allele nicht fixiert sind, fallen auch die Genotypwerte der Tiere unterschiedlich aus und eine sichere Schätzung der Markereffekte kann stattfinden.

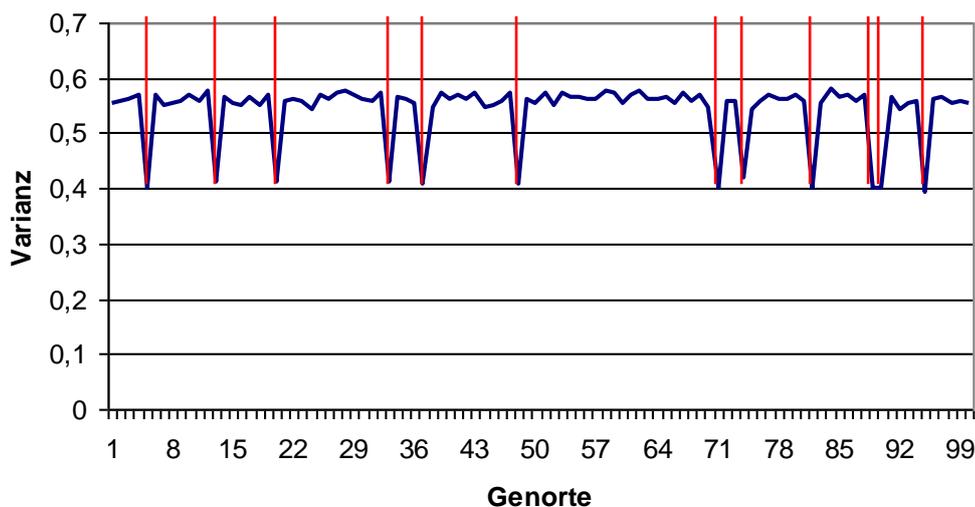


Abbildung 33: Mittlere Varianz an den Genorten für $h^2 = 0,1$ (QTL-Positionen rot)

4.3.9.3 Vergleich von zufälliger und gerichteter Selektion

In der Abbildung 34 wird der Einfluss der zufälligen Selektion im Vergleich zum Einfluss der gerichteten Selektion auf die Korrelation zwischen dem wahren und dem genomischen Zuchtwert aufgezeigt. Dargestellt ist jeweils der Wert ohne Selektion und die drei unterschiedlichen Szenarien für die Analysegeneration bei einer Heritabilität von 0,5.

Es wird gezeigt, dass sich die Korrelation für die zufällige Selektion nur sehr geringfügig von $r = 0,869$ (ohne Selektion) auf $r = 0,866$ (Szenario 3) ändert. Dagegen fällt die Korrelation bei der gerichteten Selektion von $r = 0,869$ (ohne Selektion) auf $r = 0,637$ in Szenario 3 ab. Hierbei wird deutlich, dass besonders bei der gerichteten Selektion durch die hohe Heritabilität von 0,5 QTL-Allele mit den positiven Werten fixiert werden und

keinerlei Variation mehr vorkommt. Eine sichere Schätzung der Markereffekte ist nicht mehr gegeben, die Korrelation fällt somit ab.

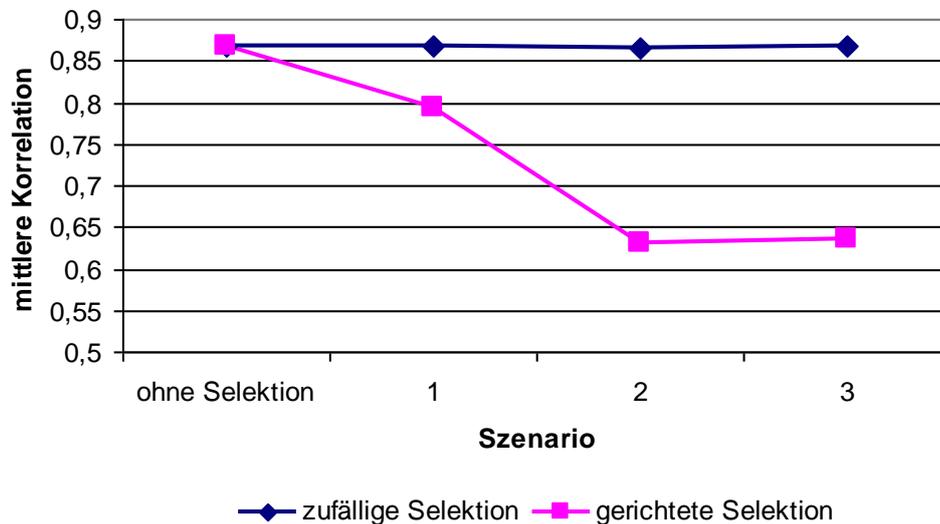


Abbildung 34: Einfluss der gerichteten und zufälligen Selektion auf die Korrelation zwischen wahren und dem genomischen Zuchtwert für die Analysegeneration und vier verschiedene Szenarien

4.3 Resampling

In der nachfolgenden Abbildung 35 wird ein Vergleich der Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die BLUP-Methode und das Resampling dargestellt. Berechnet wurden je zwei verschiedene Datensätze bei dem einer 200 SNPs, der andere 450 SNPs enthielt. In beiden Fällen wurde eine Einzel-SNP-Analyse vorgenommen und die Anzahl an QTL wurde auf 24 erhöht. Die Heritabilität lag bei 0,3.

Die erhöhten Anzahlen an SNPs wurden ausgewählt, um eine Überparameterisierung auch bei der Einzel-SNP-Analyse zu erreichen, da ein Resampling ansonsten nicht notwendig wäre. Da die Analysegeneration aus 300 Tieren bestand, war eine SNP-Anzahl von 100, woraus 200 verschiedene Haplotyp-Möglichkeiten entstehen können, nicht ausreichend. Aus diesem Grund wurde zum einen eine Anzahl von 200 SNP gewählt. Hiermit erhält man maximal 400 mögliche Haplotypen, wodurch eine leichte Überparameterisierung entsteht. In einer weiteren Einstellung wurden 450 SNPs gewählt.

Dadurch können 900 mögliche Haplotypen zustande kommen, wodurch mehr Haplotypen als Tiere zur Verfügung stehen.

In der Abbildung sind auf der linken Seite die Korrelationsergebnisse aus der Analysegeneration und auf der rechten die der Folgegeneration aufgezeigt. Die Ergebnisse für 450 SNPs sind vollfarbig, die für 200 SNPs gepunktet dargestellt.

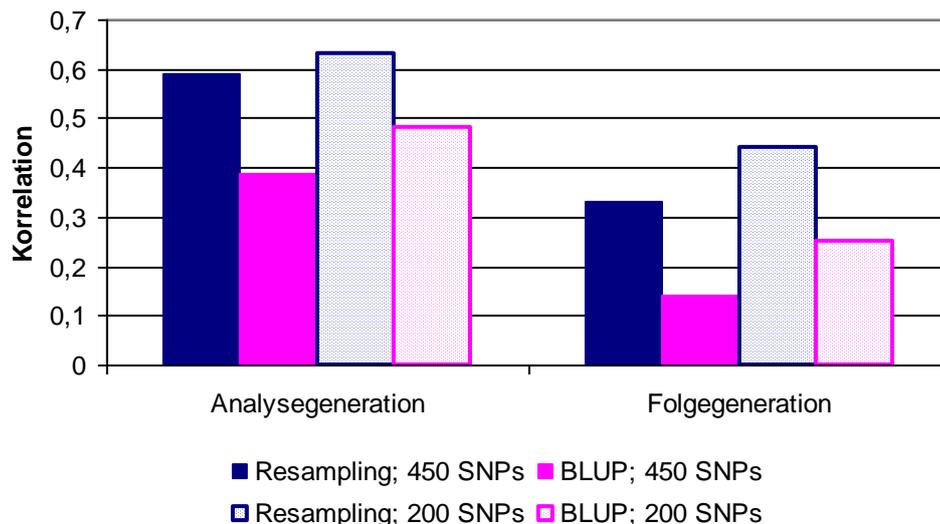


Abbildung 35: Vergleich der Korrelation zwischen dem wahren und dem genomischen Zuchtwert für die BLUP-Methode und das Resampling für zwei verschiedene Anzahlen an SNPs und Segmenten

Die Korrelationen innerhalb der Analysegeneration lassen erkennen, dass der Wert für 450 SNPs für das Resampling mit $r = 0,59$ gegenüber dem Wert für die BLUP-Methode ($r = 0,39$) deutlich höher ausfällt. Auch bei 200 SNPs ist der Resampling Wert ($r = 0,63$) höher als der BLUP-Wert ($r = 0,48$). Vergleicht man die Werte für 200 SNPs mit denen für 450 SNPs, so fällt auf, dass für 200 SNPs die höheren Korrelationswerte erreicht werden: $r = 0,63$ (200 SNPs) gegenüber $r = 0,59$ (450 SNPs) bei der Resampling-Methode und bei der BLUP-Methode $r = 0,48$ (200 SNPs) gegenüber $r = 0,39$ (450 SNPs).

Für die Folgegeneration zeigt sich eine ähnliche Konstellation, nur dass die Korrelationen hier geringer ausfallen. Bei 450 SNPs erreicht die Korrelation beim Resampling einen Wert von 0,33, während bei BLUP nur 0,14 geschätzt wird. Bei 200 SNPs liegt das Resampling mit $r = 0,44$ über dem des BLUP-Verfahrens ($r = 0,25$). Somit liegen die Korrelationen für 200 SNPs auch hier wieder über denen von 450 SNPs. Die Resampling-Methode liefert also unabhängig von der Generation für beide SNP-Anzahlen, 200 sowie

450 SNPs, höhere Korrelationen als die BLUP-Methode. Des Weiteren werden bei 200 SNPs höhere Werte als bei 450 SNPs erzielt.

Die niedrigeren Korrelationen bei 450 SNPs verglichen mit 200 SNPs für beide Methoden ist auf die veränderte Markerdichte zurückzuführen, die schon in Kapitel 4.2.6 behandelt wurde.

Die Korrelationen für die BLUP-Methode fallen geringer aus als die für das Resampling, da bei BLUP alle Haplotypeneffekte gleichzeitig geschätzt werden. Beim Resampling dagegen werden pro Runde immer nur so viele Effekte geschätzt wie auch Tiere mit Leistungen vorliegen. Durch die 10000 Runden werden so die einzelnen Haplotypen mehrmals ausgewählt und ihre Effekte geschätzt. Der so aus den einzelnen Gewichtungsfaktoren gebildete Mittelwert führt zu genaueren Ergebnissen. Ein weiterer Vorteil liegt in der variablen Varianz für die einzelnen SNPs/Haplotypen. In der Praxis haben nicht alle Haplotypen die gleiche genetische Varianz, da nur wenige QTLs mit großem Effekt beinhalten, einige QTLs mit mittleren bis geringen Effekten und andere Haplotypen enthalten gar keine QTLs. Von daher ist eine Annahme mit variablen Varianzen vorzuziehen (Goddard & Hayes, 2007; Meuwissen et al., 2001; Muir, 2007).

Ein großer Nachteil des Resamplings ist die lange Laufzeit des Programms, die schon bei diesen geringen Tier- und SNP-Zahlen zirka zwei Wochen in Anspruch nahm. Auf weitere Berechnungen wurde daher verzichtet.

5 Schlussfolgerung

Die Ergebnisse dieser Arbeit sind außerordentlich ermutigend. Die Genauigkeiten der genomischen Zuchtwerte nehmen schon bei geringen Anzahlen von genotypisierten Tieren und niedrigen Markerdichten bei hoher Heritabilität Werte um die 80% an. Durch Verwendung einer Einzel-SNP-Analyse anstelle von Haplotypen konnte die Genauigkeit auf knapp 90% angehoben werden und erreichte bei hohen genotypisierten Tierzahlen sogar Werte über 95%. Bei niedrigen Heritabilitäten fielen die Genauigkeiten der genomischen Zuchtwerte sehr gering aus, konnten allerdings durch Erhöhung der genotypisierten Tierzahlen auf über 50% angehoben werden. Diese Ergebnisse zeigen deutlich den großen Einfluss der Anzahl der Tiere, die zur Analyse der Markereffekte herangezogen werden. Je größer die Zahl der Tiere ausfällt, desto genauer werden die genomischen Zuchtwerte geschätzt, was sich besonders bei niedrigen Heritabilitäten bemerkbar macht. Somit liegen die ermittelten Genauigkeiten in einem hohen Bereich und übertreffen die Sicherheiten der Zuchtwerte aus der BLUP-Zuchtwertschätzung der Schweinezucht, die im Bereich von 72 bis 76 % liegen können, bei Prüfung in Reinzucht und im Ferkelerzeugungsbereich (Dodenhoff, 2004).

Eine Erhöhung der Populationsgröße erzielte bei Muir (2007) geringere Genauigkeiten, da durch höhere Populationsgrößen des LD herabgesetzt wird. In dieser Arbeit konnten nur geringfügig niedrigere Genauigkeiten aufgezeigt werden, die sich bei höheren Tierzahlen von 1000 wieder relativierten. Da zur Schweinefleischproduktion hauptsächlich nur vier Rassen (Deutsches Edelschwein, Deutsche Landrasse, Piétrain und Duroc) in der Reinzucht mit effektiven Populationsgrößen um die 1500 bis 2000 Tiere für Deutsches Edelschwein und Deutsche Landrasse eingesetzt werden (Bundesanstalt für Landwirtschaft und Ernährung, 2008), sollten dadurch keine geringeren Genauigkeiten erzielt werden.

Höhere Markerdichten lieferten geringere Genauigkeiten um die 75%, die auch mit erhöhten genotypisierten Tierzahlen nur geringfügig angehoben werden konnten. Hier scheint es ein optimales Marker:QTL-Verhältnis zu geben, das bei Muir (2007) bei 10:1 bzw. 2,5:1 lag, was dem Verhältnis von 8,3:1 in dieser Arbeit ungefähr entspricht. Eine Erhöhung der Markerdichte bei gleich bleibender QTL-Anzahl scheint daher kontraproduktiv zu sein, könnte allerdings auch auf die BLUP-Methode beschränkt sein und durch Anwendung anderer Methoden behoben werden (Muir, 2007).

Durch gerichtete Selektion auf den Phänotyp wurden geringere Genauigkeiten von 80% bzw. knapp 65% bei höheren Selektionsintensitäten erzielt. Allerdings wurden die Werte bei hohen Heritabilitäten geschätzt, wodurch an den Genorten der QTLs keinerlei Varianz

vorkommt. Bei geringeren Heritabilitäten und höheren Varianzen an den Genorten sollten die Genauigkeiten weniger stark abfallen.

Die Genauigkeit der genomischen Zuchtwerte fällt in den Folgegenerationen von 87% auf ca. 75%. Durch eine Erhöhung der genotypisierten Tierzahlen konnte die Genauigkeit in den Folgegenerationen stark angehoben werden und fiel nur geringfügig von knapp 97% auf 96% ab. Dieses Ergebnis unterstreicht wieder die hohe Bedeutung der Größe der analysierten Tierzahlen für die erzielte Genauigkeit. Allerdings muss man die simulierten Daten unter den Voraussetzungen betrachten, dass die Rahmenbedingungen optimiert waren. Es wurde nur ein Chromosom simuliert. Die SNPs haben die QTLs direkt abgedeckt und die simulierten QTLs hatten alle den gleichen relativ hohen Effekt. Die simulierte Population hatte nur 15 Generationen, und die Anzahl der Tiere in der Analysegruppe ist relativ gering. Diese Rahmenbedingungen sind möglicherweise optimistischer (genetische Annahmen), als in der Praxis zu finden. Es kann auch davon ausgegangen werden, dass die Anzahl der zu typisierenden Tiere größer ausfällt, auch vor dem Hintergrund stetig sinkender Typisierungskosten.

Der Einsatz der genomic selection in der Schweinezucht ist aufgrund der Ergebnisse dieser Arbeit als positiv zu bewerten. In diesen Zuchtpopulationen ist eine komplexe Merkmalsstruktur zu finden. Merkmale mit hoher Heritabilität (z.B. Fleischleistungs- und Wachstumsmerkmale) werden kombiniert mit Merkmalen mit geringerer Heritabilität (z.B. Reproduktionsleistungen). In Kombination in einem Index kommen züchterisch unvorteilhafte genetische Korrelationen dazu (z.B. Magerfleischanteil und Zunahme), die einen hohen Zuchtfortschritt erschweren.

Die genomic selection ist eine Erfolg versprechende Methode, die gerade bei Merkmalen mit hoher Heritabilität hohe Genauigkeiten erzielt. Auch bei niedrigen Heritabilitäten können hohe Genauigkeiten erzielt werden, indem die Größe der Analysegeneration erhöht wird. Ein großer Vorteil der Methode ist die Selektionsmöglichkeit der Tiere zu einem sehr frühen Zeitpunkt direkt nach der Geburt oder sogar früher. Schon zu diesem Zeitpunkt liegt ein Zuchtwert mit hoher Sicherheit für das Tier vor, der die Sicherheit der normalen BLUP-Zuchtwertschätzung übertrifft. Auf den Grundlagen des genomischen Zuchtwertes kann somit eine Vorselektion der zukünftigen Zuchttiere stattfinden. Schaeffer (2006) hat für die Rinderzuchtverbände dadurch geringere Kosten um 92% berechnet. In der Schweinezucht sind die Generationsintervalle jedoch geringer, da die Reproduktionszyklen viel kürzer als in der Rinderzucht sind. Die Einsatzzeit eines Besamungsebers ist kürzer (ca. 6 Monate), und die Anzahl der getesteten Nachkommen geringer, ca. 8-12 getestete Nachkommen pro Besamungseber auf Station. Kosteneffekte

wie in der Rinderzucht können daher in der Schweinezucht nicht erwartet werden. Der Vorteil der frühen Selektionsmöglichkeit kommt aber besonders bei Merkmalen mit später Ausprägung oder Schlachtmerkmalen zum Tragen, welche in der Schweinezucht von höchstem ökonomischen Interesse sind. Insbesondere Merkmale, die erst nach einer Schlachtung des Tieres gemessen werden können, sind für den Einsatz der genomic selection interessant. Für diese Merkmale, z.B. Fleischanteile im Bauch, Schinkenprozent und andere Merkmal des AutoFOM können nicht als Eigenleistung eines Besamungsebers angegeben werden. Diese müssen durch Nachkommenprüfungen auf Teststationen und besonders geregelter Schlachtung mit Einzeltiererkennung und Rückverfolgung der Leistung ermittelt werden. Die ökonomisch sehr bedeutsamen Merkmale können mit der genomic selection nicht nur viel früher mit einer angemessenen Genauigkeit Zuchtwertgeschätzt wird, es wird auch möglich, einen Zuchtwert als Eigenleistung eines Besamungsebers anzugeben.

Die genomic selection ermöglicht noch weitere Merkmalsgruppen abzudecken, die von Bedeutung sind. Durch den Einsatz von SNP-Chips werden tausende SNP Informationen verfügbar, die sich einfach zu weiteren Analysen eignen. In den letzten Jahren hat sich die Forschung auf dem Gebiet der Gendefekte beim Schwein rasant entwickelt. Auch in Deutschland wurde im Rahmen der Projekte FBF I und FBF II an Erbfehlern geforscht und bereits Patente zu den Merkmalen Afterlosigkeit (Atresia ani), Brüche (Hernien), Gesäugeanomalien (Stülpzitzen) und Spreizer (Splayleg syndrome) angemeldet (FBF, 2008). An den Ergebnissen waren vier Forschungsinstitute beteiligt. Anhand der Hernien lässt sich die Entwicklung der genomischen Diagnostik beim Schwein erkennen. Die anfänglichen Arbeiten wurden mit Mikrosatelliten durchgeführt, die schwer verfügbar und selten waren und teilweise für die Untersuchungen eigenst entwickelt werden mussten (Bornemann-Kolatzki, 2004). Zeitgleich wurden SNPs verfügbar und veröffentlicht (Knorr et al., 2004). In weiteren Arbeiten wurde ein potentielles Kandidatengen mit Hilfe von SNPs gesucht. Im Rahmen der Hernien konnte noch kein Gen identifiziert werden (Beck et al., 2006). Der Einsatz von SNP-Chips bei der genomischen Selektion ermöglicht eine genomweite Analyse und erhöht die Chance auf eine sichere Vorausschätzung der Übertragung von Erbkrankheiten (Daetwyler et al., 2008).

6 Zusammenfassung

Eine starke Entwicklung der Tierzucht durch fortschreitendes Verständnis der Genetik und Weiterentwicklung der Methoden fand im 20. Jahrhundert statt. Besonders bei Leistungsmerkmalen mit mittlerer/hocher Heritabilität konnte ein bemerkenswerter Zuchtfortschritt erzielt werden. Die bisher angewandten Selektionsverfahren haben aber auch ihre Grenzen, zum Beispiel bei Merkmalen mit geringer Heritabilität wie später Ausprägung und Schlachtmerkmalen. Weiterführende Methoden wie die genomic selection, die auf der Verwendung von Genominformationen basieren sind daher von großem Interesse für die Tierzucht.

Das Ziel dieser Arbeit war es, in Anlehnung an eine simulierte Schweinepopulation die genomischen Zuchtwerte für jedes Tier über das BLUP- sowie ein Resampling-Verfahren für die verschiedensten Bedingungen, wie unterschiedliche Populationsgröße, unterschiedliche Heritabilitäten, Selektionsszenarien oder Haplotypengrößen zu berechnen. Weiterhin wurden die Genauigkeiten einer BLUP-Zuchtwertschätzung mit denen der genomischen Zuchtwerte verglichen.

Simuliert wurde eine Population mit 300 Tieren. Diese Populationsgröße wurde über 15 Generationen unter Zufallspaarung konstant gehalten. In der 15. Generation wurde die Tieranzahl durch Erhöhung der Nachkommenanzahl auf 10 je Elternpaar angehoben und blieb in den folgenden Generationen konstant bei 1500 Tieren je Generation. Aus der 16. Generation wurden zufällig 300 Tiere gezogen, die für die Analyse der Marker/Haplotyp-Effekte herangezogen wurden. Für alle Tiere wurde ein Chromosomenpaar mit je einem Morgan Länge und 100 SNPs simuliert, die in der Gründergeneration zufällig vergeben wurden. 12 SNPs wurden zufällig ausgewählt und als QTLs ausgewiesen, deren Allele einen Wert zugeteilt bekamen, deren Summe den Genotypwert für jedes Tier ergab. Für jedes Tier wurde zusätzlich ein Phänotypwert simuliert. Anhand der Daten der 300 ausgewählten Tiere wurden die Marker/Haplotyp-Effekte geschätzt, mit denen die genomischen Zuchtwerte für jedes Tier berechnet wurden.

Folgende Ergebnisse wurden erzielt:

- Die Anzahl und Zusammensetzung der Analysegruppe hatte einen deutlichen Einfluss auf die Genauigkeit der genomischen Zuchtwerte. So stieg die Genauigkeit mit steigender Anzahl von Tieren von $r = 0,8 - 0,81$ bis auf $r = 0,84 - 0,87$. Aber auch die Zusammensetzung machte einen Unterschied aus. So wurde bei gleicher Gesamtanzahl mit wenigen Vollgeschwistern von mehreren Elternpaaren niedrigere Genauigkeiten erzielt, als mit wenigen Elternpaaren aber vielen Vollgeschwistern.

- Die Anzahl der Generationen vor der Analysegruppe hatte erst bei hohen Generationenzahlen einen größeren Einfluss auf die Genauigkeit der genomischen Zuchtwerte.
- Die Populationsgröße hatte in dieser Arbeit nur einen sehr geringen Einfluss auf die Genauigkeit der genomischen Zuchtwerte.
- Die Größe der zu analysierenden Chromosomensegmente hatte einen sehr ausgeprägten Einfluss auf die Genauigkeit. Hier konnten die höchsten Genauigkeiten mit einer Einzel-SNP-Analyse erreicht werden. Auch die Veränderung der Heritabilität von $h^2 = 0,1$ auf $0,5$ hatte einen Anstieg der Genauigkeit der genomischen Zuchtwerte von $r = 0,19$ auf $0,87$ zur Folge.
- Eine Veränderung der Mutationsrate von $2,5 \times 10^{-3}$ auf $2,5 \times 10^{-1}$ hatte keinerlei Einfluss auf die Genauigkeit der genomischen Zuchtwerte. Der Wert blieb konstant bei $r = 0,87$.
- Eine Erhöhung der SNP-Anzahl von 100 auf 200 SNPs konnte keine höhere Genauigkeit der genomischen Zuchtwerte bewirken. Auch durch eine größere Analysegeneration von 1000 Tieren änderte sich das Verhältnis nicht, und die Genauigkeit von 200 SNPs blieb unter der von 100 SNPs.
- Die Heritabilität zeigte einen deutlichen Einfluss auf die Genauigkeit des genomischen Zuchtwertes. Er fiel von $r = 0,87$ bei $h^2 = 0,5$ auf $r = 0,19$ bei $h^2 = 0,1$ ab. Durch eine Erhöhung der Analysegeneration auf 1500 Tiere konnte die Genauigkeit des genomischen Zuchtwertes für $h^2 = 0,1$ auf $r = 0,51$ angehoben werden.
- Die Genauigkeiten für die Folgegenerationen nach der Analysegeneration ergaben, dass die Genauigkeiten stark abnahmen je kleiner die Größe der Analysegeneration ausfiel und je größer die Chromosomensegmente waren, die analysiert wurden. Auch der Einfluss der Anzahl der Folgegenerationen nach der Analysegeneration war sehr groß und wirkte sich negativ auf die Genauigkeit aus.
- Die zufällige Selektion hatte in der Analysegeneration verglichen mit dem Wert ohne Selektion keinen Einfluss auf die Genauigkeit des genomischen Zuchtwertes. Auch in den Folgegenerationen war nur ein leichter Unterschied auszumachen. Die gerichtete Selektion auf den Phänotyp hatte dagegen einen deutlichen Einfluss auf die Genauigkeit, der besonders in den Folgegenerationen sichtbar wurde.

- Das Resampling-Verfahren ergab deutlich bessere Genauigkeiten für den genomischen Zuchtwert als das BLUP-Verfahren. Die Korrelationen fielen hier um gut 0,1 höher aus.

7 Summary

Modern developments in animal breeding in combination with advanced understanding of molecular genetics proceeded in the end of the 20th century. Most genetic gain was achieved in production traits with high/moderate heritabilities. All known selection strategies are limited for traits with low heritabilities or carcass traits. Therefore new developments as genomic selection, using direct genetic information, are of great interest in animal breeding.

Goal of this study was to estimate genomic breeding values for individuals, based on the framework of a pig population. Methods used were a BLUP and a resampling approach. Different parameters were tested, different population sizes, heritabilities, selection strategies and haplotypes. The accuracy of genomic breeding values and BLUP breeding values were compared.

A population of 300 animals was simulated. This population size has been kept constant for the first 15 generations. In generation 15 the number of animals was increased because of the higher number of offspring (10) per parents. This number (1500) was kept constant for the following generations. A subsample of 300 animals was drawn randomly from generation 16. These animals were used to estimate first marker/haplotype effects. The simulation for all animals based on one chromosome with a length of one morgan containing 100 SNP's, which were randomly spread over the whole sample. A total of 12 SNP's were randomly chosen to become QTL's, whose alleles defined the true breeding value of each animal. Additionally a phenotypic value for each animal was simulated. Based on the data of these 300 animals the haplotype/marker effects were estimated to obtain genomic breeding values for each animal.

Results were:

- The size and structure of the training population had a clear influence on the accuracy of genomic breeding values. The accuracy increased with increasing number of animals from $r = 0.8 - 0.81$ up to $r = 0.84 - 0.87$. The structure caused differences, too. The same number of animals based on high number of parents having low number of full sibs achieved lower accuracies as low number of parents with high number of full sibs.

- The number of generations simulated before defining the training generations had no influence on the accuracy, only with very high number the accuracy of genomic breeding values changed.
- Population size had low effect on the accuracy of genomic breeding values.
- The size of the analysed chromosome segment had a distinct influence on the accuracy. The most accurate values were found using single SNP's. Changing the heritability from $h^2 = 0.1$ to 0.5 caused an increased accuracy from $r = 0.19$ to 0.87 .
- Changing the mutation rate from 2.5×10^{-3} to 2.5×10^{-1} had no significant influence on the accuracy of $r = 0.87$.
- Raising the number of SNP's from 100 to 200 showed no positive effect on the accuracy. An increased size of the training population of 1000 did not change the accuracy and the values for 200 SNP's were below the values of 100 SNP's.
- The heritability had a significant influence on the accuracy. It decreased from $r = 0.87$ ($h^2 = 0.5$) to $r = 0.19$ ($h^2 = 0.1$). Increasing the size of the training population could raise the accuracy to $r = 0.51$ ($h^2 = 0.1$).
- Accuracies for later generations than the training generation showed large decrease of accuracies, the smaller the training populations and the larger the chromosome segments, the larger the decrease in accuracy. The number of later populations also showed huge influence on the accuracies.
- Random selection in comparison to no selection in the training population showed no influence on the accuracy. For later generations this caused only small changes. Selection on the phenotype had large influence on the accuracy.

The resampling method obtained higher accuracies than the BLUP method. Correlations between true and estimated breeding values were about 0.1 higher in average.

Literaturverzeichnis

- Aitken, N., Smith, S., Schwarz, C. and Morin, P. A.** (2004). Single nucleotide polymorphism (SNP) discovery in mammals: a targeted-gene approach. *Molecular Ecology* **13**, 1423-1431.
- Andersson, L.** (2001). Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* **2**, 130-138.
- Beck, J., Bornemann-Kolatzki, K., Knorr, C., Taeubert, H. and Brenig, B.** (2006). Molecular characterization and exclusion of porcine GUSB as a candidate gene for congenital hernia inguinalis/scrotalis. *BMC Veterinary Research* **28**;2:14.
- Bertsch, T. and Neumaier, M.** (2003). Wie korrelieren Genotyp und Phänoytp? *Klinikarzt* **32**, 10-14.
- Boichard, D., Fritz, S., Rossignol, M. N., Guillaume, F., Colleau, J. J. and Druet, T.** (2006). Implementation of marker assisted selection: practical lesson from dairy cattle. *8th World Congress on Genetics Applied to Livestock Production, August 13-18. Belo Horizonte, Brasil.*
- Bornemann-Kolatzki, K.** (2004). Durchführung eines Genomscans mit polymorphen DNA-Markern und Genomic Mismatch Scanning (GMS) bei *Sus scrofa* zur Detektion Hernia inguinalis/scrotalis assoziierter Genomregionen. Hannover, Tierärztliche Hochschule, Dissertation
- Botstein, D., White, R. L., Skolnick, M. and Davis, R. W.** (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *The American Journal of Human Genetics* **32**, 314-331.
- Bulmer, M. G.** (1971). Effect of selection on genetic variability. *The American Naturalist* **105**, 201-210.
- Bundesanstalt für Landwirtschaft und Ernährung** (2008). Rote Liste der gefährdeten einheimischen Nutztierassen in Deutschland. Bundesanstalt für Landwirtschaft und Ernährung. Zugriff: 28.11.2008, http://www.genres.de/downloads/publikationen/rote_liste.pdf.

- Calus, M. P., Meuwissen, T. H., de Roos, A. P. and Veerkamp, R. F.** (2008). Accuracy of genomic selection using different methods to define haplotypes. *Genetics* **178**, 553-561.
- Calus, M. P. and Veerkamp, R. F.** (2007). Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. *Journal of Animal Breeding and Genetics* **124**, 362-368.
- Cichon, S., Freudenberg, J., Propping, P. and Nöthen, M. M.** (2002). Variabilität im menschlichen Genom. Deutsches Ärzteblatt, Jg. 99, Heft **46**
- Comberg, G.** (1984). "Die deutsche Tierzucht im 19. und 20. Jahrhundert," Ulmer, Stuttgart
- Daetwyler, H. D., Villanueva, B. and Wooliams, J. A.** (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE* **3**(10): e3395. doi:10.1371/journal.pone.0003395
- Darvasi, A. and Soller, M.** (1994). Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theoretical and Applied Genetics* **89**:351-357.
- Davis, G. P. and DeNise, S. K.** (1998). The impact of genetic markers on selection. *Journal of Animal Science* **76**, 2331-2339.
- de Koning, D. J., Archibald, A. and Haley, C. S.** (2007). Livestock genomics: bridging the gap between mice and men. *Trends in Biotechnology* **25**, 483-489.
- Dekkers, J. C.** (2004). Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *Journal of Animal Science* **82** E-Suppl, E313-328.
- Dekkers, J. C.** (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *Journal of Animal Breeding and Genetics* **124**, 331-341.
- Dodenhoff, J.** (2004). Sicherheiten in der Zuchtwertschätzung für Schweine. Bayerische Landesanstalt für Landwirtschaft. Zugriff: 28.11.2008, <http://www.lfl.bayern.de/itz/schwein/09450/index.php>
- Du, F. X., Clutter, A. C. and Lohuis, M. M.** (2007). Characterizing linkage disequilibrium in pig populations. *International Journal of Biological Sciences* **3**, 166-178.

- Ellegren, H., Smith, N. G. and Webster, M. T.** (2003). Mutation rate variation in the mammalian genome. *Current Opinion in Genetics & Development* **13**, 562-568.
- Förderverein Biotechnologieforschung e.V. (FBF)** (2008). Projekt: Genomanalyse Schwein. Zugriff: 07.12.2008, http://www.fbf-forschung.de/schwein_genomanalyse.html.
- Goddard, M. E. and Hayes, B. J.** (2002). Optimisation of response using molecular data. *Proceedings 7th World Congress on Genetics Applied to Livestock Production* **33**, 3-10.
- Goddard, M. E. and Hayes, B. J.** (2007). Genomic selection. *Journal of Animal Breeding and Genetics* **124**, 323-330.
- Habier, D., Fernando, R.L. and Dekkers, J.C.** (2007). The Impact of Genetic Relationship Information on Genome-Assisted Breeding Values. *Genetics* **177**, 2389-2397.
- Hayes, B. J.** (2008). QTL Mapping, MAS and Genomic Selection; Course Notes.
- Hayes, B. J., Chamberlain, A. J., McPartlan, H., Macleod, I., Sethuraman, L. and Goddard, M. E.** (2007). Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genetic Research* **89**, 215-220.
- Hazel, L. N.** (1943). The Genetic Basis for Constructing Selection Indexes. *Genetics* **28**, 476-490.
- Henderson, C. R.** (1950). Estimation of genetic parameters. *Annals of Mathematical Statistics* **21**, 309.
- Henderson, C. R.** (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447.
- Henderson, C. R.** (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32**:69.
- Knippers, R.** (2001). "Molekulare Genetik," 8. überarbeitete Auflage, Thieme, Stuttgart
- Knorr, C., Täubert, H., Peters, U. and Brenig, B.** (2004). Characterization of two SNPs (single nucleotide polymorphisms) in the porcine INSL3 gene and their exclusion as a common genetic basis of hernia inguinalis in pigs. *Biochemical Genetics*. **42**(1-2):11-19.

- Kolbehdari, D., Schaeffer, L. R. and Robinson, J. A.** (2007). Estimation of genome-wide haplotype effects in half-sib designs. *Journal of Animal Breeding and Genetics* **124**, 356-361.
- Kräußlich, H.** (1997). "Tierzuchtungslehre," 4. Auflage, UTB
- Kumar, S. and Subramanian, S.** (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Science U S A* **99**, 803-808.
- Lande, R. and Thompson, R.** (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-756.
- Lercher, M. J. and Hurst, L. D.** (2002). Human SNP variability and mutation rate are higher in regions of high recombination. *Trends in Genetics* **18**, 337-340.
- Littmann, E., Götz, K. U. and Dodenhoff, J.** (2006). "Schweinezucht und Schweineproduktion," Schriftenreihe Bayerische Landesanstalt für Landwirtschaft (LfL). Zugriff:28.11.2008
http://www.lfl.bayern.de/publikationen/daten/schriftenreihe_url_1_39.pdf.
- Mäki-Tanila, A.** (2006). Seeing the genomic value of markers. *Journal of Animal Breeding and Genetics* **123**, 217.
- Meuwissen, T.** (2003). Genomic selection: the future of marker assisted selection and animal breeding. FAO Workshop: Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding? Zugriff: 28.11.2008
<http://www.fao.org/biotech/docs/Meuwissen.pdf>
- Meuwissen, T.** (2007). Genomic selection: marker assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics* **124**, 321-322.
- Meuwissen, T. H., Hayes, B. J. and Goddard, M. E.** (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829.
- Meuwissen, T. H. and Van Arendonk, J. A.** (1992). Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science* **75**, 1651-1659.
- Montaldo, H. H. and Meza-Herrera, C. A.** (1998). Use of molecular marks and major genes in the genetic improvement of livestock. *Electronic Journal of Biotechnology* **1:2,83-89**.
- Mrode, R. A.** (1996). "Linear Models for the Prediction of Animal Breeding Values," CAB International, 1. Auflage

- Muir, W. M.** (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics* **124**, 342-355.
- Rothschild, M. F. and Soller, M.** (1997). Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. *Probe* **8**:13-20.
- Rutten, M. J., Bijma, P., Woolliams, J. A. and van Arendonk, J. A.** (2002). SelAction: software to predict selection response and rate of inbreeding in livestock breeding programs. *Journal of Heredity* **93**, 456-458.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A.** (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
- Schaeffer, L. R.** (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics* **123**, 218-223.
- Schulman, N. F. and Dentine, M. R.** (2005). Linkage disequilibrium and selection response in two-stage marker-assisted selection of dairy cattle over several generations. *Journal of Animal Breeding and Genetics* **122**, 110-116.
- Schwerin, M.** (2003). Die funktionelle Genomanalyse - neue Möglichkeiten für Tierzucht und Tierhaltung. *Archiv für Tierzucht* **46** Sonderheft, 89-93.
- Solberg, T. R., Sonesson, A. K., Woolliams, J. A. and Meuwissen, T. H.** (2008). Genomic selection using different marker types and densities. *Journal of Animal Science* **86**, 2447-2454.
- Soller, M.** (1978). The use of loci associated with quantitative effects in dairy cattle improvement. *Animal Production* **27**:133-138.
- Watson, J. D. and Crick, F. H.** (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738.
- Williams, J. L.** (2005). The use of marker-assisted selection in animal breeding and biotechnology. *Review of Science and Technology* **24**, 379-391.
- Wolf, E.** (2008) Nutztiere - quo vadite? Neue Wege durch funktionale Genomanalyse. www.fugato-forschung.de/download.php/29/nutztiere_wolf_neu.pdf Zugriff: 03.01.2008
- Wolz, W.** (1993). Anwendung von Mikrosatelliten als genetische Marker. Diplomarbeit, Institut für Humangenetik, Bayerische Julius-Maximilians-Universität Würzburg.

- Woolliams, J. A. and Smith, C.** (1988). The value of indicator traits in the genetic improvement of dairy cattle. *Animal Production* **46**:333.
- Zhao, H., Nettleton, D., Soller, M. and Dekkers, J. C.** (2005). Evaluation of linkage disequilibrium measures between multi-allelic markers as predictors of linkage disequilibrium between markers and QTL. *Genetic Research* **86**, 77-87.
- Zhao, H. H., Fernando, R. L. and Dekkers, J. C.** (2007). Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. *Genetics* **175**, 1975-1986.

Danksagung

Bedanken möchte ich mich bei Prof. Dr. Henner Simianer für die Überlassung des Themas und die gute Betreuung und Unterstützung, die er mir während der gesamten Arbeit entgegengebracht hat, sowie bei Prof. Dr. Christoph Knorr für die spontane Übernahme des Korreferats.

Dem Fugato HeDipig-Projekt sowie dem Zentralverband der Deutschen Schweineproduktion e.V. (ZDS) danke ich für die finanzielle Unterstützung des Projektes.

Vielen Dank an meine Arbeitsgruppe: Rike, Bianca, Janet, Ding, Florence, Florian, Christian, Sven, Tamina, Frau Döring, Sven, Reza, Manfred – für seine tatkräftige Unterstützung beim Erklären und Weiterentwickeln des Programmes.

Besonderer Dank gilt meiner Bürokollegin Maria, für eine wundervolle Zeit mit ihr im Büro. Mit ihrer immer fröhlichen, aufgeschlossenen und fürsorglichen Art, war sie stets mit Rat und Tat zur Stelle.

Nicht zu vergessen sei an dieser Stelle Monique, die mir als gute Freundin vom ersten Tag des Studiums zur Seite stand. Vielen Dank, Monique, für Deine Unterstützung, Hilfe, Kritik und das immer offene Ohr.

Helge danke ich für seine unendliche Geduld, Motivation und Hilfe. Du bereicherst mein Leben. Danke für Alles.

Meinen Eltern danke ich für ihre grenzenlose Unterstützung und bedingungslose Liebe. Ich danke euch, dass ihr mich auf allen meinen Wegen begleitet habt und mir dabei geholfen habt, zu dem Menschen zu werden, der ich heute bin.

Lebenslauf

Name	Dörthe Agena
Geburtsdatum/-ort	10.Mai 1978 in Göttingen
Staatsangehörigkeit	deutsch
Familienstand	ledig

Berufliche Tätigkeit

Seit 2005	Wissenschaftlicher Mitarbeiter am Institut für Tierzucht und Haustiergenetik der Georg-August-Universität Göttingen, Arbeitsgruppe Tierzucht (Prof. Dr. H. Simianer); beschäftigt mit der Anfertigung der Dissertation
-----------	--

Studium

1997 – 2005	Studium der Agrarwissenschaften an der Georg-August-Universität Göttingen, Schwerpunkt Tierproduktion, Abschluss: Master of Science
-------------	---

Schullaufbahn

1990 – 1997	Theodor-Heuss-Gymnasium Göttingen Abschluss: Abitur
1988 – 1990	Bert-Brecht-Schule Göttingen, Orientierungsstufe
1984 – 1988	Wilhelm-Busch-Schule Göttingen, Grundschule

Praktika

2000	Praktikum in dem Friesenstall „An der Moore“, Moringen
1999	Praktikum beim Steuerberater, Northeim
1998	Praktikum auf landwirtschaftlichen Betrieb, Mollenfelde

Göttingen, im Dezember 2008