

Feinkartierung und Identifikation der kausalen Kandidatengene
für die Schwanzlänge beim Merinolandschaf

von Dominik Karl Lagler

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Tierärztlichen Fakultät der Ludwig-Maximilians-Universität
München

Feinkartierung und Identifikation der kausalen Kandidatengene
für die Schwanzlänge beim Merinolandschaf

von Dominik Karl Lagler
aus Zwettl

München 2023

Aus dem Veterinärwissenschaftlichen Department der Tierärztlichen
Fakultät der Ludwig-Maximilians-Universität München

Lehrstuhl für Molekulare Tierzucht und Biotechnologie

Arbeit angefertigt unter der Leitung von: Priv.-Doz. Dr. Ivica Medugorac

Mitbetreuung durch Dr. Maulik Upadhyay

Gedruckt mit Genehmigung der Tierärztlichen Fakultät
der Ludwig-Maximilians-Universität München

Dekan: Univ.-Prof. Dr. Reinhard K. Straubinger, Ph.D.

Berichterstatter: Priv.-Doz. Dr. Ivica Međugorac

Korreferent: Univ.-Prof. Dr. Andrea Meyer-Lindenberg
Univ.-Prof. Dr. Joris Peters
Univ.-Prof. Dr. Nicolai T. Siegel
Priv.-Doz. Dr. Andrea Stockmaier-Didier

Tag der Promotion: 11. Februar 2023

Meiner Familie

INHALTSVERZEICHNIS

ABKÜRZUNGSVERZEICHNIS	X
ABBILDUNGSVERZEICHNIS	XI
TABELLENVERZEICHNIS	XII
ANHANGSÜBERSICHT	XIII
I. EINLEITUNG	1
II. LITERATURÜBERSICHT	3
1. Die Geschichte des heutigen Hausschafes	3
2. Systematik und Einteilung	4
3. Rassebeschreibung Merinolandschaf	5
4. Der Schwanz der Schafe	6
4.1. Allgemeine Informationen über den Schwanz.....	6
4.2. Gründe für eine mögliche Langschwanzigkeit	7
5. Nachteile der Langschwanzigkeit	8
5.1. Caudophagie, Anfressen der Schwänze	9
6. Methoden des Kupierens	9
7. Unterschiedliche nationale Bestimmungen bezüglich des Schwanzkupierens bei Schafen	10
7.1. Beispiele	11
8. Nachteile des Schwanzkupierens	12
9. Zucht als Alternative des Kupierens	13
9.1. Hybridisierung und Introgression	13
9.2. Zucht innerhalb einer Rasse	15
9.3. Bisherige züchterische Misserfolge beim Schaf.....	15
10. Genetische Kartierung potentieller Kandidatengene	16
10.1. Einteilung phänotypischer Merkmale.....	16
10.2. Genkarten	17
10.3. Genetische Marker.....	17
10.3.1. Restriktionsfragment-Längenpolymorphismen, RFLPs.....	18

10.3.2.	Mikrosatelliten, SSLPs	19
10.3.3.	Einzelbasenpolymorphismen, SNPs	19
10.4.	Kopplung (Linkage), Kopplungsgleichgewicht und Kopplungsungleichgewicht.....	22
10.5.	Kopplungskartierung, LM.....	23
10.6.	Kopplungsungleichgewichtskartierung, LDM.....	24
10.7.	Kopplungsungleichgewichts- und Kopplungs-Kartierungen (<i>linkage disequilibrium and linkage mapping, cLDLA</i>).....	25
10.8.	Genomweite Assoziationsstudien GWAS.....	26
11.	Selektive Genotypisierung und Phänotypisierung	26
III.	MATERIAL UND METHODEN	28
1.	Probenahmen und Phänotypisierung von 362 Lämmern	28
2.	Durchführung der Genotypisierung	29
2.1.	Herausfiltern nicht informativer Marker	30
3.	Haplotypisierung und Imputation	32
4.	Korrektur auf Populationsstratifikation und familiäre Verwandtschaftsbeziehungen	33
5.	Erstellung einer Diplotypen-Verwandtschaftsmatrix anhand der Locus IBD	35
6.	Varianzkomponentenanalyse.....	36
7.	Likelihood Ratio Teststatistik.....	37
8.	Weitere cLDLA Modelle.....	38
9.	Genomweite Assoziationsstudie (GWAS)	39
10.	Auswahl der zu sequenzierenden Lämmer.....	40
11.	DNA-Extrahierung, Sequenzierung und Analysen der Sequenzen	41
12.	Validierung von Kandidatenmutationen mittels gezielter Genotypisierung	47
12.1.	Validierung des Kandidaten-SNPs mittels PCR	47
12.2.	Validierung der Kandidaten-Insertion mittels PCR und Sanger- Sequenzierung.....	48

13.	Resequenzierung unserer Kandidatenmutationen mittels Sanger Sequencing	50
IV.	ERGEBNISSE	51
1.	Initiale <i>cLDLA</i>	51
2.	Initiale <i>MLMA</i>	54
3.	Auswahl der Lämmer für die Target Capture Sequenzierung	57
4.	Sequenzierungsergebnisse und Detektion möglicher Kandidatenmutationen	60
4.1.	Nachweis von Punktmutationen und kurzen Indels	60
4.2.	Vergleich der SNP-Mutation mit Sequenzen verschiedenster Rassen und Spezies.....	60
4.3.	Nachweis von SV	61
5.	Genotypisierung der plausibelsten Kandidaten bei allen 362 Lämmern und erneute Kartierung der Schwanzlänge.....	63
6.	<i>MLMA</i> und <i>cLDLA</i> mit Kandidatenmutationen als zusätzlichen Marker bzw. fixen Effekt	65
V.	DISKUSSION.....	67
1.	Gene, deren eine mögliche Mitbeteiligung an der Schwanzlänge nachgesagt werden.....	68
2.	Ermittlung von Kandidatengen.....	71
3.	Ermittlung von Kandidatenmutationen.....	71
4.	Vergleich der Kartierungsmethoden	73
5.	<i>HOXB13</i> als schlüssigstes Kandidatengen	75
6.	Ausblick.....	77
VI.	ZUSAMMENFASSUNG	80
VII.	SUMMARY	82
VIII.	LITERATURVERZEICHNIS.....	84
IX.	ANHANG	102

X.	DANKSAGUNG	105
-----------	-------------------------	------------

ABKÜRZUNGSVERZEICHNIS

μ	Gesamtmittelwert	LRT	Likelihood ratio test
AS	Aminosäuren	LRTmax	LRT-Werte höher als die Signifikanzschwelle
BAM	Binary Alignment Map		
bp	Basenpaare	MAF	Minor Allel Frequency
BW	Körpergewicht	Mb	Megabasenpaar
BWA	BURROWS-WHEELER ALIGNER	MLMA	Mixed linear model association
cLDLA	Combined linkage disequilibrium and linkage mapping	NCBI	National Center for Biotechnology Information
cM	centi-Morgan	OAR	Schafchromosom
DipE	Diploeffekt(en)	ONT	Oxford Nanopore Technologie
dNTPs	Desoxyribonukleosidtriphosphate	PacBio	Pacific Biosciences
DRM	diplo relationship matrix	PCR	Polymerase chain reaction
GATK	Genome Analysis Toolkit	QTLs	Quantitative Trait Loci
GREML	Genomic-Relatedness-Based restricted maximum-likelihood	RFLPs	Restriction fragment length polymorphisms
GRM	additive genetic relationship matrices	SAM	Sequence Alignment Map
GWAS	Genomweite Assoziationsstudien	SNPs	Single nucleotide polymorphisms
IBD	Identity by descent	SRA	Sequence read archive
kb	kilobasen	SSLPs	Simple sequence length polymorphisms
KI	Konfidenzintervall	SV	Strukturelle(r) Variant(en)
LD	Linkage disequilibrium	UAR	unified additive relationships
LDM	Linkage disequilibrium mapping	VNTRs	Number of Tandem Repeats
LE	Linkage equilibrium	WGS	Whole-Genome-Sequencing
LM	Linkage mapping	WH	Widerristhöhe
LOD	logarithm of the odds		

ABBILDUNGSVERZEICHNIS

Abbildung 1: Natürliche Schwanzlängenvariation innerhalb der Merinolandschafrasse.	7
Abbildung 2: Roter Gummiring im distalen Schwanzdrittel.	10
Abbildung 3: OvineSNP50 BeadChip.	20
Abbildung 4: Drei Sonden eines Bead-Chips.	21
Abbildung 5: Jene Drei Sondenoligonukleotide wie in Abbildung 4 zu sehen.	21
Abbildung 6: Verteilung der selektiv beprobten Lämmer.	27
Abbildung 7: Basensequenz der Region 37,290,221-37,290,657 auf OAR11 (Oar_v4.0).....	48
Abbildung 8: Basensequenz der Region 37,524,721-37,525,557 auf OAR11 (ARS-UI_Ramb_v2.0).	49
Abbildung 9: LRT-Werte auf OAR11.....	51
Abbildung 10: LRT-Werte auf OAR2.....	52
Abbildung 11: Ergebnisse der durchgeführten gemischten linearen Analysen, dargestellt als Manhattan-Plots.....	55
Abbildung 12: Ergebnisse der durchgeführten gemischten linearen Analysen, dargestellt als Zoom-Plots für die Region 35 bis 40 Mb für OAR11.	56
Abbildung 13: Verteilung der sequenzierten Lämmer anhand der Schwanzlänge und des Diplotypeffektes auf Chromosom 2.	57
Abbildung 14: Verteilung der sequenzierten Lämmer anhand der Schwanzlänge und des Diplotypeffektes auf Chromosom 11.	58
Abbildung 15: Abgleich der Anfangs-Aminosäuresequenz von HOXB13 verschiedener Säugetiere.....	61
Abbildung 16: Sequenzausschnitte von OAR11.	63
Abbildung 17: LRT- Werte auf OAR11 im Modell cDLA3.....	66

TABELLENVERZEICHNIS

Tabelle 1: Einteilungsmöglichkeiten domestizierter Hausschafe.	4
Tabelle 2: Anzahl und Einteilung der beprobten Lämmer.	28
Tabelle 3: Verwendete Geräte deren Hersteller und Verwendungszweck.	29
Tabelle 4: Verwendete Programme und Softwarepakete sowie deren Verwendungszweck und Quelle.	31
Tabelle 5: Deskriptive Statistik über alle notierten Merkmale der 362 Lämmer.	37
Tabelle 6: Gemischte lineare Modelle.	39
Tabelle 7: Informationen zu den für die Mutationsdetektion verwendeten Schafen.	44
Tabelle 8: Die 19 herausgefilterten kleinen Mutationen nach Anwendung der im Abschnitt III.11 beschriebenen Filterkriterien.	46
Tabelle 9: Getestete Primerpaare.	49
Tabelle 10: Gene im Konfidenzintervall auf OAR2.	53
Tabelle 11: Gene im Konfidenzintervall auf OAR11.	53
Tabelle 12: Regressionsanalyse für OAR11.	59
Tabelle 13: Regressionsanalyse für OAR2.	59
Tabelle 14: Verteilung der Genotypen in den verschiedenen Gruppen.	64
Tabelle 15: Deskriptive Statistik der Dipolypeneffektanalyse.	64

ANHANGSÜBERSICHT

Anhang 1: Genotypfrequenzen von rs413316737 in verschiedenen Rassen mit verschiedenen Schwanzphänotypen.	102
--	------------

I. EINLEITUNG

Seit vielen Jahren werden landwirtschaftliche Nutztiere für kostengünstige Haltungsbedingungen zurechtgestutzt. Bekannte Beispiele sind das Enthornen von Kälbern, das Schnabelkürzen von Nutzgeflügel, das Abschleifen der Eckzähne bei Ferkeln oder das Kupieren der Schwänze bei Schafen und Schweinen. In der heutigen Gesellschaft, die mehr und mehr ein moderat-biozentrisches Weltbild fordert, stehen solche Praktiken unter heftiger Kritik. So fordern viele Tierschützer die Haltungsbedingungen der Nutztiere artgerechter zu gestalten, etwa durch ein erhöhtes Angebot von Platz, Bewegung oder Beschäftigung, anstatt diese schmerzhaft für industrielle Stallungen zu optimieren. Auch eine genetische Selektion der Tiere auf ein äußeres Erscheinungsbild, welches besser für heutige Haltungs- und Nutzungssysteme geeignet ist, wird von einigen Tierschutzorganisationen kritisiert, wie die Ablehnung der Verwendung von genetisch hornlosen Bullen zeigt. Zwar werden bei diesen Tieren routinemäßige operative Eingriffe reduziert, allerdings erfüllen die züchterisch veränderten Merkmale häufig ihre natürlichen Funktionen nicht oder zumindest schlechter als das ursprüngliche Merkmal, wie im Falle der Hörner der Kommunikation (KWADE, 2013). Das Kupieren der Schwänze bei Schafen nimmt hierbei eine Sonderstellung ein. Viele Studien belegen das auch Lämmer durch diesen Eingriff Schmerzen erleiden und in Folge ein verändertes Verhalten zeigen (KUHNLE et al., 2019; SMALL et al., 2018). Allerdings wird durch das Kupieren des Schwanzes ein ursprünglicher Zustand erreicht, da sämtliche Wildschafarten und somit auch der vermutete Vorfahre der heutigen domestizierten Hausschafe, das Asiatische Mouflon (*Ovis gmelini*) von Geburt an kurze Schwänze besitzen (SCOBIE und O'CONNELL, 2002). Der lange Schwanz kam erst durch menschliche Selektion zustande und ist in der heutigen Schafhaltung unerwünscht, hauptsächlich wegen der Durchtränkung und Verschmutzung durch Urin und Kot und die dadurch entstehende Begünstigung einer Myiasis, den Befall mit Fliegenlarven (Maden) (SCOBIE und O'CONNELL, 2002). Nicht alle domestizierten Schafe sind langschwänzig. So gibt es viele alte Schafrassen mit hoher Fruchtbarkeit und Vitalität, die von Natur aus kurze Schwänze besitzen, wie zum Beispiel das Romanov und das Finnschaf (MASON, 1991). Interessanterweise treten sogar

innerhalb einiger Rassen, wie beim Merinolandschaf, Variationen in der Schwanzlänge auf, sodass einige Vertreter dieser Rassen natürlich kurze Schwänze aufweisen, auch JAMES et al. (1990) beschreibt diesen Umstand in einer Merinoherde. Bei diesen Schafen ist die Indikation zur chirurgischen Entfernung eines Teils des Schwanzes weniger stark gegeben oder gar nicht mehr vorhanden.

Diese Arbeit hatte daher zum Ziel mittels genetischer Feinkartierung und anschließender Kandidatengenanalyse die für die Schwanzlänge beim Merinolandschaf verantwortlichen Gene zu lokalisieren und möglichst die kausalen Varianten zu identifizieren. Die Ergebnisse dieser Forschung sollten eine präzisere Zucht auf Kurzschwanzigkeit ermöglichen und einen großen Beitrag zum Tierschutz leisten. Dies sollte idealerweise durch die Anreicherung ursprünglicher genetischer Varianten erreicht werden, die natürlich und einheitlich bei sämtlichen Wildtirschafen und alten Kurzschwanzschafen vorkommen. Moralische Bedenken sind dabei nicht zu erwarten, da beispielsweise im Gegensatz zur Zucht auf Hornlosigkeit kein für diese Spezies neuer Phänotyp künstlich selektiert wird, sondern ein Zustand wiederhergestellt wird, der durch menschliche Intervention erst herbeigerufen wurde. Diese ethisch vertretbare Zucht macht das routinemäßige Kürzen des Schwanzes überflüssig, dadurch werden dem Tier nicht nur Leid, Schmerzen und Schäden erspart, sondern dem Landwirt fallen auch erleichterte Haltungsbedingungen in die Hände.

Die Ergebnisse dieser Forschungsarbeit wurden in zusammengefasster Form in dem Journal „Communications Biology“ unter dem Titel „Fine-mapping and identification of candidate causal genes for tail length in the Merinolandschaf breed“ am 04.04.2022 eingereicht und am 06.09.2022 veröffentlicht.

II. LITERATURÜBERSICHT

1. Die Geschichte des heutigen Hausschafes

Vor ca. 10.000-12.000 Jahren begannen sesshaft gewordene Gemeinschaften mit der Domestikation des Hausschafes (*Ovis aries*). Die Domestikation geschah vermutlich in mehreren Anläufen in einem Zeitraum längerer menschlicher Interaktionen im Gebiet, welches sich von Nordzagros bis nach Südostanatolien erstreckt (nördlicher Fruchtbare Halbmond). Neben der Haltung von Schafen wurden auch Ziegen, Rinder und Schweine gehalten, was schließlich auch zu deren Domestikation führte. Zusätzlich zur Domestikation dieser Tiere bauten diese frühneolithisch sesshaften Gemeinschaften bereits Getreide und Hülsenfrüchte an. Durch die stetige Erweiterung gewonnener Erkenntnisse in der Viehhaltung und im Ackerbau wurde vor ca. 9.500 Jahren in vielen Teilen Südwestasiens und Zyperns die Selbstversorgung durch Nahrungssuche mit Nahrungsproduktion ersetzt. Auf diesem Wege wurden die Grundlagen für die Entwicklung nahrungsproduzierender Kulturen im neolithischen Europa, Nordafrika und Zentralasien gelegt. (ARBUCKLE et al., 2014; MEADOW, 1998; PETERS et al., 2005; PETERS et al., 1999; TANNO und WILLCOX, 2006; ZEDER, 2008). Als höchstwahrscheinlicher Vorfahre des heutigen Hausschafes wird das Asiatische Mouflon (*Ovis orientalis*) genannt, welches wie alle Wildschafarten zu den kurzschwänzigen Haarschafen zählt. Sämtliche Merkmale der Wildschafe wie zum Beispiel der relativ zu anderen Tierarten kleine Körperbau, die Neigung zur Bildung von Herden, die Fähigkeit auf mageren Böden zu überleben und deren Anpassungsfähigkeit haben zu deren Domestikation und weltweiten Verbreitung beigetragen. Aufgrund der Vielfalt der klimatischen Bedingungen und des Geländes kam es zu der Entstehung von mittlerweile 1.146 Schafrassen weltweit (FAO, 2020). SHERRATT (1983) deutet darauf hin, dass die ersten domestizierten Tiere zur Fleischgewinnung dienten. Erst lange Zeit später begann deren sekundäre Verwendung als Erzeuger von Milch und Wolle oder Transport- und Arbeitsmittel (SHERRATT, 1983). Durch die langfristige Zucht auf Feinwollfasern entstand die heutzutage ökonomisch wichtigste Rasse, das Merino, welches weltweit verbreitet und oft mit heimischen Rassen gekreuzt wurde. So entstand zum Beispiel das Merinolandschaf. Alle

Merinos sind charakterisiert durch lange Schwänze und Feinwolle, was häufig zur Annahme führte, dass diese beiden Phänotypen genetisch gekoppelt oder das Ergebnis desselben künstlichen Selektors sind (SHELTON, 1977). Diese Annahmen konnten bis dato noch nicht nachgewiesen werden.

2. Systematik und Einteilung

Schafe gehören zur Ordnung der Paarhufer (*Artiodactyla*), der Unterordnung der Wiederkäuer (*Ruminantia*) und sind Mitglieder der Familie der Hornträger (*Bovidae*), wo sie sich gemeinsam mit den Ziegen den Tribus der Ziegenartigen (*Caprinae*) teilen (z.B.: BIZELIS und KOUTSOULI, 2021).

Mittlerweile gibt es weltweit rund 1,5 Milliarden Schafe (z.B.: BIZELIS und KOUTSOULI, 2021), dessen Rassen sich anhand verschiedener Charakteristika in Kategorien einteilen lassen (siehe **Tabelle 1**):

Tabelle 1: Einteilungsmöglichkeiten domestizierter Hausschafe.

Die Tabelle zeigt welches Charakteristikum in verschiedene Kategorien eingeteilt werden kann und zeigt Rassen als Beispiele der jeweiligen Kategorie.

Charakteristikum	Kategorie	Beispiele
Typ-Textur des Vlieses	Haarschafe	Wiltshire Horn, Kamerunschaf, Dorperschaf
	Grobwollschafe	Heidschnucke
	Mischwollschafe	Dorper, Waldschaf, Bentheimer Landschaf
	Feinwollschafe	Merino
Schwanzlänge und -gestaltung	Fettsteißschaf	Alai, Hissar
	Fettschwanzschafe	Awassi, Karakul
	Langschwanzschafe	Bergschaf, Merinolandschaf
	Kurzschwanzschafe	Finnschaf
Nutzungsschwerpunkt	Wollschafe	Merino
	Fleischschafe	Merinolandschaf, Merinofleischschaf, Texel, Charollais
	Milchschafe	Ostfriesisches Milchschaf, Lacaune
	Landschafe	Bentheimer Landschaf, Heidschnucke

3. Rassebeschreibung Merinolandschaf

Es handelt sich um mittelgroße bis rahmige Schafe, welche neben einem rassetypischen Wollschopf breite und leicht hängende Ohren auf dem keilförmigen und langen Kopf tragen. Auf die breite Brust folgt eine lange Mittelhand, mit straffem breiten Rücken und guter Rippenwölbung, bei angemessener Flankentiefe. Hinten zeigen sich ein breites Becken und gut befleischte Innen- und Außenkeulen. Schafe mit Hautfalten werden von der Zucht ausgeschlossen. Ebenfalls unerwünscht sind Pigmentflecken (pigmentierte Augenlieder werden toleriert) und Hörner (BAYERISCHE HERDBUCHGESELLSCHAFT FÜR SCHAFZUCHT E.V., 2021). Die weiße Merinowolle zeigt Artcharakter mit einem ausgeglichenen Sortiment von AB-AB/B 25 - 28 μ m. Aufgrund ihrer Widerstandsfähigkeit, Anpassungsfähigkeit, Weide-, Marsch- und Pferchfähigkeit eignen sie sich gut für die Landschaftspflege, zur Koppelhaltung aber auch zur Stallhaltung. Weitere Vorteile sind ein asaisonaler Brunstzyklus, ausgezeichnete Fruchtbarkeit, hervorragende Säugeleistung und eine gute Fleischleistung. Die Zuchtreife erlangen sie mit 1 - 1.5 Jahren (BEHRENS et al., 1983).

Merinolandschafe (teilweise auch Württemberger Schaf genannt) entstanden durch die Einkreuzung von Merinos in heimische Landschaftsrassen in Süd- und Mitteldeutschland (wie das Niederrheinische Marschschaf) im 19. Jahrhundert. Schwerpunkt der Zucht wird heute auf wirtschaftliche Lammfleischproduktion gesetzt aber auch Aspekte der Landschaftspflege sind in den Zuchtzielen definiert (BAYERISCHE HERDBUCHGESELLSCHAFT FÜR SCHAFZUCHT E.V., 2021).

Die durchschnittlichen täglichen Zunahmen liegen bei Mastlämmern zwischen 400 - 500 g, wobei Einlinge mit ca. fünf kg und Mehrlinge mit ca. vier kg geboren werden. Das Schlachtendgewicht wird nach ca. drei bis vier Monaten erreicht und beträgt ca. 45 - 48 kg, bei einer Schlachtausbeute von 48 - 50 %. Altböcke erreichen ein Körpergewicht (BW) von 120 - 160 kg, Mutterschafe von 75 - 115 kg (BAYERISCHE HERDBUCHGESELLSCHAFT FÜR SCHAFZUCHT E.V., 2021). Laut der Bayerischen Landesanstalt für Landwirtschaft stellt das Merinolandschaf die ökonomisch wichtigste Rasse in Bayern dar.

4. Der Schwanz der Schafe

4.1. Allgemeine Informationen über den Schwanz

Der Schwanz von Schafen und allen anderen Wirbeltieren entsteht aus einer postanaln Verlängerung des Embryos. Diese Verlängerung ist die so genannte Schwanzknospe. Bildung, Wachstum und Differenzierung dieser Struktur stehen in engem Zusammenhang mit der Aktivität einer Gruppe von Zellen, die von den axialen Vorläuferzellen abstammen. Diese bilden im Laufe der Embryonalentwicklung das Rückenmark und das Muskel-Skelett-Gebilde des Rumpfes. Die diversen Funktionen des Schwanzes sind sehr vielfältig, was dazu führt, dass sich (nicht nur bei Schafen) eine Fülle an Phänotypen entwickelt hat. Die wohl wichtigste Funktion ist eine mechanische und dient der besseren Fortbewegung durch Ausbalancierung. Beispiele nichtmechanischer Funktionen sind die der Thermoregulation, die als Kommunikationsinstrument (MALLO, 2020) und die zur Insektenabwehr (MATHERNE et al., 2018). Die Rolle der Fettschwänze bei einigen Schafrassen nimmt dabei eine Sonderfunktion ein. Die wichtigste Funktion des Fettschwanzes ist die eines Energiereservoirs für raue und schwierige Umweltbedingungen (wie zum Beispiel Nahrungsmittelknappheit, Dürre oder Kälte), ähnlich den Höckern bei Kamelen. Vorteile dieses Energiespeichers wurden auch direkt von den Haltern der Fettschwanzschafe genutzt und zwar als Konservierungsmittel für Fleisch und als traditionelle Quelle für Speisefett in Zeiten von Dürre und Hunger (MORADI et al., 2012). Schafschwänze unterscheiden sich deutlich zwischen den verschiedenen Schafrassen weltweit. So deutlich, dass sogar eine eigene Einteilung der Schafe nach Schwanzphänotypen möglich ist, wie in **Tabelle 1** zu sehen ist. Auch innerhalb einiger Rassen ist eine Variation dieses Merkmals gegeben, wie in **Abbildung 1** anhand Merinolandschafvlämmer zu sehen ist. Der Schwanz besteht aus einer unterschiedlichen Anzahl an Schwanzwirbeln, beim Schaf sind es in der Regel 16 – 24. Allerdings kann sich die Anzahl bei stummelschwänzigen oder schwanzlosen Rassen bis auf 3 reduzieren. Der Wirbelkanal endet in den ersten Schwanzwirbeln. Je weiter schwanzspitzenwärts sich die Wirbel befinden, desto mehr verlieren sie ihre typischen charakteristischen Merkmale wie z.B. die Fortsätze. Am Ende der Wirbelsäule bestehen sie nur noch aus den Wirbelkörpern, welche sich mehr und mehr verjüngen. Caudal des Lendenbereiches verjüngt sich das Rückenmark bis nur noch ein dünner Endfaden vorzufinden ist, der *Filum*

terminale. Das Nervengewebe wird dabei schwanzspitzenwärts durch Bindegewebe ersetzt, welches bis zu den ersten Schwanzwirbeln zurückverfolgt werden kann, wo es anschließend mit dem Periost der Wirbelknochen verwächst. Die Wurzeln der hinteren Lenden-, Kreuz- und Schwanznerven verlaufen aufgrund des „*Ascensus medullae spinalis*“ nach caudal und formen so das typische Bild der *Cauda equina* (NICKEL et al., 2003).

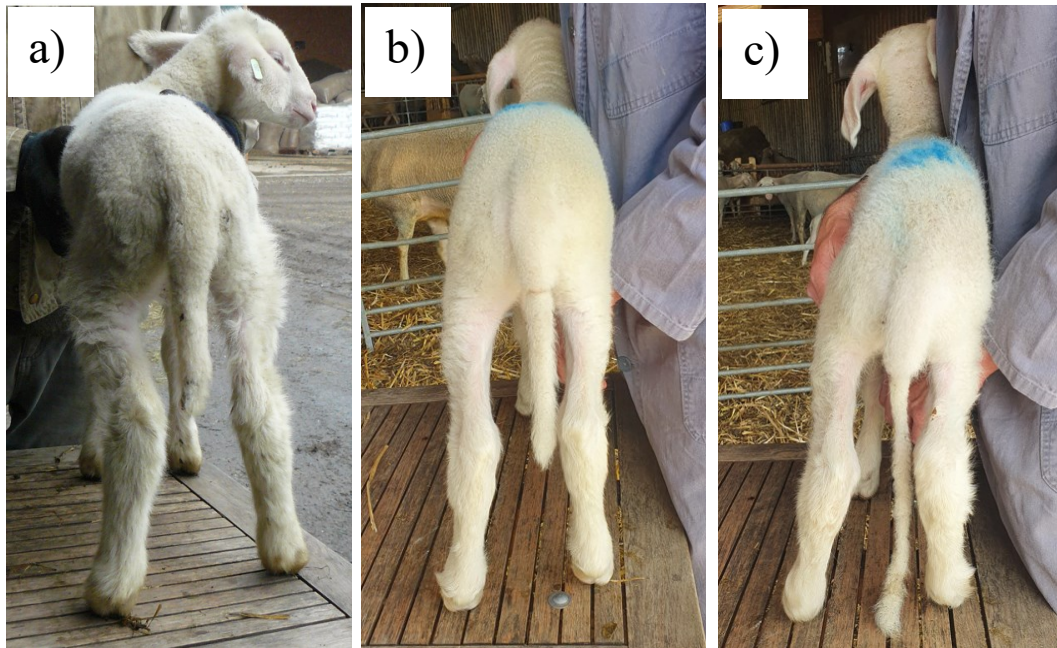


Abbildung 1: Natürliche Schwanzlängenvariation innerhalb der Merinolandschafrasse.

Abgebildete Schafe sind reinrassige Merinolandschafe und ca. 2 Wochen alt. a) Lamm mit einem natürlich kurzen Schwanz; b) Lamm mit einem kurzen bis mittellangen Schwanz. Diesem Tier wurde wenige Tage vor der Aufnahme des Fotos ein Gummiring an der oberen Schwanzhälfte zum Kupieren angelegt; c) Lamm mit einem langem bis zum Boden reichenden Schwanz. Auch diesem Lamm wurde einige Tage vor der Aufnahme ein Gummiring im proximalen Drittel des Schwanzes angelegt. Hier ist bereits der Unterschied zwischen dem vitalen proximalen Schwanzanteil und dem distalen, durch Ischämie nekrotisch werdenden Schwanzanteil, deutlich zu sehen.

4.2. Gründe für eine mögliche Langschwanzigkeit

Der in vielen modernen Hausschaf-Rassen vorhandene lange Schwanz entstand nach der Domestikation etwa im selben Zeitraum in der auch von kurzschwänzigen Haarschafen auf Wollertrag gezüchtet wurde (LYNCH et al., 1992). Dies führte, wie bereits erwähnt, zu der Annahme einer möglichen genetischen Kopplung zwischen der Schwanzlänge und dem Wollertrag (SHELTON, 1977). Dies könnte darauf zurückzuführen sein, dass die Gene, die die Schwanzlänge regulieren, in

einer genetischen Kopplung oder pleiotropen Beziehung zu Genen stehen, die für die Bewollung eine Rolle spielen. Andere Gründe für die heutige Langschwänzigkeit könnten die künstliche Zucht aufgrund eines höheren Körpergewichts bzw. Wollertrages sein. Eher unwahrscheinlich klingt die Theorie einer natürlichen Selektion aufgrund adaptiver Vorteile (SHELTON, 1977), wie zum Beispiel der erhöhte Schutz des Euters. Hinweise auf eine artifizielle Selektion befinden sich auch in den Bänden von Columella, in denen er die römische Landwirtschaft um 35 nach Christus beschreibt. Darin steht unter anderem, dass der Schwanz sehr lang und dicht behaart sein muss. Interessant ist auch, dass überlieferte Bilder darauf hin deuten, dass Schwänze im antiken Rom noch nicht kupiert wurden (CHIFFLARD und REINHARDT, 2015)

5. Nachteile der Langschwänzigkeit

Aufgrund der Lokalisation der Vagina überwiegen vor allem bei weiblichen ausgewachsenen Schafen die Nachteile eines langen bewollten Schwanzes die Vorteile. Neben durchfallbedingten Verschmutzungen begünstigen bei weiblichen Schafen auch Verklebungen durch Urin und Fruchtwasser einen Fliegenmadenbefall (Myiasis) in hinteren Körperteilen. Aber auch Tetanus, Geburtsrauschbrand und andere Infektionen können in Folge solcher Verunreinigungen auftreten (BEHRENS et al., 1983). Einzelne Schafzüchter berichten auch, dass Lämmern das Saugen und den Landwirten routinemäßige Kontrollen zur Prophylaxe oder Früherkennung einer potentiellen Mastitis durch lange bewollte Schwänze erschwert werden. Selbiges gilt für den Deckakt (MAX, 1983). Außerdem werden die Schlacht- und Schurhygiene negativ beeinflusst (STEFFENS et al., 2021). Laut STROBEL et al. (2018) treten Schwanzverletzungen (durch Drauftreten anderer Herdenmitglieder, Picken von Rabenvögel, übereifriges postpartales Trockenlecken durch Muttertiere, siehe Abschnitt II.5.1) häufiger bei Individuen mit langen Schwänzen auf als bei kurzschwänzigen Artgenossen. Bei Zuchtböcken soll die Bemuskelung der Keulen bei Fleischschafzrassen und Merinoböcken durch lange Schwänze nicht gut zur Geltung kommen, weshalb dieser bei Züchtern aus optischen Gründen kupiert wird (KUHNLE et al.). Auch der lange Fettschwanz, wie er zum Beispiel bei Karakulschafen auftritt, hat viele Nachteile, die nicht den modernen Anforderungen und Zuchtpraktiken entsprechen. Vor allem die negative Beeinflussung beim Paarungsakt und die massive Störung bei der Fortbewegung sind hier zu nennen

(KALDS et al., 2021).

5.1. Caudophagie, Anfressen der Schwänze

Vor allem bei langschwänzigen Rassen kommt es gelegentlich zum An- bzw. Abfressen der Schwänze. Dieser Umstand wird in der Veterinärmedizin als *Caudophagie* bezeichnet. Dabei knabbern sowohl Mutterschafe als auch andere Schafe in der Herde die Schwanzspitze der Neonaten bzw. der noch in den Geburtswägen befindlichen Lämmer an. Diese Verhaltensstörung wurde bislang hauptsächlich bei Schafen in ganztägiger Stallhaltung beobachtet und ist vermutlich auf eine Mangelernährung oder Mangel an Abwechslung zurückzuführen. Diese „Übermutterung“ wird häufig beim Trockenlecken beobachtet und führt häufig zu den in Abschnitt II.8 gezeigten Problemen. Neben des Schwanzes sind im Rahmen dieser Verhaltensstörung oft auch Ohren (Otophagie) und Klauen (Onychophagie) betroffen (BEHRENS et al., 1983).

6. Methoden des Kupierens

Um die im Abschnitt II.5 genannten Nachteile zu umgehen ist es üblich langschwänzigen Schafen die Schwänze zu kürzen. In europäischen Breitengraden finden hierfür vor allem die blutige und unblutige Methode (siehe unten) Anwendung. Auf welcher Höhe der Schwanz abgetrennt wird, variiert von Region zu Region, es ist aber zu beachten, dass Anus und Vulva auf jeden Fall durch den verbliebenen Schwanzanteil verdeckt bleiben sollen. Außerdem ist der Umstand erwähnenswert, dass der Blutfluss Richtung Schwanzspitze abnimmt. Das hat bei schwanzansatznahen Kupieren stärkere Blutungen und Komplikationen zur Folge als bei schwanzspitzennahen, außerdem ist der Schwanzdurchmesser Richtung Ansatz größer als Richtung Spitze, wodurch die Wundheilung umso besser voranläuft, je weiter distal man kupiert (KUHNLE et al.).

Blutige Methode: Zwischen zwei Wirbel wird mit einem scharfen Messer der Schwanz abgetrennt oder mit einer Kupierzange abgequetscht. Dabei ist streng auf eine örtliche Rasur, Reinigung und Desinfektion zu achten, da sonst sekundäre Komplikationen wie Wundheilungsstörungen, aufsteigende Infektionen oder Tetanus begünstigt werden. Auch das Schieben der Schwanzhaut nach kranial wird vor dem Messerschnitt empfohlen, damit sie sich anschließend wieder zumindest teilweise über die Wundfläche stülpt (MAX, 1983). Anschließend wird die Wunde desinfiziert und das Lamm möglichst sauber gehalten.

Unblutige Methode: Dabei wird durch das Aufsetzen eines elastischen Gummiringes (siehe **Abbildung 2**) mittels eines Elastrators eine Ischämie Richtung Schwanzspitze verursacht, wodurch dieser nekrotisch wird (siehe Abbildung 1c) und nach ca. 10 Tagen abfällt oder, wenn er nicht von alleine abfällt, manuell abgetrennt wird (MAX, 1983). Falls eine offene Wunde zurückbleibt soll diese wieder desinfiziert werden. Eine Desinfektion ist bei korrekt vernarbten Schwanzstumpf nicht notwendig.

Eine weitere Methode ist die des Mulesings, welche vor allem in Australien Anwendung findet. Dabei wird die wolltragende Haut im Steißbereich der Schafe auf schmerzhaft Art und Weise entfernt, so dass ein nackter Bereich ohne Falten entsteht. Dadurch steht weniger Wolle zur Verfügung, die mit Urin oder Kot verunreinigt werden kann. Außerdem wird das Scheren im Steißbereich erleichtert. Dadurch bleibt die Perianalregion trockener und weniger anfällig für Insekten (LEE und FISHER, 2007).



Abbildung 2: Roter Gummiring im distalen Schwanzdrittel.

Dieser führt zu einer Ischämie distal der Ansatzstelle, wodurch die Schwanzspitze nach ca. 10 Tagen abfällt.

7. Unterschiedliche nationale Bestimmungen bezüglich des Schwanzkupierens bei Schafen

In Europa sind gesetzliche Regelungen die den Tierschutz behandeln weitgehend in der Zuständigkeit der Mitgliedsstaaten. Als Anhaltspunkte dienen lediglich EU-

Richtlinien und Empfehlungen der europäischen Behörden für Lebensmittelsicherheit und des Europarates (HANNEMANN et al., 2017). Es soll hier auf die nationalen gesetzlichen Bestimmungen verwiesen werden die innerhalb der europäischen Union stark variieren. Nur die EU-Rechtsverordnungen für den ökologischen Landbau verbietet ein routinemäßiges Kupieren der Schwänze EU-weit. Aber auch bei dieser Verordnung können aus diversen Gründen Ausnahmen bei der zuständigen Behörde fallweise geltend gemacht werden, zum Beispiel wenn sie dazu dienen, die Gesundheit, das Wohlbefinden oder die Hygienebedingungen der Schafe zu verbessern (VERORDNUNG (EG) Nr. 889/2008 DER KOMMISSION, 2008). In der Regel wird ein Kupieren bei wenigen Tagen alten Lämmern vorgeschrieben, da man bei diesen ein geringeres Schmerzempfinden vermutet (STROBEL et al., 2018).

7.1. Beispiele

Skandinavische Länder: Kupieren generell verboten, in Norwegen und Schweden sind Ausnahmen möglich (STROBEL et al., 2018).

Niederlande: generell verboten, Ausnahmen für bestimmte Rassen (Suffolk, Clun Forest und Hampshire Down) vorhanden mit der Bedingung der Einführung eines effektiven Zuchtprogramms für kürzere Schwänze (BOHTE-WILHELMUS et al., 2010).

Deutschland: Kupieren mittels elastischen Gummiring erlaubt, wenn dieser zwischen 2 Wirbeln angebracht ist. Der Einsatz von Betäubungs- und Schmerzmittel ist nicht vorgeschrieben, jedoch dürfen die Lämmer den 8. Lebenstag noch nicht erreicht haben. Weitere Einzelfallregelungen zum Schutz des Tieres oder anderer Tiere etc. sind im deutschen Tierschutzgesetz ebenfalls geregelt (§ 6 Abs.1 Nr.3 in Verbindung mit § 5 Abs.3 Nr.3 und 4 des Tierschutzgesetzes).

Österreich: Kupieren nur mittels Gerät erlaubt, welches scharf schneidet und gleichzeitig verödet. Betroffene Lämmer dürfen nicht älter als 7 Tage sein und müssen einer wirksamen intra- und postoperativen Schmerzbehandlung unterzogen werden. Die Operation muss von einer sachkundigen Person oder von einem Tierarzt (dann mit Betäubung und postoperativer Schmerzausschaltung) durchgeführt werden. Es darf nicht mehr als ein Drittel des Schwanzes entfernt werden. Ausnahmen bilden weibliche Lämmer, falls eine vom Tierarzt bestätigte Notwendigkeit vorliegt und mit dem Tier gezüchtet werden soll. Hier darf

höchstens die Hälfte des Schwanzes entfernt werden (TIERHALTUNGSVERORDNUNG, 2017).

Der Ausschuss „Farm Animal Welfare Committee“ empfiehlt das Schwanzkupieren bei bis zu sieben Tage alten Lämmern mittels elastischem Gummiring durchzuführen und mit einem scharf schneidenden Gerät bei Lämmern zwischen der ersten und achten Lebenswochen (FAWC, 2008).

Weltweite Kupierverbote können als problematisch betrachtet werden, wenn der Eingriff im Einzelfall nach tierärztlicher Indikation geboten ist.

8. Nachteile des Schwanzkupierens

Mehrere Studien (zum Beispiel SMALL et al., 2018) haben gezeigt, dass Lämmer, welche kupiert oder einer Mulesing-Behandlung unterzogen wurden, Schmerzen und Leiden empfinden bzw. einer erhöhten Stresssituation ausgesetzt sind. Tierschützer und Verbraucher in aller Welt kritisieren daher diese Routinemaßnahme und fordern in einigen Fällen sogar ein Verbot dieser Eingriffe. Der Einsatz diverser Schmerzmittel und lokaler Betäubungsmittel sollen zwar in der Lage sein das Schmerzempfinden und das Stressausmaß nachvollziehbar zu reduzieren, die Restwirkung der durch das Mulesing/Kupieren verursachten Schmerzen und Beschwerden ist aber dennoch erkennbar (KUHNLE et al.; SMALL et al., 2018). Außerdem wurden auch weitere Nachteile beim Schwanzkupieren beobachtet, vor allem, wenn nur sehr kurze oder keine Schwanzstümpfe mehr übriggelassen wurden. LLOYD et al. (2016) identifizierten einen Zusammenhang zwischen der Länge des kupierten Schwanzes und bakteriellen Gelenkentzündungen. Sehr kurz kupierte Schwänze (mit bis zu zwei verbliebenen Schwanzwirbeln) sind ein erhöhter Risikofaktor als länger kupierte (mit mindestens 3 verbliebenen Wirbeln im Schwanzstumpf). Dabei wurde der Infektionserreger *Erysipelothrix rhusiopathiae* als häufigster Erreger der bakteriellen Gelenkentzündungen nachgewiesen. Aufgrund der kurzen Inspektionszeit konnten die Forscher am Schlachthof nicht herausfinden mit welchen Methoden die Schafe kupiert wurden (scharfes Messer, elastischer Ring, gasbeheiztes Messer etc.).

Untersuchungen von THOMAS et al. (2003) und WINDELS (1990) legen nahe, dass sehr kurz kupierte Schwänze eine Ursache für ein gehäuftes Auftreten von Vorfällen des Rektums sind, da die ersten Schwanzwirbeln mitunter als

Ansatzstellen der Analmuskulatur dienen (CHIFFLARD und REINHARDT, 2015). Es ist erwähnenswert, dass die Autoren zusätzlich einen Zusammenhang zwischen einer Haltung in Mastbetrieben mit hochkonzentriertem Futter und *Prolapsus recti* suggerieren.

SWAN et al. (1984) beschreiben, dass neben dem Alter auch die Schwanzlänge bzw. die Größe der Wundfläche nach dem Eingriff wichtige Determinanten für Plattenepithelkarzinome sind, welche vor allem bei Mutterschafen auftritt. Hauptsächlich betroffen ist die Vulva aber auch Anus und Schwanz zeigen häufig Läsionen, welche als Vorstufe des Plattenepithelkarzinomes betrachtet werden. Die Anzahl der Läsionen war signifikant größer bei Schafen, die einer radikalen Mulesing Operation unterzogen wurden als solche die eine moderate oder keine erhielten. Betroffene Tiere hatten auch signifikant kürzere Schwänze als nicht betroffene, falls sie kupiert wurden. Es wird auch suggeriert, dass die Häufigkeit des Auftretens dieser Krebsart reduziert werden könnte, indem man eine möglichst wenig invasive Mulesingtechnik anwendet bzw. einen längeren Schwanzstumpf nach dem Kupieren hinterlässt.

Durch das Anheben des Schwanzes während der Defäkation werden die Schwanzfalten gestrafft was dazu führt das der Kot und Urin vom Körper weggeleitet wird, wodurch die Wolle weniger mit Ausscheidungen kontaminiert wird. Bei extrem kurz kupierten Schwänzen werden die Schwanzfalten nicht mehr angezogen und es kommt eher zu Verunreinigungen. Was sich wiederum, obwohl die Schafe bereits kupiert sind, günstig für eine Myiasis auswirkt (RICHES, 1941) (RICHES, 1942).

Ein ausführlicherer Überblick über die verschiedenen Nachteile eines zu kurzen Kupierens des Schwanzes bzw. des Kupierens im Allgemeinen findet sich in dem Review von FISHER et al. (2004).

9. Zucht als Alternative des Kupierens

9.1. Hybridisierung und Introgression

Die Anhäufung von vorteilhaften Varianten in der reinrassigen Zucht ist nur dann möglich, wenn solche Varianten Teil des genetischen Pools der Rasse sind. Ist dies nicht der Fall, können neue genetische Allele durch die Kreuzung von Individuen verschiedener Rassen relativ schnell eingeführt, akkumuliert und fixiert werden.

Diese Methode wird als Hybridisierung bezeichnet und kann auf natürliche oder künstliche Weise erfolgen. In weiterer Folge kann durch Introgression, der gezielten Rückkreuzung zur Übertragung vorteilhafter Genvarianten, die genetische Umwandlung beschleunigt werden (ARNOLD, 1992). Durch die Einkreuzung kurzschwänziger (z.B.: Finnschaf) in langschwänzige Schafrassen (z.B.: Merino) ist es möglich, den Schwanz in zukünftigen Generationen züchterisch zu kürzen. Allele aus nordischen Kurzschwanzrassen sind hierbei sehr vielversprechend, da keine nachteiligen Auswirkungen auf die Lebensfähigkeit und Fruchtbarkeit zu befürchten sind (SCOBIE und O'CONNELL, 2002). OLTENACU und BOYLAN (1974) kreuzten Finnschafböcke mit Mutterschafen verschiedenster mittel- bis langschwänziger Schafrassen (Suffolk, Targhee und Kreuzungen zwischen Rambouillet, Border Leicester und Cheviot). Daraus wurden wiederum F₂-Lämmer und Lämmer aus den verschiedenen Rückkreuzungen gewonnen. Sie kamen zu der Erkenntnis, dass die Verteilung der Schwanzlängen in der Gruppe aus den F₁ × Finnschaf-Rückkreuzungen in Richtung der Verteilung der reinrassigen Finnschafen tendiert. Schafe aus Rückkreuzungen der F₁ mit mittel- bis langschwänzigen Rassen zeigten Schwanzlängenverteilungen in die Richtung der langschwänzigen Rassen. F₁- und F₂-Generation waren bei diesem Merkmal im Mittelwert und Verteilung sehr ähnlich. Weiters suggerierten sie aus deren Ergebnissen, dass Kurzschwanzigkeit dominant gegenüber Langschwanzigkeit ist. Außerdem vermuteten sie, dass die biologische Wirkung, der für dieses Merkmal verantwortlichen Gene, nicht einfach additiv ist, sondern weitaus komplexer. Dennoch zeigten deren Ergebnisse nicht nur eine hohe Heritabilität der Schwanzlängen (in Finnschaf 0.77; in den anderen Rassen 0.5), sondern lassen den Schluss zu, dass Introgression eine einfache Methode ist, die Länge des Schwanzes nachhaltig innerhalb weniger Generationen zu kürzen. Auch die Ergebnisse von SCOBIE und O'CONNELL (2002) deuten darauf hin, dass die Schwanzlänge von mehreren Genen beeinflusst wird und dass Introgression dennoch der unaufwändigste Weg sei, eine sogenannte „genetische Kupierung“ in langschwänzigen Rassen zu erreichen. Auf diese Erkenntnis kamen sie durch die Kreuzung von Cheviot- mit Finnschafen. Außerdem könnte durch diesen Weg auch die hohe Fruchtbarkeit der Finnschafe in andere Rassen eingekreuzt werden. Die Introgression nordischer Kurzschwanzallele wäre zwar eine vielversprechende Lösung, das Kupieren zu umgehen, stößt jedoch auf den Widerstand der Züchter. Diese Methode hätte nämlich zur Folge, dass auch andere rassetypische Merkmale

beeinflusst werden könnten, welche über lange Zeit mühsam angezüchtet wurden.

9.2. Zucht innerhalb einer Rasse

Durch die Reinzucht werden rassetypische Merkmale und Eigenschaften verbessert oder erhalten. Voraussetzung für neue Zuchtziele ist eine natürliche Variation in dem Merkmal welches züchterisch verändert werden soll. Im Falle der Merinolandschafe wäre das eine unterschiedliche Länge des Schwanzes innerhalb dieser Rasse, welche, wie in Abbildung 1 zu sehen ist, durchaus vorhanden ist. Im Jahre 1956 wurden bereits vermehrt kurzschwänzige Lämmer vom Herrn Don Gare von "Mt Glen View" in Südastralien beobachtet (JAMES et al., 1991). Da eine solche Variation auch in anderen Rassen gegeben ist, versuchten Züchter bereits durch eine reine merkmalsabhängige Selektion die Schwanzlänge zu reduzieren. Dabei kam es aber mitunter zu den im Abschnitt II.9.3 erwähnten Problemen.

9.3. Bisherige züchterische Misserfolge beim Schaf

CARTER (1976) hat bereits in den siebziger Jahren versucht, bei der Rasse Romney auf Kurzschwanzigkeit zu züchten. Diese Versuche führten nicht zu einer weit verbreiteten Zucht, da er eine verminderte Lebensfähigkeit der Lämmer beobachtete, die wahrscheinlich homozygot für Kurzschwanzvarianten waren. Es schien sehr wahrscheinlich, dass kurzschwänzige Allele letal, in unvollständiger Penetranz oder in Epistase mit letalen Allelen waren. Vergleichbare Versuche wurden etwa fünfzehn Jahre später von JAMES et al. (1991) bei der Merino-Rasse durchgeführt. Diese Versuche deuteten darauf hin, dass die Schwanzlänge bei Merinolämmern von nur wenigen interagierenden Genen bestimmt wird, wobei die Kurzschwanzigkeit höchstwahrscheinlich dominant ist. Die Beobachtung, dass kurzschwänzige Lämmer vermehrt verschiedene tödliche Defekte am Hinterende (z. B. *Atresia ani*) aufwiesen, verhinderte eine breitere Zucht auf Kurzschwanzigkeit.

Die Ergebnisse der beiden oben genannten Studien stehen in deutlichem Gegensatz zu der Tatsache, dass einige alte nordische Schafrassen kurze Schwänze und eine hohe Fruchtbarkeit und Vitalität aufweisen. Ebenso haben alle Wildschafe kurze Schwänze und weisen eine hohe Fitness auf. Folglich muss die Zucht auf Kurzschwänze ohne Fitnessprobleme möglich sein.

10. Genetische Kartierung potentieller Kandidatengene

10.1. Einteilung phänotypischer Merkmale

Für eine Zucht eignen sich nur erbliche Phänotypen. Der Phänotyp beschreibt das Erscheinungsbild eines Organismus, welcher durch eine oder mehrere Erbanlagen und durch Umwelteinflüsse bestimmt wird. Ein Phänotyp oder auch Merkmal wird als quantitativ oder qualitativ eingestuft (BERRY et al., 2011).

Qualitative erbliche Merkmale, auch diskrete Merkmale genannt, werden in der Regel von einem oder nur wenigen Genen beeinflusst. Umweltfaktoren haben keinen oder nur einen geringen Einfluss auf die Ausprägung des Merkmals. Die Eigenschaften können in Kategorien unterteilt werden, weshalb man sie auch kategorische Merkmale nennt. Sie folgen den Mendelschen Regeln (ARNOLD und ROLOFS, 2011). Beispiele sind die Fellfarbe, Blutgruppe oder monogene Erbkrankheiten.

Die Ausprägung von quantitativen oder auch kontinuierlichen Merkmalen geht auf die Wirkung von mehreren Genen zurück. Dabei weisen die meisten Gene kleinere unterschiedliche Effekte auf das Merkmal auf. Bei einigen quantitativen Merkmalen zeigen ein oder wenige Gene eine deutlich größere Wirkung. Diese werden oft als Hauptgene oder "*major genes*" bezeichnet. Auch Umwelteinflüsse spielen bei quantitativen Merkmalen eine wesentliche Rolle, da sie in der Lage sind die Ausprägung mehr oder weniger stark zu beeinflussen (ANDERSSON und GEORGES, 2004). Viele züchterisch wertvolle Merkmale, wie zum Beispiel die Milchleistung, tägliche Zunahme, Größe der Eier und die Widerristhöhe (WH) gehören dieser Kategorie an. Sie sind stets mess- oder zählbar (ARNOLD und ROLOFS, 2011). Aufgrund der Abhängigkeit von Umwelteinflüssen variieren diese kontinuierlichen Merkmale selbst innerhalb einer einzelnen Population oder sogar eines einzelnen Individuums (zeitlich, zum Beispiel Milchmenge). Im Gegensatz zu qualitativen Polymorphismen haben quantitative eine kontinuierliche Verteilung in Form einer Gauß'schen Normalverteilung (STEPHAN und HÖRGER, 2019). Durch das Zusammenwirken von Genen an verschiedenen Chromosomen ist es oft schwieriger die genetischen Hintergründe mittels der klassischen Vererbungslehre nachvollziehen zu können (ARNOLD und ROLOFS, 2011). Genloci welche durch ihre unterschiedlichen Varianten in der Lage sind quantitative Merkmale zu beeinflussen werden als „Quantitative Trait Loci“

(QTLs) bezeichnet (GELDERMANN, 1975)

10.2. Genkarten

Mittels molekulargenetischer Kartierungsmethoden wird die relative Position genetischer Marker und Loci, welche ein Merkmal beeinflussen, auf einem DNA-Molekül bestimmt. In weiterer Folge wird anschließend versucht ursächliche Gene zu identifizieren und ihre unterschiedlichen Varianten zu charakterisieren. Zwei molekulargenetische Kartierungsmethoden kommen hierbei häufig zum Einsatz. Es handelt sich zum einen um eine Kopplungskartierung (linkage mapping, LM) und zum anderen um eine Kopplungsungleichgewichts-Kartierungen (linkage disequilibrium mapping, LDM). Beide Kartierungsmethoden arbeiten mit sogenannten Markern und werden in verschiedenen praktischen Anwendungen auch kombiniert eingesetzt. Dabei kommt eine Markerkarte zum Einsatz, in welche der Abstand zweier Marker mittels verschiedener Einheiten gemessen wird, nämlich in Rekombinationsraten (bei genetischen Karten) und in physikalischen Einheiten (physikalische Karten). Bei der Abstandsmessung mittels Rekombinationsraten beobachtet man die relativen Häufigkeiten an Rekombinationen eines Chromosoms während des Crossing-overs pro Meiose (GELDERMANN, 2005). Die Umrechnung in centi-Morgan (cM) widerspiegelt die prozentuale Häufigkeit solcher Rekombinationen (HALDANE, 1919). Das heißt, bei einem Abstand von 1 cM liegt die Wahrscheinlichkeit einer Rekombination bei nur einem Prozent und daher liegen die beiden Gene relativ nahe beieinander. Misst man den Abstand mittels physikalischen Einheiten, verwendet man Chromosomenbänder oder die Anzahl dazwischenliegender Basenpaaren (bp), wodurch sich das Genom detaillierter analysieren lässt. Chromosomenbänder sind lichtmikroskopisch, während der Metaphase des Zellzykluses, zytochemisch erkennbar. Chromosomenkarten, die durch Bänderungsmuster abgeleitet wurden, nennt man zytogenetische Karten (GELDERMANN, 2005) und sind eher von historischer Bedeutung.

10.3. Genetische Marker

Wie bereits erwähnt arbeiten diese Kartierungsmethoden mit genetischen Markern, diese dienen als Bezugspunkte, da deren Chromosomenpositionen bereits bekannt sind. Marker sind nachweisbare Varianten im Genom, welche in der Regel keine

biologische Funktion aufweisen. Sie besitzen allerdings multiple Varianten, Allele genannt, die für die Kartierung von Loci mit biologische Funktion genutzt werden. Die Position des Markers auf dem Chromosom nennt man Locus. Nach gängiger Nomenklatur spricht man erst von einem Polymorphismus, wenn eine genetische Variante zu mindestens einem Prozent in einer Population auftritt. Es gibt verschiedene Arten von Markern, welche sich sowohl in nicht-codierende Regionen des Genoms als auch in Exons und Introns befinden und sich meistens, aber nicht immer, in der Basenpaarenlänge unterscheiden (DUBITZKY et al., 2013). Im Folgenden werden Einzelbasenpolymorphismen (single nucleotide polymorphisms, SNPs), Restriktionsfragment-Längenpolymorphismen (restriction fragment length polymorphisms, RFLPs) und Mikrosatelliten (simple sequence length polymorphisms, SSLPs) näher beschrieben.

10.3.1. Restriktionsfragment-Längenpolymorphismen, RFLPs

Die ersten Marker die für DNA-Analysen untersucht wurden, waren die RFLPs. Sie können relativ einfach, sicher und schnell nachgewiesen werden. Man vermutet etwa 10^5 dieser Marker in einem Säugetiergenom (BROWN, 2007). Bei RFLPs wird eine zuvor per Polymerase Kettenreaktion (polymerase chain reaction, PCR) amplifizierte Region im Genom, welche eine polymorphe Stelle enthält, mittels eines Restriktionsenzym verdaut. Restriktionsenzyme sind in der Lage DNA-Moleküle an bestimmten Erkennungssequenzen zu schneiden. Da aber die zuvor ausgewählten Restriktionsstellen polymorph sind, wird nur bei jenem Allel geschnitten, welches die restriktionsenzym-spezifische Erkennungssequenz aufweist. Allele mit einer anderen Sequenz an dieser Stelle werden nicht geschnitten und behalten ihre ursprüngliche Länge bei (BROWN, 2007). Anschließend werden die unterschiedlich langen Fragmente mittels Gelelektrophorese nachgewiesen (GELDERMANN, 2005). Für QTL-Kartierungen bei Tieren waren RFLPs in der Vergangenheit weniger nützlich als bei Pflanzen, da sie bei Tieren aufgrund des biallelischen Auftretens bei den meisten Individuen homozygot sind (WELLER, 2001). Zusätzlich wurden RFLPs durch die ebenfalls biallelischen SNPs, welche viel häufiger im Genom auftreten, aufgrund der schwierigeren, langsameren und umständlichen Genotypisierung abgelöst (BALL et al., 2010; KRUGLYAK, 1997). Die RFLPs spiegeln in den meisten Fällen nur einen Teil der SNPs wider, für die es ein geeignetes Restriktionsenzym gibt.

10.3.2. Mikrosatelliten, SSLPs

Bei SSLPs handelt es sich um Sequenzwiederholungen die zwischen Individuen unterschiedlich häufig vorkommen, man spricht in diesem Zusammenhang auch von Variable „Number of Tandem Repeats“ (VNTRs), zu denen auch Minisatelliten dazuzählen. Pro Locus bestehen Mikrosatelliten in der Regel aus bis zu 30 bp, können aber eine Gesamtlänge von bis zu 200 bp haben. Die Wiederholungssequenz (*Tandem Repeat*) ist meist kleiner als 4 bp, kann aber bis zu 10 bp enthalten. Im Vergleich zu SNPs, welche nur biallelisch auftreten, werden für VNTRs häufig mehr als zwei Allele beobachtet (GELDERMANN, 2005). Je mehr verschiedene Allele für einen Locus zu finden sind, desto mehr Informationen können für Kartierungsstudien gewonnen werden (BALL et al., 2010). Trotz der gleichmäßigen Verteilung im Genom und des hohen Vorkommens (alle 1000 bp) (GELDERMANN, 2005) wurden auch Mikrosatelliten weitgehend durch SNPs zur Kartierung von QTLs abgelöst. Gründe dafür sind unter anderem die leichtere Genotypisierung von SNPs (BALL et al., 2010; KRUGLYAK, 1997), die präziseren Ergebnisse (VIGNAL et al., 2002), die relativ hohe Mutationsrate von Mikrosatelliten (WEBER und WONG, 1993; WEISSENBACH et al., 1992) und der Effekt der Homoplasie (Allele, die in ihrer Größe identisch sind, sich aber durch eine Reihe von Mutationen unterscheiden) (GARZA und FREIMER, 1996). Da für die Darstellung und Auswertung von Mikrosatelliten in der Regel die mittels PCR amplifizierte Fragmentlängen verwendet werden, (GELDERMANN, 2005) können Homoplasien nicht erkannt werden.

10.3.3. Einzelbasenpolymorphismen, SNPs

Für die Durchführung verschiedener Ganzgenom-Studien wie der Identifizierung von QTLs, Bewertung der genetischen Leistung, Studien zum Kopplungsungleichgewicht und weitere, werden heutzutage überwiegend SNPs verwendet. Da die Markerdichte ein limitierender Faktor für die Auflösung der Kartierung ist (GEORGES, 2007), können QTLs durch die hohe Anzahl und gleichmäßigen Verteilung von SNPs im Genom sehr fein kartiert werden. Neben der hohen Anzahl ist die geringe Mutationsrate ein wesentlicher Vorteil gegenüber der geringen Stabilität von Mikrosatelliten. Diese beiden Eigenschaften verschaffen SNPs, trotz des nur biallelischen Auftretens, präzisere Kartierungsaufösungen als Mikrosatelliten (BALL et al., 2010). Die Entwicklung direkter Array-basierter SNP-Genotypisierungstests ermöglicht es, eine hohe Anzahl an SNPs

vergleichsweise kostengünstig und zeiteffizient zu genotypisieren (GUNDERSON et al., 2005). Der OvineSNP50 BeadChip (**Abbildung 3**) ist in der Lage mehr als 54,000 SNPs, welche relativ gleichmäßig über das gesamte Genom verteilt sind, zu genotypisieren (ILLUMINA, 2015).

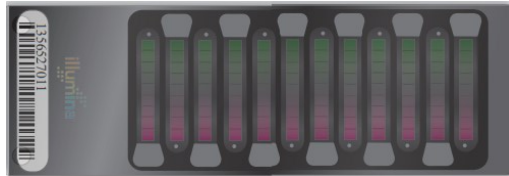


Abbildung 3: OvineSNP50 BeadChip.

Diese Version ist in der Lage 12 Proben gleichzeitig an über 54,000 Positionen zu genotypisieren. Quelle: ILLUMINA (2015)

Dabei können je nach Ausführung des Chips mehrere Proben gleichzeitig analysiert werden. Auf der Oberfläche eines Bead-Chips befinden sich mehrere tausende Silica-beads, diese sind wiederum mit mehreren Kopien einer Oligonukleotid-Sonde beschichtet, welche auf einen spezifischen Locus des Genoms abzielen. Die zuvor amplifizierten DNA-Fragmente, der zu untersuchenden Proben, binden anschließend an die komplementäre Oligonukleotidsequenz der Sonde mittels Hybridisierung (**Abbildung 4**). Die Sonden-Sequenz endet allerdings eine Base vor dem „Locus of interest“, sprich jener Position auf der sich der SNP befindet. Durch die anschließende Einzelbasenverlängerung mittels einer Polymerase wird die Allelspezifität gewährleistet. Die Basen, die zur Verlängerung verwendet werden, sind markiert und sondern ein Signal aus, welches anschließend detektiert wird (siehe **Abbildung 5**).

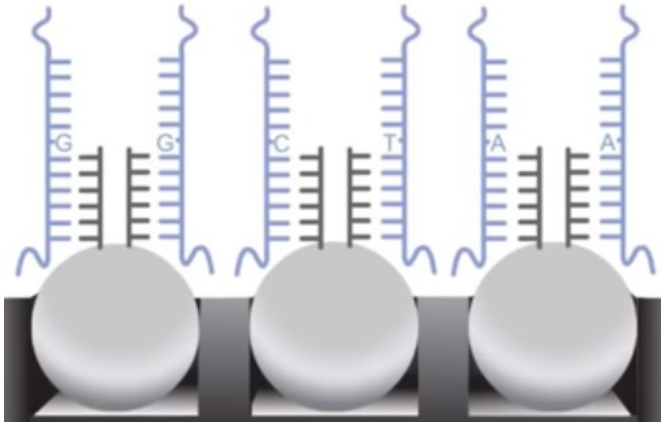


Abbildung 4: Drei Sonden eines Bead-Chips.

Diese zielen auf unterschiedliche DNA-Fragmente (blaue Stränge) des beprobten Genoms ab. Die Fragmente binden an die Oligonukleotide (graue Stränge), welche bereits an den Sonden angeheftet sind und eine Base vor dem eigentlich zu charakterisierenden SNP enden. Quelle: [Illumina Advances Genomic Research with the Infinium Assay - YouTube](https://www.youtube.com/watch?v=IVG04dAAyvY) <https://www.youtube.com/watch?v=IVG04dAAyvY>

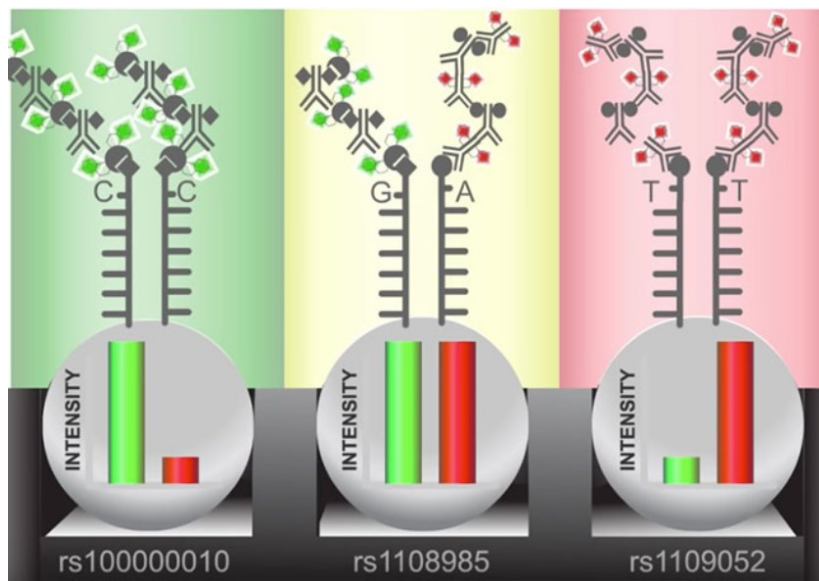


Abbildung 5: Jene Drei Sondenoligonukleotide wie in Abbildung 4 zu sehen.

Hier wurde nach der Hybridisierung, die Sonde entsprechend der fragmentierten Sequenz des zu untersuchenden Individuums verlängert. Jede Sonde genotypisiert einen Locus mit zwei Farbanzeigen, eine Farbe für jedes Allel. Die relative Intensität der beiden Farben zeigt an, ob ein Genotyp an diesem Locus heterozygot oder homozygot ist. Quelle: [Illumina Microarray Technology](https://www.illumina.com/science/technology/microarray.html) <https://www.illumina.com/science/technology/microarray.html>

10.4. Kopplung (Linkage), Kopplungsgleichgewicht und Kopplungsungleichgewicht

Zwei verschiedene Methoden wurden in den letzten zwei Jahrzehnten verwendet um Gene und deren verschiedenen Ausprägungen, welche zu Variationen im Phänotyp beitragen, zu identifizieren (GODDARD und HAYES, 2009). Die erste Methode ist die Untersuchung bestimmter chromosomaler Regionen von bereits bekannten Kandidatengen, die das untersuchte Merkmal beeinflussen könnten, anschließend wird nach kausativen Mutationen in den Kandidatengen gesucht. Die zweite Methode ist die Kartierung auf eine chromosomale Region unter Verwendung genetischer Marker. Erstere wurde in der Vergangenheit oft als attraktive Strategie für die Analyse monogener Merkmale angesehen, letztere für die Ermittlung von QTLs. Merkmale mit einem multifaktoriellen Hintergrund werden sowohl durch Umweltfaktoren als auch durch eine unbekannte Anzahl von QTLs gesteuert (ANDERSSON und GEORGES, 2004). Daher stellen diese komplexen Merkmale besondere Herausforderungen für die genetische Analyse dar. Außerdem sind Wechselwirkungen zwischen einzelnen Genen sowie zwischen Genen und Umwelt, genetische Heterogenitäten, geringe Penetranz sowie mangelnde statistische Aussagekraft weitere Hürden bei der Kartierung von QTLs (GLAZIER et al., 2002).

Die dritte Mendelsche Regel besagt, dass die Erbanlagen für zwei verschiedene Merkmale frei und unabhängig voneinander segregieren. Wie später von BATESON und PUNNETT (1911) sowie MORGAN (1911) festgestellt wurde trifft die Unabhängigkeitsregel nur auf Merkmale zu, die durch Gene codiert werden, welche an verschiedenen Chromosomen lokalisiert sind. Des Weiteren trifft sie auch auf Gene zu, die zwar auf ein und demselben Chromosom lokalisiert sind, aber weit voneinander entfernt liegen und somit nicht gekoppelt vererbt werden. Man spricht in diesem Zusammenhang auch von Kopplungsgleichgewicht oder linkage equilibrium (LE), mindestens zwei Allele sind dann im LE, wenn deren Auftreten unabhängig voneinander und somit zufällig in einer Population sind. Betrachtet man zwei benachbarte polymorphe Loci A und B mit jeweils zwei Allelen (A1, A2, B1 und B2), so sind folgende Haplotypen möglich: A1_B1, A1_B2, A2_B1, A2_B2. Ist die Frequenz eines dieser 4 Haplotypen das Produkt seiner einzelnen Allelfrequenzen so sind diese Loci im LE und somit unabhängig voneinander (GOODE, 2011). Im Gegensatz dazu treten diese Allele bei einem

Kopplungsungleichgewicht (auch linkage disequilibrium, LD, genannt) nicht mit der Frequenz auf, die zu erwarten wäre, wenn die Allele unabhängig sind. Man spricht von einem positiven Kopplungsungleichgewicht, wenn zwei Allele häufiger als erwartet gemeinsam auf demselben Haplotypen vorkommen. Unter einem negativen LD versteht man hingegen, das seltenere Auftreten zweier Allele auf ein und demselben Haplotypen. Darüber hinaus kann die LD auch durch genetischen Drift, Vermischung, Mutations- und Rekombinationsraten, Selektion, begrenzte Populationsgröße, Populationsengpässe oder andere demographische Ereignisse in einer Population beeinflusst werden (SLATKIN, 2008). Wenn eine vorteilhafte Mutation auftritt und sich bis zur Fixierung vermehrt, verschafft sie allen Allelen, mit denen sie in LD ist, eine Erhöhung der Frequenz. Wodurch die Biodiversität und durchschnittliche Heterozygotie rund um diese positiv selektierte Mutation abnimmt. Dieser Effekt wurde von SMITH und HAIGH (2009) erstmals als *'hitchhiking'* bezeichnet.

Im Folgenden wird auf die zweite Methode der Kartierung von Kandidatengen genauer eingegangen. Wie im Abschnitt II.10.2 bereits erwähnt, unterscheidet man zwischen LM und LDM. Zusätzlich gibt es noch eine Kombination von beiden, welche als combined linkage disequilibrium and linkage mapping (cLDLA) bezeichnet wird.

10.5. Kopplungskartierung, LM

Linkage Mapping ist eine Technik zur Kartierung von Genen. Diese Technik macht sich das Wissen zu Nutze, dass nahe beieinanderliegende Loci häufiger miteinander vererbt werden als jene, welche weiter weg von einander lokalisiert sind. Sprich je mehr Rekombination zwischen zwei genetischen Positionen stattfindet, desto weiter sind sie vermutlich voneinander entfernt (GELDERMANN, 2005). Bei Kopplungsanalysen wird in der Regel mit großen Halbgeschwisterfamilien gearbeitet und unter Verwendung von Markern. Die verwendeten Halbgeschwisterfamilien umfassen in der Regel zwei bis drei Generationen. Bei eng nebeneinander positionierten Markern finden in diesen zwei bis drei Generationen nur wenige Rekombinationen zwischen benachbarten Markern statt (es sei denn, die Anzahl der Individuen pro Generation ist sehr groß), so dass ein dichtes Markerset nur wenig zusätzliche Informationen über die Position des QTL

liefert (MEUWISSEN und GODDARD, 2000). Dieser Umstand führt dazu, dass QTLs in reinen Kopplungsanalysen in der Regel in einem großen Intervall von 20 cM oder mehr kartiert werden, was die Identifizierung der zugrunde liegenden Varianten und die Verwendung der markergestützten Selektion in Tierzuchtprogrammen erschwert (GODDARD und HAYES, 2009).

10.6. Kopplungsungleichgewichtskartierung, LDM

Mittels Kopplungsungleichgewichten ist eine präzisere Kartierung von Merkmalsloci möglich, da diese in Nutztierpopulationen relativ häufig auftreten (ANDERSSON und GEORGES, 2004) und dabei alle Rekombinationen seit dem Auftreten der Mutation verwendet werden. Die Kopplungsungleichgewichtskartierung verspricht eine genaue Schätzung der QTL-Positionen, da sie nicht nur Rekombinationen in der direkt beobachteten Kartierungspopulation, sondern auch historische Rekombinationen nutzt. Dadurch werden Kopplungskartierungen nur für eine grobe genomweite Suche nach QTLs eingesetzt (MEUWISSEN und GODDARD, 2000). Bei der Kopplungsungleichgewichtskartierung werden Rekombinationsereignisse genutzt um Chromosomensegmente zu einem gemeinsamen Vorfahren im Stammbaum zurückzuverfolgen. Dieser Vorfahre lebte selbst oft hunderte Generationen zuvor und vererbte an seine Nachkommen, die ebenfalls Merkmalsträger sind, die herkunftsidentische (*identity by descent*, IBD) Chromosomensegmente. Genetisch identische Genomabschnitte, welche nicht von gemeinsamen Vorfahren abstammen werden hingegen als *identical by state* (IBS) bezeichnet. Die IBD-Region wird durch eng miteinander verknüpfte Markerloci aufgespürt, die bei den Merkmalsträgern identische Allele bzw. Haplotypen in dieser Region tragen (z. B. HOUWEN et al. (1994)). Auf diese Weise werden gleichzeitige Kopplungsungleichgewichte zwischen mehreren Markern und dem Kandidatengen festgestellt. Die Größe der IBD-Region wird mit der Anzahl der Meiosen seit dem Auftreten der kausalen Mutation kleiner (MEUWISSEN und GODDARD, 2000). Theoretisch wird die LD-Kartierung umso feiner, je weiter der gemeinsame Vorfahre zurückliegt, da die historische Rekombination einen fortschreitenden Zerfall des IBD-Segments bewirkt und nur Marker, die nahe an der gesuchten Variante liegen, in starker LD bleiben. Der Grad des LD spiegelt den Abstand zwischen dem Marker und dem kausativen Locus wider und kann daher zur Kartierung der Loci verwendet werden (XIONG und JIN, 1999). Das LD ist

mitunter abhängig von der effektiven Populationsgröße (N_e). Eine kleine N_e bedeutet, dass nur eine geringe Anzahl von herkunftsverschiedenen Gameten oder Gametensegmenten in der aktuellen Population segregiert. Diese Allele bzw. Gametensegmente stammen von wenigen gemeinsamen Vorfahren ab. Das hat zur Folge, dass IBD-Chromosomenabschnitte groß sind, so dass sich die LD über eine große Entfernung erstreckt (GODDARD und HAYES, 2009). Eine interessante Information ist, dass die effektive Populationsgröße des Rindes vor der Domestizierung groß war (>50.000), ging aber nach der Domestizierung auf 1.000-2.000 zurück, aufgrund der Rassenbildung und moderner Zuchtprogramme sank sie weiter auf circa 100 (GODDARD und HAYES, 2009). Die aktuelle effektive Populationsgröße bei den meisten Schafrassen ist größer als bei Rinderrassen (KIJAS et al., 2012), aber Effekte der künstlichen Isolation (Rassegründung und Reinzucht) und künstlichen Selektion führen zu vergleichbaren Trends in verschiedenen domestizierten Spezies und deren Rassen. Gegenüber der Kopplungsungleichgewichtskartierung wurden neben der genannten Vorteile auch negative Vorbehalte geäußert, in erster Linie wegen der hohen erforderlichen Markerdichte (z.B.: KRUGLYAK (1999)).

10.7. Kopplungsungleichgewichts- und Kopplungs-Kartierungen (*linkage disequilibrium and linkage mapping, cLDLA*)

Bei dieser Kartierungsmethode werden Informationen aus der Kopplung und aus dem Kopplungsungleichgewicht in Kombination verwendet. Informationen aus der Kopplung beschränken sich auf die Generationen, welche genotypisiert wurden. Kopplungsungleichgewicht verwendet auch Informationen aus historischen Rekombinationen (FARNIR et al., 2002; MEUWISSEN et al., 2002). Eine cLDLA ist ein vielversprechender Ansatz für die Feinkartierung und hat sich als ausreichend genau erwiesen, um das QTL-Konfidenzintervall (KI) auf wenige cM des Genoms einzugrenzen (LEE und VAN DER WERF, 2004). Die in dieser Forschungsarbeit durchgeführte cLDLA wurde in Anlehnung an die in MEUWISSEN et al. (2002) beschriebene Methode durchgeführt und wird in Kapitel III.6 näher erläutert. Dabei ergibt sich die Wahrscheinlichkeit der IBD am QTL zwischen Haplotypen der ersten Generation aus der Ähnlichkeit der Markerhaplotypen, die den QTL umgeben (MEUWISSEN et al., 2002).

10.8. Genomweite Assoziationsstudien GWAS

Genomweite Assoziationsstudien suchen nach Assoziationen zwischen einem bestimmten Merkmal (welches zuvor systematisch in eine Kartierungspopulation phenotypisiert wurde) und einer genetischen Variation (meistens SNP- aber auch VNTR-Markern). Dabei testen Assoziationsstudien auf Unterschiede in der Allelhäufigkeit genetischer Varianten zwischen Individuen, die gemeinsame Vorfahren haben, sich aber phänotypisch unterscheiden (UFFELMANN et al., 2021). In der Regel wird SNP für SNP in einem linearen Modell, welches neben dem Effekt des SNPs auch fixe Effekte und polygene Effekte bewertet, auf eine Assoziation getestet. Bei GWAS nimmt man an, dass signifikante Assoziationen entstehen, weil ein SNP in LD mit und damit in der Nähe einer kausativen Mutation ist (HAYES und GODDARD, 2010). Dadurch, dass die Positionen von Markern genau bekannt sind, ist es so möglich, die Position einer kausalen Variante im Genom einzugrenzen. Die Ergebnisse dienen nicht nur dazu, die einem Phänotyp zugrundeliegende Biologie aufzuklären, sondern auch die Heritabilität eines Merkmals zu schätzen, genetische Korrelationen zu berechnen und Einblicke in viele andere Anwendungen zu gewinnen (UFFELMANN et al., 2021). GWAS-Ergebnisse fließen ebenfalls indirekt in die genomische Selektion. Wie groß die Kartierungspopulation und die Anzahl an SNPs sein muss um eine signifikante Assoziation zu detektieren ist von vielen Faktoren abhängig und generell schwierig vorauszusagen. Eine ungefähre Anlehnung ist in GODDARD und HAYES (2009) beschrieben.

11. Selektive Genotypisierung und Phänotypisierung

Bei einer selektiven Genotypisierung werden bei der Überprüfung einer möglichen Kopplung zwischen einem Markerlocus und einem QTL, Individuen genotypisiert, dessen Phänotypen die größten Abweichungen vom Mittelwert aufweisen (DARVASI und SOLLER, 1992). Wie **Abbildung 6** am Beispiel der Schwanzlänge verdeutlicht sind das jene Schafe mit extrem kurzen oder extrem langen Schwänzen. Dadurch erhöht sich zwar in der Regel die Anzahl der phänotypisierten Organismen, doch können bei gleicher statistischer Power weniger genotypisiert werden, wodurch Kosten gesenkt werden. DARVASI und SOLLER (1992) kamen zu dem Entschluss, dass es in einzelnen Merkmalsstudien nicht nützlich ist mehr als 25 % beider Extreme zu genotypisieren. Andererseits wurde ebenfalls bereits demonstriert, dass selektives Phänotypisieren eine effektive Methode ist, um die

Merkmalsausprägung bei möglichst wenigen Organismen phänotypisieren zu müssen. Das ist bei Arbeiten, bei denen die Kosten für die Phänotypisierung sehr hoch sind und daher die Anzahl der phänotypisierten Individuen gesenkt werden sollen, von besonderem Interesse (JIN et al., 2004).

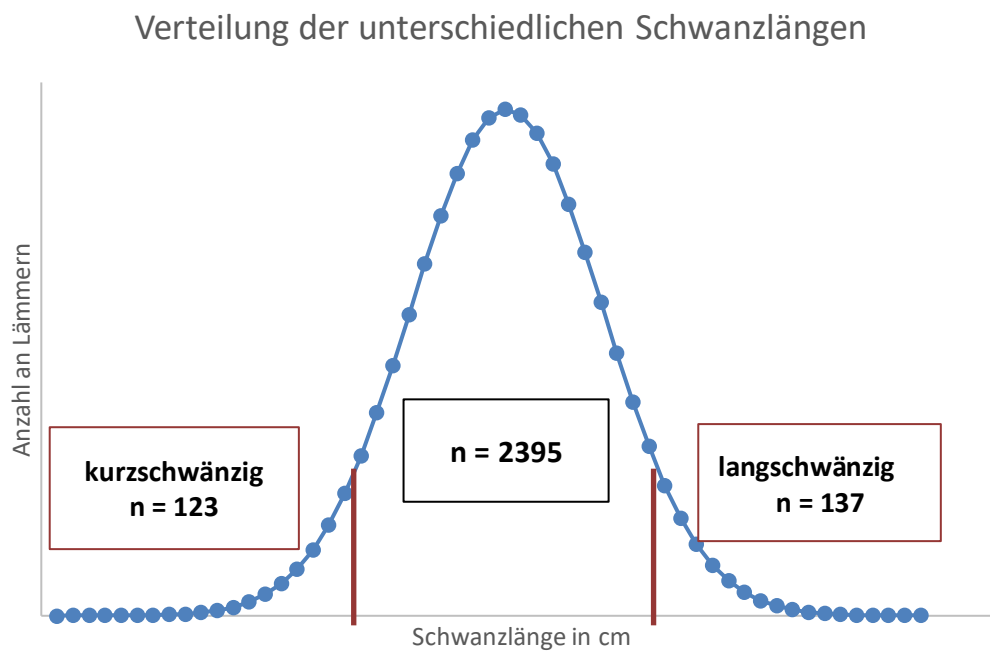


Abbildung 6: Verteilung der selektiv beprobten Lämmer.

Von 2395 visuell inspizierten Lämmern wurden 260 Lämmer von beiden Enden des phänotypischen Spektrums (123 besonders kurzschwänzige Lämmer und 137 besonders langschwänzige Lämmer) für die Genotypisierung mit dem OvineSNP50 BeadChip von Illumina ausgewählt.

III. MATERIAL UND METHODEN

1. Probenahmen und Phänotypisierung von 362 Lämmern

Für das Studiendesign wurden insgesamt 362 Lämmer der Rasse Merinolandschaf in drei verschiedenen Durchläufen beprobt und phänotypisiert (siehe **Tabelle 2**). Im ersten Durchlauf wurden 236 von insgesamt 2,293 Lämmern selektiv für die Genotypisierung beprobt, 104 Lämmer davon hatten sehr kurze und 132 sehr lange Schwänze. Im zweiten Durchlauf wurden 102 männliche Lämmer (von insgesamt 110 visuell betrachteten) nicht selektiv, also zufällig in Bezug auf deren Schwanzlänge, beprobt und vermessen. Die Lämmer des ersten und zweiten Durchlaufs befanden sich auf einem Bauernhof in Niederbayern, die restlichen 24 Lämmer des dritten Durchlaufs befanden sich auf der Lehr- und Versuchsanstalt „Oberer Hardthof“ der Justus-Liebig-Universität Gießen. Die Lämmer vom Oberen Hardthof wurden ebenfalls anhand der Schwanzlänge, aus insgesamt 102 Lämmern vorselektiert und setzten sich aus 19 kurz- und 5 langschwänzigen Merinolandschafämmern zusammen. Im ersten und dritten Durchlauf wurden also nur Tiere ausgewählt, welche eine extreme Schwanzlänge aufwiesen (sehr kurz oder sehr lang) wie von DARVASI und SOLLER (1992) empfohlen (siehe **Abbildung 6**). Sämtliche Lämmer, welche Verletzungen, Knicke oder sonstige Auffälligkeiten am Schwanz zeigten, wurden aus der Studie ausgeschlossen. Verletzungen des Schwanzes waren häufig auf dem in Abschnitt II.5.1 beschriebenen Grund zurückzuführen.

Tabelle 2: Anzahl und Einteilung der beprobten Lämmer.

Probenahme -ort	Durchlauf	Visuell betrachtet	Beprobt			
			Insgesamt	Kurz	Lang	Zufällig
Niederbayern	1. Durchlauf	2293	236	104	132	-
	2. Durchlauf	110	102	-	-	102
Oberer Hardthof	3. Durchlauf	102	24	19	5	-
Insgesamt		2505	362	123	137	102

Um die Haplotypisierung zu verbessern, wurden zusätzlich 22 mutmaßliche Väter der oben genannten Lämmer genotypisiert. Diese Widder wurden zumeist kupiert und daher nicht phänotypisiert und trugen nur indirekt zur QTL-Kartierung bei.

Alle Blutproben wurden von erfahrenen Tierärztinnen und mit Genehmigung der Regierung von Oberbayern (Genehmigungsnummer: 55.2-1-54-2532.0-47-2016), bzw. des Regierungspräsidiums Gießen, Hessen (KTV-Nummer: 19 c 20 15 h 02 Gi 19/1 KTV 22/2020) entnommen.

In **Tabelle 3** sind sämtliche Geräte die für die Messung der verschiedenen Körpermerkmale, zur Entnahme der DNA-Proben und für die DNA-Isolierung verwendet wurden aufgelistet:

Tabelle 3: Verwendete Geräte deren Hersteller und Verwendungszweck.

Gerät	Hersteller	Verwendungszweck
Babywaage	Monzana	Abwiegen der Lämmer
Personenwaage	Beurer	Abwiegen der Lämmer
Hölzernes Messbrett	Eigenmarke	Schwanzlängenmessung
Körmaß	Bromet	Messung der Widerristhöhe
Einmal-Kanülen (18 G x 1 ½)	100 Sterican	Blutentnahme
Blutentnahmeröhrchen 4ml K3E	S-Monovette	Blutentnahme
Kanüle (20G x 1 ½)	S-Monovette	Blutentnahme
Kühlschrank	Siemens	Lagerung der Blutproben und DNA
Gefrierschrank	Bosch	Lagerung der Blutproben und DNA
Autoklav	Thermo Fisher Scientific	Sterilisation
Zentrifuge	Hettich	Abzentrifugieren des Blutes
Thermomixer	Bachofer	Lyse von Zellen
NanoDrop Spektrophotometer ND-100	peqlab	Messung der DNA- Konzentration

2. Durchführung der Genotypisierung

Das Labor GeneControl der Tierzuchtforschung e.V. München in Grub genotypisierte alle 362 Lämmer und deren 22 mutmaßlichen Väter mit dem Illumina OvineSNP50 BeadChip (Illumina, San Diego, USA), entsprechend den Herstellerangaben. Dieser Chip enthält 54,241 annähernd gleichmäßig verteilte SNPs. Der durchschnittliche Abstand von 50.9 kilobasen (kb) bietet eine relativ gleichmäßige Abdeckung des gesamten Schafsgenoms (ILLUMINA, 2015). Die

durchschnittliche *Call-Rate* von > 99% und die Reproduzierbarkeit von über 99.9% sind neben der hohen SNP-Dichte gute Voraussetzungen für erfolgreiche Kartierungsstudien. Außerdem können durch die Anwendung solcher direkten Array-basierten SNP-Genotypisierungstests mehrere Bearbeitungsschritte eingespart werden, da eine PCR-Amplifikation überflüssig wird und die Probenvorbereitung in einem einzigen Röhrchen möglich ist. Dadurch minimiert sich das Risiko eines Fehlers. Die Funktionsweise des Assays ist in Abschnitt II.10.3.3 erklärt und wird hier nochmal entsprechend der Anleitung in GUNDERSON et al. (2005) kurz zusammengefasst:

- (i) Amplifikation des Ganzgenoms um genügend DNA für die Hybridisierung zu generieren;
- (ii) Hybridisierung der DNA-Fragmente an eine spezifische Oligonukleotidsonde auf dem Array;
- (iii) Biotin-markierte Desoxyribonukleosidtriphosphate (dNTPs) verlängern die Oligonukleotidsonde bevorzugt an übereinstimmenden 3'-Termini, aber nur wenn das DNA-Fragment die komplementäre Base zu der am 3'-Termini der Oligonukleotidsonde befindlichen Base (Lokalisation des SNPs) aufweist;
- (iv) empfindliche Detektions- und Signalverstärkungsschritte zum Auslesen der eingebauten dNTPs werden durchgeführt und lassen Rückschlüsse auf die Base im DNA-Fragment zu. Dadurch kann man bestimmen welchen Genotyp das jeweilige Individuum auf der jeweiligen Position trägt.

Die physikalischen Markerpositionen wurden auf dem Schaf-Referenzgenom Oar_4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2) bestimmt. Die Positionen aller Marker oder Sequenzen in dieser Arbeit beziehen sich auf das Referenzgenom Oar_4.0, sofern nicht anders angegeben.

2.1. Herausfiltern nicht informativer Marker

Nicht alle der 54,241 SNPs wurden für die Kartierung verwendet. Wir filterten die SNPs nach den folgenden Ausschlusskriterien:

- (i) erfolglose Genotypisierung des Markers bei mehr als 5 % der beprobten Tiere;
- (ii) häufige Abstammungskonflikte bei Tieren mit bekannter Vaterschaft;

- (iii) unbekannte Position im Referenzgenom;
- (iv) Minor Allel Frequency (MAF) von weniger als 0.025 und;
- (v) Lokalisierung auf einem Geschlechtschromosom, da die Analysen ausschließlich auf Autosomen durchgeführt wurden.

Infolgedessen verblieben 45,114 Marker im Markerset für die Kartierungsanalysen.

Lämmer welche eine Call-Rate von unter 95 % hatten (mehr als 5% der SNPs wurden nicht erfolgreich genotypisiert) wurden aus den Kartierungsanalysen ausgeschlossen.

Sämtliche Programme und Softwares, die in dieser Studie verwendet wurden sind in **Tabelle 4** aufgelistet

Tabelle 4: Verwendete Programme und Softwarepakete sowie deren Verwendungszweck und Quelle.

Programm	Verwendungszweck	Quelle
<i>BEAGLE</i> 5.0	Haplotypisierung und Imputation der Genotypen	BROWNING et al. (2018); BROWNING und BROWNING (2007)
MICROSOFT SQL SERVER	Verwaltungen großer Datenmengen wie zum Beispiel alle Genotypen	http://msdn.microsoft.com/de-de/sqlserver/default
MICROSOFT OFFICE 2016	Textverfassung, Tabellenanfertigungen, grafische Veranschaulichung der LRT-Werte	http://office.microsoft.com/de-de
MICROSOFT VISUAL STUDIO 2012	Erstellung von Programmen zur Datenvorbereitung, für die Durchführung von diversen Analysen, sowie zur Veranschaulichung von Analyseergebnissen	https://www.visualstudio.com/
R VERSION 4.1.2	Schätzung der genomweiten vereinheitlichten Verwandtschaftsbeziehungen	https://www.r-project.org R CORE TEAM (2020)
<i>GINVERSE</i>	Bildung der generalisierten Inverse der genomweiten Verwandtschaftsmatrix	Karin Meyer; University of New England (Australia)
<i>ASREML</i>	Varianzkomponentenanalyse, Diplotypeffekt-Analyse	BUTLER et al. (2018)
<i>GCTA</i> v1.93.2	Genomweite Assoziationsanalyse	YANG et al. (2011)
<i>SICKLE</i>	Trimmen der Adaptern von Reads und Herausfiltern von Sequenzen geringer Qualität	JOSHI NA (2011)
<i>FASTQC</i> (BABRAHAM,	Bewertung der Qualitätsparameter der gefilterten	BABRAHAM (2011)

2011)	Sequenzierdaten		
<i>BWA-MEM</i> ALIGNMENT- TOOLS	Mapping der Reads auf das Referenzgenom OAR_v4.0	LI und DURBIN (2009)	
<i>SAMTOOLS</i>	Konvertierung der SAM-Dateien in koordinatensortierte BAM-Dateien	LI et al. (2009)	
<i>PICARD</i>	Entfernung der duplizierten Reads	BROAD INSTITUTE (2019)	
<i>GATK</i> <i>HAPLOTYPECALLER</i>	SNPs- und Indels-Calling	AUWERA und O'CONNOR (2020)	
<i>SMOOVE (LUMPY)</i>	Abrufen von Structural variations	LAYER et al. (2014)	
<i>DELLY</i>	Abrufen von Structural variations	RAUSCH ET AL. (2012)	
<i>JBROWSE</i>	Visuelle Betrachtung des Genoms	BUELS et al. (2016)	
NCBI Genome Remapping Service	Zur Konvertierung von Annotierungsdaten von einem Referenzgenom in ein anderes	https://www.ncbi.nlm.nih.gov/genome/tools/remap	
<i>PRIMER3</i>	Primer-Konzipierung	UNTERGASSER ET AL. (2012)	

3. Haplotypisierung und Imputation

Die Haplotypisierung, auch *Phasing* genannt, und die Imputation wurden mit dem Programm *BEAGLE 5.0* (BROWNING et al., 2018; BROWNING und BROWNING, 2007) durchgeführt, welches auf einem *Hidden Markov Model* basiert. Diese zwei Ansätze verbessern die Leistung von Kartierungsstudien (BROWNING, 2008). Zwar erhält man nach der Genotypisierung den Genotyp für sämtliche Marker eines Individuums, jedoch bestehen diese nur aus ungeordneten Paaren von Allelen. Durch die Haplotypisierung werden diese Genotypen in Phase gebracht, d.h. alle Allele die vom selben Elternteil stammen, werden entsprechend ihrer chromosomalen Zugehörigkeit geordnet. Die dadurch entstehenden Haplotypen sind Anordnungen von Allelen innerhalb eines Chromosoms. Sowohl Ei- als auch Samenzelle sind Träger eines vollständigen haploiden Chromosomensatzes. Durch die Befruchtung der Eizelle zur Zygote verschmelzen väterlicher und mütterlicher Chromosomensatz zu einem diploiden Chromosomensatz (2n). Somit ist in der Zygote als auch in jeder weiteren Zelle bei höheren Organismen jedes Chromosom doppelt vorhanden (Diploidie) (JANNING und KNUST, 2004). Wie bereits

erwähnt können Haplotypen durch Linkage disequilibrium Informationen von Varianten liefern, welche sich zwischen bekannten Genotypen befinden und dadurch Assoziationen zu nicht genotypisierten Varianten detektieren (BROWNING, 2008).

Bei der Imputation werden, durch die Verwendung von Haplotypmustern aus einem Referenzpanel, Genotypen abgeleitet, welche in einer einzelnen Probe nicht genotypisiert werden konnten. Dieser Ansatz ermöglicht es, Genotypdaten einzelner Individuen gleichzeitig mit bekannten und unbekanntem Haplotypen von unverwandten Tieren zu analysieren und fehlende Genotypen während des Phasings zu imputieren. Das Programm *BEAGLE* vereinfacht dabei die Skalierung der Methoden, was die Analyse von Tausenden von Schafen ermöglicht. Das ist sehr vorteilhaft, da die Größe des Referenzpanels einen erheblichen Einfluss auf die Imputationsgenauigkeit hat, vor allem bei genetischen Varianten, welche in geringer Frequenz segregieren. Selbst wenn die Haplotypen des Referenzpanels nicht bekannt sind, können diese aus den Genotypdaten von unverwandten Tieren mit hoher Genauigkeit abgeleitet werden (BROWNING und BROWNING, 2009). Im Falle dieser Arbeit bestand das Referenzpanel aus ca. 5,100 zusätzlichen Schafen verschiedenster Rassen, welche für eine präzisere und genauere Haplotypisierung und Imputation sorgten. Die 5,100 zusätzlichen Schafen wurden für die weitere Kartierung nicht verwendet.

Der MICROSOFT SQL SERVER wurde verwendet um den Lämmern über Abstammungsabfragen ihre Väter zuzuweisen, falls diese sich unter den genotypisierten befanden.

4. Korrektur auf Populationsstratifikation und familiäre Verwandtschaftsbeziehungen

Klassische Kartierungsstudien wie Kopplungs- und Assoziationsstudien sind mit diversen Herausforderungen konfrontiert. In der Regel werden hierfür eine hohe Anzahl von SNPs, welche gleichmäßig über das gesamte Genom verteilt sind und viele Individuen benötigt (LIU et al., 2013; VISSCHER et al., 2014). Zusätzlich ist es notwendig Korrekturen auf Populationsstratifikation und unbekanntem Verwandtschaftsbeziehungen durchzuführen, da wie in Abschnitt II.10.4 erwähnt, die

LD von diversen populationsgenetischen Prozessen beeinflusst wird. Populationsstratifikation kann zu trügerischen Assoziationen und dadurch zu falsch positiven Kandidaten führen. Dies macht eine Bewertung und Korrektur notwendig (HIRSCHHORN und DALY, 2005) (LIU et al., 2013). Grund dafür sind unterschiedliche Allelhäufigkeiten zwischen verschiedenen Populationen, da sich die Populationen nicht gleichmäßig auf Fälle und Kontrollen verteilen. Jene Population mit der höheren Anzahl an Merkmalsträgern, bezogen auf die Gesamtpopulation, ist in der Fallgruppe überpräsentiert, dadurch werden Unterschiede aufgedeckt, die auf unterschiedliche Abstammung beruhen und nicht für das Merkmal verantwortlich sind (LIU et al., 2013).

Neben der Populationsstratifikation muss bei Kartierungsstudien auch auf unbekannte Verwandtschaftsbeziehungen korrigiert werden. Da vor allem bei unseren domestizierten Haustieren oft Verwandtschaftsbeziehungen bestehen, zum Beispiel in Form von Halbgeschwisterfamilien bei Rindern, Schafen und Ziegen oder Vollgeschwisterfamilien bei Schweinen. Unverwandte Tiere hätten bei Assoziationsstudien den Vorteil, dass eine Assoziation zwischen einem SNP und einem Merkmal in der gesamten Population besteht. Familiäre Beziehungen führen zu zusätzlichen LD zwischen Regionen die eigentlich ungekoppelt sind. Wenn beispielsweise ein Vorfahre einer Familie seltene Allele an zwei nicht miteinander verknüpften Loci trägt, ist es wahrscheinlicher, dass dessen Nachkommen diese seltenen Allele tragen als Tiere die eine andere Abstammung haben. Dies führt zu Assoziationen zwischen einem neutralen Locus und dem Merkmal (GODDARD und HAYES, 2009).

Um diese Korrekturen zu bewerkstelligen, wurden die genomweiten additivgenetischen Verwandtschaftsbeziehungen (*unified additive relationships*, UAR) mit dem R-Paket *SNPREADY* zwischen allen beprobten Lämmern geschätzt. Die dabei entstehende genomweite Verwandtschaftsmatrix wird anschließend verwendet um die Wahrscheinlichkeit zu berücksichtigen, ob ein Locus gleiche Allele aufgrund identischer Herkunft (IBD) oder durch Zufall und somit unabhängig der Vorfahren (IBS) in einer Basispopulation hat (POWELL et al., 2010). Zur Berücksichtigung linearer Gleichungssysteme wurden nicht die genomweite Verwandtschaftsmatrix, sondern deren Inverse verwendet. Hierfür wurde mit dem Programm *GINVERSE* die generalisierte Inverse der genomweiten Verwandtschaftsmatrix gebildet. Die genomweite Verwandtschaftsmatrix bzw. deren Inverse erlaubt eine Schätzung und Miteinbeziehung zufälliger polygener

Effekte in die Varianzkomponentenanalyse.

Wie im Abschnitt II.10.6 beschreibt der Begriff IBD zwei homologe Allele, welche vom selben Vorfahren aus der Basis- oder Referenzpopulation abstammen. Das Konzept des IBD findet in vielen verschiedenen Bereichen der Genetik Anwendung (POWELL et al., 2010) mitunter auch in der Kartierung von Genen (ALBRECHTSEN et al., 2009). Da in vielen Fällen Abstammungsinformationen fehlen, verwendet man üblicherweise SNPs um die genomweite Verwandtschaft bzw. genomweite IBD zu schätzen. Ebenso werden SNPs für lokale (Chromosom-Segment) Verwandtschafts- bzw. IBD-Wahrscheinlichkeitsschätzung angewendet (siehe unten). Dabei stößt man unter anderem auf das Problem, dass bis zu einem gewissen Grad alle Individuen miteinander verwandt sind, vorausgesetzt sie werden weit genug zurückverfolgt. Somit werden auch sogenannte „unverwandte“ Individuen unterschiedliche Verwandtschaftsbeziehungen aufweisen, was eine Korrektur erschwert (POWELL et al., 2010). Träger einer Mutation, welche ursprünglich in der Gründerpopulation auftauchte, weisen um diese Gründermutation einen höheren Verwandtschaftsgrad oder IBD auf. Wenn jetzt in einer Kartierungsstudie mit Fall- und Kontrollgruppe nur in den Merkmalsträger eine erhöhte Verwandtschaft in einer bestimmten Region festgestellt wird, spricht das für eine ursächliche Mutation für dieses Merkmal in dieser Region (ALBRECHTSEN et al., 2009).

5. Erstellung einer Diplotypen-Verwandtschaftsmatrix anhand der Locus IBD

Zur Kartierung eines QTLs im Genom wurden zusätzlich zu der genomweiten Verwandtschaftsmatrix lokale Verwandtschaftsmatrizen anhand der Methode von MEUWISSEN und GODDARD (2001) erstellt. Dabei wird anhand mindestens zweier verschiedener Markerhaplotypen geschätzt, mit welcher Wahrscheinlichkeit an einer bestimmten Stelle im Genom ein Locus IBD ist. Je mehr Marker pro Haplotyp verwendet werden, desto genauer ist die Schätzung der Locus IBDs und somit auch der Kartierung. Für die Kartierung potentieller QTLs der Schwanzlänge beim Merinolandschaf wurden jeweils 40 benachbarte Marker in einem Gleitfenster, welches Marker für Marker über das Genom gleitet und dabei stets die IBD zwischen dem 20. und 21. Marker schätzt, betrachtet. Für jedes Markerfenster entlang des Genoms und für jedes Paar der Tiere aus der Kartierungspopulation

(hier 362) werden vier Locus IBD-Werte (maternaler und paternaler Haplotyp vom Individuum A mit dem maternalen und paternalen Haplotyp vom Individuum B) geschätzt (LIU et al., 2002). Die Überführung der Haplotypen-LocusIBD-Matrix in eine Diplotypen-Verwandtschaftsmatrix (diplotype relationship matrix, DRM) erfolgte anhand der Methode von LIU et al. (2002) für additive-genetische Verwandtschaftsmatrizen (additive genetic relationship matrices, GRM). Hierfür wird die Summe der 4 IBD-Wahrscheinlichkeiten durch 2 dividiert.

6. Varianzkomponentenanalyse

Die Varianzkomponentenanalysen wurden mit dem Programm *ASREML* (BUTLER et al., 2018) nach der Methode von MEUWISSEN et al. (2002) durchgeführt. Dabei werden Kopplungsinformationen aus der Rekonstruktion der Haplotypen unter Einbezug der Pedigree-Daten (UAR) und Informationen des Kopplungsungleichgewichts aus der DRM berücksichtigt. *ASREML* schätzte dabei die maximale Wahrscheinlichkeit (*Maximum Likelihood*) unter Verwendung des folgenden gemischten linearen Modells:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{q} + \mathbf{e} \quad (\text{Model cLDLA1})$$

\mathbf{y} = Vektor der Phänotypen (Schwanzlängen in cm)

$\boldsymbol{\beta}$ = Vektor der fixen Effekte (BW in kg, WH in cm, Geschlecht, Alter und der Gesamtmittelwert μ ; BW- und WH-Daten wurden beide standardisiert und zentriert)

\mathbf{u} = Vektor der zufällig polygenen Effekte (mit $\mathbf{u} \sim N(0, \mathbf{G}\sigma_u^2)$),

\mathbf{q} = Vektor der zufällig additiv-genetischen QTL Effekte (mit $\mathbf{q} \sim N(0, \mathbf{D}_{\text{RM}_p}\sigma_q^2$, wobei \mathbf{D}_{RM_p} die DRM an der Position p eines angenommenen QTLs ist)

\mathbf{e} = Vektor der zufälligen Resteffekte (mit $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2$, wobei \mathbf{I} eine Identitätsmatrix ist)

\mathbf{X} , \mathbf{Z}_1 , \mathbf{Z}_2 = Designmatrizen, die die beobachteten Werte mit den fixen und zufälligen QTL-Effekten verknüpfen.

Es wird angenommen, dass \mathbf{u} , \mathbf{q} und \mathbf{e} nicht miteinander korrelieren.

ECK et al. (2019) schlagen vor, bei Kartierungsstudien zur Entschlüsselung der genetischen Grundlage der Schwanzlänge bei Schafen, die Schwanzlänge auf BW,

WH und der Kombination Wurfgröße \times Geschlecht in einem linearen Modell zu korrigieren. Da zum Zeitpunkt der Phänotypisierung nicht immer eindeutig festgestellt werden konnte, ob das untersuchte Lamm ein Einling oder zusammen mit anderen Lämmern geboren wurde, entschieden wir uns, die Wurfgröße aus dem Modell herauszunehmen und das Geschlecht einzeln zu verwenden. Außerdem empfehlen sie, aufgrund der leichten Handhabe und der Tatsache, dass es zu diesem Zeitpunkt immer noch möglich ist, zwischen natürlich kurzen Schwänzen oder kupierten Schwänzen anhand der Kupierwunde zu unterscheiden, die Messungen und Beprobungen an 5 Wochen alten Lämmern durchzuführen. Um die täglichen Arbeitsabläufe der schafhaltenden Betriebe nicht zu stören wurden teilweise auch Lämmer beprobt welche jünger oder älter waren als 5 Wochen. **Tabelle 5** zeigt neben der Altersspanne auch weitere zusammenfassende Statistiken über das Alter, das BW, WH und die Schwanzlänge aller 362 vermessenen Lämmern.

Tabelle 5: Deskriptive Statistik über alle notierten Merkmale der 362 Lämmer.

Das Alter ist in Tagen angegeben, Körpergewicht in kg und Widerristhöhe und Schwanzlänge in cm

	1.			3.		
	Minimum	Quartil	Median	Mittelwert	Quartil	Maximum
Alter	3.00	25.00	28.00	28.99	35.00	53.00
Körpergewicht	5.30	11.03	13.50	13.86	16.48	26.10
Widerristhöhe	34.00	46.80	49.50	49.42	52.30	57.70
Schwanzlänge	14.20	24.40	27.85	29.12	34.85	40.80

7. Likelihood Ratio Teststatistik

Die von *ASREML* geschätzten Wahrscheinlichkeiten wurden anschließend in einer Likelihood ratio test statistic (*LRT*) verwendet. Die *LRT*-Werte folgen einer χ^2 -Verteilung mit einem Freiheitsgrad (HEUVEN et al., 2005) und wurden wie folgt berechnet:

$$LRT_p = -2 \times (\log(L_0) - \log(L_{1p}))$$

wobei $\log(L_0)$ das Modell der Nullhypothese entspricht, indem die Likelihoods ohne QTL-Effekte geschätzt wurden. $\log(L_{1p})$ entspricht dem Modell der Alternativhypothese und QTL-Effekte wurden hier zur Schätzung der Likelihoods

in der Mitte des Fensters p miteinbezogen. Somit gilt, je höher der LRT -Wert an Position p , desto höher die Wahrscheinlichkeit, dass sich an dieser Position ein QTL befindet.

Die LRT -Statistik berücksichtigt nicht, dass entlang des Chromosoms mehrere Tests durchgeführt wurden. Daher wurde eine Bonferroni-Korrektur durchgeführt um Mehrfachtests aufgrund der 40-SNP-Gleitfenster zu berücksichtigen und eine Signifikanzschwelle zu erhalten. Bei dem zuvor ausgewählten P -Wert von 0.05 führte die Korrektur zu dem P -Wert $< 4.44 \times 10^{-5}$ ($0.05/1127$ wobei 1127 die Anzahl der nicht überlappenden 40-SNP-Fenster ist) und einem entsprechenden LRT -Wert mit genomweiter Signifikanz von 16.67.

Für Positionen deren LRT -Werte höher als die Signifikanzschwelle waren (LRT_{max}), wurde das 2-LOD (logarithm of the odd) Verfahren zur Berechnung des KI angewendet. Dieses 2-LOD KI strebt, wie in biologischen Forschungen üblich, eine mindestens 95 % Konfidenzrate an (VAN OOIJEN, 1992). In dieser Studie also befindet sich die gesuchte kausale Mutation mit einer Wahrscheinlichkeit von über 95 % in diesem 2-LOD KI, dessen Grenzen sich aus der Differenz zwischen dem LRT_{max} und 9.21 errechnen. Ein LOD entspricht 4.605 (VISSCHER und GODDARD, 2004). Bei eng beieinanderliegenden LRT -Peaks wurde angenommen, dass sie zu demselben QTL gehören, wie in MULLER et al. (2017) beschrieben.

8. Weitere cLDLA Modelle

Zusätzlich zu dem oben beschriebenen Modell (cLDLA1) wurden zwei weitere genomweite cLDLAs durchgeführt. Für diese Analysen wurde dieselbe Kartierungspopulation von 362 Tieren verwendet und beinhaltenden zusätzlich die Genotypen eines Kandidatenlocus, welcher während unserer Forschung entdeckt wurde. cLDLA2 enthält nur einen zusätzlichen Markerlocus und ist daher mit Modell 2 der MLMA vergleichbar (siehe unten). In cLDLA3 werden die Genotypen dieses Kandidatenlocus als fixer Effekt betrachtet, d. h. β ist der Vektor μ , Geschlecht, Alter, BW, WH und der Effekt des Kandidatenlocus, wobei Lämmer mit homozygotem ancestralen Genotyp als Klasse 1, heterozygote als Klasse 2 und homozygote abgeleitete als Klasse 3 klassifiziert wurden. Ein umfassender Überblick über die verschiedenen Modelle der cLDLA und der GWAS ist in **Tabelle 6** enthalten.

Tabelle 6: Gemischte lineare Modelle.

Die fixen und zufälligen Effekte sind der Gesamtmittelwert (μ), das Geschlecht (sex), das Alter (age), das Körpergewicht (BW), die Widerristhöhe (WH), Genotypen des Kandidatenlocus (cSNP), der Vektor des additiven Effekts des auf Assoziation zu testenden Kandidatenmarkers (\mathbf{a}), der Vektor der zufälligen polygenen Effekte (\mathbf{u}), der Vektor der zufälligen additiv-genetischen QTL-Effekte (\mathbf{q}) und der Vektor der zufälligen Residualeffekte (\mathbf{e})

Analyse	Model-Name	Effekte		Kommentar
		fixe	zufällige	
cLDLA	cLDLA1	$\mu, \text{sex, age, BW, WH}$	$\mathbf{u, q, e}$	Enthält zusätzlichen Kandidatenlocus als Marker
	cLDLA2	$\mu, \text{sex, age, BW, WH}$	$\mathbf{u, q, e}$	
	cLDLA3	$\mu, \text{sex, age, BW, WH, cSNP}$	$\mathbf{u, q, e}$	Enthält zusätzlichen Kandidatenlocus als fixen Effekt
MLMA	MLMA1	$\mu, \text{sex, age, BW, WH, a}$	$\mathbf{u, e}$	Enthält zusätzlichen Kandidatenlocus als Marker
	MLMA2	$\mu, \text{sex, age, BW, WH, a}$	$\mathbf{u, e}$	

Die Ergebnisse der Kartierungsanalysen wurden mittels dem R-Paket *QQMAN* (TURNER, 2017) grafisch dargestellt.

Um herauszufinden welche Gene sich in unserem KI befinden wurde diese Region mit dem ovinen Refernzgenom Oar_4.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000298735.2) verglichen. Anschließend wurden Gene-Set-Enrichment Analysen mit dem Programm ENRICH durchgeföhrt (Ontologies, MGI Mammalian Phenotype Level 4 2019 (CHEN et al., 2013; KULESHOV et al., 2016; XIE et al., 2021) und eine ausführliche Literaturrecherche begonnen.

9. Genomweite Assoziationsstudie (GWAS)

Zusätzlich zu den cLDLAs wurden zwei Mixed linear model association (MLMA)-Analysen (YANG et al., 2014) durchgeföhrt, um die Ergebnisse miteinander vergleichen zu können. Hierfür wurde das Programm *GCTA* v1.93.2 (YANG et al., 2011) mit dem Ansatz des *leave-one-chromosome-out* verwendet. Dabei wird jenes Chromosom auf welchem sich der Kandidaten-SNP befindet aus der Kalkulation der GRM ausgeschlossen. Die dafür notwendigen .ped und .map- Dateien wurden aus der hauseigenen SQL-Datenbank im *PLINK*-Format (PURCELL ET AL., 2007) bereitgestellt und mit *PLINK* in binary PED Dateien konvertiert. Mit diesem

Eingangsmaterial wurde anschließend die GRM entlang aller Lämmern geschätzt. Anschließend wurde die eigentliche MLMA, bei der jeder SNP für eine statistische Assoziation mit der Schwanzlänge getestet wird, anhand folgenden Modells durchgeführt

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_u\mathbf{u} + \mathbf{e}$$

Wobei \mathbf{y} wieder der Vektor der Schwanzlängen in cm und $\boldsymbol{\beta}$ der Vektor der fixen Effekte ist (BW in kg, WH in cm (beide wieder standardisiert und zentriert), Geschlecht, Alter und der Mittelwert). \mathbf{a} ist der Vektor der additiv-genetischen Effekte, also der fixen Effekte des Kandidaten-SNPs, der für eine Assoziation getestet wird, \mathbf{u} ist der Vektor der polygenen Effekte und \mathbf{e} ist der Vektor der zufälligen Residualeffekte. Es wird angenommen, dass \mathbf{u} und \mathbf{e} nicht miteinander korrelieren. \mathbf{X} , \mathbf{Z}_a und \mathbf{Z}_u sind Inzidenzmatrizen für $\boldsymbol{\beta}$, \mathbf{a} und \mathbf{u} .

Die suggestive Signifikanzgrenze wurde bei $P < 1/N$ gesetzt und die genomweite bei $P < 0.05/N$ (XIONG et al., 2015), bei dieser Korrektur auf Bonferroni steht N für die Anzahl der berücksichtigten Marker, welche 45,114 betrug. Dadurch ergab sich ein suggestiv P -Wert von 2.22×10^{-5} und ein genomweiter von 1.11×10^{-6} . In der zweiten MLMA wurde der bereits in Abschnitt III.8 genannte Kandidatenlocus berücksichtigt. Obwohl sich die Anzahl der Marker auf 45,115 erhöhte, blieb die suggestiv und genomweite Grenze auf zwei Kommastellen gerundet gleich.

Mittels des Programmes *GCTA* v1.93.2 wurde ebenfalls die Heritabilität der Länge des Schwanzes bei Merinolandschafen geschätzt. Mittels eines *Genomic-Relatedness-Based restricted maximum-likelihood* (GREML) Ansatzes, wurde die SNP-basierende Heritabilität berechnet (YANG et al., 2010). Diese entspricht jenem Anteil der Schwanzlängenvarianz welcher durch sämtliche in die Analyse eingegangen SNPs erklärt werden kann.

10. Auswahl der zu sequenzierenden Lämmer

Eine zur cLDLA ebenfalls durchgeführte Diplotypeffekt-Analyse mittels *ASREML* schätzte an den Loci mit den signifikantesten Assoziationen (*LRTmax*) die Effektgröße für jeden in unserer Population vorhandenen Diplotypen. Diese Diplotypen bestanden aus maternalen und paternalen Haplotypen und diese wiederum aus jeweils 40 Markern. Eine Sortierung nach den Diplotypeffekten auf

dem Schafchromosom (OAR) 11 zeigte eine starke Korrelation mit der Schwanzlänge. Dies wurde mittels einer eigens durchgeführten Regressionsanalyse anhand der Funktion *lm* in *R* bestätigt. Ein Zusammenhang zwischen dem zweiten *LRTmax* auf OAR2 und der Schwanzlänge wurde ebenfalls mittels einer Regressionsanalyse ausgewertet. Anhand der Diplotypeffekte, des Alters und der Schwanzlänge wurden 48 (23 lang- und 25 kurzschwänzige) von den insgesamt 362 Lämmer mittels *Target capture* Sequenzierung rund um beide *LRTmax* sequenziert. Die sequenzierte Region auf OAR2 erstreckte sich von Basenpaar 93,200,000 bis Basenpaar 96,700,000 und auf OAR11 von 36,600,000 bis 37,900,000 bp.

11. DNA-Extrahierung, Sequenzierung und Analysen der Sequenzen

Die DNA wurde aus den Blutproben mittels den ReliaPrep™ Blood gDNA Miniprep System extrahiert.

Whole-Genome-Sequencing (WGS)- Libraries wurden aus 250 ng genomischer DNA durch Tagmentierung mit dem NexteraFlex-Kit von Illumina vorbereitet. Anschließend wurden die Bibliotheken dual-barcodiert und mittels PCR amplifiziert, mit SPRI-Beads gereinigt und in äquimolaren Mengen gepoolt. Die gepoolten Libraries wurden für die Zielregionen durch Hybridisierung mit einem Agilent-Capture-Array mit 244k Oligo-Spots angereichert. Die Oligosonden wurden aus der wiederholungsmaskierten DNA-Sequenz ausgewählt, als alle möglichen 60mer, die sich nicht mit den wiederholungsmaskierten Basen überlappen und die in 15-nt-Kachelschritten zu ihren Nachbarn versetzt sind. Nach 65 Stunden Hybridisierung in Gegenwart von cot-I Schaf-DNA und Adapter-Blocking-Oligos bei 65 °C wurde der Capture Array gewaschen und die eingefangenen Bibliotheksmoleküle wurden bei 95 °C für 10 Minuten in einem Volumen von 500 µl Wasser (DNA-Gütegrad) eluiert. Die angereicherten Libraries wurden dann durch PCR amplifiziert, auf dem Bioanalyzer analysiert und im 2*110 bp Paired-End-Modus auf einer P2-Fließzelle eines NextSeq1000-Sequenzers von Illumina sequenziert.

Die in FASTQ-Dateien enthaltenen Sequenzen wurden anhand folgender Programme bearbeitet: *SICKLE* (JOSHI NA, 2011) wurde zum Trimmen der Adaptern und zum Herausfiltern von Sequenzen geringer Qualität verwendet. *FASTQC* (BABRAHAM, 2011) bewertete die Qualitätsparameter der gefilterten

Sequenzierdaten. Die gefilterten Reads wurden mit den Standardparametern des *BWA-MEM* Alignment-Tools (LI und DURBIN, 2009) auf das Schaf-Referenzgenom OAR_v4.0 gemappt. Zur Konvertierung der SAM-Dateien in koordinatensortierte BAM-Dateien und zur Entfernung der duplizierten Reads wurden *SAMTOOLS* (LI et al., 2009) und *PICARD* (BROAD INSTITUTE, 2019) verwendet. Die Rekalibrierung der Basenqualität und das Indel-Realignment wurden mit *GATK* (AUWERA UND O'CONNOR, 2020) durchgeführt.

Anschließend wurden SNPs und Indels mittels dem *GATK HAPLOTYPECALLER* (AUWERA und O'CONNOR, 2020) abgerufen. Diesbezüglich verwendeten wir nicht die Sequenzen der einzelnen Lämmer, sondern:

- (i) zwei Sequenzpools, die die kurz- und langschwänzigen Gruppen der im Rahmen dieser Doktorarbeit sequenzierten 48 Merinolandschafslämmer repräsentieren.
- (ii) 16 Genomen von Hausschafen (jeweils 8 Proben von kurz- und langschwänzigen Rassen) und
- (iii) 37 Genomen von Wildschafen.

Die Reads der Merinolandschafslämmer wurden gepoolt, um die Sequenzierungsabdeckung beider Gruppen zu erhöhen. Die Genome der Wild- und Hausschafe, die aus dem NCBI sequence read archive (SRA) heruntergeladen wurden, sind in **Tabelle 7** aufgeführt. Um die erhaltenen *GATK* - short variants zu filtern, wurde folgender Ablauf eingehalten:

- i) Das ancestrale Allel wurde anhand von 37 Wildschafproben (**Tabelle 7**) bestimmt; dabei wurde das häufigste Allel als das ancestrale festgelegt.
- ii) Nach der erfolgreichen Bestimmung des ancestralen Alleles in Schritt i) wurde geprüft, ob die Anzahl der Wildschafe, die das abgeleitete Allel enthalten, kleiner oder gleich 5 ist. Dieser weiche Schwellenwert für das abgeleitete Allel wurde so festgelegt, um die Möglichkeit der Introgression zwischen Hausschafen und Wildschafen zu berücksichtigen.
- iii) Ermittlung ob die Anzahl der Reads, welche das Referenzallel tragen, und die Anzahl der Reads, die das alternative Allele tragen, signifikant unterschiedlich sind (Fisher-exact-Test, P -Wert < 0.05 ; nach Korrektur

der mehrfachen Testung) zwischen der gepoolten Probe der Kurzschwanz- und der Langschwanzgruppe.

- iv) Ermittlung ob die Anzahl der Allele, welche als ancestrally definiert wurden und die Anzahl der Allele, welche als abgeleitet definiert wurden, signifikant unterschiedlich (Fisher-exact-Test, P-Wert < 0.05; nach Korrektur der mehrfachen Testung) zwischen den NCBI-Proben der domestizierten Langschwanz- und Kurzschwanzrassen sind.
- v) Anschließend wurde geprüft, ob die Anzahl der Reads, die abgeleitete Allele unterstützen, signifikant höher ist als die Anzahl der ancestralen Allele in der gepoolten Stichprobe des langschwänzigen Pools.

Nach Anwendung dieser 5 Kriterien blieben 19 Varianten über, die in der Tabelle 8 aufgelistet sind. Wir untersuchten dann diese 19 Varianten in unseren 48 sequenzierten Lämmern separat. Da ihre BAM-Dateien nur eine geringe Abdeckung aufwiesen, wählten wir nur die Varianten aus, die zu 90 % in Individuen vorhanden waren, die an dieser bestimmten Position in einer Gruppe gemappte Reads hatten und nicht mehr als 10 % in der anderen Gruppe. Zusätzlich wurden alle 19 Varianten mit dem Ensembl Variant Effect Predictor (MCLAREN et al., 2016) nach dem Remapping auf Oar_3.1 mit NCBI Remap bestimmt (die Ergebnisse der Bestimmung sind in der Tabelle 8 aufgeführt).

Tabelle 7: Informationen zu den für die Mutationsdetektion verwendeten Schafen.

Run numbers und BioSample ID von WGS aus verschiedenen Schafrassen und deren Genotyp für die Kandidatenvarianten der „Structural Variation“ (SV) und den SNP. Diese Schafe wurden unter anderem auch für das Abrufen von SNPs und kleinen Indels verwendet. Genotypen sind homozygot anzestral (A/A), homozygot abgeleitet (D/D) und heterozygot (A/D)

Rasse	Probe		Genotyp		Veröffentlicht von
	Run number	BioSample ID	SV	SNP	
<i>Ovis gmelini</i>	ERR157938	SAMEA2012637	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR157930	SAMEA2012638	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR157939	SAMEA2012639	A/A	A/D	Genoscope *
<i>Ovis gmelini</i>	ERR157942	SAMEA2012640	A/A	A/D	Genoscope *
<i>Ovis gmelini</i>	ERR157931	SAMEA2012641	A/A	D/D	Genoscope *
<i>Ovis gmelini</i>	ERR157932	SAMEA2012642	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR157944	SAMEA1967031	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR157935	SAMEA2012643	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR332589	SAMEA2065600	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR332575	SAMEA2065601	A/A	D/D	Genoscope *
<i>Ovis gmelini</i>	ERR332587	SAMEA2065602	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR332582	SAMEA2065603	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR332573	SAMEA1972234	A/A	A/D	Genoscope *
<i>Ovis gmelini</i>	ERR315509	SAMEA2065604	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR466546	SAMEA2395410	A/A	A/A	Genoscope *
<i>Ovis gmelini</i>	ERR466544	SAMEA2395411	A/A	A/A	Genoscope *
Finnschaf	SRR11657543	SAMN14590314	A/A	A/A	LI et al. (2020)
Finnschaf	SRR11657544	SAMN14590313	A/D	A/D	LI et al. (2020)
Finnschaf	SRR11657545	SAMN14590312	A/A	A/A	LI et al. (2020)
Finnschaf	SRR11657546	SAMN14590311	A/A	A/A	LI et al. (2020)
Romanov	SRR12396891	SAMN15517583	A/A	A/D	DENG et al. (2020)
Romanov	SRR4291219	SAMN05216760	A/A	A/A	HEATON et al. (2017)
Romanov	SRR4291223	SAMN05216759	A/A	D/D	HEATON et al. (2017)
Romanov	SRR4291160	SAMN05216766	A/A	A/A	HEATON et al. (2017)
Weißes Alpenschaf	ERR3086436	SAMEA5239874	D/D	D/D	University of Bern**
Weißes Alpenschaf	ERR3086440	SAMEA5239878	D/D	D/D	University of Bern**
Weißes Alpenschaf	ERR3086476	SAMEA5239914	A/D	A/D	University of Bern**
Weißes Alpenschaf	ERR3086477	SAMEA5239915	A/D	A/D	University of Bern**
Rambouillet	SRR4291242	SAMN05216757	D/D	D/D	HEATON et al. (2017)
Rambouillet	SRR4291257	SAMN05216755	***	D/D	HEATON et al. (2017)
Rambouillet	SRR4291268	SAMN05216753	D/D	D/D	HEATON et al. (2017)
Rambouillet	SRR6305143	SAMEA104496890	A/D	A/D	Baylor College of Med.
8,000 Jahre altes Schaf	ERR3861593	SAMEA6516192	***	A/A	YURTMAN et al. (2021)
8,000 Jahre altes Schaf DNA Seq	ERR3861592	SAMEA6516191	***	A/A	YURTMAN et al. (2021)

Tabelle 7: Fortsetzung

Rasse	Probe		Genotyp		Veröffentlicht von
	Run number	BioSample ID	SV	SNP	
<i>Ovis ammon</i>	SRR8560952	SAMN10915547	A/A	A/A	CAAS****
<i>Ovis ammon</i>	SRR8560953	SAMN10915548	A/A	A/A	CAAS****
<i>Ovis ammon</i>	SRR9222805	SAMN11979390	A/A	A/A	CAAS****
<i>Ovis ammon</i>	SRR9222806	SAMN11979389	A/A	A/A	CAAS****
<i>Ovis ammon</i>	SRR9222807	SAMN11979391	A/A	A/A	CAAS****
<i>Ovis canadensis</i>	SRR501858	SAMN01000748	A/A	A/A	Baylor College of Med.
<i>Ovis canadensis</i>	SRR501895	SAMN01000746	A/A	A/A	Baylor College of Med.
<i>Ovis canadensis</i>	SRR501898	SAMN01000747	A/A	A/A	Baylor College of Med.
<i>Ovis dalli</i>	SRR501847	SAMN01000785	A/A	A/A	Baylor College of Med.
<i>Ovis dalli</i>	SRR501897	SAMN01000764	A/A	A/A	Baylor College of Med.
<i>Ovis vignei</i>	ERR454945	SAMEA2358291	A/A	A/A	Genoscope *
<i>Ovis vignei</i>	ERR454946	SAMEA2358287	A/A	A/A	Genoscope *
<i>Ovis vignei</i>	ERR454947	SAMEA2358290	A/A	A/A	Genoscope *
<i>Ovis vignei</i>	ERR454948	SAMEA2358289	A/A	A/A	Genoscope *
<i>Ovis vignei</i>	ERR454950	SAMEA2358291	A/A	A/A	Genoscope *
<i>Ovis nivicola</i>	ERR4161992	SAMEA6833340	A/A	A/A	UPADHYAY et al. (2020)
<i>Ovis nivicola</i>	ERR6667562	SAMEA8657697	A/A	A/A	UPADHYAY et al. (2021)
<i>Ovis nivicola</i>	ERR6668200	SAMEA8657699	A/A	A/A	UPADHYAY et al. (2021)
<i>Ovis nivicola</i>	ERR6668794	SAMEA8657700	A/A	A/A	UPADHYAY et al. (2021)
<i>Ovis nivicola</i>	ERR6667561	SAMEA8657698	A/A	A/A	UPADHYAY et al. (2021)
<i>Ovis nivicola</i>	ERR5858461	SAMEA8657696	A/A	A/A	UPADHYAY et al. (2021)

* Sequenziert im Rahmen des NextGen Projektes; ** Institut für Genetik; *** Keine gemappten Reads vorhanden; **** Institut des Animal Science of CAAS

Tabelle 8: Die 19 herausgefilterten kleinen Mutationen nach Anwendung der im Abschnitt III.11 beschriebenen Filterkriterien.

Spalten „ST_ref“, „ST_var“, „LT_ref“ und „LT_var“ zeigen die Anzahl der kurz- und langschwänzigen Lämmer mit ancestralen und abgeleiteten Allelen (ST_ref: Anzahl der kurzschwänzigen Tiere mit homozygotem ancestralen Genotyp; ST_var: Anzahl der kurzschwänzigen Tiere mit entweder heterozygotem oder homozygotem abgeleiteten Genotyp; LT_ref: Anzahl der langschwänzigen Tiere mit homozygotem ancestralen Genotyp; LT_var: Anzahl der langschwänzigen Tiere mit entweder heterozygotem oder homozygotem abgeleiteten Genotyp. Außerdem werden die Varianten-IDs und die Ergebnisse des Ensembl Variant Effect Predictor angezeigt. Die Bedeutung der Spalte „Konsequenz“ und der Spalte „Auswirkung“ sind beschrieben auf: <https://www.ensembl.org/Help/Glossary?id=535>

Position	ST_ref	ST_var	LT_ref	LT_var	Variant ID	Konsequenz	Auswirkung	SIFT
36921089	3	5	10	2	rs420125730	intron_variant (B4GALNT2)	MODIFIER	-
37079829	10	7	3	13	rs426813381	downstream_gene_variant (GIP)	MODIFIER	-
37081200	7	1	1	7	rs413378768	downstream_gene_variant (GIP)	MODIFIER	-
37081999	19	1	7	8	rs421540535	downstream_gene_variant (GIP)	MODIFIER	-
37082002	18	1	7	7	rs399930658	downstream_gene_variant (GIP)	MODIFIER	-
37274710	7	7	15	1	rs406224770	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37275045	14	9	0	18	rs424357928	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37275989	14	8	0	18	rs407002893	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37277182	16	6	0	18	rs410624664	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37278058	15	7	0	19	rs429611961	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37278438	5	19	21	1	rs419262514	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37278757	15	8	0	19	rs398973702	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37278805	15	7	0	17	rs414342629	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37279115	15	7	0	21	rs405338697	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37279536	16	5	0	19	rs430325358	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37279822	11	2	1	7	rs421532300	intergenic_variant (TTLL6-HOXB13)	MODIFIER	-
37285884	13	7	0	17	rs414085759	upstream_gene_variant (HOXB13)	MODIFIER	-
37288535	12	9	0	17	rs413678645	upstream_gene_variant (HOXB13)	MODIFIER	-
37290361	17	2	0	15	rs413316737	missense_variant (HOXB13)	MODERATE	0.54

Um größere strukturelle Varianten (SV) zu entdecken, wurde *SMOOVE (LUMPY)* (LAYER et al., 2014) und *DELLY* (RAUSCH ET AL., 2012) mit Standardparametern in den gepoolten Reads der kurz- und langschwänzigen Lämmern verwendet. Um sicherzustellen, dass wir keine Kandidatenvarianten übersehen, wurde eine visuelle Prüfung der erfassten Regionen mit *JBROWSE* (BUELS et al., 2016) durchgeführt. Dabei wurde der Focus auf die Region des hier entdeckten Kandidatengens gelegt.

12. Validierung von Kandidatenmutationen mittels gezielter Genotypisierung

Detektierte Kandidatenmutationen wurden im institutseigenen Labor an allen 362 beprobten Lämmern und deren 17 bestätigten Vätern mittels folgenden Ansätzen auf ihre Plausibilität hin überprüft.

12.1. Validierung des Kandidaten-SNPs mittels PCR

Für einen, während unserer Studie entdeckten SNP führten wir die Genotypisierung mittels PCR-RFLP und Elektrophorese auf einem 2 %igen Agarosegel durch. Mittels *PRIMER3* (UNTERGASSER ET AL., 2012) wurden folgende PCR-Primer konzipiert:

- Forward: TTTAAAACGCTTTGGATT
- Reverse: CACTCGGCAGGAGTAGTA

Diese Primer dienten als Start- und Endpunkt unserer zu amplifizierenden Region (37,290,230 bp - 37,290,637 bp auf dem OAR11), welche in Abbildung 7 zu sehen ist. Das verwendete Restriktionsenzym war BsrI. Es handelt sich um ein Restriktionsenzym vom Typ IIS, welches asymmetrische DNA-Sequenzen erkennt. In dieser Studie wurde die mutierte Sequenz $\rightarrow 3' \dots \text{TGAC/CN} \dots 5' \leftarrow$ erkannt. Die Base G repräsentiert die Mutation und ersetzt die ursprüngliche Base C. Das „/“ zeigt die Position, an der das Restriktionsenzym schneidet. Die DNA-Amplifikation wurde in 35 Zyklen durchgeführt. Die gesamte Reaktionsmischung bestand aus 15.0 μl und enthielt 3.0 μl 5X-Puffer, 1.5 μl dNTPs (10 mM), 0.6 μl jeweils 10 μM Vorwärts- und Rückwärtsprimer, 1.0 μl DNA (15 ng/ μl), 0.07 μl GoTaq®G2 DNA Polymerase (Promega, Madison, Wisconsin, USA) und destilliertes Wasser. 1.5 U des Enzyms BsrI, 3.0 μl DNA (PCR-Produkt), 2 μl Cut Smart Buffer und destilliertes Wasser wurden verwendet für ein Gesamtreaktionsvolumen von 20 μl . Die Reaktionsmischung wurde anschließend

für 3 h bei 65 °C inkubiert. Im letzten Schritt trennten wir die DNA-Fragmente nach Größe und machten sie durch GelRed™-gefärbte Agarosegel-Elektrophorese sichtbar. Es wurden nur Sequenzen geschnitten, die das abgeleitete Allel (SNP G) enthielten. Die beiden resultierenden Fragmente hatten eine Länge von 120 bp und 259 bp. Die Sequenz mit dem ursprünglichen Allel (SNP C) behielt ihre Länge von 379 bp bei.

```
TGCCAGCGATTTAAAACGCTTTGGATTCCCCGGGCCTGGGTGGGGAGAGAGAGCT
GGGTGCCCCCTGTATCCCCACCCCAGCGCCTCATGAGCCGATCCTCGGCCCCA
TGGAGCCCGCAATTATAC/CACTTTGGACGGCGCCAAGGAGATTGAAGGCTTGC
TGGGAGCTGGAGGGAGTCGGAACCTGGTCACCCACTCGCCACTGACCAGCCATCC
AGCGTCGGCGCCTACGCTGTTGCCTGCTGTCAACTACGGCCCCCTGGATCTGCCA
GGCTCTGCGGAACCGCCAAAGCAGTGCCACCCATGTCCCGGGATGCCCCAGGGGG
CGTCCCCAGCTCCTGTGCCTCACGGCTACTTTGGAGGCGGGTACTACTCCTGCCG
AGTGTCTCGAAGCTCGCTGAAGCC
```

Abbildung 7: Basensequenz der Region 37,290,221-37,290,657 auf OAR11 (Oar_v4.0).

Der Sequenzausschnitt beinhaltet die amplifizierte Region, welche vom Primerpaar flankiert wird (rot und unterstrichen). Die Erkennungssequenz vom Restriktionsenzym ist fett markiert und beinhaltet den SNP an der Position 37,290,361 (unterstrichen), wenn diese Position die Mutation G trägt wird an der Stelle mit dem „/“ geschnitten

12.2. Validierung der Kandidaten-Insertion mittels PCR und Sanger-Sequenzierung

Zusätzlich zu dem oben erwähnten SNP wurde eine Insertion entdeckt, welche ebenfalls mittels PCR und Elektrophorese auf einem 2%igen Agarosegel genotypisiert wurde. Mehrere unterschiedliche PCR-Primersequenzen, die mittels *PRIMER3* (UNTERGASSER et al., 2012) entworfen wurden, wurden getestet da die Amplifizierungsergebnisse im Labor bei den meisten nicht zufriedenstellend waren. In **Tabelle 9** sind alle verwendeten Primer aufgelistet. Jenes Primerpaar, welche die qualitativsten Amplikons erzeugte, waren:

- Forward: TTTATGAGCTTCTCTCCGCCA und
- Reverse: CACTCGGCAGGAGTAGTA

Dieses Primerpaar amplifiziert die in **Abbildung 8** dargestellte Region. Die DNA-Amplifikation wurde mit 35 Zyklen durchgeführt. Die gesamte Reaktionsmischung umfasste 25.0 µl, bestehend aus 5.0 µl 5X-Puffer, 2.5 µl dNTPs (10 mM), jeweils 1 µl 10 µM Forward- und Reverseprimer, 1.0 µl DNA (15 ng/µl), 0.07 µl

GoTaq®G2 DNA Polymerase (Promega, Madison, Wisconsin, USA) und destilliertes Wasser. Im letzten Schritt trennten wir die Amplikons nach Größe und machten sie durch GelRed™-gefärbte Agarosegel-Elektrophorese sichtbar.

Tabelle 9: Getestete Primerpaare.

Das erste Primerpaar wurde zur Genotypisierung aller 362 Lämmer und ihrer Väter des SNPs rs413316737 verwendet, das zweite zur Genotypisierung der Insertion, die sich in der Promotorregion von HOXB13 befindet. Alle anderen Primerpaare, die für die Genotypisierung der Insertion bestellt wurden, erfüllten nicht unsere Qualitätskriterien

Genotypisierte Mutation	Forward Primer	Reverse Primer
SNP	TTTAAAACGCTTTGGATT	CACTCGGCAGGAGTAGTA
Insertion	TTTATGAGCTTCTCTCCGCCA	CACTCGGCAGGAGTAGTA
Insertion	TTTATGAGCTTCTCTCCGCCA	GACACTCGGCAGGAGTAGTAC
Insertion	GGAAGGAGGAGGGAGTGAGC	CTCCCAGCAAGCCTTCAATC
Insertion (Verwendet in LI et al. (2022))	TTTATGAGCTTCTCTCCGCCA	AAGTGGTATAATTGCCGGGCT
Insertion	TTTATGAGCTTCTCTCCGCCA	TGGTATAATTGCCGGGCTCC

CAGGGGTAAAGTATTTTCGCAATTTCTGCAGTGAAGATTTTATGAGCTTCTCTCCGCCAGGCCGCAGCCAATCAGCCCGCGAGCCCGGGCACCTGCGGCTCCTGCGTCACGACGGCCGGGCTGAGCGAATGCAGGCGCCAGCGAG**CTGCCAGCGATTTA** [TTATTTTTTTTTTTTAAAATTTTTATTTATTTATTTGGCTGCTCTGGGTCTTAGTTGTAGCATGTGGGATCTAGCTCCCTGACCAGGGATCGAACCTGGGCCCCCTGCACTGGGAGCTCAGAGTCTTAGCCACTGGACCACCAGGGAATCC**CTGCCAGCGATTTA**] AAACGCTTTGGATTCCCCGGGCCTGGGTGGGGAGAGAGAGCTGGGTGCCCCCTGTATCCCCACCTCCAGCGCCTCATGAGCCGATCCTCGGCCCATGGAGCCCGGCAATATATACTTTGGACGGCGCCAAGGAGATTGAAGGCTTGCTGGGAGCTGGAGGGAGTCGGAACCTGGTCACCCACTCGCCACTGACCAGCCATCCAGCGTCGGCGCCTACGCTGTTGCCTGCTGTCAACTACGGCCCCCTGGATCTGCCAGGCTCTGCGGAACCCGCCAAAGCAGTGCCACCCATGTCCCAGGATGCCCCAGGGGGCGTCCCAGCTCCAGTGCCTTACGGCTACTTTGGAGGCGGGTACTACTCCTGCCGAGTGTCTCGAAGCTCGCTGAAGCCCTGTGCCAGGCGGCCACCCTGGCCGCCT

Abbildung 8: Basensequenz der Region 37,524,721-37,525,557 auf OAR11 (ARS-UI_Ramb_v2.0).

Der Sequenzausschnitt beinhaltet die amplifizierte Region, welche vom Primerpaar flankiert wird (rot und unterstrichen). Ein Teil der genotypisierten Schafe enthält eine Insertionssequenz (in eckige Klammern gesetzt, [Insertion]). Beachte, dass hier das aktuelle Referenzgenom ARS-UI_Ramb_v2.0, aufgrund eines während der Studie detektierten Assemblyfehlers verwendet wurde, auf diesen Assemblyfehler wird später noch näher eingegangen. Beachte ebenfalls, die idente Sequenz welche die Insertion flankiert (fett markiert).

13. Resequenzierung unserer Kandidatenmutationen mittels Sanger Sequencing

Zwei Lämmer, die homozygot für die SV sind, ein Lamm, das homozygot für das ancestrale Allel ist und zwei heterozygote Lämmer wurden mittels Sanger-Sequenzierung mit den in III.12.2 erwähnten Primern neu sequenziert. Die Amplikons wurden mit der *Cycle Sequencing* Technologie (dideoxy chain termination / cycle sequencing) auf ABI 3730XL Sequenziermaschinen (Eurofins Genomics, Deutschland) gewonnen. Die sequenzierten Daten wurden mit der Software *SNAPGENE* (von Insightful Science; erhältlich unter <https://www.snapgene.com/>) analysiert.

IV. ERGEBNISSE

1. Initiale cLDLA

Die erste haplotypbasierte cLDLA Kartierung erzielte zwei genomweit signifikante QTLs, welche mit der Schwanzlänge beim Merinolandschaf assoziiert sind. Der größte und eindeutigste Peak befand sich hierbei auf OAR11 an der Position 37,111,462 bp mit einem LRT_{max} von 29.460, das entspricht einen P -Wert von $P = 5.71 \cdot 10^{-8}$ (Bonferroni korrigiert: $P = 6.43 \cdot 10^{-5}$). Der zweite genomweit signifikante Peak war weniger scharf und unspezifisch. Er befand sich auf OAR2 an der Position 94,538,115 bp mit einem LRT_{max} von 19.356, das entspricht einen P -Wert von $P = 1.08 \cdot 10^{-5}$ (Bonferroni korrigiert: $P = 1.22 \cdot 10^{-2}$). Die LRT -Werte sämtlicher getesteter Positionen auf OAR11 sind in Abbildung 9 grafisch dargestellt, jene auf OAR2 in Abbildung 10.

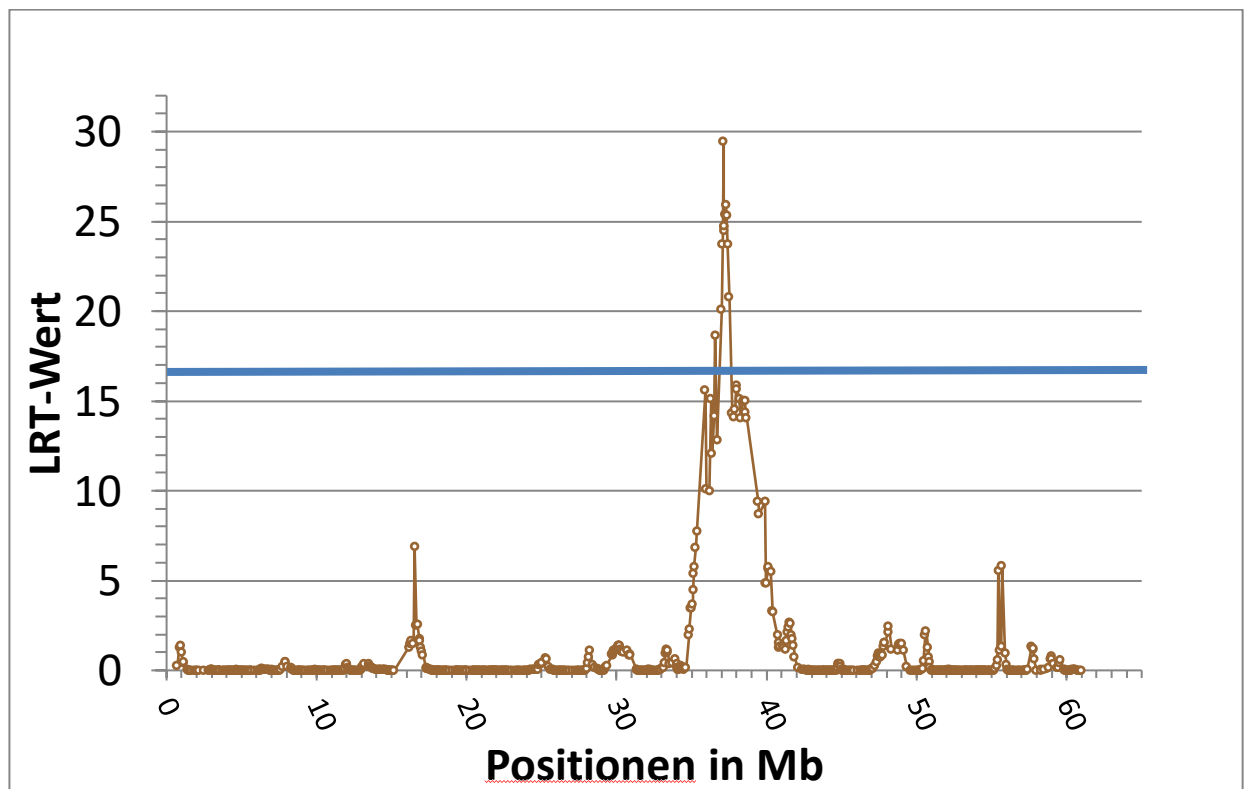


Abbildung 9: LRT-Werte auf OAR11.

Der höchste LRT -Wert befindet sich auf Position 37,111,462 bp mit einem Wert von 29.460. Die blaue Linie stellt die genomweite Signifikanzschwelle dar.

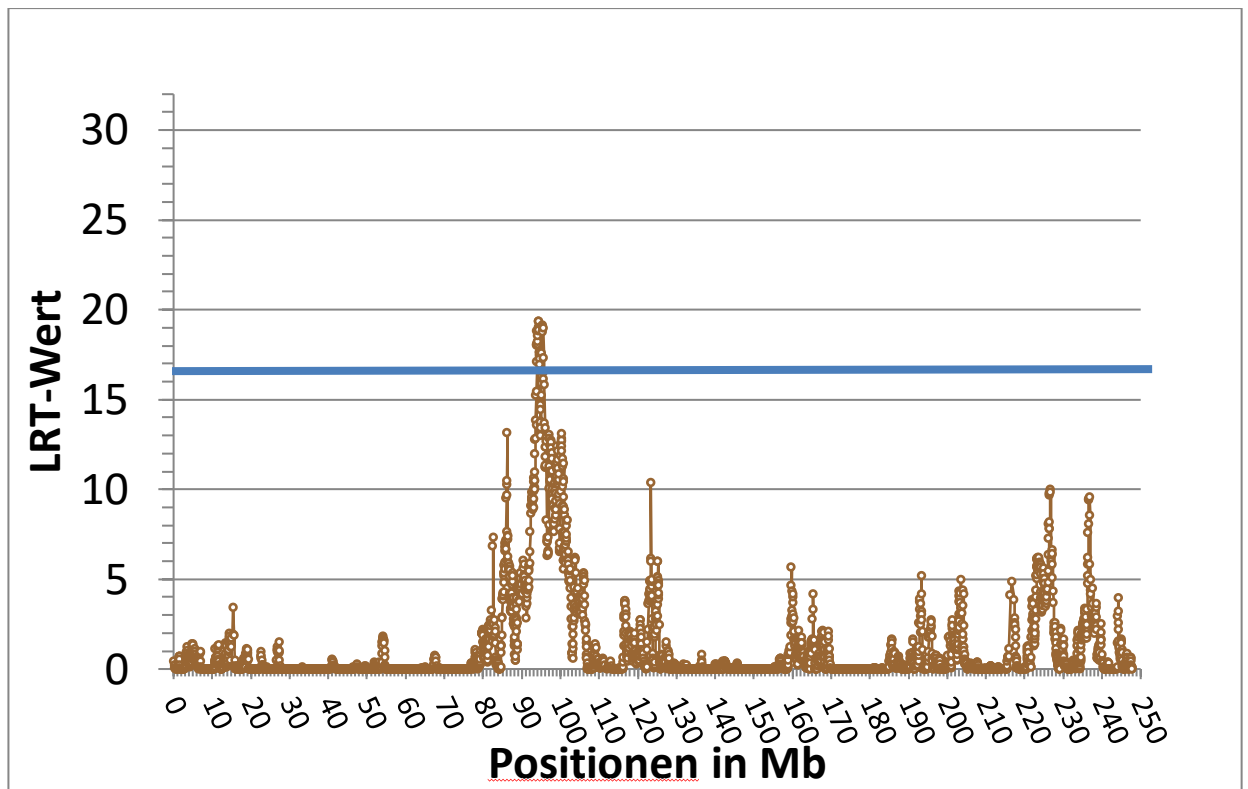


Abbildung 10: LRT-Werte auf OAR2.

Der höchste LRT-Wert befindet sich auf Position 94,538,115 bp mit einem Wert von 19.356. Dieser Peak ist weniger hoch und weniger scharf als jener auf OAR11. Die blaue Linie stellt die genomweite Signifikanzschwelle dar.

Chromosomweite signifikante Peaks befanden sich außerdem auf OAR2, OAR3, OAR10, OAR14 und OAR17.

Das 2-LOD KI wurde für das LRT_{max} auf OAR11 zwischen 37,000,925 bp und 37,521,490 bp und für das LRT_{max} auf OAR2 zwischen 93,441,900 bp und 96,402,884 bp berechnet. Laut dem Genome Browser auf NCBI für das Referenzgenom Oar_v4.0 befinden sich in diesen KI die in **Tabelle 10** und **Tabelle 11** gelisteten Genen.

Tabelle 10: Gene im Konfidenzintervall auf OAR2.

Zwischen den Positionen 93,441,900 bp und 96,402,884 bp, der genomweit signifikante Peak von cLDLA1 ist ebenfalls gezeigt.

Gene	Name	Start (bp)	Ende (bp)
<i>CAAP1</i>	caspase activity and apoptosis inhibitor 1	94,480,790	94,556,081
QTL-Peak	Peak der cLDLA1		94,538,115
<i>PLAA</i>	phospholipase A2 activating protein	94,567,698	94,605,168
<i>IFT74</i>	intraflagellar transport 74	94,607,523	94,708,596
<i>LRRC19</i>	leucine rich repeat containing 19	94,629,884	94,643,326
<i>TEK</i>	TEK receptor tyrosine kinase	94,739,266	94,838,005
<i>EQTN</i>	Equatorin	94,900,676	94,917,170
<i>MOB3B</i>	MOB kinase activator 3B	94,934,882	95,077,788
<i>IFNK</i>	interferon kappa	95,145,920	95,147,256
<i>C9ORF72</i>	C9orf72-SMCR8 complex subunit	95,183,641	95,202,857
<i>LINGO2</i>	leucine rich repeat and Ig domain containing 2	95,627,231	95,629,051

Tabelle 11: Gene im Konfidenzintervall auf OAR11.

Zwischen den Positionen 37,000,925 bp und 37,521,490 bp, die genomweit signifikanten Peaks von cLDLA1 und 2 sind ebenfalls gezeigt

Gene	Name	Start (bp)	End (bp)
<i>IGF2BP1</i>	insulin like growth factor 2 mRNA binding protein 1	37,009,695	37,032,963
<i>GIP</i>	gastric inhibitory polypeptide	37,074,497	37,079,791
<i>SNF8</i>	SNF8, ESCRT-II complex subunit	37,093,635	37,101,565
<i>UBE2Z</i>	ubiquitin conjugating enzyme E2 Z	37,104,244	37,117,532
QTL-Peak	Peak der cLDLA1		37,111,462
<i>ATP5G1</i>	ATP synthase	37,125,809	37,128,277
<i>CALCOCO2</i>	calcium binding and coiled-coil domain 2	37,179,738	37,202,548
<i>TLL6</i>	tubulin tyrosine ligase like 6	37,221,445	37,263,708
<i>HOXB13</i>	homeobox B13	37,290,203	37,292,513
QTL-Peak	Peak der cLDLA2		37,311,842
<i>HOXB9</i>	homeobox B9	37,365,122	37,369,564
<i>HOXB8</i>	homeobox B8	37,376,709	37,378,201
<i>HOXB7</i>	homeobox B7	37,381,261	37,384,010
<i>HOXB6</i>	homeobox B6	37,391,399	37,395,989
<i>HOXB5</i>	homeobox B5	37,397,814	37,404,968
<i>HOXB3</i>	homeobox B3	37,437,808	37,439,623
<i>HOXB2</i>	homeobox B2	37,445,188	37,448,075
<i>HOXB1</i>	homeobox B1	37,457,621	37,460,240

Die anschließende Literaturrecherche offenbarte eine hochgradige Assoziation für den Homeobox-B-Gencluster (OAR11:37,290,203-37,460,240) auf OAR11 mit der Ausprägung einzelner Körpermerkmale. *HOXB13* ist bereits als Kandidatengene für die Schwanzlänge in anderen Spezies beschrieben und ist nur 179 kb vom *LRTmax* aus Modell cLDLA1 entfernt.

2. Initiale MLMA

Die GCTA-GREML Analyse berechnete eine sehr hohe Heritabilität der Schwanzlänge von 0.992 (bei einem Standardfehler von 0.12). Dadurch lässt sich ein sehr hoher Varianzanteil, der für die Schwanzlänge beim Merinolandschaf verantwortlich ist, alleine durch die genomweiten SNPs erklären. Trotz der hohen Heritabilität war die MLMA nicht in der Lage genomweit signifikante Assoziationen zwischen SNPs und untersuchten Merkmal zu enthüllen. **Abbildung 11** zeigt die Ergebnisse von MLMA1, MLMA2, cLDLA1 und cLDLA3 als Manhattan-Plot, die 4 signifikantesten SNPs der MLMA1 sind mit deren *P*-Werten im ersten Panel dieser Abbildung dargestellt. Die Ergebnisse die in **Abbildung 11** dargestellt sind, sind auch als Zoom-Plots für die die Region 35 Megabasenpaar (Mb) bis 40 Mb auf OAR11, in **Abbildung 12** zu sehen. Hier sind die Veränderungen in den Ergebnissen der einzelnen Analysen herangezoomt. Zusätzlich sind auf Panel e) die Gene des KI auf OAR11 entsprechend Ihrer Lage und Größe dargestellt.

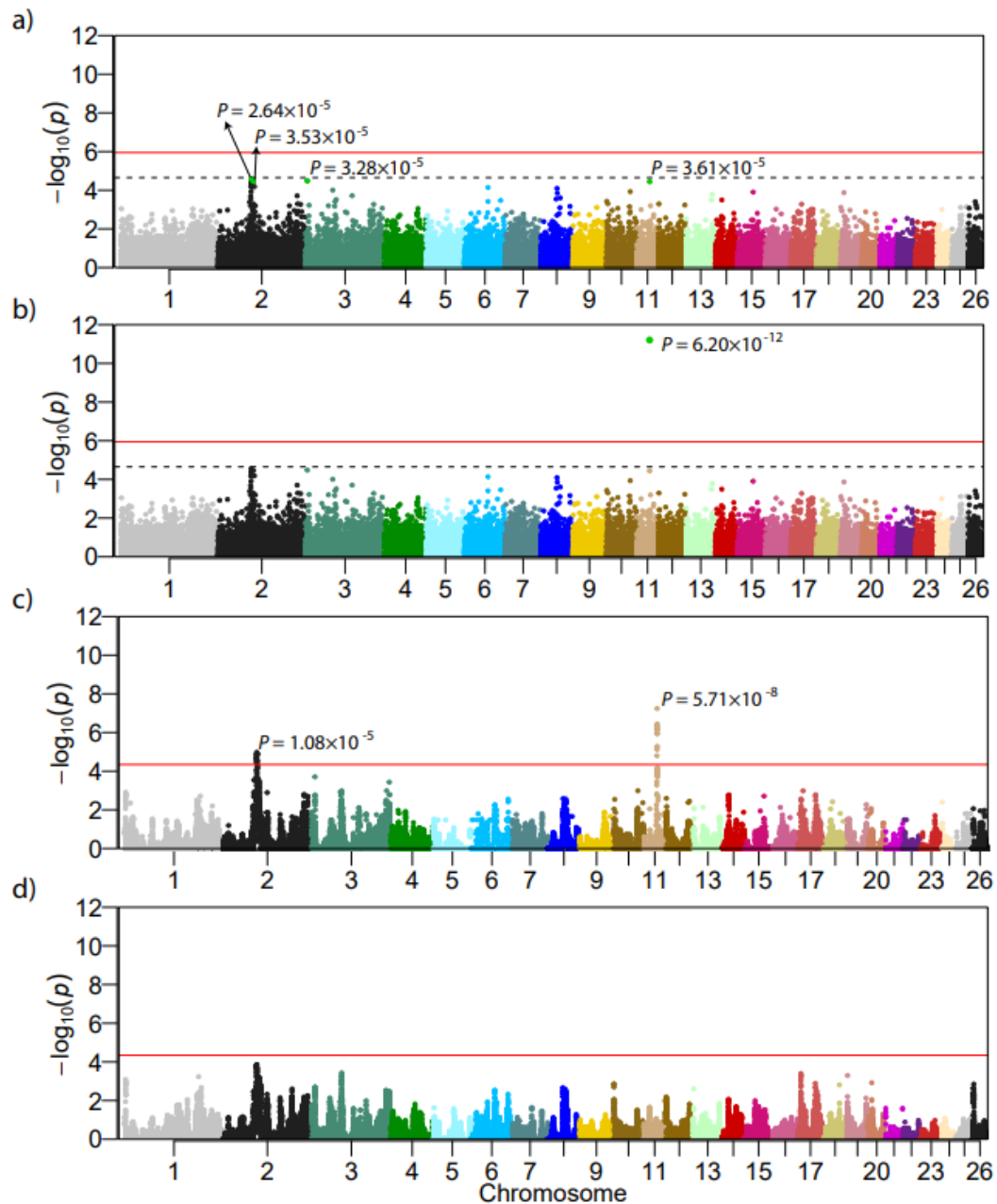


Abbildung 11: Ergebnisse der durchgeführten gemischten linearen Analysen, dargestellt als Manhattan-Plots.

(a) MLMA1 mit 45,114 Markern, kein Marker lag oberhalb der Suggestivlinie, die vier Marker mit den signifikantesten P -Werten sind dargestellt; (b) MLMA2 mit 45,115 Markern, der zusätzlich hinzugefügte Kandidatenlocus auf OAR11 zeigt eine genomweite Signifikanz; (c) cLDLA1 mit 45,114 Markern resultierte in zwei genomweit signifikanten Peaks auf OAR2 und OAR11; (d) cLDLA3 mit 45,115 Markern und zusätzlichem Kandidatenlocus als fixem Effekt, der Peak auf OAR2 sinkt unter die genomweite Signifikanz und der Peak auf OAR11 verschwindet vollständig.

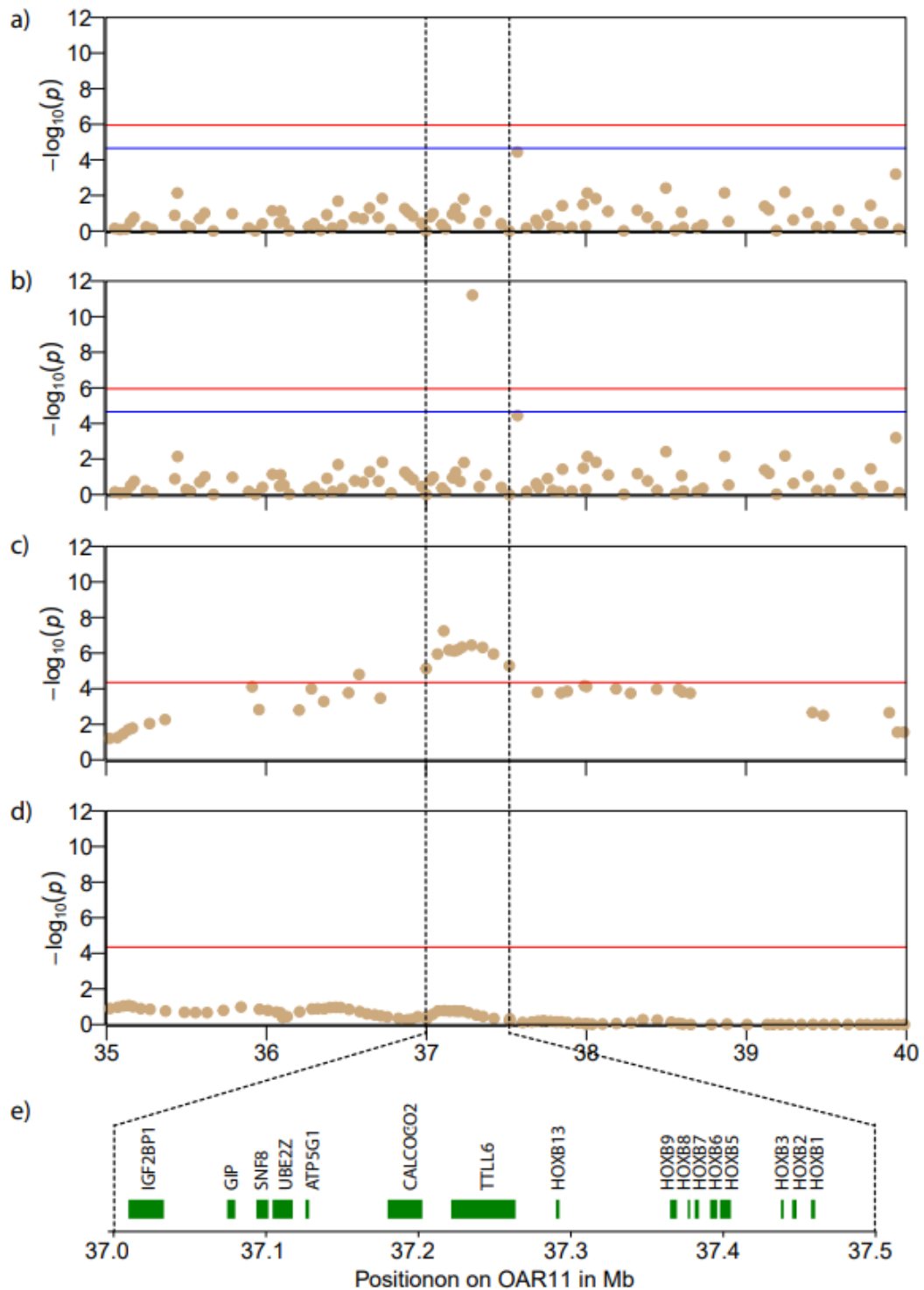


Abbildung 12: Ergebnisse der durchgeführten gemischten linearen Analysen, dargestellt als Zoom-Plots für die Region 35 bis 40 Mb für OAR11.

(a) MLMA1 mit 45,114 Markern, kein Marker lag oberhalb der Suggestivlinie; (b) MLMA2 mit 45,115 Markern, der zusätzlich hinzugefügte Kandidatenlocus auf OAR11 zeigt eine genomweite Signifikanz; (c) cDLA1 mit 45,114 Markern resultierte in zwei genomweit signifikanten Peaks auf OAR2 (hier nicht dargestellt) und OAR11; (d) cDLA3 mit 45,115 Markern und zusätzlichem Kandidatenlocus als fixem Effekt, der Peak verschwindet vollständig; e) Gene des Konfidenzintervalles entsprechend Ihrer Lage und Größe.

3. Auswahl der Lämmer für die Target Capture Sequenzierung

Anhand der Diplotypenanalyse und der Regressionsanalyse wurden 48 Tiere für eine *Target Capture* Sequenzierung ausgewählt.

Abbildung 13 und **Abbildung 14** zeigen die Verteilung der einzelnen genotypisierten Lämmer bezüglich der Diplotypenwirkungen an den *LRTmax* und der Schwanzlänge. Dabei ist eine deutliche Korrelation zwischen Diplotypeneffekt (DipE) auf OAR11 und Schwanzlänge zu erkennen. Eine zusätzlich durchgeführte Regressionsanalyse, bei der die Schwanzlänge als abhängige Variable und der DipE als unabhängige Variable definiert wurden, errechnete einen angepassten Determinationskoeffizienten von $R^2=0.58$ für das *LRTmax* auf OAR11 und nur $R^2=0.15$ für das *LRTmax* auf OAR2. Die multiple lineare Regression bei der neben dem DipE auch Alter, Geschlecht, BW und WH als zusätzliche unabhängige Variablen definiert wurden errechnete einen R^2 von 0.78 für das QTL auf OAR11 und $R^2=0.45$ für das QTL auf OAR2. Nach der Form der LRT-Kurven, der Signifikanz der Kartierungen und dem Determinationskoeffizienten sind die Effekte der Haplotypen, die mit den mutmaßlich ursächlichen Allelen assoziiert sind, bei dem QTL auf OAR11 bestimmender als auf OAR2. Ergebnisse der Regressionsanalysen sind in der **Tabelle 12** und **Tabelle 13** zu sehen.

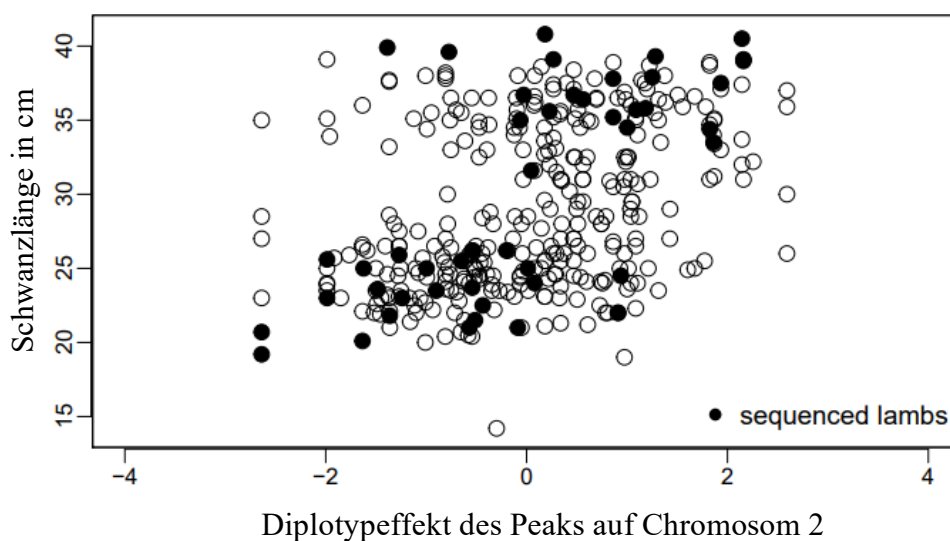


Abbildung 13: Verteilung der sequenzierten Lämmer anhand der Schwanzlänge und des Diplotypeneffektes auf Chromosom 2. Jeder Punkt entspricht eines der 362 untersuchten Lämmer. Sequenzierte Lämmer

sind als ausgefüllte Kreise präsentiert und wurden anhand des Alters, Schwanzlänge und Diplotypeffektes auf Chromosom 11 (Position beim höchsten *LRT*-Wert) ausgewählt. Die Abbildung zeigt keine deutliche Assoziation zwischen der Schwanzlänge und dem Diplotypeffekt.

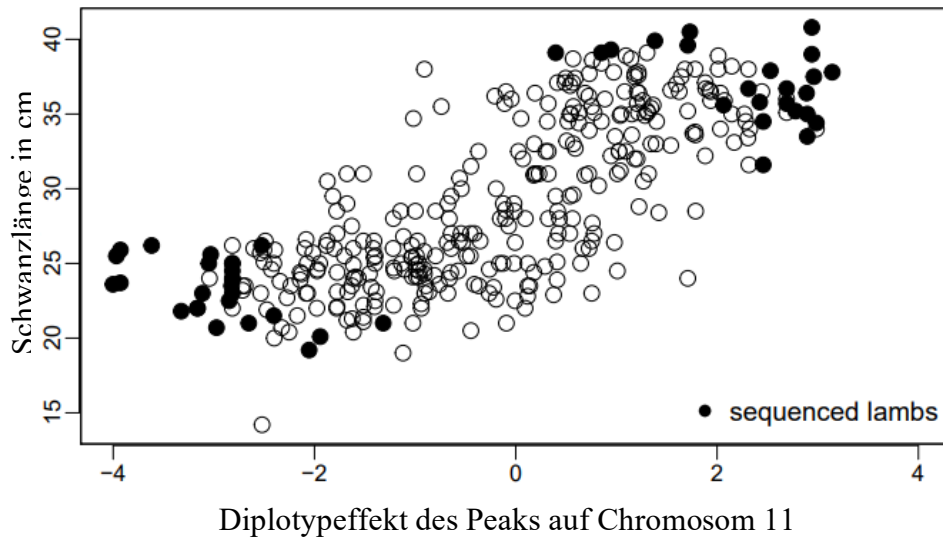


Abbildung 14: Verteilung der sequenzierten Lämmer anhand der Schwanzlänge und des Diplotypeffektes auf Chromosom 11.

Jeder Punkt entspricht eines der 362 untersuchten Lämmer. Sequenzierte Lämmer sind als ausgefüllte Kreise präsentiert und wurden anhand des Alters, Schwanzlänge und Diplotypeffektes auf Chromosom 11 (Position beim höchsten *LRT*-Wert) ausgewählt. Die Abbildung zeigt eine erkennbare Assoziation zwischen der Schwanzlänge und dem Diplotypeffekt.

Tabelle 12: Regressionsanalyse für OAR11.

In Spalte (A) ist der Diplotypeffekt (DipE) die unabhängige Variable und in Spalte (B) sind DipE, Alter, Geschlecht, Körpergewicht (BW) und Widerristhöhe (WH) die unabhängigen Variablen. In beiden Modellen ist die Schwanzlänge die abhängige Variable.

	<i>Abhängige Variable: Schwanzlänge</i>	
	(A)	(B)
DipE	2.661*** (0.120)	2.447*** (0.087)
Alter		-0.077*** (0.022)
Geschlecht		-0.993*** (0.298)
BW		0.621*** (0.064)
WH		0.164** (0.066)
Konstante	29.665*** (0.196)	16.452*** (2.524)
Observations	362	362
R ²	0.577	0.785
Adjusted R ²	0.576	0.782
Residual Std. Error	3.698 (df = 360)	2.649 (df = 356)
F Statistic	491.510*** (df = 1; 360)	260.561*** (df = 5; 356)
<i>Note:</i>		* p<0.1; ** p<0.05; *** p<0.01

Tabelle 13: Regressionsanalyse für OAR2.

In Spalte (A) ist der Diplotypeffekt (DipE) die unabhängige Variable und in Spalte (B) sind DipE, Alter, Geschlecht, Körpergewicht (BW) und Widerristhöhe (WH) die unabhängigen Variablen. In beiden Modellen ist die Schwanzlänge die abhängige Variable.

	<i>Abhängige Variable: Schwanzlänge</i>	
	(A)	(B)
DipE	2.038*** (0.252)	2.047*** (0.204)
Alter		-0.078** (0.036)
Geschlecht		-0.755 (0.472)
BW		0.723*** (0.102)
WH		0.212** (0.105)
Konstante	29.046*** (0.275)	11.816*** (3.987)
Observations	362	362
R ²	0.153	0.461
Adjusted R ²	0.151	0.454
Residual Std. Error	5.232 (df = 360)	4.197 (df = 356)
F Statistic	65.213*** (df = 1; 360)	60.959*** (df = 5; 356)
<i>Note:</i>		* p<0.1; ** p<0.05; *** p<0.01

4. Sequenzierungsergebnisse und Detektion möglicher Kandidatenmutationen

Aufgrund technischer Probleme im Labor wurden nur sehr geringe Sequenzabdeckungen erreicht. Die durchschnittliche Abdeckung betrug zwischen 0.41 und 2.45 für die abgezielte Region auf OAR2 und zwischen 1.01 und 2.39 auf OAR11.

4.1. Nachweis von Punktmutationen und kurzen Indels

Aufgrund der schlechten Qualitäten der *Target capture* Sequenzierungsergebnisse wurden die Sequenzen aller kurzschwänzigen Individuen gepoolt und als ein Individuum betrachtet. Selbiges wurde bei den Sequenzen der langschwänzigen Lämmern durchgeführt. Der *GATK HAPLOTYPECALLER* detektierte so 40,433 sogenannte Short Variants auf OAR11. Die anschließend angewendeten Filterungsmethoden (siehe Abschnitt III.11) reduzierten die Anzahl auf 19 (siehe **Tabelle 8**), von denen wiederum nur ein SNP das erforderliche frequenzbasierte Kriterium erfüllte (siehe Abschnitt III.11). Bei diesem SNP handelt es sich um einen g.37,290,361C>G Basenaustausch. Auch die visuelle Untersuchung der Region mittels *JBROWSE* bestätigte diese Mutation. Diese Missense-Mutation befindet sich im ersten Exon des Gens *HOXB13* und hat einen Aminosäuren (AS)-Austausch p.(Thr8Ser) zur Folge. Alle sequenzierten langschwänzigen Lämmer sind homozygot für diese Missense-Mutation (G/G). Allerdings trugen auch 4 kurzschwänzige Lämmer den Genotyp G/G, 6 waren heterozygot (C/G) und 15 trugen den Genotyp C/C. Eine Bewertung dieser Mutation mittels dem *ENSEMBL* Variant Effect Predictor (MCLAREN et al., 2016) klassifizierte diese Mutation als einen "tolerated" missense Varianten und bestimmte einen SIFT-score von 0.54.

4.2. Vergleich der SNP-Mutation mit Sequenzen verschiedenster Rassen und Spezies

Anschließend wurde die AS-Sequenz des mutierten HOXB13-Proteins mit der AS-Sequenz des Wildtyp-HOXB13-Proteins verschiedener Säugetiere, einschließlich aller wildlebenden Schafarten verglichen (**Abbildung 15**). Diese speziesübergreifende Abgleichung enthüllte, dass die AS, an der die entdeckte Variante auftritt sowie die 5 benachbarten downstream gelegenen AS (LDGAK) konserviert sind.

```

Ovis aries(DA)  MEPGNYTSLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis aries(AA)  MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis vignei     MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis canadensis MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis dalli      MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis nivicola   MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Ovis ammon      MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Bos taurus      MEPSNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPTSAPTLLMPAVNY-APLDLPGSA
Equus caballus  MEPGNYATLDGAKDIEGLLGAGGSRNLVAHSPLTSHPAAAPTLLMPAVNY-APLDLPGSA
Sus scrofa      MEPGNYATLDGAKEIEGLLGAGGSRNLVAHSPLTSHPAAAPTLLMPAVNY-APLDLPGSA
Capra hircus    MEPGNYTLDGAKEIEGLLGAGGSRNLVTHSPLTSHPASAPTLLPAVNY-GPLDLPGSA
Mus musculus    MEPGNYATLDGAKDIEGLLGAGGSRNLVSHSSPLASHPAAAPTLLMPTVNY-APLDLPGSA
Homo sapiens    MEPGNYATLDGAKDIEGLLGAGGSRNLVAHSPLTSHPA-APTLLMPAVNYAAPLDLPGSA
***.*.:*****:*****.****:* :   *****:*** .*****

```

Abbildung 15: Abgleich der Anfangs-Aminosäuresequenz von HOXB13 verschiedener Säugetiere.

Die mutierte Aminosäure ist gelb markiert. Die Sternchen stehen für einheitliche, die Doppelpunkte für sehr ähnliche und der einzelne Punkt für mäßig ähnliche Aminosäuren in jeder Spezies an der jeweiligen Position. Zur Darstellung der Sequenz wurde Clustal Omega (GOUJON et al., 2010; SIEVERS et al., 2011) verwendet; DA: Abgeleitetes Allel, AA: Anzestralles Allel

Der Vergleich der Allelhäufigkeit bei lang- und kurzschwänzigen Schafen mehrerer Rassen offenbarte, dass das oben erwähnte abgeleitete Allel G häufiger bei langschwänzigen Schafrassen vorkommt. Darüber hinaus beobachteten wir an dieser Stelle bei Urial-, Argali-, Schnee-, Dall- und Canadensis-Schafen sowie in zwei ca. ~8,000 Jahre alten Schaf-Genomen nur das ancestrale (Tabelle 7). Auf der anderen Seite waren von 16 untersuchten asiatischen Mufflons 3 heterozygot C/G und 2 homozygot G/G für die Punktmutation (siehe **Anhang 1**). Hierbei sei erwähnt, dass ein asiatisches Mufflon (G/G) und die beiden ~8,000 Jahre alten WGS eine geringe Abdeckung an diesem Locus aufweisen.

4.3. Nachweis von SV

Darüber hinaus identifizierten wir SVs in der Zielregion. Unter Verwendung des folgenden Schwellenwertkriteriums: SV, die in der gepoolten Probe der kurzschwänzigen Lämmer nicht vorhanden sind, jedoch in der gepoolten Probe der langschwänzigen, identifizierten wir 27 und 32 SVs jeweils mittels *SMOOVE* und *DELLY*. Es ist jedoch anzumerken, dass aufgrund der sehr uneinheitlichen und relativ niedrigen Abdeckung der Sequenzen möglicherweise viele echte-positive übersehen wurden und eine hohe Anzahl von falsch-positiven Ergebnissen vorhanden sind. Auch die visuelle Überprüfung mittels *JBROWSE* ergab Hinweise auf die schlechte Qualität der identifizierten SVs. Die visuelle Inspektion jener Region, an der unser detektierter SNP lokalisiert ist, zeigte allerdings zwei unterschiedliche Cluster von Soft-Clip-Reads (**Abbildung 16a**) etwa 130 bp

upstream des Kandidaten-SNPs. Beide Cluster waren nebeneinander angeordnet. Einer war in allen sequenzierten Tieren vorhanden, was auf einen Assembly-Fehler hindeutet, der andere hatte eine unterschiedliche Häufigkeitsverteilung zwischen den Gruppen der kurzschwänzigen und langschwänzigen Lämmer. An der Position 37,290,229 (OAR11) wies die Gruppe der langschwänzigen Lämmer 34 von 35 soft-clipped Reads auf, während in der Gruppe der kurzschwänzigen Lämmer 13 von 54 Reads soft-clipped waren. Auch bei den WGS-Daten, die von NCBI SRA heruntergeladen wurden (Tabelle 7), variierte die Häufigkeit der soft-clipped Reads zwischen kurz- und langschwänzigen Schafrassen an dieser Position signifikant. Daher wurden im nächsten Schritt die gepoolten Reads der Kurzschwanz- und der Langschwanz-Gruppe sowie die WGS-Daten von NCBI SRA gegen das neueste Schaf-Assembly „ARS-UI_Ramb_v2.0“ (DAVENPORT et al., 2022) gemapped. Die erneute visuelle Untersuchung dieser Region zeigte bei diesem Referenzgenom nur ein Cluster von Soft-Clip-Reads an der Position 37,524,996 (**Abbildung 16b**). Dies deutet darauf hin, dass der verschwundene Cluster, welcher ursprünglich bei der Kartierung gegen das Oar 4.0-Assembly identifiziert wurde, auf einen fehlenden Sequenzabschnitt von etwa 40 bp im Oar 4.0-Assembly zurückzuführen ist. Darüber hinaus waren am noch vorhandenen Cluster am Bruchpunkt überlappende Alignments zwischen Vorwärts- und Rückwärts-Reads zu sehen. Aufgrund der Präsenz der Soft-Clip-Reads und des nicht übereinstimmenden Alignments vermuteten wir an dieser Position eine Insertion oder eine Translokation. Während wir diese Region in der ARS-UI_Ramb_v2.0-Assembly untersuchten, stießen wir auf eine Preprint-Veröffentlichung von LI et al. (2022), die von einer Insertion in derselben Region berichtet. Um dieses Erkenntnis zu ergänzen wurden 5 Lämmer mit der Sanger Methode sequenziert. Durch die Analyse der Sangersequenzierungsergebnisse und ein anschließendes Alignment auf das OAR11 im Referenzgenom ARS-UI_Ramb_v2.0 mittels *BLAST* (ALTSCHUL et al., 1990) identifizierten wir die SV als eine 167 bp lange Insertion, welche von einer 14 bp langen Wiederholungssequenz (CTGCCAGCGATTTA) flankiert wird. Ausschnitte der Sangersequenzen sind in **Abbildung 16c** zu sehen. Die komplette Sequenz der Insertion zeigt **Abbildung 8**. Da solche flankierenden Wiederholungssequenzen ein typisches Charakteristikum von Transposons sind, suchten wir nach dieser Sequenz in der DFAM repeat Datenbank (STORER et al., 2021) und identifizierten es als Mitglied der OviAri-1.113 SINE Familie.

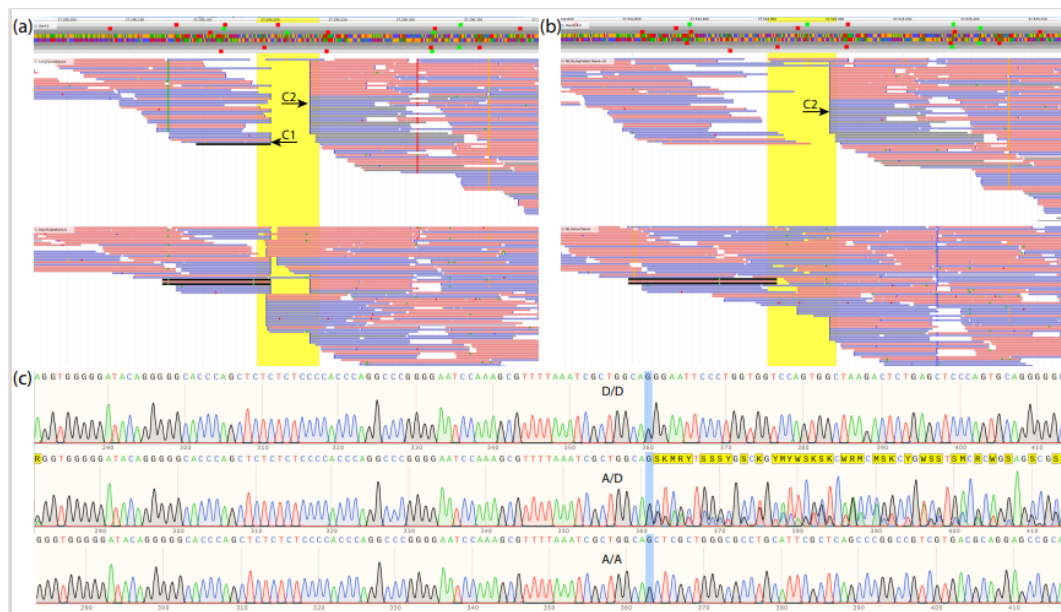


Abbildung 16: Sequenzausschnitte von OAR11.

Mehrere Screenshots aus *JBROWSE* und der Software *SNAPGENE*, welche zwei unterschiedliche Cluster, die Position des Kandidaten-SNPs an Position 37,290,631 auf OAR11 und Auszüge aus den .ab1-Trace-Dateien der Sanger-Sequenzierung zeigen; a) Reads der gepoolten sequenzierten langschwänzigen (oben) und kurzschwänzigen (unten) Merinolandschaftflämmer, die auf das Assembly Oar_4.0 gemappt wurden. Auffallend sind die beiden Cluster. Cluster eins (C1) ist wahrscheinlich auf ein Assembly-Problem zurückzuführen, Cluster zwei (C2) stellt eine Insertion dar, die in 34 der 35 Reads in der langschwänzigen Gruppe und in 13 der 54 Reads in der kurzschwänzigen Gruppe zu finden sind; b) zeigt dieselben Gruppen, gemappt gegen das neueste Assembly ARS-UI_Ramb_v2.0, man beachte das Verschwinden von C1; c) Sanger-Sequenzen des Reverse Primers für ein homozygotes abgeleitetes (D/D), ein heterozygotes (A/D) und ein homozygotes ancestrales (A/A) Lamm, etwa 80 bp vor und 50 bp nach dem Bruchpunkt (blauer Balken) der Insertion. Da die Sequenzen des Reverse Primers qualitativ hochwertiger sind, wurden diese abgebildet. Die Basen der ersten Sangersequenz nach dem Bruchpunkt entsprechen den letzten Basen des Reverse Complements der Insertionssequenz.

5. Genotypisierung der plausibelsten Kandidaten bei allen 362 Lämmern und erneute Kartierung der Schwanzlänge

Um den Zusammenhang zwischen den entdeckten Varianten und der Schwanzlänge zu bestätigen, wurde eine Genotypisierung dieser beiden Kandidaten in der gesamten Kartierungspopulation durchgeführt. Anschließend verwendeten wir die erhaltenen Genotypen in den Modellen MLMA2, cLDLA2 und cLDLA3 (Tabelle 6). Die PCR-Genotypisierung des Kandidaten-SNPs ergab 220 mal den Genotyp G/G, 118 mal C/G und 24 mal C/C. Die PCR-Genotypisierung der 132 bp stromaufwärts gelegenen Kandidaten-Insertion von 167 bp zeigte eine identische Verteilung der Genotypen über die gesamte Kartierungspopulation, d. h. die

Insertion trat in allen Haplotypen auf, die die Base G an der Position 37,290,361 enthielten, und nie in Haplotypen, die die Base C enthielten. Die Verteilung der Allele in den Unterarten der Wildschafe und der Hausschafe erlaubt es uns, auf ancestrale und abgeleitete Allele an beiden Kandidatenloci zu schließen. Das Fehlen der Insertion und das Vorhandensein der Base C am c.C23G SNP in HOXB13 sind ancestrale Allele, während die 167 bp-Insertion und die Base G im Folgenden als abgeleitete Allele bezeichnet werden. **Tabelle 14** zeigt die Verteilung der Genotypen gruppiert in die Lämmer die selektiv beprobt und vermessen wurden und in jene Lämmer die im zweiten Durchlauf zufällig selektiert wurden. Aus der lässt sich erkennen, dass durch selektives Genotypisieren vor allem homozygote Varianten bevorzugt werden. Während sich aus der zufälligen Auswahl an Tieren die wahrscheinlichere Verteilung in einer Population ableiten lässt.

Tabelle 14: Verteilung der Genotypen in den verschiedenen Gruppen.

A/A steht für den ancestralen Genotyp der mit Kurzschwanzigkeit assoziiert ist. A/D ist der gemischterbige Genotyp und D/D ist der abgeleitete Genotyp, welcher vor allem in langschwänzigen Rassen vorzufinden ist. Angaben sind in %.

	A/A	A/D	D/D
Genotypfrequenz aller 362 untersuchten Tiere in %	7	32	61
Genotypfrequenz bei den 102 zufällig ausgewählten Tieren in %	6	40	54
Genotypfrequenz bei den 236 selektiv ausgewählten Tieren in %	7	30	63

Aufgrund der vollständigen Kopplung zwischen der Insertion und dem Missense-SNP werden beide Kandidaten in der MLMA2, cLDLA2 und cLDLA3 synchron berücksichtigt. Alle homozygot ancestralen Lämmer, wurden als kurzschwänzig klassifiziert mit einer mittleren Schwanzlänge von 24.1 cm (± 1.34) und einem mittleren QTL-Effekt von -2.92 cm (± 0.71). Auf der anderen Seite wurden Lämmer, die homozygot für abgeleitete Allele waren, sowohl als langschwänzig als auch als kurzschwänzig eingestuft. Folglich zeigen Lämmer mit abgeleiteten homozygoten Genotypen eine höhere durchschnittliche Schwanzlänge von 31.5 cm und eine 4.15-mal höhere Standardabweichung der Schwanzlänge (± 5.56). Lämmer mit heterozygotem Genotyp zeigen eine durchschnittliche Schwanzlänge von 25.7 und eine 2.73-fach höhere Standardabweichung (± 3.66). **Tabelle 15** fasst Phänotyp, QTL und polygene Effekte der Insertions- und SNP-Kandidaten zusammen, die sich in einem populationsweiten Kopplungsungleichgewicht befinden.

Tabelle 15: Deskriptive Statistik der Dipolypeneffektanalyse.

Dargestellt sind Anzahl der Individuen, Mittelwert (MW) und Standardabweichung (SD) der Schwanzlänge, QTL-Effekte und polygene Effekte der verschiedenen

Genotyp-Gruppen. Die Gruppen sind homozygot anezstral (A/A), homozygot abgeleitet (D/D) und heterozygot (A/D)

Geno -typ	Anzahl			Schwanz- länge		QTL- Effekt		Polygene Effekte	
	Alle	Kurz	Lang	MW	SD	MW	SD	MW	SD
D/D	220	70	150	31.50	5.56	0.68	1.27	1.12	3.56
A/D	118	94	24	25.70	3.66	-1.30	0.89	-1.61	2.96
A/A	24	24	0	24.10	1.34	-2.92	0.71	-2.19	1.52

6. *MLMA* und *cLDLA* mit Kandidatenmutationen als zusätzlichen Marker bzw. fixen Effekt

Um die Auswirkungen der abgeleiteten Allele auf die Ergebnisse der SNP-basierten Assoziationsanalyse und der haplotypbasierten Kartierung zu untersuchen, haben wir beide entdeckten Mutationen als einen Kandidatenlocus zusammengefasst und als zusätzlichen Marker auf Chr11:37,290,361 berücksichtigt (siehe *MLMA2* und *cLDLA2*, **Tabelle 6**). Dadurch erhöht sich im Vergleich zu den ursprünglichen Analysen nur die Anzahl der berücksichtigten Marker um einen, während sich die anderen Eingabedaten sowie die Parameter und das Modell nicht änderten. Diese minimale Änderung führte zu einer enormen Änderung in den Ergebnissen der Assoziationsanalyse und zu geringgradigen Änderungen in den Ergebnissen der haplotypbasierten Kartierung. Das Modell *MLMA2* bestätigte diesen Locus als einen eindeutig genomweit signifikanten ($P = 6.2 \cdot 10^{-12}$), während in der *cLDLA2* eine leicht veränderte Signifikanz ($LRT_{max}=29.112$) an der leicht veränderten Position 37,311,842 bp zu sehen war. Bei der ursprünglichen haplotypbasierten Kartierung (*cLDLA1*) war das LRT_{max} jedoch 179 Kb von *HOXB13* entfernt, und das Modell *cLDLA2* platzierte das LRT_{max} zwischen *HOXB13* und *HOXB9* (**Tabelle 11**). Daher verringerte sich der Abstand zwischen LRT_{max} und dem Kandidatengen *HOXB13* von 179 Kb auf 19 Kb, und somit wurde *HOXB13* zum nächstgelegenen Gen zu LRT_{max} .

Die obigen Ergebnisse weisen auf rs413316737 und/oder die Insertion als plausible Kandidaten für kausale Variationen hin. Gemischte lineare Modelle (*MLMAs* oder *cLDLAs*) ermöglichen es, wichtige kausale Kandidaten zu modellieren und so die

Kartierung der Restvarianz (falls vorhanden) in der Kartierungspopulation zu verbessern. Um das Vorhandensein zusätzlicher Loci zu untersuchen, welche die Schwanzlänge bei Merinoschafen beeinflussen, modellierten wir die Genotypen an der Kandidateninsertion und den SNP als fixe Effekte, während die anderen Eingabedaten, Parameter und das Modell unverändert blieben. Wir schätzten zunächst die SNP-basierte Heritabilität der Schwanzlänge, welche von $h^2=0.992$ auf 0.898 sank. Diese minimale Änderung zeigt, dass der Anteil der additiven genetischen Varianz an der phänotypischen Varianz nach Korrektur für die plausibelsten Mutationskandidaten immer noch hoch ist. Wie in **Abbildung 11d** gezeigt, ist die cLDLA jedoch nicht in der Lage zusätzliche Kandidaten hervorzuheben, obwohl ein relativ hoher Anteil der additiven genetischen Varianz in der hier untersuchten Kartierungspopulation noch vorhanden ist ($h^2=0.898$). Wie von dem wahren Kandidaten erwartet, verschwand das Kartierungssignal auf OAR11 vollständig (**Abbildung 11d** und **Abbildung 17**). Außerdem sank der *LRT*-Wert am Peak auf OAR2 von 19.356 auf 14.476 und änderte seine Position geringfügig von 94,538,115 bp auf 94,345,619 bp. Mit dieser Änderung verlor der mögliche QTL auf OAR2 seine genomweite Signifikanz.

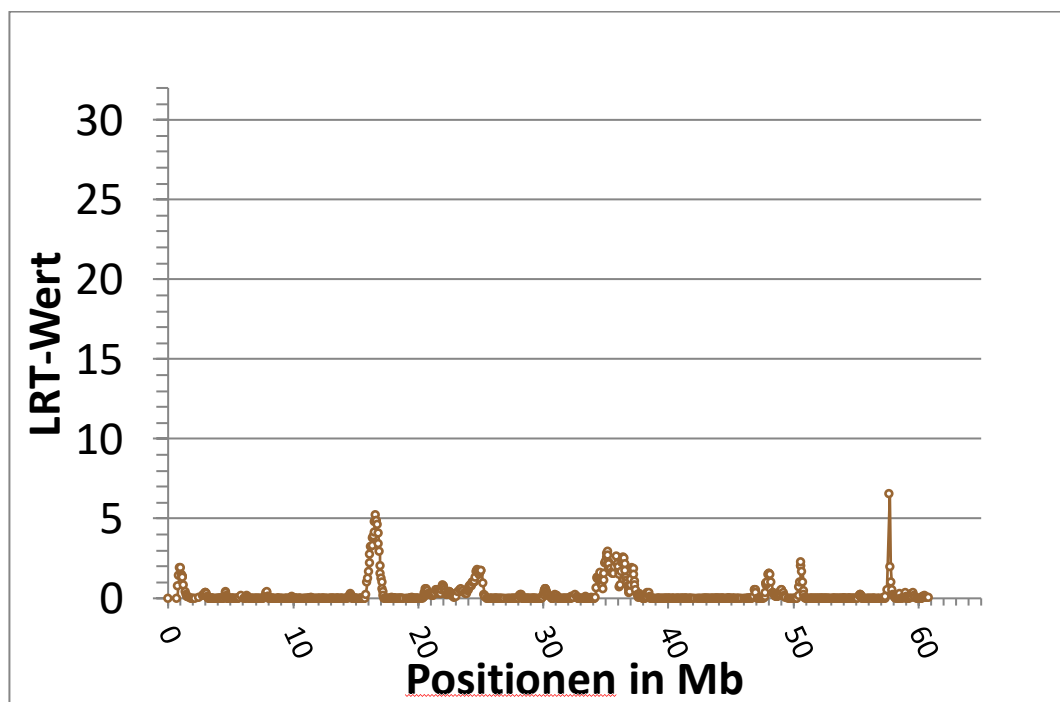


Abbildung 17: LRT- Werte auf OAR11 im Modell cLDLA3.
Der Peak vom cLDLA1 ist verschwunden

V. DISKUSSION

Wie bereits von mehreren Autoren beschrieben (GREEFF et al., 2015; OLTENACU und BOYLAN, 1974) kommen auch unsere Ergebnisse zu dem Entschluss, dass die Heritabilität der Schwanzlänge beim Schaf zum größten Teil von der additiven genetischen Varianz abhängt ($h^2 = 0.992$). Es ist jedoch anzumerken, dass alle Lämmer, die bei der Phänotypisierung Verletzungen, Knicke oder sonstige Anomalien aufwiesen, von der Studie ausgeschlossen wurden. Eine Miteinbeziehung solcher natürlich vorkommenden und umweltbedingter Phänokopien würde höchstwahrscheinlich eine etwas geringere Heritabilität zur Folge haben und Varianzen in der Umwelt mehr Gewichtung verleihen. Außerdem ist eine phänotypische Korrelation bei verwandten Individuen häufig zu einem größeren Teil als geschätzt auf eine gemeinsame Umwelt als auf SNP-assoziierte genetische Effekte zurückzuführen, was zu überhöhten Schätzungen der Heritabilität führt (XIA et al., 2016; ZAITLEN et al., 2013). Da alle hier untersuchten Lämmer aus nur zwei verschiedenen Betrieben (d.h. gemeinsame Umwelt) stammen, ist auch aus diesem Grund die tatsächliche Heritabilität höchstwahrscheinlich niedriger als die berechnete, selbst wenn Verwandtschafts- und Populationsstrukturen berücksichtigt wurden. Trotz der relativ kleinen Anzahl von 362 in die Heritabilitätsberechnung eingegangenen Lämmern, ergab diese einen vergleichsweise geringen Standardfehler von 0.12. Dies liegt wahrscheinlich daran, dass die Mehrzahl der untersuchten Lämmer mehr oder weniger miteinander verwandt waren. Laut VISSCHER et al. (2014) hängt der Standardfehler nur von der Stichprobengröße ab und liegt unter 0.1, wenn mehr als ~3000 unabhängige Individuen verwendet werden.

GWAS, die Standardmethode zur Kartierung von Loci mit Einfluss auf dem Phänotyp, war bei der ersten Anwendung (mit 45,114 Markern) nicht in der Lage, weder genomweit noch suggestiv signifikante Ergebnisse zu liefern. Dies ist nicht überraschend, da die meisten GWAS mehrere hunderttausend Marker in großen Kartierungsdesigns verwenden. Eine Lösung wäre also gewesen, sämtliche Lämmer nochmal mit dem HD SNP-Chip für Schafe zu genotypisieren. Einen ähnlichen Effekt hätte die Erhöhung der Probenanzahl. Da beide Ansätze mit erhöhten Kosten oder massiven Aufwand verbunden gewesen wären, entschieden wir uns für eine Kartierungsmethode, welche aus dem vorhandenen Daten mehr

Informationen verwendet. Es handelt sich hierbei um eine haplotypbasierte Kartierung der cLDLA, welche Kopplungs- und Kopplungsungleichgewichtsinformationen kombiniert. Diese erzielte ein hochsignifikantes Ergebnis ($P = 5.71 \cdot 10^{-8}$) auf OAR11 und ein weiteres genomweitsignifikantes Ergebnis ($P = 1.08 \cdot 10^{-5}$) auf OAR2. Dies zeigt, dass die Varianzkomponentenanalyse im haplotypbasierten gemischten linearen Modell erfolgreicher sein kann als Standardmethoden wie GWAS, wenn die Anzahl der phänotypisierten Individuen und genotypisierten Marker für GWAS nicht groß genug sind.

1. Gene, deren eine mögliche Mitbeteiligung an der Schwanzlänge nachgesagt werden

In der Literatur ist eine hohe Anzahl vorangegangener Forschungsprojekte, welche sich mit dem Thema Schwanzlänge in verschiedenen Tierarten auseinandergesetzt haben, nachzulesen. Viele davon identifizierten verschiedene Varianten im Brachyury-Gen, besser bekannt unter den Namen *T*-Gen, als Ursache von abweichenden Schwanzlängen. Das Brachyury-Gen ist ein Mitglied der T-Box-Familie und codiert für einen Transkriptionsfaktor, welcher wiederum die Expression eines Zielgenes beeinflusst. Die mehr als 20 Gene in der T-Box-Familie übernehmen wesentliche Aufgaben bei der Embryogenese (SHOWELL et al., 2004). Es existieren viele Allele die die Expression oder die AS-Sequenz des Transkriptionsfaktors verändern können, was wiederum Auswirkungen auf die Ausprägung diverser Körpermerkmale hat. So sollen homozygote Mausmutanten im Brachyury-Gen extreme morphologische Abnormitäten in der posterioren Körperregionen aber auch im Notochord, Neuralrohr und den Somiten zeigen. Auch die Allantois kann sich nicht entsprechend ausbilden was zur Folge hat, dass betroffene homozygote Embryonen am zehnten Tage nach der Befruchtung sterben, da sie nicht ausreichend mit Blut versorgt werden können. Heterozygote Mutanten zeigen, neben mehrerer anderen Auffälligkeiten, kürzere Schwänze als Träger homozygoter Wildtyp-Allelen (BEDDINGTON et al., 1992; GLUECKSOHN-SCHOENHEIMER, 1944; GRUNEBERG, 1958).

Auch die Ergebnisse von BUCKINGHAM et al. (2013) deuten darauf hin, dass multiple Allele mit verschiedenen Mutationen (drei 1-bp Deletionen und eine kleine Duplikation/Deletion) im *T*-Gen die Kurzschwanzigkeit in diversen Katzenrassen

(Manx, American Bobtail etc.) verursachen. Der Phänotyp wird bei allen vier Allelen voraussichtlich durch eine Leserasterverschiebung, die zu einer vorzeitigen Verkürzung des Brachyury-Proteins führt, verursacht. Dieser Transkriptionsfaktor ist anschließend nicht in der Lage die Expression eines downstream-gelegenen Gens im gleichen Maße wie im Wildtyp aufrechtzuerhalten. BUCKINGHAM et al. (2013) sprechen in diesem Zusammenhang von einer Haploinsuffizienz. XU et al. (2016) entdeckten zusätzlich einen Zusammenhang zwischen dem Gen *HES7* und der Kurzschwanzigkeit in domestizierten Katzen Asiens und vermuten neben Varianten in *HES7* und *T-Gen* eine dritte unabhängige Varianz in der Evolution der Hauskatze, die zu den kurzschwänzigen Merkmalen führte.

HAWORTH et al. (2001) waren die ersten, die eine genetische Ursache der natürlichen Kurzschwanzigkeit bei Hunden identifizierten. Sie entdeckten einen Basenaustausch im ersten Exon des Gens *T* in mehreren untersuchten Hunderassen mit Stummelschwänzen. Sie vermuteten, dass diese Mutation zu einem Funktionsausfall bei Hunden führt, die heterozygot für die Mutation sind. Das hat zur Folge, dass das T-Protein nicht an sein DNA-Ziel binden kann, was zu einem Kurzschwanz-Phänotyp bei Welsh Corgi Pembroke und deren Nachkommen führt. Ebenfalls schließen sie aus deren Ergebnissen, dass der homozygote Genotyp dieser Mutation letal in der frühen fötalen Entwicklung ist. Bei der Genotypisierung dieser Mutation zweier verschiedener Würfen mit insgesamt 10 frischgeborenen Welpen, derselben Elterntieren, entdeckten INDREBO et al. (2008) zwei Welpen mit diesem mutierten Genotyp in homozygotem Zustand (G/G). Diese Welpen besaßen keinen Schwanz und wiesen mehrere Missbildungen wie *Atresia Ani* und Defekten an der Wirbelsäule auf, weshalb sie unmittelbar *post partum* euthanasiert wurden. Diese Genotypisierungsergebnisse widersprechen zwar der Annahme, dass sämtliche homozygote Mutanten *in utero* sterben, schließen aber dennoch eine Überlebensfähigkeit der Neonaten aus. Neben den zwei schwanzlosen Welpen waren beide Eltern und die kurzschwänzigen Welpen heterozygot für die dominante Mutation im *T-Gen* (C/G). Zwei langschwänzige Welpen hatten den ancestralen Genotyp C/C. HYTONEN et al. (2009) bestätigte diese Mutation als die natürliche Ursache für Kurzschwanzigkeit in 17 anderen Rassen. Bei weiteren 6 kurzschwänzigen Rassen wurde diese Mutation allerdings nicht gefunden und vermuten daher weitere genetische Faktoren die den Phänotyp des Schwanzes

beeinflussen. Bei der Rasse Schwedischer Wallhund (Västgötaspets) beobachteten HYTONEN et al. (2009) einen Rückgang der Wurfgröße um 29 %, wenn beide Elternteile kurzschwänzig waren. Diese Tatsache spricht für eine *in utero* Letalität der Mutation im *T-Gen* (C295G).

Bei der Suche nach der kausalen Variante für die Kurzschwänzigkeit bei der Hunderasse Donggyeongi stieß YOO et al. (2017) ebenfalls auf das *T-Gen*. Dabei vermuten sie eine nicht-synonyme Variante in der kodierenden Sequenz des *T-Gens* aber auch eine Variante des *SFRP2-Gens* (in der CpG Insel) als kausale Varianten für die Anury, das völlige Fehlen des Schwanzes. Für die Kurzschwänzigkeit vermuten sie die Mitwirkung der Gene *ANKRD11* und *ACVR2B* und somit unterschiedliche kausale Varianten als für den völligen Schwanzverlust.

Auch in Schafen gibt es Forschungsergebnisse, die belegen, dass eine Mutation im *T-Gen* für die Kurzschwänzigkeit in bestimmten Schafrassen verantwortlich sein kann. ZHI et al. (2018) untersuchte Hulunbuir-Kurzschwanzschafe, welche in China beheimatet sind. Die Autoren entdeckten eine nonsynonymous Mutation (c.G334T) im *T-Gen* auf OAR8, welche zur Substitution von p.G112W führt. Auch in dieser Studie konnten keine homozygoten Mutanten nachgewiesen werden, was für eine letale Komponente während der Trächtigkeit spricht. Andererseits behaupten HAN et al. (2019), dass die nicht-synonyme c.G334T-Mutation bei chinesischen Fettsteiß- und Hulunbuir-Schafen überwiegend oder ausschließlich homozygot sind. Außerdem segregiert dieser Genotyp auch bei anderen Rassen und ist immer mit der synonymen c.G333C-Mutation assoziiert. Trotz dieser beiden widersprüchlichen Aussagen von ZHI et al. (2018) und HAN et al. (2019) kamen beide Studien zu dem Schluss, dass ein Zusammenhang zwischen dem (extrem) kurzen Schwanzphänotyp und der c.334G>T-Mutation vorliegt.

In diesem Forschungsprojekt konnte kein Anstieg der *LRT*-Kurve in der *T-Gen*-Region (OAR8: 87,717,306-87,727,483 bp) festgestellt werden. Außerdem zeigten weder *ANKRD11* (OAR14: 13,810,611-13,882,911 bp), *ACVR2B* (OAR19: 11,794,562-11,802,479 bp) noch *SFRP2* (OAR17: 3,727,698-3,736,241 bp) einen signifikanten Anstieg des *LRT*-Wertes. *HES7* befindet sich zwar auf OAR11 (27,284,897-27,287,414 bp), ist jedoch ca. 10 Mb vom QTL-KI entfernt. Daher wurde eine Beteiligung als QTL für die Schwanzlänge bei der Rasse Merinolandschaf ausgeschlossen.

2. Ermittlung von Kandidatengen

Das KI auf OAR11 beinhaltet das gesamte HOXB-Cluster des Schafes und sieben weitere Gene (**Tabelle 11**). Unter diesen Genen konnte für *HOXB6* (CASACA et al., 2016), *HOXB8* (YOUNG et al., 2009) und *HOXB13* ein potenzieller Einfluss auf die Schwanzlänge vorhergesagt werden. Außerdem gibt es auch hier Hinweise, dass sie die embryonale Lebensfähigkeit beeinflussen können. Mittels der Capture-Sequenzierung wurden keine genetischen Varianten in *HOXB6* und *HOXB8* gefunden, die mit der Schwanzlänge in Verbindung stehen. Daher werden diese Kandidaten nicht weiter diskutiert. *HOXB13* ist das Gen, das dem maximalen *LRT*-Wert am nächsten liegt und eine entsprechende Literaturrecherche (siehe DIAZ-CUADROS et al., 2021 für eine Übersicht) weist dieses Gen als jenes Gen mit den größten Effekt auf die Schwanzlänge aus. Die Literaturrecherche nach plausiblen Kandidatengen auf OAR2 (**Tabelle 10**) und die Suche mittels *ENRICH*, einem Tool zur Klassifizierung und Organisation phänotypischer Informationen von Säugetierarten (CHEN et al., 2013; KULESHOV et al., 2016; XIE et al., 2021), ergaben keine brauchbaren Ergebnisse. Nach Korrektur für Mehrfachtests (angepasster $P < 0.05$) wurden 29 Ontologien mittels *ENRICH* signifikant angereichert, aber eine plausible Verbindung zur Schwanzlänge konnte nicht festgestellt werden.

Um die geeignetsten Lämmer für die Sequenzierung der Konfidenzintervalle auf OAR11 und OAR2 auszuwählen, wurde zusätzlich zur cDLA eine Diplotypeffekt-Analyse durchgeführt. Die Diplotypeffekten, die Ausprägung des Phänotyps und zum Teil auch die fixen und zufälligen Effekten bei beiden *LRTmax* wurden anschließend gemeinsam betrachtet. Auch hier bestätigten die visuellen Inspektionen (**Abbildung 13** und **Abbildung 14**) sowie die linearen Regressionsanalysen (angepasste R^2 von 0.58 auf OAR11 und nur 0.15 auf OAR2), dass ein QTL auf OAR2 unschlüssig erscheint.

3. Ermittlung von Kandidatenmutationen

In Übereinstimmung mit den oben diskutierten Hinweisen, die für einen ausschlaggebenderen QTL auf OAR11 sprechen, ergab die Capture-Sequenzierung eine plausible Punktmutation und eine SINE-Insertion in der QTL-Region. Bei allen Merinolandschaftlammern und ihren Vätern besteht ein vollständiges Kopplungsungleichgewicht zwischen diesen beiden Varianten. Dies ist nicht

überraschend, da diese Mutationen nur 132 Basenpaare voneinander entfernt sind und eine Rekombination in einem so kurzen Segment selten ist. Diese Kopplung wurde ebenfalls bei einigen typisch lang- und kurzschwänzigen Rassen (Tabelle 7) untersucht und konnte für die langschwänzigen Rassen Weißes Alpenschaf und Rambouillet bestätigt werden. Bei einigen Individuen der kurzschwänzigen domestizierten Schafrassen sowie bei fünf asiatischen Mufflons wurde das Auftreten der Punktmutation ohne Insertion beobachtet. Die Analyse der WGS-Daten (Tabelle 7) und des Aminosäuresequenz-Alignments (**Abbildung 15**) zeigt, dass das Allel C ancestral ist. Daher ist das Allel G abgeleitet, aber relativ alt, sodass es neben *Ovis aries* auch in *Ovis gmelini* (den angenommenen Vorfahre aller domestizierten Hausschafe) segregiert. Die 167 bp-Insertion in der Promotorregion ist ebenfalls abgeleitet, aber jünger und segregiert ausschließlich bei Hausschafen und hier überwiegend bei langschwänzigen Schafrassen. Da die Insertion nie ohne das Allel G, aber das ältere Allel G ohne Insertion beobachtet wurde schließen wir, dass die Insertion im Haplotyp mit dem älteren Missense-Allel G auftrat, und beide abgeleiteten Allele als Block segregieren. Die Abwesenheit der Insertion bei *Ovis gmelini* und ihre Lage in der Promotorregion von *HOXB13* lassen vermuten, dass die Insertion die ursächliche Variante für den Langschwanzphänotyp ist. Höchstwahrscheinlich moduliert die Insertion aufgrund ihrer Lage die Promotoraktivität von *HOXB13* und führt zu einem längeren Schwanz, indem sie die Expression des *HOXB13*-Gens reduziert. In einer kürzlich durchgeführten Studie von LI et al. (2022) wurde ebenfalls eine Insertion von 169 bp in der Nähe der 5'-UTR von *HOXB13* (Position 37,525,005, ARS-UI_Ramb_v2.0) entdeckt. Sie verwendeten ein Graph-Assembly basierend auf Pacific Biosciences (PacBio)-Sequenzierungsdaten von 13 verschiedenen Schafrassen einschließlich Merino. Es ist sehr wahrscheinlich, dass diese Studie und LI et al. (2022) die gleiche Insertion identifizierten. Die geringen Unterschiede in Länge und Position dieser Insertion sind eventuell auf Sequenzierungsfehler aufgrund des Vorhandenseins eines langen Homopolymers von "T"-Basen im identifizierten SINE-Repeat-Element zurückzuführen.

Short-Read-Sequenzierungstechnologien der zweiten Generation können, vor allem in Verbindung mit Assembly-Fehlern, das Auffinden vielversprechender SV-Kandidaten erschweren, weshalb immer eine zusätzliche visuelle Untersuchung der Zielregion mittels eines interaktiven Genomebrowsers in Betracht gezogen werden

sollte. Long-Read-Sequenzierungstechnologien wie PacBio und Oxford Nanopore Technologie (ONT) haben sich bereits als wirksam bei der Identifizierung solcher SVs erwiesen und scheinen daher besser geeignet für den Zweck der SV-Detektion.

Interessanterweise ist die Missense-Mutation (rs413316737) im ersten Exon von *HOXB13* (37,290,361 C→G) auf dem OvineHD-Array als SNP-Marker oar3_OAR11_37337253 enthalten. Dies bietet die Möglichkeit, die Allelverteilung in veröffentlichten Open-Source-Genotypdaten zu überprüfen. Es ist jedoch zu beachten, dass das Top-Allel G von oar3_OAR11_37337253 dem ancestralen C in der kodierenden Sequenz von *HOXB13* entspricht. Dieser Umstand ist in der Nomenklatur von Illumina begründet, in der eine einheitliche Darstellung von Genotyp-Calls, unter Berücksichtigung der Bezeichnung und Ausrichtung des DNA-Strangs, gewährleistet ist (näheres dazu siehe ILLUMINA, 2006).

4. Vergleich der Kartierungsmethoden

In diesem Forschungsprojekt hat sich gezeigt, dass die Varianzkomponentenanalyse im haplotypbasierten gemischten linearen Modell erfolgreicher sein kann als Methoden wie GWAS, besonders wenn die Anzahl der phänotypisierten Individuen und genotypisierten Marker für GWAS nicht groß genug sind.

Die Genotypisierung des Kandidaten-SNPs und der Insertion im gesamten Kartierungsdesign bietet die Möglichkeit, die Effizienz von GWAS und cLDLA mit den Kandidaten als zusätzlichen Marker oder als fixen Effekt zu testen. Die Einbeziehung der Kandidaten als Marker bestätigt nicht nur eine sehr starke Assoziation ($P = 6.2 \cdot 10^{-12}$) mit dem kausativen Locus, sondern stärkt auch den Verdacht, dass einer bzw. beide Kandidaten zusammen die kausativen Varianten sein könnten. Darüber hinaus zeigt es die Leistungsfähigkeit von GWAS, wenn das Design kausale Varianten oder Marker mit populationsweiter LD einschließt. Obwohl die GWAS-Methode im ersten Anlauf keine signifikanten Ergebnisse erzielte, war sie im zweiten Ansatz, bei der Bestätigung der Kandidatenmutationen, sehr erfolgreich. Andererseits führt die Einbeziehung der Kandidaten als zusätzlichen Marker in der haplotypbasierten Kartierung nur zu einer kleinen Veränderung in der Höhe des Kartierungspeaks, dieser liegt aber 9-mal näher am plausibelsten Kandidatengen. Darüber hinaus sei nochmal erwähnt, dass die initiale cLDLA in der Lage war, mit derselben Anzahl an SNPs, den kausativen Locus auf

eine enge Region im Genom einzugrenzen.

Noch schlüssiger ist die Auswirkung dieser Mutationen als fixer Effekt im Modell, denn die Korrektur für die wahre kausale Variante sollte den *LRT*-Peak auf OAR11 auslöschen, und wenn dies eine vollständige additive genetische Varianz erklärt, sollte auch die Heritabilität gegen Null reduziert sein. Tatsächlich löscht dieses Modell die *LRT*-Spitze auf OAR11 aus (**Abbildung 11d** und **Abbildung 17**), aber die Heritabilität bleibt relativ hoch (0.898). Damit haben wir weitere Hinweise auf eine Missense-Mutation im ersten Exon von *HOXB13* und/oder eine SV in der Promotorregion von *HOXB13* als plausible kausale Mutationen gesammelt. Da die Heritabilität nur geringfügig reduziert wurde, konnten wir das Vorhandensein anderer, noch unbekannter kausaler Varianten, die einen großen Teil der phänotypischen Varianz erklären, nicht ausschließen. Allerdings ist auch hier zu beachten, dass die Schätzung der Heritabilität in der vorliegenden Studie aufgrund der Pedigree assoziierten genetischen Varianten und der gemeinsamen Umwelt unserer Proben überschätzt ist (XIA et al., 2016; ZAITLEN et al., 2013) und daher die Variationen um *HOXB13* eventuell mehr zur Schwanzlänge beitragen, als diese Ergebnisse vermuten lassen.

Bei der Beurteilung der Umweltvarianz und eventuell vorhandener Phänokopien sind vier Bemerkungen angebracht.

1. Wir haben uns, falls möglich, an die Vorgaben von ECK et al. (2019) gehalten und Lämmer, die Verletzungen, Knicke oder sonstige Anomalien oder Zeichen einer Kupierung aufwiesen von der Studie ausgeschlossen.
2. Uns sind keine Umwelteffekte bekannt, die in der Lage sind den Schwanz zu verlängern.
3. Allerdings sind uns mehrere Umwelteffekte (Abschnitt II.5) bekannt, die in der Lage sind den Schwanz zu verkürzen.
4. Spätestens bei fünf Wochen alten Lämmern (ECK et al., 2019) sind Verkürzungs-Phänokopien nicht mehr von natürlich kürzerem Schwanz unterscheidbar.

Diese Tatsachen müssen zu einer asymmetrischen Verteilung der Phänokopien führen. Ein genauerer Blick auf Tabelle 15 unterstützt diese Annahme. Wir beobachten nämlich kein einziges langschwänziges Lamm, das homozygot für die ancestrale Variante ist. Andererseits, obwohl die meisten kurzschwänzigen

Lämmer eine (94) oder zwei (70) Kopien des abgeleiteten Allels tragen, finden wir unter ihnen 24 homozygote-anzestrals Lämmer. Diese asymmetrische Verteilung kann auf genetische oder umweltbedingte Varianzquellen zurückzuführen sein und eine Analyse in einem viel größeren Kartierungsdesign sollte hier Klarheit bringen. Ungeachtet dessen führt die Homozygotie für das anzestrale Allel zweifelsohne zu einer signifikanten Schwanzverkürzung.

5. *HOXB13* als schlüssigstes Kandidatengen

HOXB13 gehört zur Familie der Homeobox-Gene, die erstmals in *Drosophila melanogaster* beschrieben wurden (LEWIS, 1978). Dieses Gen kodiert für Transkriptionsfaktoren und spielt eine wichtige Rolle bei der Strukturierung des Körperbaus während der Embryogenese (Das Review von DIAZ-CUADROS et al., 2021 fasst wichtige Erkenntnisse dieses Genes übersichtlich zusammen). Bei Säugetieren gibt es 39 Hox-Gene, die in vier Clustern und 13 paralogen Gruppen organisiert sind (WELLIK, 2007). Zwischen den paralogen Hox-Genen besteht eine funktionelle Redundanz (MCINTYRE et al., 2007), und die paralogen Allele können sich bis zu einem gewissen Grad gegenseitig kompensieren (MACONCHIE et al., 1996). Daher wird der Funktionsverlust eines Hox-Gens aus einem Cluster unter Umständen durch die Funktionalität eines intakten paralogen Gens aus einem anderen Cluster kompensiert, und nur der Funktionsverlust mehrerer paraloger Gene führt zu schwerwiegenderen Folgen für die Achsenstrukturierung (WELLIK, 2007). Die synergistische Wechselwirkung paraloger HOX-Gene könnte auch eine mögliche Verbindung zwischen Schwanzlänge und dem Feinheitsgrad der Wolle bei Schafen nahelegen. Es ist durchaus bekannt, dass *HOXB13* (KOMUVES et al., 2003) und *HOXC13* (SANDER und POWELL, 2004), durch die Regulierung der Expression von Haarstrukturproteinen (Keratin) und Keratin-assoziierten Proteinen (WU et al., 2010) einen Einfluss auf die Morphogenese der Haarfollikel haben und somit die Haarbildung und die Eigenschaften der Wolle steuern. Dies würde einen bereits diskutierten Zusammenhang zwischen Schwanzlänge und feiner Wolle erklären. Darüber hinaus würde es auch die Behauptung von SHELTON (1977), dass lange Schwänze die korrelierte Antwort auf die artifizielle Selektion auf Wollproduktion sei, untermauern. Weitere Studien sind erforderlich, um eine Korrelation zu

beweisen oder zu widerlegen, indem die Expressionsmuster der verschiedenen HOX13-Gene berücksichtigt werden.

Darüber hinaus werden Hox-Genen eine räumliche (entlang der Körperachse) und zeitliche Kollinearität während der Embryogenese nachgesagt, entsprechend der Anordnung der paralogen Hox-Gene auf dem jeweiligen Chromosom (LEWIS, 1978). Da sich Wirbeltierembryonen schrittweise vom Kopf zum Schwanz entwickeln, wird angenommen, dass die paraloge Gruppe HOX13 die Achsenbeendigung am posterioren Ende steuert (DIAZ-CUADROS et al., 2021), was hauptsächlich durch die Regulierung der Proliferation und Apoptoseaktivität in den hinteren Embryonalregionen erreicht wird (ECONOMIDES et al., 2003). Bei Mäusen führen Loss-of-Function-Mutationen in *Hoxb13* zu einem Überwachstum des Rückenmarks und der Schwanzwirbeln bei homozygoten Mäusen. Diese Tiere weisen folglich längere und dickere Schwänze auf, während Lebensfähigkeit und Fruchtbarkeit unbeeinträchtigt bleiben (ECONOMIDES et al., 2003). Die vorzeitige Expression von Genen der paralogen Hox13-Gruppe wirkt sich dagegen negativ auf die Verlängerung der Schwanzachse aus und führt zu einem verkürzten Phänotyp (YOUNG et al., 2009). Auch LI et al. (2022) kamen zu dieser Erkenntnis bei Schafen nachdem sie RNA-Sequenzierungsdaten aus dem Dickdarm von Schafen verwendeten und einen Luciferase-Reporter-Assay durchführten. Dabei erkannten sie einen Zusammenhang zwischen einer geringeren Expression von *HOXB13* und einer Insertion in dessen Promotorregion. Höchstwahrscheinlich wird durch diese Insertion die Expression von *HOXB13* negativ beeinflusst, was eine schwächere Hemmung des Schwanzlängenwachstums zur Folge hat. Dies steht im Einklang mit den Erkenntnissen bei Mäusen. Unter allen Kandidatengenen aus der Hox-Familie ist *HOXB13* somit das schlüssigste funktionelle Kandidatengen. Aufgrund der vollständigen Kopplung können wir jedoch eine gemeinsame Kausalität zwischen der Insertion und der Missense-Mutation im Exon 1 nicht ausschließen.

Ein interessanter Nebenaspekt ist, dass *HOXB13* in der Prostata erwachsener Menschen exprimiert und intensiv als Biomarker-Kandidat für die Prognose von Prostatakrebs untersucht wird (siehe OUHTIT et al., 2016 für einen Überblick). Studien zeigen, dass Missense-Mutationen in der kodierenden Sequenz von *HOXB13* die Affinität (VANOPSTALL et al., 2020) oder die Halbwertszeit (JOHNG et al., 2019) von Heterodimeren zwischen HOXB13 und z. B. MEIS1-

Proteinen verändern können. Es lag jedoch nicht im Rahmen dieser Studie, die Affinität zwischen verschiedenen Transkriptionsfaktoren in wachsenden Schafembryonen oder Zelllinien zu untersuchen. Allerdings ist bereits durchaus bekannt, dass SINE-Insertionen allein die Genexpression auf vielfältige Weise verändern können. Die Hypothese einer möglichen Kausalität der Kombination aus veränderter Expression und veränderter Aminosäuresequenz sollte dennoch in zukünftigen Forschungsprojekten getestet werden.

6. Ausblick

Wie bereits erwähnt, wurden in den 1970er und 1990er Jahren Versuche unternommen, Romney (CARTER, 1976) und Merino (JAMES et al., 1990) (JAMES et al., 1991) Schafe mit kurzen Schwänzen zu züchten. Diese Versuche scheiterten an der nachweislich verminderten Lebensfähigkeit von vermutlich homozygoten Romney-Embryonen mit kurzem Schwänzen und am vermehrten Auftreten von Hinterenddefekten bei kurzschwänzigen Merinos. Diese Beobachtungen sind jedoch nicht in Einklang mit der Tatsache, dass der kurze Schwanz das ancestrale Merkmal ist und dass alte nordische Kurzschwanzrassen, wie das Romanov und Finnschaf sehr lebensfähig und hoch fruchtbar sind. Daher ist davon auszugehen, dass bei diesen Studien genetische Varianten im Zentrum der Forschung standen, die einen anderen genetischen Hintergrund hatten (und deswegen schädlich waren) und nicht jene die auch in den Wildschafarten segregieren. JAMES et al. (1991) schrieben, dass die Schwanzlängen der vier Vatertiere, die vor der Paarung gemessen wurden, alle weniger als 5 cm betragen. Das ist viel kürzer als der Schwanz erwachsener nordischen Kurzschwanzrassen, wie zum Beispiel von Romanov-Schafböcken (~20 cm). Auch die Romney-Schafe, die von CARTER (1976) für die experimentelle Paarung verwendet wurden, wurden als "schwanzlos" beschrieben. Daher sind die kurzen oder schwanzlosen Phänotypen, die in den ersten Schafzuchtversuchen (CARTER, 1976; JAMES et al., 1990) beschrieben wurden, eher mit schädlichen Mutationen vergleichbar, die in bestimmten Hunde- und Katzenrassen (siehe V.1) beschrieben wurden. Möglich ist auch, dass JAMES et al. (1991) und/oder CARTER (1976) für ihre Experimente Tiere mit extrem kurzen Schwänzen und daher mit Mutationen im *T*-Gen, die mit denen der Hulunbuir-Schafe (ZHI et al. (2018)) vergleichbar sind, ausgesucht haben. Die in dieser Studie entdeckten kausalen Kandidaten für kurze Schwänze sind bei einigen sehr lebensfähigen und hochfruchtbaren Rassen (z. B. Finnschaf,

Romanov, Dalapaels, Soay, siehe **Anhang 1**) sehr häufig oder sogar fixiert. Daher hat die Selektion auf das ancestrale Allel keine nachteiligen Auswirkungen auf die Fruchtbarkeit oder der Entstehung von Missbildungen und kann ohne die bei Züchter unbeliebte Introgression erreicht werden. Allerdings bietet die Introgression nordischer Kurzschwanzrassen den Vorteil, dass Kurzschwanzigkeit innerhalb weniger Generationen erreicht werden kann. Der größte Nachteil ist wahrscheinlich, dass viele der erwünschten Produktionsmerkmale, die über viele Jahre bei Merinos selektiert wurden, beeinträchtigt werden (JAMES, 2006). Dem gegenüber bleiben Produktionsmerkmale bei der Zucht auf Kurzschwanzigkeit innerhalb der Merinolandschaftsrasse erhalten, wird aber viele Jahre in Anspruch nehmen. Für die Rasse Suffolk gehen beispielsweise DE HAAS und VEERKAMP (2004) von einer maximalen genetischen Verkürzung um 0.62 cm pro Jahr aus, was eine Schwanzlängenverkürzung um 50 % innerhalb von geschätzten 15 Jahren ermöglicht. Allerdings ist bei der Annahme von DE HAAS und VEERKAMP (2004) nur die rein phänotypische Selektion berücksichtigt. Durch das Auffinden der mutmaßlich kausativen Mutation wird dieser Zeitraum sehr wahrscheinlich stark verkürzt und kann sogar durch das Auffinden weiterer kausativer Mutationen, mit kleineren Effekten, noch weiter reduziert werden. Voraussetzung dafür ist die Einführung und breite Anwendung entsprechender genetischer Tests. Offen ist allerdings noch ein Punkt der bereits von JAMES (2006) erwähnt wurde, er hält es für unwahrscheinlich, dass die Zucht allein in der Lage sein wird das Maß an Schutz zu bieten, welches durch chirurgisches Mulesing oder Schwanzkupieren erreicht wird. Dennoch wird die Zucht auf Kurzschwanzigkeit Leiden und Schäden bei den Lämmern verringern und den Arbeitsaufwand und die Haltungskosten schafhaltender Betriebe senken. Ein Teil der Gesellschaft, der sich aktiv für den Tierschutz einsetzt, hat häufig Vorurteile gegenüber genetischen Methoden im Allgemeinen. Deshalb soll noch einmal darauf hingewiesen werden, dass es sich um eine natürliche, ursprüngliche genetische Variante handelt und die Selektion zugunsten dieser Variante der Wiederherstellung des natürlichsten ursprünglichen Merkmals dient. Wir würden es nicht als "Reparatur" bezeichnen, sondern als "*back to the roots*"-Strategie.

In unserem Design konnten alle homozygot ancestralen und die Mehrheit (79,7 %) der heterozygoten Lämmer als kurzschwänzig eingestuft werden, während homozygote abgeleitete Lämmer meist (71,4 %), aber nicht immer langschwänzig

sind (**Tabelle 15**). Dies könnte umweltbedingt sein oder auf eine Interaktion (Epistase) mit Allelen an anderen Loci im Schafgenom zurückzuführen sein oder einfach auf den polygenen Charakter des hohen Anteils an additiver genetischer Varianz, der durch die hier entdeckten Mutationen noch nicht erklärt wurde (0.898). Darüber hinaus könnten funktionelle Redundanz (MCINTYRE et al., 2007) und synergistische Interaktionen (CHEN und CAPECCHI, 1997) zwischen den paralogen HOX-Genen zu der zusätzlichen Komplexität des Phänotyps beitragen. Die genomischen Regionen, die die Gencluster HOXA (OAR4, 68 Mb), HOXC (OAR3, 132 Mb) und HOXD (OAR2, 132 Mb) enthalten, zeigen jedoch keine Signale in den cLDLA-Analysen ohne (cLDLA1) und mit den (cLDLA3) Kandidatenmutationen als fixem Effekt (Abbildung 11c und Abbildung 11d). Zusammenfassend lässt sich sagen, dass wir für die Kartierung weiterer kausaler Varianten bei langschwänzigen Rassen ein Design mit wesentlich höherer statistischer Aussagekraft benötigen, als das in der vorliegenden Studie durchgeführte. Diese Ergebnisse bieten allerdings einen umfassenden Einblick in die genetische Varianz der Schwanzlänge bei Merinoschafen und liefern Informationen für eine direkte genetisch unterstützte Selektion auf kürzere Schwänze und tragen somit zum Tierschutz bei. Neben kommerziellen und tierschützerischen Aspekten könnten diese und nachfolgende Studien auch zu einem besseren Verständnis der embryonalen Entwicklung beitragen. Nach AIRES et al. (2019) könnten quantitative Unterschiede innerhalb des *Gdf11-Lin28-Hoxb13-Hoxc13*-Genetzwerks für die bei Wirbeltieren beobachtete Variabilität der Schwanzlänge verantwortlich sein. Dabei könnte die Schwanzlänge aus der relativen Intensität oder der Sequenz der individuellen Netzwerkkomponenten bestimmt werden. Die Mechanismen, die die Schwanzlänge regulieren, sind noch nicht vollständig geklärt, insbesondere die den *Lin28*- und *Hox13*-Genen im *Gdf11-Lin28-Hoxb13-Hoxc13*-Netzwerk nachgeschalteten Signalwege. Daher schlugen AIRES et al. (2019) vor, die Parameter des Gen-Netzwerks an Embryonen von Wirbeltieren mit unterschiedlichen Schwanzlängen zu testen. Die Embryonen einer Schafrasse mit einem ancestralen kurzen Schwanz und einem langen Schwanz sind geeignete Kandidaten für die Untersuchung der Kontrolle der Achsenbeendigung. In der Zwischenzeit könnte das Kartierungsdesign erweitert und verbessert werden, um Gene zu kartieren, die auf weitere Netzwerkkandidaten hinweisen, möglicherweise nachgeschaltete Mediatoren von HOXB13.

VI. ZUSAMMENFASSUNG

Ziel dieser Arbeit war es, jene Gene zu identifizieren, die die Schwanzlänge bei Merinoschafen beeinflussen um so eine genetische Alternative zu der inzwischen weit verbreiteten und tierschutzrelevanten Praxis des Schwanzkupierens zu bieten. Dieses Ziel sollte durch eine umfassende und objektive Erhebung von Merkmals- und Umwelteinflüssen in Kombination mit genomweiter Genotypisierung, Feinkartierung und Kandidatengenanalyse erreicht werden.

Die Datengrundlage für die Kartierung mittels kombinierter Kopplungsungleichgewichts- und Kopplungsanalyse (cLDLA) besteht aus 45,114 informativen SNP-Markern (Illumina OvineSNP50-Chip), die bei 362 Merinolandschaftslämmern genotypisiert wurden. Bei diesen genomweit genotypisierten Tieren wurden die Schwanzlänge (cm), das Geschlecht, das Körpergewicht (kg), die Widerristhöhe (cm), das Alter in Tagen sowie Umwelteinflüsse objektiv erfasst. Die cLDLA-Kartierung ergab zwei QTLs mit signifikantem Einfluss auf die Schwanzlänge bei Merinolandschafen in Form von zwei Peaks, die deutlich über der genomweiten Signifikanzschwelle lagen. Diese Peaks befinden sich auf den Schafschromosomen 11 und 2. Die beiden Chromosomensegmente mit insgesamt 4.8 Megabasen wurden dann in 48 selektierten Merinolandschaftslämmern mit dem Capture-Ansatz sequenziert und anschließend mit Hilfe einer vergleichenden Analyse auf das Vorhandensein von Kandidatenvarianten untersucht. Weitere öffentlich verfügbare Schafgenome wurden ebenfalls in die vergleichende Analyse einbezogen. Auf diese Weise konnten zwei plausible Kandidatenvarianten identifiziert werden, die anschließend in allen 362 verfügbaren Proben mittels PCR genotypisiert wurden. Dabei handelt es sich um zwei vielversprechende Kandidaten, die die Schwanzlänge beeinflussen könnten: ein SNP im ersten Exon von *HOXB13* und eine SINE-Insertion im Promotor von *HOXB13*. Eine anschließende Assoziationsanalyse mit einem linearen gemischten Modell (MLMA) zeigte eine hochsignifikante Assoziation der beiden hier entdeckten Kandidatenvarianten mit der Schwanzlänge der 362 Lämmer.

Aus den Ergebnissen dieser gut strukturierten Kartierungspopulation lässt sich schließen, dass die Schwanzlänge von Merinoschafen ein Merkmal mit sehr hoher Heritabilität ist und von einem Hauptlocus und wahrscheinlich mehreren weiteren Loci mit geringem Effekt abhängt. Variationen in und um *HOXB13* verursachen den Haupteffekt auf die Schwanzlänge. Das Vorhandensein mehrerer anderer Loci mit geringem Effekt wird aus der nur geringfügigen Verringerung der Heritabilität nach

Korrektur für den Hauptlocus abgeleitet. Ebenfalls spricht die Tatsache, dass die cLDLA, die unsere Kandidatenmutationen als fixen Effekt enthielt, keine weiteren genomweit signifikanten LRT-Peaks liefern konnte, dafür. Um die zusätzlichen kausalen Loci zu entdecken, ist ein leistungsstärkeres Design erforderlich. Unsere Ergebnisse demonstrieren kraftvoll, dass die Varianzkomponentenanalyse im haplotypbasierten gemischten linearen Modell deutlich erfolgreicher sein kann als andere Methoden wie GWAS, insbesondere wenn die Zahl der phänotypisierten Individuen und genotypisierten Marker für eine GWAS nicht groß genug sind. Unsere Ergebnisse deuten außerdem darauf hin, dass die Short-Read-Sequenzierungstechnologie der zweiten Generation in Verbindung mit Assembly-Fehlern das Auffinden vielversprechender struktureller Kandidatenvarianten massiv erschweren kann. Long-Read-Sequenzierungstechnologien wie PacBio und ONT haben sich jedoch bereits als effektiv bei der Identifizierung solcher strukturellen Varianten erwiesen. Trotz des Umstandes, dass die in dieser Arbeit verwendeten Schafe mittels einer Short-Read-Sequenzierungsmethode sequenziert wurden, konnten wir eine Insertion innerhalb der Promotorregion und einen SNP im ersten Exon von *HOXB13* als plausibelste, aber möglicherweise nicht alleinige Ursachen für den großen Effekt nachweisen. Ein Assemblierungsfehler des Referenzgenoms Oar_v4.0 in unmittelbarer Nähe der Kandidatenvarianten erschwerte den Nachweis der kausalen Insertion massiv, konnte aber schließlich nur durch eine nicht-automatische, visuelle Überprüfung der Sequenzen mit Hilfe eines interaktiven Genombrowsers entdeckt werden. Darüber hinaus empfehlen unsere Ergebnisse kurz- und langschwänzige Schafe als Modelltiere für die Entschlüsselung von Mechanismen von allgemeinem Interesse, z. B. in der Entwicklungsbiologie und der Krebsforschung (Prostata- und Dickdarmkrebs). Nicht zuletzt zeigen unsere Ergebnisse eindrucksvoll, dass eine Selektion auf kürzere Schwänze bei den wirtschaftlich wichtigsten langschwänzigen Merino-Rassen ohne unliebsame Introgression und ohne negative Nebeneffekte möglich ist. Ebenso zeigen unsere Ergebnisse, dass die direkten Gentests und die Selektion auf kürzere Schwänze ethisch unproblematisch sind, da sie nur zu einer erhöhten Häufigkeit des ancestralen Allels führen, welches die natürlichste genetische Variante bei Schafen ist.

VII. SUMMARY

The aim of this work was to identify those genes that influence tail length in Merino sheep in order to provide a genetic alternative to the now widespread and welfare-relevant practice of tail docking. This aim was achieved through a comprehensive and objective survey of trait and environmental influences combined with genome-wide genotyping, fine mapping and candidate gene analysis.

The dataset consisted of 45,114 informative SNP markers (Illumina OvineSNP50 chip) genotyped in 362 Merinolandschaf lambs. Tail length (cm), sex, body weight (kg), height at withers (cm), age in days, and environmental factors were objectively recorded in these genome-wide genotyped animals. Mapping, carried out using combined linkage and linkage disequilibrium analysis (cLDLA) revealed two QTLs with significant influence on tail length in Merino sheep in form of two peaks that were well above the genome-wide significance threshold. These peaks were located on sheep chromosomes 11 and 2. The two chromosome segments, totaling 4.8 megabases, were then sequenced in 48 selected Merinolandschaf lambs using the capture approach and subsequently analyzed for the presence of candidate variants using comparative analysis. Other publicly available sheep genomes were also included in the comparative analysis. In this way, two plausible candidate variants were identified, which were subsequently genotyped by PCR in all 362 available samples. These two promising candidates that could influence tail length include: a SNP in the first exon of *HOXB13* and a SINE insertion in the promoter of *HOXB13*. Subsequent association analysis with a linear mixed model (MLMA) revealed a highly significant association of the two candidate variants discovered here with tail length in the 362 lambs.

From the results of this well-structured mapping population, it can be concluded that tail length of merino sheep is a trait with very high heritability and depends on a major locus and probably several other loci with small effects. Variations in and around *HOXB13* cause the main effect on tail length. The presence of several other low-effect loci is inferred from the only slight reduction in heritability after correction for the main locus. Likewise, the fact that the cLDLA, which contained our candidate mutations as a fixed effect, failed to yield any other genome-wide significant LRT peaks supports this. A more powerful design is required to detect

the additional causal loci. Our results powerfully demonstrate that variance component analysis in the haplotype-based mixed linear model can be significantly more successful than other methods such as GWAS, especially when the number of phenotyped individuals and genotyped markers are not large enough. Our results also suggest that second-generation short-read sequencing technology, in combination with assembly errors, can massively complicate the discovery of promising structural candidate variants. However, long-read sequencing technologies such as PacBio and ONT have already proven effective in identifying such structural variants. Despite the fact that the sheep used in this work were sequenced using a short-read sequencing method, we were able to demonstrate an insertion within the promoter region and a SNP in the first exon of *HOXB13* as the most plausible, but possibly not sole, causes for the large effect. An assembly error of the reference genome Oar_v4.0 in close proximity to the candidate variants massively hampered the detection of the causal insertion, but could finally only be detected by a non-automatic visual inspection of the sequences using an interactive genome browser. Moreover, our results recommend short- and long-tailed sheep as model animals for deciphering mechanisms of general interest, e.g. in developmental biology and cancer research (prostate and colon cancer). Finally, our results impressively demonstrate that selection for shorter tails in the economically most important long-tailed Merino breeds is possible without unwelcome introgression and negative side effects. Likewise, our results show that direct genetic testing and selection for shorter tails are ethically unproblematic, as they only lead to an increased frequency of the ancestral allele, which is the most natural genetic variant in sheep

VIII. LITERATURVERZEICHNIS

Aires R., de Lemos L., Novoa A., et al. (2019): Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev Cell*, 48(3), 383-95 e8.

Albrechtsen A., Sand Korneliussen T., Moltke I., et al. (2009): Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. *Genet Epidemiol*, 33(3), 266-74.

Altschul S. F., Gish W., Miller W., et al. (1990): Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-10.

Andersson L. und Georges M. (2004): Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet*, 5(3), 202-12.

Arbuckle B. S., Kansa S. W., Kansa E., et al. (2014): Data sharing reveals complexity in the westward spread of domestic animals across Neolithic Turkey. *PLoS One*, 9(6), e99845.

ARNOLD A. und ROLOFS H. (2011). *Pferdewirtprüfung [Bd 5] – Genetik* (Vol. 2. Auflage): Norderstedt Books on Demand.

Arnold M. L. (1992): Natural Hybridization as an Evolutionary Process. *Annual Review of Ecology and Systematics*, 23, 237-61.

Auwera G. d. und O'Connor B. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra* (1st Edition ed.): O'Reilly Media.

Babraham B. (2011): FastQC: a quality control tool for high throughput sequence data.

Ball A. D., Stapley J., Dawson D. A., et al. (2010): A comparison of SNPs and

microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). *BMC genomics*, 11, 218.

Bateson W. und Punnett R. C. (1911): On the inter-relations of genetic factors. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 84, 3-8.

Bayerische Herdbuchgesellschaft für Schafzucht e.V. (2021). ZUCHTPROGRAMM MERINOLANDSCHAF. In.

Beddington R. S. P., Rashbass P. und Wilson V. (1992): Brachyury-a gene affecting mouse gastrulation and early organogenesis. *Development*, 157-65.

Behrens H., Scheelje R. und Waßmuth R. (1983). *Lehrbuch der Schafzucht* (Vol. 6.): Paul Parey Hamburg und Berlin.

Berry D. P., Bermingham M. L., Good M., et al. (2011): Genetics of animal health and disease in cattle. *Irish Veterinary Journal*, 64(1), 5.

Bizelis I. und Koutsouli P. (2021). *Greek Domestic Breeds A hidden treasure* (S. Nika Ed.). Athen: AMALTHIA - Network for the Protection of Greek Indigenous Farm Animals.

Bohte-Wilhelmus D. I., De Haas Y., Veerkamp R. F., et al. (2010). Proceedings of the 9th world congress on genetics applied to livestock production, Leipzig, Germany.

Broad Institute. (2019): Picard toolkit. *Broad Institute, GitHub repository*.

Brown T. A. (2007). *Genomes* (3rd ed.). New York: Garland Science Publishing.

Browning B. L. und Browning S. R. (2009): A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated

individuals. *American journal of human genetics*, 84(2), 210-23.

Browning B. L., Zhou Y. und Browning S. R. (2018): A One-Penny Imputed Genome from Next-Generation Reference Panels. *American journal of human genetics*, 103(3), 338-48.

Browning S. R. (2008): Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum Genet*, 124(5), 439-50.

Browning S. R. und Browning B. L. (2007): Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American journal of human genetics*, 81(5), 1084-97.

Buckingham K. J., McMillin M. J., Brassil M. M., et al. (2013): Multiple mutant T alleles cause haploinsufficiency of Brachyury and short tails in Manx cats. *Mamm Genome*, 24(9-10), 400-8.

Buels R., Yao E., Diesh C. M., et al. (2016): JBrowse: a dynamic web platform for genome visualization and analysis. *Genome biology*, 17, 66.

Butler D. G., Cullis B. R., Gilmour A. R., et al. (2018). *ASReml-R Reference Manual*

Version 4. Retrieved from

Carter A. H. (1976): Inherited taillessness in sheep. *New Zealand Ministry of Agriculture and Fisheries. Annual report of the research division.*, 44-5.

Casaca A., Nóvoa A. und Mallo M. (2016): Hoxb6 can interfere with somitogenesis in the posterior embryo through a mechanism independent of its rib-promoting activity. *Development*, 143, 437-48.

Chen E. Y., Tan C. M., Kou Y., et al. (2013): Enrichr: interactive and collaborative

HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14, 128.

Chen F. und Capecchi M. R. (1997): Targeted mutations in *hoxa-9* and *hoxb-9* reveal synergistic interactions. *Developmental biology*, 181(2), 186-96.

Chiffard H. und Reinhardt M. (2015). Vortragstagung der Deutschen Gesellschaft für Züchtungskunde und der Gesellschaft für Tierzuchtwissenschaften.

Darvasi A. und Soller M. (1992): Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics*, 85, 353-9.

Davenport K. M., Bickhart D. M., Worley K., et al. (2022): An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *GigaScience*, 11.

de Haas Y. und Veerkamp R. F. (2004). *Selectie op kortere staarten bij Suffolk schapen*. Retrieved from Wageningen:

Deng J., Xie X. L., Wang D. F., et al. (2020): Paternal Origins and Migratory Episodes of Domestic Sheep. *Curr Biol*, 30(20), 4085-95 e6.

Diaz-Cuadros M., Pourquie O. und El-Sherif E. (2021): Patterning with clocks and genetic cascades: Segmentation and regionalization of vertebrate versus insect body plans. *PLoS Genet*, 17(10), e1009812.

Dubitzky W., Wolkenhauer O., Cho K.-H., et al. (2013). *Encyclopedia of Systems Biology* (W. Li Ed.). New York: Springer.

Eck K., Kunz E., Mendel C., et al. (2019): Morphometric measurements in lambs as a basis for future mapping studies. *Small Ruminant Research*, 181, 57-64.

Economides K. D., Zeltser L. und Capecchi M. R. (2003): *Hoxb13* mutations cause

overgrowth of caudal spinal cord and tail vertebrae. *Developmental Biology*, 256, 317-30.

FAO C. o. g. r. f. f. a. a. (2020). *Status and trends of animal genetic resources*. Retrieved from

Farnir F., Grisart B., Coppieters W., et al. (2002): Simultaneous mining of linkage and linkage disequilibrium to fine map quantitative trait loci in outbred half-sib pedigrees: revisiting the location of a quantitative trait locus with major effect on milk production on bovine chromosome 14. *Genetics*, 161(1), 275-87.

FAWC. (2008). *FAWC report on the implications of castration and tail docking for the welfare of lambs*. Retrieved from

Fisher M., Gregory N., Kent J., et al. (2004). PROCEEDINGS-NEW ZEALAND SOCIETY OF ANIMAL PRODUCTION.

Garza J. C. und Freimer N. B. (1996): Homoplasmy for size at microsatellite loci in humans and chimpanzees. *Genome Res*, 6(3), 211-7.

Geldermann H. (1975): Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 46(7), 319-30.

Geldermann H. (2005). *Tier-Biotechnologie*. Stuttgart: 1. Aufl. Verlag Eugen Ulmer.

Georges M. (2007): Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annual review of genomics and human genetics*, 8, 131-62.

Glazier A. M., Nadeau J. H. und Aitman T. J. (2002): Finding genes that underlie complex traits. *Science (New York, N.Y.)*, 298(5602), 2345-9.

Gluecksohn-Schoenheimer S. (1944): The Development of Normal and Homozygous Brachy (T/T) Mouse Embryos in the Extraembryonic Coelom of the Chick. *Proceedings of the National Academy of Sciences of the United States of America*, 30(6), 134-40.

Goddard M. E. und Hayes B. J. (2009): Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 10(6), 381-91.

Goode E. L. (2011). Linkage Disequilibrium. In M. Schwab (Ed.), *Encyclopedia of Cancer* (pp. 2043-8). Berlin, Heidelberg: Springer Berlin Heidelberg.

Goujon M., McWilliam H., Li W., et al. (2010): A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic acids research*, 38(Web Server issue), W695-9.

Greeff J. C., Karlsson L. J. E. und Schlink A. C. (2015): Inheritance of tail length in Merino sheep. *Proc. Assoc. Advmt. Breed. Genet.*, 21, 237-40.

Gruneberg H. (1958): Genetical studies on the skeleton of the mouse. XXIII. The development of brachyury and anury. *J Embryol Exp Morphol*, 6(3), 424-43.

Gunderson K. L., Steemers F. J., Lee G., et al. (2005): A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet*, 37(5), 549-54.

Haldane J. (1919): The combination of linkage values and the calculation of distance between the loci fo linked factors. *Journal of Gnetics*, 8, 299-309.

Han J., Yang M., Guo T., et al. (2019): Two linked TBXT (brachyury) gene polymorphisms are associated with the tailless phenotype in fat-rumped sheep. *Animal genetics*, 50(6), 772-7.

Hannemann R., Bauer B., Ganter M., et al. (2017): Schmerzhaftige Eingriffe beim Schaf-Schwanzkupieren (Painful interventions in sheep - tail docking).

Tierärztliche Praxis Ausgabe G: Großtiere/Nutztiere, 45(05), 302-11.

Haworth K., Putt W., Cattanaach B., et al. (2001): Canine homolog of the T-box transcription factor T; failure of the protein to bind to its DNA target leads to a short-tail phenotype. *Mamm Genome*, 12(3), 212-8.

Hayes B. und Goddard M. (2010): Genome-wide association and genomic selection in animal breeding. *Genome*, 53(11), 876-83.

Heaton M. P., Smith T. P. L., Freking B. A., et al. (2017): Using sheep genomes from diverse U.S. breeds to identify missense variants in genes affecting fecundity. *F1000Res*, 6, 1303.

Heuven H. C., Bovenhuis H., Janss L. L., et al. (2005): Efficiency of population structures for mapping of Mendelian and imprinted quantitative trait loci in outbred pigs using variance component methods. *Genet Sel Evol*, 37(6), 635-55.

Hirschhorn J. N. und Daly M. J. (2005): Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*, 6(2), 95-108.

Houwen R. H. J., Baharloo S., Blankenship K., et al. (1994): Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genetics*, 8(4), 380-6.

Hytonen M. K., Grall A., Hedan B., et al. (2009): Ancestral T-box mutation is present in many, but not all, short-tailed dog breeds. *The Journal of heredity*, 100(2), 236-40.

Illumina. Data Sheet: Agrigenomics. OvineSNP50 Genotyping BeadChip. Available at: <https://www.illumina.com>

Illumina I. (2006). "TOP/BOT" Strand and "A/B" Allele. Retrieved from https://www.illumina.com/documents/products/technotes/technote_topbot.pdf

Indrebo A., Langeland M., Juul H. M., et al. (2008): A study of inherited short tail and taillessness in Pembroke Welsh corgi. *J Small Anim Pract*, 49(5), 220-4.

James P., Ponzoni R., Gare D., et al. (1991). Proceedings of the Association for the Advancement of Animal Breeding and Genetics.

James P. J. (2006): Genetic alternatives to mulesing and tail docking in sheep: a review. *Australian Journal of Experimental Agriculture*, 46(1), 1-18.

James P. J., Gare D. R., Singh A. W., et al. (1990): Studies of the potential for breeding short tail Merinos. *Wool technology and sheep breeding*, 38(3-4), 106-11.

Janning W. und Knust E. (2004). *Genetik*. Stuttgart.: 1. Aufl, Georg Thieme Verlag.

Jin C., Lan H., Attie A. D., et al. (2004): Selective phenotyping for increased efficiency in genetic mapping studies. *Genetics*, 168(4), 2285-93.

Johng D., Torga G., Ewing C. M., et al. (2019): HOXB13 interaction with MEIS1 modifies proliferation and gene expression in prostate cancer. *Prostate*, 79(4), 414-24.

Joshi NA F. J. (2011): Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33).

Kalds P., Luo Q., Sun K., et al. (2021): Trends towards revealing the genetic architecture of sheep tail patterning: Promising genes and investigatory pathways. *Animal genetics*, 52(6), 799-812.

Kijas J. W., Lenstra J. A., Hayes B., et al. (2012): Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*, 10(2), e1001258.

VERORDNUNG (EG) Nr. 889/2008 DER KOMMISSION vom 5. September 2008

mit Durchführungsvorschriften zur Verordnung (EG) Nr. 834/2007 des Rates über die ökologische/biologische Produktion und die Kennzeichnung von ökologischen/biologischen Erzeugnissen hinsichtlich der ökologischen/biologischen Produktion, Kennzeichnung und Kontrolle, (ABl. L 250 vom 18.9.2008, S. 1), (2008).

Komuves L. G., Ma X. K., Stelnicki E., et al. (2003): HOXB13 homeodomain protein is cytoplasmic throughout fetal skin development. *Dev Dyn*, 227(2), 192-202.

Kruglyak L. (1997): The use of a genetic map of biallelic markers in linkage studies. *Nature genetics*, 17(1), 21-4.

Kruglyak L. (1999): Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, 22(2), 139-44.

Kuhnle E., Weiler U. und Mendel C. (2019): Auswirkungen verschiedener Methoden des Schwanzkupierens bei Lämmern.

Kuleshov M. V., Jones M. R., Rouillard A. D., et al. (2016): Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), W90-7.

Kwade V. (2013). Bauern, lasst den Kühen ihre Hörner, denn sie brauchen sie! *PROVIEH Magazin*.

Layer R. M., Chiang C., Quinlan A. R., et al. (2014): LUMPY: a probabilistic framework for structural variant discovery. *Genome biology*, 15(6), R84.

Lee C. und Fisher A. D. (2007): Welfare consequences of mulesing of sheep. *Aust Vet J*, 85(3), 89-93.

Lee S. H. und van der Werf J. H. J. (2004): The efficiency of designs for fine-

mapping of quantitative trait

loci using combined linkage disequilibrium and linkage. *Genetics Selection Evolution*, 36(2), 145-61.

Lewis E. B. (1978): A gene complex controlling segmentation in *Drosophila*. *Nature*, 276(5688), 565-70.

Li H. und Durbin R. (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, 25(14), 1754-60.

Li H., Handsaker B., Wysoker A., et al. (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078-9.

Li R., Gong M., Zhang X., et al. (2022): The first sheep graph-based pan-genome reveals the spectrum of structural variations and their effects on tail phenotypes.

Li X., Yang J., Shen M., et al. (2020): Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun*, 11(1), 2815.

Liu L., Zhang D., Liu H., et al. (2013): Robust methods for population stratification in genome wide association studies. *BMC bioinformatics*, 14, 132.

Liu Y., Jansen G. B. und Lin C. Y. (2002): The covariance between relatives conditional on genetic markers. *Genet Sel Evol*, 34(6), 657-78.

Lloyd J., Kessell A., Barchia I., et al. (2016): Docked tail length is a risk factor for bacterial arthritis in lambs. *Small Ruminant Research*, 144, 17-22.

Lynch J. J., Hinch G. N. und Adams D. B. (1992). *The behaviour of sheep: biological principles and implications for production* (1 ed.): C.A.B. international.

Maconochie M., Nonchev S., Morrison A., et al. (1996): Paralogous Hox genes:

function and regulation. *Annual review of genetics*, 30, 529-56.

Mallo M. (2020): The vertebrate tail: a gene playground for evolution. *Cellular and molecular life sciences : CMLS*, 77(6), 1021-30.

Mason I. L. (1991). Classification and distribution of sheep breeds. In K. Majjala (Ed.), *Genetic resources of pig, sheep and goat* (1 ed., Vol. 8, pp. 179-94). Amsterdam: Elsevier Science Publishers B. V.

Matherne M. E., Cockerill K., Zhou Y., et al. (2018): Mammals repel mosquitoes with their tails. *J Exp Biol*, 221(Pt 20).

Max B. (1983). *Praktische Schafhaltung*: Verlagsunion Agrar.

McIntyre D. C., Rakshit S., Yallowitz A. R., et al. (2007): Hox patterning of the vertebrate rib cage. *Development*, 134(16), 2981-9.

McLaren W., Gil L., Hunt S. E., et al. (2016): The Ensembl Variant Effect Predictor. *Genome biology*, 17(1), 122.

Meadow R. H. (1998): PIG EXPLOITATION AT NEOLITHIC ÇAYÖNÜ TEPEŞİ. *Ancestors for the Pigs: Pigs in Prehistory*, 15, 77.

Meuwissen T. H. und Goddard M. E. (2000): Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*, 155(1), 421-30.

Meuwissen T. H. und Goddard M. E. (2001): Prediction of identity by descent probabilities from marker-haplotypes. *Genet Sel Evol*, 33(6), 605-34.

Meuwissen T. H., Karlsen A., Lien S., et al. (2002): Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, 161(1), 373-9.

Moradi M. H., Nejati-Javaremi A., Moradi-Shahrabak M., et al. (2012): Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet*, 13, 10.

Morgan T. H. (1911): RANDOM SEGREGATION VERSUS COUPLING IN MENDELIAN INHERITANCE. *Science*, 34(873), 384.

Muller M. P., Rothammer S., Seichter D., et al. (2017): Genome-wide mapping of 10 calving and fertility traits in Holstein dairy cattle with special regard to chromosome 18. *Journal of dairy science*, 100(3), 1987-2006.

Nickel R., Schummer A. und Seiferle E. (2003). *Lehrbuch der Anatomie der Haustiere ; 4*. Stuttgart: Parey im MVS.

Oltenacu E. A. und Boylan W. J. (1974): Inheritance of tail length in crossbred Finnsheep. *The Journal of heredity*, 65(6), 331-4.

Ouhtit A., Al-Kindi M. N., Kumar P. R., et al. (2016): Hoxb13, a potential prognostic biomarker for prostate cancer. *Front Biosci (Elite Ed)*, 8(1), 40-5.

Peters J., A D. V. D. und D H. (2005). The Upper Euphrates-Tigris Basin: cradle of agro-pastoralism?[Vigne, J. et al.(eds.)][96–124] The first steps of animal domestication. New Archaeological Approaches. In: Oxbow Books, Oxford, UK.

Peters J., von den Driesch A., Helmer D., et al. (1999): Early animal husbandry in the Northern Levant. *Paléorient*, 27-48.

Powell J. E., Visscher P. M. und Goddard M. E. (2010): Reconciling the analysis of IBD and IBS in complex trait studies. *Nat Rev Genet*, 11(11), 800-5.

Purcell S., Neale B., Todd-Brown K., et al. (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), 559-75.

R Core Team (2020). R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. Retrieved from <https://www.R-project.org/>

Rausch T., Zichner T., Schlattl A., et al. (2012): DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England)*, 28(18), i333-i9.

Riches J. (1941): The relation of tail length to the incidence of blowfly strike of the breech of merino sheep. *Journal of the Council for Scientific and Industrial Research*, 14: 88-93.

Riches J. H. (1942): Further observations on the relation of tail length to the incidence of blowfly strike of the breech of Merino sheep. *Journal of the Council for Scientific and Industrial Research*, 15: 3-9.

Sander G. R. und Powell B. C. (2004): Structure and expression of the ovine Hoxc-13 gene. *Gene*, 327(1), 107-16.

Scobie D. und O'Connell D. (2002). Proceedings of the New Zealand Society of Animal Production.

Shelton M. (1977): Studies on tail length of Rambouillet and Mouflon sheep. *The Journal of heredity*, 68(2), 128-30.

Sherratt A. J. W. a. (1983): The secondary exploitation of animals in the Old World. *15(1)*, 90-104.

Showell C., Binder O. und Conlon F. L. (2004): T-box genes in early embryogenesis. *Dev Dyn*, 229(1), 201-18.

Sievers F., Wilm A., Dineen D., et al. (2011): Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular*

systems biology, 7, 539.

Slatkin M. (2008): Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nature reviews. Genetics*, 9(6), 477-85.

Small A. H., Marini D., Dyall T., et al. (2018): A randomised field study evaluating the effectiveness of buccal meloxicam and topical local anaesthetic formulations administered singly or in combination at improving welfare of female Merino lambs undergoing surgical mulesing and hot knife tail docking. *Res Vet Sci*, 118, 305-11.

Smith J. M. und Haigh J. (2009): The hitch-hiking effect of a favourable gene. *Genetical research*, 23(1), 23-35.

Steffens M., Jäger L. und Klinkel N. (2021). Die Haltung von unkupierten Schafen - Modell- und Demonstrationsvorhaben (MuD) Tierschutz. In B. f. L. u. Ernährung (Ed.): Dorothe Heidemann.

Stephan W. und Hörger A. C. (2019). *Molekulare Populationsgenetik*: Springer Berlin Heidelberg.

Storer J., Hubley R., Rosen J., et al. (2021): The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA*, 12(1), 2.

Strobel H., Ganter M., Bauer B., et al. (2018): Schmerzhaftes Eingriffe beim Schaf – Schwanzkupieren. *Tierärztliche Praxis Ausgabe G: Großtiere / Nutztiere*, 45(05), 302-11.

Swan R. A., Chapman H. M., Hawkins C. D., et al. (1984): The epidemiology of squamous cell carcinoma of the perineal region of sheep: abattoir and flock studies. *Aust Vet J*, 61(5), 146-51.

Tanno K. und Willcox G. (2006): How fast was wild wheat domesticated? *Science*

(New York, N.Y.), 311(5769), 1886.

Thomas D. L., Waldron D. F., Lowe G. D., et al. (2003): Length of docked tail and the incidence of rectal prolapse in lambs. *J Anim Sci*, 81(11), 2725-32.

Verordnung der Bundesministerin für Gesundheit und Frauen über die Mindestanforderungen für die Haltung von Pferden und Pferdeartigen, Schweinen, Rindern, Schafen, Ziegen, Schalenwild, Lamas, Kaninchen, Hausgeflügel, Straußen und Nutzfischen 1. Tierhaltungsverordnung (Austrian Animal Husbandry Ordinance), (2017).

Turner S. (2017). qqman: Q-Q and Manhattan Plots for GWAS Data. (Version R package version 0.1.4.). Retrieved from <https://CRAN.R-project.org/package=qqman>

Uffelmann E., Huang Q. Q., Munung N. S., et al. (2021): Genome-wide association studies. *Nature Reviews Methods Primers*, 1(1).

Untergasser A., Cutcutache I., Koressaar T., et al. (2012): Primer3--new capabilities and interfaces. *Nucleic acids research*, 40(15), e115.

Upadhyay M., Hauser A., Kunz E., et al. (2020): The First Draft Genome Assembly of Snow Sheep (*Ovis nivicola*). *Genome Biol Evol*, 12(8), 1330-6.

Upadhyay M., Kunz E., Sandoval-Castellanos E., et al. (2021): Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep. *Mol Ecol*, 30(24), 6701-17.

van Ooijen J. W. (1992): Accuracy of mapping quantitative trait loci in autogamous species. *Theor Appl Genet*, 84(7-8), 803-11.

VanOpstall C., Perike S., Brechka H., et al. (2020): MEIS-mediated suppression of human prostate cancer growth and metastasis through HOXB13-dependent

regulation of proteoglycans. *Elife*, 9.

Vignal A., Milan D., SanCristobal M., et al. (2002): A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, selection, evolution : GSE*, 34(3), 275-305.

Visscher P. M. und Goddard M. E. (2004): Prediction of the confidence interval of quantitative trait Loci location. *Behav Genet*, 34(4), 477-82.

Visscher P. M., Hemani G., Vinkhuyzen A. A., et al. (2014): Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet*, 10(4), e1004269.

Weber J. L. und Wong C. (1993): Mutation of human short tandem repeats. *Hum Mol Genet*, 2(8), 1123-8.

Weissenbach J., Gyapay G., Dib C., et al. (1992): A second-generation linkage map of the human genome. *Nature*, 359(6398), 794-801.

Weller J. I. (2001). *Quantitative Trait Loci Analysis in Animals*. Wallingford, United Kingdom: 1. Aufl. CABI Publishing

Wellik D. M. (2007): Hox patterning of the vertebrate axial skeleton. *Dev Dyn*, 236(9), 2454-63.

Windels H. (1990). 62nd Annual Sheep and Lamb Feeders

Day, Univ. of Minnesota, Morris.

Wu J. H., Yan Z. W., Husile, et al. (2010): [Hoxc13 and the development of hair follicle]. *Yi Chuan*, 32(7), 656-62.

Xia C., Amador C., Huffman J., et al. (2016): Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric

and Cardiometabolic Trait Variation. *PLoS Genet*, 12(2), e1005804.

Xie Z., Bailey A., Kuleshov M. V., et al. (2021): Gene Set Knowledge Discovery with Enrichr. *Curr Protoc*, 1(3), e90.

Xiong M. und Jin L. (1999): Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. *American journal of human genetics*, 64(2), 629-40.

Xiong X., Liu X., Zhou L., et al. (2015): Genome-wide association analysis reveals genetic loci and candidate genes for meat quality traits in Chinese Laiwu pigs. *Mamm Genome*, 26(3-4), 181-90.

Xu X., Sun X., Hu X. S., et al. (2016): Whole Genome Sequencing Identifies a Missense Mutation in HES7 Associated with Short Tails in Asian Domestic Cats. *Sci Rep*, 6, 31583.

Yang J., Benyamin B., McEvoy B. P., et al. (2010): Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*, 42(7), 565-9.

Yang J., Lee S. H., Goddard M. E., et al. (2011): GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics*, 88(1), 76-82.

Yang J., Zaitlen N. A., Goddard M. E., et al. (2014): Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*, 46(2), 100-6.

Yoo D., Kim K., Kim H., et al. (2017): The Genetic Origin of Short Tail in Endangered Korean Dog, DongGyeonggi. *Sci Rep*, 7(1), 10048.

Young T., Rowland J. E., van de Ven C., et al. (2009): Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Dev Cell*, 17(4), 516-26.

Yurtman E., Ozer O., Yuncu E., et al. (2021): Archaeogenetic analysis of Neolithic sheep from Anatolia suggests a complex demographic history since domestication. *Commun Biol*, 4(1), 1279.

Zaitlen N., Kraft P., Patterson N., et al. (2013): Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet*, 9(5), e1003520.

Zeder M. A. (2008): Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), 11597-604.

Zhi D., Da L., Liu M., et al. (2018): Whole Genome Sequencing of Hulunbuir Short-Tailed Sheep for Identifying Candidate Genes Related to the Short-Tail Phenotype. *G3 (Bethesda)*, 8(2), 377-83.

IX. ANHANG

Anhang 1: Genotypfrequenzen von rs413316737 in verschiedenen Rassen mit verschiedenen Schwanzphänotypen.

Rassen, für die der untersuchte Phänotyp nicht bekannt ist, wurden entfernt. Die Tabelle besteht aus Informationen vom "International Sheep Genome Consortium und aus der hauseigenen Datenbank. Die letzte Spalte zeigt die Anzahl der Tiere aus der hauseigenen Datenbank, aus der die Genotypen an der Kandidatenposition extrahiert wurden. Diese Datenbank enthält die am Institut genotypisierten Proben und zusätzlich ungefilterte Genotypdaten, die von der Vetsuisse-Fakultät, Institut für Genetik, Universität Bern, zur Verfügung gestellt wurden; online verfügbare Genotypdaten von Rochus et al, BMC Genomics, 2018 (<https://doi.org/10.1186/s12864-018-4447-x>), Rochus et al, BMC Genet, 2020 (DOI: 10.1186/s12863-020-0827-8), Stoffel et al, Nat Commun 2021 (10.1038/s41467-021-23222-9) und Genotypen, die aus online verfügbaren Whole-Genome-Sequenzen mit unten stehenden Run numbers und BioSample-ID extrahiert wurden:

Rasse	Phänotyp des Schwanzes	Summe			Hauseigene SNP-Datenbank
		G/G	G/C	C/C	
Afshari	fettschwänzig	1	1	0	0
Awassi	fettschwänzig	1	0	2	0
Bangladeshi	kurz und dünn	0	0	2	0
Castellana	lang und dünn	1	1	0	0
Cheviot	lang und dünn	2	0	0	0
Changthangi	kurz und dünn	0	1	1	0
Churra	lang und dünn	2	0	0	0
Coopworth	lang und dünn	30	5	2	0
D'Man	lang und dünn	20	3	3	0
Finnsheep	kurz und dünn	2	6	11	15
Garut	thin-tailed	1	1	0	0
Gulf Coast Native	lang und dünn	2	0	0	0
Karakas	fettschwänzig	0	1	1	0
Merino Horned	lang und dünn	1	3	0	0
Merino Polled	lang und dünn	3	3	0	0
Merino	lang und dünn	2	1	0	0
Norduz	fettschwänzig	1	1	0	0
Norwegian White Sheep	kurz und dünn	2	0	0	0
Ojalada	lang und dünn	1	1	0	0
Ouled Djellal	lang und dünn	1	2	5	0
Romney	lang und dünn	40	3	2	0
Ronderib Afrikaner	fettschwänzig	0	1	1	0
Salz	kurz und dünn	0	1	1	0
Santa Ines	thin-tailed	0	1	1	0

Sardinian	Anzestral	lang und dünn	19	5	0	0
Black						
Sumatran		kurz und dünn	0	0	2	0
Swiss White Alpine		lang und dünn	25	2	0	24
Texel		kurz und dünn	12	14	8	24
Wiltshire		lang und dünn	2	0	0	0
Alai		fettsteißig	7	12	1	20
Aykol		fettsteißig	1	7	11	19
Gissar		fettsteißig	1	12	7	20
Kerry Hill		sprunggelenkslang	2	0	0	2
Alpines Steinschaf		sprunggelenkslang	5	10	1	16
Merinolandschaf		lang und dünn	225	161	72	458
Asiatic Mouflon		kurz und dünn	11	3	2	17
Tien-Shan		lang und dünn	6	13	1	20
Berrichon		lang und dünn	19	0	0	19
Blanc du Massif Central		lang und dünn	19	1	0	20
Causses du Lot		lang und dünn	20	0	0	20
Mouton Charollais		lang und dünn	24	0	0	24
Charmoise		lang und dünn	23	0	0	23
Corse		lang und dünn	12	4	0	16
European Mouflon		kurz und dünn	0	2	0	2
Île de France		lang und dünn	23	0	0	23
Limousine		lang und dünn	18	0	0	18
Lacaune meat		lang und dünn	33	1	0	34
Lacaune milk		lang und dünn	36	0	0	36
Mérinos d'Arles		lang und dünn	12	6	0	18
Mourerous		lang und dünn	15	1	0	16
Manech Tête Rouge		lang und dünn	22	3	0	25
Noire du Velay		lang und dünn	16	3	0	19
Ouessant		kurz und dünn	4	6	8	18
Préalpes du Sud		lang und dünn	15	2	0	17
Rava		lang und dünn	20	0	0	20
Mérinos de Rambouillet		lang und dünn	0	2	25	27
Roussin de la Hague		lang und dünn	21	0	0	21
Romanov		kurz und dünn	2	7	5	14
Rouge de l'Ouest		lang und dünn	17	0	0	17
Tarasconnaise		lang und dünn	14	1	0	15
Mouton Vendéen		lang und dünn	22	0	0	22
Dalapaels(Dala fur)		kurz und dünn	0	1	20	21
Fjaellnaes		kurz und dünn	10	0	0	10
Gotland		kurz und dünn	11	7	1	19
Gute		kurz und dünn	11	11	0	22
Kloevsjoe		kurz und dünn	14	7	0	21
Swiss Mirror Sheep		lang und dünn	19	0	0	19
Valais Blacknose Sheep		lang und dünn	24	0	0	24
Valais Red Sheep		lang und dünn	24	0	89	113
Soay		kurz und dünn	1	50	137	188
Argali		kurz und dünn	0	0	5	5
Canadensis (Bighorn)		kurz und dünn	0	0	3	3

Dalli	kurz und dünn	0	0	2	2
Urial	kurz und dünn	0	0	5	5
SnowSheep	kurz und dünn	0	0	6	6
Cyprus Mouflon	kurz und dünn	0	0	5	5
Jezersko-Solcava sheep	lang und dünn	33	3	0	36

X. DANKSAGUNG

Als erstes möchte ich mich bei der Tierzuchtforschung e.V. München für die Anstellung und für die Finanzierung meines Doktorats bedanken. Ebenfalls möchte ich auch meinen Dank für die Mitfinanzierung des Besuchs des Kongresses WCGALP in Rotterdam aussprechen.

An zweiter Stelle bedanke ich mich sehr herzlich bei meinem Doktorvater PD Dr. Ivica Međugorac für die Aufnahme als Doktorand und die Überlassung des erfolgreichen Themas sowie für die umfassende Betreuung und Unterstützung während der gesamten Dauer der Dissertation.

Besonderer Dank gilt auch meinen hilfsbereiten Kollegen Jürgen Klawatsch, Edson Sandoval Castellanos und vor allem Maulik Upadhyay für die ausführlichen Diskussionen, die amüsanten Meinungsäustausche und zuvorkommenden Hilfestellungen zu den verschiedensten Themen meiner Doktorarbeit. Die verschiedenen Konversationen während unserer Mittagspausen und sonstigen Treffen werde ich sehr vermissen.

Meiner Mitdotorandin in der Anfangsphase Dr. Nina Dachs danke für die nette Aufnahme und großzügige Beratung gerade in der Zeit in der ich sie am meisten brauchte. Die Arbeitsatmosphäre mit dir war sehr angenehm.

Vielen Dank möchte ich auch den AG Populationsgenomik Mitarbeiter/in Renate Damian und Martin Dinkel aussprechen für die tatkräftige Unterstützung und netten Gespräche.

Schöne Grüße und vielen lieben Dank richte ich auch an die Justus-Liebig-Universität in Giessen, ganz besonders an Frau Prof. Gesine Lühken. Es hat mich sehr gefreut mit Kollegen zusammenzuarbeiten die an jener Universität arbeiten, die mir die Approbation als Tierarzt erteilte. Deine kritische Durchsicht meines Manuskriptes hat das finale Paper qualitativ höhergestellt.

Meinen besten Freund/innen, die mir stets mit Rat und Tat zur Verfügung standen, mich niemals hängen ließen und immer ein offenes Ohr für mich hatten. Ich werde hier keine Namen nennen, da ich mir sicher bin, dass jeder weiß der angesprochen ist und ich niemanden vergessen möchte. Ich hoffe, dass auch in Zukunft unsere Freundschaften bestehen bleiben.

Die letzte Worte meiner Doktorarbeit richte ich an meine Familie, den meiner Meinung nach der größte Dank gebührt. Ich danke euch vom ganzen Herzen für die großzügige Unterstützung während meiner Dissertation, meiner Zeit als Student und der Zeit davor. Mit euren zahlreichen Handgriffen, sei es auch nur das Herumkutschieren gewesen, wart ihr eine große Stütze und Motivation in meinem bisherigen Leben. Auch nach einem langen Wiedersehen fühlte ich mich stets so als wäre ich nie weg gewesen.