

THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Genomic basis of sexually divergent traits in red deer (*Cervus elaphus*)

Lucy Peters



Thesis submitted for the degree of Doctor of Philosophy

School of Biological Sciences

University of Edinburgh

September 2021

Declaration

The work described in this thesis has been carried out by myself under the guidance from my supervisors, unless otherwise stated and detailed below. I have written this thesis and this work has not been submitted for any other degree or professional qualification.

Chapters 2 and 3: The genomic data and all antler measure data were collected by others as part of the ongoing Rum red deer research project. Access to the data was provided by Josephine Pemberton. The pedigree used was constructed by Jisca Huisman. I conducted the statistical analysis and wrote the chapter.

Chapter 4: Additional quality control regarding SNP clustering was carried out by Susan Johnston. All data on individual survival including birth and death data were collected as part of the long term Rum red deer study. Access to the data was provided by Josephine Pemberton. I conducted the statistical analysis and wrote the chapter.

Lucy Peters

Abstract

Sexually selected traits show large variation and rapid evolution across the animal kingdom. The genetic variation underlying sexually selected traits is often maintained despite apparent directional selection, with proposed explanations including sexually antagonistic selection or trade-offs with other life-history traits, such as survival. A key step in uncovering evolutionary drivers and constraints of these traits is to understand them at the genomic level. However, few empirical studies have attempted to investigate this aspect in the wild. Using long-term data from a population of red deer on the Isle of Rum, Scotland, this thesis both investigates the genomic basis of antler morphology, an important sexually selected trait, and genome-wide signals of sexually antagonistic selection through viability selection across the genome.

Antlers are a form of sexual weaponry in male red deer and are associated with increased breeding success. I show that antler morphology is highly repeatable over an individual's lifetime, heritable, and that almost all examined antler morphology traits are positively genetically correlated. The investigation of causative loci underlying antler morphology showed that it has a highly polygenic architecture, with some loci identified as having potential pleiotropic effects. These findings suggest that a large mutational target and genetic covariances among antler traits, in part maintained by pleiotropy, are likely to contribute to the maintenance of genetic variation in antler morphology in this population.

In species with two sexes, males and females largely share the same genome. However, due to differing biology and life histories, the sexes can have divergent fitness optima for the same traits, thus resulting in sexually antagonistic selection, i.e., sexual conflict. Evidence for sex-specific transmission distortion was found in the deer, which pointed towards sexually

divergent selection on parental gametes and prenatal viability. No genetic loci underlying differential mortality between males and females at later life stages could be identified. Among the loci under sex-specific selection, several showed evidence of balancing selection and could be linked to candidate genes. These results support ongoing sex-specific, but not sexually antagonistic, selection on parental haplotypes and offspring genotype during prenatal development, with potential trans-generational effects. These selection patterns likely contribute to the maintenance of genetic variance and thus evolutionary potential in this population of red deer.

Lay abstract

Some traits are selected because they increase an animal's chances to find a mate and reproduce. These so called sexually selected traits show large variation and rapid evolution across the animal kingdom. Sexually selected traits vary among individuals on the level of the phenotype, meaning the observable characteristics of the trait, as well as on the genetic level, meaning the genotypes of the genes that underlie the trait. This genetic variation is expected to be reduced over time due to strong selection, but is often maintained in the wild. Some explanations for this phenomenon include opposing selection in the two sexes or trade-offs with other life-history traits, such as survival. A key step in uncovering evolutionary drivers and constraints of sexually selected traits is to understand them at the level of the genome, the whole of an individual's genetic material. However, few research studies have attempted to investigate this aspect in the wild. Using long-term data from a population of red deer on the Isle of Rum, Scotland, this thesis both investigates the genomic basis of antler morphology, an important sexually selected trait, and genome-wide signals of sexually antagonistic selection by identifying signals of sexually divergent selection on viability.

Antlers are a form of sexual weaponry in male red deer that are cast and regrown every year before the rut, the mating season, and they are associated with increased success in siring offspring. I show that antler morphology traits are highly repeatable over an individual's lifetime, meaning a stag's antlers are recognisable and unique to the individual. I also show that antler morphology is heritable, which means substantial genetic variation is associated with the examined antler traits and that almost all of these traits are positively correlated on the genetic level. The investigation of causative loci underlying antler morphology showed that it has a highly polygenic

architecture, meaning that the genetic basis consists of many genes with small effects on the phenotype. Some genetic loci were identified as having potential pleiotropic effects, which means one locus influences more than one antler trait. These findings suggest that many new mutations that are introduced through the polygenic architecture and genetic covariances among antler traits, in part maintained by pleiotropy, are likely to contribute to the maintenance of genetic variation in antler morphology in this population.

In species with two sexes, males and females largely share the same genome. However, due to differing biology and life histories, the sexes can have different optima for the same traits that increase their fitness and this can result in opposing selection in the two sexes, so-called sexual conflict. If the inheritance of genetic variants at a locus deviates from expected Mendelian inheritance patterns of 1:1, this is called transmission distortion and can be a signal of sexual conflict. Evidence for sex-specific transmission distortion was found in the deer, which pointed towards sexually divergent selection on parental gene variants and prenatal viability of offspring. No genetic loci underlying differential mortality between males and females at later life stages could be identified. Among the loci under sex-specific selection, several showed evidence of balancing selection, which can be an indicator for sexual conflict, and could be linked to candidate genes. These results support ongoing sex-specific, but not sexually antagonistic, selection on parental genetic variants and offspring genotype during prenatal development, with potential trans-generational effects. These selection patterns likely contribute to the maintenance of genetic variance and thus evolutionary potential in this population of red deer.

Acknowledgements

This PhD project was funded by a NERC E3 Doctoral Training Programme PhD Studentship. Permission to work was granted by Nature Scot (formerly Scottish Natural Heritage). Philip Ellis prepared samples for DNA extraction and the Wellcome Trust Clinical Research Facility Genetics Core in Edinburgh performed the genotyping. This work made extensive use of the University of Edinburgh Compute and Data Facility (http://www.ecdf.ed.ac.uk/). The long-term project on Rum red deer is funded by the UK Natural Environment Research Council (NERC) and the SNP genotyping was funded by a European Research Council Advanced Grant to Josephine Pemberton.

First and foremost I want to thank my supervisor Susan Johnston. Susie is not only a great researcher and scientific role model to me but also a wonderful mentor. I genuinely think I could not have asked for a better supervisor to guide me on this PhD journey. Susie has always been patient, supportive, understanding and helpful. I always felt like I could ask her anything and tell her even my most silly ideas without having to be embarrassed. Susie was always there to help with coding problems and scientific advice, but she also always encouraged my own ideas. She has always made time for me even when it was difficult, like when she was on maternity leave for her first child, and she was always available for urgent (and sometimes only perceived urgent) questions via late-night emails. Even a global pandemic did not stop her from supporting me as if she didn't have a lively two year old to look after at home while the world was also on fire. Thank you so much for everything you have done for me Susie, I could not have done this PhD with anyone else as my supervisor, especially not while having so much fun and having a good gossip now and then as well. My second supervisor, Josephine Pemberton, is another great role model and a scientist I greatly admire. Her extensive

knowledge on the deer and on all things ecology and evolution have been invaluable to me and her enthusiasm for my work has always encouraged me. I would also like to thank Jisca Huisman and Loeske Kruuk for commenting and improving my work and Jisca especially for being my go-to deer person to ask questions when I first started and now being my twitter stats and coding advice guru.

This project would not have been possible if not for the incredible long-term Rum deer data set and the dedicated work of managing it all of Josephine Pemberton especially. I also want to thank Alison Morris, Sean Morris, Martin Baker, Fiona Guinness, Tim Clutton-Brock and many others for collecting field data and DNA samples over the course of the long-term study. Ali and Sean were my guides to field work and the deer when I joined my first calving season and I'm incredibly grateful for their knowledge and advice and for making tough field work an enjoyable time, even when I was covered in midges.

I'd like to thank the Johnston lab group members who joined later in my PhD journey and became my lab siblings and support system. Thank you to John, Anna, Martin and our honourable lab member Cathrine for all the laughs, the serious and silly conversations, the stats and coding advice and for keeping me sane during a global pandemic. I also want to thank my office mates, Cat, Ed, Jess, Wei, Jamie, Libbi and Kirsty for all the wonderful and random conversations we had while we were supposed to work, but also for helpful advice and for commiserating with me. I am also grateful for the postdoc office next door, for Eryn, Joel and Julie who have all helped me out with stats and general advice about life in academia and who have encouraged me when the imposter syndrome was strong.

I was fortunate to have a wonderful PhD cohort. In particular I want to thank, Heather, Jess, Cammy, Camille and Koorosh for being with me through the highs and lows of the last four years. Thank you for listening to me complain and rant, for commiserating with me and for supporting and encouraging me.

Thanks for the laughs, the weird and wonderful conversations (looking at you Koorosh) and the many coffee breaks and pub evenings. Thank you in particular to Jess who has been my long-term office desk neighbour and has listened to my worries and frustrations and indulged me in obsessive stats chats (are evolvability and heritability correlated?). And a special huge thank you to Heather, who has been my flatmate and closest friend for most of my PhD and was one of my biggest supporters. I don't know what I would have done without you Heather, I wouldn't have wanted to do endless lock-downs with anyone else. Thank you for always being there for me, for cheering me on and for nursing me back to health when I was really ill. And also thank you to Heather's wonderful parents, Sue and Garry, who sent me flowers when I was in hospital and offered me to stay with them if I didn't find a job after my PhD.

Finally, I want to thank my family, my parents and my sister, from the bottom of my heart, for always supporting and believing in me. Thank you Mami und Papi for always having my back and for raising me with so much love and the knowledge that I can always count on you. And especially thank you to my wonderful twin sister Stella, my best friend and soulmate, mein Schnecki, who has always been there for me the most, even thousands of miles away, and believed in me when I didn't. Thank you for being my rock and my anchor, I couldn't have done it without you! Ich hab dich unglaublich lieb, Schneckilein! This thesis is for all of you.

Contents

Declaration ii

Abstract iii

Lay abstract v

Acknowledgements vii

1 General Introduction: The genomic basis of sexually divergent traits in red deer (*Cervus elaphus*). 1

1.1 Introduction 1

1.2 Quantitative genetics 3

1.2.1 Fundamentals of quantitative genetics 3

1.2.2 Genetic architecture 6

1.2.3 Quantitative genetics and genomics in the wild 9

1.3 Sexually selected traits 12

1.3.1 Types of sexually selected traits and their evolutionary importance 12

1.3.2 The genetic architecture of male weaponry 15

1.4 Sexual conflict 17

1.4.1 Types of sexual conflict and their evolutionary dynamics . 18

1.4.2 Genomics of sexual conflict 20

1.5 The red deer study population on the Isle of Rum 26

1.5.1 Study system 26

1.5.2 Data collection 27

1.5.3 Red deer antlers 29

1.5.4 Life history and sex differences 31

1.5.5 Quantitative genetics tools 32

1.6 Thesis aims and outline 33

2 Pedigree and genomic approaches to estimate quantitative genetic parameters of antler traits in a wild population of red deer (*Cervus elaphus*). 35

2.1 Summary 35

2.2 Introduction 36

2.3 Methods 42

2.3.1 Study system 42

2.3.2 Antler measures 42

2.3.3 Genomic data-set 44

2.3.4 Principal components of antler measures 45

2.3.5 Estimating heritability using the animal model 46

2.3.6 Data Statement 49

2.4 Results 49

2.4.1 Comparison of pedigree and genomic relatedness 49

2.4.2 Principal component analysis of antler measures 49

2.4.3 Animal models of antler traits 51

2.4.4 Genetic correlations between antler measures 56

2.5 Discussion 58

2.5.1 Genomic versus pedigree relatedness 58

2.5.2 Heritability and repeatability of antler morphology. 59

2.5.3 Genetic correlations and constraints on antler morphology. 60

2.6 Conclusions 61

3 Genomic approaches to understanding the genetic architecture of antlers in red deer (*Cervus elaphus*). 63

3.1 Summary 63

3.2 Introduction 64

3.3 Methods 67

3.3.1 Study system and antler morphology 67

3.3.2 Genomic data-set 67

3.3.3 Genome-wide association studies 68

3.3.4 GWAS power analysis 69

3.3.5 Regional heritability analysis 70

3.3.6 Estimation of SNP effect size distribution 71

3.4 Results 73

3.4.1 Genome-wide and regional heritability studies 73

3.4.2 Distribution and quantification of SNP effect sizes 77

3.5 Discussion 80

3.5.1 The polygenic architecture of antler morphology. 81

3.5.2 Genomic regions associated with antler morphology. . . . 83

3.5.3 Shared genetic architecture and constraints on antler morphology. 85

3.6 Conclusions 87

4 Genomic signatures of sexual conflict in red deer (*Cervus elaphus*). 88

4.1 Summary 88

4.2 Introduction 90

4.3 Methods 98

4.3.1 Study system and genomic data-set 98

4.3.2 SNP clustering 99

4.3.3 Haplotype phasing 102

4.3.4 Quantifying transmission distortion 103

4.3.5 Gene-drop simulations 105

4.3.6 Imputation of missing genotypes 106

4.3.7 Bayesian mixed model of early life survival 107

4.3.8 Signatures of selection 109

4.3.9 Candidate locus annotation and functional enrichment analysis 110

4.4 Results 112

4.4.1 Signatures of transmission distortion 112

4.4.2 Selection signals of sexual conflict loci 116

4.4.3 Genomic signals of sex differences in early life survival . 119

4.4.4 Genomic regions associated with sexual antagonism . . . 120

4.5 Discussion 125

4.5.1 Sex-differences in prenatal viability selection 125

4.5.2 Genetic basis of early life survival 131

4.5.3 Candidate genes associated with differential viability selection. 133

4.5.4 Advances, limitations and future work. 135

5 General Discussion. 139

5.1 Aims and main findings 139

5.2 Novel analysis approaches and findings 141

5.3 Limitations of quantitative genetics and genomics studies in wild populations 143

5.4 Future directions 145

Bibliography 171

Supplemental Information 172

Chapter 2 172

Supplemental Figures 172

Supplemental Tables 172

Chapter 3 177

Supplemental Figures 177

Supplemental Tables 180

Chapter 4 192

Supplemental Figures 192

Supplemental Tables 198

Chapter 1

General Introduction: The genomic basis of sexually divergent traits in red deer (*Cervus elaphus*).



Stag with harem, Rum.

1.1 Introduction

Sexually selected traits show large variation across the animal kingdom, ranging from morphology and behaviours that increase attractiveness to the opposite sex to traits that increase intersexual competitiveness for access to mates (Andersson, 1994). These traits are frequently subject to strong directional selection and evolve rapidly between species, with consequences for adaptation and speciation (Kingsolver *et al.*, 2001; Lorch *et al.*, 2003; Ritchie, 2007; Servedio & Bürger, 2014; Wilkinson *et al.*, 2015; Martínez-Ruiz & Knell, 2017). Evolutionary theory predicts that strong sexual selection will reduce genetic variation within populations, yet empirical studies in the wild

often reveal substantial underlying genetic variation, presenting a key paradox in evolutionary biology (Pomiankowski & Moller, 1995; Kotiaho *et al.*, 2001; Kruuk *et al.*, 2002; Svensson & Gosden, 2007). However, resolving this paradox has been limited due to a lack of knowledge of genetic architecture of sexually selected traits - that is, the loci underpinning this variation, as well as their number, effect size distribution and the extent of pleiotropy, epistasis and other interactions (Timpson *et al.*, 2018; Chenoweth & McGuigan, 2010). By identifying the genetic architecture, we can better understand the underlying molecular mechanisms and evolutionary processes that drive their variation (Dobzhansky, 1971; Lewontin, 1974; Kuijper *et al.*, 2012; Wilkinson *et al.*, 2015). Recent genomic advances in natural populations have led to several studies characterising genetic architectures using genome-wide association studies (GWAS, reviewed in Santure & Garant 2018), yet relatively few studies exist for sexually selected traits.

In this thesis, I will explore two aspects of the genetic architecture of sexually selected traits, using long-term genomic data from a wild population of red deer (*Cervus elaphus*) on the Isle of Rum, Scotland. First, I use genomic approaches to investigate the heritability and genetic architecture of antler morphology, an important sexually selected trait. Second, I characterise signals of sexually antagonistic selection across the genome by tracking the transmission of alleles from parents and offspring of different sexes, and their association with male and female juvenile survival. In this general introduction, I will discuss the background to the approaches used and the broader evolutionary questions that are presented. I will begin by discussing quantitative genetics, from the fundamental statistical basis, to how this is adapted for determining genetic architectures, including in wild populations. I will then discuss the evolutionary importance of sexual selection and what is known about its genetic architecture, particularly in relation to sexual weaponry, an important trait for intra-male competition for access to females. I will then discuss types of sexual conflict and their evolutionary dynamics

within the genome. Finally, I will introduce the red deer study population and present the detailed aims and objectives of this PhD thesis.

1.2 Quantitative genetics

1.2.1 Fundamentals of quantitative genetics

Quantitative genetics is the study of the genetic variance of complex traits (Hill, 2010). Classic Mendelian traits are typically categorical and controlled by one or very few genes, whereas quantitative traits are continuously varying, with a phenotype P in a given individual being the sum of both its genetics G (predicted to be many loci of small effect) and the environment E that it experiences (i.e., $P = G + E$). The study of quantitative genetics generally dissects the overall phenotypic variance of a trait within a population, V_P, into its genetic (V_G) and environmental components (V_E):

$$V_P = V_G + V_E \tag{1.1}$$

The genetic variance V_G can be further divided into variance due to additive, dominance and epistatic effects; typically, the genetic variance component of interest in quantitative genetics is the additive genetic variance V_A, which is the genic variance passed from parent to offspring, disregarding dominance variance and epistasis. The core quantity derived from the additive genetic variance of a trait is the (narrow-sense) heritability, h^2, which is the proportion of overall phenotypic variance that is represented by additive genetic variance (Falconer & Mackay, 1996; Lynch & Walsh, 1998; Hill, 2010):

$$h^2 = V_A/V_P \tag{1.2}$$

The heritability of a trait is an important variable in both animal and plant breeding and evolutionary biology, as it allows for the prediction of change in the mean of a trait of interest under selection. The formula to predict change in mean phenotype per generation is the so-called breeder's equation, first proposed by Lush in 1937:

$$\Delta Z = h^2 S \tag{1.3}$$

where ΔZ is the change in mean phenotype, h^2 is the trait heritability and S is the selection differential. Thus, based on this equation, higher heritability values can result in a stronger and rapid response to selection that is sustained over many generations (Barton & Keightley, 2002). In artificial selection, S is calculated as the difference between the trait mean of the parents chosen for reproduction and the mean of the rest of the population, while in wild populations it is derived from the covariance of fitness and the trait of interest. This latter selection differential is known as the Robertson-Price identity, given by

$$S = \sigma(z, w) \tag{1.4}$$

where S is the selection differential and $\sigma(z, w)$ is the covariance between the phenotypic trait z and relative fitness w, which represents the difference between the trait mean of the population before and after selection (Falconer & Mackay, 1996; Arnold & Wade, 1984; Price, 1970). Thus, heritability is an essential quantity in quantitative genetics and evolutionary biology for understanding how much variation in traits is due to genetic variants and for

inferring how traits might respond to a given selection pressure.

Traditionally, trait heritability was calculated via parent-offspring or sometimes full-sib or half-sib regressions (Falconer & Mackay, 1996; Lynch & Walsh, 1998). In these regressions, the phenotypic values of a trait of e.g., the parents are regressed on the phenotypic values of their offspring, with the slope of the regression representing the heritability estimate. However, simple regressions can over-estimate heritabilities as they cannot account for common environmental factors or other indirect effects. This is because relatives are more likely to share their environment, especially during early life (i.e., their brood or maternal environment) which might make them appear more similar to each other independently of what is due to their shared genes (Kruuk, 2004; Kruuk & Hadfield, 2007). Other indirect effects that can have similar consequences for heritability estimates are temporal environmental effects that increase similarity among individuals due to shared experience at a particular period in time.

These issues have been addressed by more complex statistical methods, including mixed-effects models that integrate relatedness information between measured individuals (Kruuk, 2004; Kruuk & Hadfield, 2007). One such model, the ”animal model” was first developed for animal breeding studies; it decomposes phenotypic variation into additive genetic and multiple environmental sources, fitting a genetic relationship matrix to estimate the additive genetic variance component of a trait (Henderson, 1975). The model structure is typically the following:

$$y = X\beta + Z_1 a + Z_r u_r + e \tag{1.5}$$

where y is a vector of the trait of interest, X is an incidence matrix relating individual measures to the vector of fixed effects β; Z_1 and Z_r are incidence

matrices relating individual measures with additive genetic and other random effects, respectively; a is a vector of additive genetic effects; u_r is a vector of additional random effects; and e is a vector of residual effects. The relatedness matrix is traditionally derived from pedigree information, which gives the expected relatedness between individuals. Furthermore, models can also be expanded to investigate genetic correlations between traits, to understand to what extent heritable variation is shared (Kruuk, 2004). More recently, this has been estimated using genomic relationship matrices (GRM), which estimates relatedness based on identity by state (IBS) of genome-wide markers (VanRaden, 2008). The relatedness matrix is used to quantify the additive genetic variance component from the animal model, which can then be used to calculate the heritability of a trait in a given population. Over the last two decades, the use of pedigree reconstruction techniques using molecular markers in a wide range of species means that the animal model has become widely used in quantitative genetics analysis of wild populations (Kruuk *et al.*, 2002; Johnston *et al.*, 2009, 2016, 2018; Bérénos *et al.*, 2014; Berenos *et al.*, 2015; Miller *et al.*, 2018). This development represents an important advancement for quantitative genetics in natural populations, since, as mentioned above, there are several confounding factors in the wild that can be controlled in domestic or lab animals, such as pedigree structure, population density and cohort effects due to fluctuating environments, which can be accounted for using the animal model.

1.2.2 Genetic architecture

Knowing trait heritabilities allows for the prediction of evolutionary responses, but it is based on the assumption that the genetic architecture of complex traits is characterised by an infinite number of unlinked loci of infinitesimal small additive genetic effects. This model was first proposed by Fisher (1919) and later formalised by Bulmer (1987). Although its assumptions appear

unrealistic, it provides a solution to predict evolutionary trajectories without knowing the causative loci of a trait and has been used successfully in quantitative genetics for decades (Hill, 2010). This is because the infinitesimal model implies that allele frequency changes at individual loci are negligible, so that genetic variance stays virtually constant across generations unless selection is very intense and will recover if selection relaxes (known as the 'Bulmer effect'; Hill 2010; Barton & Keightley 2002; Barton *et al.* 2017). This has meant that trait responses to selection can be estimated without identifying the underlying loci, their effects, and allele frequencies.

However, with the advancement of genetic and genomic tools, it has been made possible to identify the underlying causative loci of traits. Most traits are likely to have a complex genetic architecture: they may be controlled by genes of different allele frequencies and effect sizes, including loci with smaller or larger effects on phenotype. Identifying the distribution of causal loci, and large effect loci in particular, can have important real-life impacts. First, we can detect loci that underlie diseases, especially Mendelian disorders or those within simpler (oligogenic) architectures, and use this information to develop targeted treatments (Alkuraya, 2016). Second, we can use our knowledge of the genetic basis of traits to improve livestock or crop species via targeted genomic selection or even gene editing (Wakchaure & Ganguly, 2015; Collard *et al.*, 2005; Jiang & Shen, 2019). Finally, characterising the genetic architecture of a trait can help to improve our understanding of trait evolution and its constraints and potential, in particular of traits with major quantitative trait loci (QTLs). Identifying the number and distribution of QTLs can also help to understand the importance of dominance, pleiotropy, and epistasis occurring within and between these loci, which models of additive genetic variance alone will fail to capture. Non-additive variance is largely ignored in quantitative genetics, but has been shown to be of significant importance and can inflate estimates of additive genetic variance (Mackay, 2014; Huang & Mackay, 2016; Johnson & Barton, 2005).

Classic QTL studies have relied on genetically distinct strains or lines that are crossed to achieve segregating sites linked to known genetic markers, which until the mid 2000s were typically allozymes (allelic variants of single proteins) or mini- and microsatellites (sections of highly repeated DNA motifs of different lengths). These were used to detect QTLs based on linkage with either single markers or pairs of flanking markers ('interval mapping'; Thoday 1961; Paterson *et al.* 1988; Simpson 1989; Lander & Botstein 1989; Haley & Knott 1992; Shaw *et al.* 2007). These methods were then adapted to make use of available relatedness/pedigree information to map QTLs using markers segregating in populations of livestock and humans (Soller & Genizi, 1978; Soller & Beckmann, 1983; Geldermann *et al.*, 1985; Weller *et al.*, 1990; Comuzzie *et al.*, 1997). This led to the development of more complex maximum likelihood approaches enabled QTL mapping in wild populations with complex pedigrees (George *et al.*, 2000; Beraldi *et al.*, 2006; Johnston *et al.*, 2009). These methods, though effective, were also costly, requiring not only the slow development of new molecular markers that were almost always species-specific, but also multiple generations of breeding to built up linkage and increase power to detect QTLs.

With the rapid advancement of affordable and high-throughput sequencing and genotyping technology over the last two decades, genome-wide markers such as single nucleotide polymorphisms (SNPs) are increasingly used to detect and fine-map QTLs using Genome-Wide Association Studies (GWAS), which can also be carried out in unrelated individuals (Timpson *et al.*, 2018; Flint & Mackay, 2009). Association studies rely on genotyping SNPs in linkage disequilibrium (LD), i.e., close physical association with a causative locus. The extent of LD can differ between mapping populations, but generally GWAS requires high marker densities to ensure linkage with causal loci (Bush & Moore, 2012; Pritchard & Przeworski, 2001), although this can be particularly difficult in non-model organisms (Ekblom & Galindo, 2011). GWAS has been very successful as an approach in mapping human traits for

example, having identified tens of thousands causative trait loci for several thousand traits, including many health and disease traits and morphological traits such as height and BMI (MacArthur *et al.*, 2017). Characterising genetic risk factors underlying disease has been particularly successful, resulting in the development of targeted drugs for conditions like type 2 diabetes and autoimmune diseases such as rheumatoid arthritis and ankylosing spondylitis (Visscher *et al.*, 2017). Still, a large part of the genetic architecture remains unresolved for many traits including human height, one of the most highly heritable human traits ($h^2 \approx 0.8$), for which only a fraction of causative loci has been identified in GWASs, despite having data from millions of SNPs and hundreds of thousands of people (Wood *et al.*, 2014; Yengo *et al.*, 2018). This is because the power to detect QTL depends on the magnitude of its effect size (Lynch & Walsh, 1998; Sham *et al.*, 2000), and these effect sizes are often inflated ("Beavis effect"), especially when sample sizes are small (Beavis, 1998; Xu, 2003; Slate, 2013). This makes it difficult to detect causative loci of moderate or small effect, despite evidence pointing towards polygenic architectures (i.e., many genes of small effect) to be more common across many species, including model organisms, livestock, humans and natural populations (Hayes & Goddard, 2001; Flint & Mackay, 2009; Yang *et al.*, 2011; Berenos *et al.*, 2015; Santure *et al.*, 2013, 2015).

1.2.3 Quantitative genetics and genomics in the wild

Quantitative genetic studies in wild populations have addressed important questions in ecology and evolution, from sexual selection (Emlen, 1994; Kruuk *et al.*, 2002; Coltman *et al.*, 2005; O'Sullivan *et al.*, 2019), behaviour and life-history variation (Berthold & Pulido, 1994; Facon *et al.*, 2008; Evans *et al.*, 2020), to responses to climate change (Charmantier & Gienapp, 2014; Bonnet *et al.*, 2019a; Kingston *et al.*, 2018), immune variation (Hammerschmidt *et al.*, 2011; Sparks *et al.*, 2019; Gauzere *et al.*, 2021) and

ageing (Nussey *et al.*, 2009; Froy *et al.*, 2019; Asghar *et al.*, 2015). These studies have ranged from basic investigations of trait heritability to more detailed examinations of genetic correlations between phenotypic traits of interest and individual fitness (Kruuk *et al.*, 2002; Coltman *et al.*, 2005; O'Sullivan *et al.*, 2019). However, for many years, quantitative genetic studies in wild populations were (and often still are) limited to intensely monitored long-term studies of mammals and birds in which pedigrees could be constructed from observational or molecular data for parentage analyses over several generations of individuals (Pemberton, 2008; Jones *et al.*, 2010). Pedigree construction was based mostly on species-specific molecular markers such as microsatellites, often aided by field observations and information on individuals' age, which makes parentage assignment more powerful (Walling *et al.*, 2010). Trait mapping studies were also challenging due to small numbers of offspring per mating, longevity, conservation concerns and feasibility (Slate *et al.*, 2010). However, QTL mapping studies had some success in pedigreed populations with larger numbers of microsatellite markers (i.e., 200 – 300) in several hundred individuals (Beraldi *et al.*, 2006; Slate *et al.*, 2002; Poissant *et al.*, 2012), although these studies were comparatively low in power and only able to map loci of large effect, typically Mendelian traits such as horn type and coat colour in Soay sheep (Beraldi *et al.*, 2006).

The modern era of tractable high-throughput genotyping and sequencing means that, over the last decade, genome-wide approaches are more readily available at much lower costs for greater numbers of individuals in almost any species (Santure & Garant, 2018). In wild populations, genome-wide SNP markers are now widely used to understand trait genetic variance using SNP-based pedigrees and GRMs (Bérénos *et al.*, 2014; Huisman, 2017) and to identify QTLs using GWAS approaches (Ellegren, 2014; Schielzeth & Husby, 2014; Santure & Garant, 2018). Markers can be developed from reduced representation sequencing approaches, such as

genotyping-by-sequencing or restriction site associated DNA sequencing (RADseq) to discover and characterise segregating SNPs (Davey *et al.*, 2011). In addition, population- and species-specific SNP genotyping arrays (or 'SNP-chips') are used to genotype individuals at thousands to millions of polymorphic SNP loci (Pavan *et al.*, 2020). These genome-wide approaches have revolutionised quantitative genetics in the wild and have resulted in many more studies contributing to our understanding of quantitative genetics.

GWASs in wild systems have faced a number of challenges, as they often have much fewer individuals and lower marker densities compared to studies in humans or livestock, which often have access to several hundred thousand individuals and arrays with millions of genetic markers (Santure & Garant, 2018; Uffelmann *et al.*, 2021). These constraints have been overcome in some intensely studied populations with large number of individuals and markers (Berenos *et al.*, 2015; Johnston *et al.*, 2011, 2016, 2018; Santure *et al.*, 2015; Robinson *et al.*, 2013; Malenfant *et al.*, 2018). Another challenge that must be addressed in wild GWAS studies are confounding factors such as pronounced relatedness and population structure, and correlated environmental effects, which can lead to spurious associations with the trait of interest (Santure & Garant, 2018; Campbell *et al.*, 2005). These challenges can be tackled with statistical approaches described above, such as using mixed models (i.e., Equation 1.5) that fit a relationship matrix to account for close relatedness and population structure, and other random effects to account for common environmental effects on phenotype (Aulchenko *et al.*, 2007; Sul *et al.*, 2018). Other methods to account for structure include the construction of a genetic distance matrix using a multi-dimensional scaling approach to account for distinct populations within a data set (Price *et al.*, 2006).

To date, GWAS has been successful in identifying medium- to large-effect QTLs in several wild populations: In Soay sheep (*Ovis aries*), specific genes

and genomic regions have been associated with horn development (Johnston *et al.*, 2011), hindleg length (Berenos *et al.*, 2015), recombination rate (Johnston *et al.*, 2016) and immunoglobulin levels (Sparks *et al.*, 2019); and in red deer (*Cervus elaphus*), genomic regions have been associated with recombination rate (Johnston *et al.*, 2018). Another example of a large effect QTL detected in a wild species without pedigree information is that of the *Vgll3* locus, a gene that controls age at maturity in Atlantic salmon (*Salmo salar*) and accounts for $\sim$ 39% of all phenotypic variation (Barson *et al.*, 2015). Even moderate effect size loci of quantitative traits have been identified in natural populations, such as body length in polar bears (*Ursus maritimus*), for which a candidate region was identified that is likely involved in regulatory cascades of cyclin expression with known effects on body size in mice (Malenfant *et al.*, 2018). However, in many cases GWASs still fail to identify significant phenotype-marker associations (Robinson *et al.*, 2013; Berenos *et al.*, 2015; Santure *et al.*, 2015; Yang *et al.*, 2010). Nonetheless, GWAS is still a relatively easy and accessible exploratory tool for understanding the genetic framework underlying traits of interest in wild populations.

1.3 Sexually selected traits

1.3.1 Types of sexually selected traits and their evolutionary importance

Sexually selected traits such as exaggerated ornaments, court-ship behaviour and male weaponry show great variety and complexity across the animal kingdom; they are typically characterised by strong directional selection and in the most extreme cases this can lead to speciation via rapid population divergence (Lande, 1981; Civetta & Singh, 1998). Two of the main drivers of

the evolution of sexual traits are intra-male competition and female choice. In natural systems where males compete with each other for mating access to females, male weaponry such as horns in bovids (Bro-Jørgensen, 2007) and antlers in deer (Clutton-Brock *et al.*, 1982) can evolve via sexual selection to allow for the resolution of male-male conflict (Andersson, 1994). Female mate choice, acting on a range of different traits, can drive the evolution of elaborate male sexual traits via genetic correlations between female preference and male traits (Lande, 1981). There are several possible mechanisms explaining the evolution of mate choice: direct phenotypic benefits reflected by the preferred trait or ornament, such as good-quality territory, nutrition, parental care or protection (Moller & Jennions, 2001); sensory bias where female preference favours an ornament that initially evolved under natural selection, for example foraging ability or predator avoidance (Endler & Basolo, 1998); and indicator mechanisms such as 'good genes' or genic capture (Rowe & Houle, 1996). These latter indicator mechanisms predict that attractive males reflect overall good genetic quality, i.e., high genetic variance (Folstad & Karter, 1992), favourable alleles or low mutational load (Lorch *et al.*, 2003; Rowe & Houle, 1996); traits associated with good quality are preferred by females as they provide genetic benefits to their offspring (Neff & Pitcher, 2005; Bonduriansky & Rowe, 2005). In both male-male competition scenarios and female choice scenarios, evolution should erode the genetic variance underlying sexually selected traits over time, especially considering that sexual selection is thought to be the strongest form of directional selection in the wild (Kingsolver *et al.*, 2001). Thus, when a trait optimum value is eventually reached, it is implied that the alleles conferring this value are fixed within a population, with little phenotypic or genetic variance remaining (Chenoweth *et al.*, 2012).

These predictions are based on a long-standing theoretical framework (Darwin, 1981; Fisher, 2011; Lande, 1981), yet empirical studies commonly find that sexually selected traits in the wild have substantial genetic variation

despite assumed directional sexual selection (Pomiankowski & Moller, 1995; Rowe & Houle, 1996; Kruuk *et al.*, 2008), and may show no evolutionary response to selection pressure (Merila *et al.*, 2001; Svensson & Gosden, 2007). These findings directly contradict evolutionary theory on sexual selection and hence represent an interesting evolutionary paradox. A number of possible explanations for this phenomenon have been explored, including differences in selection between the sexes, developmental stages or environmental conditions (Bourret *et al.*, 2017; Barson *et al.*, 2015), phenotypic plasticity (Charmantier & Gienapp, 2014), condition dependence (Dugand *et al.*, 2019) and trade-offs with survival (Johnston *et al.*, 2013a). Some of these observations could be due to genetic correlations between traits under opposing selection pressures and linkage disequilibrium between alleles at causal loci and deleterious alleles (Lande, 1982; Lande & Arnold, 1983; Connallon & Hall, 2018). Furthermore, the genetic architecture of a trait, in particular how many genes contribute to its genetic variation, plays an important role in the persistence of genetic variance. As polygenic traits present a large mutational target, the introduction of new variants contributes to the maintenance of genetic variance (Rowe & Houle, 1996). Additionally, the evolutionary response to selection can be sustained by a large mutational input, so that the distribution of genetic effects and thus genetic variance is left unperturbed (Barton *et al.*, 2017; Sella & Barton, 2019).

Quantitative genetic studies have provided some valuable insights into the different explanations for evolutionary stasis, by estimating the relative contributions of genetic (i.e., the heritability, h^2) and environmental effects to phenotypic variance, as well as examining phenotypic and genetic correlations with other traits, including fitness (Emlen, 1994; Griffith *et al.*, 1999; Kruuk *et al.*, 2002; Robinson *et al.*, 2006). Some studies have demonstrated the importance of exploring multivariate selection on several sexually selected traits simultaneously: studies in the fruit flies *Drosophila bunnanda* and *D. serrata* analysing the genetic variance of multiple cuticular

hydrocarbons and the strength and direction of selection acting on them, discovered that while substantial genetic variance was present, it was not aligned with the direction of sexual selection for most traits and thus unable to respond to selection pressure (Van Homrigh *et al.*, 2007; Reddiex & Chenoweth, 2021). Another study in black field crickets found that while single male call traits were associated with large genetic variances, the genetic variance for specific combinations of genetically correlated call traits under directional or stabilising selection was depleted (Hunt *et al.*, 2007). Thus, multivariate sexual selection can constrain evolution if it is not aligned with the genetic variance, or if the genetic variance is already depleted in a specific trait combination under selection, even if single traits show high levels of genetic variance.

1.3.2 The genetic architecture of male weaponry

Despite evolutionary biologists' long-standing and intense interest in sexual selection and sexually selected traits, and despite the rapid acceleration in our knowledge of the genomic basis of variation in phenotypes, we know surprisingly little about the detailed genetic architecture of sexually selected traits in general, and sexual weaponry in particular. Most of our current knowledge is restricted to fundamental quantitative genetic parameters, such as their heritabilities. Focusing on male weaponry, several studies in wild bighorn and thinhorn sheep have found substantial additive genetic variance underlying horn circumference ($h^2 =$ 0.30 - 0.43) and weight ($h^2 =$ 0.23 - 0.36; Coltman *et al.* 2005; Poissant *et al.* 2012; Sim & Coltman 2019), which appears to be unconstrained and available to selection in some cases (Coltman *et al.*, 2005). In Cervids, most information about the inheritance of antler traits was previously generated by studies on captive bred populations, showing moderate to high levels of heritability for various antler traits in white-tailed deer (*Odocoileus virginianus*, beam circumference h^2 = 0.29,

beam length h^2 = 0.14, antler mass h^2 = 0.43; Lukefahr & Jacobson 1998; Williams *et al.* 1994), wapiti (*Cervus canadensis*, antler velvet mass h^2 = 0.27; Wang *et al.* 1999) and red deer (*Cervus elaphus*, antler velvet mass h^2 = 0.36; Van Den Berg & Garrick 1997). As populations experience different environments, selection pressures and demographic processes in the wild compared to captivity, genetic and residual variance and thus heritabilities are expected to differ. Evidence from the long-term study population of wild red deer on the Isle of Rum suggests similar moderate levels of heritabilities as in captive populations (*Cervus elaphus*, antler mass h^2 = 0.38, h^2 brow length = 0.37, h^2 coronet circumference = 0.36) (Kruuk *et al.*, 2002, 2014). Another study in wild white-tailed deer estimated heritability values for antler traits using genomic marker information on largely unrelated individuals (*Odocoileus virginianus*, antler spread h^2 = 0.07, number of points h^2 = 0.33, beam circumference h^2 = 0.14)) (Jamieson *et al.*, 2020), but large standard errors around these values illustrate the need for suitable data sets for quantitative genetic analyses of such populations.

Despite the availability of genomic tools, little is still known about the underlying genetic architecture of male weaponry. Like other sexually selected traits, weaponry is sometimes polymorphic, suggesting major gene effects, and sometimes continuously varying, suggesting that it is under the control of many genes of small effect. However, the latter assumption might not always hold true. In insects, the sex determination gene *doublesex* seems to play a major role in both the development and size of male sexual weaponry, such as in species of rhinoceros beetles, dung beetles, and stag beetles (Kijimoto *et al.*, 2012; Ito *et al.*, 2013; Gotoh *et al.*, 2014). Similarly, in feral Soay sheep, the presence, or absence of horns is a Mendelian trait determined by a major effect locus *RXFP2*, but the same locus also has large effects on horn size in this population, which follows a quantitative trait distribution (Johnston *et al.*, 2011). Yet, in related wild species of bighorn sheep, attempts to identify major effect loci controlling horn size traits have

been inconclusive (Poissant *et al.*, 2012; Miller *et al.*, 2018; Sim & Coltman, 2019); similarly, a study trying to identify the underlying genetic architecture of antler morphology in white-tailed deer was inconclusive to date (Anderson *et al.*, 2019). While a lack of hits could indicate a polygenic architecture, these studies may also have reduced power to detect large effect loci. Therefore, more work is needed to better understand the genetic underpinnings of weaponry in wild living ruminants and of sexually-selected traits in general. However, the persistence of large genetic variance and the widespread condition dependency of sexually-selected traits make it likely that polygenic architectures with a large mutational target should be common (Houle *et al.*, 1996; Rowe & Houle, 1996).

1.4 Sexual conflict

In species with separate sexes, males and females can have very different life history strategies. If an increase in fitness in one sex reduces the fitness of the other sex, selection will act in opposing directions, leading to sexual antagonism or conflict. For example, females often invest more heavily in offspring development and parental care per single reproductive event compared to males. This begins at the level of gametes, as oocytes are much larger than sperm (anisogamy), often less numerous, and contain metabolites and other molecules essential for early embryonic development (Hayward & Gillooly, 2011). This asymmetry has further cascading effects so that female investment may continue with increased incubation/gestation, as well as postnatal food and care provision (Clutton-Brock, 1991; Møller & Birkhead, 1993; Schwagmeyer *et al.*, 1999; Schärer *et al.*, 2012). As a result, females are often ‘choosy’ about whom they mate with, while males often try to mate with a large number of females, leading to deviating interests in female-male interactions such as courtship, fertilisation, and offspring investment. This

division can lead to divergent fitness optima in the two sexes and result in males and females differing drastically in a range of physical and behavioural traits (Andersson, 1994; Rowe *et al.*, 2018).

Whilst sexual conflict can readily arise from sexual selection, it is not an inevitable consequence of it and extends to traits not under sexual selection, as evidenced by extensive sexual dimorphism in non-sexual traits; conflict arises if optima cannot be achieved simultaneously (Parker, 1979, 2006; Arnqvist & Rowe, 2005; Chapman *et al.*, 2003; Kokko & Jennions, 2014). While most sexual conflict studies focus on reproductive traits, since the sexes have distinct life histories in many species, conflict could conceivably also arise through natural rather than sexual selection and affect other aspects of fitness, such as survival. Sexual dimorphism, which can arise as a consequence of resolved sexual conflict, has been proposed to evolve due to ecological competition in some circumstances (Slatkin, 1984; Shine, 1989) and an increasing number of empirical studies show evidence in support of sexual conflict over prenatal viability and survival (Flanagan & Jones, 2017; Wright *et al.*, 2018; Lin *et al.*, 2021), including in humans (Lucotte *et al.*, 2016).

1.4.1 Types of sexual conflict and their evolutionary dynamics

Sexual conflict is often differentiated according to its underlying genetic architecture: inter-locus or intra-locus sexual conflict. **Inter-locus sexual conflict** typically involves separate, sex-specific traits that are mediated by separate loci in each sex (Chapman *et al.*, 2003). Conflict arises when a mutation at a given locus that is beneficial to one sex but harmful to the other increases in frequency. The disadvantaged sex will then favour a mutation at a separate locus that counteracts the harmful action of the mutation at the first

locus, leading to an evolutionary arms race between loci (Rice & Holland, 1997). These across sex correlated traits are typically involved in reproduction or mating: in *Drosophila melanogaster*, seminal fluid increases female mortality, leading to the evolution of female resistance to multiple matings (Chapman *et al.*, 1995; Wigby & Chapman, 2005; Rice *et al.*, 2006). In the Dytiscinae subfamily of diving beetles, males have developed grasping adhesive discs that permit forced mating, while females, in turn, have evolved a structured dorsum to impede attachment (Bergsten *et al.*, 2001). The genetic covariance between the sexes that evolves as a consequence of inter-sexual conflict is an important driver of antagonistic co-evolution of male and female traits, much like runaway sexual selection of female preference and male secondary sexual traits (Chapman *et al.*, 2003; Lande, 1981; Rice *et al.*, 2006) or host-pathogen interactions (Brockhurst *et al.*, 2014; Kasimatis *et al.*, 2017). Consequently, sexual conflict is believed to be of major importance for the emergence and diversification of sexually dimorphic traits, contributing to the great diversity of sexually selected traits and the evolution of highly exaggerated phenotypes (Mank, 2017b; Rowe *et al.*, 2018).

On the other hand, **intra-locus sexual conflict** occurs when the same genes control a specific trait in males and females, where the optimum trait values are different between the sexes (Rice, 1998; Parker, 2006). While evolving towards different trait variant optima, the sexes still share the same species-specific genomic background. Therefore, the combination of shared genetic architecture and diverging fitness interests of the sexes should make intra-locus sexual conflict common throughout the genome. Indeed, several studies have identified loci under sexually antagonistic selection and increasingly genome-wide signals of intra-locus sexual conflict are documented (Cheng & Kirkpatrick, 2016; Dutoit *et al.*, 2018; Ruzicka *et al.*, 2019). Notable examples include: the locus *Vgll3* controls age at sexual maturity in Atlantic salmon, with fitness in males and females associated with the alleles for earlier and later maturation, respectively (Barson *et al.*, 2015);

the locus inducing DDT insecticide resistance *Cyp6g1* in *Drosophila melanogaster* is female-beneficial, but reduces male fitness (Rostant *et al.*, 2015).

The presence of sexual conflict on the one hand can constrain evolution, preventing each sex from reaching its respective fitness optima, and interfere with global population fitness and adaptive potential (Rice, 1998; Bonduriansky & Rowe, 2005; Arbuthnott & Rundle, 2012; Chenoweth *et al.*, 2015). On the other hand, sexual conflict may also drive speciation in cases where antagonistic coevolution between the sexes generates evolutionary divergence of reproductive traits (Arnqvist *et al.*, 2000; Gavrilets, 2014); it may also contribute to the evolutionary potential of a population by maintaining genetic variance via balancing selection (Hall *et al.*, 2010; Rostant *et al.*, 2015; Barson *et al.*, 2015). Regardless of whether conflict is inter- or intra-locus, the genetic aspect of fitness components are expected to be negatively correlated across the sexes if true sexual antagonism is present (Rice & Chippindale, 2001; Chippindale, 2001; Gavrilets, 2014; Kim & Velando, 2016; Okada *et al.*, 2021). Despite our understanding of the selection acting on these traits, some cases for which the genetics of intra-locus sexual conflict are well understood, often if major effect loci are involved (see *Vgll3* and *Cyp6g1* examples above) (Roberts *et al.*, 2009; Rostant *et al.*, 2015; Barson *et al.*, 2015; Ducret *et al.*, 2016), in most cases the genetic architecture of sexual antagonism remains unknown.

1.4.2 Genomics of sexual conflict

Next-generation sequencing methods have allowed the generation of an increasing number of datasets containing genome-wide marker and whole genome sequence and transcriptome information. These data allow the genetic architecture of sexual conflict to be explored in more detail, including

the genomic distribution and number of loci, their effect sizes, the kinds of genes involved and how common sexual conflict may be. In particular, the genomic patterns generated by the presence of intra-locus sexual conflict can be explored this way, as the same loci can be compared directly for signals of antagonism or association with trait variation and fitness in both sexes. As a result, many genomic studies to date have focused more on intra-locus conflict (Mank, 2017a) and/or the extent to which males and females differ in the genetic architectures of common traits (Bernabeu *et al.*, 2021).

A common signature of intra-locus sexual conflict is the evolution of sex-specific gene expression, which can result in phenotypic sexual dimorphism (Mank, 2017b; Ellegren & Parsch, 2007; Wright *et al.*, 2018). The degree to which expression is sex-biased can be used as an indicator of whether conflict is resolved or ongoing (Cheng & Kirkpatrick, 2016; Dutoit *et al.*, 2018). Numerous studies have explored sex-specific transcriptome divergence and identified several loci related to sexual dimorphism in gene expression across a number of taxa (Derome *et al.*, 2008; Yang *et al.*, 2006; Jiang & Machado, 2009; Wright *et al.*, 2018; Bagchi *et al.*, 2021). A study assessing the genetic basis of gene expression regulation in lake whitefish *Coregonus* species for example, found evidence for extensive sex-linked dimorphism in the number of expression QTLs that was associated with the adaptive divergence of the *dwarf* and normal whitefish species pair (Derome *et al.*, 2008). Another study in mice found wide-spread sexual dimorphism in expression levels of thousands of genes across several tissues and sex-biased gene pathways were highly tissue and function specific, for example female biased genes involved in immune/inflammatory response (Yang *et al.*, 2006). Evidence from these kinds of studies implicates large numbers of loci spread throughout the genome, suggesting that sexual conflict can have highly polygenic architectures (Derome *et al.*, 2008; Yang *et al.*, 2006; Jiang & Machado, 2009). This is contradictory to previous findings that sexual conflict involves few loci of large effect, and that they

mainly aggregate on the sex chromosomes in order to resolve conflict (Rice, 1984; Gibson *et al.*, 2002). In Lake Malawi cichlids for example, sexual conflict over the orange-blotch phenotype is resolved by tight linkage of the *OB* locus with a novel female sex determination locus, so that the *OB* haplotype that is beneficial to females but detrimental in males is only carried by females (Roberts *et al.*, 2009). Although sex chromosomes have theorised and shown to play an important role in the resolution of sexual conflict loci (Roberts *et al.*, 2009; Mank & Ellegren, 2009; Jaquiéry *et al.*, 2013), growing evidence suggests that clusters of antagonistic loci are ubiquitous throughout the genome (Wright *et al.*, 2018; Ruzicka *et al.*, 2019; Cheng & Kirkpatrick, 2016). Furthermore, while highly sexually dimorphic transcriptomes are widespread, they are not necessarily representative of sexual antagonism, as sex-biased gene expression does not always necessarily result in sexual dimorphic phenotypes (Mank, 2017b) and some indication of associated fitness consequences is needed (Parker, 2006). As a result, additional lines of evidence, such as sexually divergent allele frequencies (F_{ST}) or elevated patterns of genetic diversity after correction for confounding population dynamics (Wright *et al.*, 2018, 2019), or a connection with sex-specific selection (Immonen *et al.*, 2014; Moghadam *et al.*, 2012) can lend additional support to sexual antagonism as a driving force.

In addition to these population level studies, some studies have leveraged individual level genomic data to look at differences in genetic architectures between the sexes, with a foundation in the quantitative genetic and GWAS approaches outlined above. Such studies can detect sexually antagonistic or sex-specific effects by quantifying genetic correlations between the sexes and conducting GWAS stratified by sex. For example, a study on the genetic architecture of recombination rate in feral Soay sheep found tentative evidence for some degree of sex-specific genetic architecture based on a cross-sex genetic correlation that deviated from r = 1 (i.e., perfect correlation), although standard errors were high (Johnston *et al.*, 2016). The same study

also found sex-limited effects of a genomic region containing the *RNF212* locus on female crossover count in Soay sheep and a similar effect was documented in red deer *Cervus elaphus* where a homologous gene *RNF212B* was associated with female but not male recombination rate variation (Johnston *et al.*, 2018). Such gene-by-sex (*GxS*) interactions seem to be wide-spread throughout the genome: a large scale study by Bernabeu et al. 2021 in humans using UK Biobank data investigated sex effects on the genetic architecture of 530 traits, using over one million SNPs and 450,000 participants, and found evidence for extensive sex-specific effects. In total, 69 traits showed cross-sex correlations that departed from r = 1, and 100 traits had at least one sexually dimorphic SNP in sex-stratified GWASs, including metric traits such as waist:hip ratio and height, and diseases such as ankylosing spondylitis and hypothyroidism. Stratifying by sex to account for *GxS* interactions also enabled the identification of the sexually antagonistic *Vgll3* locus underlying age at sexual maturity in salmon (see above; Barson *et al.* 2015). While antagonistic *GxS* effects indicate ongoing sexual antagonism, sex-specific effects point towards resolved conflict (Wright *et al.*, 2018, 2019). As GWASs become even more prolific, patterns of trait effect size differences between the sexes and their association with fitness-related traits such as reproductive success and survival could be investigated to further our knowledge of sexually antagonistic and sex-limited effects throughout the genome.

Most studies focus on sexual antagonism over reproductive fitness, since sexual dimorphism, a common consequence of sexual conflict, is believed to arise primarily in response to selection on traits related to reproductive fitness (Andersson, 1994). However, in recent years, more attention has been given to the possibility of sexual conflict over viability and survival. Differential survival between the sexes (from the level of the zygote to adulthood) should lead to sex-specific allele frequency divergence, such that females will be enriched for female-beneficial alleles, while males should show an excess of

male-beneficial alleles (Rice & Chippindale, 2001). To detect sexual conflict over survival, a common approach has been to compare sex-specific fixation index (F_{ST}) values, which is a determinant of allelic divergence that can be applied at the level of individual loci (Weir & Cockerham, 1984; Mank, 2017a). Often, this analysis is further complemented by determining Tajima's D of differentiated markers (i.e., where F_{ST} is higher), as it is an indicator of balancing selection, which is another hallmark of sexual conflict (Mank, 2017a; Wright *et al.*, 2018). Theory predicts that the mortality rates necessary to generate significant allelic differences would be too extreme to be detected, as harmful alleles are purged from the population too quickly, (Kasimatis *et al.*, 2017) and empirical evidence suggests that conflict over viability produces much weaker selection than antagonism involving reproductive fitness (Cox & Calsbeek, 2009). Despite these expectations, several studies have identified loci with sex-specific allele frequency divergence consistent with sexual antagonism that acts on survival, ongoing or resolved through sex-specific genetic architecture (Wright *et al.*, 2018; Lucotte *et al.*, 2016; Dutoit *et al.*, 2018; Cheng & Kirkpatrick, 2016; Flanagan & Jones, 2017; Lin *et al.*, 2021).

However, patterns of widespread ongoing sexual conflict (Cheng & Kirkpatrick, 2016; Dutoit *et al.*, 2018; Ruzicka *et al.*, 2019) have been called into question, and there have been suggestions of a more transient model of sexual antagonism, where conflict is resolved quickly and only low levels of ongoing sexual conflict are maintained (Mank *et al.*, 2020; Wright *et al.*, 2019). Supporting this model is the fact that reliably identifying genomic patterns of sexual conflict remains problematic, particularly when using population-based statistics such as Tajima's D and F_{ST}), as there are several other demographic and selective processes that could produce similar genomic patterns of divergence and balancing selection, such as heterozygote advantage and life history trade-offs within one or across both sexes, frequency-dependent selection, or temporal variation in fitness (Huang

et al., 2014; Mokkonen *et al.*, 2011; Fijarczyk & Babik, 2015). In addition, genotyping errors associated with sequence similarities between autosomal regions and sex chromosomes can also lead to false signals of sexual conflict (Kasimatis *et al.*, 2021; Bissegger *et al.*, 2020).

Nevertheless, using a quantitative genetic approach could help overcome some of the issues faced by population genetics methods. In particular, there are few empirical examples that directly examine this issue at the individual level, rather than the population level, and that are supported by direct associations with male and female fitness. The few examples of analyses that combine reproductive fitness or early life/prenatal viability data with tests for sexual antagonism are exclusively in humans or model organisms such as *Drosophila* (Healey *et al.*, 2000; Friedrichs *et al.*, 2006; Ruzicka *et al.*, 2019). Individual fitness information on reproductive success and survival at different life history stages can help disentangle true signals of sex-specific or sexually antagonistic selection from confounding factors such as demographic and other selective processes that could produce similar genomic patterns of divergence and balancing selection (Huang *et al.*, 2014; Mokkonen *et al.*, 2011; Fijarczyk & Babik, 2015). This could be achieved by investigating sex differences in allelic transmission between parents and offspring to examine signatures of gametic and viability selection before birth and at different life stages (Knief *et al.*, 2015; Ducret *et al.*, 2016; Finseth *et al.*, 2021), and by conducting sex-stratified GWAS and sex-by-gene interaction studies for fitness traits within populations (Bernabeu *et al.*, 2021). Long-term studies of natural populations with access to individual genome and fitness data could play a crucial role in elucidating the genetic architecture of sexual antagonism and the extent to which it is resolved or ongoing (Rowe *et al.*, 2018).

1.5 The red deer study population on the Isle of Rum

1.5.1 Study system

The Isle of Rum is one of the islands comprising the Small Isles archipelago, which is part of the Inner Hebrides, and is situated 30 miles (ca. 48 km) off the west coast of Scotland. Rum is the largest of the Small Isles, which also contain the Isles of Eigg, Muck and Canna; it measures 40.4 square miles and is largely volcanic in origin. Up until 1825, the island had about 425 inhabitants, but the population was removed during the Highland Clearances, and in 1845 Rum became a sporting estate. Today, the island's population consists of about 30 people, including field researchers of the Rum red deer research project. In 1958, the Isle of Rum was bought by the Nature Conservancy Council (later Scottish National Heritage and currently NatureScot) with the intention to use the island as an outdoor research laboratory for long-term ecological studies, particularly of red deer. There have been three major phases of research on red deer on Rum: ecological research (1958-1972), research on reproductive physiology (1968 – 1973), and research on behaviour, life histories, population dynamics, and evolutionary genetics (1972—present). Since 1972, the project has focused on intense individual based monitoring of the population in the 'North Block' of Rum (see Figure 1.1). The annual cull was terminated in the same year to allow exploration of population dynamics and individual breeding success in a naturally regulated population, as well as enable the habituation of animals to close observation. Since 1984, tissue samples for genetic analysis have been collected from calves born in the North Block, which are used to assign paternity and conduct population and quantitative genetic analyses.

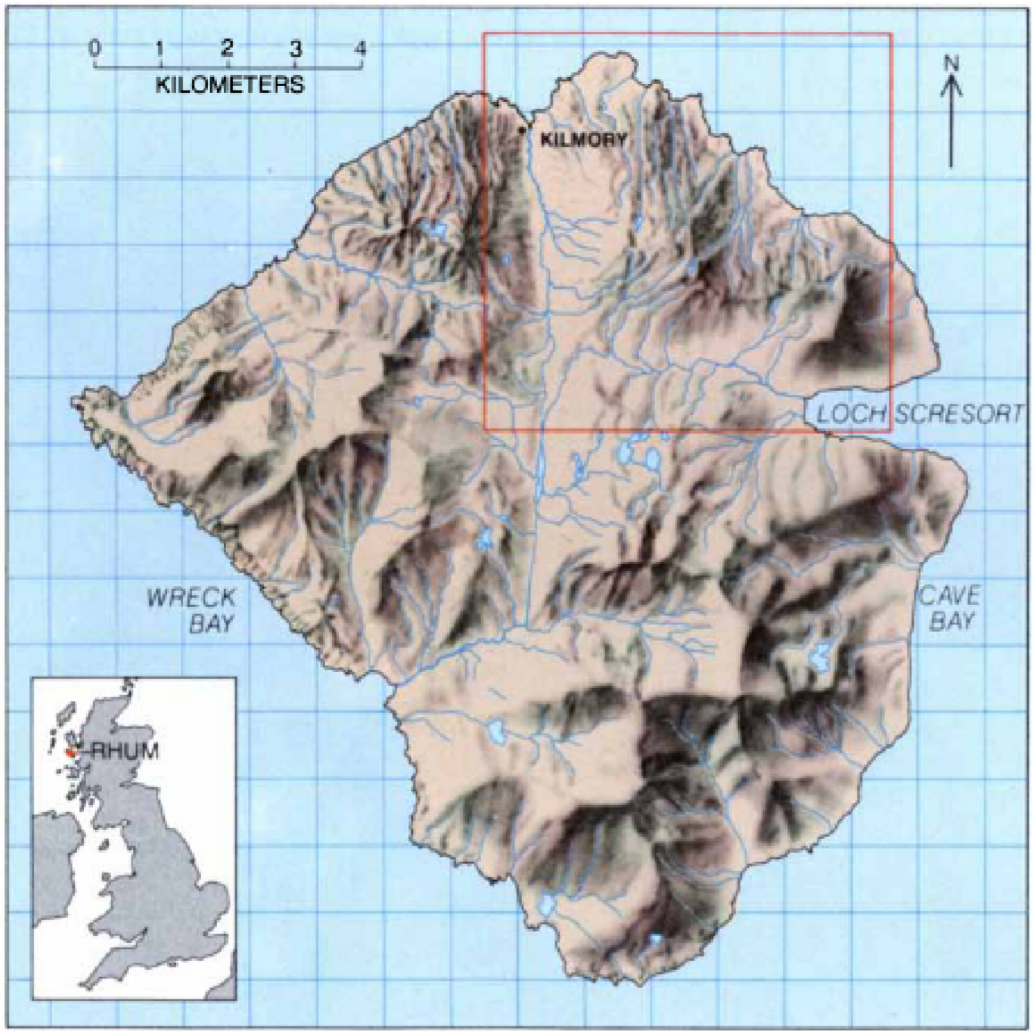


Figure 1.1: Map of the Isle of Rum, the small grey map in the left corner indicates the island's position relative to the UK mainland. The red box marks the part of Rum that is part of the red deer project study area ('North Block'). The map was adapted from Clutton-Brock (1985)

1.5.2 Data collection

The current project comprises data obtained from deer observed in the North Block of Rum since 1972. A key requirement for the research conducted on the deer is the collection of individual based data, and so natural markings and unique tags are used to distinguish individual deer. This way, deer can be

observed throughout their life, from birth to death, and their behaviour, movement, and reproduction can be recorded in detail. There are four major types of data collection that take place throughout the year.

First, there are year-round study area censuses conducted by a field researcher that take place about five times a month and during which identities, location, and behaviour of individual deer in all groups seen on a set route within the study area are recorded.

Second, there is a major sampling effort during calving season, when females (hinds) give birth. Calving season usually starts in late May and lasts through June, with some calves still being born in July and sometimes even August. A team of volunteers coordinated by field researchers monitors the study area daily and track females when they give birth. Calves are captured within the first few days of birth and marked using ear tags, ear punches, and in the case of female calves a collar that expands as the animal grows; collars cannot be used for males as their necks grow too large to accommodate the collar. In addition, a blood and tissue sample is taken (for maternal antibody and genetic analysis respectively); the calf is measured for birth weight and hind leg length; and other condition markers and behaviour during capture as well as birth location are recorded. Usually, around 80% of calves born in a given year in the study area are captured.

Third, rut censuses are conducted during the red deer mating season, which lasts from about mid-September to mid-November, when the males (stags) compete for access to females and gather groups of hinds in harems that they defend against rivals. Daily censuses are carried out to record the identities of hinds in each stag's harem, as well as the outcome of fights and challenges between males and any matings that take place. This behavioural information can be used to inform paternity assignment and pedigree construction.

Fourth, mortality searches are carried out during the winter months to locate

the bodies of missing study animals. Post-mortem examinations are performed in the field and jaw and leg bones are taken for measurements. Additional tissue samples are also taken for genetic analyses to supplement the data set with animals that were not caught as calves or to help with identification in cases where this is not clear. During this sampling effort, cast antlers are also collected when found and various measurements are taken; DNA can be extracted from cast antlers to again expand the genetic data set to include stags born outside the study area, which helps increase the success of paternity assignment.

1.5.3 Red deer antlers



Stag (Zaphod) with fully grown antlers during the rut and hind.

Antlers are an important secondary sexual trait in red deer and are only present in males. They function as male weaponry and are used in inter-male competition for access to mates. Antler size has been shown to be associated with male fertility in red deer (Malo *et al.*, 2005) and is important for reproductive success in the red deer of Rum (Kruuk *et al.*, 2002, 2014). Red deer antlers are cast and regrown every year and are one of the fastest

growing tissues of any animal (Goss, 1983). Antlers are made of solid bone but are covered with a soft layer of tissue called 'velvet' during the growth period, which supplies the antler with blood and essential nutrients. The velvet is discarded or 'cleaned' after the antler has stopped growing and only the bone remains. Antlers are frequently damaged during fights that take place during the rut season, some stags even losing one antler altogether, but are completely regrown the following year. The growth cycle of antlers is closely linked to fluctuating sex hormone levels such as testosterone that change seasonally over the course of a year (Bubenik *et al.*, 1987; Suttie *et al.*, 1984) so that all stags go through the same stages of the antler cycle at about the same time. Stags on Rum usually cast their antlers from the beginning of March to mid-April and have completed cleaning by mid-September in time for the beginning of the rut (see Figure 1.2). Antlers grow larger with age until they reach their full adult size, but due to their unique shape, they can be attributed to an individual when found after casting, when males have reached the age of three.



Figure 1.2: Timing of important life history processes throughout the year for male (right cycle) and female (left cycle) red deer on Rum.

1.5.4 Life history and sex differences

Red deer have a polygynous mating system with intense inter-male competition so that few males mate with many females during a single mating season. Thus, male mating success is highly skewed with few very successful and many unsuccessful individuals, whereas most females will produce one calf every year at most after reaching sexual maturity. Therefore, the strategy for investment in reproductive fitness is different between stags and hinds. Stags need to invest in fighting ability to secure as many matings as possible each rut, while hinds need to balance survival and fecundity costs of gestation and in particular lactation with the benefit of increasing their lifetime reproductive success by producing a calf each year (Clutton-Brock, 1985; Clutton-Brock *et al.*, 1989; Froy *et al.*, 2016). Maternal investment is higher in sons than in daughters in the red deer, but does not appear to result in fewer males being born (Clutton-Brock *et al.*, 1981). Consequently, producing male calves reduces the mother's subsequent survival and reproductive success, but this costs seems to depend on the social rank of a female, with dominant hinds suffering no apparent costs for rearing sons (Gomendio *et al.*, 1990; Froy *et al.*, 2016).

Differences between the sexes can be observed early on in life, as male calves are heavier at birth and have higher growth rates compared to female calves (Clutton-Brock, 1985; Kruuk & Hadfield, 2007). However, males are also more susceptible to limited resource conditions and have lower levels of fat storage, which contribute to higher mortality rates compared to females within the first two years of life (Clutton-Brock, 1985). On Rum, males are usually at least five years old when they first successfully hold a harem and sire offspring, which is considerably older than the age at which females start to reproduce, usually at around three years of age. Life expectancy is also lower in males than in females, with stags on average reaching just under nine years, while females live to about just over ten years. Overall, there are clear differences in male

and female life history strategies and thus resource distribution (see Figure 1.2 for a breakdown of the physiological stages males and females cycle through within a year).

1.5.5 Quantitative genetics tools

The Rum red deer research is well suited for quantitative genetic analyses. First, the project has access to multi-generational pedigree information. The pedigree was first built using microsatellite markers to assign paternity, as well as observational data collected during the rut to better inform assignments (Walling *et al.*, 2010). In 2015, the Cervine Illumina BeadChip, a 50K SNP chip array was published using red deer and smaller number of sika deer (*Cervus nippon*) derived markers (Brauning *et al.*, 2015). The current pedigree is constructed based on 440 core SNPs, and paternity is assigned using the R package *Sequoia*. This package was originally developed using Rum red deer data (Huisman, 2017) and is especially designed to construct multi-generational pedigrees for wild populations. Currently, around 3,000 individuals in the Rum deer population are genotyped at over 39,000 polymorphic SNP loci and life-history, morphological and pedigree information is available for over 5,000 deer. These data have enabled several quantitative genetic analyses including quantifying additive genetic variance (heritability), genetic correlations, and selection pressures on fitness related traits such as antler size (Kruuk *et al.*, 2002, 2014), birth weight (Kruuk & Hadfield, 2007), endocranial volume (Logan *et al.*, 2016), inbreeding depression (Slate *et al.*, 2000; Huisman *et al.*, 2016), testosterone levels (Pavitt *et al.*, 2014) and parturition date (Froy *et al.*, 2019; Bonnet *et al.*, 2019b). In addition, the SNP data has successfully been used to build a linkage map for red deer (Johnston *et al.*, 2017), which in turn has provided the means to characterise the genetic architecture of recombination rate in red deer (Johnston *et al.*, 2018) and allowed us to understand how the SNPs are ordered relative to each other, as

a reliable and annotated genome assemble for *Cervus elaphus* is not yet available at the time of writing.

1.6 Thesis aims and outline

The main aim of this PhD was to investigate the genomic basis of sexually divergent traits in red deer by (a) characterising the genetic architecture of antler morphology using genomic approaches and (b) identifying genomic signals of ongoing sexual conflict using classic quantitative genetic and genomic approaches. The aims of the project are addressed in three data chapters, which can be summarised as follows:

- **Chapter 2: Pedigree and genomic approaches to estimate quantitative genetic parameters of antler traits in a wild population of red deer (*Cervus elaphus*)**. In this chapter, I characterise the genetic (co)-variances of several antler morphology traits using traditional pedigree and genomic approaches.
- **Chapter 3: Genomic approaches to understanding the genetic architecture of antlers in red deer (*Cervus elaphus*)**. In this chapter, I identify causative loci underlying the genetic architecture of antler morphology and characterise their effect size distribution through a combination of single marker and haplotype association mapping.
- **Chapter 4: Genomic signatures of sexual conflict in red deer (*Cervus elaphus*)**. In this chapter, I explore genomic signals of ongoing sexual conflict by investigating patterns of sex-specific transmission distortion and modelling SNP effects on sex-specific survival at later life stages.

Finally, in Chapter 5, I present a general discussion of the findings of this thesis and future directions in the investigation of antler morphology and sexual conflict in the Red deer, and more broadly across the field of wild evolutionary genomics.

Chapter 2

Pedigree and genomic approaches to estimate quantitative genetic parameters of antler traits in a wild population of red deer (*Cervus elaphus)*.



Stag (Hellboy05) resting.

2.1 Summary

Sexually selected traits show large variation and rapid evolution across the animal kingdom, yet genetic variation often persists within populations despite apparent directional selection. A key step in solving this long-standing

paradox is to determine the genetic architecture of sexually selected traits to understand evolutionary drivers and constraints at the genomic level. Antlers are a form of sexual weaponry in male red deer. On the island of Rum, Scotland, males with larger antlers have increased breeding success, yet there has been no evidence of any response to selection at the genetic level. To try to understand the mechanisms underlying this observation, I started by investigating the genetic variances of ten antler traits from >1,000 antlers and their principal components (PCs). Traditionally, researchers have used pedigree data from long-term ecological studies to estimate trait heritability in wild populations. In this chapter, I estimate heritabilities of the antler traits and their PCs, as well as the genetic correlations among antler traits, using and comparing relatedness matrices derived from the red deer pedigree and from a genomic relatedness matrix constructed using data from >38,000 single nucleotide polymorphisms (SNPs). I show that antler morphology is highly heritable (h^2 of antler traits ranging from 0.211 to 0.436) and repeatable over an individual's lifetime, and that genomic approaches can be a reliable alternative to pedigrees in quantitative studies of in this population. Antler traits are also highly positively correlated on the phenotypic and genetic level, indicating some degree of shared genetic architecture. These findings form the basis for more detailed work on the genetic architecture of antler morphology, and indicate that the complex genetic structure of antler morphology might contribute to maintaining the genetic variance of antler traits.

2.2 Introduction

Sexually selected traits show great variety and complexity across the animal kingdom, ranging from morphology and behaviours that increase attractiveness to the opposite sex, to traits which increase inter-sexual

competitiveness for access to mates (Andersson, 1994). Such traits are frequently under directional strong selection (Kingsolver *et al.*, 2001), with phenotypic differences between related species suggesting that they can evolve rapidly, with downstream consequences for other phenomena such as adaptation, speciation, and extinction probability (Lorch *et al.*, 2003; Ritchie, 2007; Servedio & Bürger, 2014; Wilkinson *et al.*, 2015; Martínez-Ruiz & Knell, 2017). But despite the evolutionary importance of sexually selected traits and despite the rapid acceleration in our knowledge of the genomic basis of variation in phenotypes, we know surprisingly little about their detailed genetic architecture. Understanding in the first instance the heritability and genetic correlations among traits can provide valuable insights into evolutionary mechanisms.

Theory predicts that the strong selection experienced by secondary sexual traits reduce genetic variation within populations, yet empirical studies often show that sexually selected traits have substantial underlying genetic variation despite evidence of selection (Pomiankowski & Moller, 1995; Kotiaho *et al.*, 2001; Kruuk *et al.*, 2002; Svensson & Gosden, 2007). This contradiction presents an evolutionary paradox, for which several explanations have been proposed. Sexually selected traits are often condition dependant, (Dugand *et al.*, 2019) and using quantitative genetic approaches to estimate genetic variance (e.g., heritability) and other variance components can inform about the relative importance of environment over genetic parameters (Emlen, 1994; Griffith *et al.*, 1999; Kruuk *et al.*, 2014). Additionally, estimation of genetic correlations with indicators of body condition and correlations among suits of sexually selected traits can shed further light on how variation is maintained and expose genetic constraints on evolution. For example, a study in fruit flies *Drosophila bunnanda* and *D. serrata* analysing multiple genetically correlated cuticular hydrocarbons under sexual selection found that while substantial genetic variance was present, most of it was not aligned with the direction of selection and thus unavailable for evolutionary change

(Van Homrigh *et al.*, 2007). Another study in black field crickets found that while single male call traits were associated with large genetic variances, genetic variance for specific combinations of genetically correlated call traits was actually depleted by directional or stabilising selection(Hunt *et al.*, 2007). Some other the potential mechanisms include genetic correlations between traits under opposing selection pressures and linkage disequilibrium between alleles at causal loci and deleterious alleles (Lande, 1982; Lande & Arnold, 1983; Connallon & Hall, 2018). Several other explanations for the persistence of genetic variance have been proposed, such as differences in selection between the sexes, developmental stages or environmental conditions (Bourret *et al.*, 2017; Barson *et al.*, 2015), phenotypic plasticity (Charmantier & Gienapp, 2014) and trade-offs with survival (Johnston *et al.*, 2013a). It is however crucial to also understand the genetic relationship of a trait with fitness, as genetic covariance between the trait and fitness is needed for a response to selection, even if the trait is associated with mating success on a phenotypic level (Morrissey *et al.*, 2010).

In ruminants, most sexually selected weaponry appear to conform to a classical quantitative trait, but there is relatively little information on the inheritance of weaponry or the underlying genetic architectures. Earlier studies in antlers in red deer *(Cervus elaphus*) and horns in Soay, bighorn and thinhorn sheep (*Ovis aries*, *O. canadensis* and *O. dalli*, respectively) found that weaponry can have substantial additive genetic variance (i.e., heritabilities, ranging from $h^2 = 0.23$ - 0.43). This included horn circumference and weight in bighorn and thinhorn sheep (which appears to be unconstrained and available to selection in some cases; Coltman *et al.* 2005; Poissant *et al.* 2012; Sim & Coltman 2019), horn length and circumference in Soay sheep (Johnston *et al.*, 2009), and antler weight in red deer (Kruuk *et al.*, 2002). An early genome-wide association study in Soay sheep led to the discovery of the *RXFP2* gene, involved in primary sexual trait determination in mammals, as a major effect locus underlying horn type and size (Johnston *et al.*, 2011).

This remains one of the few examples of known genetic architecture of quantitative variation in male weaponry in a wild living ruminant. However, studies attempting to map specific trait loci underlying heritable variation in the other systems above have failed to identify specific genetic architectures underlying male weaponry. These studies often have small samples sizes in terms of individuals and genetic markers, making it difficult to discern whether a lack of genetic association is due to many loci of small effect (i.e., a polygenic architecture) or low power to detect loci of medium to large effect. Nevertheless, the recent development of tractable genomic resources has facilitated the study of the genetic architecture of quantitative traits in other natural systems (Santure & Garant, 2018) and provides a renewed opportunity to understand the genetic architecture of sexually selected traits both within and outwith these systems.

Antlers are a form of sexual weaponry in deer (Cervidae) that are generally only present in males and are shed and regrown annually (Davis *et al.*, 2011). Like horns, they are used as weaponry in inter-male competition for access to females, with larger antlers often associated with increased reproductive success (Kruuk *et al.*, 2002; Malo *et al.*, 2005). Most information about the inheritance of antler traits has traditionally been generated by studies on captive bred populations, which found moderate to high levels of heritability for various antler traits in species like white-tailed deer (*Odocoileus virginianus*; beam circumference h^2 = 0.29, beam length h^2 = 0.14, antler mass h^2 = 0.43; Lukefahr & Jacobson 1998; Williams *et al.* 1994), wapiti (*Cervus canadensis*; antler velvet mass h^2 = 0.27; Wang *et al.* 1999) and red deer (antler velvet mass h^2 = 0.36; Van Den Berg & Garrick 1997). As populations experience different selection pressures and demographic processes in the wild compared to captivity, genetic variance and thus heritabilities might be expected to differ. Evidence from a pedigreed wild population of red deer suggests similarly moderate levels of heritabilities (see below, Kruuk *et al.* 2002, 2014); however, high uncertainty in heritability

estimates for antler traits in other wild populations (such as a study in wild white-tailed deer using genomic marker information on largely unrelated individuals; Jamieson *et al.* 2020) illustrates the need for more suitable data sets for quantitative genetic analyses of wild populations.

In male red deer (*Cervus elaphus*) on the island of Rum, Scotland, there is directional selection for increased antler weight and number of antler points (known as "form"). Both traits are substantially heritable (h^2 = 0.38 & 0.24, respectively) and positively genetically correlated, but no phenotypic response to selection has been observed over a 30-year study period (Kruuk *et al.*, 2002, 2014). Using annual measures of breeding success, quantitative genetic analysis indicate that the selection gradients on the genetic components were not different from zero (for weight) and negative (for form), while the environmental components of selection were both positive (Kruuk *et al.*, 2002, 2014). This suggests that the apparent phenotypic selection for increased antler weight is mainly driven by environmental associations, with males who had experienced favourable environmental conditions having both larger antlers and individual breeding success. In contrast, antler morphology would be constrained by the zero or negative genetic selection gradients (indicative of antagonistic selection at the genetic level), or genetic associations with one or more unknown traits. These results indicate that genetic constraints may contribute to the maintenance of genetic variance. To better understand the mechanisms maintaining genetic variance, a logical next step is to determine the genetic architecture of antler morphology. Further, as antlers present a multi-dimensional phenotype, adding more information from different measures may contribute to our understanding of potential evolutionary conflict and constraints within the antler and may allow us to characterise the specific genetic variants underpinning heritable variation (Chenoweth & McGuigan, 2010).

In this study, I used an extensive antler morphology data set with 948 to 3,972

observations in 336 to 891 unique males and genomic data from 39,000 polymorphic SNPs to estimate the heritability and genetic correlations of ten antler traits. I also use principal components (PCs) to condense correlated aspects of antler measures and enhance differentiation among unrelated characteristics. The use of PCs and related methods for reducing the dimensionality of a data set are common in the analysis of multidimensional morphological structures, including in analysis of their genetic architecture (Buczek *et al.*, 2016; Pallares *et al.*, 2014). I then estimate the heritability of antler traits and PCs by fitting animal models, mixed effects models first developed for animal breeding studies, which are now widely used in quantitative genetics analysis of wild populations to calculate genetic correlations and trait heritability (Kruuk *et al.*, 2002; Johnston *et al.*, 2009, 2018; Berenos *et al.*, 2015; Miller *et al.*, 2018; Husby *et al.*, 2015). The animal model estimates the additive genetic variance component of a trait by fitting a relatedness matrix (Henderson, 1975), which is either derived from a pedigree and based on identity by descent (IBD), or from genome-wide markers which estimate relatedness based on identity by state (IBS; VanRaden 2008). In this chapter, I compare both pedigree based and GRM based heritability estimates of antler traits and PCs to explore the informativeness of both approaches and of the GRM in particular, as an alternative for quantitative genetic studies in wild systems where long-term pedigrees are unavailable (also see Bérénos *et al.* 2014). I find that all antler traits are heritable and repeatable over an individual's lifetime, and show similar results for the PCs, which represents distinct aspects of antler morphology. My results also indicate high genetic correlation across all antler traits, which likely contributes to evolutionary constraints within this population.

2.3 Methods

2.3.1 Study system

The red deer study population is situated in the North Block of the Isle of Rum, Scotland (57°02'N, 6°20'W) and has been subject to individual monitoring since 1971 (Clutton-Brock *et al.*, 1982). Deer calves are marked with ear tags shortly after birth to enable recording of detailed life histories of individuals. DNA is routinely extracted from neonatal ear punches, post-mortem tissue and/or cast antlers (see Huisman *et al.* 2016). A pedigree of 4,515 individuals is available for the population, constructed using single nucleotide polymorphism (SNP) data in the R package Sequoia (Huisman, 2017). Research was conducted following approval of the University of Edinburgh's Animal Welfare and Ethical Review Body and under appropriate UK Home Office licences.

2.3.2 Antler measures

Male red deer cast and regrow their antlers every year from the age of one or two (Kruuk *et al.*, 2002). Ten antler measures are routinely taken from cast antlers and antlers from deceased individuals between 1971 and 2017 (see Figure 2.1 and Table 2.1 for full details and sample sizes of each measure). All measures of length were taken following the curve of the antler, and measures of circumference were taken at the narrowest point between antler tines (points). Total antler length was defined as the distance from the coronet (base) to the furthest point of the antler. All length and circumference measures were taken in centimetres. Antler weight was measured as the total dry mass of the antler in grams. Antler form was defined as the number of tines, and as this trait can be determined from observations in the field it has

the greatest sample size (3,972 observations from 891 stags). The form data contained non-integers for tines that weren't fully formed and those were rounded to the closest integer or 0.5 measure (e.g 2.25 to 2.5 and 2.75 to 3). Where measurements from both the left and right antlers were available, the mean was taken. Individual measurements were excluded if the antler part was broken, and antler weight was discarded if any part of the antler was broken. Only antlers from stags aged 3 years or older were considered, as cast antlers recovered in the field can be reliably assigned to known individuals by their shape from this age onwards (Kruuk *et al.*, 2002, 2014), as has been confirmed by genetic analysis (Huisman *et al.*, 2016).

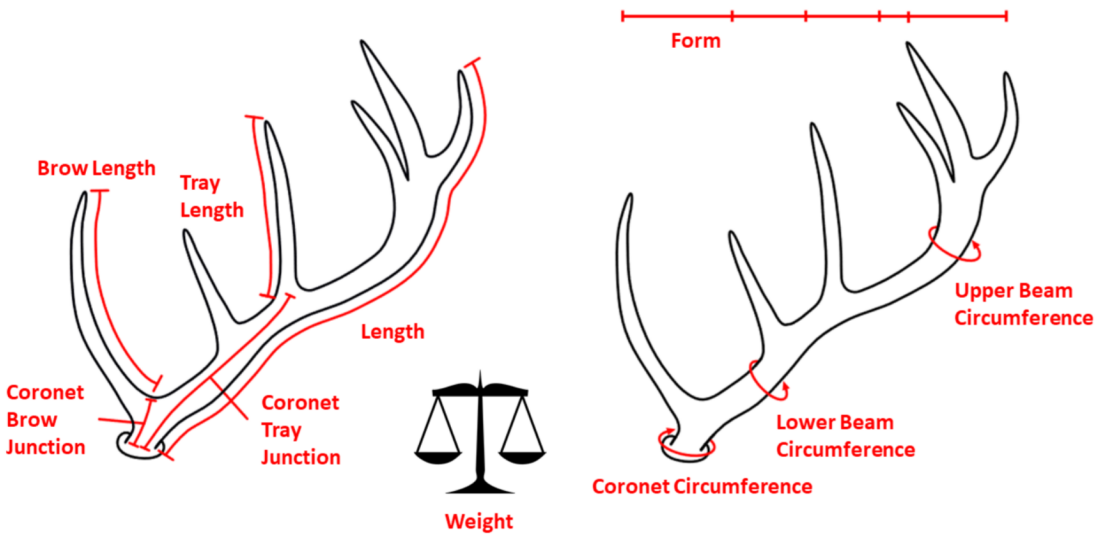


Figure 2.1: A schematic of the antler measures used in this study (measured in *cm*). All traits were measured in *cm*, except for antler weight (measured in *g*)and form, which is the total number of points on an antler. Further details for each measurement are given in the main text and in Table 2.1.

Table 2.1: Summary statistics of all 10 antler measures as shown in Figure 2.1 in SNP genotyped males. *N Obs* is the number of observations, *N IDs* is the number of unique individuals for each measure. All lengths and circumferences were measured in *cm*, weight was measured in *g*. *SD* denotes the standard deviation on the raw date (not corrected for age). *min* and *max* refer to the minimum and maximum value, respectively

Trait	N Obs	N IDs	Mean	SD	min	max
Antler Length	1060	444	65.168	12.029	15.50	93.80
Coronet Circumference	1224	471	15.626	2.013	4.60	21.70
Lower Beam Circumference	1204	460	10.504	1.533	4.30	18.00
Upper Beam Circumference	1123	436	9.446	1.358	5.85	15.00
Coronet-Brow Junction	1162	466	5.720	1.025	3.00	9.80
Coronet-Tray Junction	1184	455	27.313	5.398	7.30	55.90
Brow Length	1139	457	20.568	5.461	1.00	35.20
Tray Length	948	388	14.484	4.177	1.00	27.40
Antler Weight	1003	336	649.858	243.098	55.00	1505.50
Form (Total Number of Points)	3972	891	4.498	1.072	1.00	8.00

2.3.3 Genomic data-set

DNA samples from 2,870 individuals have been genotyped at 51,248 SNP markers (Huisman *et al.*, 2016) on the Cervine Illumina BeadChip (Brauning *et al.*, 2015) using an Illumina genotyping platform and Illumina GenomeStudio software (Illumina Inc., San Diego, CA, USA). All SNPs on the Cervine Illumina BeadChip are named based on their synteny with the cattle genome BTA vUMD 3.0 (e.g., SNP ID *cela1_red_15_1479373* is orthologous to position 1479373 on cattle chromosome 15). Quality control was carried out in PLINK v1.9 (Chang *et al.*, 2015) with the following thresholds: SNP genotyping success >0.99, minor allele frequency >0.01 and individual genotyping success >0.99. The final SNP data-set consisted of 2,936

individuals and 39,493 markers. In addition, a linkage map specific to the Rum population is available, with 38,083 SNPs assigned to linkage groups corresponding to the 33 deer autosomes and X chromosome (Johnston *et al.*, 2017). Genome-wide linkage disequilibrium (LD) was calculated between all SNPs within 1Mb of each other using Spearman's Rank correlation (ρ). Further quality control of mapped SNPs and X-linked markers (i.e., heterozygous state in males) was conducted using the *check.marker* function with default thresholds in the R library GenABEL v1.80 in R v3.4.2 (Aulchenko *et al.*, 2007). A total of 2,138 individuals and 38,006 markers were used for pairwise LD estimation using the squared coefficient of correlation r^2. Based on a linear regression of r^2 on the log base pair distance between SNPs, LD decayed at a relatively low rate of 0.031 ρ per Mb (SE = 5.56×10^{-4}, Figure S3.1).

2.3.4 Principal components of antler measures

A principal component analysis (PCA) was conducted to create a second dataset combining information from the different antler measures, while also increasing the differentiation among the different principal components (PCs). As PCA does not allow for missing data, I imputed missing antler measures using the Bayesian *bpca* algorithm in the packages pcaMethods v1.7 in R v3.4.2 (Stacklies & Redestig, 2018; Oba *et al.*, 2003). I used the default settings which assumes a flat prior distribution for imputation, and the most appropriate number of PCs was determined using the *kEstimate* function. To improve imputation accuracy, 'form' was divided into lower and upper values for each antler, to represent the number of tines (branches) on the lower part of the antler ('lower beam', 1-3 tines) and the upper part ('upper beam', 1-5 tines) respectively, resulting in 11 antler measures. Imputation accuracy was quantified by calculating the error of prediction (E) from a complete subset of the data with no missing values and the same subset with randomly missing

data at a similar level to the whole data set ($\sim$ 9%). The error of prediction was calculated as follows:

$$E = \sum(V - I)^2 / \sum(V)^2 \tag{2.1}$$

where I refers to the imputed values of the data subset with missing values and V to their counterpart in the complete data subset (Stacklies & Redestig, 2018). My analysis found that the imputation of missing values using a Bayesian PCA approach achieved high accuracy and hence low error (E = 0.015). To account for variation among antlers due to age, antler measure values were modelled using a linear model approach (following Pallares *et al.* 2014). All models had the same structure:

$$y = Age + Age^2 + e \tag{2.2}$$

where y is a vector of the antler measure and e is a residual error term. Age effects were fitted as a fixed quadratic term to account for the non-linear change in antler measures with age (see also Nussey *et al.* 2009 and Kruuk *et al.* 2002). All antler measures were modelled using a Gaussian distribution. All models were fitted using the *lm* function in R v3.4.2. The residuals of these models were then used in a standard PCA using the *prcomp* function in R and the scores of the PCs used as trait values in the downstream analysis.

2.3.5 Estimating heritability using the animal model

All antler measures and PCs were modelled using a restricted maximum-likelihood (REML) approach within the mixed 'animal model' framework (Henderson, 1975) in ASReml-R v4.0 (Butler *et al.*, 2009) in R v3.4.2. The animal model estimates the effect sizes of fixed effects and partitions phenotypic variance (V_P) into several random effects, including the

variance attributed to additive genetic effects (V_A). Previous studies have estimated V_A using a pedigree-derived relatedness matrix (Kruuk *et al.*, 2002, 2014); here, I wished to compare estimates of V_A using both pedigree and genomic relatedness information. Therefore, all models were carried out estimating V_A in one of two ways: 1. using a numerator relationship matrix A based on the pedigree, using the *ainv* function in ASReml-R; and 2. using a genomic relatedness matrix (GRM) (VanRaden, 2008) calculated using autosomal SNPs (N = 37,271) in the *--make-grm function* in GCTA v1.24.3 (Yang *et al.*, 2011). The GRM was adjusted to assume similar frequency spectra of genotyped and causal loci with the argument *--grm-adj 0*.

All 10 antler measures and 11 PCs were modelled in univariate animal models with the following structure:

$$y = X\beta + Z_1 a + Z_r u_r + e \tag{2.3}$$

where y is a vector of the antler measure or PC, X is an incidence matrix relating individual measures to the vector of fixed effects β; Z_1 and Z_r are incidence matrices relating individual measures to additive genetic and other random effects respectively; a is a vector of pedigree relatedness matrix A or GRM additive genetic effects; u_r is a vector of additional random effects; and e is a vector of residual effects. Fixed effects included age in years as a both a linear and quadratic term for the antler measures and an intercept only for the PC models (as age structure was accounted for prior to PC estimation). Hindleg length at birth was fitted as a proxy for adult body size in initial models, but was later excluded due to concerns about potential genetic correlations with antler traits. Random effects included: the additive genetic effect; permanent environment (i.e., individual identity) to account for pseudoreplication due to repeated measures in the same individual; and birth year and year of antler growth to account for common environmental effects between individuals. The narrow sense heritability (h^2) was calculated as V_A/V_P, where V_P was defined as the sum of the variance attributed to all

random effects, including the residual variance (Falconer & Mackay, 1996). The significance of fixed effects was calculated with a Wald test, while the significance of the random effects was tested using a likelihood ratio test (LRT) between models with and without random effect of interest (i.e., 2 $\times$ the difference between the model log-likelihoods, assuming a χ^2 distribution with 1 degree of freedom).

Bivariate models were run to determine genetic correlations between the 10 antler measures, with the following structure:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_{a1} \\ a_{a2} \end{bmatrix} + \begin{bmatrix} Z_{r1} & 0 \\ 0 & Z_{r2} \end{bmatrix} \begin{bmatrix} u_{r1} \\ u_{r2} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \quad (2.4)$$

All variables are as defined in equation (2.3), with subscripts referring to antler traits 1 and 2, respectively. The pedigree-derived relatedness matrix and the GRM gave very similar estimates for the additive genetic variance in the univariate animal models for estimation of trait heritability (see results section). I decided to use the GRM to model the additive genetic covariance to capture any small deviation of genome-sharing from the expected proportions used by the pedigree based matrix. The genetic correlation r^2 was obtained from the genetic covariance:

$$r_a = cov_a(1,2)/\sigma_1\sigma_2 \quad (2.5)$$

where $cov_a(1,2)$ is the covariance between trait 1 and 2, and σ represents the respective standard deviations for traits 1 and 2. The significance of r_a was determined using an LRT as above, by comparing the model to another where r_a was constrained to either zero or one.

2.3.6 Data Statement

Data for this study, which includes genomic data after quality control measures, pedigree, and linkage map information and autosomal GRMs, is available in a Dryad repository (doi:10.5061/dryad.612jm643c)([dataset] Peters *et al.*, 2021). All scripts for the analysis are provided at `https://github.com/Lucy-Peters/Red\_deer\_antler\_genetic\_architecture`.

2.4 Results

2.4.1 Comparison of pedigree and genomic relatedness

Pairwise relatedness in the pedigree was centred around a mean of 2.4 x 10^{-2} (SD = 4.0 x 10^{-2}); relatedness as estimated by the GRM had a mean of -3.4 x 10^{-4} (SD = 4.3 x 10^{-2}). Pedigree and genomic pairwise relatedness estimates were correlated (linear regression, intercept = -2.1 x 10^{-2}, slope = 0.87; Figure S2.1). There was still some substantial unexplained variation, given that the regression represents a correlation of different estimates of the same variable (R^2 = 0.66), indicating that pedigree and genomic approaches differed in their ability to capture relatedness to some extent. This incongruity however did not seem to much affect estimates of heritability (see Section “Animal models of antler traits” below).

2.4.2 Principal component analysis of antler measures

A principal component analysis (PCA) of 11 antler measures resulted in the maximum number of principal components (i.e., 11 PCs). The composition of

the PCs showed that PC1, which explained around 41% of the variance, combined approximately equal amounts of information from all 11 measures, while PCs 2 to 11 explained increasingly less variance, mostly representing one or two antler measures (Figure 2.2). Almost all antler measures were also significantly positively phenotypically correlated (using Pearson's product moment correlation coefficient), except for the correlation of tray length with coronet-tray junction. The phenotypic correlation coefficients ranged from r = 0.07 for the correlation of coronet-tray junction with form to r = 0.85 for the correlation of lower beam with upper beam circumference (see Table S2.1).

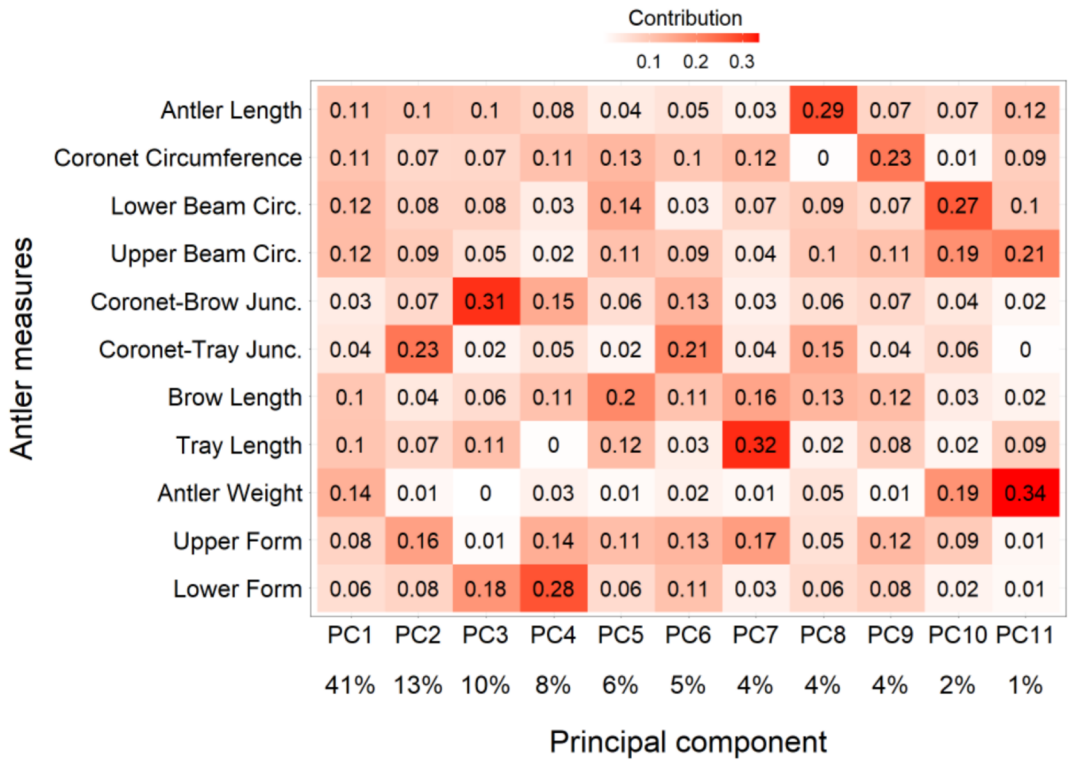


Figure 2.2: Heat-map showing the contribution (as a proportion) of each antler measurement to each of the 11 principal components.

2.4.3 Animal models of antler traits

Antler measures were significantly heritable, with estimates ranging from h^2 = 0.211 to 0.436 for the pedigree estimates, and h^2 = 0.229 to 0.414 for the genomic estimates (Figure 2.3; Tables 2.2 & S2.2). Heritability estimates for antler weight, antler length, coronet circumference, brow length and form were generally consistent with previous findings by Kruuk *et al.* (2014, Figure 2.3). All antler PCs were also significantly heritable, although estimates decreased substantially for higher order PCs, which explained relatively small proportions of variance in antler morphology (Figures 2.4 & 2.2, Tables 2.2 & S2.2). For all antler measures and PCs, confidence intervals between pedigree and genomic relatedness estimates were very similar; there was no trend when comparing estimates of h^2 for the same trait, suggesting that both the pedigree and genomic relatedness matrices capture the additive genetic variance to a similar degree in this population. Due to the pedigree and GRM data sets not being completely identical, it was not possible to test differences between pedigree and GRM heritability estimates formally here. Therefore, all results described from this point onwards are from models fitting a GRM, unless otherwise stated.

The permanent environmental effect (which includes, among others, dominance and epistatic effects) was generally significant for all antler measures, explaining up to 28.0% (upper beam circumference) of the phenotypic variance (Table 2.2). Year of antler growth explained a significant proportion of phenotypic variance for most antler measures and PCs (Table 2.2). Conversely, birth year was not significant for any antler measure or PC; nevertheless, I retained this random effect in all models to account for potential cohort effects (Table 2.2). Trait repeatabilities, calculated as the sum of contributions from the additive genetic, the permanent environment and birth year components, were high for all antler measures, ranging from 38.5% (antler length) to 66.1% (tray length; Table 2.2). Age as both linear and

quadratic fixed effect terms was significantly associated with all antler measures (positive and negative, respectively, in all cases; Table S2.3).

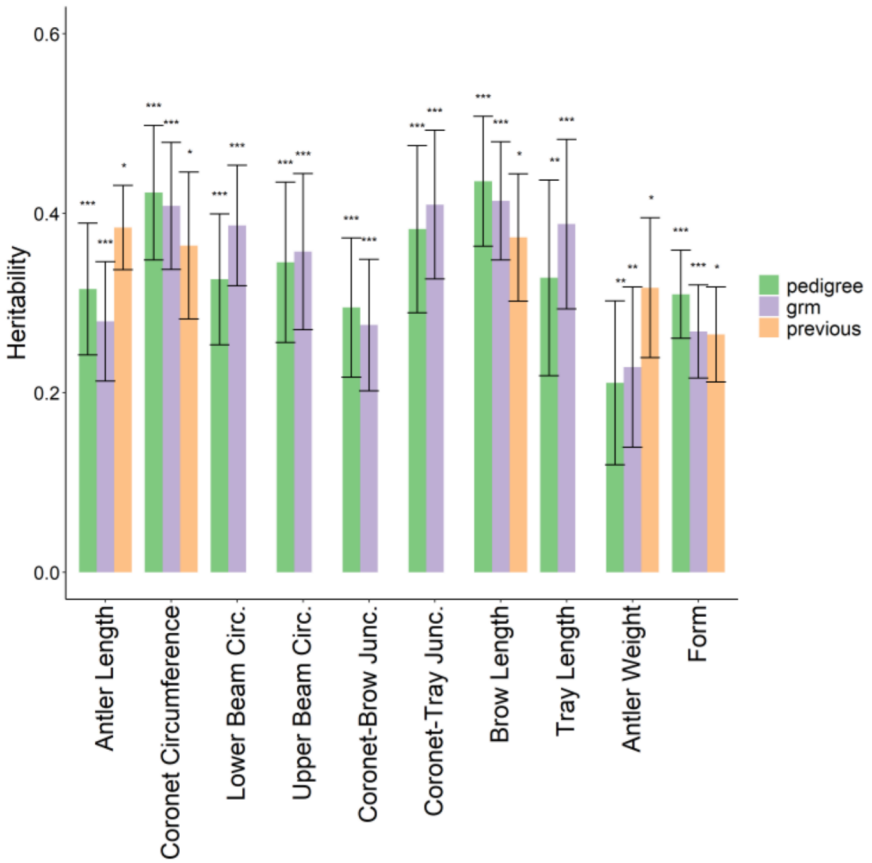


Figure 2.3: Heritability estimates for all 10 antler measures. Both estimates from the pedigree and the GRM models are shown, as well as results from a previous study by Kruuk *et al.* (2014) for antler weight and form. $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$. Underlying data is provided in Table S2.2

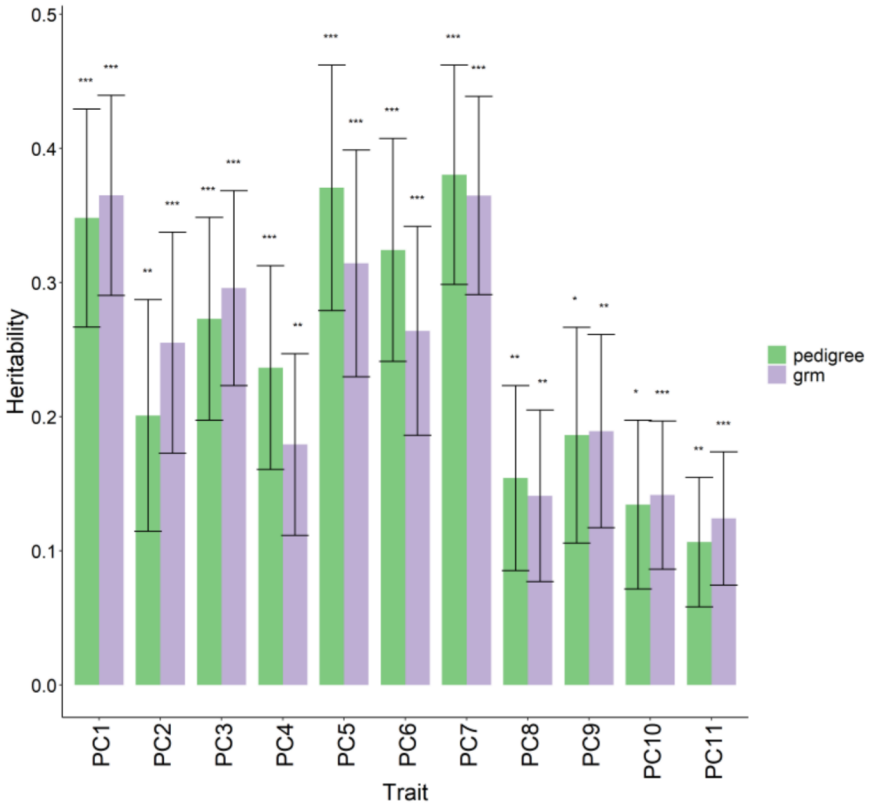


Figure 2.4: Heritability estimates for all 11 PCs. Both estimates from the pedigree and the GRM models are shown, h^2 refers to heritability, both standard error (*SE*) and p value for each heritability estimate are given in brackets. $^{*}P \leq 0.05$, $^{**}P \leq 0.01$ and $^{***}P \leq 0.001$.

Table 2.2: Proportions of phenotypic variance (V_P) explained by random effects in animal models of the ten antler measures, from models in which the additive genetic variance was estimated using the GRM. h^2 and *PE* indicate the proportions explained by the additive genetic and permanent environment effects, respectively. Proportions explained by birth year, growth year and the proportion of residual variance are given by *Birth Year*, *Growth Year* and *Residual*. Standard errors are given in brackets. Information on sample sizes and mean measures is provided in Table 2.1.

Trait	V_P	h^2	**PE**	**Birth Year**	**Growth Year**	**Residual**	**Repeatability**
Antler Length	61.155	0.279 (0.066)	0.105 (0.058)	1.04E-08 (0.014)	0.222 (0.047)	0.393 (0.035)	0.385
Coronet Circ.	3.047	0.408 (0.071)	0.128 (0.056)	0.018 (0.018)	0.243 (0.047)	0.203 (0.020)	0.554
Lower Beam Circ.	1.323	0.386 (0.067)	0.096 (0.054)	2.22E-07 (1.64E-08)	0.211 (0.042)	0.307 (0.027)	0.483
Upper Beam Circ.	1.084	0.357 (0.087)	0.280 (0.081)	5.16E-08 (0.017)	0.164 (0.035)	0.199 (0.018)	0.637
Coronet-Brow Junc.	0.872	0.275 (0.073)	0.254 (0.068)	0.026 (0.022)	0.152 (0.037)	0.292 (0.025)	0.555
Coronet-Tray Junc.	20.930	0.410 (0.083)	0.148 (0.072)	0.001 (0.016)	0.019 (0.009)	0.422 (0.032)	0.559
Brow Length	21.555	0.414 (0.066)	0.068 (0.050)	6.33E-09 (0.013)	0.269 (0.050)	0.250 (0.025)	0.481
Tray Length	13.809	0.388 (0.094)	0.273 (0.088)	7.65E-08 (5.34E-09)	0.028 (0.012)	0.311 (0.028)	0.661

Antler Weight	30727.710	0.229 (0.089)	0.263 (0.086)	0.023 (0.027)	0.252 (0.050)	0.233 (0.025)	0.514
Form	0.907	0.268 (0.052)	0.249 (0.046)	0.024 (0.015)	0.044 (0.011)	0.416 (0.019)	0.541
PC1	5.742	0.365 (0.075)	0.203 (0.064)	8.980E-09 (0.015)	0.264 (0.049)	0.169 (0.018)	0.568
PC2	1.492	0.255 (0.082)	0.311 (0.078)	0.009 (0.020)	0.063 (0.019)	0.362 (0.029)	0.575
PC3	1.145	0.296 (0.073)	0.136 (0.067)	0.038 (0.025)	0.078 (0.024)	0.451 (0.035)	0.470
PC4	0.830	0.179 (0.068)	0.205 (0.068)	0.017 (0.021)	0.061 (0.021)	0.538 (0.037)	0.401
PC5	0.694	0.314 (0.085)	0.253 (0.080)	4.015E-04 (0.017)	0.058 (0.018)	0.375 (0.030)	0.567
PC6	0.641	0.264 (0.078)	0.157 (0.072)	4.717E-07 (2.712E-08)	0.025 (0.012)	0.554 (0.039)	0.421
PC7	0.515	0.365 (0.074)	0.059 (0.065)	0.023 (0.019)	0.019 (0.011)	0.533 (0.038)	0.448
PC8	0.437	0.141 (0.064)	0.223 (0.069)	0.010 (0.018)	0.053 (0.018)	0.573 (0.038)	0.374
PC9	0.400	0.189 (0.072)	0.269 (0.072)	0.003 (0.018)	0.039 (0.015)	0.500 (0.035)	0.461
PC10	0.225	0.142 (0.055)	0.086 (0.056)	0.024 (0.020)	0.062 (0.021)	0.686 (0.039)	0.252
PC11	0.156	0.124 (0.050)	0.035 (0.047)	4.552E-07 (2.272E-08)	0.070 (0.023)	0.771 (0.039)	0.159

2.4.4 Genetic correlations between antler measures

Most antler measures were positively genetically correlated, suggesting some degree of shared genetic architecture (Table 2.3). A high degree of correlation was also reflected in the disproportionate contribution of all antler measures to PC1 (Figure 2.2). Antler weight was significantly positively genetically correlated with almost all other measures (except for coronet-tray junction), with genetic correlations ranging from r_a = 0.40 for its correlation with coronet-brow junction to r_a = 0.94 with upper beam circumference. The only significant negative genetic correlations were observed between coronet-tray junction and tray length, form, and lower beam circumference (Table 2.3). The phenotypic correlation between coronet-tray junction and tray length was non-significant, and that between coronet-tray junction and form was the smallest among all phenotypic correlations; however, the phenotypic correlation between coronet-tray junction and lower beam circumference was substantial r_a = 0.29 (see Table S2.1).

Table 2.3: Genetic correlations among all 10 antler measurements. Standard errors are given in brackets, *$P \leq$ 0.05, **$P \leq$ 0.01 and ***$P \leq$ 0.001. AL = Antler Length, CC = Coronet Circumference, LBC = Lower Beam Circumference, UBC = Upper Beam Circumference, CBJ = Coronet-Beam Junction, CTJ = Coronet-Tray Junction, BL = Brow Length, TL = Tray Length, AW = Antler Weight, F = Form.

	CC	**LBC**	**UBC**	**CBJ**	**CTJ**	**BL**	**TL**	**AW**	**F**
AL	0.678 (0.068) ***	0.640 (0.077) ***	0.664 (0.093) ***	0.674 (0.103) ***	0.665 (0.096) ***	0.549 (0.075) ***	0.447 (0.109) ***	0.854 (0.063) ***	0.406 (0.108) ***
CC	—	0.800 (0.054) ***	0.745 (0.084) ***	0.168 (0.109)	0.369 (0.089) ***	0.650 (0.067) ***	0.603 (0.088) ***	0.913 (0.065) ***	0.565 (0.095) ***
LBC	—	—	0.940 (0.024) ***	0.442 (0.102) ***	-0.322 (0.091) ***	0.592 (0.068) ***	0.810 (0.059) ***	0.928 (0.059) ***	0.780 (0.064) ***
UBC	—	—		0.266 (0.132) *	-0.179 (0.112)	0.541 (0.073) ***	0.733 (0.069) ***	0.943 (0.059) ***	0.902 (0.041) ***
CBJ	—	—	—	—	0.492 (0.108) ***	-0.079 (0.109)	-0.006 (0.137)	0.396 (0.185) *	-0.163 (0.126)
CTJ	—	—	—	—	—	0.055 (0.065)	-0.713 (0.073) ***	0.188 (0.146)	-0.521 (0.099) ***
BL	—	—	—	—	—	—	0.739 (0.070) ***	0.710 (0.073) ***	0.608 (0.065) ***
TL	—	—	—	—	—	—	—	0.811 (0.082) ***	0.822 (0.064) ***
AW	—	—	—	—	—	—	—	—	0.890 (0.061) ***

2.5 Discussion

In this study, I have used both a long-term pedigree and a genomic approach to quantify genetic (co-)variances of several antler morphology in the red deer of Rum. I have shown that antler morphology is heritable and repeatable over an individual's lifetime, and that antler traits are highly correlated on the phenotypic and genetic level, indicating some degree of shared genetic architecture. Here, I discuss how my findings build on previous quantitative genetic studies in the Rum red deer, and how they inform the broader question of the importance of and constraints on genetic variation of sexually selected male weaponry and the consequences for its evolution.

2.5.1 Genomic versus pedigree relatedness

The pairwise relatedness estimates from the pedigree and the GRM were very similar (linear regression slope = 0.87; Figure S2.1), and the degree of correlation was comparable to previous findings in another well studied wild ungulate, the Soay sheep of St Kilda (Bérénos *et al.*, 2014). For both antler measures and their PCs, pedigree and genomic relatedness resulted in very similar and generally large estimates of narrow sense trait heritability (h_2). It should be noted that the heritabilities could be over-estimated due to associations caused by shared environments on a fine spatio-temporal scale which might not have been captured by the model structure used (e.g., fitting only growth year and birth year; Stopher *et al.* 2012), but this would affect both the pedigree and GRM model in the same way and is thus difficult to ascertain. The strong agreement of estimates from the pedigree and GRM approaches indicate that the SNPs present on the red deer SNP chip are likely to be in sufficiently high LD with causative loci to allow accurate estimation of trait heritabilities in this population (Yang *et al.* 2010; Figures 2.3

& 2.4). Indeed, LD is maintained at a relatively constant level of ρ >0.1 over a distance of up to 1Mb, after which it starts to slowly decay (see Figure S3.1 in Chapter 3 Supplementary Material). Therefore, although the pedigree performs very well in a well-pedigreed population like the Rum red deer, the GRM can represent a reliable alternative when LD is high, thus encouraging its use in suitable populations with shallow or no pedigrees.

2.5.2 Heritability and repeatability of antler morphology.

All 10 antler measures were significantly heritable (ranging from 0.229 to 0.414, Table 2.2) and were similar to previous heritability estimates for antler weight and form in the same population (Kruuk *et al.*, 2014). All antler traits were also highly repeatable, with between 38.5% and 66.1% of the phenotypic variance explained by additive genetic, permanent environment and birth year effects (after correcting for age). This indicates that antler morphology is temporally stable over an individual's adult life, despite the annual shedding and regrowth of antlers. Similar findings were found for the 11 principal components, which showed substantial heritabilities, despite explaining as little as 1% of phenotypic variance in some cases (Figures 2.4 & 2.2). A study of skull morphology in wild mice found comparable patterns, as principal components used as phenotypic traits explained up to 50% of genetic variance (Pallares *et al.*, 2014). Generally, lower order PCs exhibited higher heritabilities and repeatabilities (Table 2.2); I discuss why I believe this is the case in the following Section. In addition to individual effects, there was significant contribution of the year of growth to phenotypic variance in antler traits, except for the coronet-tray junction, tray length and antler form, explaining 14% to 25% of the phenotypic variance (Table 2.2). Overall, my findings support previous work showing that traits under sexual selection can have substantial underlying genetic variation in wild populations (Kruuk *et al.*, 2008; Merila *et al.*, 2001; Pomiankowski & Moller, 1995), and of the same

magnitude as other morphological traits associated with fitness (Johnston *et al.*, 2013a; Bérénos *et al.*, 2014; Bourret *et al.*, 2017; Malenfant *et al.*, 2018).

2.5.3 Genetic correlations and constraints on antler morphology.

Almost all antler traits were positively genetically and phenotypically correlated, except for the coronet-tray junction, which was negatively genetically correlated with tray length, lower beam circumference and antler form. These findings were reflected in the PC analysis, where the PC explaining the most variance in antler morphology (PC1, $\sim$41%) combined equal information from all antler measures, whereas those explaining declining amounts of variation represented one or two antler traits.

The large heritabilities of individual antler traits suggest there is potential for response to selection, but my results further add to previous findings that constraints at the genetic level may affect how the population may respond to selection (Kruuk *et al.*, 2014). Previous work showed that a large part of genetic variance in antler weight is not available to selection due to the lack of genetic covariance between weight and fitness (Kruuk *et al.*, 2002, 2014). In the current study, antler weight was strongly correlated with most other antler measures at the genetic level, which may likely indicate that large parts of the genetic variance of these antler measures are also unavailable for selection, thus limiting the evolutionary potential of antler morphology despite large heritability estimates. Furthermore, positive genetic correlations can constrain the evolution of traits with large proportions of genetic variances if selection patterns among traits are antagonistic (Teplitsky *et al.*, 2014). In the PC analysis, higher order PCs explained much smaller amounts of variation and had very low additive genetic variation. It is possible that these PCs could

represent morphologically stable aspects of the antlers, where lower heritabilities could indicate past strong stabilising selection on the combination of trait aspects represented by that PC. Other studies exploring multivariate sexual selection on a suite of traits found that despite large genetic variance in univariate analyses, there can be very little genetic variance available in the trait composition that is the target of selection (Hunt *et al.*, 2007; Van Homrigh *et al.*, 2007). Overall, conclusions about what the evolutionary constraints on antler morphology are can only be reliably made when more is known about their genetic architecture. Therefore, in Chapter 3, I will address this by a fine-scale dissection of the genetic architecture of all aspects of antler morphology investigated above.

2.6 Conclusions

In this study, I have shown that antler morphology has substantial genetic variance, is highly repeatable over an individual's lifetime and that pedigree-free genomic approaches can be a reliable alternative in quantitative studies of wild populations. I also show that different antler traits are highly correlated genetically and phenotypically and can be partitioned into representative components. Overall, these findings indicate some degree of shared genetic architecture between different antler measures. The substantial heritabilities found here confirm previous findings in this population that antler morphology should be able to respond to selection, despite apparent evolutionary stasis. The highly interwoven genetic covariance structure of antler traits might contribute to maintain this stable phenotypic space. In the same line, reduced heritability of some antler PCs might indicate the aspects of antler morphology whose genetic variances were depleted by past selection, constraining any further evolution. A more detailed understanding of the genetic loci controlling antler morphology as

well as the selection pressure acting upon them is carried out in Chapter 3, in order to better dissect constraints on trait evolution within this population.

Chapter 3

Genomic approaches to understanding the genetic architecture of antlers in red deer (*Cervus elaphus)*.



Stag (Rumpel03) with vegetation stuck in antlers.

3.1 Summary

Sexually selected traits show large variation and rapid evolution across the animal kingdom, yet genetic variation often persists within populations despite apparent directional selection. A key step in solving this long-standing paradox is to determine the genetic architecture of sexually-selected traits to

understand evolutionary drivers and constraints at the genomic level. Here, I build on the findings in Chapter 2 to explore the genetic architecture of antler morphology in Rum red deer in more detail. Using the same genomic data set, I make use of genome-wide association and haplotype-based regional heritability approaches to identify regions of the genome underlying antler morphology. In addition, I use an Empirical Bayes approach to estimate the underlying distributions of allele effect sizes. Almost all traits were underpinned by SNPs with uniformly small effect sizes, with few traits showing distributions that included SNPs with larger effects on phenotype. A single region on linkage group 21 was associated with antler PC9 in the regional heritability analysis and linked to candidate genes that potentially have an effect on antler morphology. I show that overall, antler morphology has a polygenic architecture, with some loci identified as having pleiotropic effects. Both pleiotropy and linkage disequilibrium (LD) are expected to add to the widespread marker sharing and subsequent genetic covariance among antler traits. These findings, together with the results of Chapter 2, suggest that a large mutational target and extensive genetic correlations among antler traits are likely to contribute to the maintenance of genetic variation in antler morphology in this population.

3.2 Introduction

The maintenance of genetic variation of sexually selected traits remains an evolutionary paradox (Pomiankowski & Moller, 1995; Kotiaho *et al.*, 2001; Kruuk *et al.*, 2002; Svensson & Gosden, 2007) that has been the topic of numerous quantitative genetic studies (Emlen, 1994; Griffith *et al.*, 1999; Kruuk *et al.*, 2002; Robinson *et al.*, 2006). As discussed in Chapters 1 and 2, a key limitation of most studies to date is that most aspects of the genetic architecture of secondary sexual traits are generally unknown - that is, the

underlying loci, their number, distribution and relative effect sizes, and the extent of pleiotropy, epistasis and other interactions (Timpson *et al.*, 2018; Chenoweth & McGuigan, 2010). By identifying the genetic architecture of secondary sexual traits, we can better understand the underlying molecular mechanisms and evolutionary processes that drive their variation (Dobzhansky, 1971; Lewontin, 1974; Kuijper *et al.*, 2012; Wilkinson *et al.*, 2015).

Recent genomic advances in natural populations have led to a number of studies characterising genetic architectures using genome-wide association studies (GWAS, reviewed in Santure & Garant 2018). Yet, relatively few studies exist for sexually selected traits, with much of the focus on discrete traits with Mendelian or relatively simple genetic architectures, including horn development in Soay sheep (Johnston *et al.*, 2011), age of sexual maturity in Atlantic salmon (Barson *et al.*, 2015), male mating morphs in ruffs (Küpper *et al.*, 2015), and exaggerated reproductive morphs in the spider *Oedothorax gibbosus* (Hendrickx *et al.*, 2021). In these rare cases, mapping specific genomic variants associated with sexual trait variation can allow investigation of sex, age and environment-specific effects at individual loci. As such, they have revealed compelling cases of heterozygote advantage due to trade-offs between reproductive success and survival (Johnston *et al.*, 2013a), or due to differences in optimal trait expression between the sexes (Barson *et al.*, 2015). However, in most other cases, secondary sexual traits are likely to have oligogenic or polygenic architectures (i.e., moderate to large numbers of underlying loci), particularly in cases where they are condition dependent (Rowe & Houle, 1996). Therefore, despite the availability of genomic tools, we still know very little about the underlying genes that control sexually selected traits such as male weaponry, as attempts to identify major effect loci often have been inconclusive (Poissant *et al.*, 2012; Miller *et al.*, 2018; Sim & Coltman, 2019), with some evidence pointing towards a polygenic architecture (Anderson *et al.*, 2019).

As the number of loci underlying a trait increases and their relative effect sizes decrease, it becomes more difficult to implicate individual loci in trait variation; for example, in analysis of heights of people of European ancestry, only a fraction of the loci underpinning variation has been identified, despite sample sizes of >700,000 individuals and >16 million SNPs (Yengo *et al.*, 2018). On the other hand, being able to determine that a trait has a polygenic architecture can still shed light on how sexual traits evolve for the following reasons. First, polygenic traits present a large mutational target, contributing to the maintenance of genetic variance via the introduction of new variants (Rowe & Houle, 1996). Second, allele frequency changes due to selection at a great number of loci is expected to result in a rapid change in trait mean, and thus trait evolution. This evolutionary response to selection is then sustained by the aforementioned large mutational input, so that the distribution of genetic effects and thus genetic variance is left unperturbed (Barton *et al.*, 2017; Sella & Barton, 2019). Third, pleiotropy and/or linkage between loci could maintain variation through conflicts between traits sharing a similar or linked polygenic architecture (Ruzicka *et al.*, 2019; Teplitsky *et al.*, 2014; Morrissey *et al.*, 2012). Therefore, studies of sexual traits should aim to identify specific genetic variants with large effects on phenotype, and should also aim to determine the distribution of polygenic effect sizes and the degree to which these underlying loci are shared between traits. This will not only shed light on potential evolutionary processes and mechanisms in empirical studies, but will also inform the mechanistic details of theoretical models to allow better assumptions to account for the complexities of natural populations (McNamara & Houston, 2009; Wilkinson *et al.*, 2015).

In the previous Chapter, I showed that antler morphology in the Red deer on Rum is moderately heritable. In this Chapter, I build on this finding by determining the genetic architecture of antler morphology in this population. I aim to identify loci underlying genetic variation in antler traits and their principal components (PCs) by conducting genome-wide association studies

(GWAS), which identify SNP loci linked to genetic variation of a trait (Yang *et al.*, 2011). I also conduct a regional heritability analysis, an approach that attempts to link trait variation to haplotypic variation within defined genomic regions (Nagamine *et al.*, 2012). Using both single locus and haplotype based approaches to identify causal genetic markers allows for the identification of both large effect size and medium to small effect size loci (Nagamine *et al.*, 2012). In addition, I use an Empirical Bayes method to estimate the underlying distributions of allele effect sizes outwith genome-wide significance (Stephens, 2016). This approach has been used successfully to characterise smaller effect loci contributing to genetic variation in human height that were not previously identified by GWAS (Boyle *et al.*, 2017). My findings suggest that genetic variation in antler morphology in the Rum red deer is highly polygenic, and that it is maintained via a large mutational target and pleiotropy with traits sharing similar complex polygenic architectures in this population.

3.3 Methods

3.3.1 Study system and antler morphology

Information on the study system and the data-set of antler measures and their principal components (PCs) are identical to those described in Chapter 2.

3.3.2 Genomic data-set

The genomic data was generated as outlined in Chapter 2. In this Chapter, the data consisted of the SNP markers of known position in the red deer linkage map (Johnston *et al.*, 2017), as described in Chapter 2. The SNP positions and ordering in both the GWAS and regional heritability analyses were based

on the linkage map data. The quality control of mapped SNPs and X-linked markers (i.e., heterozygous state in males) was as detailed previously, using the *check.marker* function with default thresholds in the R library GenABEL v1.80 in R v3.4.2 (Aulchenko *et al.*, 2007). The final SNP dataset consisted of 2,138 individuals and 38,006 markers. The total number of observations per trait are given in Table 3.1.

Table 3.1: Number of observations (including multiple observations per individual across years) for each antler trait used for GWAS after quality control.

Trait	N Observations
Antler Length	860
Coronet Circumference	979
Lower Beam Circumference	967
Upper Beam Circumference	904
Coronet-Brow Junction	929
Coronet-Tray Junction	950
Brow Length	910
Tray Length	785
Antler Weight	773
Form (Total Number of Points)	2699

3.3.3 Genome-wide association studies

Genome-wide association studies (GWAS) (Uffelmann *et al.*, 2021) were conducted in RepeatABEL v1.1 (Rönnegård *et al.*, 2016) implemented in R v3.4.2. First, the *prefitModel* function was used to fit a linear mixed model (without fixed SNP effects), specifying the same fixed and random effect structure as Equation 2.3 (see Chapter 2). The resulting covariance matrix of the random effects was then input to the *rGLS* function, which fits each SNP

genotype as an additive linear covariate. This approach accounts for population structure by fitting the genomic relatedness matrix (GRM, as calculated in Chapter 2) as a random effect and allows for repeated measures per individual. The significance of each SNP was determining using Wald tests, distributed as χ^2 with 1 degree of freedom. These statistics were corrected for potential inflation due to population structure not captured by the GRM by dividing them by the genomic inflation factor λ, defined as the observed median χ^2 statistic divided by the null expectation median χ^2 statistic (Devlin & Roeder, 1999). This was done separately for each antler measure. After correction, the genome-wide significance threshold was set to a p-value of 1.42×10^{-6}, equivalent to α=0.05 after correcting for multiple-testing and accounting for non-independence due to LD between SNP markers (see Johnston *et al.* 2018).

3.3.4 GWAS power analysis

To better distinguish whether any null results in the GWAS were driven by the absence of large effect QTLs and/or power issues due to small sample sizes, I conducted a simulation-based power analysis. First, I used a permutation approach to assign the phenotypes of a given antler trait (along with their associated values for age, rut, and birth year) randomly across individuals. This was done to break any existing correlations between SNPs and phenotypes in the data. Then, I picked a SNP at random and assigned values to the three genotypes at this SNP such that the additive genetic variance explained by the SNP equalled a predetermined SNP heritability. In this way, I was able to determine the power to detect SNP effects of a chosen magnitude. I assumed a fully additive model, with the additive genetic value of the heterozygote exactly intermediate between the two homozygotes. I then removed the simulated causal SNP from the dataset so that the QTL effect would have to be detected via a linked SNP, as is usually the case in GWAS.

Finally, I carried out a GWAS exactly as described above for the original empirical data for each antler trait (excluding antler PCs) and calculated power as the proportion of simulations that resulted in the detection of one or more significant SNPs. I conducted this analysis for a range of SNP heritabilities, representing very small to very large effects: 0.01, 0.05, 0.1, and 0.4, for each of which a total of 100 simulation runs were performed.

3.3.5 Regional heritability analysis

In addition to GWAS, I used a regional heritability approach to identify regions of the genome associated with antler trait variation. This method uses information from multiple loci to determine the proportion of V_P explained by defined genomic regions, and has increased power to detect variants of small effect sizes and low minor allele frequencies (Yang *et al.*, 2011; Nagamine *et al.*, 2012). Regions were defined using a 'sliding window' approach with SNPs of known position on the Rum deer linkage map (Johnston *et al.*, 2017). Each window was 20 SNPs wide and overlapped the preceding window by 10 SNPs. If the last window in the linkage group contained less than 20 SNPs (i.e., the number of SNPs on the chromosome was not divisible by 20), the last 20 SNPs of that linkage group were taken instead. This resulted in a total of 3,608 autosomal genomic windows. For each window, a GRM was constructed using the *--make-grm function* in GCTA v1.24.3 (Yang *et al.*, 2011). The GRM was adjusted to assume similar frequency spectra of genotyped and causal loci with the argument *--grm-adj 0*. In addition to this, I attempted to conduct regional heritability analyses using windows in Linkage Group 34, which corresponds to the X chromosome, by calculating the GRMs using the GCTA function *--make-grm-xchr*; however, no subsequent models using X-linked markers converged and so these windows were not investigated in the current study.

The contribution of each genomic region to V_A and V_P for each antler measure and PC was modelled as follows:

$$y = X\beta + Z_1(a - v_i) + Z_2 v_i + Z_r u_r + e \tag{3.1}$$

with variables defined as in Equation (2.3), but with the additive genetic components split into two terms: $Z_1(a - v_i)$ and $Z_2 v_i$, where Z_1 is an incidence matrix relating individual measures with additive genetic effects for the genome excluding window i, $(a - v_i)$ is the vector of the additive genetic effects explained by all autosomal markers not in window i; Z_2 is an incidence matrix relating individual measures to additive genetic effects for window i and v_i is the vector of additive genetic effects explained by all autosomal markers in window i. The significance of an association between a window i and an antler trait was determined using a likelihood ratio test (LRT) comparing models containing and omitting the term $Z_2 v_i$ (i.e., 2 $\times$ the difference between the model log-likelihoods, assuming a χ^2 distribution with 1 degree of freedom). The distribution of χ^2 statistics from the LRTs across all windows was corrected using the genomic control parameter λ, calculated using the same approach as above (Devlin & Roeder, 1999). A genome-wide significance threshold was calculated using a Bonferroni correction, where the α significance level (here 0.05) was divided by the number of effective tests. For this, I divided the number of windows by two to account for the overlap of half the number of total SNPs within each window, resulting in a significance threshold of P = 2.77×10^{-5}.

3.3.6 Estimation of SNP effect size distribution

I investigated the distribution of allele effect sizes and false discovery rates (FDR) for all antler measures and PCs using the *ash* function in the R package ashR v2.2-32 (Stephens, 2016). This applies “adaptive shrinkage”,

within an Empirical Bayes framework that uses the slopes and standard errors of the additive SNP effects from the GWAS models above to compute a posterior distribution of SNP effect sizes across all loci. This approach estimates the local false discovery rate (*lfdr*), which is the probability that the SNP effect is zero. The significance of a SNP effect was then determined by a local false sign rate (*lfsr*), defined as the probability of making an error when assigning a sign (positive or negative) to an effect, with a cut-off at α = 0.05. The prior distribution was specified to be any symmetric unimodal distribution when applying the *lfdr* estimation. In order to separate the effects of LD between markers, this analysis was carried out on the full SNP dataset and on LD-pruned datasets where the threshold of pairwise r^2 between SNPs was 0.2, 0.5 or 0.8. Pruning was carried out in PLINK v1.9 (**?**) using the *--indep-pairwise* function, with a window size of 20 SNPs and an overlap of 10 SNPs. This resulted in datasets of 13459, 30634 and 35924 SNPs, respectively.

I then re-estimated the effect sizes of SNPs with the highest non-zero effects (based on adaptive shrinkage; Stephens 2016) using the animal model framework (Equation 2.3 in Chapter 2) in ASREML-R v4.0 (Butler *et al.*, 2009). A maximum number of 10 SNPs per trait were taken. SNP genotype was fit as a three level factor and significance was tested using a Wald test. Fitting SNP genotype as a factor allowed quantification of both the dominance deviation and the variance-covariance structure of the genotypes, which are needed to estimate the variance attributed to each SNP, calculated as follows (Falconer & Mackay, 1996):

$$V_{SNP} = 2pq(a + d(q - p))^2 \tag{3.2}$$

where p and q are the allele frequencies of alleles A and B, respectively; a is the additive genetic effect defined as the mid-point between the effect sizes of the genotypes AA and BB; and d is the dominance deviation defined as the

difference between a and the effect size of the heterozygote AB. The proportion of V_A attributed to a SNP was calculated as the ratio of V_{SNP} to the sum of V_{SNP} and the V_A obtained from an animal model where the SNP effect was omitted. Standard errors of V_{SNP} were estimated using the *deltamethod* function in the R library msm v1.6.7 (Jackson, 2011) in R v3.4.2.

Data Statement

Data for this study, which includes genomic data after quality control measures, pedigree and linkage map information, and autosomal GRM, are available in a Dryad repository (doi:10.5061/dryad.612jm643c) ([dataset] Peters *et al.*, 2021). All scripts for the analysis are provided at `https://github.com/Lucy-Peters/Red\_deer\_antler\_genetic\_architecture`. Due to their size, some of the Supplementary Tables have only been provided with their caption information in the printed version of this thesis; they can be accessed using the private Dryad link above.

3.4 Results

3.4.1 Genome-wide and regional heritability studies

Genome-wide association studies did not identify any SNP loci that were significantly associated with any antler measure (Figure 3.1, Table S3.1) or PC (Figure S3.2, Table S3.2). According to the power analysis, the power to detect very small effect QTLs that have a SNP heritability of 0.01 was very low for most antler traits, from $\sim$0% for tray length to 6% for brow length. Slightly larger effect QTLs with SNP heritability of 0.05 had a higher probability of detection, with power ranging from around 40% for coronet tray

junction length to over 60% for coronet circumference. For QTLs with a SNP heritability of 0.10, the power was substantial across most traits, ranging from around 60% for upper beam circumference to 90% for brow length. Very large QTLs with a SNP heritability of 0.40 were extremely likely to be identified across all traits, with power ranging from 98% to 100%. Antler form, likely in part due to having the largest sample size, showed the highest statistical power of detection of all antler traits across all categories of QTL effect size, even for the smallest SNP heritability of 0.01, which at 23% was several times larger than that of any other antler trait at that effect size (see Table S3.3 for full results).

Furthermore, the regional heritability analysis found no regions of the genome significantly associated with any antler measure (Figure 3.2, Table S3.4) or PC (Figure S3.3, Table S3.5), except for PC9, which was significantly associated with three overlapping windows, corresponding to a $\sim$4.4 Mb region on Red deer linkage group 21. The most strongly associated window explained 16.8% (SE = 8.0%) of the phenotypic variance and 66.3% (SE = 22.0%) of the additive genetic variance. However, it should be noted that PC9 accounts for only 4% of overall phenotypic variance among all antler measures (see Figure 2.2 in Chapter 2). Homology with the cattle genome (version ARS-UCD1.2) suggested that there are a total of 19 coding regions within the $\sim$4.4 Mb region covered by all three significant windows. Details of SNPs within this region can be found in Table S3.6 and associated gene ontology (GO) terms can be found in Table S3.7.

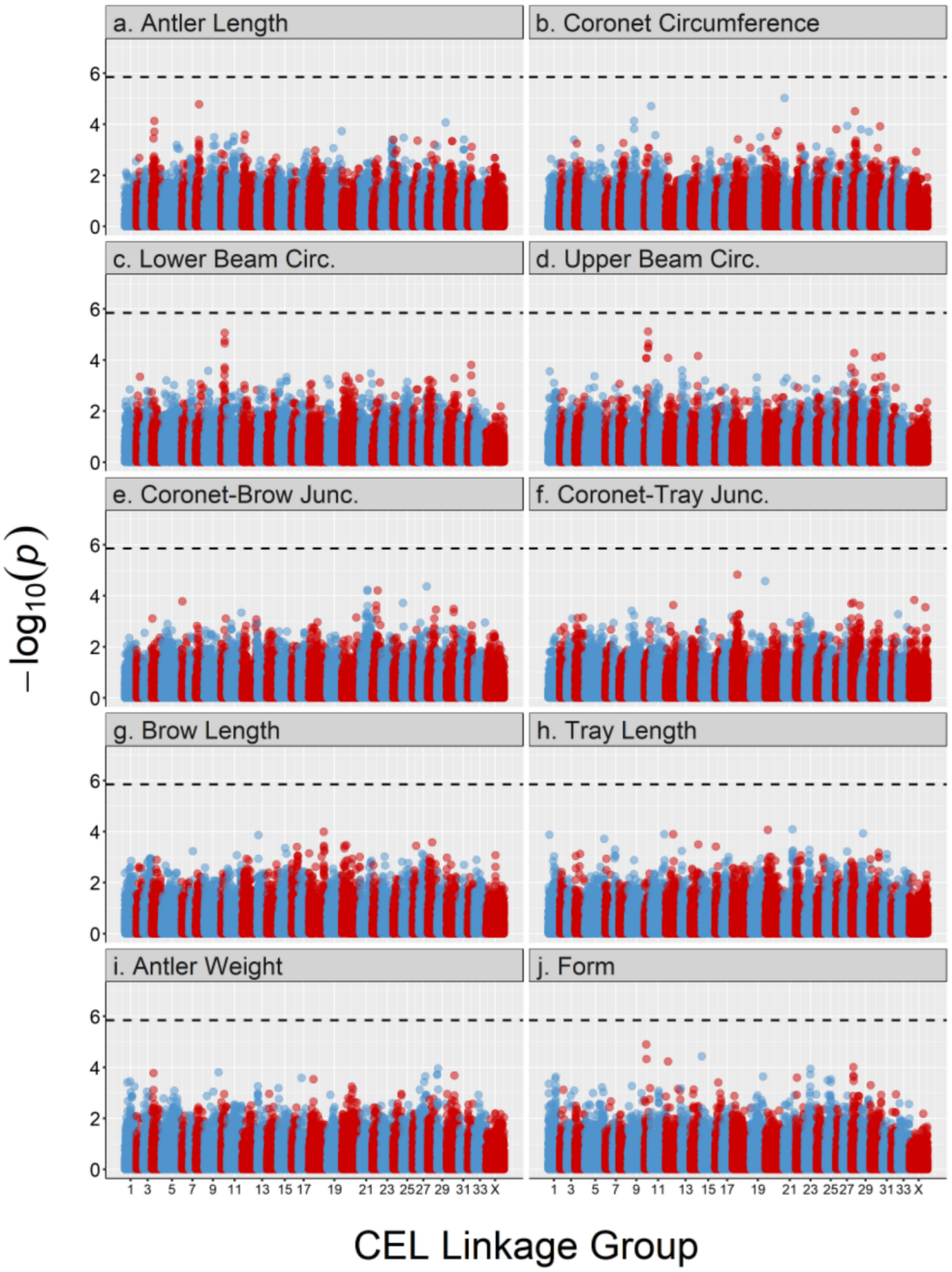


Figure 3.1: Genome-wide association study for antler measures. The dashed line indicates the significance threshold, equivalent to α = 0.05. Points are colour-coded by chromosome. Underlying data is provided in Table S3.1.

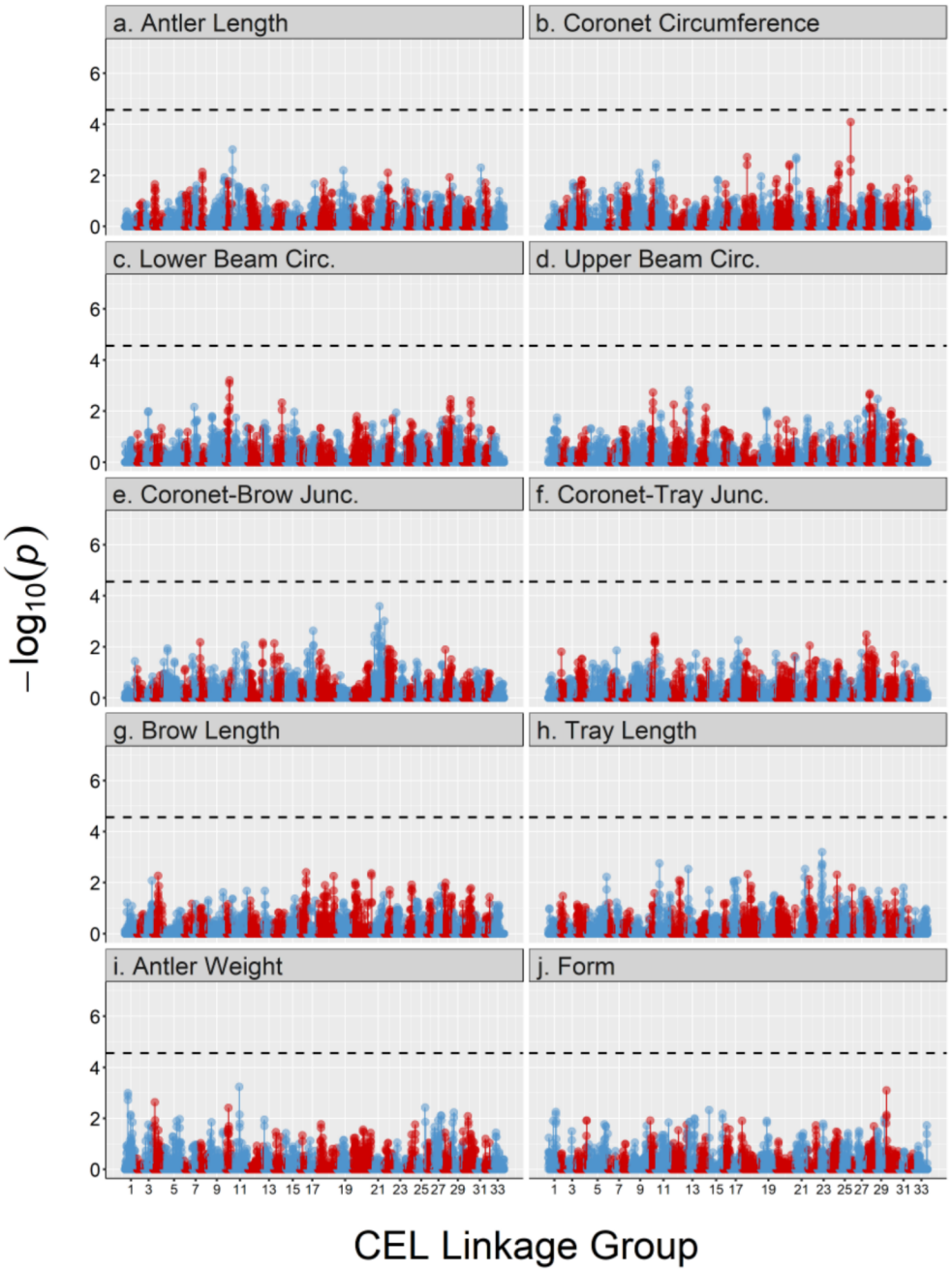


Figure 3.2: Regional heritability analysis for antler measures. The dashed line indicates the significance threshold, equivalent to α = 0.05. Points are colour-coded by chromosome. Underlying data is provided in Table S3.4

3.4.2 Distribution and quantification of SNP effect sizes

A total of 897 unique SNPs had significant non-zero effects across the ten antler measures (Table 3.2; full results are provided in Table S3.11). The number of significant non-zero SNPs ranged from 15 SNPs (antler length) to 279 SNPs (lower beam circumference; Table 3.2), although antler weight and upper beam circumference had no significant non-zero effect SNPs. Several SNPs showed pleiotropic effects, i.e., they were associated with more than one measure (Table S3.11), with the underlying proportion ranging from 6% (coronet-tray junction) to 68% (tray length; Table 3.2). Lower beam circumference and antler form showed distributions that included more extreme SNPs with large effects on phenotype outside the 5% to 95% quantile boundaries, whereas most other measures had more uniformly distributed effect sizes (Table 3.2). For the 11 antler PCs, only 159 unique SNPs had non-zero effects (Table S3.10; full results in Table S3.12). No pleiotropic SNPs were observed, most likely due to the independence of each PC. PC4 and PC9 had no non-zero effect SNP associations. Only about 25% of the 159 SNPs were in common with the 897 SNPs in the antler measure analysis. This was mainly due to the higher order PCs (PC6 to PC11) sharing no SNP associations with any antler measures, while other PCs had similar or greater numbers of shared and unique SNPs (PC1 and PC2). The LD-pruned datasets yielded lower numbers of non-zero SNPs compared to the full SNP dataset, resulting in 299, 806 and 896 unique SNPs across all antler measures at r^2 = 0.2, 0.5 and 0.8, respectively. The distribution of non-zero SNPs, SNP effect sizes and the proportions of pleiotropic SNPs for each trait were similar across all data sets, giving qualitatively similar results to the full dataset. Full summaries of these results for all antler measures and PCs are provided in Tables S3.8 and S3.9.

I examined the contribution of SNPs with the highest non-zero effects in the FDR analysis to the additive genetic and phenotypic variance by fitting the

SNP genotypes as a fixed effect within an animal model framework. Some SNPs appear to explain large proportions of overall genetic and phenotypic variance for their respective traits. For example, the SNP *cela1_red_8_56767953* explained 29% of the overall additive genetic and 13% of phenotypic variance in coronet circumference (SNP variance = 0.4499, SE = 0.0760), while for PC8 (which is strongly representative of antler length; Figure 2.2) the marker *cela1_red_4_41756873* explained 59% of the additive genetic and 9% of the phenotypic variance (SNP variance = 0.0413, SE = 0.0077). However, we noted that the standard errors of these variance estimates were also large (see Tables S3.14 and S3.15 for full results for the antler measures and PCs, respectively). Furthermore, it is likely that effect sizes are overestimated due to the Beavis effect, which describes the inflation of effect sizes of quantitative trait loci, particularly within small sample sizes (Beavis, 1998; **?**; **?**). Therefore, whilst it can be determined that these SNPs have significantly non-zero effects on phenotype, it remains challenging to determine the precise proportion of the genetic variance that they explain.

Table 3.2: Summary of SNPs with non-zero effects on antler measures. N SNPs is the number of SNPs with non-zero effects. Proportion pleiotropic is the proportion of SNPs with non-zero effects on other antler measures. Maximum and minimum effects are given relative to the data scale (units = cm for lengths, g for weight). Lower and upper quantiles refer to the 5% and 95% boundaries of the effect size distribution, respectively.

Antler Measure	**N SNPs**	**Proportion Pleiotropic**	**Maximum Effect**	**Minimum Effect**	**Lower Quantile**	**Upper Quantile**
Antler Length	15	0.400	0.668	-0.622	-0.619	0.653
Coronet Circ.	56	0.214	0.151	-0.159	-0.152	0.146
Lower Beam Circ.	279	0.136	0.251	-0.151	-0.120	0.144
Upper Beam Circ.	0	-	-	-	-	-
Coronet-Brow Junc.	29	0.138	0.092	-0.084	-0.084	0.091
Coronet-Tray Junc.	110	0.055	0.460	-0.490	-0.435	0.441
Brow Length	272	0.081	0.538	-0.532	-0.484	-0.477
Tray Length	6	0.667	0.357	-0.357	-0.355	0.354
Antler Weight	0	-	-	-	-	-
Form	193	0.124	1.018	-0.097	-0.091	0.096

Table 3.3: Summary of SNPs with the highest additive genetic effect sizes for antler measures. The Wald statistics and associated p-value are taken from the animal models fitting SNP genotypes as categorical variable. The additive effect (*a*)is taken from the calculation of SNP variance based on (Falconer & Mackay, 1996). The phenotypic and genetic proportion refer to the proportion of total genetic and phenotypic variance attributed to the SNP.

Antler Measure	**SNP Name**	**Wald statistic**	**P-value**	a	V_{SNP}	**SE**	**Prop.** V_A	**Prop.** V_P
Antler Length	cela1_red_2_101997097	20.143	4.227E-05	2.351	0.325	1.094	0.025	0.005
Coronet Circ	cela1_red_8_56767953	13.889	9.640E-04	0.591	0.450	0.076	0.289	0.133
Lower Beam Circ.	cela1_red_25_28715823	16.540	2.561E-04	0.655	0.073	0.015	0.140	0.056
Upper Beam Circ.	-	-	-	-	-	-	-	-
Coronet-Brow Junc.	cela1_red_12_49003594	10.522	0.005	0.251	0.088	0.018	0.288	0.094
Coronet-Tray Junc.	cela1_red_x_20398798	20.851	4.9645E-06	1.726	0.462	0.630	0.045	0.002
Brow Length	cela1_red_11_5634923	13.782	0.001	1.596	2.759	0.505	0.211	0.114
Tray Length	cela1_red_14_63829569	16.757	2.297E-04	1.303	0.195	0.305	0.035	0.015
Antler Weight	-	-	-	-	-	-	-	-
Form	cela1_red_25_15869276	17.126	1.911E-04	0.837	0.017	0.011	0.070	0.019

3.5 Discussion

In this Chapter, I have used genomic approaches to determine the genetic architecture of antler morphology in the red deer of Rum. In the previous chapter, I had shown that antler morphology is heritable and repeatable over an individual's lifetime. Building on those findings, I find that heritable variation is likely to be highly polygenic, with a moderate degree of shared genetic architecture across different antler traits. Genome-wide association and

regional heritability studies failed to identify any genomic regions associated with antler measures and their principal components, except for a single region on Linkage Group 21 associated with variation in PC9. Rather, most traits were underpinned by SNPs with uniformly small effect sizes, whereas others, such as lower beam and antler form, showed distributions that included some SNPs with larger effects on phenotype. Overall, my findings suggest that antler morphology has a genetic architecture that largely consists of many loci of small effect. In this section, I discuss how my findings build on results of Chapter 2, how they contribute to our understanding of genetic architectures of sexually selected male weaponry, and address its influence on evolutionary outcomes.

3.5.1 The polygenic architecture of antler morphology.

Genome-wide association studies and regional heritability analyses across all antler traits and PCs showed no significant associations, with the exception of PC9 (discussed in the next section). It is not uncommon for GWAS in natural systems to fail to find single locus associations. Common issues include insufficient marker densities, as wild populations generally have smaller numbers of SNPs (1—10s of thousands) compared to humans, livestock, and model systems (100s to 1000s of thousands), which means that genotyped SNPs may be less likely to be in LD with causative trait loci. In addition, the sample size of individuals is often limited in natural populations as well, hundreds of individuals compared to 100s of thousands in humans for examples, which further reduces the ability to detect significant associations in these studies (Miller *et al.*, 2018). Often however, null results of GWAS in wild populations are due to the highly quantitative nature of traits meaning that they are likely to be polygenic (Berenos *et al.*, 2015; Santure *et al.*, 2015). As discussed previously, GWAS can only detect loci with moderate to large effects on phenotype, which only partially explain the trait heritabilities, with

the remainder termed the “missing heritability” (Manolio *et al.*, 2009; Golan *et al.*, 2014). For example, a meta-analysis of human GWAS found that the heritability attributed to all common SNP variants was significantly higher than that of only the SNPs that achieved genome-wide significance (Shi *et al.*, 2016), meaning that large numbers of “non-significant” SNPs will contribute to the additive genetic variation.

To characterise the genetic architecture beyond GWAS alone, I employed two additional approaches. The first, regional heritability (Nagamine *et al.*, 2012), incorporated the haplotypic diversity within genomic regions. This approach detected a large contribution of defined genomic regions to a single PC, but not to any of the other antler measures or PCs. The second, the Empirical Bayes false discovery rate and effect size estimation, incorporated information from the GWAS effect size estimates and their error to show that a substantial number of SNPs had non-zero effects on antler morphology (Stephens, 2016). This is in line with a study revisiting GWAS data for human height that used the same method to detect contributions from common SNPs and found that 62% of all markers (including linked SNPs) had non-zero effects on height (Boyle *et al.*, 2017).

I also quantified the SNP variances of markers with the largest additive genetic effects after Empirical Bayes FDR correction, and found for antler measures that they accounted for up to 17% of the genetic and up to 5% of the phenotypic variance (antler length and tray length, respectively, see Table 3.2) and up to 38% of the genetic and up to 6% of the phenotypic variance for PCs (PC8 and PC2, respectively; see Table S3.10). Although these estimates appear substantial, the associated standard errors were large, and it is very likely that the effect sizes were overestimated due to the Beavis effect (Slate, 2013), which describes the inflation of effect sizes of significant QTL hits and is especially problematic for small sample sizes (Beavis, 1998; Xu, 2003; Slate, 2013; Husby *et al.*, 2015); however, they are still considerable for

sample sizes of the order used in this study (King & Long, 2017). Furthermore, the GWAS power analysis clearly confirmed that medium to large effect QTLs with SNP heritabilities of 0.10 and 0.40 had very high probabilities of detection, and even QTLs with a SNP heritability of 0.05 were associated with non-trivial statistical power for some antler traits.

Taken together, these analyses provide compelling evidence that most aspects of antler morphology have a highly polygenic architecture. This is in line with findings from other studies of wild organisms that have identified polygenic architectures for morphological and life history traits (Robinson *et al.*, 2013; Santure *et al.*, 2013; Pallares *et al.*, 2014; Berenos *et al.*, 2015; Husby *et al.*, 2015). In this study, I modelled antler morphology using both independent measures of antler morphology and a principal component framework to characterise different dimensions of shape variation. An advantage of using both approaches was that it allowed us to identify a greater number of potentially causal loci, which supports the usefulness of this approach when trying to characterise the genetic architecture of a complex morphological trait (such as in Pallares *et al.* 2014).

3.5.2 Genomic regions associated with antler morphology.

Three adjacent genomic windows within CEL linkage group 21 were associated with PC9 at the genome-wide level (Figure S3.3), which was mainly associated with coronet circumference. The region covered by the most highly associated window explained ∼66% of the additive genetic variance, representing a large part of the overall heritability estimated using the whole GRM (∼19%; Chapter 2). The region contains a number of candidate genes among which are gasdermin C (*GSDMC*); MYC proto-oncogene (*MYC*); ArfGAP with SH3 domain, ankyrin repeat and PH domain 1 (*ASAP1*) and cellular communication network factor 4 (*CCN4*).

Whilst none of these have previously been implicated in antler morphology, they have associated functions that make them potential candidate genes. Both the enhancer protein *ASAP1* and *CCN4*, which is a type of connective tissue growth factor, are linked to bone ossification and bone cell differentiation in mice (The Jackson Laboratory, 2019; Schreiber *et al.*, 2019; Maeda *et al.*, 2015), processes which are likely to be vital to antler regeneration, rapid growth and the ability of antlers to withstand impact (Goss, 1983). Upregulation of *GSDMC* is implicated in carcinogenesis in mice, as a consequence of an interrupted growth factor signalling pathway (Miguchi *et al.*, 2016) and *MYC* is a potent oncogene that is implicated in many human cancers and promotes rapid tumour cell proliferation (Beroukhim *et al.*, 2010; Lin *et al.*, 2012); recent work suggests that rapid regeneration of bony antlers has evolved by upregulating cell proliferation pathways while suppressing tumorigenesis (Wang *et al.*, 2019).

Despite being linked to compelling candidate gene regions, with my current data it is virtually impossible to determine exactly which genes in this region drive the association with PC9. Furthermore, as PC9 only represents $\sim$4% of the overall phenotypic variance among all antler PCs, I expect effect sizes of causal loci to be very small. Validating the findings for this association would be challenging, as replication of a similar PC in other deer populations would have to consider its main contributing antler measures (e.g., coronet circumference); additionally, the same variants may not be associated with this trait in different populations. A more feasible approach may be to type a higher density of SNP loci to characterise more variation at (or in tight linkage with) potential causal loci in the Rum deer population.

3.5.3 Shared genetic architecture and constraints on antler morphology.

The Empirical Bayes analysis identified varying degrees of marker pleiotropy associated with the antler measures (Table 3.2) that were consistent with the observed genetic correlations in Chapter 2. Nevertheless, there were some exceptions, such as for brow length and antler weight, which both showed strong positive genetic correlations (see Chapter 2) yet had small proportions of shared loci (brow length) or no associated loci at all (antler length). This incongruity may be explained by the large variation in the number of non-zero effect SNPs detected between antler measures, which could be due to differences in the effect size distribution of markers.

The sharing of non-zero SNPs among antler traits suggests that some aspect of genetic correlations between antler traits is due to pleiotropic effects of these shared loci, meaning that they share a functional basis. This is perhaps to be expected, given that these traits are all part of the same structure and tissue type, and pleiotropic mutations are ubiquitous in highly polygenic traits (Lande, 1980). However, as there is extensive LD observed within this population (Figure S3.1) maintained by significant levels of inbreeding (Huisman *et al.*, 2016; Slate *et al.*, 2000; Coulson *et al.*, 1999; Pemberton *et al.*, 1999), we cannot rule out that SNP sharing and genetic correlations may also be partly explained by signals of linkage disequilibrium between SNPs underpinning different aspects of antler morphology. It is difficult to distinguish between these two mechanisms with certainty outwith an experimental setting, study replication or extensive simulation, which are beyond the scope of this thesis. However, the results using LD-pruned data sets suggest that although the absolute number of non-zero SNPs may be overestimated due to LD, the proportion of shared SNPs stayed consistent, supporting the presence of some level of pleiotropy. Furthermore, as I find no marker sharing of non-zero SNPs among PCs, these SNPs conceivably

represent a distinct set of genetic loci that affect functionally similar aspects of the antler traits a particular PC represents.

The discovery that the genetic architecture of antler morphology is highly polygenic suggests that particular evolutionary mechanisms may be partially responsible for maintaining genetic variation in this trait. As discussed in the introduction, traits with many genes of small effect can present a large mutational target for the introduction of novel genetic variation, which can contribute to the genetic variation in a trait (Rowe & Houle, 1996). Consequently, although selection on polygenic traits can lead to rapid changes in trait mean, under the infinitesimal model, the distribution of underlying genetic effects is expected to remain relatively constant, counteracting the loss of genetic variation (Barton *et al.*, 2017; Sella & Barton, 2019). This explanation is plausible in the red deer, as I have shown high polygenicity of antler traits which could provide this mutational target. In addition, I have shown that the genetic architecture of antler traits are likely to possess some level of pleiotropy with each other. This is also expected for traits with polygenic architectures, as many genes affecting a trait will increase their likelihood of being affected by pleiotropy and/or linked selection with loci that are associated with other traits, including those related to fitness (Lande, 1980). This in turn could maintain genetic variation through widespread genomic conflicts and trade-offs (Lande, 1982). Finally, empirical evidence generally supports the idea that sexually-selected traits are often condition dependent (Rowe & Houle, 1996). In the case of the deer, I can hypothesise that the combined effects of polygenicity, linkage, and pleiotropy may couple antler variation with individual condition, which itself is likely to be driven by the effects of many loci of small effect throughout the genome affecting a wealth of fitness-related traits.

3.6 Conclusions

Considering the findings of both Chapter 2 and Chapter 3, I show that antler morphology is heritable, has a polygenic genetic architecture, and some degree of shared genetic architecture between different antler measures. A single region on linkage group 21 associated with antler PC9 is linked to candidate genes that potentially have an effect on antler morphology, but more work is required to validate this finding in this and other populations. Future work in this system will integrate knowledge of the genomic architecture of antler morphology with fitness measures to further dissect constraints on trait evolution within this population. Ultimately, these findings corroborate the expectation for a quantitative trait, such as multidimensional weaponry, to conform to a polygenic genetic architecture of many genes with small effects.

Chapter 4

Genomic signatures of sexual conflict in red deer (*Cervus elaphus)*.



Stag with hind and calf.

4.1 Summary

In species with two sexes, males and females largely share the same genome. However, due to differences in biology and life history, the sexes can have divergent fitness optima for the same traits, resulting in sexual conflict. If these traits share a genetic basis, this can result in intra-locus sexual conflict. The resolution of sexual conflict often leads to the development of dimorphisms in phenotypes between sexes, but there remains little empirical

research on the genetic architecture of ongoing conflict in the wild. Here, I explore evidence for sexual conflict on fitness in Rum red deer, a species that shows extensive sexual dimorphism, using the genomic dataset described in Chapters 2 and 3. First, I use evidence for sex-specific transmission distortion of SNP loci from parents to offspring as an indication of differential gametic and viability selection between the sexes. Specifically, sex-specific transmission distortion was said to occur if parents of one sex are more likely to transmit one allele over the other, or offspring of one sex are more likely to inherit one allele more often than the other if they survive until birth. Second, I conduct genetic mapping of sex-specific early life survival using a Bayesian mixed model approach. I find 136 unique SNPs that exhibit sex-limited transmission distortion, compatible with sexually divergent gametic and viability selection. This was more common in fathers, with a number of SNPs exhibiting signatures of balancing selection, a common feature of sexual conflict loci. Regions of sex-specific transmission distortion included oncogenes and genes involved in prenatal development. No SNPs were significantly associated with sex-specific postnatal survival to the second year, but males were on average $\sim$14% more likely to die before the age of two and early life survival was significantly heritable (h^2 = $\sim$0.1). In summary, my findings point toward a sexually dimorphic genetic architecture of prenatal viability across several loci in the red deer genome that possibly is the result of resolved sexual conflict, and I propose a number of candidate genes affecting prenatal gamete segregation and viability. Overall, this study gives valuable insight into the genetic architecture of unresolved sexual conflict in a species exhibiting strongly sexually-dimorphic traits and life history strategies.

4.2 Introduction

In species with separate sexes, males and females often follow very different life history strategies. In many species, females invest more heavily in parental care compared to males, starting at the level of gametes, since oocytes are much larger than spermatozoa (anisogamy) and contain metabolites and other molecules essential for early embryonic development (Hayward & Gillooly, 2011). This fundamental difference has a cascading effect on the development of sex roles such that females may provide more additional care in the form of gestation and increased incubation, as well as food provisioning (Clutton-Brock, 1991; Møller & Birkhead, 1993; Schwagmeyer *et al.*, 1999; Schärer *et al.*, 2012). Thus, as for a single reproductive event, female investment is typically larger than that of males and so females are often 'choosy' about whom they mate with, while males often try to attract and mate with a large number of females (Lehtonen *et al.*, 2016). This leads to deviating interests in female-male interactions, including courtship, fertilisation, and investment in offspring (Parker, 1979, 2006). This division can lead to divergent fitness optima in the two sexes and result in males and females drastically differing in a range of physical and behavioural traits. If an increase in fitness in one sex reduces the fitness of the other sex, selection will act in opposite directions, leading to sexual antagonism or conflict (Andersson, 1994; Rowe *et al.*, 2018).

Sexual conflict is often differentiated according to its underlying genetic architecture: inter-locus or intra-locus sexual conflict. **Inter-locus sexual conflict** typically involves separate, sex-specific traits that are mediated by separate loci in each sex (Chapman *et al.*, 2003). Conflict arises when a mutation at a given locus that is beneficial to one sex but harmful to the other increases in frequency. The disadvantaged sex will then favour a mutation at a separate locus that counteracts the harmful action of the mutation at the first locus, leading to an evolutionary arms race between loci (Rice & Holland,

1997). There are numerous examples of trait pairs involved in sexual conflict: in *Drosophila melanogaster*, seminal fluid increases female mortality, leading to the evolution of female resistance to multiple matings (Chapman *et al.*, 1995; Wigby & Chapman, 2005; Rice *et al.*, 2006). In the Dytiscinae subfamily of diving beetles, males have developed grasping adhesive discs that allow forced matings, whereas females in turn evolved a structured dorsum to impede attachment (Bergsten *et al.*, 2001). Genetic covariance between sexes that evolves as a consequence of inter-sexual conflict is an important driver of antagonistic co-evolution of male and female traits, much like runaway sexual selection of female preference and male secondary sexual traits (Chapman *et al.*, 2003; Lande, 1981; Rice *et al.*, 2006). Consequently, sexual conflict is thought to be of major importance for the emergence and diversification of sexually dimorphic traits, contributing to the great diversity of sexually selected traits and the evolution of highly exaggerated phenotypes (Mank, 2017b; Rowe *et al.*, 2018). Furthermore, sexual antagonism can both hinder and advance evolution: on the one hand, it acts to inhibit both sexes achieving their respective trait optima, leading to decreased population fitness and reduced adaptive potential (Rice, 1998; Bonduriansky & Rowe, 2005; Arbuthnott & Rundle, 2012; Chenoweth *et al.*, 2015); on the other hand, antagonistic selection can drive speciation and maintain a population's evolutionary capacity by increasing variance in fitness through remodelling of genetic architecture, balancing selection and shifting fitness optima (Gavrilets, 2014; Hall *et al.*, 2010; Rostant *et al.*, 2015; Rowe *et al.*, 2018).

Conversely, **intra-locus sexual conflict** occurs when the same genes control a specific trait in males and females, but the optimum trait values are different between the sexes (Rice, 1998; Parker, 2006; Chenoweth & McGuigan, 2010). Although evolving towards different variant optima of traits, the sexes still share the same species-specific genomic background. Therefore, the combination of shared genetic architecture and diverging fitness interests of the sexes should make intra-locus sexual conflict common throughout the

genome. Indeed, several studies have identified loci under sexually antagonistic selection, and increasingly genome-wide signals of intra-locus sexual conflict are documented (Cheng & Kirkpatrick, 2016; Dutoit *et al.*, 2018; Ruzicka *et al.*, 2019). Notable examples include: the locus *Vgll3* controls the age of sexual maturity in Atlantic salmon, with fitness in males and females associated with the alleles for earlier and later maturation, respectively (Barson *et al.*, 2015); the locus that induces resistance to DDT insecticides *Cyp6g1* in *Drosophila melanogaster* is female-beneficial, but reduces male fitness (Hawkes *et al.*, 2016). Thus, intra-locus sexual conflict can be an important factor in maintaining genetic polymorphism, preventing antagonistic loci from reaching their fitness optima within each sex. However, numerous mechanisms have been proposed and/or shown to resolve sexual conflict, through sex-specific gene expression (Mank, 2017b; Ellegren & Parsch, 2007) and sex-specific masking of gene effects through accumulation of antagonistic loci on sex chromosomes or linkage with sex-determination systems (Rice, 1984; Gibson *et al.*, 2002). In Lake Malawi cichlids for example, sexual conflict over the orange-blotch phenotype is resolved by tight linkage of the *OB* locus with a novel female sex determination locus, so that the *OB* haplotype that is beneficial to females but detrimental in males is only carried by females (Roberts *et al.*, 2009). Nevertheless, autosomal loci are often implicated in intra-locus sexual antagonism, in both partially resolved (Barson *et al.*, 2015) and ongoing conflict (Hawkes *et al.*, 2016). Yet, despite several well-characterised examples of major effect loci involved in sexual conflict (Roberts *et al.*, 2009; Rostant *et al.*, 2015; Barson *et al.*, 2015; Ducret *et al.*, 2016), in most cases the genetic architecture of sexual antagonism is unknown.

In recent years, the increased availability of genomic resources has allowed studies to examine genome-wide patterns of sexual conflict, rather than focussing on specific loci or traits, contributing to our understanding of the genetic architecture of sexual conflict. Such studies often characterise

sexually antagonistic selection using differential gene expression, which, as mentioned above, is a mechanism to resolve intra-locus sexual conflict and contributes to sexual dimorphism (Mank, 2017b; Ellegren & Parsch, 2007). Indeed, sex-biased gene expression is prevalent throughout the genome (Derome *et al.*, 2008; Yang *et al.*, 2006; Jiang & Machado, 2009) and has repeatedly been associated with sexual antagonism, particularly resolved conflict (Dutoit *et al.*, 2018; Wright *et al.*, 2018; Cheng & Kirkpatrick, 2016). Traditionally, sexual conflict has been investigated in the context of reproductive fitness, excluding any survival component, as sexual dimorphism, which is a common consequence of sexual conflict, is thought to arise primarily in response to selection on traits relating to reproductive fitness (Andersson, 1994). Classic studies of inter-locus sexual conflict such as those in *Drosophila melanogaster* and diving beetles (see above for details) (Chapman *et al.*, 1995; Bergsten *et al.*, 2001) for example exclusively examine sexual conflict over mating success. The same is often true for intra-locus sexual conflict, for example an early experimental study in *Drosophila melanogaster* assessed sexual antagonism over several components of male and female mating success and fecundity in relation to a haplotype that was previously allowed to segregate as male-limited (Rice, 1998); a more recent large-scale study in *Drosophila melanogaster* using GWAS identified a large number of loci associated with sexually antagonistic effects on male fertilisation and female fecundity (Ruzicka *et al.*, 2019).

However, sexually antagonistic viability selection during ontogeny (prenatal or postnatal) should result in sex-specific allele differences in adults, such that females are enriched for female beneficial alleles, while males should show an excess of male beneficial alleles (Rice & Chippindale, 2001). A common approach to detect sexual conflict over survival is to compare sex-specific values of F_{ST}, which is a determinant of allelic divergence. Some studies have complemented this by determining a positive Tajima's D statistic of sequence variation (Tajima, 1989) at differentiated markers, as it is an indicator of

balancing selection, another hallmark for sexual conflict (Mank, 2017a; Wright *et al.*, 2018). However, it should be noted that phenomena other than sexual conflict can generate a positive Tajima's D, such as heterozygote advantage and life-history trade-offs within one or across both sexes, frequency dependent selection, or temporal variation in fitness (Mank, 2017a; Hedrick, 2007). Several studies have now confirmed that there can be widespread genomic differentiation in allele frequencies between adult males and females, both in wild populations such as Gulf pipefish and Trinidadian guppies (Flanagan & Jones, 2017; Lin *et al.*, 2021), and in humans (Lucotte *et al.*, 2016). The implication of many genomic regions points towards a polygenic architecture of sexual conflict, be it over reproductive interests or survival, but more evidence is needed. In particular, there are few empirical examples that directly examine this issue at the individual level, rather than the population level, and that are supported by direct associations with male and female fitness. The few examples of analyses that combine reproductive fitness or early life/prenatal viability data with tests for sexual antagonism are exclusively in model organisms such as *Drosophila melanogaster* or in humans (Healey *et al.*, 2000; Friedrichs *et al.*, 2006; Ruzicka *et al.*, 2019). Fitness information can help disentangle the true signals of sex-specific or sexually antagonistic selection from confounding factors such as demographic and other selective processes that could produce similar genomic patterns of divergence and balancing selection (Huang *et al.*, 2014; Mokkonen *et al.*, 2011; Fijarczyk & Babik, 2015). Long-term studies of natural populations with access to this kind of data could play a crucial role in elucidating the genetic architecture of sexual antagonism and the extent to which it is resolved or ongoing.

One approach to investigate this question is to use long-term data from natural populations with genomic and pedigree information to characterise sexually antagonistic prenatal selection affecting gamete segregation and prenatal viability. Specifically, sex differences in prenatal selection can be

detected by investigating signals of transmission distortion of individual loci from parents to offspring, which can be defined as any deviation from 1:1 Mendelian segregation or inheritance of alleles. If this distortion favours alternate alleles depending on the sex, this can be a signal of sexual conflict. Sex differences in transmission distortion can be quantified at two levels: that of the parents (i.e., are there differences in allele transmission between parents of different sexes?) and that of the offspring (i.e., are there differences in the alleles inherited by offspring of different sexes?).

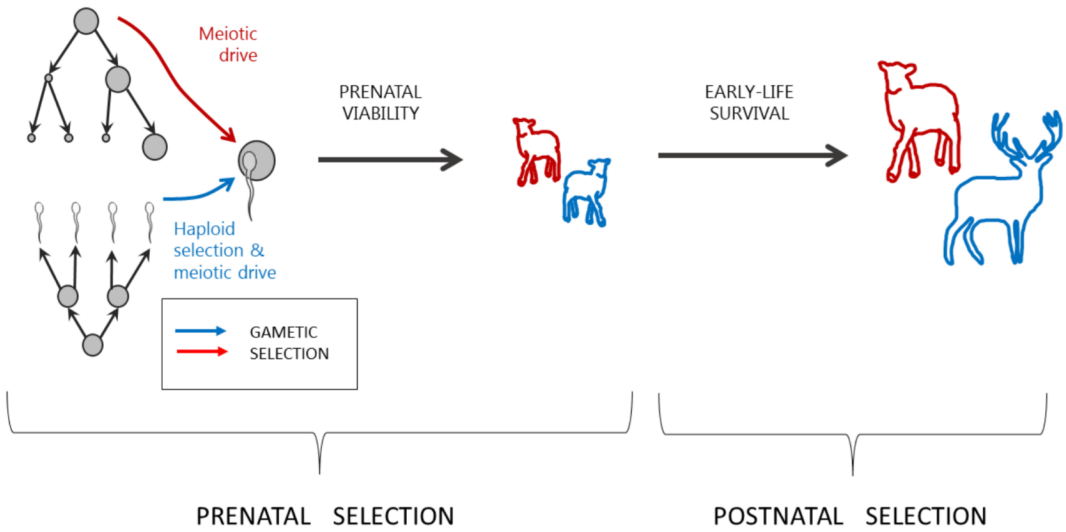


Figure 4.1: Flow chart illustrating drivers of transmission distortion via prenatal selection, and sex-specific allele frequencies in adults via postnatal selection in an example of a mammalian population. The colours represent sex, red for female and blue for male.

At the **parental level**, deviation from Mendelian expectation can be caused by segregation distortion mediated by meiotic drive or haploid selection of the gamete (see Figure 4.1). Segregation distortion means that there is an allelic bias in which alleles are passed on by the parents. Meiotic drive is typically associated with female gametogenesis, as meiosis during oogenesis is asymmetrical, i.e., one progenitor germ cell produces one large gamete and

three small non-functional polar bodies, instead of four functional gametes as in males. An allele subject to meiotic drive is more likely to remain in the mature oocyte, whereas the alternate allele will end up in the polar bodies, which are degraded and represent an evolutionary dead end (Pardo-Manuel De Villena & Sapienza, 2001; Lyttle, 1993). On the other hand, haploid selection may occur more often in males (e.g., in mammals, where the haploid phase is relatively short in females); in such cases, it can affect which sperm is able to fertilise the oocyte via sperm viability, motility, and competition (Immler, 2019). Thus, transmission bias of parental alleles can lead to deviation from expected Mendelian ratios, potentially driving sexual conflict if it acts in opposing directions in mothers and fathers.

At the **offspring level**, transmission distortion can be driven by differential prenatal viability of the zygote or later embryonic stages and postnatal viability at different life stages (Figure 4.1). Sexual antagonism over viability is present if different alleles are beneficial or detrimental to viability in either sex, which means that male and/or female offspring inherit a particular allele more frequently compared to the other sex (Healey *et al.*, 2000; Henckaerts *et al.*, 2010). Ongoing sexually divergent selection over survival should generate strong effects, as the mortality rates needed to generate significant allelic differences are quite high (Kasimatis *et al.*, 2017). Nonetheless, viability selection seems to generate much weaker sexual antagonism compared to conflict over reproductive success (Cox & Calsbeek, 2009), which is in line with sexual selection often exceeding viability selection in the wild (Kingsolver *et al.*, 2001). Overall, investigation of the prevalence and importance of sexual antagonism and conflict over viability requires the ability to characterise a large number of transmission events at large numbers of loci across populations, as well as the ability to directly link the extent of transmission distortion with different life-stages and fitness measures within individuals.

Red deer are a species with a polygynous mating system that show extensive

sexual dimorphism in their morphology and life-histories. Males compete for access to many mates and can sire multiple offspring per year, while females give birth to one calf per year at most. Male mating success is highly skewed, with few highly successful and many unsuccessful males. Due to intense inter-male competition, traits such as body size and fighting ability are important in stags, leading to significantly higher birth weight and faster growth rates in male calves (Clutton-Brock, 1985; Kruuk *et al.*, 1999). However, male calves also have lower fat storage compared to female calves and as a consequence often much lower first winter survival (Clutton-Brock, 1985; Gauzere *et al.*, 2020). As a result, male and female red deer have different strategies for investing in fitness, and thus optima for a number of traits, including those related to survival.

The red deer study population on the Isle of Rum, Scotland, has been subject to intense individual-based monitoring for over 40 years and offers long-term data on life histories as well as a multi-generational pedigree and extensive genomic marker information. This framework puts the study in the rare position of examining sexual conflict in the context of fitness effects, something that has only been done in humans or model lab species before (Friedrichs *et al.*, 2006; Healey *et al.*, 2000; Ruzicka *et al.*, 2019). Here, I quantify genomic signatures of ongoing sexual conflict over prenatal and early life survival in red deer. First, I test for signals of transmission distortion, deviation from Mendelian inheritance, to discover patterns of sexually dimorphic and antagonistic gametic and prenatal viability selection. I do this by testing if *parents of different sexes transmit* one allele more often than the other, and if *offspring of different sexes inherit* one allele more often than the other. I also use pedigree-based simulations of neutral inheritance patterns to establish a genome-wide significance threshold for distortion signals. Second, I use a Bayesian mixed model approach to detect signals of sexually antagonistic loci affecting early life survival of calves. Overall, I show that there are a substantial number of loci that exhibit transmission distortion in

line with gametic and viability selection in both males and females, but in a sex-specific and not an antagonistic manner. I also show that early life survival is likely polygenic and, although there are strong sex effects, I find no evidence that the genetic architecture is sexually dimorphic. This is valuable information, as this is one of the first studies to look at sex-specific genetic effects on early life survival in the wild. This evidence only supports sexual conflict acting at the prenatal stage in the deer, which is at least partially resolved via sex-specific genetic architecture. However, this sex-specific selection, potentially evolved from sexual conflict, likely contributes to maintaining genetic variance in this population.

4.3 Methods

4.3.1 Study system and genomic data-set

This study uses the same study system described in Chapters 2 and 3 of this thesis. The genomic dataset was produced using the same Illumina Cervine BeadChip as in these chapters, but modifications had to be made to the SNP clustering and genotyping method, meaning that this dataset was slightly different (see subsection on 'SNP clustering'). A SNP-based pedigree of 4,515 individuals was available for the population. After quality control below, a total of 3067 individuals were genotyped at 39,581 polymorphic SNP loci. A linkage map specific to the Rum population is available, with 38,083 SNPs assigned to linkage groups corresponding to the 33 deer autosomes and the X chromosome (Johnston *et al.*, 2017).

4.3.2 SNP clustering

Analyses that require comparisons of genotypes or allele frequencies between groups can be sensitive to genotyping error in cases where genotype scoring is carried out independently between the groups being compared. In the Rum deer population, this issue can arise when investigating differences in genotype or allele frequencies between generations that were genotyped on the Cervine SNP50 BeadChip at different times. This is due to variation in automatic genotype cluster positioning and subsequent genotype calling in the Illumina GenomeStudio software (Illumina Inc., San Diego, CA, USA). This issue is particularly acute for rare alleles when clustering is carried out using smaller numbers of individuals, making them more sensitive to genotyping error (Johnston *et al.*, 2013b). This has a downstream effect of exacerbating false signals at rarer loci where informative sample sizes will already be smaller (see subsection on “Quantifying transmission distortion”).

The original red deer genotyping experiment on the Cervine SNP50 BeadChip included 2,880 red deer genotyped from samples collected between 1982 and 2014 (referred to as Plates 1-29) at 51,388 SNP loci. Subsequent cohorts of red deer were run across annual batches of 96 deer (Plates 30 – 35), with each of these plates comprising a separate genotyping run. This has resulted in seven independent GenomeStudio projects. Genotype clustering and scoring was carried out in each project using the ‘Cluster all SNPs’ command after removing all individuals with a genotyping call rate of < 0.9. The GenomeStudio software creates a ’SNP table’, containing information on all parameter values used below; these were exported from each project individually. In the initial stages of data exploration, it was observed that genotype cluster positions could vary between projects, meaning that one could neither import standard cluster positions for each SNP across all plates, nor merge all projects into one. As a result, it was necessary to conduct quality control using the following criteria

for each SNP locus, in order to standardise and minimise error between runs:

1. The minor allele frequency (MAF) was calculated across all 7 projects; loci with MAF < 0.001 were discarded (N = 7,795 SNPs).

2. The variance in genotyping call rate was calculated across all 7 projects. Those with a variance > 0.01 were visually inspected, and subsequently all SNPs were removed (N = 175 SNPs; see example Figure S4.1).

3. The call rate was calculated across all 7 projects and those < 0.99 were removed (N = 2,670 SNPs).

4. If any locus on an individual plate had a call rate of < 0.99 (N = 2,055 SNPs), the clusters of all 7 projects were visually examined. Any cluster position mismatch that visually would result in different genotypes between plates were classified as a fail (e.g., for a locus with alleles A and B, a BB genotype obviously falling in a position that would be AB on the other plates). Only a small number of these SNPs were discarded (N = 57 SNPs; see example Figures S4.2, S4.3 and S4.4).

5. The variance in the following parameters was calculated across all seven plates: AA genotype frequency, AB genotype frequency, BB genotype frequency, AA genotype mean Theta (Θ), AB genotype mean Θ and BB genotype mean Θ (calculated as $\Theta = (2/\pi) \times \arctan(B/A)$, where A and B are the probe raw intensity values for each SNP allele from the Illumina run. Note that data generally cluster into the three genotypes separated by their Θ values.). In cases where the variance in any of these parameters was > 0.01, loci were visually inspected (N = 705 SNPs). Any cluster position mismatch that visually would result in different genotypes between plates were a fail, e.g., a BB genotype obviously falling in a position that would be AB on the other plates (N = 398 SNPs; see example Figures S4.5 and S4.6).

6. Genotype data from each project was exported to PLINK format from Illumina GenomeStudio and merged into a single file using the *–bmerge* function in PLINK v1.90b3.45 (Purcell *et al.*, 2007). Duplicate IDs were removed, and parental information was added using the *–update-parents* function. Mendelian errors per locus were determined using the *–mendel-multigen* function. As this stage was before pedigree construction, any SNP errors occurring in parent-offspring pairs with > 25 mismatches were disregarded. Any SNPs with more than 5 Mendelian errors remaining were removed from the dataset (N = 665 SNPs).

7. Loci were visually examined that had a high heterozygote excess (> 1.065) or deficit (< 0.93) or X chromosome excess (> 0.55), where the score was calculated as the ratio of the observed number of heterozygotes over the expected number of heterozygotes, assuming Hardy Weinberg equilibrium (N = 1,187 SNPs). These were visually inspected for ambiguous clustering (N = 56 SNPs removed; see example Figures S4.7 and S4.8).

8. Each plate contains two wells for a positive control, known as the “Golden Stag” (Deer ID = SIG09). The maximum mismatch across 63 positive controls was one SNP, so all 43 SNPs that had any mismatches were retained.

After this process, a total of 39,581 SNPs passed the SNP clustering quality control. It is acknowledged that visual inspection of SNPs can lead to difficulties in reproducibility and may be subject to the individual biases of researchers. The reclustering was carried out by a single person using R scripts to calculate variances and export the plots for inspection. All plots and their QC outcomes have been archived on the University of Edinburgh *rumdeer* DataStore archive for future reference.

4.3.3 Haplotype phasing

Transmission distortion analysis requires knowledge of the parental origin of each transmitted allele. This can be easily determined in cases where one parent is heterozygous and the other is homozygous at a given SNP. For example, an AB mother and a AA father will produce AA or AB offspring, indicating maternal transmission of A and B, respectively. However, this is less simplistic in cases where both parents are heterozygous (AB $\times$ AB) and the offspring are also AB. This scenario becomes more frequent when SNPs are common, potentially reducing sample sizes for determining transmission distortion. Therefore, to increase the number of informative parent-offspring trios, the parental phase, i.e., whether an allele was maternally or paternally inherited, was determined using a genome-wide approach. Parental phases of the 33 autosomal linkage group haplotypes were identified using AlphaPeel version 0.0.1 (Whalen *et al.*, 2018). AlphaPeel uses hybrid peeling, which determines genotypes and underlying haplotypes of an individual based on their own genotype information and the genotype information of their ancestors and descendants. Phasing was conducted using the red deer pedigree and genomic dataset, specifying the following criteria: five peeling cycles; multilocus peeling, which takes into account that nearby loci are more likely to be inherited jointly from a parental haplotype; and setting the calling threshold for phased genotypes to 0.95, meaning that the probability of a genotype had to be greater than 95% to be called. Phasing success was very high overall, with an individual mean phasing rate (i.e., the proportion of an individual's genotypes that were phased) of approximately 99%. The phasing rate improved with an individual's kin-depth, which is their depth in the pedigree set to 0 for a founder and otherwise defined as 1 + the maximum depth of either parent. Therefore, the 'founding' individuals born at the beginning of the study had lower phasing rates than the animals in subsequent birth cohorts (see Figure 4.2). There was no substantial difference in phasing rate or kin-depth between males and females (see

Figure S4.9).

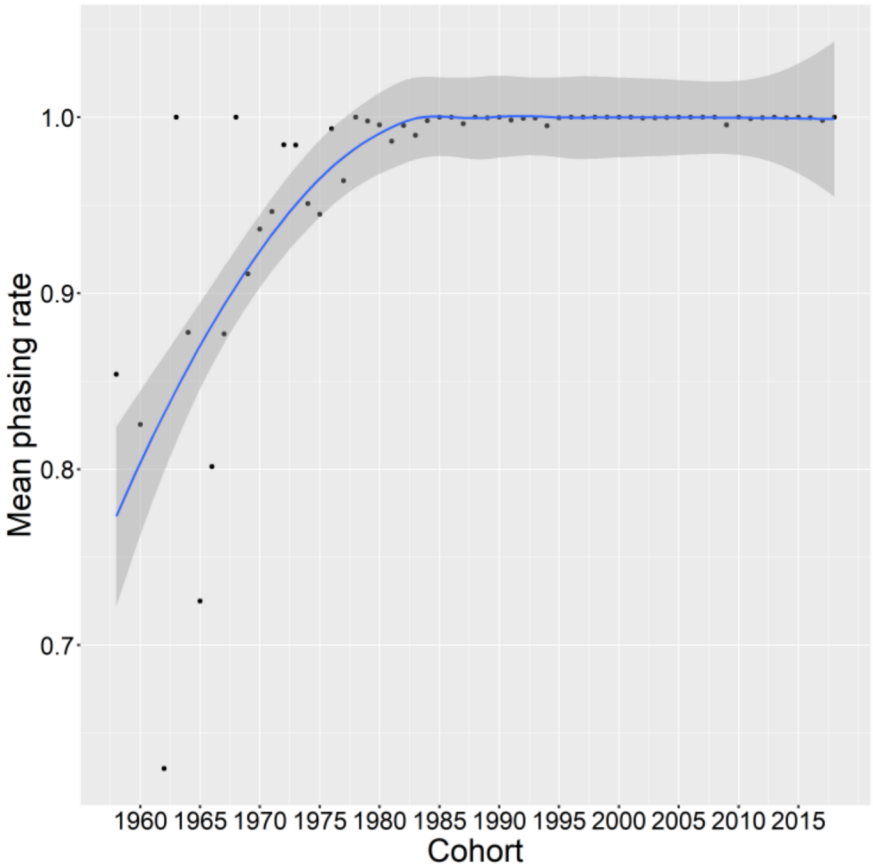


Figure 4.2: Mean AlphaPeel phasing rate (i.e., proportion of genome phased) against birth cohorts. The curve represent the best fitting relationship with the 95% confidence interval indicated by the shaded grey area.

4.3.4 Quantifying transmission distortion

In order to identify signals of prenatal viability and gametic selection, transmission distortion was quantified by tracing the inheritance of alleles transmitted from parents to offspring using the phased genomic and pedigree data. The minor and major alleles of each SNP were randomly assigned the labels “A” and “B” using a binomial sampling process. At least one parent had to be heterozygous to detect transmission distortion at a particular locus, as AA $\times$ BB pairings are not informative since there is no alternative allele for

either parent to transmit or the offspring to inherit, making it impossible to identify any deviation from Mendelian expectations. As all genotypes had known phases, the inheritance, or transmission of each allele from each heterozygous parent to their offspring could be determined, including cases where AB $\times$ AB parents had AB offspring. The total counts of A and B allelic transmission were tallied across all parent-offspring pairs and tested for deviation from the expected counts under Mendelian inheritance, i.e., equal numbers of each of the two variants, using Fisher's exact tests for each SNP. Unless otherwise stated, all following statistical analyses were conducted in R version 4.0 (R Core team, 2020).

Allele counts were tallied in two different ways. First, allele counts were determined according to the sex of the offspring that inherited the allele, not accounting for parent sex. Second, alleles were counted according to the sex of the parent who transmitted the allele, not accounting for the sex of the offspring. For each SNP, the p-value was calculated based on Fisher's exact test from a $2x2$ contingency table using the following formula:

$$p = ((A+E)!(B+E)!(A+B)!(2E)!)/A!B!(E!)^2N! \tag{4.1}$$

where A and B refer to the allele counts of the two variants at each SNP, E refers to the allele count of each variant under Mendelian inheritance and N is the sum of contingency table. Distortion effect sizes were quantified based on the following calculation: $(A-E)/E$; effect sizes represent the proportion of alleles that deviate from Mendelian segregation (in the case of parents) or inheritance (in the case of offspring).

A genome-wide significance threshold for the obtained per SNP p-values was determined based on a null-distribution of p-values under Mendelian inheritance expectations. The null-distribution was generated by first simulating phased haplotypes of all individuals based on the same pedigree structure, but without transmission distortion or selection - only Mendelian

inheritance and recombination (see section 'Gene-drop simulations' below for full details). The p-values for signals of transmission distortion were calculated using Fisher's exact test, as in equation 4.1. These simulations were conducted 1,000 times, with the p-values produced this way representing the null-distribution. The genome-wide significance threshold for the empirical data set was determined by selecting the 0.1% quantile of the null-distribution of the p value, which was ≈ 0.02. Therefore, any p-value in the empirical data set > 0.02 shows significant signs of transmission distortion with a false positive rate of 0.1%.

4.3.5 Gene-drop simulations

As outlined above, only informative parent-offspring trios can be used to detect transmission distortion, which means that the sample sizes for these crosses can differ greatly across loci depending on allele frequencies. For example, a locus with a small minor allele frequency (MAF) will have more homozygotes and fewer informative crosses, increasing this likelihood of spurious associations due to smaller sample sizes. Thus, the significance threshold for distortion needs to take this by-locus heterogeneity into account by employing a tailored simulation approach.

To determine significance thresholds for the transmission distortion analysis, the distributions of p-values under the null expectation of Mendelian inheritance with recombination and no transmission distortion were generated using a gene-drop simulation approach. For reasons already described, phased haplotypes are more useful for detecting transmission distortion, and thus for the simulation, individual phased haplotypes were generated using the R package *GeneDrop* version 0.1.0 (Hunter, 2020), which uses a known pedigree structure and founder haplotypes to simulate individual haplotypes of descendants with recombination. GeneDrop requires the pedigree to be

complete; therefore, unknown parents were assigned 'dummy' IDs, which were given empty haplotypes treated as NAs for the downstream analysis. As phasing rates improved markedly for birth cohorts after 1980 (see Figure 4.2), only individuals born after 1980 were used as their genomes could be phased accurately. Parents of individuals that were born before 1980 were removed from the pedigree, and founders were defined as individuals with unknown parents. The haplotypes of founders with no available genotype information were simulated based on the allele frequencies of the individual's birth cohort. Recombination frequencies between adjacent loci in centiMorgans were taken from the red deer linkage map (Johnston *et al.*, 2017) and originally calculated using the Kosambi mapping function (Kosambi, 1943). Whole-genome gene-drop simulations were conducted 1000 times, and each SNP locus in each simulated data set was tested for transmission distortion as described above. The genome-wide significance threshold was determined from the 0.1% quantile of the distribution of p-values from all simulations (see section "Quantifying transmission distortion" above for details).

4.3.6 Imputation of missing genotypes

The survival analysis in the following subsection does not allow for missing values in the predictor variables. Therefore, any missing genotypes in the genomic dataset were imputed using AlphaImpute version 1.98 (Hickey *et al.*, 2012). AlphaImpute needs a group of individuals with a high genotyping rate for imputation, and the threshold for non-missing SNPs for an animal to be included in this high-density group was set to 50%. Before imputation, an individual's haplotype needs to be phased, which is done using non-overlapping genomic regions called cores and trailing segments on each side of these cores called tails, which can overlap. The threshold for allowed disagreement between cores of two surrogate parents was set to 0.5, five phasing iterations were run with the size of cores including tails set to 100,

200, 300, 400 and 500 and the respective core sizes set to 50, 150, 250, 350,and 450.

4.3.7 Bayesian mixed model of early life survival

To identify sexually antagonistic SNPs that affect survival after birth, a Bayesian general linear mixed model approach was adopted using the R package *BGLR* version 1.0.8 (Pérez & De Los Campos, 2014). The *BGLR* package implements whole genome regression for trait analysis and prediction and offers a wide range of flexible Bayesian regression models which allow regression coefficients to have different types of prior densities. This framework facilitates the detection of small and/or non-zero effect loci underlying complex traits by fitting all SNP effects in one model, avoiding issues related to multiple testing. Most of the mortality in the red deer study population occurs within the first two years of life, and previous work has established substantial differences in early life survival between the sexes (Clutton-Brock *et al.*, 1982). Therefore, survival to the second year of life was used as a binomial response variable, coded 1 if the animal survived and 0 if it died. Only individuals that were identified as core animals of the study area by the field researchers were used, which means they are frequently sighted and thus have reliable lifetime records to avoid coding animals that disappeared as having died. Core animal status is recorded in the Rum red deer database. Calves that died before the first winter were also excluded, as such deaths are often accidental and likely not due to inherent traits of the calf (Gauzere *et al.*, 2020). All models employed the following formula:

$$y = X\beta + Z_1 a + Z_r u_r + e \tag{4.2}$$

where y is a vector of survival to the second year, X is an incidence matrix

relating individual measures to the vector of fixed effects β; Z_1 and Z_r are incidence matrices relating individual measures to additive genetic and other random effects respectively; a is a vector of SNP effects or the genomic covariance matrix; u_r is a vector of additional random effects; and e is a vector of residual effects. Models were run both with sex as a fixed effect and as separate models for each sex.

Model selection was carried out by first creating a reference model fitting all SNP effects as a random effect and sex as a fixed effect. Then, the model was adapted by fitting different genetic priors for the SNP effects and/or sequentially adding additional random effects. Models were selected based on the prediction accuracy of their genomic values, which means that if an alteration to the model significantly improved the ability of an individual's genotypes to predict its phenotype, the change to the model was kept. Genomic prediction precision was calculated as the area under the curve (AUC) of a Receiver Operating Characteristic (ROC) curve, which has been proven to be an effective method for assessing the genomic prediction accuracy of binary traits, such as survival (Wray *et al.*, 2010). AUC calculation and significance testing were performed in R using the *pROC* package version 1.17.0.1 (Robin *et al.*, 2011). Different models were compared by testing if their AUCs were significantly different from each other using the *roc.test* function, and a model was selected if the resulting p-value was statistically significant (alpha = 0.05). Genetic priors that were tested were the BayesA model (Meuwissen *et al.*, 2001), which assumes that all markers have non-zero effects and uses a scaled-t distribution, and the BayesB model (Meuwissen *et al.*, 2001) which implements a mixture of a distribution with the highest density at 0 (i.e., SNPs with zero effect) and a scaled-t distribution (i.e., SNPs with non-zero effect).

Additional random effects that were tested included birth year, the year of the second winter, and mother ID to account for any maternal effects. The model

used for the non-genetic random effects was the Bayesian Ridge Regression (BRR), which uses a Gaussian prior and induces shrinkage of all effects to a similar extent. Additionally, a model fitting the principal components of a genomic covariance matrix was run to determine the heritability of early life survival, employing a Bayesian Reproducing Kernel Hilbert Spaces Regressions (RKHS) model, which uses phenotypic and genomic data simultaneously. The RKHS model also included sex as a fixed effect and any random effects selected in the model selection process described above. All models were run with the default parameters except for the number of iterations, which was set to 120,000 with a burn-in of 20,000; the thinning interval was kept the default value of 5.

4.3.8 Signatures of selection

All SNPs that showed significant sexually-dimorphic effects on prenatal or early life survival were also tested for signals of selection using the simulated allele frequencies from the gene-drop analysis. Three measures were considered: to test for directional selection, I determined 1) the total change in minor allele frequency (MAF) over the simulated time period (1980 to 2018) and 2) the slope of the regression of MAF on birth cohort (i.e., year); and to test for balancing selection, I determined 3) the cumulative absolute distance of each SNP trajectory, calculated as the sum of the absolute MAF change between each consecutive birth cohort. The observed value in the true data was then compared with the distribution of values from the gene-drop simulated data; empirical p-values were calculated per SNP as the proportion of gene-drop iterations that produced a value equal to or more extreme than the observed value. The empirical p-values were then adjusted for multiple testing using Benjamini and Hochberg's false discovery rate (FDR) (Benjamini & Hochberg, 1995). Iterations where the MAF reached values of 1 in any cohort, which occurred at random, likely due to small sample sizes, were

excluded from these calculations. The total number of iterations used for the empirical calculation of the p-value was based on the number of iterations after this exclusion.

4.3.9 Candidate locus annotation and functional enrichment analysis

At present, there is no annotated deer genome sequence available; as a result, candidate loci and functional enrichment in this chapter was conducted based on homology with the cattle genome version ARS_UCD1.2. Note that this is a newer version of the genome than used in Chapters 2 and 3, meaning that the new SNP positions were unknown. I determined their homologous positions by aligning the flanking sequence for each deer SNP against the cattle genome using the *blastn* command with default parameters in BLAST version 2.10.1 (Camacho *et al.*, 2009). Each SNP had 150 base pairs of sequence flanking each side of the polymorphism. The top positional hit for each SNP was retained where there were $>$ 85% matching bases and $>$ 100 bases matched (NB. disparity between these two measures can arise due to flanking sequences harbouring a large number of N values).

Cattle gene IDs and their associated gene ontology (GO annotations) within 70kbp (the average distance between SNP markers) up- and downstream of all SNP markers with significant transmission distortion or early life survival effects were extracted from Ensembl (gene build ID ARS_UCD1.2) using the function *getBM* in the R package biomaRt v2.44.4 (Durinck *et al.*, 2005, 2009). Gene orthologues in humans (*Homo sapiens*), mouse (*Mus musculus*) and associated GO terms were also extracted using the function *getLDS*. Next, a gene ontology (GO) enrichment analysis was conducted separately for each set of SNPs significant either for transmission distortion in daughters, sons, mothers or fathers, or significant in the Bayesian regression analysis for

early life survival, using the software Gowinda v1.12 (Kofler & Schlötterer, 2012). An annotated genome file of *Bos taurus* version ARS_UCD1.2 in .gtf format was used as a reference file. The information relating GO terms to genes obtained from biomaRt was used to create a mapping file. Gowinda was used in both *snp* mode, which assumes that SNPs are in linkage equilibrium, and *gene* mode, which assumes that SNPs are in complete linkage disequilibrium. Gowinda determines the significance of over-representation for any GO term by simulating an empirical null distribution of randomly drawn genes and their GO terms based on the number of candidate SNPs and all markers used in the analysis. The number of simulations was set to 100,000. The generated p-values are then corrected for multiple testing using FDR calculated from the empirical null distribution.

Data Statement

The raw data for this study are not yet available on a public repository, but is being held on the University of Edinburgh *rumdeer* DataStore archive. All scripts for the analysis will be provided in a repository at `https://github.com/Lucy-Peters/Red\_deer\_genomics\_of\_sexual\_conflict`. Due to their size, some of the Supplementary Tables have only been provided with their caption information in the printed version of this thesis; they have been uploaded to a Dryad repository, the DOI is not live yet, but they can be accessed privately using the following link: `https://datadryad.org/stash/share/oszYV8OAptfbK1U3hR6BaQAzb\_eYlopIIeU5qeCJmRQ`.

4.4 Results

4.4.1 Signatures of transmission distortion

The gene-drop simulations of neutral Mendelian inheritance with recombination using the red deer pedigree structure showed that only 0.01% of Fisher's exact test p-values for any category tested (i.e., transmission distortion by offspring sex or parent sex) was smaller than the genome-wide significance threshold of 0.02. Based on this cut-off, 136 unique SNPs showed transmission distortion at the level of either the offspring or parents across both sexes. There were 21 SNPs in mothers and 43 SNPs in fathers that showed distortion in transmission from parents, and 29 SNPs in daughters and 47 SNPs in sons that showed distortion in inheritance (see Figure 4.3 and Table S4.1 for full results and Table S4.2 for significant SNPs). Of all significantly distorted loci, four SNPs showed transmission distortion in both a parent and an offspring. In these cases and as expected, the direction of distortion was the same, i.e., the same allele was over- or under-represented in both the parent and offspring (see Table 4.1).

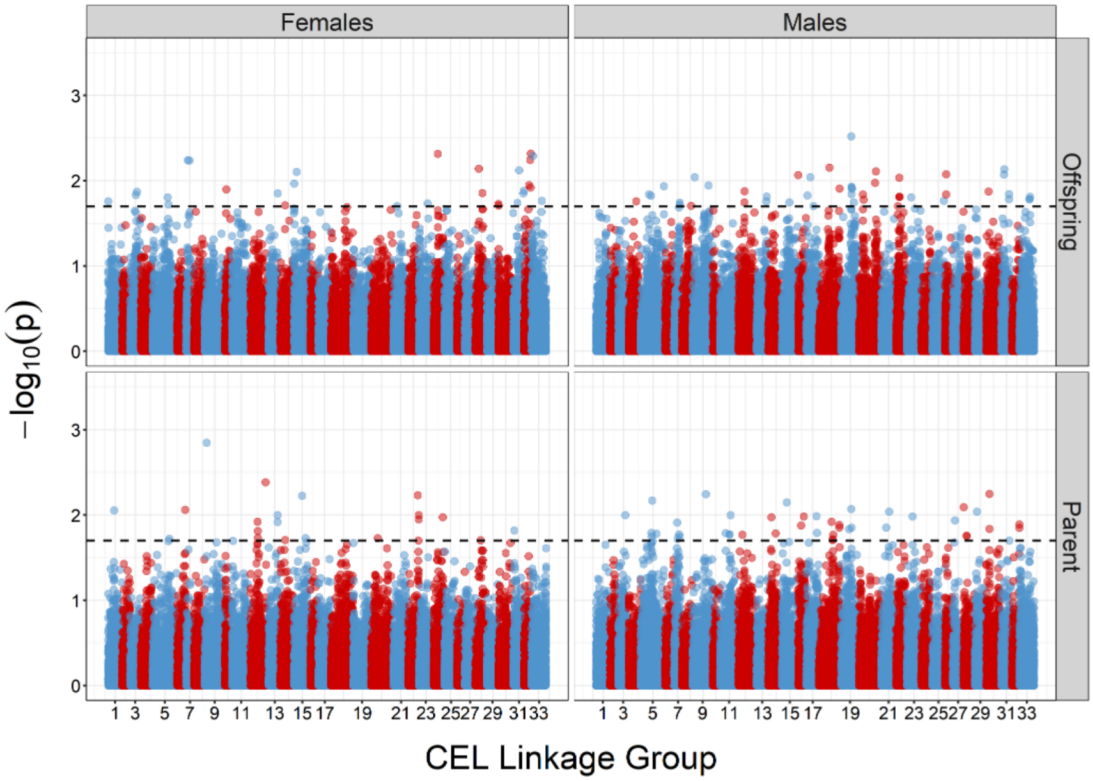


Figure 4.3: Manhattan plot showing the results of the transmission distortion analysis. The figure is divided by sex by the vertical panels and by status - parent or offspring - by the horizontal panel, so that the upper left panel shows the results for female offspring and so forth. The X axis shows the red deer linkage groups, and every dot represents a SNP ordered along those linkage groups. The Y-axis shows the p-values from the Fisher's exact test for transmission distortion on a negative log_10 scale. The colours of each dot represent the linkage group the SNP belongs to, which alternate in colour along the X-axis. The dotted line represents the genome wide significance threshold (p = 0.02), obtained from the gene-drop simulations.

Table 4.1: SNPs showing significant distortion in both a parent and offspring. Status indicates if allele counts refer to transmission from a parent or inheritance in an offspring. A and B show the counts for the two alleles at each SNP locus. The minor allele indicates which label (A or B) was assigned to the minor allele for each locus. The p-value refers to the p-value obtained from Fisher's exact test to detect TD. CEL.LG refers to the deer linkage group, and MAF refers to the minor allele frequency of each locus in the population.

SNP name	sex	status	A	B	minor allele	p-value	CEL.LG	MAF
cela1_red_1_78437038	male	parent	355	261	B	0.0085	19	0.28
	male	offspring	402	291	-	0.0030	-	-
cela1_red_16_32005570	female	parent	353	269	A	0.0198	14	0.20
	female	offspring	335	254	-	0.0194	-	-
cela1_red_23_18947043	male	parent	270	193	B	0.0122	7	0.19
	female	offspring	314	223	-	0.0058	-	-
cela1_red_7_15671293	female	parent	318	443	B	0.0014	9	0.29
	male	offspring	319	421	-	0.0091	-	-

All significant distortions were sex-specific within their generational class, so that distorted SNPs in male offspring did not show distortion in female offspring and vice versa. The same was true for distorted SNPs in mothers and fathers. Deviations from Mendelian inheritance were quantified as a proportion of the difference between expected and observed allele counts of allele "A" over the expected allele counts (see Figure 4.4 and Table S4.1 for full results). The distribution of deviations of significant SNPs showed some loci with quite large distortion effects, especially for segregation distortion on the parental level, which were up to ~34% in fathers. The minimum distortion effects were generally similar in size to the mean, with the smallest deviation of ~0.11% in male offspring, which is a moderate effect for viability selection (see Discussion). Overall magnitudes of distortion were very similar across sexes, offspring, and parents (Table 4.2).

Table 4.2: Summary of magnitudes of transmission distortion of significant SNPs only, quantified as deviations from expected allele counts under Mendelian inheritance. All values are based on absolute numbers, since the sign of the distortion effect is only dependent on the reference allele and is not relevant to the magnitude of the distortion. Status indicates whether the values refer to transmission distortion from a parent or to an offspring, while the sex is given in a separate column. The minimum and maximum values represent the extremes of the distribution. The mean is reported with the standard deviation (SD) in parentheses.

Status	Sex	Minimum	Maximum	Mean (SD)
parent	female	0.1085	0.2598	0.1379 (0.0342)
offspring	female	0.1096	0.2151	0.1505 (0.0289)
parent	male	0.1142	0.2583	0.1507 (0.0382)
offspring	male	0.1078	0.2353	0.1393 (0.0299)

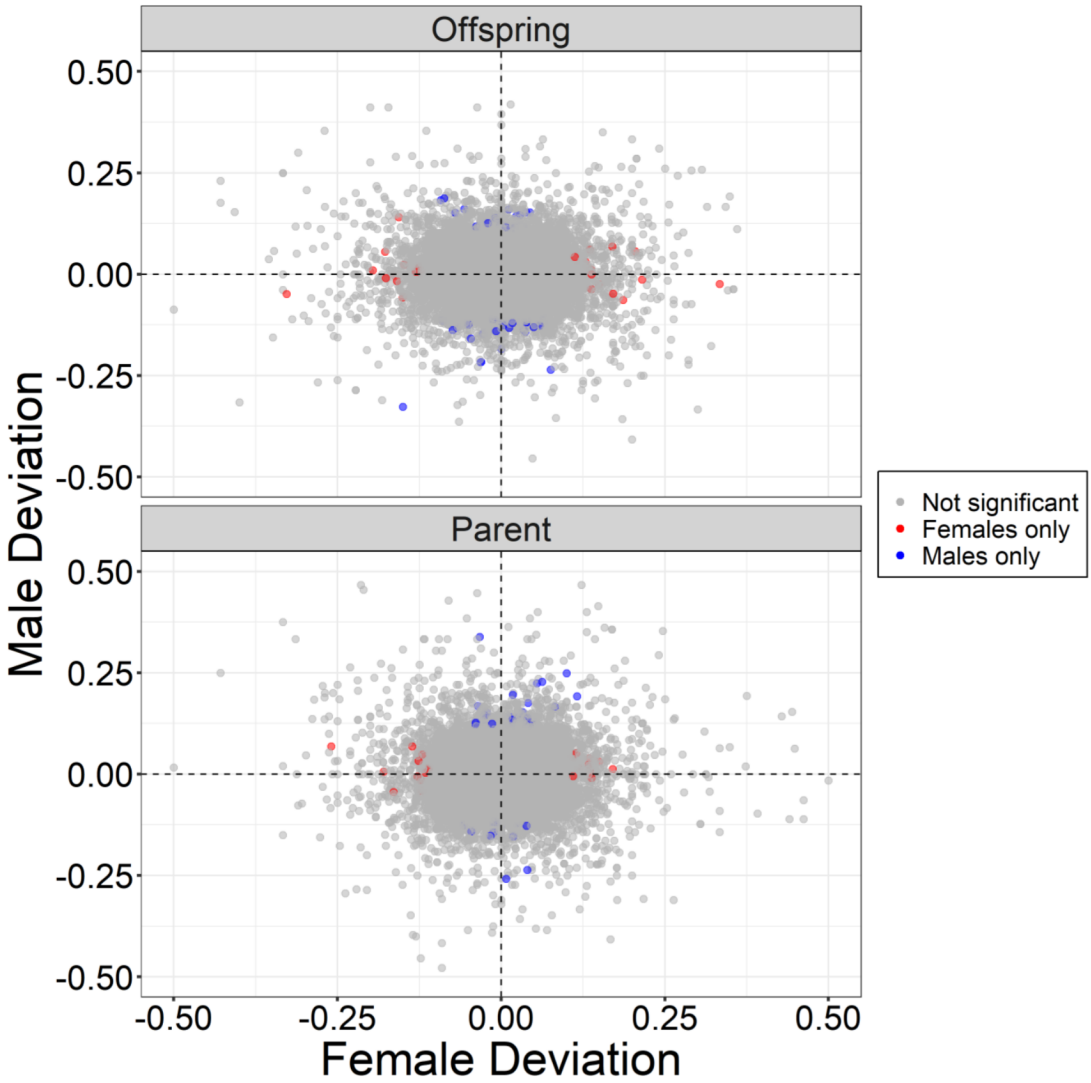


Figure 4.4: Deviation plot showing the results of the transmission distortion analysis. Deviations represent proportions of deviation in allele count of the "A" allele over the expected allele count under Mendelian inheritance

4.4.2 Selection signals of sexual conflict loci

All SNP markers that were identified to exhibit significant sexual antagonism were also tested for signals of selection using empirical p-values based on allele frequencies in the gene-drop simulations. Of all the 136 unique SNPs that were found to show significant transmission distortion patterns, none

were associated with significant signals for directional selection after FDR correction for multiple testing (see Table S4.3 for complete results). However, seven loci were associated with patterns of significant balancing selection (see Table 4.3 and Figure 4.5). Of these SNPs, five were loci with transmission distortion patterns in fathers, while two were SNPs with transmission distortion signals in male offspring. None of the SNPs identified as transmission distortion loci in mothers or female offspring showed significant signals of selection.

Table 4.3: All SNPs exhibiting significant signals of selection based on gene-drop simulations. All SNPs were under significant balancing but none under directional selection and all SNPs were significant in the transmission distortion analysis. SNP Name refers to the SNP locus ID, Status indicates if SNPs showed transmission distortion from the parents or to the offspring and sex refers to the sex of the parent or offspring. Total Dist. indicates the absolute cumulative change in the minor allele frequency (MAF) or the distance travelled by the SNP marker during the time period used in the gene-drop simulation (1980 to 2018) and p-balance is the associated empirical p-value. Slope refers to the slope of the regression of MAF on birth cohorts, and p-slope is the respective empirical p-value. All p-values were corrected for multiple testing using FDR.

SNP Name	Status	Sex	Total Dist.	p-balance	Slope	p-slope
cela1_red_17_15881624	parent	male	1.7997	0.0417	0.0021	0.4561
cela1_red_17_11825067	parent	male	1.6240	0.0417	0.0001	0.2882
cela1_red_23_33896016	offspring	male	1.7110	0.0417	0.0012	0.2085
cela1_red_11_59417377	parent	male	1.4586	0.0417	0.0006	0.7979
cela1_red_21_59324469	offspring	male	1.3994	0.0417	0.0002	0.3567
cela1_red_8_98827226	parent	male	1.3254	0.0417	0.0002	0.3044
cela1_red_13_33854335	parent	male	1.2861	0.0477	0.0009	0.5237

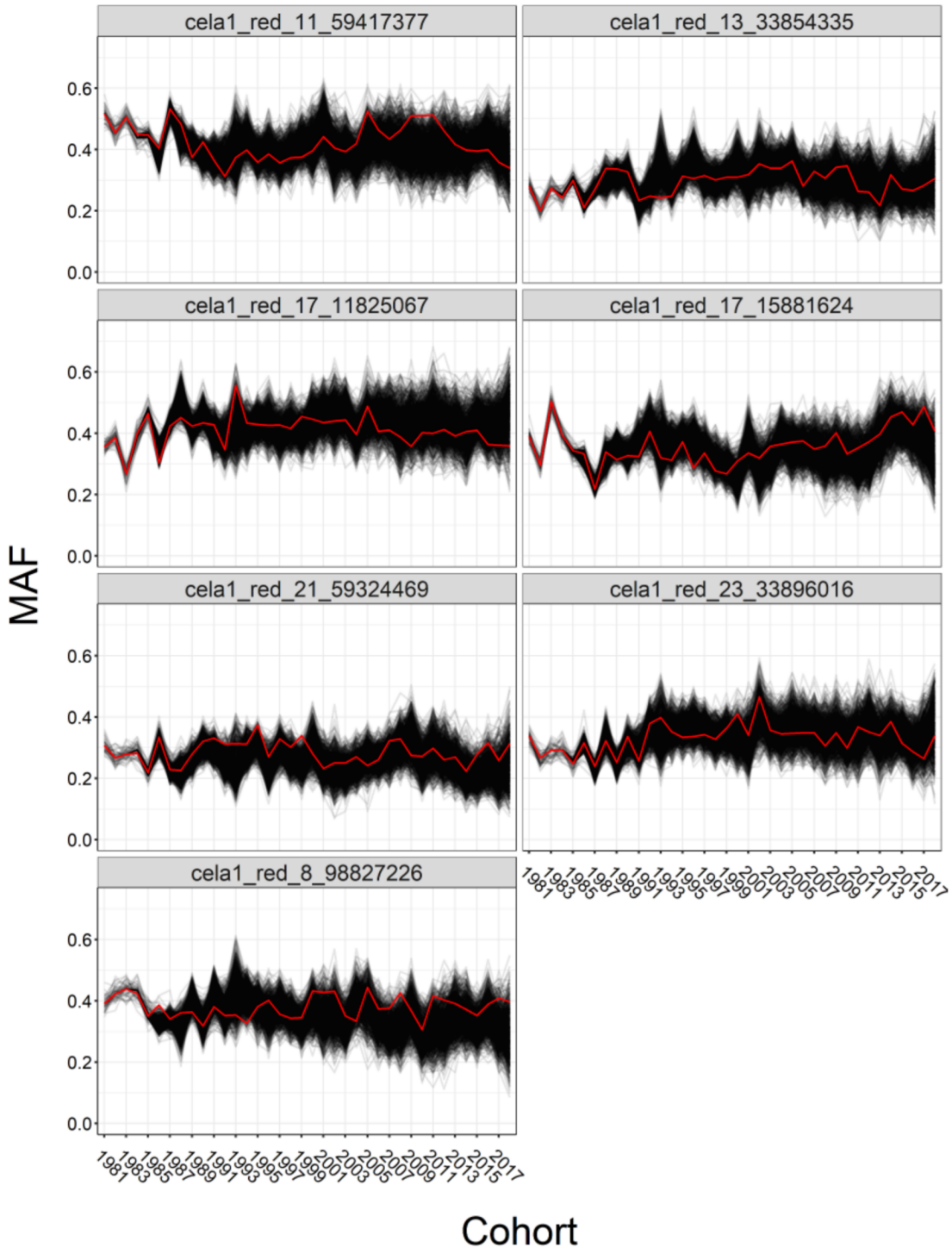


Figure 4.5: MAF change trajectories of SNPs, showing evidence of balancing selection. The red line shows the MAF change trend in the empirical dataset, while the black lines represent the trajectories generated in the gene-drop simulations. Cohort refers to a particular birth cohort, and the SNP name is given in the grey panel headings. All SNPs depicted also showed significant transmission distortion.

4.4.3 Genomic signals of sex differences in early life survival

After model selection using the area under the Receiver Operating Characteristic (ROC) curves, the best model was a BayesB model with additional random effects to account for the influence of common environment (birth year and year of second winter) and maternal effects (mother ID). The effect of sex was such that males had on average a ~14% lower probability of survival to the second year compared to females (effect = -0.1382, 95% CI = [-0.1917; -0.0776]). The heritability of early life survival was substantial at 0.1026 (95% CI = [0.0505; 0.1690], estimated by the RKHS model that included sex as a fixed effect and environmental and maternal random effects. The maternal effect on survival was also substantial and larger than the additive genetic effect, while the environmental effects had larger uncertainty associated with their estimates (see Table 4.4). None of the effect sizes of the SNP loci in the model containing sex as a fixed effect or in the two sex-specific models had 95% CIs that did not contain zero, and therefore no SNP could clearly be identified as having an effect on survival (see Figure 4.6 and Tables S4.4 and S4.5 for full results on SNP effects and predicted phenotypes, respectively).

Table 4.4: Variance components of the RKHS BGLR model for early life survival with sex fitted as a fixed effect. The variance component is indicated by variance with the 95% CI given in brackets, and the proportion of variance explained by each component is referred to by proportion V_{total}, again giving the 95% CIs in brackets.

variable	variance (95% CI)	proportion V_total (95% CI)
additive genetic	0.3054 (0.1367; 0.5405)	0.1026 (0.0505; 0.1690)
mother	0.3966 (0.2220; 0.6030)	0.1365 (0.0761; 0.2063)
birth year	0.6171 (0.0732; 1.6605)	0.2006 (0.0248; 0.4682)
second year	0.6136 (0.0739; 1.7031)	0.2103 (0.0246; 0.4751)

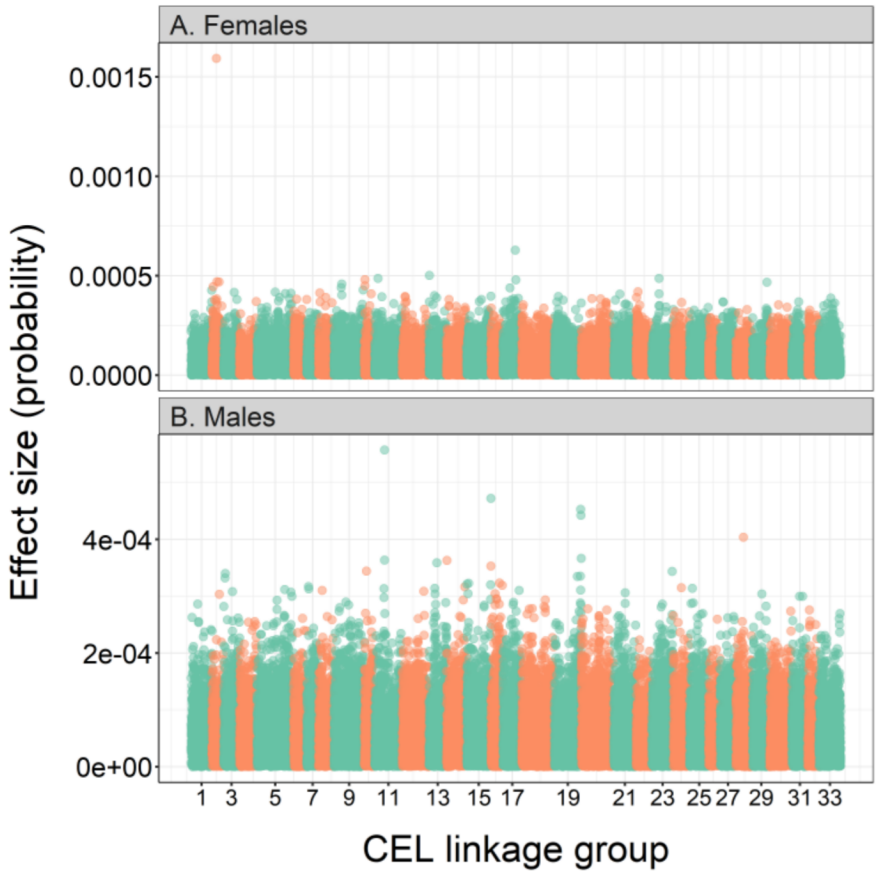


Figure 4.6: SNP effect sizes from sex-specific BGLR models for females (panel A) and males (panel B). Effect sizes are given as absolute change on a probability scale (obtained by evaluating the effect sizes, which are given on a probit scale by the BGLR model, at the standard normal probability density function).

4.4.4 Genomic regions associated with sexual antagonism

In homology with the cattle genome (version ARS_UCD1.2), several coding sequences were identified within regions that covered 70kbp each side of the 136 unique SNP markers that exhibited significant transmission distortion. For transmission to offspring, 76 unique genes were identified, consisting of 25 in females and 51 in males; for transmission from parents, 66 genes were identified, consisting of 24 in females and 42 in males. A total of 49 SNPs were found to be within the coding region of 42 distinct genes: in female

offspring, 11 genes had SNPs within coding regions; in male offspring, 13 genes; in mothers, 7 genes; and in fathers, 14 genes were identified. Note that these numbers include genes that were shared among offspring and parents; see Table S4.6 for complete information about all identified genes, including comprehensive GO terms and orthologues in humans and mice. Below is a more detailed description of some of the candidate genes associated with high-confidence SNPs. These SNPs included all markers with significant trans-generational transmission distortion effects and loci that showed signatures of balancing selection, see Table 4.5 for more detail on candidate SNPs and associated GO terms.

Of the four SNPs that showed significant transmission distortion in both a parent and offspring, three markers were in close proximity to several coding regions. The SNP cela1_red_16_32005570 which showed significant patterns in transmission distortion in both daughters and mothers was located within the gene *SMYD3* (SET And MYND Domain Containing 3), which is a methyltransferase involved in the development of muscle tissue (heart and skeletal muscle) (Codato *et al.*, 2019; Fujii *et al.*, 2011), but also in the carcinogenesis of multiple cancers, including breast cancer Sarris *et al.* (2016); Fenizia *et al.* (2019); Bottino *et al.* (2020). The marker cela1_red_7_15671293 for which significant transmission distortion was found in both mothers and sons was associated with several coding regions. One of these regions contained *PIN1* (peptidyl-prolyl isomerase), an important regulator of cell-cycle progression that has been implicated as a major oncogene in several cancers (Bao *et al.*, 2004; Yeh & Means, 2007; Dubiella *et al.*, 2021) but also in accelerated bone development (Nakashima *et al.*, 2021) (see Table 4.5 for additional candidate genes and relevant GO terms). The SNP cela1_red_23_18947043 associated with transmission distortion in both fathers and daughters was found in gene *RUNX2* (Runt-Related Transcription Factor 2), a transcription factor that is essential for skeletal morphogenesis and proper embryonic development (Komori &

Kishimoto, 1998; Schroeder *et al.*, 2005; Komori, 2011) and also involved in the formation of osteosarcoma (bone cancer) and other cancers (Martin *et al.*, 2011; Blyth *et al.*, 2010).

Of the SNPs that exhibited signatures of balancing selection, five were found in close proximity to candidate genes (see Table 4.5). One SNP, cela1_red_17_11825067, which showed segregation distortion in fathers, was associated with two genes of interest, *POU4F2* (POU class 4 homeobox 2) and *TTC29* (tetratricopeptide repeat domain 29). Mouse models show that *POU4F2* plays an essential role in normal cardiac development and function, as disruption of expression levels are associated with early embryonic lethality and heart defects as well as abnormal remodelling of cardiac tissue after stress or injury, especially in males (Budhram-Mahadeo *et al.*, 2021; Maskell *et al.*, 2017). *TTC29* is an evolutionarily conserved protein expressed in the testes that is required for sperm motility, and homozygous loss-of-function mutations have been shown to result in male sterility (Lorès *et al.*, 2019; Liu *et al.*, 2019). Two other SNP markers, cela1_red_17_15881624 and cela1_red_8_98827226 also showed distortion in fathers, and were associated with genes involved in tumour suppression, namely *RAD23B* (RAD23 Homolog B, Nucleotide Excision Repair Protein) and *INPP4B* (inositol polyphosphate 4-phosphatase type II). *RAD23B* is a tumour suppressor important for tumour progression and invasion especially in breast and lung cancer (Linge *et al.*, 2014; Han *et al.*, 2019) while *INPP4B* also plays a major role in tumour suppression in breast, ovary and prostate cancer, which at least in the case of prostate cancer is regulated by male hormones (Agoulnik *et al.*, 2011; Kofuji *et al.*, 2015; Hodgson *et al.*, 2011).

Analysis for GO enrichment using Gowinda identified five GO terms that were over-represented in any of the sets of SNPs with significant transmission distortion in either mothers, fathers, or male or female offspring. Four were identified in the SNP set that showed transmission distortion in mothers

(GO:0032970, GO:0003383, GO:0098592, GO:0035332) and one in the SNP set that showed transmission distortion in male offspring (GO:1905885; see Table S4.7 for full results). However, all significantly enriched GO terms were obtained using the *snp* mode, which assumes that all SNPs are in linkage equilibrium; this resulted in an overestimation of genes associated with the respective GO terms, when in reality there was only one unique gene with that specific GO term in the data set. As this was the case for all significantly enriched terms, any signal of enrichment is likely a statistical artefact. Thus, no over-arching gene function can be associated with the signals of transmission distortion I identified.

Table 4.5: Genes associated with loci exhibiting trans-generational transmission distortion signals (SNPs one to five in the top half of the table) and genes associated with loci showing patterns of transmission distortion and balancing selection (SNPs six to eleven in the bottom half of the table).

SNP name	external gene name	GO IDs	GO terms
cela1_red_16_32005570	*SMYD3*	GO:0006334; GO:0014904; GO:0034968	nucleosome assembly; myotube cell development; histone lysine methylation
cela1_red_23_18947043	*RUNX2*	GO:0002062; GO:0002063; GO:0045669; GO:0005654	skeletal system development; ossification; embryonic forelimb morphogenesis; embryonic cranial skeleton morphogenesis
cela1_red_7_15671293	*FBXL12*	GO:0000086; GO:0016567; GO:0043153	G2/M transition of mitotic cell cycle; protein ubiquitination; entrainment of circadian clock by photoperiod
cela1_red_7_15671293	*PIN1*	GO:0030512; GO:0032465; GO:0090263	negative regulation of TGF-β receptor signalling pathway; regulation of cytokinesis; upregulation of Wnt signalling pathway
cela1_red_7_15671293	*UBL5*	GO:0000398; GO:0006464; GO:1903955	mRNA splicing via spliceosome; cellular protein modification process; positive regulation of protein targeting to mitochondrion
cela1_red_13_33854335	*ARHGAP12*	GO:0002011; GO:0006911; GO:0043087	morphogenesis of epithelial sheet; phagocytosis; actin filament organization
cela1_red_17_11825067	*POU4F2*	GO:0031290; GO:0007507; GO:0030520; GO:0045672; GO:1904178	retinal ganglion cell axon guidance; heart development; intracellular estrogen receptor signalling pathway; positive regulation of osteoclast differentiation; negative regulation of adipose tissue development
cela1_red_17_11825067	*TTC29*	GO:0036126; GO:0003341; GO:0031514	sperm flagellum; cilium movement; motile cilium
cela1_red_17_15881624	*INPP4B*	GO:0016311; GO:0006661; GO:0016021	dephosphorylation; phosphatidylinositol biosynthetic process; integral component of membrane
cela1_red_21_59324469	*ASB2*	GO:0016567; GO:0006511; GO:0035914	protein ubiquitination; ubiquitin-dependent protein catabolic process; skeletal muscle cell differentiation
cela1_red_8_98827226	*RAD23B*	GO:0006281; GO:000628; GO:0003684	DNA repair; nucleotide-excision repair; damaged DNA binding

4.5 Discussion

In this chapter, I investigated genomic signatures of ongoing sexual conflict on prenatal and early life survival in the Rum red deer using two approaches. First, I tested for signals of transmission distortion from Mendelian inheritance to identify sexually dimorphic and antagonistic gametic and prenatal viability selection; I identified a number of loci which exhibit transmission distortion in line with gametic and viability selection in both males and females in a sex-specific manner, but did not identify loci with sexually-antagonistic effects. Second, I used a Bayesian mixed model approach to detect signals of sexually antagonistic selection at loci affecting early life survival of red deer calves. I determined that early life survival is likely to be polygenic, and despite a strong effect of sex on survival, I found no evidence of a sexually dimorphic genetic architecture. Here, I will discuss the trends, effect sizes and potential mechanisms driving sex-differences in prenatal viability and early-life survival, before considering the potential function of candidate regions significantly associated with these traits. I will then discuss the advances, limitations, and future directions for the Rum red deer study and for this topic across taxa more broadly.

4.5.1 Sex-differences in prenatal viability selection

I identified 136 loci exhibiting sexually dimorphic transmission distortion compatible with sexually divergent prenatal selection on Mendelian segregation of gametes and offspring viability. This included loci that exhibited sex-specific gametic selection from parents (21 distorted in mothers and 43 in fathers) and those that showed sex differences in prenatal viability (29 affecting female and 47 male offspring survival). No loci showed transmission distortion in both males and females within the same generation; of the four

loci showing trans-generational effects, transmission always favoured the same allele. In addition, seven of these loci showed signatures of balancing selection over the study period (five in fathers and two in sons). This suggests that there is no evidence for ongoing sexual antagonism or conflict for prenatal viability selection in the red deer population.

Transmission distortion is more common from fathers and to sons.

A higher number of distorted loci from fathers rather than mothers suggests that haploid selection on sperm is a major driver of transmission distortion. Some of the associated candidate genes, such as *TTC29*, which is associated with sperm motility, may directly influence sperm phenotype and competitiveness (Lorès *et al.*, 2019; Liu *et al.*, 2019). It should be noted that in mammals, spermatozoa can retain haplotype-specific transcriptomes and can share gene products through cytoplasmic bridges (Braun *et al.*, 1989); however, recent evidence suggests that this latter exchange can be restricted and there is positive selection on some haplotypes that results in allelic bias and transmission distortion (Bhutani *et al.*, 2021). As outlined in the introduction, loci distorted from mothers to offspring are likely to be driven by effects of meiotic drive, which I discuss in more detail below. In offspring, a higher number of loci showed distortion in sons compared to daughters. If the effects of distorted loci on viability are relatively similar, this suggests that prenatal viability selection may be stronger in males in this population. Evidence in humans suggests that male foetuses are generally more vulnerable to adverse environmental effects in the womb (Catalano & Bruckner, 2006; DiPietro & Voegtline, 2017), suggesting potential sex-by-environment interactions, which might be extended to a sex-by-environment-by-genotype dependency. In the red deer of Rum, male calves have lower survival probabilities early in life (Clutton-Brock 1985; Gauzere *et al.* 2020; see section on early life survival below); however, there is no evidence that the sex ratio is significantly skewed at birth.

Trans-generational patterns of transmission distortion.

Four loci showed significant distortion in both a parent and offspring, each representing different combinations (mother and daughter, mother and son, father and daughter and father and son; Table 4.1) with both generations showing bias toward the same allele at each locus. At present, the mechanisms driving these sex-specific effects are not clear, but some mechanisms can be proposed. In sperm, there is evidence from mouse models that the haplotype specific transcriptomic environment is maintained even after meiosis, and that this in turn influences sperm phenotype (Bhutani *et al.*, 2021), and could conceivably influence post-zygotic viability depending on whether it carries an X or Y chromosome. At the same time, differences in sperm phenotype could also influence its competitiveness, conditional on the type of sex chromosome that is present. Similarly, the oocyte environment is dependent on the maternal haplotype and important for embryonic development (Svoboda *et al.*, 2015; Kocabas *et al.*, 2006) and could influence offspring viability in a sex dependent manner after zygote formation.

Meiotic drive as a mechanism of transmission distortion.

When considering meiotic drive as a mechanism of transmission distortion more generally, it should be noted that examples of drive in females or male spermatozoa generally have very large effects (Lyttle, 1993; Pardo-Manuel De Villena & Sapienza, 2001), sometimes leading to 100% transmission of the driving allele (Finseth *et al.*, 2021). This dynamic would result in rapid fixation unless there is a substantial fitness cost associated with the driver. This could potentially make stable polymorphisms potentially difficult to detect (Traulsen & Reed, 2012); alternatively, meiotic drive may be transient in nature, and therefore the probability of sampling a population in a period of extreme drive may be relatively small. It is also possible that male and female drive may act in synergy, as there is evidence that in some systems sperm genotype can actually affect female drive, as in mammals the sperm and

oocyte fuse when the oocyte has not fully completed meiosis (Pardo-Manuel De Villena & Sapienza, 2001; Agulnik *et al.*, 1993). Thus, if a sperm is more competitive when carrying a particular allele that interacts with the oocyte in a way that drives the retaining of the same allele upon completion of meiosis, both the father and offspring could show evidence of transmission distortion. If this mechanism is also conditional on the sex chromosome present in the sperm, the offspring sex specific trans-generational transmission effects observed in fathers could be accounted for this way. However, without confirmatory experimental manipulation, it will be difficult to determine which of these mechanisms might be operating. One possibility to test for viability effects could be to obtain early or late stage embryos from culled red deer and genotype them for the candidate SNPs and genes found in this study. Testing meiotic drive or haploid selection is more difficult, but could be done in a laboratory environment in setups similar to in vitro fertilisation and related fertility studies. However, samples for this type of work are difficult to collect; whilst it could be possible on farmed deer, these populations may experience different selection pressures and will have a unique demographic history, meaning that not at all certain the same signals would be identified.

Effect sizes of transmission distortion.

Loci that showed significant transmission distortion from mothers and fathers and to daughters and sons had a mean deviation from Mendelian inheritance of between 13.8% and 15.1%, showing that both parent and offspring effects and the sex effects within were comparable in magnitude (Table 4.2 and Figure 4.4). For these loci, it can be expected that sex-specific selection is substantial in the Rum deer in order to produce detectable allelic differences (Mank, 2017a; Kasimatis *et al.*, 2017, 2019). Other studies have found similar effects for sex differences in segregation distortion in male and female offspring at a small number of loci (Friedrichs *et al.*, 2006; Ducret *et al.*, 2016). In humans, for example, the *DLG5* locus shows sex-specific distortion for the

Q variant, which is associated with Crohn's disease: In healthy individuals 4.9% fewer men carry the Q allele compared to men with the disease, while frequencies in healthy and ill women are almost identical (Friedrichs *et al.*, 2006). Evidence of allele distortion was also found in barn owls (*Tyto alba*), where the *MC1R* locus, which controls coat colour and has one female preferred and one male preferred variant, showed a deficit in the proportion of heterozygotes of 7.5% and 12.9% in males and females, respectively (Ducret *et al.*, 2016). Segregation distortion in parents with effects comparable to those found in red deer has also been documented by other studies. One such analysis in Zebra finches (*Taeniopygia guttata*) found segregation distortion at three linked loci on chromosome *Tgu2* due to meiotic drive in both parents, resulting in a deviation from Mendelian segregation of $\sim$6.7% (Knief *et al.*, 2015). Nevertheless, these studies have mostly focused on a limited number of loci; while other studies of transmission distortion at the whole genome level are in progress (for example, Elise Lucotte and colleagues have investigated this issue in humans: see `https://tinyurl.com/3ahsaxut`), few have been published to my knowledge. Generally, the effect sizes for segregation distortion found here and in other studies are moderate, while theory predicts that mortality loads should be extreme to result in detectable allelic differences between the sexes (Kasimatis *et al.*, 2019). However, such large effects have rarely been documented, with an exception being the *Puf60b* locus in Trinidadian guppies *Poecilia reticulata*, which was associated with 39% mortality load in females and 54% in males (Lin *et al.*, 2021). Nonetheless, as segregation distortion analyses here and above do not rely on allelic frequency differences between the sexes, but rather track inheritance patterns sometimes over multiple generations, they may have more power to detect loci with smaller effects on viability that would not be purged from the population as quickly as loci with large effects on viability.

Signatures of selection in transmission-distorted loci.

A subset of seven markers with transmission distortion *from fathers* and *to male offspring* displayed signatures of balancing selection based on GeneDrop simulations of neutral Mendelian inheritance patterns. Signatures of balancing selection can be expected under ongoing sexual antagonism because opposing selection pressures between the sexes can maintain genetic variation (Wright *et al.*, 2018; Mank, 2017a; Wright *et al.*, 2019; Lin *et al.*, 2021). However, this is incompatible with the sex-limited effects I find here, which are more compliant with sex-specific genetic architecture that could arise from the resolution of sexual conflict, and should result in reduced genetic diversity (Wright *et al.*, 2018, 2019). One explanation is that the pattern of allele frequency maintenance observed in the GeneDrop simulations might actually stem from relaxed purifying selection rather than balancing selection. This is in line with expectations that sex-limited genes are under relaxed selective constraint, especially male-specific ones such as sperm competition genes (Dapper & Wade, 2016; Gershoni & Pietrokovski, 2014); indeed, one locus where I observed transmission distortion from fathers is associated with the sperm motility gene *TTC29* (see the section on candidate genes below). Another seemingly paradoxical finding is that none of the 136 loci that showed sex-specific transmission distortion were associated with signals of directional selection. If an allele is favoured in one sex and doesn't experience antagonistic selection in the opposite sex, one might expect these loci to be on a trajectory towards fixation (Falconer & Mackay, 1996). One possibility is that as distortion effects, and thus mortality loads, are not extreme for any of these loci, selection might be weak and counteracted by genetic drift. In the same avenue, per-cohort sample sizes in the GeneDrop simulations may be too small to distinguish signals of weak directional selection from allelic trajectories produced by genetic drift. These small sample sizes might also have led to the minor allele frequency jumping to 0 in some years, only to fall back to previous levels in the next birth cohort. These simulations were removed from the analysis due to concerns they might lead to false signals of selection, when it is more likely these patterns

were caused by stochasticity due to very small numbers of individuals in some birth cohorts. Nonetheless, the removal of these simulations, although there were very few, might in turn have skewed the null distribution of allele frequency change in a way that could bias the detection of selection signals, which must thus be interpreted with caution. Overall, the sex-limited signals of transmission distortion discovered here indicate sexually dimorphic genetic architectures that could be the result of resolved sexual antagonism, rather than ongoing conflict. This interpretation is in agreement with evidence suggesting that sexual conflict is rapidly resolved through regulatory evolution and, therefore, evidence of unresolved conflict is difficult to detect (Wright *et al.*, 2019; Mank *et al.*, 2020).

4.5.2 Genetic basis of early life survival

I found that survival to the second year was significantly heritable, with additive genetic variance explaining approximately 10% of the total phenotypic variance. This was similar to previous findings in the deer that showed first and second winter survival separately, each having a heritability of about 5% (Gauzere *et al.*, 2020). However, unlike the analysis of prenatal viability above, I did not identify evidence of individual loci with significant effects on post-natal early life survival. Therefore, although early life survival probability clearly has a genetic component, the trait is likely polygenic with many loci of small effect rather than a small number of loci explaining most of the genetic variance. This kind of genetic architecture is typical for fitness-related traits, where additive genetic variance is thought to be maintained by high mutational variability due to a large number of loci influencing survival (Houle *et al.*, 1996; Merilä & Sheldon, 1999). Thus, as discussed in previous chapters, genetic variance can be maintained even in the face of strong selection if a trait is highly polygenic (Barton & Keightley, 2002). However, fitness heritability estimates are usually small, as most

fitness traits may be highly phenotypically and genetically correlated with other traits, leading to the evolution of condition dependence and contributing to small heritability values by increasing residual variance (Kruuk *et al.*, 2000; Merilä & Sheldon, 1999; Rowe & Houle, 1996). In this context, the heritability of early life survival in the deer seems larger than expected for this kind of trait, but the large uncertainty associated with the effects of the environmental components means that they might contribute more to the residual variance than is captured by the point estimate, thus underestimating residual variance and overestimating heritability.

As expected, there was a strong sex effect on early life survival probability, with female offspring around 14% more likely to survive to their second year compared to their male counterparts, confirming earlier findings on sex differences in early life survival in the red deer (Clutton-Brock, 1985). Additionally, maternal effects were also found to play a significant role in offspring survival, again explaining approximately 14% of the overall phenotypic variance, more than the additive genetic component. This is somewhat different from recent findings in the red deer of Rum showing much smaller (non-genetic) maternal effects for first and second winter survival (survival to age 1 = 3.2%, survival to age 2 = 3.1%) (Gauzere *et al.*, 2020), but as that study considered first and second year survival separately, the traits are not directly comparable. However, it is evident that maternal investment plays a non-negligible role in early life survival and is probably mediated in part by birth weight, which is also heavily influenced by maternal effects (Kruuk *et al.*, 2000; Gauzere *et al.*, 2020). As birth weight is markedly different between the sexes and is of great importance for male lifetime reproductive success (Clutton-Brock, 1985; Kruuk & Hadfield, 2007; Kruuk *et al.*, 1999), there might be potential for correlated sex-specific selection on birth weight to be uncovered by future studies. Due to time constraints, models incorporating cross-sex covariance for survival were not run. The cross-sex genetic correlation of early life survival could give further insight into sex-specific

effects and the genetic architecture and will be addressed in a future analysis.

4.5.3 Candidate genes associated with differential viability selection.

The GO enrichment analysis did not reveal significant enrichment for particular biological processes or functions associated with the transmission distorted loci. Nevertheless, I identified a number of candidate genes that are promising avenues for further study (Table 4.5). Of the four SNPs with trans-generational effects, three were associated with well-known oncogenes (*SMYD3*, *PIN1* and *RUNX2*); and of the male distorted loci under balancing selection, two were associated with major tumour suppressors (*RAD23B* and *INPP4B*). Oncogenes have been associated with transmission distortion in other species such as humans, mice, and cattle, due to embryonic lethality and paternal-specific segregation effects, particularly in genes associated with breast and prostate cancer (Healey *et al.*, 2000; Casellas *et al.*, 2017), of which *SMYD3* and *PIN1* have been implicated (Fenizia *et al.*, 2019; Bao *et al.*, 2004) and suppressors *RAD23B* and *INPP4B* are known to act against (Linge *et al.*, 2014; Han *et al.*, 2019; Agoulnik *et al.*, 2011; Kofuji *et al.*, 2015; Hodgson *et al.*, 2011). There is evidence that the role of *INPP4B* in suppressing prostate cancer is regulated by sex hormones (Hodgson *et al.*, 2011), providing a potential mechanism for sex specific effects of transmission distortion loci. The role of sex-specific selection on oncogenes is particularly interesting in deer, as a recent study found evidence for positive selection on several oncogenes and tumour suppressor genes expressed in the developing antler in Cervids, indicating that the rapid proliferation of cells required for fast antler growth is functionally similar to that of cancer cell growth (Wang *et al.*, 2019). This link to a key sex-limited trait makes these loci a promising target for sex-biased (and particularly male-biased) selection.

Several candidate genes were associated with embryonic development, in particular the morphogenesis of the heart (*SMYD* and *POU4F2*; Fujii *et al.* 2011; Budhram-Mahadeo *et al.* 2021; Maskell *et al.* 2017) and the skeleton (*RUNX2*, *POU4F2* and *PIN1*; Komori & Kishimoto 1998; Schroeder *et al.* 2005; Komori 2011; Nakashima *et al.* 2021). *POU4F2* has been identified as a homozygous embryonic lethal based on mouse models, with particularly adverse effects on heart tissue regeneration in males (Budhram-Mahadeo *et al.*, 2021), and here showed signals of transmission distortion and balancing selection in fathers. This locus is also in close proximity to *TTC29*, which is essential for sperm motility, with loss-of function mutations associated with male sterility in humans (Lorès *et al.*, 2019; Liu *et al.*, 2019). It is possible that haploid selection could act on an underperforming allele of *TTC29*, but it may be advantageous if linked to a deleterious mutation at *POU4F2* affecting heart development and maintenance of cardiac tissue, particularly in children. This hypothesis is favourable when considering the skewed fitness distribution in male red deer, so successful stags benefit the most from having successful sons themselves, increasing their genetic contribution to the population (Gomendio *et al.*, 2006).

Finally, two candidate genes were associated with transcription regulation and chromatin modification. A locus showing transmission distortion in mothers, *UBL5*, plays an important role in sister chromosome adhesion during mitosis (Oka *et al.*, 2014), which means that function-reducing mutations might adversely impact embryonic cell proliferation through abnormal chromosome segregation. Another function of this gene is regulation of transcription and mRNA modification via splicing (Oka *et al.*, 2015), which is of interest since oocyte-specific mRNAs initiate cell differentiation of the zygote into the embryo (Svoboda *et al.*, 2015). *SMYD3* was also distorted in mothers and is associated with chromatin organisation and epigenetic modification, which are essential processes for reprogramming zygote chromatin into embryo-specific chromatin (Beaujean, 2014). Thus, although none of these genes have

previously been implicated in prenatal viability or segregation distortion, they may be important in early embryonic development, especially via the transcriptomic landscape of the oocyte.

4.5.4 Advances, limitations and future work.

This study is one of the few to date that has quantified transmission distortion from parents and to offspring of different sexes at birth. Previous work in wild populations has either relied on F_{ST} based approaches between the sexes (often in adults) to identify sexual conflict in survival (Wright *et al.*, 2018), or has investigated this question at a small number of loci (such as the barn owl and Crohn's disease examples above; Friedrichs *et al.* 2006; Ducret *et al.* 2016). The Rum red deer research project is in the unique position of having access to genetic data of virtually all animals born in the study area in recent years, which, combined with high quality pedigree information, enabled this study to examine potential drivers of sexual conflict at different life stages. Furthermore, although there are some studies in humans and livestock that examine segregation distortion in parents or transmission distortion in offspring at birth separately (Henckaerts *et al.*, 2010; Casellas *et al.*, 2017), here I look at both of these processes. Considering both distortion in parents and offspring not only facilitates distinguishing between these different drivers of allelic bias, but also the identification of trans-generational distortion effects, where distortion at the same locus happens in both a parent and offspring.

Despite the advantages of the red deer data set, there are still some potential limitations and biases that are still to be explored. For example, patterns of sexually divergent allelic bias can be falsely mistaken for signals of sexual conflict when in reality they originate from a sex chromosome but are taken for an autosomal region due to sequence similarity, especially if these loci are on the Y chromosome, as most systems have no or incomplete sequences for

the Y (Kasimatis *et al.*, 2021; Bissegger *et al.*, 2020; Ruzicka *et al.*, 2020). Such spurious associations might be limited by excluding loci with biased sex-specific minor allele frequencies and read depth if genome sequences are available, but cannot be excluded completely due to, for example, recent duplication events from the autosome into a sex chromosome (Bissegger *et al.*, 2020). The imminent publication of the deer genome may highlight cases in our data where this may potentially have an effect. Another factor that can introduce bias and increase false positive rates is the fact that sample sizes in transmission distortion analysis can vary widely among loci due to allele frequency differences. Loci with rare alleles will have fewer informative crosses (because of fewer heterozygous parents where transmission distortion can be quantified) and therefore smaller sample sizes, which might increase the probability of spurious distortion patterns. However, I believe that this is less of an issue in the present study for the following reasons: I conducted GeneDrop simulations to give a realistic null distribution for each locus based on its allele frequency; I found no consistent trend between MAF and Fisher's exact test p-values; and all 136 significant hits for transmission distortion had a MAF $>$ 0.01 (see Table S4.2).

Nonetheless, it would be informative to explicitly investigate the effect of MAF on observed patterns of allele transmission and consequently on the p-value threshold determined from the null distribution and number of loci identified as showing evidence of transmission distortion. As mentioned above, the MAF of a locus can influence sample size and from the empirical p-value distribution, it appears that small sample sizes are more likely to result in larger p-values. This may be because small numbers of meioses make it more difficult to distinguish deviations from Mendelian inheritance from the sample variance, and thus these small sample sizes are uninformative. Similarly, although linkage was accounted for in the GeneDrop simulations, SNPs were modelled independently in the downstream Fisher's exact test analysis used to generate the null distribution of p-values, when in reality they might be linked

by LD, potentially introducing bias. Pruning the empirical data sets based on linkage and MAF results in a reduction of larger p-values compared to the full dataset (see figure S4.10). Thus, both MAF and LD might lead to a p-value cut-off for significance that is too stringent, by skewing the p-value distribution towards the right-hand tail (i.e., larger values). This effect is illustrated by comparing the quantiles of the complete empirical dataset against pruned versions: the 0.1%, 1% and 5% quantiles (representing reasonable significance thresholds) are 0.020, 0.068 and 0.167 respectively. And while the least stringent pruning (r^2 = 0.8, MAF = 0.05) led to very similar quantiles, 0.019, 0.067 and 0.166, the most stringent pruning (r^2 = 0.2, MAF = 0.1) resulted in quantiles that were slightly less biased towards the right-hand tail of the distribution: 0.023, 0.071 and 0.170. Furthermore, this bias can be observed in the results of this study: At the selected threshold of 0.1%, I only find 21 significant SNPs in mothers, 43 in fathers, 29 SNPs in daughters and 47 SNPs in sons, when with 35204 SNPs in each dataset even under the null model ~ 35 SNPs are expected to be identified as significant. Thus, in the case of the mother and daughter data sets, fewer SNPs than expected under the null hypothesis were significant and in the other two data sets the number of significant SNPs was not much greater than that expected under the null. Therefore, these effects should be explored in more detail.

There are several future analyses that would build on the findings of this study. There are two immediate next steps that were omitted due to time constraints. First, quantifying the cross-sex genetic correlation for early life survival would make it possible to identify if early life survival, similar to prenatal viability, has a sex-specific genetic architecture. Second, running the early life survival models within each sex would allow modelling of genotype-by-sex interactions based on the sex-specific effect sizes and genetic correlations (see Randall *et al.* 2013; Bernabeu *et al.* 2021). This approach could indicate non-zero causal associations with juvenile survival that not detected by the model above. In terms of other future analyses, it

appears that F_{ST} based studies generally identify more loci associated with sexually divergent viability compared to studies quantifying transmission distortion, yet only a small subset of these loci are assigned viability effects with high confidence (Flanagan & Jones, 2017; Lin *et al.*, 2021). Therefore, although not strictly independent from the transmission distortion approach, a complementary analysis of sexually divergent allele frequencies using F_{ST} could be useful to compare which signals, based on effect size or degree of divergence, are being detected between these two methods. Other future directions would leverage sequencing and transcriptomics, particularly in sperm, to test for potential haploid selection in males and specifically if allele ratios observed here are different from that in the ejaculate, and if sperm transcriptomes are haplotype specific. Such approaches are becoming more tractable with the advent of novel linked-read technology that facilitates linkage map construction and haplotype phasing (see Dréau *et al.* 2019 for a recent example in mice and sticklebacks).

Overall, with extensive genomic data becoming more readily available, genomic mapping studies that attempt to identify genotype-phenotype associations, such as those described here for viability and early life survival, are more common (Slate *et al.*, 2010; Santure & Garant, 2018). Such studies could add to the findings on genomic patterns of transmission distortion described here and give valuable insights into the pervasiveness and extent of resolution of sexual conflict loci in natural populations. The findings presented in this study highlight the utility of population genetic studies in wild populations and lend some support to the theory that sexual conflict is resolved quickly and loci with large differences in fitness effects between the sexes may be rare and/or transient (Wright *et al.*, 2019; Mank *et al.*, 2020).

Chapter 5

General Discussion.



Kilmory in the 'North Block' of Rum.

5.1 Aims and main findings

The main aim of this PhD was to investigate the genomic basis of sexually divergent traits in the red deer of the Isle of Rum by characterising the genetic architecture of antler morphology and identifying genomic signals of sexual conflict using classic quantitative genetic and genomic approaches. The main findings of this thesis are the following:

1. Antler morphology is heritable, and antler traits are highly phenotypically and genetically correlated. Various antler measurements have been

shown to have substantial additive genetic variation, contributing to the high repeatability of antler traits over an individual's lifetime. The high level of both phenotypic and genetic correlation among antler traits indicate some degree of shared genetic architecture. The extensive genetic covariances and smaller genetic variance of some antler PCs suggest that the genetic variance of antler morphology might be constrained along some and reduced along other phenotypic axes. These mechanisms may help explain apparent maintenance of genetic variation despite strong directional selection in antler traits in this population of red deer and more generally. Antler morphology also has a strong environmental component that influences overall phenotypic variation, therefore condition dependency is also likely to play a role in retaining substantial heritabilities of antler traits and of sexually selected traits more broadly.

2. Antler morphology is likely to be highly polygenic, with a moderate degree of pleiotropy across different antler traits. A single region on linkage group 21 was associated with antler PC9 using regional heritability analysis, but with the current data it is essentially impossible to identify the causal gene underlying the association with PC9. Furthermore, as PC9 only represents $\sim$4% of the overall phenotypic variance among all antler PCs, effect sizes of causal loci are expected to be very small. Most traits were underpinned by SNPs with uniformly small effect sizes, as expected under the infinitesimal model. The expectations of the infinitesimal model that genetic variance remains relatively constant and the fact that polygenic architecture represents a large mutational target might help counteract loss of genetic variation. Additionally, antler traits show putative evidence of extensive pleiotropy among them, as is common for polygenic traits. If these pleiotropic effects extend to other traits and/or are in LD with loci associated with fitness, they could maintain genetic variation through conflicts and trade-offs.

3. Red deer show evidence of at least partially resolved sexual conflict. Several loci exhibit transmission distortion compatible with sexually divergent prenatal selection on Mendelian segregation of gametes and offspring viability. Of these, four loci showed trans-generational effects, meaning SNPs that deviated from Mendelian expectations in a parent also deviated in an offspring while always favouring the same allele. As none of the SNPs exhibiting transmission distortion showed opposing effects in males and females, the evidence is more compliant with sex-specific genetic architectures as a consequence of the resolution of sexual conflict. SNPs showing segregation distortion in fathers were prevalent among the high confidence SNPs for transmission distortion, especially among the markers that also exhibited signatures of balancing selection. Thus, haploid selection on sperm might a major driver of transmission distortion and resolution of sexual conflict through the emergence of sex specific architecture via sex dependent selection. A number of candidate genes near distorted loci were identified, several of which were involved in carcinogenesis or early development, and one *TTC29* was directly linked to sperm mobility and fertility. Early life survival was a strongly influenced by sex with male calves much less likely to survive to two years compared to females, but no causal genetic loci were identified. Survival was significantly heritable and strongly influenced by maternal effects, but due to its polygenic architecture, no potential genetic driver for sexual conflict could be determined.

5.2 Novel analysis approaches and findings

Modelling multidimensionality of antler morphology. Antler morphology is a multidimensional trait and although previous studies in the Rum red deer have characterised genetic (co)-variances of various antler traits (Kruuk *et al.*,

2002, 2014), several more antler measures remained unexplored. In Chapters 2 and 3, I examined several more antler traits to help understand the genetic basis and evolutionary constraints of antler morphology. Additionally, I modelled antler morphology by conducting a principal component analysis to identify phenotypic axes representing distinct characteristics of antlers. Principal component analyses can be helpful in reducing complex traits to specific trait aspects that might be more informative in association studies, for example for the identification of causal loci (Pallares *et al.*, 2014). To my knowledge, no other study has used principal components to study the genetic basis of antler morphology, especially while considering all identified PCs, including higher order PCs, which are usually disregarded in studies using principal component analysis. This approach resulted in the only significant association of any antler trait with a genomic region, namely PC9, with a region on linkage group 21.

Novel association approaches for QTL identification. In Chapter 3, I used regional heritability analysis and an empirical Bayesian FDR approach to map antler morphology QTLs. Both of these approaches have mainly been used in human genetics studies and are designed to identify loci with small effects that are difficult to detect using GWAS (Nagamine *et al.*, 2012; Stephens, 2016). Regional heritability has previously been used successfully in the red deer to identify a QTL for recombination frequency, (Johnston *et al.*, 2018) and using this method I was able to identify a genomic region explaining genetic variation in one of the antler PCs when GWAS failed to detect any causal loci. The empirical Bayesian FDR method proposed by Stephens (2016) was developed to identify small or non-zero effect loci based on estimates from GWAS and has been shown to be capable of identifying previously unknown loci such as for human height (Boyle *et al.*, 2017). Using this method to identify non-zero SNPs underlying antler morphology provided additional lines of evidence supporting a polygenic architecture; it also helped elucidate the extent of pleiotropy and linkage disequilibrium among antler traits, which in turn provided insights into

potential evolutionary constraints. As far as I am aware, no other study has used this approach to identify QTLs in a natural, non-human population and the analysis conducted in this thesis could act as a guide on how to characterise the genetic architecture of complex traits in non-model organisms.

Evidence of prenatal viability and gametic selection in a wild population. In Chapter 4, I used genomic marker and pedigree information to detect evidence of transmission distortion as an indicator of sexual conflict in the red deer of Rum. I found several loci with significant deviations from Mendelian expectation in line with gametic and prenatal viability selection. Research trying to detect genomic signals of sexual conflict are still relatively new, and most focus on sexually dimorphic gene expression and reproductive fitness (Mank, 2017b; Cheng & Kirkpatrick, 2016; Dutoit *et al.*, 2018; Ruzicka *et al.*, 2019). Very few studies have explored sexual conflict over survival (Lucotte *et al.*, 2016; Flanagan & Jones, 2017; Lin *et al.*, 2021), even fewer have examined prenatal viability or segregation distortion and those that do are exclusively in humans and livestock or model organisms (Healey *et al.*, 2000; Friedrichs *et al.*, 2006; Casellas *et al.*, 2017). Thus, the analysis presented in Chapter 4 is one of the first attempts to identify drivers of prenatal selection on viability and gamete segregation in a wild organism and contributes novel findings about the extent and types of genes involved in sexual conflict and sex-specific selection.

5.3 Limitations of quantitative genetics and genomics studies in wild populations

Quantitative genetics and genomics studies in the wild face several limitations that studies in domesticated or model species in the lab do not. As the pedigree structure in wild populations is much more complex compared to

livestock populations, the power to detect quantitative trait loci is much reduced. Furthermore, the much more variable environment natural populations experience means the environmental component of phenotypic variance is much greater for most traits, and thus heritabilities are often smaller and harder to detect. Genomic studies in the wild, such as GWAS, often have small sample sizes and insufficient LD due to limited genomic coverage, which can impede the detection of causal loci. This is a consequence of reduced funding compared to similar research in livestock or humans. The Rum red deer study is in a comparably much better position, with relatively large sample sizes as the result of long-term monitoring that most other studies in natural populations don't have access to, demonstrating the need for long-term study systems in the wild. Nonetheless, compared to studies in humans or domesticated species the sample sizes are still relatively small, although they should be sufficient to detect large effect loci (see Chapter 3). Adequate LD also seems less of a problem in the red deer, and it is possible that an increase in typed markers would not necessarily result in the identification of loci with large effects underlying antler morphology (see Chapter 3 for more details). Another limitation in studies of wild populations is that more invasive approaches that are possible in the lab or a breeding program are not achievable because of ethical concerns, general feasibility or clashing research interests, particularly as long-term research projects often need to balance the needs of multiple different studies. Even approaches like mRNA sequencing would be difficult to do in the deer if it required the capturing of adult deer, which would be the case, for example, if gene expression differences between mature males and females should be determined as an extension of the sexual conflict study laid out in Chapter 4. Adult deer can only be captured if anaesthetised by a dart and certain tissues such as reproductive organs would be hard to sample without harming the animal and the sampling effort required generally would reduce sample size considerably. However, there is some managed culling of adult deer on other areas of Rum, which could allow such studies of genetically similar animals.

Other methods that might be of interest for future directions of the sexual conflict analysis conducted in this thesis include sampling of naturally aborted or unborn fetuses (from deceased individuals or from culled females outside the study area) or the study of fertilisation and early development *in vitro*; however, these are virtually impossible to carry out in wild animals within the scope of the current project.

5.4 Future directions

Selection analysis of antler morphology. After having characterised the genetic architecture of antler morphology in detail, a logical next step would be to explore the selection pressures acting on it. Previous work has identified selection gradients on antler weight and number of tines, and found selection to be constrained by condition dependency and lack of genetic covariance with fitness, partly driven by opposing genetic selection gradients of the two traits (Kruuk *et al.*, 2002, 2014). In this thesis, I have identified extensive genetic covariances and pleiotropy among several additional antler traits, and so it would be elucidating to determine if similar constraints as documented for antler weight and form are acting on other antler measures as well. Additionally, analysing selection gradients of antler PCs could help characterise selection on those parts of antler morphology that are unconstrained by genetic covariances, as the lack of pleiotropy among PCs described in this thesis suggests they are genetically independent. Furthermore, as antler morphology clearly has a condition dependent component, other traits closely linked to fitness and condition could play a role in constraining its evolution. Birth weight is such a trait, and although it is strongly influenced by maternal effects, it also has substantial additive genetic variance (Kruuk *et al.*, 2000; Gauzere *et al.*, 2020) and is an important factor for lifetime reproductive success in males. These approaches could help

further advance our understanding of the genetic constraints and evolutionary potential of antler morphology, and thus further our knowledge of the evolution of sexually selected traits in general.

Sexually dimorphic gene expression patterns. As mentioned previously, detecting genomic signals of sexual conflict is often done by exploring differences in gene expression patterns between the sexes. This approach could complement the analysis of signals of sexual antagonism in the red deer using transmission distortion presented in this thesis. The research shown in Chapter 4 focused on sexual conflict over survival, both at the prenatal and early life stages, but genomic signatures of antagonism over reproductive fitness remain unexplored. As the degree of sex bias in gene expression is correlated with the extent of resolution of sexual conflict (Cheng & Kirkpatrick, 2016), examining sex biases in the transcriptome could shed further light on the extent of ongoing versus resolved sexual antagonism in the deer, since findings in Chapter 4 point more towards resolved conflict. Results from such analysis contributes to our understanding of the persistence of sexual antagonism, which is still hotly debated (Mank *et al.*, 2020), and can inform our understanding of its role in the maintenance of genetic variance (Dutoit *et al.*, 2018) and evolution in general. In addition, gonad specific transcriptomes could help explore some of the potential mechanisms contributing to differential prenatal viability driven by gametic haplotype proposed in Chapter 4. In particular, gene expression profiles of sperm in combination with genotyping could test some of the proposed hypotheses on haploid selection in males, specifically if allele ratios observed in Chapter 4 are present in the ejaculate and if sperm transcriptomes are haplotype specific. And if possible, both the genotype and transcriptome of naturally aborted fetuses could help confirm and/or extend the findings of this thesis on transmission distortion drivers that influence prenatal viability.

Bibliography

Agoulnik IU, Hodgson MC, Bowden WA, Ittmann MM (2011) INPP4B: the New Kid on the PI3K Block. *Oncotarget*, **2**, 321.

Agulnik SI, Sabantsev ID, Ruvinsky AO (1993) Effect of sperm genotype on chromatid segregation in female mice heterozygous for aberrant chromosome 1. *Genetical Research*, **61**, 97–100.

Alkuraya FS (2016) Discovery of mutations for Mendelian disorders.

Anderson S, Côté S, Richard J, Shafer A (2019) Genomic architecture of artificially and sexually-selected traits in a wild cervid. *bioRxiv*, p. 841528.

Andersson MB (1994) *Sexual Selection*. Princeton University Press.

Arbuthnott D, Rundle HD (2012) Sexual selection is ineffectual or inhibits the purging of deleterious mutations in drosophila melanogaster. *Evolution*, **66**, 2127–2137.

Arnold SJ, Wade MJ (1984) On the Measurement of Natural and Sexual Selection: Theory. *Evolution*, **38**, 709–719.

Arnqvist G, Edvardsson M, Friberg U, Nilsson T (2000) Sexual conflict promotes speciation in insects. *Proceedings of the National Academy of Sciences*, **97**, 10460–10464.

Arnqvist G, Rowe L (2005) *Sexual conflict* . Princeton University Press, Princeton, NJ.

Asghar M, Bensch S, Tarka M, Hansson B, Hasselquist D (2015) Maternal and genetic factors determine early life telomere length. *Proceedings of the Royal Society B: Biological Sciences*, **282**.

Aulchenko YS, de Koning DJ, Haley C (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, **177**, 577–85.

Bagchi B, Corbel Q, Khan I, *et al.* (2021) Sexual conflict drives micro- and macroevolution of sexual dimorphism in immunity. *BMC Biology*, **19**, 1–19.

Bao L, Kimzey A, Sauter G, Sowadski JM, Lu KP, Wang DG (2004) Prevalent Overexpression of Prolyl Isomerase Pin1 in Human Cancers. *American Journal of Pathology*, **164**, 1727–1737.

Barson NJ, Aykanat T, Hindar K, *et al.* (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, **528**, 405–408.

Barton N, Etheridge A, Véber A (2017) The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology*, **118**, 50–73.

Barton NH, Keightley PD (2002) Understanding quantitative genetic variation. *Nature Reviews Genetics*, **3**, 11–21.

Beaujean N (2014) Histone post-translational modifications in preimplantation mouse embryos and their role in nuclear architecture.

Beavis WD (1998) QTL analyses: power, precision, and accuracy. In *Molecular Dissection of Complex Traits* (edited by AH Paterson), pp. 145–162. CRC Press, New York.

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing . *Journal of the Royal Statistical Society*, **57**, 289–300.

Beraldi D, McRae AF, Gratten J, Slate J, Visscher PM, Pemberton JM (2006) Development of a linkage map and mapping of phenotypic polymorphisms in a free-living population of soay sheep (Ovis aries). *Genetics*, **173**, 1521–1537.

Berenos C, Ellis PA, Pilkington JG, Lee SH, Gratten J, Pemberton JM (2015) Heterogeneity of genetic architecture of body size traits in a free-living population. *Molecular Ecology*, **24**, 1810–1830.

Bérénos C, Ellis PA, Pilkington JG, Pemberton JM (2014) Estimating quantitative genetic parameters in wild populations: A comparison of pedigree and genomic approaches. *Molecular Ecology*, **23**, 3434–3451.

Bergsten J, Töyrä A, Nilsson AN (2001) Intraspecific variation and intersexual correlation in secondary sexual characters of three diving beetles (Coleoptera: Dytiscidae). *Biological Journal of the Linnean Society*, **73**, 221–232.

Bernabeu E, Canela-Xandri O, Rawlik K, Talenti A, Prendergast J, Tenesa A (2021) Sex differences in genetic architecture in the UK Biobank. *Nature Genetics 2021 53:9*, **53**, 1283–1289.

Beroukhim R, Mermel CH, Porter D, *et al.* (2010) The landscape of somatic copy-number alteration across human cancers. *Nature*, **463**, 899–905.

Berthold P, Pulido F (1994) Heritability of migratory activity in a natural bird population. *Proceedings of the Royal Society B: Biological Sciences*, **257**, 311–315.

Bhutani K, Stansifer K, Ticau S, *et al.* (2021) Widespread haploid-biased gene expression enables sperm-level natural selection. *Science*, **371**, p. eabb1723.

Bissegger M, Laurentino TG, Roesti M, Berner D (2020) Widespread intersex differentiation across the stickleback genome – The signature of sexually antagonistic selection? *Molecular Ecology*, **29**, 262–271.

Blyth K, Vaillant F, Jenkins A, *et al.* (2010) Runx2 in normal tissues and cancer cells: A developing story. *Blood Cells, Molecules, and Diseases*, **45**, 117–123.

Bonduriansky R, Rowe L (2005) Sexual selection, genetic architecture, and the condition dependence of body shape in the sexually dimorphic fly Prochyliza xanthostoma (piophilidae). *Evolution*, **59**, 138–151.

Bonnet T, Morrissey MB, Kruuk LEB (2019a) Estimation of Genetic Variance in Fitness, and Inference of Adaptation, When Fitness Follows a Log-Normal Distribution. *Journal of Heredity*, **110**, 383–395.

Bonnet T, Morrissey MB, Morris A, *et al.* (2019b) The role of selection and evolution in changing parturition date in a red deer population. *PLoS Biology*, **17**, e3000493.

Bottino C, Peserico A, Simone C, Caretti G (2020) SMYD3: An oncogenic driver targeting epigenetic regulation and signaling pathways. *Cancers*, **12**.

Bourret A, Belisle M, Pelletier F, Garant D (2017) Evolutionary potential of morphological traits across different life-history stages. *Journal of Evolutionary Biology*, **30**, 616–626.

Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, **169**, 1177–1186.

Braun RE, Behringer RR, Peschon JJ, Brinster RL, Palmiter RD (1989) Genetically haploid spermatids are phenotypically diploid. *Nature*, **337**, 373–376.

Brauning R, Fisher PJ, McCulloch AF, *et al.* (2015) Utilization of high throughput genome sequencing technology for large scale single nucleotide polymorphism discovery in red deer and Canadian elk. *bioRxiv*.

Bro-Jørgensen J (2007) The intensity of sexual selection predicts weapon size in male bovids. *Evolution*, **61**, 1316–1326.

Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GD (2014) Running with the Red Queen: The role of biotic conflicts in evolution. *Proceedings of the Royal Society B: Biological Sciences*, **281**.

Bubenik GA, Schams D, Coenen G (1987) The effect of artificial photoperiodicity and antiandrogen treatment on the antler growth and plasma levels of LH, FSH, testosterone, prolactin and alkaline phosphatase in the male white-tailed deer. *Comparative Biochemistry and Physiology – Part A: Physiology*, **87**, 551–559.

Buczek M, Okarma H, Demiaszkiewicz AW, Radwan J (2016) MHC, parasites and antler development in red deer: no support for the Hamilton & Zuk hypothesis. *Journal of Evolutionary Biology*, **29**, 617–632.

Budhram-Mahadeo VS, Solomons MR, Mahadeo-Heads EA (2021) Linking metabolic dysfunction with cardiovascular diseases: Brn-3b/POU4F2 transcription factor in cardiometabolic tissues in health and disease. *Cell Death and Disease*, **12**, 1–13.

Bulmer MG (1987) The Mathematical Theory of Quantitative Genetics. *Biometrics*, **43**, 258.

Bush WS, Moore JH (2012) Chapter 11: Genome-Wide Association Studies. *Plos Computational Biology*, **8**, e1002822.

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) Mixed Models for S language Environments: ASReml-R reference manual.

Camacho C, Coulouris G, Avagyan V, *et al.* (2009) BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**.

Campbell CD, Ogburn EL, Lunetta KL, *et al.* (2005) Demonstrating stratification in a European American population. *Nature Genetics*, **37**, 868–872.

Casellas J, Cañas-Álvarez JJ, González-Rodríguez A, *et al.* (2017) Bayesian analysis of parent-specific transmission ratio distortion in seven Spanish beef cattle breeds. *Animal Genetics*, **48**, 93–96.

Catalano R, Bruckner T (2006) Secondary sex ratios and male lifespan: Damaged or culled cohorts. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1639–1643.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**, s13742–015–0047–8.

Chapman T, Arnquist G, Bangham J, Rowe L (2003) Sexual conflict. *Trends in Ecology and Evolution*, **18**, 41–47.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L (1995) Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. *Nature*, **373**, 241–244.

Charmantier A, Gienapp P (2014) Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications*, **7**, 15–28.

Cheng C, Kirkpatrick M (2016) Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLoS Genetics*, **12**, e1006170.

Chenoweth SF, Appleton NC, Allen SL, Rundle HD (2015) Genomic Evidence that Sexual Selection Impedes Adaptation to a Novel Environment. *Current Biology*, **25**, 1860–1866.

Chenoweth SF, Hunt J, Rundle HD (2012) Analyzing and Comparing the Geometry of Individual Fitness Surfaces. *Adaptive Landscape in Evolutionary Biology*, pp. 126–149.

Chenoweth SF, McGuigan K (2010) The genetic basis of sexually selected variation. *Annual Review of Ecology, Evolution, and Systematics*, **41**, 81–101.

Chippindale AK (2001) Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in Drosophila. *Proceedings of the National Academy of Sciences*, **98**, 1671–1675.

Civetta A, Singh RS (1998) Sex and speciation: genetic architecture and evolutionary potential of sexual versus nonsexual traits in the sibling species of the Drosophila melanogaster complex. *Evolution*, **52**, 1080–1092.

Clutton-Brock T, Guiness F, Albon S (1982) *Red Deer. Behaviour and Ecology of Two Sexes*. University of Chicago Press.

Clutton-Brock TH (1985) Reproductive Success in Red Deer. *Scientific American*, **252**, 86–92.

Clutton-Brock TH (1991) *The Evolution of Parental Care*. Princeton University Press.

Clutton-Brock TH, Albon SD, Guinness FE (1981) Parental investment in male and female offspring in polygynous mammals. *Nature 1981 289:5797*, **289**, 487–489.

Clutton-Brock TH, Albon SD, Guinness FE (1989) Fitness costs of gestation and lactation in wild mammals. *Nature 1989 337:6204*, **337**, 260–262.

Codato R, Perichon M, Divol A, *et al.* (2019) The SMYD3 methyltransferase promotes myogenesis by activating the myogenin regulatory network. *Scientific Reports*, **9**, 1–14.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, **142**, 169–196.

Coltman DW, O'Donoghue P, Hogg JT, Festa-Bianchet M (2005) Selection and genetic (co)variance in bighorn sheep. *Evolution*, **59**, 1372–1382.

Comuzzie AG, Hixson JE, Almasy L, *et al.* (1997) A major quantitative trait locus determining serum leptin levels and fat mass is located on human chromosome 2. *Nature Genetics*, **15**, 273–276.

Connallon T, Hall MD (2018) Genetic constraints on adaptation: a theoretical primer for the genomics era. *Annals of the New York Academy of Sciences*, **1422**, 65–87.

Coulson T, Albon S, Slate J, Pemberton J (1999) Microsatellite Loci Reveal Sex-Dependent Responses to Inbreeding and Outbreeding in Red Deer Calves. *Evolution*, **53**, 1951.

Cox RM, Calsbeek R (2009) Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *American Naturalist*, **173**, 176–187.

Dapper AL, Wade MJ (2016) The evolution of sperm competition genes: The effect of mating system on levels of genetic variation within and between species. *Evolution*, **70**, 502–511.

Darwin C (1981) *The Descent of Man, and Selection in Relation to Sex*. Princeton University Press, Princeton.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Davis EB, Brakora KA, Lee AH (2011) Evolution of ruminant headgear: A review.

Derome N, Bougas B, Rogers SM, *et al.* (2008) Pervasive sex-linked effects on transcription regulation as revealed by expression quantitative trait loci mapping in lake whitefish species pairs (Coregonus sp., Salmonidae). *Genetics*, **179**, 1903–1917.

Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.

DiPietro JA, Voegtline KM (2017) The gestational foundation of sex differences in development and vulnerability. *Neuroscience*, **342**, 4–20.

Dobzhansky T (1971) *Genetics of the evolutionary process*, vol. 139. Columbia University Press.

Dréau A, Venu V, Avdievich E, Gaspar L, Jones FC (2019) Genome-wide recombination map construction from single individuals using linked-read sequencing. *Nature Communications*, **10**.

Dubiella C, Pinch BJ, Koikawa K, *et al.* (2021) Sulfopin is a covalent inhibitor of Pin1 that blocks Myc-driven tumors in vivo. *Nature Chemical Biology*, pp. 1–10.

Ducret V, Gaigher A, Simon C, Goudet J, Roulin A (2016) Sex-specific allelic transmission bias suggests sexual conflict at MC1R. *Molecular ecology*, **25**, 4551–4563.

Dugand RJ, Tomkins JL, Kennington WJ (2019) Molecular evidence supports a genic capture resolution of the lek paradox. *Nature Communications*, **10**, 1–8.

Durinck S, Moreau Y, Kasprzyk A, *et al.* (2005) BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nature Protocols*, **4**, 1184–1191.

Dutoit L, Mugal CF, Bolívar P, *et al.* (2018) Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (Ficedula albicollis) genome. *Molecular Ecology*, **27**, 3572–3581.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, **107**, 1–15.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

Ellegren H, Parsch J (2007) The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, **8**, 689–698.

Emlen DJ (1994) Environmental control of horn length dimorphism in the beetle Onthophagus acuminatus (Coleoptera: Scarabaeidae). *Proceedings of the Royal Society B-Biological Sciences*, **256**, 131–136.

Endler JA, Basolo AL (1998) Sensory ecology, receiver biases and sexual selection. *Trends in Ecology & Evolution*, **13**, 415–420.

Evans SR, Postma E, Sheldon BC (2020) It takes two: Heritable male effects on reproductive timing but not clutch size in a wild bird population*. *Evolution*, **74**, 2320–2331.

Facon B, Pointier JP, Jarne P, Sarda V, David P (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Current Biology*, **18**, 363–367.

Falconer DS, Mackay TFC (1996) *Introduction to quantitative genetics*. Longman, Harlow, fourth edi edn..

Fenizia C, Bottino C, Corbetta S, *et al.* (2019) SMYD3 promotes the epithelial-mesenchymal transition in breast cancer. *Nucleic Acids Research*, **47**, 1278–1293.

Fijarczyk A, Babik W (2015) Detecting balancing selection in genomes: Limits and prospects. *Molecular Ecology*, **24**, 3529–3545.

Finseth FR, Nelson TC, Fishman L (2021) Selfish chromosomal drive shapes recent centromeric histone evolution in monkeyflowers. *PLoS Genetics*, **17**, e1009418.

Fisher RA (1919) XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, **52**, 399–433.

Fisher RA (2011) *The genetical theory of natural selection.* The Clarendon Press, Oxford.

Flanagan SP, Jones AG (2017) Genome-wide selection components analysis in a fish with male pregnancy. *Evolution*, **71**, 1096–1105.

Flint J, Mackay TFC (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, **19**, 723–733.

Folstad I, Karter AJ (1992) Parasites , Bright Males , and the Immunocompetence Handicap. *The American Naturalist*, **139**, 603–622.

Friedrichs F, Brescianini S, Annese V, *et al.* (2006) Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Human Genetics*, **119**, 305–311.

Froy H, Martin J, Stopher KV, *et al.* (2019) Consistent within-individual plasticity is sufficient to explain temperature responses in red deer reproductive traits. *Journal of Evolutionary Biology*, **32**, 1194–1206.

Froy H, Walling CA, Pemberton JM, Clutton-Brock TH, Kruuk LE (2016) Relative costs of offspring sex and offspring survival in a polygynous mammal. *Biology Letters*, **12**.

Fujii T, Tsunesumi Si, Yamaguchi K, Watanabe S, Furukawa Y (2011) Smyd3 is required for the development of cardiac and skeletal muscle in zebrafish. *PLoS ONE*, **6**, 23491.

Gauzere J, Pemberton JM, Morris S, Morris A, Kruuk LE, Walling CA (2020) The genetic architecture of maternal effects across ontogeny in the red deer. *Evolution*, **74**, 1378–1391.

Gauzere J, Walling CA, Pick JL, *et al.* (2021) The role of maternally transferred antibodies in maternal performance in red deer. *Ecology Letters*, **24**, 2065–2076.

Gavrilets S (2014) Is sexual conflict an “Engine of speciation”? *Cold Spring Harbor Perspectives in Biology*, **6**.

Geldermann H, Pieper U, Roth B (1985) Effects of marked chromosome sections on milk performance in cattle. *Theoretical and Applied Genetics*, **70**, 138–146.

George AW, Visscher PM, Haley CS (2000) Mapping quantitative trait loci in Complex pedigrees: A two-step variance component approach. *Genetics*, **156**, 2081–2092.

Gershoni M, Pietrokovski S (2014) Reduced selection and accumulation of deleterious mutations in genes exclusively expressed in men. *Nature Communications*, **5**, 1–10.

Gibson JR, Chippindale AK, Rice WR (2002) The X chromosome is a hot spot for sexually antagonistic fitness variation. *Proceedings of the Royal Society B: Biological Sciences*, **269**, 499–505.

Golan D, Lander ES, Rosset S (2014) Measuring missing heritability: inferring the contribution of common variants. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 5272–81.

Gomendio M, Clutton-Brock TH, Albon SD, Guinness FE, Simpson MJ (1990) Mammalian sex ratios and variation in costs of rearing sons and daughters. *Nature*, **343**, 261–263.

Gomendio M, Malo AF, Soler AJ, *et al.* (2006) Male Fertility and Sex Ratio at Birth in Red Deer. *Science*, **314**, 1445–1447.

Goss RJ (1983) Developmental Anatomy of Antlers. In *Deer Antlers: Regeneration, Function and Evolution*, pp. 133–136. Academic Press.

Gotoh H, Miyakawa H, Ishikawa A, *et al.* (2014) Developmental Link between Sex and Nutrition; doublesex Regulates Sex-Specific Mandible Growth via Juvenile Hormone Signaling in Stag Beetles. *PLoS Genetics*, **10**, e1004098.

Griffith SC, Owens IP, Burke T (1999) Environmental determination of a sexually selected trait. *Nature*, **400**, 358–360.

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, **69**, 315–324.

Hall MD, Lailvaux SP, Blows MW, Brooks RC (2010) Sexual conflict and the maintenance of multivariate genetic variation. *Evolution*, **64**, 1697–1703.

Hammerschmidt K, Deines P, Wilson AJ, Rolff J (2011) Quantitative genetics of immunity and life history under different photoperiods. *Heredity 2012 108:5*, **108**, 569–576.

Han W, Wang L, Zhang L, Wang Y, Li Y (2019) Circular RNA circ-RAD23B promotes cell growth and invasion by miR-593–3p/CCND2 and miR-653–5p/TIAM1 pathways in non-small cell lung cancer. *Biochemical and Biophysical Research Communications*, **510**, 462–466.

Hawkes MF, Gamble CE, Turner EC, Carey MR, Wedell N, Hosken DJ (2016) Intralocus sexual conflict and insecticide resistance. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20161429.

Hayes B, Goddard ME (2001) The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, **33**, 209–229.

Hayward A, Gillooly JF (2011) The Cost of Sex: Quantifying Energetic Investment in Gamete Production by Males and Females. *PLOS ONE*, **6**, e16557.

Healey CS, Dunning AM, Dawn Teare M, *et al.* (2000) A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability. *Nature Genetics*, **26**, 362–364.

Hedrick PW (2007) Balancing selection. *Current Biology*, **17**.

Henckaerts L, Vlietinck R, Derom C, Boonen S, Rutgeerts P, Vermeire S (2010) Transmission ratio distortion of DLG5 R30Q: Evidence for prenatal selection? *Inflammatory Bowel Diseases*, **16**, 910–911.

Henderson CR (1975) Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, **31**, 423–447.

Hendrickx F, Corte ZD, Sonet G, Belleghem SMV, Köstlbacher S, Vangestel C (2021) A masculinizing supergene underlies an exaggerated male reproductive morph in a spider. *bioRxiv*, p. 2021.02.09.430505.

Hickey JM, Kinghorn BP, Tier B, Van Der Werf JH, Cleveland MA (2012) A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics Selection Evolution*, **44**, 1–11.

Hill WG (2010) Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **365**, 73–85.

Hodgson MC, Shao Lj, Frolov A, *et al.* (2011) Decreased Expression and Androgen Regulation of the Tumor Suppressor Gene INPP4B in Prostate Cancer. *Cancer Research*, **71**, 572–582.

Houle D, Morikawa B, Lynch M (1996) Comparing mutational variabilities. *Genetics*, **143**, 1467–1483.

Huang W, Mackay TF (2016) The Genetic Architecture of Quantitative Traits Cannot Be Inferred from Variance Component Analysis. *PLoS Genetics*, **12**, 1006421.

Huang Y, Wright SI, Agrawal AF (2014) Genome-Wide Patterns of Genetic Variation within and among Alternative Selective Regimes. *PLoS Genetics*, **10**, 1004527.

Huisman J (2017) Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular Ecology Resources*, **17**, 1009–1024.

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM (2016) Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 3585–3590.

Hunt J, Blows MW, Zajitschek F, Jennions MD, Brooks R (2007) Reconciling strong stabilizing selection with the maintenance of genetic variation in a natural population of black field crickets (Teleogryllus commodus). *Genetics*, **177**, 875–880.

Hunter D (2020) GeneDrop.

Husby A, Kawakami T, Rönnegård L, Smeds L, Ellegren H, Qvarnström A (2015) Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proceedings of the Royal Society B: Biological Sciences*, **282**.

Immler S (2019) Haploid Selection in "diploid" Organisms. *Annual Review of Ecology, Evolution, and Systematics*, **50**, 219–236.

Immonen E, Snook RR, Ritchie MG (2014) Mating system variation drives rapid evolution of the female transcriptome in Drosophila pseudoobscura. *Ecology and Evolution*, **4**, 2186–2201.

Ito Y, Harigai A, Nakata M, *et al.* (2013) The role of doublesex in the evolution of exaggerated horns in the Japanese rhinoceros beetle. *EMBO Reports*, **14**, 561–567.

Jackson CH (2011) Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software*, **38**, 1–28.

Jamieson A, Anderson SJ, Fuller J, Côté SD, Northrup JM, Shafer AB (2020) Heritability estimates of antler and body traits in white-tailed deer (odocoileus virginianus) from genomic-relatedness matrices. *Journal of Heredity*, **111**, 429–435.

Jaquiéry J, Rispe C, Roze D, *et al.* (2013) Masculinization of the X Chromosome in the Pea Aphid. *PLoS Genetics*, **9**, 1003690.

Jiang S, Shen QW (2019) Principles of gene editing techniques and applications in animal husbandry. *3 Biotech*, **9**.

Jiang ZF, Machado CA (2009) Evolution of sex-dependent gene expression in three recently diverged species of Drosophila. *Genetics*, **183**, 1175–1185.

Johnson T, Barton N (2005) Theoretical models of selection and mutation on quantitative traits. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1411–1425.

Johnston SE, Beraldi D, McRae AF, Pemberton JM, Slate J (2009) Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity*, **104**, 196–205.

Johnston SE, Bérénos C, Slate J, Pemberton JM (2016) Conserved genetic architecture underlying individual recombination rate variation in a wild population of soay sheep (Ovis aries). *Genetics*, **203**, 583–598.

Johnston SE, Gratten J, Berenos C, *et al.* (2013a) Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, **502**, 93–95.

Johnston SE, Huisman J, Ellis PA, Pemberton JM (2017) A High-Density Linkage Map Reveals Sexual Dimorphism in Recombination Landscapes in Red Deer (Cervus elaphus). *G3-Genes Genomes Genetics*, **7**, 2859–2870.

Johnston SE, Huisman J, Pemberton JM (2018) A Genomic Region Containing REC8 and RNF212B Is Associated with Individual Recombination Rate Variation in a Wild Population of Red Deer (Cervus elaphus). *G3: Genes|Genomes|Genetics*, **8**, 2265–2276.

Johnston SE, Lindqvist M, Niemelä E, *et al.* (2013b) Fish scales and SNP chips: SNP genotyping and allele frequency estimation in individual and pooled DNA from historical samples of Atlantic salmon (Salmo salar). *BMC Genomics*, **14**, 1–13.

Johnston SE, McEwan JC, Pickering NK, *et al.* (2011) Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, **20**, 2555–2566.

Jones AG, Small CM, Paczolt KA, Ratterman NL (2010) A practical guide to methods of parentage analysis. *Molecular Ecology Resources*, **10**, 6–30.

Kasimatis KR, Abraham A, Ralph PL, Kern AD, Capra JA, Phillips PC (2021) Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics*, **217**.

Kasimatis KR, Nelson TC, Phillips PC (2017) Genomic Signatures of Sexual Conflict. *Journal of Heredity*, **108**, 780–790.

Kasimatis KR, Ralph PL, Phillips PC (2019) Limits to Genomic Divergence Under Sexually Antagonistic Selection. *G3: Genes|Genomes|Genetics*, **9**, 3813.

Kijimoto T, Moczek AP, Andrews J (2012) Diversification of doublesex function underlies morph-, sex-, and species-specific development of beetle horns. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 20526–20531.

Kim SY, Velando A (2016) Genetic conflict between sexual signalling and juvenile survival in the three-spined stickleback. *Bmc Evolutionary Biology*, **16**.

King EG, Long AD (2017) The beavis effect in next-generation mapping panels in Drosophila melanogaster. *G3: Genes, Genomes, Genetics*, **7**, 1643–1652.

Kingsolver JG, Hoekstra HE, Hoekstra JM, *et al.* (2001) The strength of phenotypic selection in natural populations. *American Naturalist*, **157**, 245–261.

Kingston SE, Martino P, Melendy M, Reed FA, Carlon DB (2018) Linking genotype to phenotype in a changing ocean: inferring the genomic architecture of a blue mussel stress response with genome-wide association. *Journal of Evolutionary Biology*, **31**, 346–361.

Knief U, Schielzeth H, Ellegren H, Kempenaers B, Forstmeier W (2015) A prezygotic transmission distorter acting equally in female and male zebra finches Taeniopygia guttata. *Molecular Ecology*, **24**, 3846–3859.

Kocabas AM, Crosby J, Ross PJ, *et al.* (2006) The transcriptome of human oocytes. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 14027–14032.

Kofler R, Schlötterer C (2012) Gowinda: Unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, **28**, 2084–2085.

Kofuji S, Kimura H, Nakanishi H, *et al.* (2015) INPP4B Is a PtdIns(3,4,5)P3 Phosphatase That Can Act as a Tumor Suppressor. *Cancer Discovery*, **5**, 730–739.

Kokko H, Jennions MD (2014) The relationship between sexual selection and sexual conflict. *Cold Spring Harbor Perspectives in Biology*, **6**, 17517–17518.

Komori T (2011) Signaling networks in RUNX2-dependent bone development. *Journal of Cellular Biochemistry*, **112**, 750–755.

Komori T, Kishimoto T (1998) Cbfa1 in bone development. *Current Opinion in Genetics and Development*, **8**, 494–499.

Kosambi DD (1943) The estimation of map distances from recombination values. *Annals of Eugenics*, **12**, 172–175.

Kotiaho JS, Simmons LW, Tomkins JL (2001) Towards a resolution of the lek paradox. *Nature*, **410**, 684–686.

Kruuk LE (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 873–890.

Kruuk LE, Clutton-Brock TH, Rose KE, Guinness FE (1999) Early determinants of lifetime reproductive success differ between the sexes in red deer. *Proceedings of the Royal Society B: Biological Sciences*, **266**, 1655–1661.

Kruuk LE, Clutton-Brock TH, Slate J, Pemberton JM, Brotherstone S, Guinness FE (2000) Heritability of fitness in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 698–703.

Kruuk LE, Hadfield JD (2007) How to separate genetic and environmental causes of similarity between relatives. *Journal of Evolutionary Biology*, **20**, 1890–1903.

Kruuk LEB, Clutton-Brock T, Pemberton JM (2014) Case study: quantitative genetics and sexual selection of weaponry in a wild ungulate. In *Quantitative genetics in the wild* (edited by A Charmantier, D Garant, LEB Kruuk), chap. 10, pp. 160–176. Oxford University Press, 1st edn..

Kruuk LEB, Slate J, Pemberton JM, Brotherstone S, Clutton-Brock TH (2002) Antler size in red deer: hertability and selection but no evolution. *Evolution*, **56**, 1683–1695.

Kruuk LEB, Slate J, Wilson AJ (2008) New Answers for Old Questions: The Evolutionary Quantitative Genetics of Wild Animal Populations. *Annual Review of Ecology Evolution and Systematics*, **39**, 525–548.

Kuijper B, Pen I, Weissing FJ (2012) A guide to sexual selection theory. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 287–311.

Küpper C, Stocks M, Risse JE, *et al.* (2015) A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, **48**, 79–83.

Lande R (1980) The genetic covariance between characters maintained by pleiotropic mutations. *Genetics*, **94**, 203–215.

Lande R (1981) Models of Speciation by Sexual Selection on Polygenic Traits. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, **78**, 3721–3725.

Lande R (1982) A Quantitative Genetic Theory of Life History Evolution. *Ecology*, **63**, 607–615.

Lande R, Arnold SJ (1983) The Measurement of Selection on Correlated Characters. *Evolution*, **37**, 1210.

Lander ES, Botstein S (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, **121**, 185.

Lehtonen J, Parker GA, Schärer L (2016) Why anisogamy drives ancestral sex roles. *Evolution*, **70**, 1129–1135.

Lewontin R (1974) *The genetic basis of evolutionary change*, vol. 560. Columbia University Press New York.

Lin CY, Lovén J, Rahl PB, *et al.* (2012) Transcriptional amplification in tumor cells with elevated c-Myc. *Cell*, **151**, 56–67.

Lin Y, Darolti I, S Furman BL, *et al.* (2021) Sexual conflcit over survival in Trinidaian guppies. *bioRxiv*.

Linge A, Maurya P, Friedrich K, *et al.* (2014) Identification and functional validation of rad23b as a potential protein in human breast cancer progression. *Journal of Proteome Research*, **13**, 3212–3222.

Liu C, He X, Liu W, *et al.* (2019) Bi-allelic Mutations in TTC29 Cause Male Subfertility with Asthenoteratospermia in Humans and Mice. *American Journal of Human Genetics*, **105**, 1168–1181.

Logan CJ, Kruuk LEB, Stanley R, Thompson AM, Clutton-Brock TH (2016) Endocranial volume is heritable and is associated with longevity and fitness in a wild mammal. *Royal Society Open Science*, **3**.

Lorch PD, Proulx S, Rowe L, Day T (2003) Condition-dependent sexual selection can accelerate adaptation. *Evolutionary Ecology Research*, **5**, 867–881.

Lorès P, Dacheux D, Kherraf ZE, *et al.* (2019) Mutations in TTC29, Encoding an Evolutionarily Conserved Axonemal Protein, Result in Asthenozoospermia and Male Infertility. *American Journal of Human Genetics*, **105**, 1148–1167.

Lucotte EA, Laurent R, Heyer E, Sé Gurel L, Toupance B (2016) Detection of Allelic Frequency Differences between the Sexes in Humans: A Signature of Sexually Antagonistic Selection. *Genome biology and evolution*, **8**, 1489–1500.

Lukefahr SD, Jacobson HA (1998) Variance Component Analysis and Heritability of Antler Traits in White-Tailed Deer. *The Journal of Wildlife Management*, **62**, 262.

Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits* . Sinauer Associates, Sunderland, Massachusetts.

Lyttle TW (1993) Cheaters sometimes prosper: distortion of mendelian segregation by meiotic drive. *Trends in Genetics*, **9**, 205–210.

MacArthur J, Bowler E, Cerezo M, *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, **45**, D896–D901.

Mackay TFC (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature Reviews Genetics*, **15**, 22–33.

Maeda A, Ono M, Holmbeck K, *et al.* (2015) WNT1-Induced Secreted Protein-1 (WISP1), a novel regulator of bone turnover and Wnt signaling. *Journal of Biological Chemistry*, **290**, 14004–14018.

Malenfant RM, Davis CS, Richardson ES, Lunn NJ, Coltman DW (2018) Heritability of body size in the polar bears of Western Hudson Bay. *Molecular Ecology Resources*, **18**, 854–866.

Malo AF, Roldan ER, Garde J, Soler AJ, Gomendio M (2005) Antlers honestly advertise sperm production and quality. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 149–157.

Mank JE (2017a) Population genetics of sexual conflict in the genomic era. *Nature Reviews Genetics*, **18**, 721–730.

Mank JE (2017b) The transcriptional architecture of phenotypic dimorphism. *Nature Ecology and Evolution*, **1**.

Mank JE, Ellegren H (2009) Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution*, **63**, 1464–1472.

Mank JE, Shu JJ, Wright AE (2020) Signature of sexual conflict is actually conflict resolved.

Manolio TA, Collins FS, Cox NJ, *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

Martin JW, Zielenska M, Stein GS, Van Wijnen AJ, Squire JA (2011) The role of RUNX2 in osteosarcoma oncogenesis.

Martínez-Ruiz C, Knell RJ (2017) Sexual selection can both increase and decrease extinction probability: reconciling demographic and evolutionary factors. *Journal of Animal Ecology*, **86**, 117–127.

Maskell LJ, Qamar K, Babakr AA, Hawkins TA, Heads RJ, Budhram-Mahadeo VS (2017) Essential but partially redundant roles for POU4F1/Brn-3a and POU4F2/Brn-3b transcription factors in the developing heart. *Cell Death & Disease*, **8**, e2861.

McNamara JM, Houston AI (2009) Integrating function and mechanism. *Trends in Ecology and Evolution*, **24**, 670–675.

Merilä J, Sheldon BC (1999) Genetic architecture of fitness and nonfitness traits: empirical patterns and development of ideas. *Heredity 1999 83:2*, **83**, 103–109.

Merila J, Sheldon BC, Kruuk LEB, Merilä J, Sheldon BC, Kruuk LEB (2001) Explaining stasis: microevolutionary studies in natural populations. *Genetica*, **112**, 199–222.

Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Miguchi M, Hinoi T, Shimomura M, *et al.* (2016) Gasdermin C is upregulated by inactivation of transforming growth factor β receptor type II in the presence of mutated Apc, promoting colorectal cancer proliferation. *PLoS ONE*, **11**.

Miller JM, Festa-Bianchet M, Coltman DW (2018) Genomic analysis of morphometric traits in bighorn sheep using the Ovine Infinium (R) HD SNP BeadChip. *Peerj*, **6**.

Moghadam HK, Pointer MA, Wright AE, Berlin S, Mank JE (2012) W chromosome expression responds to female-specific selection. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 8207–8211.

Mokkonen M, Kokko H, Koskela E, *et al.* (2011) Negative frequency-dependent selection of sexually antagonistic alleles in Myodes glareolus. *Science*, **334**, 972–974.

Møller AP, Birkhead TR (1993) Certainty of paternity covaries with paternal care in birds. *Behavioral Ecology and Sociobiology*, **33**, 261–268.

Moller AP, Jennions MD (2001) How important are direct fitness benefits of sexual selection? *Naturwissenschaften*, **88**, 401–415.

Morrissey MB, Kruuk LE, Wilson AJ (2010) The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, **23**, 2277–2288.

Morrissey MB, Walling CA, Wilson AJ, Pemberton JM, Clutton-Brock TH, Kruuk LEB (2012) Genetic Analysis of Life-History Constraint and Evolution in a Wild Ungulate Population. *American Naturalist*, **179**, E97–E114.

Nagamine Y, Pong-Wong R, Navarro P, *et al.* (2012) Localising Loci underlying Complex Trait Variation Using Regional Genomic Relationship Mapping. *Plos One*, **7**.

Nakashima H, Yasunaga M, Yoshida M, *et al.* (2021) Low Concentration of Etoposide Induces Enhanced Osteogenesis in MG63 Cells via Pin1 Activation. *Journal of Hard Tissue Biology*, **30**, 175–182.

Neff BD, Pitcher TE (2005) Genetic quality and sexual selection: An integrated framework for good genes and compatible genes. *Molecular Ecology*, **14**, 19–38.

Nussey DH, Kruuk LE, Morris A, Clements MN, Pemberton JM, Clutton-Brock TH (2009) Inter- And intrasexual variation in aging patterns across reproductive traits in a wild red deer population. *American Naturalist*, **174**, 342–357.

Oba S, Sato M, Takemasa I, Monden M, Matsubara K, Ishii S (2003) A Bayesian missing value estimation method for gene expression profile data. *Bioinformatics*, **19**, 2088–2096.

Oka Y, Bekker-Jensen S, Mailand N (2015) Ubiquitin-like protein UBL 5 promotes the functional integrity of the Fanconi anemia pathway . *The EMBO Journal*, **34**, 1385–1398.

Oka Y, Varmark H, Vitting-Seerup K, *et al.* (2014) UBL5 is essential for pre-mRNA splicing and sister chromatid cohesion in human cells. *EMBO reports*, **15**, 1330–1330.

Okada K, Katsuki M, Sharma MD, *et al.* (2021) Natural selection increases female fitness by reversing the exaggeration of a male sexually selected trait. *Nature Communications*, **12**, 1–10.

O'Sullivan RJ, Aykanat T, Johnston SE, *et al.* (2019) Evolutionary stasis of a heritable morphological trait in a wild fish population despite apparent directional selection. *Ecology and Evolution*, **9**, 7096–7111.

Pallares LF, Harr B, Turner LM, Tautz D (2014) Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. *Molecular Ecology*, **23**, 5756–5770.

Pardo-Manuel De Villena F, Sapienza C (2001) Nonrandom segregation during meiosis: The unfairness of females. *Mammalian Genome*, **12**, 331–339.

Parker GA (1979) Sexual Selection and Sexual Conflict. In *Sexual Selection and Reproductive Competition in Insects* (edited by MS Blum, NA Blum), pp. 123–166. Academic Press, INC.

Parker GA (2006) Sexual conflict over mating and fertilization: An overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 235–259.

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of Mendelian factors by using a complete map of restriction fragment length polymorphism. *Nature*, **335**, 721–726.

Pavan S, Delvento C, Ricciardi L, Lotti C, Ciani E, D'Agostino N (2020) Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies. *Frontiers in Genetics*, **0**, 447.

Pavitt AT, Walling CA, Pemberton JM, Kruuk LEB (2014) Heritability and cross-sex genetic correlations of early-life circulating testosterone levels in a wild mammal. *Biology Letters*, **10**.

Pemberton J (2008) Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 613–621.

Pemberton JM, Coltman DW, Coulson TN, Slate J (1999) Using microsatellites to measure the fitness consequences of inbreeding and outbreeding. In *Microsatellites: Evolution and Aplications* (edited by DB Goldstein, C Schlötterer), p. 151–164. Oxford University Press, Oxford.

Pérez P, De Los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, **198**, 483–495.

[dataset] Peters L, Huisman J, Kruuk LE, Pemberton JM, Johnston SE (2021) Genomic analysis reveals a polygenic architecture of antler morphology in wild red deer (Cervus elaphus). *Dryad*, p. https://doi.org/10.5061/dryad.612jm643c.

Poissant J, Davis CS, Malenfant RM, Hogg JT, Coltman DW (2012) QTL mapping for sexually dimorphic fitness-related traits in wild bighorn sheep. *Heredity*, **108**, 256–263.

Pomiankowski A, Moller AP (1995) A Resolution of the Lek Paradox. *Proceedings of the Royal Society B-Biological Sciences*, **260**, 21–29.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

Price GR (1970) Selection and Covariance. *Nature*, **227**, 520–521.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans:models and data. *American Journal of Human Genetics*, **69**, 1–14.

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**, 559–575.

R Core team (2020) R: A language and environment for statistical computing.

Randall JC, Winkler TW, Kutalik Z, *et al.* (2013) Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. *PLOS Genetics*, **9**, e1003500.

Reddiex A, Chenoweth A (2021) Integrating genomics and multivariate evolutionary quantitative genetics: A case study of constraints on sexual selection in Drosophila serrata. *Proceedings of the Royal Society B: Biological Sciences*, **in press**.

Rice WR (1984) Sex chromosomes and the evolution of sexual dimorphism. *Evolution*, **38**, 735–742.

Rice WR (1998) Male fitness increases when females are eliminated from gene pool: Implications for the Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 6217–6221.

Rice WR, Chippindale AK (2001) Intersexual ontogenetic conflict. *Journal of Evolutionary Biology*, **14**, 685–693.

Rice WR, Holland B (1997) The enemies within: Intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behavioral Ecology and Sociobiology*, **41**, 1–10.

Rice WR, Stewart AD, Morrow EH, Linder JE, Orteiza N, Byrne PG (2006) Assessing sexual conflict in the Drosophila melanogaster laboratory model system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 287–299.

Ritchie MG (2007) Sexual selection and speciation. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 79–102.

Roberts RB, Ser JR, Kocher TD (2009) Sexual conflict resolved by invasion of a novel sex determiner in lake malawi cichlid fishes. *Science*, **326**, 998–1001.

Robin X, Turck N, Hainard A, *et al.* (2011) pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**, 1–8.

Robinson MR, Pilkington JG, Clutton-Brock TH, Pemberton JM, Kruuk LE (2006) Live fast, die young: trade-offs between fitness components and sexually antagonistic selection on weaponry in Soay sheep. *Evolution*, **60**, 2168–2181.

Robinson MR, Santure AW, DeCauwer I, Sheldon BC, Slate J (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, **22**, 3963–3980.

Rönnegård L, McFarlane SE, Husby A, Kawakami T, Ellegren H, Qvarnstrom A (2016) Increasing the power of genome wide association studies in natural populations using repeated measures - evaluation and implementation. *Methods in Ecology and Evolution*, **7**, 792–799.

Rostant WG, Kay C, Wedell N, Hosken DJ (2015) Sexual conflict maintains variation at an insecticide resistance locus. *BMC Biology*, **13**, 34.

Rowe L, Chenoweth SF, Agrawal AF (2018) The genomics of sexual conflict. *American Naturalist*, **192**, 274–286.

Rowe L, Houle D (1996) The lek paradox and the capture of genetic variance by condition dependent traits. *Proceedings of the Royal Society B-Biological Sciences*, **263**, 1415–1421.

Ruzicka F, Dutoit L, Czuppon P, *et al.* (2020) The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, **4**, 398–415.

Ruzicka F, Hill MS, Pennell TM, *et al.* (2019) Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLOS Biology*, **17**, e3000244.

Santure AW, De Cauwer I, Robinson MR, Poissant J, Sheldon BC, Slate J (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Molecular Ecology*, **22**, 3949–3962.

Santure AW, Garant D (2018) Wild GWAS-association mapping in natural populations. *Molecular Ecology Resources*, **18**, 729–738.

Santure AW, Poissant J, De Cauwer I, *et al.* (2015) Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Molecular Ecology*, **24**, 6148–6162.

Sarris ME, Moulos P, Haroniti A, Giakountis A, Talianidis I (2016) Smyd3 Is a Transcriptional Potentiator of Multiple Cancer-Promoting Genes and Required for Liver and Colon Cancer Development. *Cancer Cell*, **29**, 354–366.

Schärer L, Rowe L, Arnqvist G (2012) Anisogamy, chance and the evolution of sex roles. *Trends in Ecology and Evolution*, **27**, 260–264.

Schielzeth H, Husby A (2014) Challenges and prospects in genome-wide quantitative trait loci mapping of standing genetic variation in natural populations. *Year in Evolutionary Biology*, **1320**, 35–57.

Schreiber C, Saraswati S, Harkins S, *et al.* (2019) Loss of ASAP1 in mice impairs adipogenic and osteogenic differentiation of mesenchymal progenitor cells through dysregulation of FAK/ Src and AKT signaling. *PLoS Genetics*, **15**, e1008216.

Schroeder TM, Jensen ED, Westendorf JJ (2005) Runx2: A master organizer of gene transcription in developing and maturing osteoblasts. *Birth Defects Research Part C - Embryo Today: Reviews*, **75**, 213–225.

Schwagmeyer PL, St Clair RC, Moodie JD, Lamey TC, Schnell GD, Moodie MN (1999) Species differences in male parental care in birds: A reexamination of correlates with paternity. *Auk*, **116**, 487–503.

Sella G, Barton NH (2019) Thinking about the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, **20**, 461–493.

Servedio MR, Bürger R (2014) The counterintuitive role of sexual selection in species maintenance and speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 8113–8118.

Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *American Journal of Human Genetics*, **66**, 1616–1630.

Shaw KL, Parsons YM, Lesnick SC (2007) QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket Laupala. *Molecular Ecology*, **16**, 2879–2892.

Shi H, Kichaev G, Pasaniuc B (2016) Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *The American Journal of Human Genetics*, **99**, 139–153.

Shine R (1989) Ecological causes for the evolution of sexual dimorphism: a review of the evidence. *Quarterly Review of Biology*, **64**, 419–461.

Sim Z, Coltman DW (2019) Heritability of Horn Size in Thinhorn Sheep. *Frontiers in Genetics*, **10**, 959.

Simpson SP (1989) Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. *Theoretical And Applied Genetics*, **77**, 815–819.

Slate J (2013) From beavis to beak color: A simulation study to examine how much QTL mapping can reveal about the genetic architecture of quantitative traits. *Evolution*, **67**, 1251–1262.

Slate J, Kruuk LE, Marshall TC, Pemberton JM, Clutton-Brock TH (2000) Inbreeding depression influences lifetime breeding success in a wild population of red deer (Cervus elaphus). *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1657–1662.

Slate J, Santure AW, Feulner PG, *et al.* (2010) Genome mapping in intensively studied wild vertebrate populations. *Trends in Genetics*, **26**, 275–284.

Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML, Pembertont JM (2002) A genome scan for quantitative trait loci in a wild population of red deer (Cervus elaphus). *Genetics*, **162**, 1863–1873.

Slatkin M (1984) Ecological causes of sexual dimorphism. *Evolution*, **38**, 622–630.

Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. *Theoretical and Applied Genetics*, **67**, 25–33.

Soller M, Genizi A (1978) The Efficiency of Experimental Designs for the Detection of Linkage between a Marker Locus and a Locus Affecting a Quantitative Trait in Segregating Populations. *Biometrics*, **34**, 47.

Sparks AM, Watt K, Sinclair R, *et al.* (2019) The genetic architecture of helminth-specific immune responses in a wild population of Soay sheep (Ovis aries). *PLoS Genetics*, **15**, e1008461.

Stacklies W, Redestig H (2018) Imputing missing values using the pcaMethods package.

Stephens M (2016) False discovery rates: a new deal. *Biostatistics*, **18**, 275–294.

Stopher KV, Walling CA, Morris A, *et al.* (2012) Shared Spatial Effects on Quantitative Genetic Parameters: Accounting for Spatial Autocorrelation and Home Range Overlap Reduces Estimates of Heritability in Wild Red Deer. *Evolution*, **66**, 2411–2426.

Sul JH, Martin LS, Eskin E (2018) Population structure in genetic studies: Confounding factors and mixed models. *PLoS Genetics*, **14**.

Suttie JM, Lincoln GA, Kay RN (1984) Endocrine control of antler growth in red deer stags. *Journal of reproduction and fertility*, **71**, 7–15.

Svensson EI, Gosden TP (2007) Contemporary evolution of secondary sexual traits in the wild. *Functional Ecology*, **21**, 422–433.

Svoboda P, Franke V, Schultz RM (2015) Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse. In *Current Topics in Developmental Biology*, vol. 113, pp. 305–349. Academic Press.

Tajima F (1989) Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics*, **123**, 585.

Teplitsky C, Tarka M, Møller AP, *et al.* (2014) Assessing multivariate constraints to evolution across ten long-term avian studies. *PLoS ONE*, **9**, e90444.

The Jackson Laboratory (2019) Mouse Genome Database (MGD) at the Mouse Genome Informatics website.

Thoday JM (1961) Location of Polygenes. *Nature*, **191**, 368–370.

Timpson NJ, Greenwood CM, Soranzo N, Lawson DJ, Richards JB (2018) Genetic architecture: The shape of the genetic contribution to human traits and disease.

Traulsen A, Reed FA (2012) From genes to games: Cooperation and cyclic dominance in meiotic drive. *Journal of Theoretical Biology*, **299**, 120–125.

Uffelmann E, Huang QQ, Munung NS, *et al.* (2021) Genome-wide association studies. *Nature Reviews Methods Primers 2021 1:1*, **1**, 1–21.

Van Den Berg GH, Garrick DJ (1997) Inheritance of adult velvet antler weights and live weights in farmed red deer. *Livestock Production Science*, **49**, 287–295.

Van Homrigh A, Higgie M, McGuigan K, Blows MW (2007) The depletion of genetic variance by sexual selection. *Current Biology*, **17**, 528–532.

VanRaden P (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, **91**, 4414–4423.

Visscher PM, Wray NR, Zhang Q, *et al.* (2017) 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, **101**, 5–22.

Wakchaure R, Ganguly S (2015) Marker Assisted Selection (MAS) in Animal Breeding: A Review. *Journal of Drug Metabolism & Toxicology*, **06**.

Walling CA, Pemberton JM, Hadfield JD, Kruuk LE (2010) Comparing parentage inference software: Reanalysis of a red deer pedigree. *Molecular Ecology*, **19**, 1914–1928.

Wang Y, Zhang C, Wang N, *et al.* (2019) Genetic basis of ruminant headgear and rapid antler regeneration. *Science (New York, N.Y.)*, **364**, eaav6335.

Wang Z, Yang RC, Goonewardene LA, Huedepohl C (1999) Genetic analysis of velvet antler yield in farmed elk (Cervus elaphus). *Canadian Journal of Animal Science*, **79**, 569–571.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, **38**, 1358.

Weller JI, Kashi Y, Soller M (1990) Power of Daughter and Granddaughter Designs for Determining Linkage Between Marker Loci and Quantitative Trait Loci in Dairy Cattle. *Journal of Dairy Science*, **73**, 2525–2537.

Whalen A, Ros-Freixedes R, Wilson DL, Gorjanc G, Hickey JM (2018) Hybrid peeling for fast and accurate calling, phasing, and imputation with sequence data of any coverage in pedigrees. *Genetics Selection Evolution*, **50**, 67.

Wigby S, Chapman T (2005) Sex peptide causes mating costs in female Drosophila melanogaster. *Current Biology*, **15**, 316–321.

Wilkinson GS, Breden F, Mank JE, *et al.* (2015) The locus of sexual selection: Moving sexual selection studies into the post-genomics era. *Journal of Evolutionary Biology*, **28**, 739–755.

Williams JD, Krueger WF, Harmel DH (1994) Heritabilities for antler characteristics and body weight in yearling white-tailed deer. *Heredity*, **73**, 78–83.

Wood AR, Esko T, Yang J, *et al.* (2014) Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics 2014 46:11*, **46**, 1173–1186.

Wray NR, Yang J, Goddard ME, Visscher PM (2010) The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genetics*, **6**, 1000864.

Wright AE, Fumagalli M, Cooney CR, *et al.* (2018) Male-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evolution Letters*, **2**, 52–61.

Wright AE, Rogers TF, Fumagalli M, Cooney CR, Mank JE (2019) Phenotypic sexual dimorphism is associated with genomic signatures of resolved sexual conflict. *Molecular Ecology*, **28**, 2860–2871.

Xu S (2003) Theoretical Basis of the Beavis Effect. *Genetics*, **165**, 2259–2268.

Yang J, Benyamin B, McEvoy BP, *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565–569.

Yang J, Manolio TA, Pasquale LR, *et al.* (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, **43**, 519–U44.

Yang X, Schadt EE, Wang S, *et al.* (2006) Tissue-specific expression and regulation of sexually dimorphic genes in mice. *Genome Research*, **16**, 995–1004.

Yeh ES, Means AR (2007) PIN1, the cell cycle and cancer. *Nature Reviews Cancer*, **7**, 381–388.

Yengo L, Sidorenko J, Kemper KE, *et al.* (2018) Meta-analysis of genome-wide association studies for height and body mass index in ˜700 000 individuals of European ancestry. *Human Molecular Genetics*, **27**, 3641–3649.

Supplemental Information

Chapter 2

Supplemental Figures

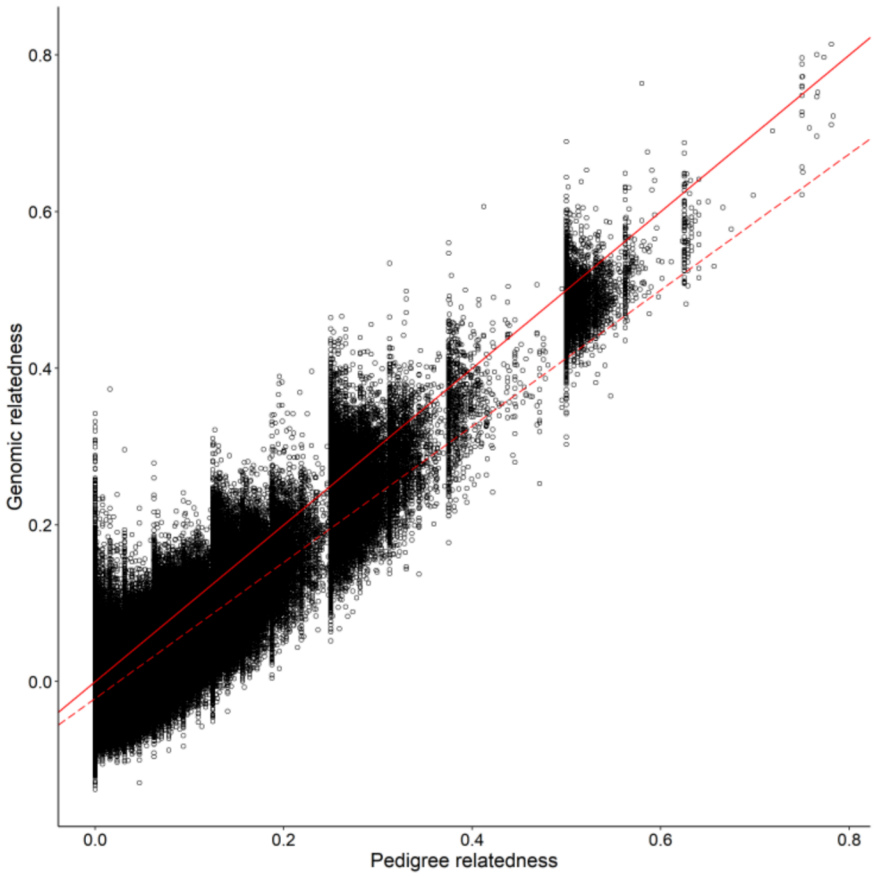


Figure S2.1: Pedigree plotted against genomic pairwise relatedness. The dashed red line represents a fitted line from a linear regression, the solid line represents a 1:1 relationship.

Supplemental Tables

Table S2.1: Phenotypic correlations among all 10 antler measurements. Pearson's product moment correlation coefficient is given by *r*, the t-test statisitc by *t* and the associated degrees of freedom and p value by *DF* and *p-value*, respectively. The 95 % confidence interval of the correlation coefficient is indicated by columns *Conf.Lower* (for the lower limit of the confidence interval) and *Conf.Upper* (for the upper limit of the confidence interval).

Comparison	t	DF	p-value	r	Conf.Lower	Conf.Upper
Length_CoronetCirc	31.33	1057	5.94E-153	0.69	0.66	0.72
Length_LowerBeam	34.58	1045	2.66E-175	0.73	0.70	0.76
Length_UpperBeam	29.55	1007	1.11E-138	0.68	0.65	0.71
Length_CoronetBrowJunc	17.15	1009	5.00E-58	0.47	0.43	0.52
Length_CoronetTrayJunc	21.71	1034	2.09E-86	0.56	0.52	0.60
Length_BrowLength	28.11	988	3.15E-128	0.67	0.63	0.70
Length_TrayLength	18.68	864	1.08E-65	0.54	0.49	0.58
Length_AntlerWt	38.30	682	3.88E-172	0.83	0.80	0.85
Length_Form	15.99	1007	1.85E-51	0.45	0.40	0.50
CoronetCirc_LowerBeam	35.09	1198	4.23E-186	0.71	0.68	0.74
CoronetCirc_UpperBeam	26.37	1117	1.45E-119	0.62	0.58	0.65
CoronetCirc_CoronetBrowJunc	10.68	1160	1.86E-25	0.30	0.25	0.35
CoronetCirc_CoronetTrayJunc	18.85	1182	1.68E-69	0.48	0.44	0.52
CoronetCirc_BrowLength	29.27	1136	8.38E-141	0.66	0.62	0.69
CoronetCirc_TrayLength	20.75	941	5.11E-79	0.56	0.51	0.60
CoronetCirc_AntlerWt	29.70	743	2.22E-128	0.74	0.70	0.77
CoronetCirc_Form	17.88	1146	3.13E-63	0.47	0.42	0.51
LowerBeam_UpperBeam	55.10	1120	3.38E-321	0.85	0.84	0.87
LowerBeam_CoronetBrowJunc	15.87	1145	2.03E-51	0.42	0.38	0.47
LowerBeam_CoronetTrayJunc	10.69	1179	1.65E-25	0.30	0.24	0.35
LowerBeam_BrowLength	26.35	1119	1.84E-119	0.62	0.58	0.65
LowerBeam_TrayLength	27.08	942	6.33E-120	0.66	0.62	0.70
LowerBeam_AntlerWt	34.91	739	1.69E-158	0.79	0.76	0.81
LowerBeam_Form	22.16	1128	1.34E-90	0.55	0.51	0.59
UpperBeam_CoronetBrowJunc	12.07	1072	1.55E-31	0.35	0.29	0.40
UpperBeam_CoronetTrayJunc	11.05	1112	5.20E-27	0.31	0.26	0.37
UpperBeam_BrowLength	21.86	1044	1.50E-87	0.56	0.52	0.60
UpperBeam_TrayLength	21.98	889	7.14E-86	0.59	0.55	0.63
UpperBeam_AntlerWt	38.27	713	4.70E-175	0.82	0.79	0.84
UpperBeam_Form	25.19	1053	5.48E-110	0.61	0.57	0.65
CoronetBrowJunc_CoronetTrayJunc	13.73	1129	9.06E-40	0.38	0.33	0.43
CoronetBrowJunc_BrowLength	9.06	1081	5.85E-19	0.27	0.21	0.32
CoronetBrowJunc_TrayLength	5.91	890	4.80E-09	0.19	0.13	0.26
CoronetBrowJunc_AntlerWt	10.29	705	3.18E-23	0.36	0.30	0.42
CoronetBrowJunc_Form	7.53	1086	1.07E-13	0.22	0.17	0.28
CoronetTrayJunc_BrowLength	14.49	1103	1.06E-43	0.40	0.35	0.45
CoronetTrayJunc_TrayLength	1.30	940	1.95E-01	0.04	-0.02	0.11
CoronetTrayJunc_AntlerWt	13.68	728	4.52E-38	0.45	0.39	0.51
CoronetTrayJunc_Form	2.45	1111	1.43E-02	0.07	0.01	0.13

BrowLength_TrayLength	23.80	926	4.32E-98	0.62	0.57	0.65
BrowLength_AntlerWt	23.45	697	4.41E-90	0.66	0.62	0.70
BrowLength_Form	17.65	1068	2.14E-61	0.48	0.43	0.52
TrayLength_AntlerWt	20.55	562	2.17E-70	0.65	0.61	0.70
TrayLength_Form	19.32	945	2.49E-70	0.53	0.48	0.58
AntlerWt_Form	20.72	846	1.97E-77	0.58	0.53	0.62

Table S2.2: Heritability for all 10 antler measures and 11 PCs. Both estimates from the pedigree and the GRM models are shown. h^2, *SE* and *p-value* refer to the heritability, its standard error and p value, respectively.

Antler Measure	**Pedigree h^2**	**SE**	**p-value**	**GRM h^2**	**SE**	**P-value**
Antler Length	0.316	0.073	3.69E-07	0.279	0.067	2.47E-07
Coronet Circ.	0.423	0.075	2.92E-10	0.408	0.071	1.64E-10
Lower Beam Circ.	0.326	0.073	6.32E-08	0.386	0.067	1.24E-10
Upper Beam Circ.	0.345	0.089	1.81E-05	0.357	0.087	4.86E-05
Coronet-Brow Junc.	0.295	0.078	9.44E-07	0.275	0.073	1.43E-06
Coronet-Tray Junc.	0.382	0.093	4.50E-05	0.410	0.083	1.92E-07
Brow Length	0.436	0.072	1.17E-13	0.414	0.066	2.49E-15
Tray Length	0.328	0.109	0.002	0.388	0.094	8.65E-06
Antler Weight	0.211	0.091	0.006	0.229	0.089	0.002
Form	0.310	0.049	5.17E-19	0.268	0.052	6.57E-12
PC1	0.348	0.081	1.53E-06	0.365	0.075	1.13E-07
PC2	0.201	0.086	0.005	0.255	0.082	3.626E-04
PC3	0.273	0.076	4.14E-06	0.296	0.073	4.52E-07
PC4	0.237	0.076	2.655E-04	0.179	0.068	0.001
PC5	0.371	0.092	5.17E-06	0.314	0.085	1.13E-05
PC6	0.324	0.083	1.16E-05	0.264	0.078	1.419E-04
PC7	0.380	0.082	3.59E-08	0.365	0.074	2.72E-09
PC8	0.154	0.069	0.007	0.141	0.064	0.008
PC9	0.186	0.080	0.0.012	0.189	0.072	0.002
PC10	0.134	0.063	0.011	0.142	0.055	9.142E-04
PC11	0.106	0.048	0.002	0.124	0.050	4.658E-04

Table S2.3: Estimates of fixed effects from the animal models estimating the heritability of antler measures using the GRM. The fixed effect is indicated by variable, where *MeCaAge* refers to an animal's age during antler growth and *MeCaAge.2* indicates the squared effect of age. *SE* refers to the standard error of the effect size and the effect size standardized by the mean value of the respective antler measure is indicated by *Effect Mean Std.*. The Z ratio statistics and associated p value for each fixed effect are given by *Z ratio* and *p-value*, respectively. Trait refers to the antler measure (phenotype) modelled.

Trait	Variable	Effect Size	SE	Z Ratio	p-value	Effect Mean Std
Antler Length	Intercept	12.731	1.567	8.122	0	0.195
	Age	10.919	0.377	28.941	0	0.168
	Age^2	-0.478	0.023	-20.978	0	-0.007
Coronet Circ.	Intercept	7.451	0.266	28.048	0	0.477
	Age	1.907	0.058	32.626	0	0.122
	Age^2	-0.096	0.004	-26.884	0	-0.006
Lower Beam Circ.	Intercept	4.419	0.195	22.649	0	0.421
	Age	1.309	0.046	28.416	0	0.125
	Age^2	-0.059	0.003	-21.094	0	-0.006
Upper Beam Circ.	Intercept	4.505	0.168	26.791	0	0.477
	Age	0.987	0.039	25.445	0	0.104
	Age^2	-0.040	0.002	-17.453	0	-0.004
Coronet-Brow Junc.	Intercept	3.561	0.160	22.200	0	0.623
	Age	0.407	0.038	10.697	0	0.071
	Age^2	-0.016	0.002	-6.925	4.354E-12	-0.003
Coronet-Tray Junc.	Intercept	13.185	0.804	16.398	0	0.483
	Age	2.838	0.208	13.642	0	0.104

	Age^2	-0.116	0.013	-9.132	0	-0.004
Brow Length	Intercept	2.470	0.767	3.222	0	0.120
	Age	3.828	0.174	22.056	0	0.186
	Age^2	-0.169	0.011	-16.026	0	-0.008
Tray Length	Intercept	2.017	0.690	2.925	0	0.139
	Age	2.649	0.172	15.400	0	0.183
	Age^2	-0.118	0.010	-11.536	0	-0.008
Antler Weight	Intercept	-336.438	32.885	-10.231	0	-0.518
	Age	205.130	7.587	27.037	0	0.316
	Age^2	-8.811	0.482	-18.286	0	-0.014
Form	Intercept	2.894	0.064	45.461	0	0.643
	Age	0.296	0.010	28.394	0	0.066
	Age^2	-0.008	0.001	-14.528	0	-0.002

Chapter 3

Supplemental Figures

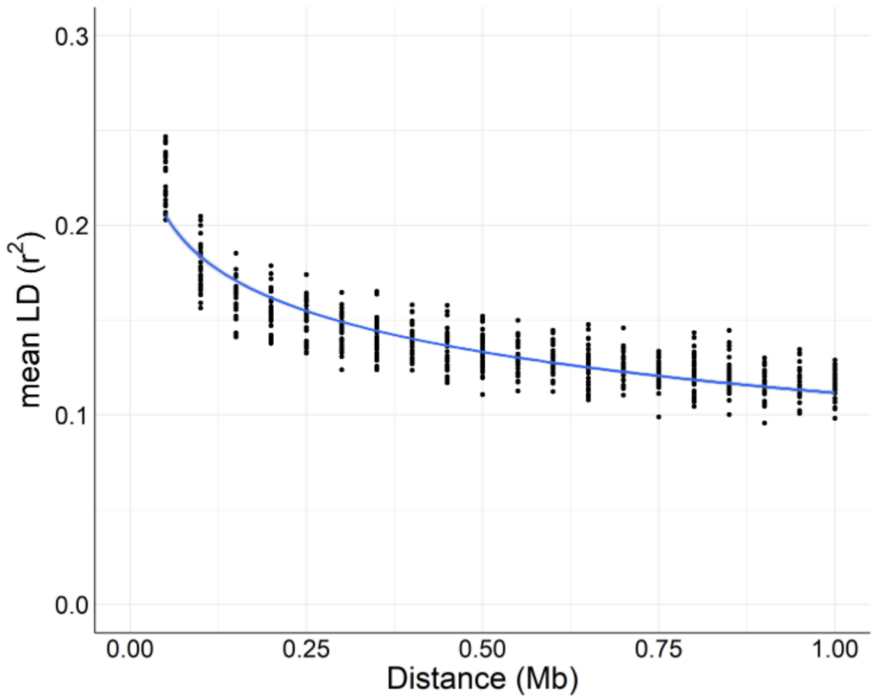


Figure S3.1: Genomic LD decay over a 1 Mb distance. Each point represents an average r^2 taken over 0.05 Mb increments and for each linkage group separately.

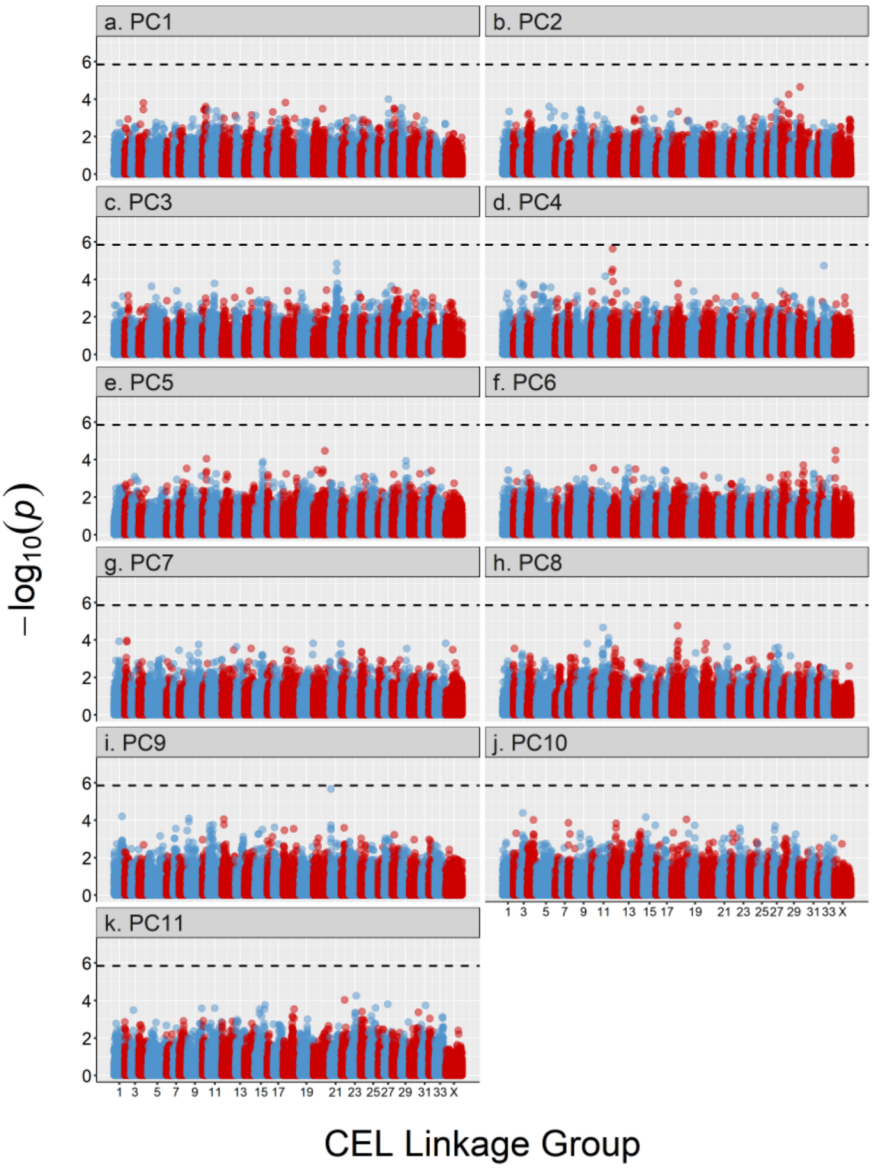


Figure S3.2: Genome-wide association study for antler PCs. The dashed line indicates the threshold for statistical significance after multiple testing (equivalent to $\alpha = 0.05$). Points are colour-coded by chromosome. Association statistics were corrected for population structure using the genomic control parameter λ. Underlying data is provided in Table S3.2.

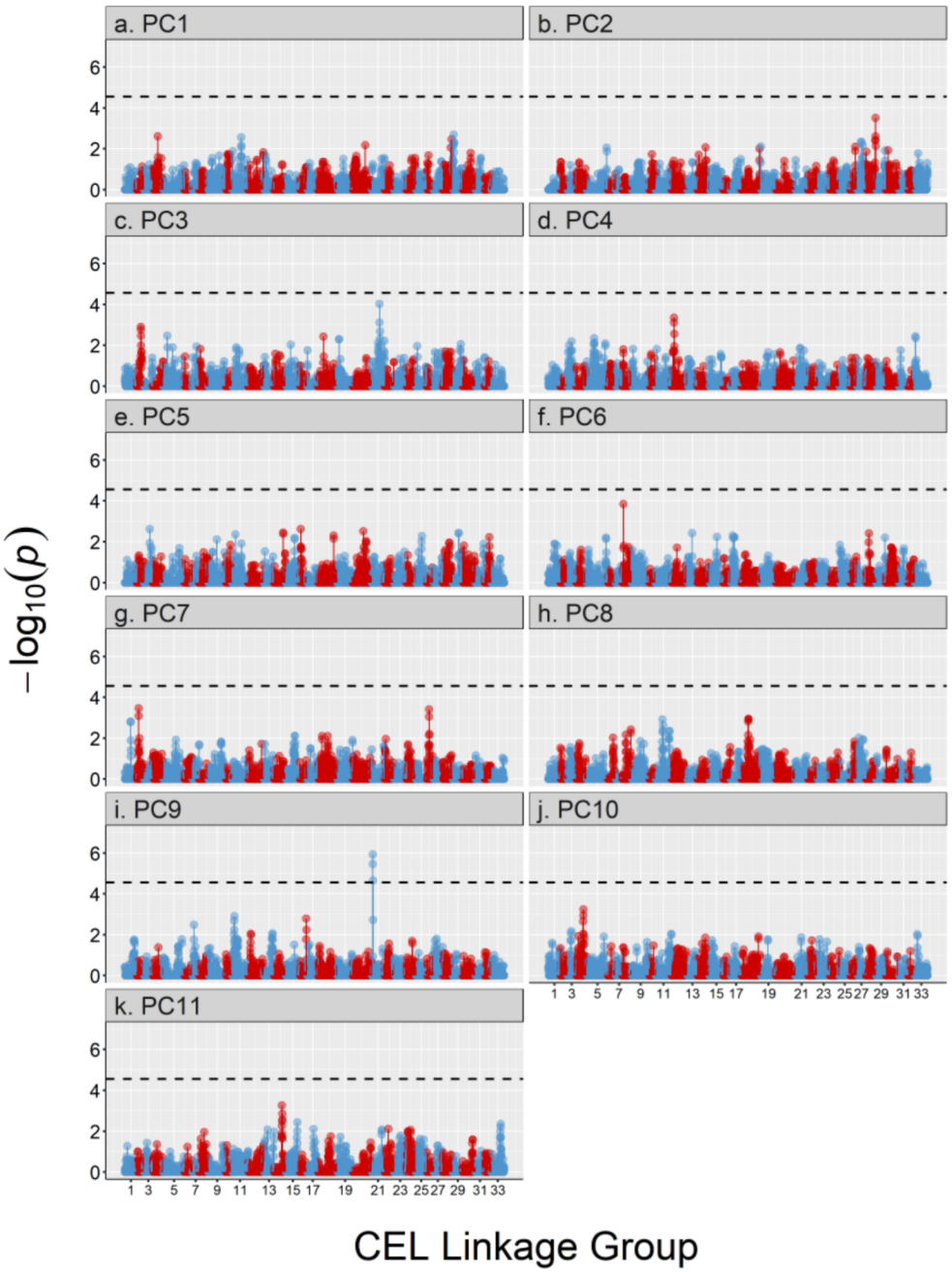


Figure S3.3: Regional heritability analyses for antler PCs. The dashed line indicates the significance threshold, equivalent to α = 0.05. Points are colour-coded by chromosome. Underlying data is provided in Table S3.5

Supplemental Tables

Table S3.1: Complete results for all GWAS of all 10 antler measures. SNP is the name of the SNP marker. Chromosome is the red deer linkage group. Position is the linkage map cM position. A1 and A2 are the major allele and minor alleles, respectively. Measure is the antler measure (phenotype) modelled. effB is the odds ratio estimate of SNP, and se_effB = standard error of this estimate. chi2.1df is the chi^2 estimate for SNP effect with 1 degree of freedom with the p-value P1df; chi1.1df_corrected and Pc1df are the same, but corrected by inflation factor λ. NoMeasured is the number of genotypes scored. CallRate is the proportion of genotypes scored in the full genomic dataset for each locus. Q.2 is the minor allele frequency. P.11 is the major allele homozygote genotype count, P.12 is the heterozygote genotype count and P.22 is the minor allele homozygote genotype count. Pexact is the p-value for Hardy-Weinberg-Equilibrium. Fmax is the direction of deviation from Hardy-Weinberg-Equilibrium. lambda is the genomic inflation factor λ for that phenotype measure.

Table S3.2: Complete results for all GWAS of all PCs. SNP is the name of the SNP marker. Chromosome is the red deer linkage group. Position is the linkage map cM position. A1 and A2 are the major allele and minor alleles, respectively. Measure is the PC (phenotype) modelled. effB is the odds ratio estimate of SNP, and se_effB = standard error of this estimate. chi2.1df is the chi^2 estimate for SNP effect with 1 degree of freedom with the p-value P1df; chi1.1df_corrected and Pc1df are the same, but corrected by inflation factor λ. NoMeasured is the number of genotypes scored. CallRate is the proportion of genotypes scored in the full genomic dataset for each locus. Q.2 is the minor allele frequency. P.11 is the major allele homozygote genotype count, P.12 is the heterozygote genotype count and P.22 is the minor allele homozygote genotype count. Pexact is the p-value for Hardy-Weinberg-Equilibrium. Fmax is the direction of deviation from Hardy-Weinberg-Equilibrium. lambda is the genomic inflation factor λ for that phenotype measure.

.

Table S3.3: Summary of GWAS power analysis. Power is the proportion of simulations that resulted in one or more significant GWAS hit, out of 100 simulations. The SNP heritability (SNP h^2) is an indicator of the simulated QTL effect size. SNP proportion Va represents the proportion of additive genetic variance Va explained by the simulated QTL.

Antler measure	Power	SNP h^2	SNP proportion Va
Antler Length	0.05	0.01	0.035
Antler Length	0.50	0.05	0.15
Antler Length	0.78	0.10	0.26
Antler Length	0.99	0.40	0.59
Coronet Circ.	0.03	0.01	0.024
Coronet Circ.	0.62	0.05	0.11
Coronet Circ.	0.86	0.10	0.2
Coronet Circ.	0.99	0.40	0.49
Lower Beam Circ.	0.04	0.01	0.025
Lower Beam Circ.	0.61	0.05	0.11
Lower Beam Circ.	0.89	0.10	0.21
Lower Beam Circ.	0.99	0.40	0.51
Upper Beam Circ.	0.06	0.01	0.027
Upper Beam Circ.	0.59	0.05	0.12
Upper Beam Circ.	0.90	0.10	0.22
Upper Beam Circ.	1.00	0.40	0.53
Coronet Brow Junc.	0.03	0.01	0.035
Coronet Brow Junc.	0.49	0.05	0.15
Coronet Brow Junc.	0.82	0.10	0.27
Coronet Brow Junc.	0.99	0.40	0.59
Coronet Tray Junc.	0.03	0.01	0.024
Coronet Tray Junc.	0.40	0.05	0.11
Coronet Tray Junc.	0.71	0.10	0.2
Coronet Tray Junc.	0.98	0.40	0.49
Brow Length	0.06	0.01	0.024
Brow Length	0.47	0.05	0.11
Brow Length	0.90	0.10	0.19
Brow Length	0.99	0.40	0.49
Tray Length	0.00	0.01	0.025
Tray Length	0.42	0.05	0.11
Tray Length	0.64	0.10	0.2
Tray Length	0.98	0.40	0.51
Antler Weight	0.05	0.01	0.042
Antler Weight	0.51	0.05	0.18
Antler Weight	0.86	0.10	0.3
Antler Weight	1.00	0.40	0.64
Form	0.23	0.01	0.036
Form	0.92	0.05	0.16
Form	0.99	0.10	0.27

Form	1.00	0.40	0.6

Table S3.4: Complete results from the regional heritability analysis for all antler measures. The table shows results for all variance component per model (full model including genomic window and relatedness matrix and partial model excluding genomic window). Model indicates partial or full model; model_variable describes the random effect (MeCaYear refers to the rut year, BirthYear to birth year, ide(ID) to the permanent environment, giv(ID) to the relatedness matrix without the genomic window, giv(ID2) to the genomic window and R to the residual variance). The scaling parameter for the variance components is given by gamma. The variance components and associated standard errors on a gamma scale are given by component and std.error, respectively. The associated Z ratio statistic is given by z.ratio. The constraint indicates convergence status. The Effect and SE refer to the variance component on the original data scale, presented as a proportion of total phenotypic variance, and the associated standard error, respectively. The model log-likelihood is given by loglik and the p-value of the log-likelihood ratio test between the full and partial model of each genomic window and PC by p_value. Trait indicates the antler measure (phenotype) modelled. The window_name refers to the genomic window ID. The uncorrected chi squared value associated with the log-likelihood ratio test is given by chisq, lambda indicates the genomic inflation factor and chisq_corr and p_corrected refer to the chi squared statistic and associated p-value corrected for genomic inflation.

Table S3.5: Complete results from the regional heritability analysis for all PCs. The table shows results for all variance component per model (full model including genomic window and relatedness matrix and partial model excluding genomic window). Model indicates partial or full model; model_variable describes the random effect (MeCaYear refers to the rut year, BirthYear to birth year, ide(ID) to the permanent environment, giv(ID) to the relatedness matrix without the genomic window, giv(ID2) to the genomic window and R to the residual variance). The scaling parameter for the variance components is given by gamma. The variance components and associated standard errors on a gamma scale are given by component and std.error, respectively. The associated Z ratio statistic is given by z.ratio. The constraint indicates convergence status. The Effect and SE refer to the variance component on the original data scale, presented as a proportion of total phenotypic variance, and the associated standard error, respectively. The model log-likelihood is given by loglik and the p-value of the log-likelihood ratio test between the full and partial model of each genomic window and PC by p_value. Trait indicates the PC (phenotype) modelled. The window_name refers to the genomic window ID. The uncorrected chi squared value associated with the log-likelihood ratio test is given by chisq, lambda indicates the genomic inflation factor and chisq_corr and p_corrected refer to the chi squared statistic and associated p-value corrected for genomic inflation

Table S3.6: All SNPs within the region covered by the three genomic windows significantly associated with PC9 in the regional heritability analysis. The SNP ID is given by SNP.name, BTA.Chr and BTA.Position refer to the cattle genome (version BTA vUMD 3.0) chromosome and base pair position, respectively. CEL.LG and CEL.order indicate the red deer linkage group and associated SNP order, respectively. The Estimated Mb.Position refers to an estimate of the SNP position in Mb in relation to the red deer genome. The position in base pairs and associated chromosome in relation to the current version of the cattle genome (ARS-UCD1.2) are given by cow_ARS_UCD1.2_pos and cow_ARS_UCD1.2_Chr, respectively. CallRate is the proportion of genotypes scored in the full genomic dataset for each locus. Q.2 is the minor allele frequency. The ID of the genomic windows containing the relevant SNP is given by window1_ID and window2_ID, if a SNP is present in more than one genomic window, as windows overlap by 10 SNPs.

Table S3.7: Complete list of all genes found in region covered by the three genomic windows significantly associated with PC9 in regional heritability analysis homologous to the the cattle genome (version ARS-UCD1.2). The gene ID is given by external_gene_name, and a long version of the gene ID is given by description. The start and end positions of the gene in base pairs are given by transcript_start and transcript_end, respectively. The Gene Ontology (GO) ID is given by go_id, name_1006 refers to the associated GO term and description_1006 is the associated description of that term.

Table S3.8: Summary of SNPs with non-zero effects on antler measures using data sets pruned for LD. N SNPs is the number of SNPs with non-zero effects. Proportion pleiotropic is the proportion of total SNPs of each measure with non-zero effects on other antler measures. Maximum and minimum effects are given relative to the data scale (units = cm for lengths, g for weight). Lower and upper quantiles refer to the 5% and 95% boundaries of the effect size distribution, respectively. The threshold value for LD for each of the three pruned data set is given by r^2. LD was pruned in windows of 20 markers that overlapped by 10 SNPs.

Antler measure	N SNPs	Proportion Pleiotropic	Minimum Effect	Maximum Effect	Lower Quantile	Upper Quantile	r^2
Brow Length	95	0.05	-0.52	0.56	-0.50	0.49	0.20
Coronet Brow Junc.	11	0.09	-0.08	0.09	-0.08	0.09	0.20
Coronet Circ	16	0.25	-0.22	0.17	-0.18	0.16	0.20
Coronet Tray Junc.	27	0.04	-0.48	0.46	-0.45	0.44	0.20
Form	62	0.10	-0.09	0.85	-0.09	0.09	0.20
Antler Length	3		-0.66	0.62	-0.66	0.49	0.20
Lower Beam Circ.	97	0.12	-0.13	0.55	-0.12	0.13	0.20
Tray Length	3	0.33	-0.39	0.39	-0.39	0.31	0.20
Upper Beam Circ.	3	0.67	-0.10	0.10	-0.08	0.10	0.20
Brow Length	237	0.08	-0.55	0.55	-0.50	0.49	0.50
Coronet Brow Junc.	25	0.08	-0.08	0.09	-0.08	0.09	0.50
Coronet Circ.	56	0.21	-0.35	0.21	-0.23	0.17	0.50
Coronet Tray Junc.	94	0.06	-0.48	0.43	-0.42	0.42	0.50
Form	168	0.14	-0.10	1.02	-0.09	0.10	0.50
Antler Length	8	0.38	-0.64	0.66	-0.64	0.23	0.50
Lower Beam Circ.	269	0.14	-0.13	0.55	-0.11	0.13	0.50
Tray Length	5	0.60	-0.35	0.33	-0.34	0.33	0.50
Upper Beam Circ.	2	0.50	-0.11	0.12	-0.10	0.11	0.50
Brow Length	266	0.08	-0.54	0.54	-0.49	0.48	0.80
Coronet Brow Junc.	30	0.07	-0.08	0.09	-0.08	0.09	0.80
Coronet Circ	53	0.23	-0.16	0.15	-0.15	0.15	0.80
Coronet Tray Junc.	112	0.06	-0.49	0.45	-0.43	0.44	0.80
Form	195	0.13	-0.10	1.02	-0.09	0.10	0.80
Antler Length	12	0.33	-0.62	0.67	-0.62	0.66	0.80
Lower Beam Circ.	284	0.14	-0.13	0.55	-0.11	0.13	0.80
Tray Length	6	0.67	-0.37	0.37	-0.36	0.36	0.80

Table S3.9: Summary of SNPs with non-zero effects on antler PCs using data sets pruned for LD. N SNPs is the number of SNPs with non-zero effects. Maximum and minimum effects are given relative to the data scale (units = cm for lengths, g for weight). Lower and upper quantiles refer to the 5% and 95% boundaries of the effect size distribution, respectively. N SNPs unique refers to the number of non-zero SNPs not shared with any of the 10 antler measures (based on data pruned for the same LD threshold value). The threshold value for LD for each of the three pruned data set is given by r^2. LD was pruned in windows of 20 markers that overlapped by 10 SNPs.

Antler PC	N SNPs	Maximum Effect	Minimum Effect	Lower Quantile	Upper Quantile	N SNPs unique	r^2
PC1	3	-0.21	0.20	-0.21	0.16	0.00	0.20
PC11	30	-0.04	0.04	-0.03	0.04	30.00	0.20
PC2	13	-0.12	0.11	-0.12	0.11	5.00	0.20

PC3	9	-0.10	0.11	-0.10	0.11	7.00	0.20
PC5	2	-0.08	-0.07	-0.08	-0.07	0.00	0.20
PC7	4	-0.06	0.06	-0.06	0.06	4.00	0.20
PC1	10	-0.18	0.18	-0.18	0.17	10.00	0.50
PC10	3	-0.04	0.04	-0.04	0.03	3.00	0.50
PC11	57	-0.03	0.04	-0.03	0.03	0.00	0.50
PC2	33	-0.11	0.11	-0.11	0.11	33.00	0.50
PC3	19	-0.10	0.11	-0.10	0.10	19.00	0.50
PC5	10	-0.08	0.07	-0.07	0.07	10.00	0.50
PC6	2	-0.05	0.06	-0.05	0.06	2.00	0.50
PC7	10	-0.06	0.06	-0.06	0.06	10.00	0.50
PC8	1	-0.05	-0.05	-0.05	-0.05	1.00	0.50
PC1	10	-0.17	0.17	-0.17	0.17	2.00	0.80
PC10	5	-0.04	0.04	-0.04	0.03	5.00	0.80
PC11	60	-0.03	0.03	-0.03	0.03	60.00	0.80
PC2	36	-0.12	0.11	-0.11	0.11	16.00	0.80
PC3	27	-0.10	0.11	-0.10	0.11	19.00	0.80
PC5	13	-0.07	0.07	-0.07	0.07	7.00	0.80
PC6	2	0.06	0.06	0.06	0.06	2.00	0.80
PC7	8	-0.06	0.06	-0.06	0.05	8.00	0.80
PC8	1	-0.05	-0.05	-0.05	-0.05	1.00	0.80

Table S3.10: Summary of SNPs with non-zero effects on antler PCs. N SNPs is the number of SNPs with non-zero effects. Maximum and minimum effects are given relative to the data scale. Lower and upper quantiles refer to the 5% and 95% boundaries of the effect size distribution, respectively.N SNPs unique refers to the number of non-zero SNPs not shared with any of the 10 antler measures.

Antler PC	N SNPs	Maximum Effect	Minimum Effect	Lower Quantile	Upper Quantile	N SNPs unique
PC1	10	0.165	-0.171	-0.170	0.164	2
PC2	34	0.114	-0.118	-0.112	0.111	16
PC3	26	0.108	-0.104	-0.100	0.106	19
PC4	0	-	-	-	-	-
PC5	13	0.072	-0.074	-0.070	0.070	7
PC6	2	0.056	0.054	0.054	0.056	2
PC7	8	0.055	-0.056	-0.055	0.054	8
PC8	1	-0.047	-0.047	-0.047	-0.047	1
PC9	0	-	-	-	-	-
PC10	5	0.043	-0.044	-0.043	0.026	5
PC11	60	0.035	-0.034	-0.033	0.034	60

Table S3.11: Full results for Empirical Bayes FDR correction on GWAS results for all 10 antler measures. The table shows the original estimates for effect size per SNP, as well as the posterior mean or corrected estimate and associated probabilities and test statistics regarding confidence in the sign of the posterior estimate. SNP.name = name of SNP marker; measure = antler measure (phenotype); betahat and sebetahat = estimate of SNP effect size and associated standard error from GWAS; NegativeProb = probability of effect size being negative; PositiveProb = probability of effect size being positive; lfsr = local false sign rate (used for significance); svalue = S-value (binary log of p-value); lfdr = local false discovery rate; qvalue = q-value; PosteriorMean and PosteriorSD = posterior mean and associated standard deviation of SNP effect size; lambda = estimate of genome-wide inflation factor lambda.

Table S3.12: Full results for Empirical Bayes FDR correction on GWAS results for all PCs. The table shows the original estimates for effect size per SNP, as well as the posterior mean or corrected estimate and associated probabilities and test statistics regarding confidence in the sign of the posterior estimate. SNP.name = name of SNP marker; measure = antler measure (phenotype); betahat and sebetahat = estimate of SNP effect size and associated standard error from GWAS; NegativeProb = probability of effect size being negative; PositiveProb = probability of effect size being positive; lfsr = local false sign rate (used for significance); svalue = S-value (binary log of p-value); lfdr = local false discovery rate; qvalue = q-value; PosteriorMean and PosteriorSD = posterior mean and associated standard deviation of SNP effect size; lambda = estimate of genome-wide inflation factor lambda.

Table S3.13: Summary of SNPs with the highest additive genetic effect sizes for antler PCs. The Wald statistics and associated p-value are taken from the animal models fitting SNP genotypes as categorical variable. The additive effect (*a*) is taken from the calculation of SNP variance based on (Falconer & Mackay, 1996). The phenotypic and genetic proportion refer to the proportion of total genetic and phenotypic variance attributed to the SNP.

Antler PC	**SNP Name**	Wald statistic	**p-value**	a	V_{SNP}	**SE**	**Prop.** V_A	**Prop.** V_P
PC1	cela1_red_11_60390816	10.581	0.005	0.521	0.179	0.083	0.069	0.032
PC2	cela1_red_9_60396904	24.915	3.888E-06	0.436	0.098	0.032	0.150	0.064
PC3	cela1_red_17_48010785	16.242	2.972E-04	0.333	0.084	0.029	0.228	0.071
PC4	-	-	-	-	-	-	-	-
PC5	cela1_red_8_39886627	19.413	6.089E-05	0.379	0.002	0.012	0.010	0.003
PC6	cela1_red_x_62877526	19.371	1.076E-05	0.177	0.013	0.011	0.056	0.020
PC7	cela1_red_29_16028833	13.399	0.001	0.230	1.476E-04	0.007	0.001	2.982E-04
PC8	cela1_red_4_41756873	22.019	1.655E-05	0.194	0.041	0.008	0.585	0.089
PC9	-	-	-	-	-	-	-	-
PC10	cela1_red_18_43824609	21.139	2.568E-05	0.153	0.024	0.005	0.444	0.093
PC11	cela1_red_5_85872147	15.170	5.081E-04	0.145	3.219E-04	0.002	0.018	0.002

Table S3.14: Complete results from categorical mixed animal model estimating the effect and variance of the top SNPs relating the antler measures after Bayesian FDR. *Variable* = fixed effect of SNP split into factor levels (genotypes), with A indicating the reference allele and B the alternative allele; *solution* = effect size of factor level; *SE* = standard error of effect size; *z_ratio* and *p.z_ratio* = Z ratio and associated p-value for factor level respectively; *measure* = antler measure; *SNP.name* = name of SNP marker; *effect.solution* = solution for factor level (genotype) if homozygote with largest (positive) effect size is the intercept level; *effect* = genotype coded using A as increasing and B as reducing allele; *neg.allele* = allele with reducing effect coded as base pair; *Df* = degrees of freedom for SNP effect; *Sum.of.Sq* = Sum of Squares of SNP effect; *Wald.statistic* = Wald statistic for SNP effect; *Pr.Chisq.* = Wald statistic associated p-value; *major.allele* = major allele of SNP coded as base pair; *minor.allele* = minor allele of SNP coded as base pair; *minor.allele.freq* = frequency of minor allele; *additive_effect* = additive genetic effect; *domiance_effect* = dominance deviation; *SNP.var* and *SNP.var.SE* = SNP variance and associated standard error, estimated as described in the methods under section 'Estimation of SNP effect size distribution' respectively; *SNP.prop.gen* = proportion of total additive genetic variance (of measure) attributed to SNP; *SNP.prop.phen* = proportion of total phenotypic variance (of measure) attributed to SNP; *Var.add* and *SE.var.add* = total additive genetic variance (of measure) and associated standard error, respectively; *Var.total* = total phenotypic variance (of measure).

Table S3.15: Complete results from categorical mixed animal model estimating the effect and variance of the top SNPs relating to antler PCs after Bayesian FDR. Variable = fixed effect of SNP split into factor levels (genotypes), with A indicating the reference allele and B the alternative allele; solution = effect size of factor level; SE = standard error of effect size; z_ratio and p.z_ratio = Z ratio and associated p-value for factor level respectively; measure = PC; SNP.name = name of SNP marker; effect.solution = solution for factor level (genotype) if homozygote with largest (positive) effect size is the intercept level; effect = genotype coded using A as increasing and B as reducing allele; neg.allele = allele with reducing effect coded as base pair; Df = degrees of freedom for SNP effect; Sum.of.Sq = Sum of Squares of SNP effect; Wald.statistic = Wald statistic for SNP effect; Pr.Chisq. = Wald statistic associated p-value; major.allele = major allele of SNP coded as base pair; minor.allele = minor allele of SNP coded as base pair; minor.allele.freq = frequency of minor allele; additive_effect = additive genetic effect; domiance_effect = dominance deviation; SNP.var and SNP.var.SE = SNP variance and associated standard error, estimated as described in the methods under section 'Estimation of SNP effect size distribution' respectively; SNP.prop.gen = proportion of total additive genetic variance (of PC) attributed to SNP; SNP.prop.phen = proportion of total phenotypic variance (of PC) attributed to SNP; Var.add and SE.var.add = total additive genetic variance (of PC) and associated standard error, respectively; Var.total = total phenotypic variance (of PC).

Chapter 4

Supplemental Figures

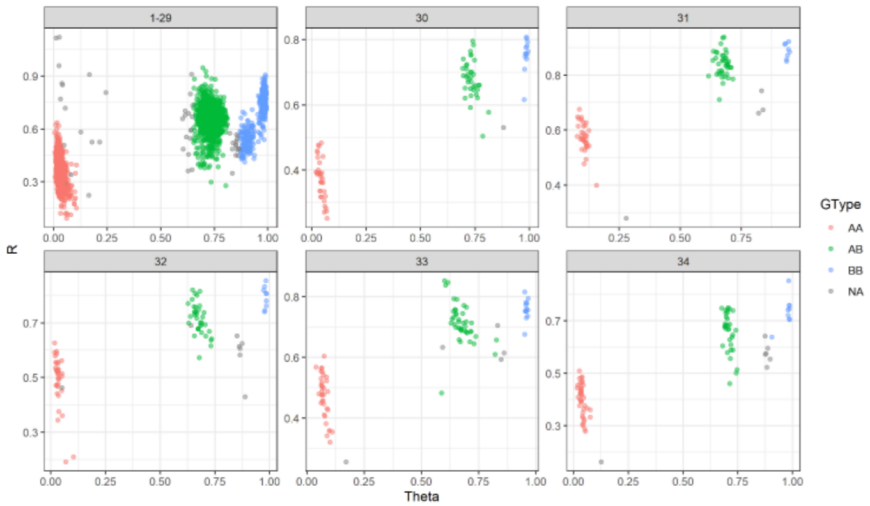


Figure S4.1: Example genotyping clusters for locus cela1_red_1_115520931, removed for high variance in genotyping call rate across plates. Note ambiguous clustering between plates for the AB and BB genotypes.

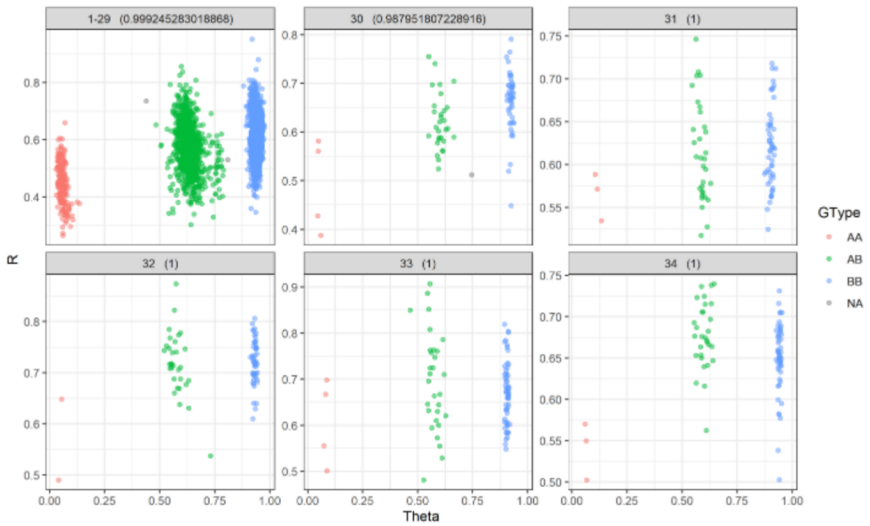


Figure S4.2: Example genotyping clusters for locus cela1_red_1_32077481, passing inspection for low within-plate genotyping call rate across plates.

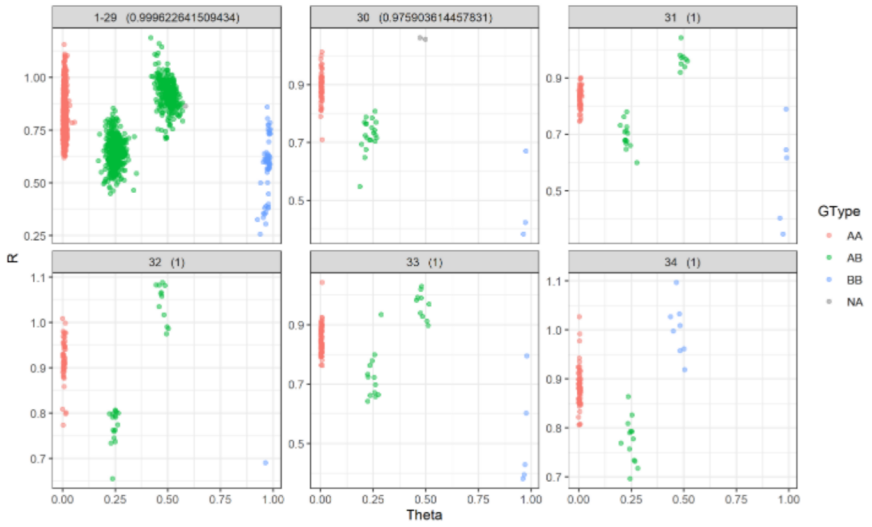


Figure S4.3: Example genotyping clusters for locus cela1_red_6_91235005, failing inspection for low within-plate genotyping call rate across plates. Note the misplaced BB cluster on Plate 34.

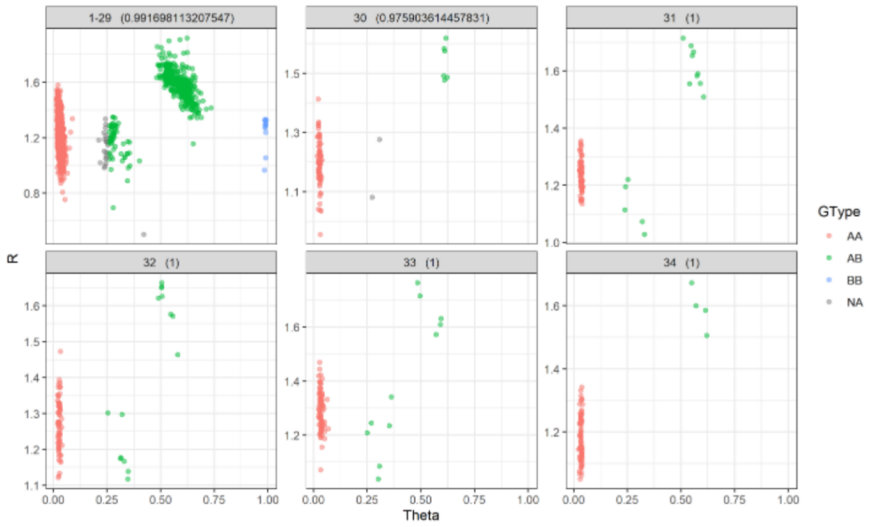


Figure S4.4: Example genotyping clusters for locus cela1_red_4_91623770, failing inspection for low within-plate genotyping call rate across plates. Note ambiguous AB cluster on Plates 1-29 and 30.

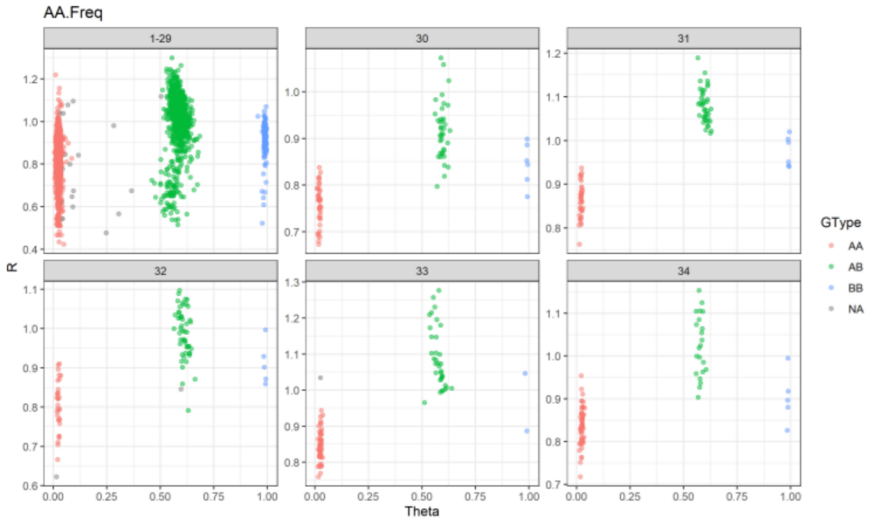


Figure S4.5: Example genotyping clusters for locus cela1_red_2_4891682, passing inspection for variance in the frequency of AA genotypes across plates.

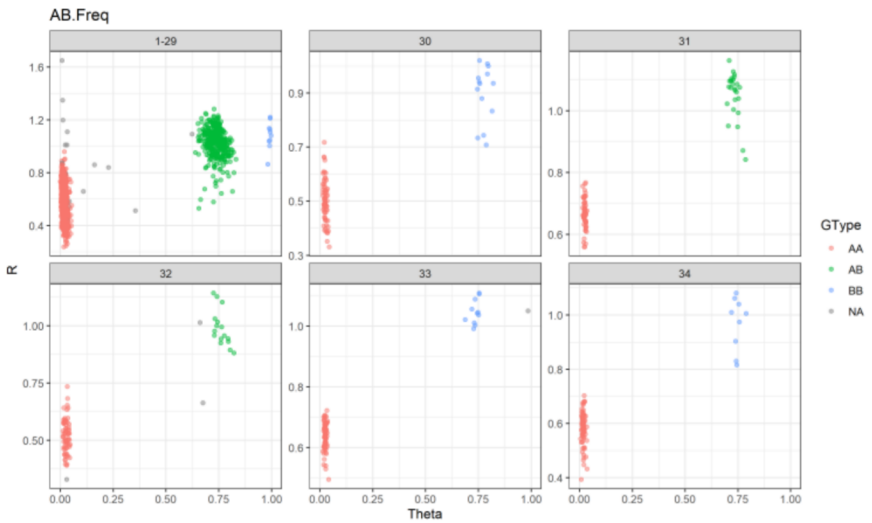


Figure S4.6: Example genotyping clusters for locus cela1_red_2_6910963, failing inspection for variance in the frequency of AB genotypes across plates. Note the switches of AB and BB scoring between plates at the same cluster position.

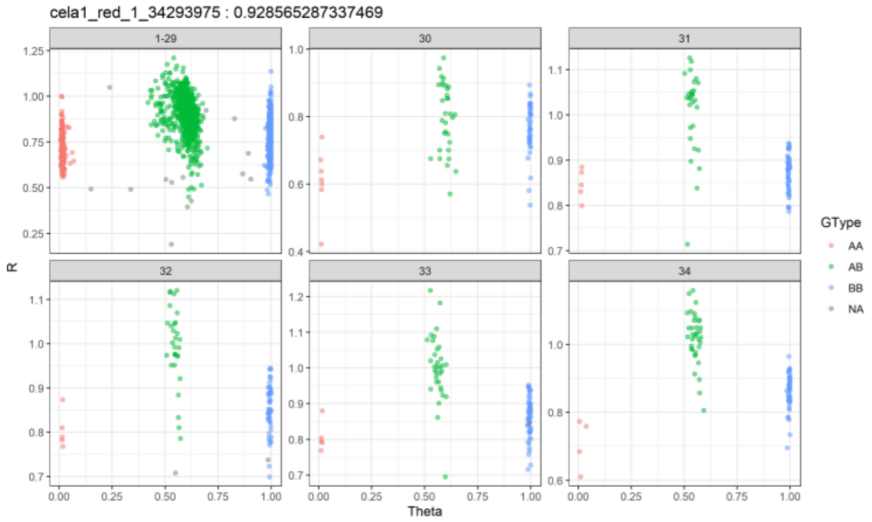


Figure S4.7: Example genotyping clusters for locus cela1_red_1_34293975, passing inspection for heterozygote excess/deficit.

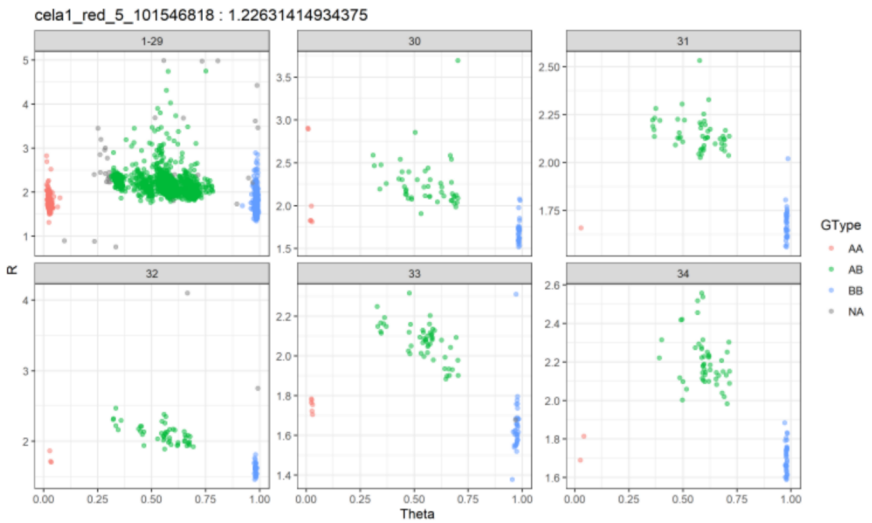


Figure S4.8: Example genotyping clusters for locus cela1_red_5_101546818, failing inspection for heterozygote excess/deficit. The wide distribution of the AB genotype indicates that this could be a duplicated polymorphic locus, and is therefore discarded as a precaution.

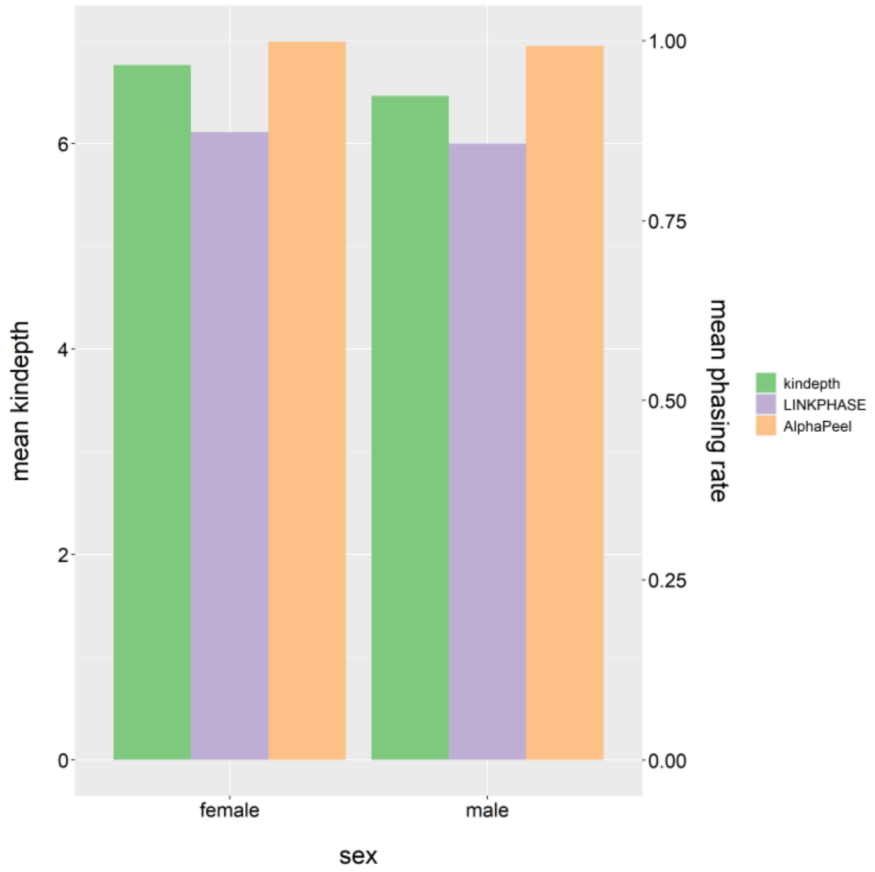


Figure S4.9: Figure showing the mean phasing rate and mean kindepth for males and females. Phasing rates are shown based on phasing using the programmes AlphaPeel and LINKPHASE. LINKPHASE was not used for downstream analysis and only shown for comparing performance of phasing with Alphapeel.

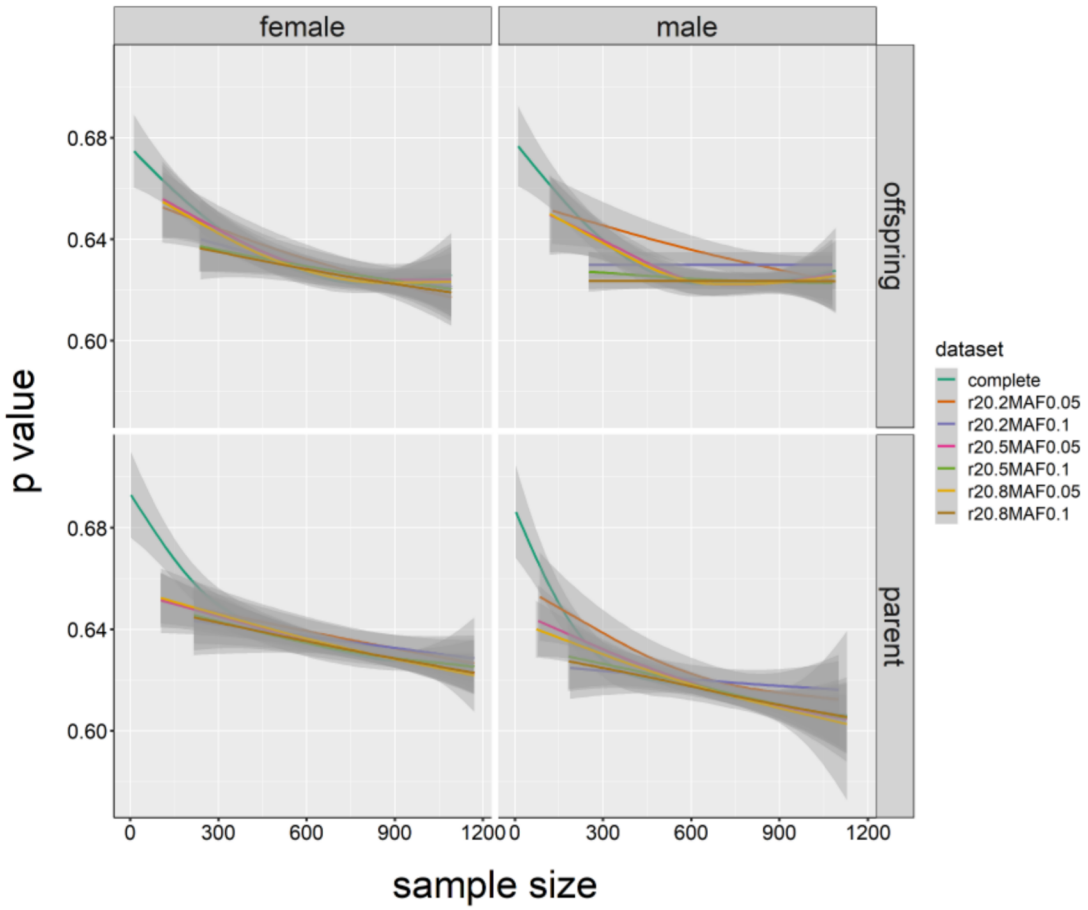


Figure S4.10: Figure showing the correlation of sample size, i.e., number of informative matings, with p-values from the Fisher's exact test to detect transmission distortion. The colours represent different datasets, the complete data, labelled 'complete', and six datasets pruned according to different LD and MAF thresholds. The label r20.2MAF0.05 refers to the dataset pruned using r^2 >= 0.2 as LD threshold and 0.05 as the MAF cut-off value, with the other labels following the same pattern.

Supplemental Tables

Table S4.1: Complete results of transmission distortion analysis. SNP Name refers to the SNP locus ID. The status indicates if alleles were counted in the offspring (inheritance) or the parent (transmission). A and B are randomly assigned labels for the two alternative alleles and represent counts of inherited/transmitted alleles. The expected frequency refers to the expected allele count under Mendelian inheritance. The minor allele indicates which label was used for the minor allele for each SNP. The deviation shows the proportion of allele counts that deviated from Mendelian inheritance. The sign indicates the direction of distortion of the “A” allele. CEL.LG refers to the red deer linkage group, and CEL.order to the order of the SNP within that linkage group. cow_ARS_UCD1.2_Chr and cow_ARS_UCD1.2_pos refer to the equivalent chromosome and SNP position in base pairs in the cattle genome version ARS_UCD1.2. MAF represents the minor allele frequency in the population.

Table S4.2: SNPs that show significant signals of transmission distortion. The status indicates if distortion occurs in the offspring (distorted inheritance) or the parent (distorted transmission); A and B are randomly assigned labels for the two alternative alleles and represent counts of inherited/transmitted alleles; the minor allele indicates which label was used for the minor allele for each SNP; the expected frequency refers to the expected allele count under Mendelian inheritance; CEL LG refers to the red deer linkage group and CEL order to the order of the SNP within that linkage group; MAF represents the minor allele frequency in the population.

SNP name	sex	status	A	B	minor allele	expected freq.	p-value	CEL LG	CEL order	MAF
cela1_red_15_2083615	female	offspring	204	140	B	172.00	0.018	1	13	0.11
cela1_red_15_32873496	female	parent	289	205	B	247.00	0.0088	1	493	0.13
cela1_red_5_31220603	female	offspring	161	230	B	195.50	0.015	3	481	0.13
cela1_red_5_36061813	female	offspring	316	411	A	363.50	0.014	3	556	0.25
cela1_red_5_39129553	male	parent	211	142	B	176.50	0.01	3	607	0.15
cela1_red_18_40822403	male	offspring	326	244	A	285.00	0.017	4	605	0.2
cela1_red_17_29008881	male	offspring	526	419	B	472.50	0.015	5	708	0.48
cela1_red_17_19944725	male	offspring	257	342	A	299.50	0.015	5	863	0.23
cela1_red_17_15881624	male	parent	367	467	B	417.00	0.016	5	911	0.36
cela1_red_17_14205606	male	parent	461	364	B	412.50	0.018	5	934	0.39
cela1_red_17_11825067	male	parent	387	503	A	445.00	0.0068	5	980	0.4

cela1_red_19_7091142	female	offspring	447	557	A	502.00	0.016	5	1246	0.45
cela1_red_19_11487628	female	offspring	489	389	B	439.00	0.019	5	1309	0.49
cela1_red_19_13559483	male	parent	323	416	A	369.50	0.017	5	1335	0.24
cela1_red_19_16923111	female	parent	526	423	B	474.50	0.019	5	1380	0.43
cela1_red_19_56855784	male	offspring	514	405	B	459.50	0.012	5	1935	0.4
cela1_red_6_117221935	female	parent	386	498	A	442.00	0.0087	6	650	0.35
cela1_red_23_10064689	female	offspring	162	241	B	201.50	0.0058	7	154	0.12
cela1_red_23_18947043	female	offspring	314	223	B	268.50	0.0058	7	274	0.19
cela1_red_23_18947043	male	parent	270	193	B	231.50	0.012	7	274	0.19
cela1_red_23_20157649	male	parent	201	274	B	237.50	0.019	7	288	0.24
cela1_red_23_30976844	male	parent	473	373	B	423.00	0.017	7	427	0.32
cela1_red_23_33896016	male	offspring	372	469	A	420.50	0.019	7	462	0.33
cela1_red_23_33916413	male	offspring	428	334	B	381.00	0.018	7	463	0.3
cela1_red_23_34142616	male	parent	365	462	A	413.50	0.018	7	468	0.44
cela1_red_2_129070839	male	offspring	394	494	A	444.00	0.02	8	741	0.36
cela1_red_7_15671293	male	offspring	319	421	B	370.00	0.0091	9	184	0.29
cela1_red_7_15671293	female	parent	318	443	B	380.50	0.0014	9	184	0.29
cela1_red_7_74496459	male	parent	89	151	B	120.00	0.0057	9	1110	0.14
cela1_red_7_87470136	male	offspring	286	379	B	332.50	0.011	9	1322	0.29
cela1_red_25_9992359	female	offspring	40	79	A	59.50	0.013	10	105	0.028
cela1_red_11_32767246	male	parent	224	157	B	190.50	0.016	11	482	0.12
cela1_red_11_53825197	male	parent	386	296	B	341.00	0.017	11	797	0.27
cela1_red_11_53826073	male	parent	386	296	B	341.00	0.017	11	798	0.27
cela1_red_11_59417377	male	parent	328	430	B	379.00	0.01	11	879	0.42
cela1_red_10_26376198	male	parent	118	71	A	94.50	0.017	12	335	0.091
cela1_red_10_37402940	male	offspring	381	480	A	430.50	0.018	12	492	0.42
cela1_red_10_38160902	male	offspring	315	410	B	362.50	0.013	12	503	0.29
cela1_red_10_43096556	female	parent	339	251	A	295.00	0.012	12	590	0.2
cela1_red_10_44502505	female	parent	376	287	A	331.50	0.015	12	611	0.25
cela1_red_10_44572589	female	parent	337	430	B	383.50	0.019	12	612	0.32
cela1_red_10_85688278	female	parent	94	160	A	127.00	0.0041	12	1239	0.092
cela1_red_21_53354661	female	parent	384	291	A	337.50	0.012	13	744	0.27
cela1_red_21_53374425	female	parent	171	246	B	208.50	0.01	13	745	0.13
cela1_red_21_54825083	female	offspring	366	468	A	417.00	0.014	13	773	0.4
cela1_red_21_59324469	male	offspring	271	357	B	314.00	0.017	13	820	0.28
cela1_red_21_61756749	male	offspring	393	497	A	445.00	0.015	13	848	0.37
cela1_red_16_21028246	male	parent	293	388	A	340.50	0.011	14	248	0.33
cela1_red_16_32005570	female	offspring	335	254	A	294.50	0.019	14	392	0.2
cela1_red_16_32005570	female	parent	353	269	A	311.00	0.02	14	392	0.2
cela1_red_16_46999075	male	parent	335	252	B	293.50	0.016	14	611	0.27
cela1_red_28_10731433	female	offspring	381	489	B	435.00	0.011	15	32	0.48
cela1_red_28_16921558	female	offspring	86	43	B	64.50	0.0079	15	209	0.031
cela1_red_28_30132248	male	parent	459	350	A	404.50	0.0071	15	406	0.35
cela1_red_28_45611754	male	offspring	284	209	A	246.50	0.018	15	621	0.15
cela1_red_26_3853654	female	parent	489	374	A	431.50	0.006	15	678	0.35
cela1_red_26_24623653	female	parent	486	386	B	436.00	0.019	15	972	0.32
cela1_red_8_67479466	male	offspring	496	385	B	440.50	0.0086	16	1	0.33

cela1_red_8_84526490	male	parent	460	359	A	409.50	0.013	16	239	0.41
cela1_red_8_98827226	male	parent	483	376	A	429.50	0.01	16	467	0.38
cela1_red_6_7863314	male	offspring	511	405	B	458.00	0.015	17	86	0.43
cela1_red_6_22967480	male	offspring	347	453	B	400.00	0.0091	17	339	0.34
cela1_red_6_41019581	male	offspring	37	73	B	55.00	0.02	17	620	0.036
cela1_red_6_54701086	male	parent	159	100	B	129.50	0.01	17	844	0.083
cela1_red_6_57133932	male	parent	84	136	B	110.00	0.016	17	900	0.064
cela1_red_4_54651931	male	offspring	443	335	A	389.00	0.007	18	861	0.34
cela1_red_4_68988501	male	parent	353	264	B	308.50	0.012	18	1050	0.22
cela1_red_4_73941083	male	parent	400	310	A	355.00	0.019	18	1125	0.27
cela1_red_4_74380080	male	parent	438	344	A	391.00	0.02	18	1135	0.42
cela1_red_4_77995342	male	parent	437	342	B	389.50	0.017	18	1179	0.45
cela1_red_4_110936970	male	parent	367	278	A	322.50	0.014	18	1683	0.32
cela1_red_4_111057809	male	parent	271	360	B	315.50	0.013	18	1685	0.29
cela1_red_4_111904468	male	offspring	217	299	B	258.00	0.012	18	1698	0.17
cela1_red_1_67521172	male	parent	242	325	A	283.50	0.015	19	665	0.25
cela1_red_1_78437038	male	offspring	402	291	B	346.50	0.003	19	836	0.28
cela1_red_1_78437038	male	parent	355	261	B	308.00	0.0085	19	836	0.28
cela1_red_1_78819201	male	parent	350	449	A	399.50	0.014	19	843	0.32
cela1_red_1_78973880	male	offspring	391	498	A	444.50	0.012	19	847	0.38
cela1_red_1_79282842	male	offspring	310	228	B	269.00	0.014	19	852	0.18
cela1_red_1_79306035	male	offspring	325	239	B	282.00	0.012	19	853	0.19
cela1_red_1_82082704	male	offspring	107	166	A	136.50	0.012	19	896	0.084
cela1_red_1_90822334	male	offspring	256	185	A	220.50	0.018	19	1044	0.14
cela1_red_3_24627800	male	offspring	507	402	A	454.50	0.015	20	309	0.39
cela1_red_3_28263984	male	offspring	340	436	B	388.00	0.017	20	371	0.3
cela1_red_3_38988182	female	parent	370	467	A	418.50	0.019	20	532	0.39
cela1_red_3_87332400	male	offspring	363	468	A	415.50	0.011	20	1324	0.34
cela1_red_3_91714164	male	offspring	244	167	B	205.50	0.0078	20	1389	0.14
cela1_red_14_28234577	female	offspring	403	314	B	358.50	0.02	21	335	0.32
cela1_red_14_35784443	male	parent	150	95	B	122.50	0.014	21	427	0.065
cela1_red_14_49464991	male	parent	345	254	A	299.50	0.0091	21	653	0.29
cela1_red_5_69043278	male	offspring	395	494	A	444.50	0.02	22	141	0.37
cela1_red_5_77945428	male	offspring	378	487	A	432.50	0.0093	22	247	0.39
cela1_red_5_79042894	male	offspring	478	377	A	427.50	0.015	22	264	0.46
cela1_red_5_79058925	male	offspring	377	478	B	427.50	0.015	22	265	0.46
cela1_red_5_79091651	male	offspring	377	478	B	427.50	0.015	22	266	0.46
cela1_red_5_116144268	female	parent	416	538	B	477.00	0.0059	22	800	0.46
cela1_red_5_118556565	female	parent	412	514	A	463.00	0.02	22	834	0.37
cela1_red_5_120107214	female	parent	315	414	B	364.50	0.01	22	861	0.23
cela1_red_5_120534995	female	parent	401	511	B	456.00	0.011	22	866	0.34
cela1_red_13_25757841	male	offspring	99	154	B	126.50	0.016	23	367	0.087
cela1_red_13_33854335	male	parent	386	291	A	338.50	0.01	23	486	0.29
cela1_red_13_51907018	female	offspring	211	286	A	248.50	0.019	23	738	0.17
cela1_red_22_25387708	female	offspring	315	425	A	370.00	0.0049	24	362	0.28
cela1_red_22_52782018	female	parent	379	486	B	432.50	0.011	24	759	0.23
cela1_red_20_63606593	male	offspring	443	550	A	496.50	0.017	25	954	0.41

cela1_red_9_72238025	male	offspring	370	471	A	420.50	0.015	26	102	0.36
cela1_red_9_72379976	male	offspring	367	477	A	422.00	0.0084	26	104	0.37
cela1_red_24_13903651	male	parent	274	366	B	320.00	0.012	27	177	0.35
cela1_red_9_2015763	male	parent	87	43	B	65.00	0.0081	28	22	0.038
cela1_red_9_17931387	male	parent	407	315	A	361.00	0.018	28	260	0.31
cela1_red_9_17931483	male	parent	315	407	B	361.00	0.018	28	261	0.31
cela1_red_9_18507609	female	offspring	446	338	A	392.00	0.0073	28	268	0.37
cela1_red_9_28533108	female	parent	512	411	B	461.50	0.02	28	423	0.36
cela1_red_9_38759028	female	offspring	539	430	A	484.50	0.014	28	569	0.43
cela1_red_8_13208565	male	parent	254	345	B	299.50	0.0091	29	191	0.23
cela1_sika_12_1305312	female	offspring	233	165	A	199.00	0.019	30	8	0.13
cela1_red_12_1921626	female	offspring	146	209	B	177.50	0.02	30	23	0.11
cela1_red_12_11917408	male	offspring	392	497	B	444.50	0.013	30	181	0.43
cela1_red_12_16592749	male	parent	419	527	B	473.00	0.015	30	248	0.44
cela1_red_12_16832622	male	parent	326	232	A	279.00	0.0057	30	251	0.23
cela1_red_1_7261063	female	parent	495	391	B	443.00	0.015	31	97	0.4
cela1_red_1_17185806	male	offspring	249	172	B	210.50	0.0085	31	249	0.14
cela1_red_1_18214314	male	offspring	104	168	A	136.00	0.0074	31	268	0.086
cela1_red_1_22355321	female	offspring	300	222	B	261.00	0.018	31	333	0.19
cela1_red_1_31628012	female	offspring	209	138	B	173.50	0.0076	31	497	0.1
cela1_red_1_40812687	male	offspring	354	451	B	402.50	0.017	31	646	0.37
cela1_red_1_40846958	male	offspring	353	452	B	402.50	0.014	31	648	0.36
cela1_red_1_54809345	female	offspring	416	321	B	368.50	0.014	31	855	0.26
cela1_red_1_55613382	female	offspring	198	275	A	236.50	0.013	31	869	0.15
cela1_red_27_26716954	female	offspring	222	306	B	264.00	0.011	32	421	0.19
cela1_red_27_35114965	female	offspring	209	135	B	172.00	0.0057	32	511	0.13
cela1_red_27_37895237	female	offspring	363	467	B	415.00	0.012	32	549	0.37
cela1_red_27_38731584	female	offspring	484	367	A	425.50	0.0048	32	561	0.4
cela1_red_27_44388163	male	parent	205	139	B	172.00	0.014	32	608	0.1
cela1_red_27_45245171	male	parent	402	509	A	455.50	0.013	32	619	0.42
cela1_red_2_8941291	female	offspring	490	373	A	431.50	0.0051	33	116	0.44
cela1_red_2_55319851	male	offspring	321	239	A	280.00	0.016	33	837	0.18
cela1_red_2_55319979	male	offspring	239	321	B	280.00	0.016	33	838	0.18
cela1_red_2_57349673	female	offspring	205	281	A	243.00	0.017	33	880	0.17
cela1_red_2_57628629	male	offspring	144	209	B	176.50	0.016	33	884	0.12

Table S4.3: Complete results of selection analysis using empirical p-values derived from GeneDrop simulations. SNP Name refers to the SNP locus ID, Status indicates if SNPs showed transmission distortion in parents or offspring, and sex refers to the sex of the parent or offspring. Total change shows the total change in minor allele frequency (MAF) over the time period used in the GeneDrop simulation (1980 to 2018) and p-direct is the related uncorrected empirical p-value. Total dist. refers to the cumulative absolute MAF change or distance travelled by the SNP marker, and p-balance is the associated uncorrected empirical p-value. Finally, slope refers to the slope of the regression of MAF on birth cohorts and p-slope is the respective uncorrected empirical p-value. P-values corrected for FDR are given by p-direct.fdr, p-balance.fdr and p-slope.fdr.

Table S4.4: SNP effects and variances from all BGLR models fitting SNP effects with a BayesB prior. The SNP ID is given by SNP.name, SNP effects are given by the posterior mean of each SNP and are indicated by bhat while SD.bhat refers to the standard deviation of SNP effects. The values for the 95% confidence interval are given by 95LowerBound and 95UpperBound. SNP variance and associated standard deviation are given by var.b and SD.var.b, respectively. Finally, model indicates whether the values refer to the complete model with sex as fixed effect ("both_sexes"), the female only model ("female_specific"), or the model containing only males (male_specific").

Table S4.5: Predicted phenotypes from all BGLR models fitting SNP effects with a BayesB prior on data and link function (*probit*) scale. ID refers to the ID of an individual deer and y represents the phenotypic response variable, which was survival to the second year coded as 1 for survival and 0 for death. The posterior mean of the response variable represents the phenotypic value predicted by the model and is given by yhat on a probit scale. The values of yhat_prob are the yhat values on a probability scale derived by evaluating them as quantiles at the standard normal cumulative distribution function (using *pnorm() in R*. The cumulative effect of all SNPs within an individual on the phenotype is referred to by ghat again on the probit scale of the model and ghat_prob gives the values of ghat on a probability scale, calculated the same way as yhat_prob for yhat. Finally, model indicates whether the values refer to the complete model with sex as fixed effect ("both_sexes"), the female only model ("female_specific"), or the model containing only males (male_specific").

Table S4.6: Comprehensive list of all genes found within 70kbp each side of all 136 SNPs showing significant transmission distortion. SNP.name refers to the name of the significant SNP marker, sex refers to the sex the SNP showed transmission distortion in, and status to whether it was in a parent or offspring. The gene ID of the ensmbl database is indicated by ensembl_gene_id, and external_gene_name refers to the commonly used identifier of the gene. The full descriptive gene name is given by description. The ensembl gene ID of the mouse ortholog and the common gene name are indicated by mmusculus_homolog_ensembl_gene and mmusculus_homolog_associated_gene_name, respectively. The ensembl gene ID of the human ortholog and the common gene name are given by hsapiens_homolog_ensembl_gene and hsapiens_homolog_associated_gene_name, respectively. The basepair position of the start and end of the coding region of each gene are referred to by start_position and end_position. The basepair position of each SNP in relation to the cattle genome (version ARS_UCD1.2) is indicated by cow_ARS_UCD1.2_pos and cow_ARS_UCD1.2_Chr refers to the respective cattle chromosome.

Table S4.7: Complete results of Gowinda gene ontology enrichment analysis. GO_term refers to the tested GO term. No.genes indicates the average number of genes found per simulation for the given GO category; in *gene* mode every gene is only counted once, whereas in *snp* mode a single gene may be counted several times dependent on the SNP. No.genes_candidateSNPs indicates the same as No.genes, but based on the candidate SNPs only instead of the average of the simulations. The uncorrected p-value is given by p-value, and the p-value after FDR correction by p.fdr. No.genes_found refers to the number of unique genes found for the given GO term. No.genes.max indicates the maximum number of genes that could be found for the given GO category, i.e., genes of the given GO category that have a corresponding entry in the annotation file and contain at least one SNP. No.genes.GO refers to the total number of genes for the given GO category in the GO association file. A description of the respective GO term is given by GO.description and ensembl_gene_ids lists the gene IDs found for the given GO category. The assumed linkage model is given by mode. Finally, sex refers to which sex candidate SNPs showed transmission distortion in and status indicates if it was in parents or offspring.