Evolution and Recent Adaptation of the Domestic Water Buffalo

Luke William David Davies

A thesis submitted to Cardiff University for the degree of Doctor of Philosophy

February 2024

Table of Contents

Table of Contents: Figures

Table of Contents: Tables

Table of Contents: Supplementary Materials

Acknowledgments

I would like to thank the BBSRC SWBio DTP for giving me the opportunity and, along with the Global Challenge Research Fund – HEFCW, funding to undertake such an amazing PhD for the past 4 years. The knowledge and skills that I have gained are invaluable.

It is impossible to overstate the amount of help and support given to me by my amazing supervisory team. After meeting Pablo in in my second year of undergraduate where I undoubtably impressed him knowing the biogeographic layout of Madagascar; I still and always will possess the David Attenborough DVDs that gave me that knowledge. They clearly come in handy. Your expertise, and the confidence and trust you place in your students is truly the best support a student could ask for. I will never forget it. Thank you to Mark who had the difficult task of teaching me ABC. Your thoughtfulness, patience and consideration for work and students is something I will strive to achieve throughout my career. Benoit: the opportunity you gave me to study at DGFC for my PTY was a springboard to the rest of my life. Your dedication, commitment, and investment in people is admirable, and something that is rare to come across. Your patience for my lack of emails is also quite rare 😉.

I wish to thank Jonathan Corpe of WCWB; Dagan James of BWB; Dr Muralidhar Metta, and Sabah's DVS for their participation in this project. Your help was fundamental to making this project happen. Thank you to Kennedy Juani who has been the most tremendously brilliant collaborator myself and my supervisors could have hoped for. I know between you, Pablo, and Ben that the Malaysian buffalo project is in excellent hands.

A special thank you to all my friends and family for all the enjoyable times over the past years. My conversation topics have expanded from Borneo to now include buffaloes! I had the fortune of meeting my wonderful girlfriend Jess at the start and I have had the most incredible time with her since. Despite Jess's (human) and Agatha's (Maine coon) attempts to thwart my productivity, I will always endeavour to support Jess throughout her career as she has for me. Moira (the other cat) is less needy, I couldn't take three. Finally, thank you to my mum and dad for the support that has made my whole education, experiences, and career possible. My mum always nagged me for when she could call me Dr. and show everyone my work. I would like to think that she'll be watching when I graduate and publish my research.

Summary

Domestication has underpinned the development of human civilization. A growing human population and increasingly complex societies have created intensification pressures upon the livestock industry. Favouring of highly productive breeds at the expense of the loss of locally adapted rarer breeds will have implications for future food security due to the loss of genomic resources. With the recent availability of accessible genomic resources for domestic water buffalo (*Bubalus bubalis*), this thesis seeks to contribute towards quantifying their genetic variation and adaptive potential. In Chapter Two, ancestry of two UK populations was tracked to Italian origins and was shown to have retained the majority of genetic diversity since importation. Such levels of genetic diversity provide the opportunity for effective selection programmes focused on production-based QTLs. Chapter Three analysed and compared novel data from an Indian murrah buffalo population to wider global populations to detect selective sweeps indicative of recent selection. Loci under selection in non-Indian murrah were associated with genes linked to draught and meat traits instead of the typical focus on milk. SNP array data is incredibly useful for genomic selection; however, inherent ascertainment bias creates challenges for evolutionary studies. Chapter Four attempted to understand the bias in the Axiom^{TM} Buffalo Genotyping Array when analysing river and swamp buffaloes. Results showed that patterns of variation were inconsistent between river and swamp species, and between array and sequencing data. Shared polymorphic markers in the array are likely ancestral SNPs and targets of balancing selection, thus distorting evolutionary inferences between the two species. Chapter Five modelled the evolutionary history of domestic buffaloes revealing further support for a Pleistocene divergence of river and swamp species. Dispersal of river buffalo from India is likely linked to maritime trade early in the Common Era. Furthermore, artificial selection appears to be a driving force for regions of divergence between river and swamp buffalo.

Common Abbreviations

ABC	Approximate Bayesian Computation
ANOVA	Analysis of Variance
AMOVA	Analysis of Molecular Variance
ART	Assisted Reproductive Technologies
BF	Bayes Factor
f/F	Inbreeding Coefficient
FDR	False Discovery Rate
FID	Family ID
F_{ROH}	Runs of Homozygosity Inbreeding Coefficient
F_{ST}	Fixation Index
H_E	Expected Heterozygosity
H_O	Observed Heterozygosity
IBD	Identity by Descent
K	Defined number of genetic clusters for analysis
KYP	Thousand Years Before Present
LAMP-ANC	Local Ancestry in Admixed Populations (ancestral)
LD	Linkage Disequilibrium
Ma	Million Years Ago
MAF	Minor Allele Frequency
MDS	Multi-Dimensional Scaling
N_e	Effective Population Size
NGS	Next Generation Sequencing

PC	Principal Component
PCA	Principal Component Analysis
QTL	Quantitative Trait Loci
R^2	Coefficient of Determination
RFLP	Restriction Fragment Length Polymorphism
ROH	Runs of Homozygosity
RV_	Prefix indicating population is of river buffalo species
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SW_	Prefix indicating population is of swamp buffalo species
THI	Temperature Humidity Index
UK	United Kingdom
WGS	Whole Genome Sequence
XP-EHH	Cross Population Extended Haplotype Homozygosity
YBP	Years Before Present

Chapter One

General Introduction

1.1. Domestication

The transition from Late Pleistocene into the Holocene is underpinned by changes in the Earth's climatic conditions. The Pleistocene is characterized by a series of major glaciation events with the Last Glacial Maximum occurring approximately 21,000 years before present (YBP) (Clark and Mix, 2002; Clark *et al.*, 2009; Shakun and Carlson, 2010; Hughes, Gibbard and Ehlers, 2013). Minor cooling events culminating in the end of the Younger Dryas marked the onset of the Holocene 11,700 YBP (Cheng *et al.,* 2020; Shakun and Carlson, 2010). An environmental shift to a warmer, more stable climate alongside a period of megafaunal extinctions facilitated a change in human (*Homo sapiens*) behaviour from nomadic hunter-gatherers to agrarian societies (Diamond, 2002). Reductions in food availability caused a subsequent broadening of diets to include more plants and small game (Diamond, 2002). Eventually, the use of gathered plants and control of animals outside of natural habitats generated agricultural systems in what is now known as domestication (Diamond, 2002). Evidence suggests early farming was incredibly difficult as humans displayed poor health. However, the ability to generate a sustainable source of food gave great competitive advantages to these populations as a whole to grow and expand (Diamond, 2002). As a result, 88% of humans alive today speak a language that can be traced back to the earliest centres of domestication in Eurasia (Diamond, 2002). Domestication can therefore be considered as the most important technological advancement in human evolution, and a prerequisite for the development of civilization and innovative societies found today.

The first known example of domestication of an animal is the dog (*Canis familiaris*). Domestication from a now-extinct lineages of grey wolves (*Canis lupus*) likely occurred in the Near or Far East (Vilà *et al.*, 1997; Savolainen *et al.*, 2002; vonHoldt *et al.*, 2010; Ding *et al.*, 2012; Freedman *et al.*, 2014; Frantz *et al.*, 2016; Bergström *et al.*, 2020). The interspecies relationship between humans and prehistoric dogs may date back to ca. 31,000 YBP, with full

domestication occurring by 14,000 YBP (Germonpré *et al.*, 2009; Udell, Dorey and Wynne, 2010; Galibert *et al.*, 2011). Dog domestication is unique as dogs are the only large carnivore to have been domesticated (Frantz *et al.*, 2016). It has been hypothesised that dogs were domesticated via a commensal pathway, in which prehistoric dogs were sensitive to human sociability, before becoming conditioned (Udell, Dorey and Wynne, 2010; Vigne, 2011; Larson and Fuller, 2014). Subsequent major animal domestication events followed a more human centric approach, with wild individuals being caught, tamed, and then bred for the benefit of human societies.

Sixteen species of large mammals (>40 kg) have been domesticated, the majority of which are Bovidae and constrained to a few biogeographic areas (Figure 1.1) (Larson and Fuller, 2014). Major centres of domestication include the Fertile Crescent (taurine cattle, sheep, goats, pigs), Indus Valley (indicine cattle, river buffalo), Mainland Southeast Asia (swamp buffalo, gayal, pigs, yak), and Andes (alpaca, llama) with sporadic additional domestications elsewhere (Bactrian camel, horse, reindeer, at northernly latitudes; dromedary camel at Arabian Peninsula). Successful domestication of animals is rare. Diamond (2002) sets out six barriers to domestication with these being; i) inaccessible diet, ii) slow growth rate and long birth spacing, iii) high aggression, iv) reluctance to breed in captivity, v) lack of follow-the-leader dominance, and vi) likely to panic in enclosures or when faced with predators. In practice, humans were able to domesticate large herbivorous bovids due to the wide availability of grass as feed, and the ability to maintain herds within defined areas of land.

The domestication of dogs proved that that other species could be controlled by humans, however this case likely did not define an outcome of domestication (Diamond, 2002). Instead, later domestication of herbivorous mammals may have been a strategic response to overhunting, a common theme in the extinction of megafauna of the Late Pleistocene (Larson and Fuller, 2014; Bergman *et al.*, 2023). Archaeological studies have revealed that overhunting was prevalent prior to domestication, and that humans began altering hunting strategies of wild sheep, goats, pigs, and cows (Zeder, 2012; Marom and Bar-Oz, 2013). By 10,000 YBP, humans were preferentially killing young males of a variety of species and allowing females to survive, probably to produce more offspring (Zeder, 2012;

Marom and Bar-Oz, 2013). Early domestication events of large mammals likely orientated around maintaining a sustainable source of food (Larson and Fuller, 2014).

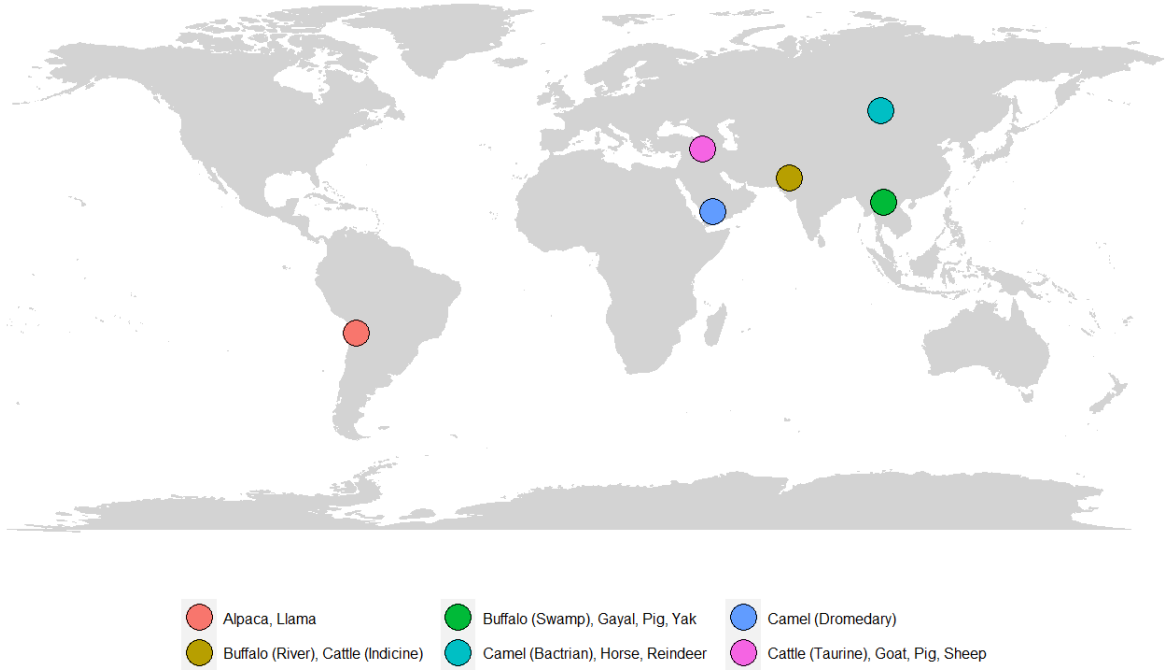


Figure 1.1: Domestication centres of large (>40kg) mammalian livestock species. 16 species of large mammals (excluding dogs) have been domesticated with the majority being bovids from Eurasia.

1.2. Livestock Production Systems

Development of agriculture has progressed to support an expanding human population that now numbers eight billion. Livestock contribute 40% of the global value of agricultural output, directly providing 15% of energy needs and 25% protein to people (FAO, 2009). The popularity of livestock products is rapidly increasing. Since the 1960s milk consumption per capita has doubled, meat consumption tripled, and egg consumption increased by a factor of five (FAO, 2009). Post 1980s, these increases have rapidly outmatched that of other major food commodity groups (e.g., cereals) (FAO, 2009). Economic growth and the development of countries (e.g., China, Brazil) appears to be the driving force behind the expansion of livestock industries (Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009; Godde *et al.*, 2018). As a result, humans have bred livestock in huge numbers. For example (as of 2021), the number of chickens exceeds 25 billion, meanwhile cattle, sheep, goat, and pigs exceed 1 billion individuals each (FAOSTAT). These population sizes translate into 62% of mammalian

biomass on Earth being accounted for by livestock, while 71% of biomass of avian species is poultry (Bar-On, Phillips and Milo, 2018). To accommodate such scales of production, humans have radically altered the Earth's biosphere for agricultural and urban development.

Several production systems have been developed based upon agro-ecological opportunities that vary in intensity of output. Grazing systems occupy the largest land cover, estimated up to 26% of ice-free surface and occupying 33 million square kilometres (Asner *et al.*, 2004; FAO, 2009). Farms within grazing systems are differentiated between intensive and extensive systems. Extensive farms are typically characterized by grazing ruminants on communal or open-access areas and frequently occur among marginal lands and sparsely populated areas that are unsuitable for reliable crop production (Sere & Steinfeld, 1996; Gandini and Villa, 2003; FAO, 2009; Kleppel and Frank, 2022). Livestock are an incredibly important source of food security in these areas to meet energy needs of people (Herrero *et al.*, 2013). In contrast, intensive grazing systems operate nearby areas of increased human densities typically in temperate zones where high-quality grassland and fodder production can support a larger number of animals (Sere & Steinfeld 1996). Extensive and intensive farming systems are dependent upon the surrounding environment and therefore livestock that persist in these regions are typically intrinsically adapted to the conditions they persist in (Mirkena *et al.*, 2010; Marshall, 2014; Biscarini *et al.*, 2015). As a result, livestock adapted to these local environments have small ranges and often have strong cultural and societal links to the communities where they occur (Gandini and Villa, 2003; Biscarini *et al.*, 2015). Livestock within these systems show high environmental resilience and provide a sustainable source of food.

Greater productivity of livestock rearing can be achieved with an increased input from humans. Mixed farming systems link livestock rearing and cropping activities, being defined as "those where more than 10% of dry matter fed to animals comes from crop by-product or stubble or where more than 10% of total value of production comes from non-livestock farming activities" (FAO, 2009; Sere & Steinfeld 1996). Mixed farming systems can broadly be categorized into two forms and biogeographic areas. Rainfed systems, those in which more than 90% of non-livestock farm production comes from rainfed land use, are mostly distributed across temperate regions (Sere & Steinfeld 1996; Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009). Meanwhile, irrigation systems, defined as more than 10% of the value of

non-livestock farm production, are commonly found in the East and South Asia surrounding areas of high population density (Sere & Steinfeld 1996; Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009). These farming systems contribute the bulk of animal products to humans, shifting focus towards increased productivity (Steinfeld, Wassenaar and Jutzi, 2006; Herrero *et al.*, 2013). Extending this theme further, agricultural industries are now capable of regulating all farming conditions. Industrialized systems are typically dedicated to a single species and used to supply large dense urban centres (Sere & Steinfeld 1996; FAO, 2009). A rapid increase in poultry and pork production occurred alongside development of industrialized systems as these monogastric animals have the highest growth rates and lowest costs per unit of output (Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009). Despite occupying small amounts of land, meeting the feed requirements of industrialised systems means that approximately 33% of global agricultural cropland is used to produce animal feed (Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009). Intensive breeding of livestock means that environmental adaptations become less relevant according to resource allocation theory (Beilharz, Luxford and Wilkinson, 1993; Mignon-Grasteau *et al.*, 2005; Mirkena *et al.*, 2010). The result is that industrialized livestock are highly productive and therefore valuable across an urban landscape.

1.3. Threats to Food Security

The development and expansion of agricultural industries to support the demanding human population growth has profoundly altered the biosphere. Despite agricultural land cover not increasing since 1991, development of the livestock industry has continued growing through intensification (O'Mara, 2012). The consequences of current food production effects an array of vital ecosystem processes. The agricultural industry is the largest contributor to biodiversity loss through habitat destruction, intensification, and release of pollutants leading to downstream effects such as disruption of nutrient cycling (Dudley and Alexander, 2017; Tully and Ryals, 2017). Therefore, without proper management, nutrient rich soils are quickly depleted from grazing of livestock and extraction of crops, reducing future produce yields (Tully and Ryals, 2017). Farms overcome nutrient deficiency with the application of fertilizers. However, excess macronutrients (e.g., phosphorus) typically leech into waterways leading to detrimental effects such as eutrophication (Dudley and Alexander, 2017; Tully and Ryals, 2017). Intensification of farming provides the ideal conditions for the spread of diseases. The

dependence upon antibiotic use in the livestock sector is a leading cause of antibiotic-resistant pathogens that has major consequences for food production, public health, and wide ecological impact (Woolhouse *et al.*, 2015; Koch, Hungate and Price, 2017). Among the most prominent causes of concern for people and livestock is climate change (Baumgard *et al.*, 2012). The vast numbers of livestock provide a significant contributor to global greenhouse gases (GHGs), accounting for 14.5% of global emissions of carbon dioxide (Gerber *et al.,* 2013). Potent GHGs such as methane (CH_4) and nitrous oxide (NO_2) are also emitted with CH_4 emissions nearing double that of CO_2 within livestock supply chains (Gerber *et al.,* 2013; Cheng, McCarl and Fei, 2022). The result of being such a contributor produces a feedback loop that will become increasingly negative for the health and productivity of livestock.

Maintaining livestock productivity for future generations will require flexibility against increasing climatic instability. The most prominent effect of climate change is the increasing average global temperatures and increased frequency of extreme heat. Exposure to higher temperatures pushes livestock out of their thermal comfort zone for longer and more often, increasing the risk of temperature-related illness and death. To compensate for this increased stress, extra energy of the individual is required to maintain thermoregulation resulting in declines in productivity (Cheng, McCarl and Fei, 2022). The economic consequences and threat to food security from heat stress are huge. In the U.S alone, the livestock industry was estimated to have lost between $1.7 - $2.4 billion in 2003 (St-Pierre, Cobanov and Schnitkey, 2003). Since 2003, average global temperatures have continued to increase (IPCC, 2021). Heat stress increases metabolic load on the individual enhancing reactive oxidative substance production and thus generating oxidative stress that is damaging to cells (Belhadj Slimen *et al.*, 2016). The energy expended (e.g., increased sweating and respiration rate) controlling these stresses means that other biological processes suffer as a result. Under heat stress, feed intake has been observed to decrease in all livestock species further reducing energy availability for production (Kadzere *et al.*, 2002; Collier, Renquist and Xiao, 2017; Cheng, McCarl and Fei, 2022). Decline in feed intake can explain approximately 35% of decrease in milk production in dairy cows, with additional decreases in quality by reduced milk protein and fat content (Kadzere *et al.*, 2002; Collier, Renquist and Xiao, 2017; Cheng, McCarl and Fei, 2022). Comparatively, meat producing livestock fail to grow to their expected weights as body

size, carcass weight, and fat thickness are all reduced (Gonzalez-Rivas *et al.*, 2020; Cheng, McCarl and Fei, 2022). Furthermore, reproductive capabilities are reduced, and immune systems fail with the implication that maintaining a healthy stock becomes more difficult (Kadzere *et al.*, 2002; Bagath *et al.*, 2019; Chauhan *et al.*, 2021). Certain breeds are adapted to higher temperatures, and farmers can utilize these alternative breeds for mitigating heat stress, however their production yields are often lower (Rojas-Downing *et al.*, 2017). Therefore, finding the key genetic components associated with resistance to heat stress and productivity could prove key.

In tandem with increased temperature, many countries are seeing a decrease in water availability. Agriculture is the single largest global water user, accounting for 69% of fresh water withdrawals (Thornton, 2010; Cheng, McCarl and Fei, 2022). As temperatures increase, greater water consumption will be required for livestock (Thornton, 2010). Due to the uneven distribution of freshwater sources, increased competition and unsustainable use will be prevalent across water stressed regions (Thornton, 2010; Leng and Hall, 2021). Further strains on freshwater resources will be observed via other means. As sea levels rise, an influx of saltwater will occur into coastal freshwater sources, while inadequate waste disposal from urbans areas contaminate waterways (Rojas-Downing *et al.*, 2017). For example, the accumulation of heavy metals in water bodies presents a major source of concern for general health. Heavy metals can have severe detrimental effects on individuals, impairing many biological processes, and causing serious developmental and health problems (Gupta *et al.,* 2001; Vardhan, Kumar and Panda, 2019). Combined with other stresses from increased temperature and water issues, vulnerable individuals present opportunities for pathogenic organisms. Changing environmental conditions, highly connected human societies, intensification, and growing human-wildlife contact all facilitate easy transfer of pathogenic organisms (Tomley and Shirley, 2009; Gummow, 2010). Livestock are now encountering new pathogens that they are not resilient against (Thornton, 2010; Rojas-Downing *et al.*, 2017). For example, African swine fever outbreaks have devastated global wild and domestic pig populations since spreading worldwide (Sánchez-Cordón *et al.*, 2018; Luskin *et al.*, 2021; You *et al.*, 2021). In livestock that are resistant to pathogens, constant exposure still has negative consequences for production. For example, in a study on Australian livestock, ticks are responsible for an 18% decrease in body weight (White *et al.*, 2003).

The popularity of livestock depends on their accessibility and profitability. Livestock breeds typically fall into one of three categories (local, regional, or international) defined by the range that they occupy (FAO, 2015). Local breeds occur in only one country and are typically historic breeds that have adapted to their surrounding environment over time and are used by local rural communities. These livestock breeds are often a very important source of nutrition, labour, and income that can be maintained over many generations (Herrero *et al.*, 2013; Mapiye *et al.*, 2020). Regional transboundary breeds can be found in many countries within the same continental region. These breeds have likely become the most productive and popular local breed becoming traded across larger regions. International transboundary breeds are found globally, and these breeds have been commercially developed to produce great yields. For mammalian livestock species, 4,127 local breeds comprise 83.3% of total breeds (FAO, 2015). Regional and international transboundary breeds make up small proportions each at 8.7% and 8.0%, respectively (FAO, 2015).

Across most livestock, most local breeds are found across Europe, Caucasus, and Asia, spanning biogeographic strongholds for domestication (FAO, 2015). In livestock, genetic diversity increases i) closer to the domestication origin of a species, and ii) in those breeds not under intense artificial selection (Bruford, Bradley and Luikart, 2003). Most of the genetic diversity is held across local breeds as each breed has evolved unique adaptations to their environments. To maintain genetic diversity for future adaptability, we need to understand the distribution of variation across livestock and how traits are defined by underlying genetic interactions. However, 16% of mammalian livestock breeds are classified as at risk of extinction with a further 35% of unknown status (FAO, 2015). 11% of livestock breeds are already extinct (FAO, 2015). Breeds at risk of extinction are spread across both local and transboundary categories, although transboundary breeds are being conserved at a greater rate than local breeds. From 1999 to 2006, 31% of transboundary breeds at risk were reclassified to not at risk, while only 7.4% of local breeds were reclassified (FAO, 2015). The loss of such breeds may have major consequences on future livestock adaptability, and therefore security of food production.

1.4. Genetic Technologies for Livestock

The idea of applied selective breeding is considered to have begun with Sir Robert Bakewell (1725 – 1795) in 18th century England with the development of livestock for increased carcass traits. Changes across farming were taking place as the conversion of arable to grassland and the greater emergence of enclosed fields were becoming more prominent across the country (Wykes, 2004). The greater ability to manage livestock meant farmers could experiment with improving livestock more easily. Bakewell used 'in-and-in' breeding (inbreeding) to maintain desired traits within his stock and is attributed as the founder of the Dishley Leicester sheep, Dishley Longhorn, and Black Cart horse (Wykes, 2004). Bakewell's original breeds have since died out, though lineages persist within English Leicester, English Longhorn, and Shire breeds respectively. Although selection as a process is a relatively recent concept, artificial selection has been occurring throughout domestication. Humans have linked reproductive fitness of livestock to desirable traits, whether that be environmental adaptation or productive output, resulting in phenotypic divergence of livestock from their wild ancestors (Proudfoot *et al.*, 2020). Now in the modern day, humans are continually understanding the complex relationship between genetic variation and phenotypic output behind artificial selection (Meuwissen, Hayes and Goddard, 2013; Proudfoot *et al.*, 2020; Saravanan *et al.*, 2020).

Molecular data in livestock has been present since the early 1990s allowing scientists to characterize genetic diversity, map evolutionary history, and manage future development (FAO, 2023). In 1993, the FAO established a global program for characterization of farm animal genetic resources (FanGR), raising awareness and providing recommendations for the molecular monitoring of FAnGR diversity (FAO, 2023). Initial studies relied upon mitochondrial and microsatellite markers (Bruford and Wayne, 1993; Bruford, Bradley and Luikart, 2003). Due to its haploid nature and absence of recombination, mitochondrial DNA is an ideal marker to easily track phylogeny and, in doing so, has revealed extensive knowledge surrounding the evolutionary history and domestication of livestock (Bruford, Bradley and Luikart, 2003; Hanotte and Jianlin, 2005). Its shortfalls of being a maternally inherited single locus means that mitochondrial DNA is a poor predictor of overall genomic diversity (Bruford, Bradley and Luikart, 2003). The use of microsatellite markers counterbalanced this, allowing quantification of genetic variation across livestock, the capability of discriminating between

all levels of genetic units (e.g., population, breed, species), and detection of evolutionary processes (e.g., bottlenecks, admixture) that impacts variation (Bruford, Bradley and Luikart, 2003; Hanotte and Jianlin, 2005). Molecular data in livestock transformed our understanding away from thinking of breeds as static homogenous units, instead towards domestication as being a continually shifting and dynamic process.

Molecular tools have been extensively developed in the past 15 years. The emergence of next generation sequencing (NGS) technology has revolutionized the genomics industry, allowing the production of accessible large-scale extensive datasets and complete sequencing of entire genomes. NGS has also provided access to more molecular markers in the form of, for example, single nucleotide polymorphisms (SNPs), structural variations, copy number variation (CNVs), and epigenetic factors among others (FAO, 2023). The increasing use of such genetic technology has also been matched with developments in increased computing power, processing software, and statistical analysis. The application of these technological developments means that considerable increases in livestock productivity has been realized, as precision breeding methods such as genomic selection programmes can increase productivity beyond that attained from conventional breeding techniques. For example, since the 1960s, average milk yields in Holstein cattle have more than doubled from 5,000 kg to more than 11,000 kg per 305 days of lactation. More than 50% of these gains are attributed solely to genetic improvement, and genomic selection typically enables a further 1-3% annual gain (Thornton, 2010; Georges, Charlier and Hayes, 2019; Brito *et al.*, 2021)

SNP genotyping has become the primary method of quantifying genetic diversity and studying selection in livestock. SNPs have several advantages over microsatellite markers including; i) greater abundance across the genome, providing greater genomic resolution, ii) evolutionarily stable generation to generation (i.e., change less frequently than microsatellites), iii) simpler nomenclature and greater suitability for high throughput automated analysis and data interpretation, iv) found in coding regions of genes, and v) can be more easily linked to QTLs and traits of interest for generation of highly accurate and targeted genomic selection programmes (Vignal *et al.*, 2002; Fernández *et al.*, 2013; Cortes, Cañon and Gama, 2022; Saravanan *et al.*, 2022; FAO 2023). Although the cost of whole genome sequencing (WGS) is dramatically decreasing, it is still expensive to carry out large scale studies. In order to commercialize genomic selection, genotyping microarrays have been

used as the vector of choice for SNP analysis. SNP arrays are produced by fixing a subset of probes relating to highly polymorphic SNPs from a population to a platform. These arrays allow livestock breeders to avoid the expense of WGS whilst capturing a high resolution of the genome for a large number of individuals (Fan *et al.*, 2010; Gurgul *et al.*, 2014). The highly polymorphic state of SNPs chosen allows differentiation between individuals, and these differences can be attributed to variation in production outputs and phenotypic changes (Fan *et al.*, 2010; Gurgul *et al.*, 2014). Therefore, SNP arrays are an important method in guiding breeding management plans and development of livestock, as well as finding target genes for future livestock function.

Although the choice of SNPs is incredibly useful for livestock management, arrays result in ascertainment bias and do not represent the natural levels of genetic variation (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; McTavish and Hillis, 2015; Quinto-Cortés *et al.*, 2018). Selection of SNPs for microarrays is often based upon a small number of sequenced DNA samples. In the livestock sector, SNP selection for microarrays is usually restricted to a small number of breeds of interest. The deliberate selection of highly polymorphic SNPs known from characterization of only a few breeds means that these markers likely represent ancestral variation maintained since before breed divergence. Furthermore, genotyping of only previously known polymorphic sites means that no additional genetic variation can be discovered (and hence genotyped) across the species (Nielsen, 2004; Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Geibel *et al.*, 2021). Ascertainment bias within arrays lead to distortions in allele frequencies, and overinflation of statistics compared to methods based on sequencing data, and underrepresentation of genetic variation in populations not captured in SNP discovery for array production (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018; Benjelloun *et al.*, 2019; Geibel *et al.*, 2021). Therefore, when conducting evolutionary studies, ascertainment bias needs to be accounted for.

1.5. Application of Genomic Data in Livestock

Assessment of genetic variation is essential for livestock management. Understanding genetic variation allows humans to mitigate harmful genetic effects such as inbreeding or genetic erosion whilst maintaining genetically resilient and adaptable livestock (Bruford *et al.*,

2015; Leroy *et al.*, 2018; FAO, 2023). Estimates of genetic variation across individuals and populations are typically calculated using statistics such as observed heterozygosity (H_o) displaying actual distributions of measured genotypes, expected heterozygosity (H_e) that is calculated on allele frequencies, and nucleotide diversity (π) that assesses the number of differences across sequences (FAO, 2023). Maintaining a high genetic diversity typically provides a species with the necessary genetic flexibility for adapting to different external pressures, with livestock being no exception (Notter, 1999; Barker, 2001; Ollivier and Foulley, 2009; Kristensen *et al.*, 2015). Species with low genetic variation are often vulnerable to extinction. Due to intensive artificial selection, many livestock breeds feature reduced genetic variation (Leroy *et al.*, 2018; FAO, 2019; Gicquel *et al.*, 2020). This will eventually lead to high levels of inbreeding which can be monitored by the inbreeding coefficient (*f*). Estimates can be calculated SNP by SNP, comparing the proportion of observed homozygous SNPs to expected, such as F implemented in PLINK (Li and Horitz, 1953; Chang *et al.*, 2015; Dadousis *et al.*, 2022). In livestock, inbreeding is now preferentially monitored by runs of homozygosity (via inbreeding coefficient for ROHs, F_{ROH}) that calculates the length of stretches of the genome with an absence of genetic variation (Curik, Ferenčaković and Sölkner, 2014; Peripolli *et al.*, 2017; Ceballos *et al.*, 2018; Meyermans *et al.*, 2020). Increased inbreeding increases the chance that recessive alleles become phenotypically active through homozygous genotypes, which in turn may result in harmful physical effects. Detrimental inbreeding among purebred livestock is often realised economically through reduced yields and reproductive issues (Leroy, 2014; Doekes, Bijma and Windig, 2021). Recovering from inbreeding depression is easily achieved through the introduction of new genetic variation.

Further statistical analysis can be carried out to understand how genetic variation compares, and how gene flow operates between populations. The outputs of genetic differentiation analysis give great insight into how best to manage interactions between breeds and populations. Several methods are often used to delineate individuals into genetically defined groups. Among the most used are F_{ST}, principal component analysis (PCA) and multidimensional scaling (MDS). F_{ST} evaluates the differences in allelic frequencies at each locus and can be summarized across all markers. PCA and MDS methods meanwhile allow for the condensing of large multi-locus genetic data into the generation of a few meaningful synthetic variables, termed components (FAO, 2023). These components will

capture variation between individuals. The level of genetic similarity between individuals is shown via the size of distance between points with admixture shown through intermediate individuals between ancestral populations. Model-based clustering algorithms (e.g., ADMIXTURE) that test the number of appropriate genetic clusters are also available to deduce ancestry, admixture, and relatedness between populations (Alexander, Novembre and Lange, 2009; FAO, 2023).

Although the methods above allow inferences to be made about species' history via patterns and abundance of genetic variation, they don't provide capacity for testing complex evolutionary processes. Incorporating genetic processes such as drift, mutation rate, and recombination facilitates modelling changes in genetic variation over time. For mitochondrial DNA and microsatellite markers, the use of tests such as Tajima's D, or Bottleneck and MSVar aid in identifying historic demographic trends such as population expansions and bottlenecks (Tajima, 1989; Beaumont, 1999; Piry, Luikart and Cornuet, 1999). Following the production of dense genomic SNP datasets, the popularity in using linkage disequilibrium (LD) has increased due to its sensitivity towards genetic processes (e.g., genetic drift, selection) (Slatkin, 2008; Qanbari, 2020). LD is typically higher between nearby loci, as the likelihood of recombination increases with physical distance along the chromosome. In evaluating the rate of decline in LD between loci at varying distances apart and encompassing recombination rate, N_e can be tracked backwards through time and reveal demographic trends and changes in genetic diversity (Corbin *et al.*, 2012). Among livestock SNP genotyping arrays, reconstruction of recent demographic history is frequently accomplished using N_e calculations such as that found in Sved (1971).

Exploring more ancient trends in N_e has been accomplished with WGS and the generation of pairwise sequential Markovian coalescence (PSMC) (Li and Durbin, 2011). With this, demographic trends have been generated for a variety of species through the Upper Pleistocene, revealing contraction and expansion events in populations often in accordance with glaciation periods (Frantz *et al.*, 2013; Qiu *et al.*, 2015; Gautier *et al.*, 2016; Weldenegodguad *et al.*, 2019; Luo *et al.*, 2020; Upadhyay *et al.*, 2021; Robin *et al.*, 2022). With increased computational power, additional genome sequences could be included leading to the creation of MSMC (Schiffels and Durbin, 2014). Other methods to model complex demography included using the allele frequency spectrum (AFS). Here, tree models

could be defined with fluctuating population sizes, divergence times, and gene flow and the tree most likely to match the real data was calculated with a likelihood approach (Excoffier *et al.*, 2013). Likelihood approaches are computationally difficult as the defined model becomes increasingly more complex (Marjoram and Tavaré, 2006; Csilléry *et al.*, 2010). Approximate Bayesian Computation (ABC) offers a flexible alternative. This method works under the assumption that the patterns of genetic variation for the sampled data are reflective of that of the real populations. Defined models are generated depicting historic events using coalescent simulations, providing a neutral model for comparison (Beaumont, 2010; Csilléry *et al.*, 2010). For example, in livestock, divergence times between domestic and wild species may represent time of domestication, and gene flow into domestic species may represent wild introgression. Millions of simulations are generated with each simulation, sampling different parameter values for population sizes, divergence times etc, and subsequent outputs are then statistically summarized. A rejection algorithm is then applied to obtain the most closely matched simulations (based on distances) to the observed data, and using this sample of simulations, generate posterior distributions for each parameter showing the most likely true values (Beaumont, 2010; Csilléry *et al.*, 2010). A high number of different tree models can be generated and compared using Bayesian statistics to identify the most appropriate model (Beaumont, 2010).

Studying demography typically uses neutrally evolving loci so that genetic data fits the assumptions of evolutionary theories. However, the patterns of genetic variation found across populations is not solely due to background processes of neutral evolution. Populations are continually interacting with changing conditions, be that of changing environments or interaction with new biota, or even alterations in behaviour within species. All these place selective pressures upon populations to adapt. Individuals with greater fitness are more likely to successfully reproduce, contributing to future gene pools. Genetic variation across the genome will fluctuate depending on favourable alleles at any given time. The result of selective pressure on a genome can be identified by the distinct footprints they leave, which are termed selection signatures. Understanding these signatures and the subsequent genes that are under selection is important for characterizing livestock and conserving important economic traits.

Positive, purifying, and balancing selection are three types of selection (Figure 1.2). All three alter allelic frequencies across the genome resulting in phenotypic variation. Under positive selection, the advantageous allele increases in frequency (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). The rate of which this occurs determines the extent that neutral variants physically linked to the advantageous allele also increase in frequency (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). The additional selection of linked variants either side of the advantageous allele is called genetic hitchhiking (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). These stretches of DNA under positive selection can be detected, as they present longer regions of increased LD (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). The prevalence of the advantageous allele in the population can provide us with information surrounding the strength of the selection pressure. This is considered under what is known as a selective sweep. Rapid selection of an allele resulting in near fixation within a population is considered a hard selective sweep, identifiable by fixated hitchhiking regions (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Softer sweeps that occur over a longer period allow for greater variation in neutral flanking regions (Gouveia *et al.*, 2014; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Purifying selection operates upon disadvantageous alleles whereby these alleles are selected against whilst generally not affecting the surrounding genetic variation (Gouveia *et al.*, 2014; Saravanan *et al.*, 2020). The third main form of selection, balancing selection, maintains multiple alleles in a locus and keeps high genetic diversity (Gouveia *et al.*, 2014; Saravanan *et al.*, 2020).

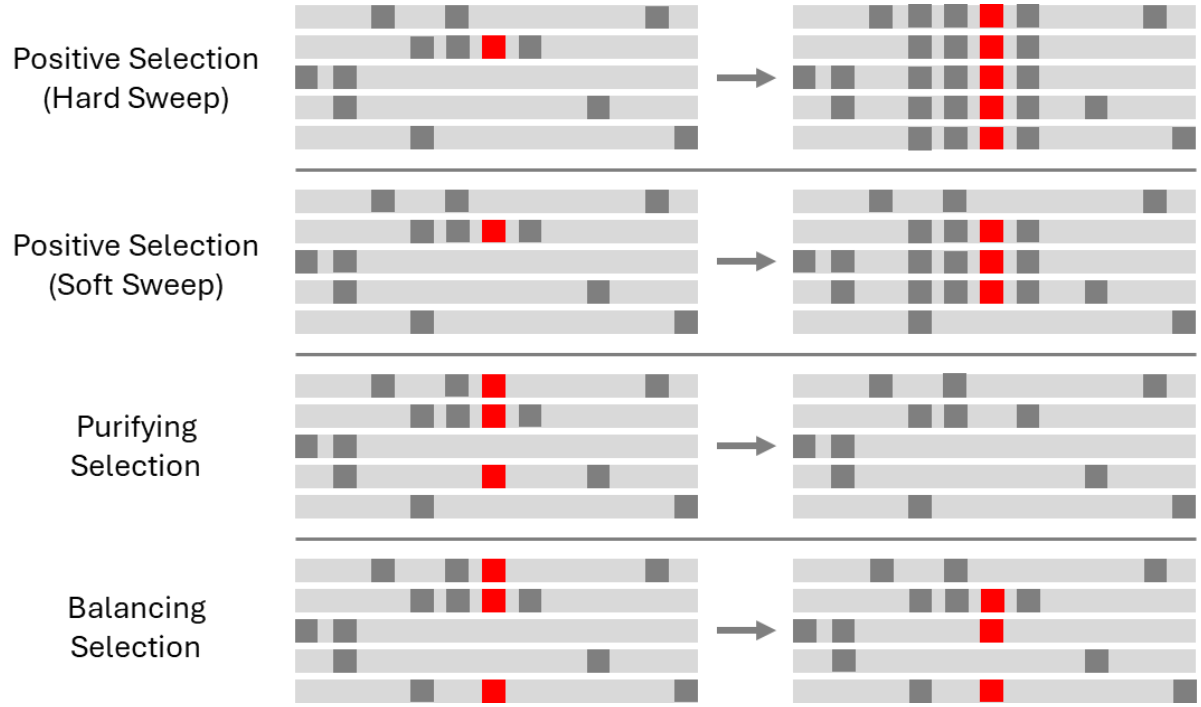


Figure 1.2: Figure displaying how different types of selection alter allele frequencies in a population. Red loci are alleles under selection while grey loci are not under selection. Positive selection causes an advantageous allele to increase in frequency. The rate at which this happens dictates whether a hard sweep (top) or soft sweep (2nd from top) has occurred. Positive selection will also cause nearby linked loci to increase in frequency leaving an extended region appearing under selection. Hard sweeps push allele frequencies towards fixation across a population whereas soft sweeps leave greater variation at the loci. Purifying selection (2nd from bottom) selects against a disadvantageous allele leaving surrounding variation. Balancing selection (bottom) maintains multiple alleles under selection and keeps genetic diversity high.

Selective breeding in livestock typically results in the increased frequency of breeding favourable individuals due to a beneficial trait. Therefore, artificial selection commonly operates under positive selection (Saravanan *et al.*, 2020). Negative and balancing selection also operate within artificial selection to a lesser extent. Due to the nature of positive selection, the haplotypes maintained in the genome can be easily detected within and between populations. Detection of positive selection for these two comparisons works on the basis of detecting haplotypes of increased frequency and LD from areas of normal genetic variation (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Within- or intra-population statistics do this across areas within a genome, whereas between- or inter-population

statistics detect this between the same area of different genomes. Several statistics are available to detect positive selection within populations. Site frequency spectrum (SFS) statistics such as Tajima's D and Fay and Wu's H-statistic can differentiate between areas of genome with an increased number of high and low frequency variants generated by positive selection, compared to areas of medium frequency variants (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Methods such as the extended haplotype homozygosity (EHH) and the integrated haplotype score (iHS) use linkage disequilibrium to find areas of high linkage (Sabeti *et al.*, 2002; Voight *et al.*, 2006). Runs of homozygosity can offer a method of detecting positive selection as these locate long homozygous tracts of depressed genetic variation within the genome that are maintained by selection (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Without selection, these tracts would be disrupted by gene flow, mutation, and recombination. Methods have been developed to identify differences between populations. F_{ST} is based on single site differences in allelic frequency and can detect older and fine scale selection meanwhile cross population (XP-EHH) uses linkage disequilibrium again to detect long haplotypes (Sabeti *et al.*, 2007; Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). The benefits of using inter-population statistics are that no prior history surrounding the population is required, such as determining the ancestral alleles (e.g., required for iHS) which can be difficult to identify without ancestral populations.

Soft selective sweeps are more difficult to differentiate from background genetic variation. Selection pressures for softer sweeps typically occur over a longer period of time and in response to a variable with a large variation over a spatial area (Saravanan *et al.*, 2020). Therefore, methods have been developed to identify regions under selection in comparison with metadata. In livestock, populations adapt to either environmental conditions or pressure from farmers developing production traits. Isolated populations naturally diverge via mechanisms such as genetic drift. Selection enacting differently on each population can further diverge populations. Methods such as PCAdapt can be used to detect loci that greatly contribute to major axis of variation and therefore, are likely to represent loci under selection (Luu, Bazin and Blum, 2017). However, such analysis does not include any further information as to why these are under selection. Methods detecting selection in relation to environmental variables include R-Samβada (Duruz *et al.*, 2019). With these, loci that enable individuals to be more greatly adapted to different climatic variables can be identified. For example, genes

relating to wax secretion in Moroccan sheep and heat stress in Lidia cattle were found associated with changes in precipitation and temperature respectively (Duruz *et al.*, 2019). Finding loci relating to production traits is important for maintaining development of livestock for greater yields and efficiency. Though, determining precise genetic functions underpinning complex biological processes is incredibly difficult due pleiotropic effects, and that phenotypes are often determined by numerous loci (Boyle, Li and Pritchard, 2017; Sella and Barton, 2019; Barton, 2022). As such, genes under selection may not make immediate sense in relation to livestock functions. Genome wide association studies are frequently conducted in livestock allowing alleles to be identified that account for greater effects in a desired trait (Sharma *et al.*, 2015; Saravanan, Panigrahi, *et al.*, 2022).

1.6. Domestic Water Buffalo

Asian water buffaloes (Genus *Bubalus*) evolved across Eurasia since divergence from their African counterparts, *Syncerus*, approximately 5 – 8.5 million years ago (Tanaka *et al.*, 1996; MacEachern, McEwan and Goddard, 2009a; Bibi, 2013). Today, four species of extant wild buffaloes (all endangered) are present comprising of the wild Asian water buffalo (*B. arnee*) in Northern India, Tamaraw (*B. mindorensis*) on the Philippines, and two species of Anoa (Mountain: *B. quarlesi* & Lowland: *B. depressicornis*) on the Indonesian island of Sulawesi. Two additional domestic species of buffalo (*B. bubalis*) can be found, both of which were independently domesticated from the wild Asian water buffalo. The riverine form (*B. bubalis bubalis*) was domesticated in the Indus valley, Northwestern India, approximately 6,000 years ago, while the swamp form (*Bubalus bubalis carabanensis*) was domesticated 4,000 years ago in Northern Thailand (Kumar *et al.*, 2007; Wang *et al.*, 2017). The two domestic forms can be phenotypically (e.g., size, coat colour markings) and genetically separated. The most prominent difference between river and swamp buffaloes is that the two species feature different chromosomal numbers (riverine 2n = 24; swamp 2n = 23), with swamp buffalo displaying a fusion event between chromosomes 2 and 3 in river buffalo (Ulbrich & Fischer 1967; Fischer & Ulbrich 1968; Iannuzzi 1998). Despite chromosomal differences, river and swamp buffalo are fully capable of interbreeding to produce fertile offspring (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). Less than 5,000 wild Asian water buffaloes remain, 90% of which reside along wetlands habitats in Northern India (Kaul *et al.,* 2019). Remnant populations may exist in Thailand, Cambodia, and Myanmar,

however the progenitor populations to the swamp buffalo may now be extinct (Kaul *et al.,* 2019).

Domestic buffaloes number 200 million globally (FAO STAT). The distribution of worldwide buffalo populations can be found in Figure 1.3. River buffaloes account for three quarters of total buffalo population, with 110 million buffaloes found in India alone. The river buffalo was domesticated primarily for dairy production as a vital source of nutrition and energy for families (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). The historic range of the river buffalo stretches west across the Middle East and coastal Mediterranean Europe reaching Italy and Romania, and with populations occurring around wetlands areas (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020; Wordsworth *et al.*, 2021). River buffaloes have been further transported over the course of the 20th century, spreading further into Europe, and exported to South America and Southeast Asia (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). Few sporadic populations are established in Africa. Swamp buffalo make up a smaller proportion as these were primarily domesticated for draught (Zhang, Colli and Barker, 2020). Their domestication appears to occur in tandem with the spread of rice cultivation across Southeast Asia (Setyaningsih *et al.*, 2019; Sun, Wang, *et al.*, 2020). The bulk of swamp buffalo population resides in China with their distribution spreading down through continental and islandic Southeast Asia.

Despite the relatively small population size compared to cattle (1.5bn), domestic buffaloes are an incredibly important livestock species for people across Southern Asia (Mishra *et al.*, 2015). Buffaloes are renowned for their environmental hardiness. Being native to the tropics, they are inherently adapted to the local conditions through a combination of physiological and behavioural traits. Most importantly, domestic buffaloes are productive under high temperatures and humidities (Marai and Haeeb, 2010; Yáñez *et al.*, 2020). Buffalo skin is covered with a thick epidermis with high concentrations of melanin (Marai and Haeeb, 2010). This combination captures UV rays to prevent penetration and damage to lower tissues (Marai and Haeeb, 2010). However, their black coloration and poor sweating capabilities make them vulnerable to heat stress from excessive exposure to sunlight (Marai and Haeeb, 2010). Buffaloes are adapted to mitigating heat stress by bathing in freshwater as a sebum

excretion provides a protective layer against water and chemical absorption (Marai and Haeeb, 2010). This explains their historical distributions surrounding wetlands areas.

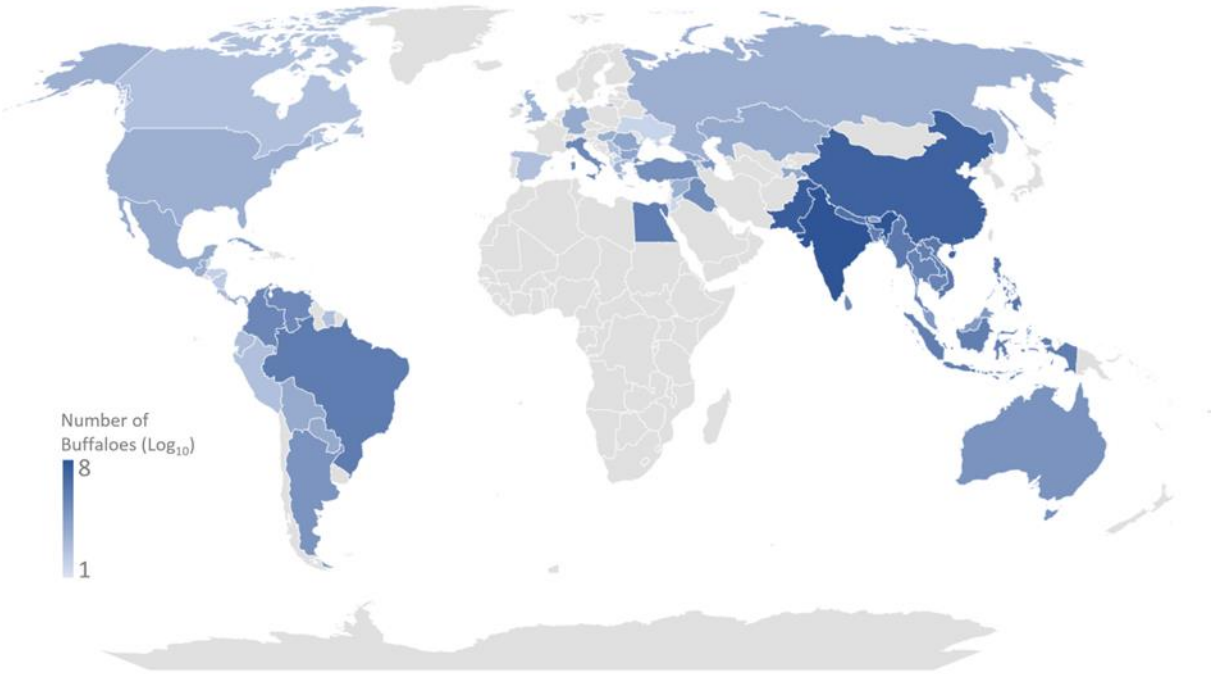


Figure 1.3: Global distribution of domestic water buffaloes. Blue shading indicates the absolute number of domestic water buffalo found in each country. Lighter blue indicates a fewer number of buffaloes with darker blue indicating a greater number of buffaloes. Number of buffaloes in the key is given in $\log_{10}$ scale. Grey countries indicate countries with zero domestic water buffaloes or no data present.

Tropical environments host a great diversity of pathogenic organisms that buffaloes are frequently exposed to via freshwater sources and rural environments. The ability to resist pathogens is a major asset for buffaloes that are frequently observed to be less vulnerable to disease than cattle (Cockrill, 1981; Villanueva *et al.*, 2018; Bertoni *et al.,* 2020; Kamaruddin *et al.*, 2021). This is particularly noticeable for mastitis as river buffalo appear to have a lower predisposition to infection due to longer, thicker teats with a narrower canal and tighter sphincter than dairy cows (Bertoni *et al*., 2020). Consequently, this makes milk ejection more difficult (Bertoni *et al.,* 2020). Wallowing behaviour may also be an important defensive mechanism against ectoparasites by disrupting their life cycles (Bertoni *et al.,* 2020). Livestock, including buffaloes, are common among marginal lands where reliable crop growth is not possible. Here, buffaloes excel in extracting nutrients from unproductive environments

due to differences in digestive system morphology and microbial composition in comparison to cattle (Bertoni *et al.,* 2020).

The traits outlined above make swamp buffalo a resilient species for draught, while river buffalo maintain milk production as environmental quality decline. Therefore, buffaloes are important assets for rural families across Southern Asia. Focusing on the river buffalo, this domestic species provides a source of high energy milk for nutrition, an economic asset down generations, and a buffer to times of crises (Nanda and Nakao, 2003; Hegde, 2019). Buffalo milk is rich in fat and protein with greater mineral and vitamin concentrations than cattle milk (Guo and Hendricks, 2010). Therefore, buffalo milk is energy dense which is ideal for these rural communities (Guo and Hendricks, 2010). Families use this milk for sustenance, however in times of plentiful supply, buffalo milk provides an additional source of income (Nanda and Nakao, 2003). Livestock are an important source of wealth, particularly for rural communities, as they are an asset that can be maintained down generations (Nanda and Nakao, 2003). In times of crises (e.g., droughts) buffaloes can be sold and traded in order to obtain more food for families (Clingingsmith and Williamson, 2008; Venot, Reddy and Umapathy, 2010). This has led to buffaloes being called a 'living bank' and the highly productive murrah and nili-ravi breeds have earned the name of 'black gold' (Bilal, Suleman and Kakar, 2006; Hegde, 2019; Kumar *et al.*, 2023).

Domestic buffaloes are traditionally farmed by small holding landowners. Buffaloes are largely farmed in extensive or semi-extensive systems as farmers allow them to freely graze upon common land or within crop rotations (Chantalakhana & Skunmun, 1999; Kumar and Singh, 2010; Bertoni *et al.,* 2020). As a result, buffaloes have not been subject to intense management and have few defined breeds in comparison to more abundant livestock species (e.g. cows). In India, home to the river buffalo, there are 13 native breeds, yet most of the buffalo population has been subject to admixture as 43% are classed non-type, with a further 39% classed as graded (crossbreed) (DHAD 2013). Swamp buffalo rarely feature defined breeds, falling under a ubiquitous term of carabao. Instead, swamp buffaloes are frequently classed by location (e.g., Chinese provinces). In theory, the lack of intense management should provide buffaloes with high levels of genetic variation with scope to improve productivity through genomic selection.

Intensification of buffalo farming has yet to occur as few modern large-scale farms exist with more than 300 individuals. This is likely due to the establishment of more popular and productive livestock species prior to the arrival of river buffaloes. In India and Pakistan, river buffaloes display their capabilities as livestock. River buffaloes contribute 57% and 68% respectively of the milk produced despite only composing of 36% and 50% of the bovine populations (Murtaza, Pandya and Khan, 2017; FAOSTAT). However, with average milk yields of 1,000 – 2,500 kg milk per 305-day lactation period, river buffaloes fall short of the productivity of commercial Western taurine cattle breeds (>5,000kg) (van Arendonk and Liinamo, 2003; Oltenacu and Broom, 2010; Borghese, 2013a). Despite this, river buffaloes have recently been incorporated into modern farming systems. This is often under the assumption that the greater protein, fat, and mineral content of buffalo milk gives a higher quality product than cattle. For example, Italy farms highly productive Mediterranean buffaloes as high-quality mozzarella can be produced from their milk (Zicarelli, 2004; Borghese, 2013). River buffalo numbers are increasing due to their importance to India and Pakistan and additional uptake in other countries(Zhang, Colli and Barker, 2020). In contrast, swamp buffalo are in decline as their draught use is replaced by mechanization (Zhang, Colli and Barker, 2020). Outside of draught, swamp buffalo are not highly valued as livestock as their smaller carcasses produce less meat, and their low milk yield (600kg) is not competitive (Borghese, 2011; Borghese and Moioli, 2016). Therefore, farming enterprises focused on swamp buffalo are rare. However, across China and Philippines, swamp buffalo are being repurposed into meat and milk producers via assistance through crossbreeding with river buffaloes (Zhang, Colli and Barker, 2020).

Without demand on a commercial scale, research into domestic water buffaloes and the generation of resources has lagged compared to high-profile livestock. In the field of genetics, studies began evaluating the phylogenetic relationships between domestic buffaloes, and various wild species in the late 1990s through mitochondrial haplotypes and restriction fragment length polymorphisms (Barker *et al.,* 1996; Tanaka *et al.*, 1996). Through the 2000s, genetic research began quantifying the levels of diversity across buffaloes using microsatellite markers and expanded exploration of their domestication origins with mitochondrial DNA (Kierstein *et al.*, 2004; Lei *et al.*, 2007; Kumar *et al.*, 2007; Kumar *et al.*,

2007; Zhang *et al.*, 2007). However, studies had not found their way into guiding future management and development of buffaloes as livestock.

Around 2010, the NGS revolution came into effect and the generation of genomic resources became accessible. For livestock, the production of SNP genotyping microarrays underpinned the development of genomic selection programs. For cattle, the Illumina BovineSNP50 BeadChip featured approximately 54,000 SNPs and genotyped 24 individuals at any one time (Matukumalli *et al.*, 2009). Sheep and pigs both had arrays in 2009, while a goat array was developed in 2014 (Kijas *et al.*, 2009; Ramos *et al.*, 2009; Tosser-Klopp *et al.*, 2014). High density SNP arrays are now available with the Illumina BovineHD Genotyping BeadChip delivering more than 777,000 SNPs, though human genotyping arrays now exceed 4 million SNPs (Verlouw *et al.*, 2021). Alongside this development, increasing WGS of livestock is continuously being undertaken revealing a wealth of information for genomic research.

For buffaloes, the introduction of genomic resources began with the formation of a completed draft genome in 2017, before a high quality genome was developed a year later (J. L. Williams *et al.*, 2017; Low *et al.*, 2019). The first commercial genotyping array for buffaloes became available in 2017 featuring approximately 90,000 SNP markers (Iamartino *et al.*, 2017). The production of this array enabled the formation of the first genomic selection programs led mostly by Italian and Brazilian farms (de Camargo *et al.*, 2015; Iamartino *et al.*, 2017; J. J. Liu *et al.*, 2018; Cesarani *et al.*, 2021; Lázaro *et al.*, 2021). Since then, characterization of genomic resources across domestic buffaloes has begun to take place, exploring levels of genetic diversity worldwide, their evolutionary history, and gene under selection or those that are commercially relevant (Wang *et al.*, 2017; Colli, Milanesi, Vajana, *et al.*, 2018; Mokhber *et al.*, 2018, 2019; Fallahi *et al.*, 2020; Luo *et al.*, 2020; Sun, Huang, *et al.*, 2020; Sun, Shen, *et al.*, 2020; Macciotta *et al.*, 2021). The caveat though is that genomic studies were biased towards the more commercially relevant river buffaloes, though recent generation of a swamp specific reference genome and WGS has given genomic accessibility to swamp buffalo (Iamartino *et al.*, 2017; Colli, Milanesi, Vajana, *et al.*, 2018; Luo *et al.*, 2020).

The incorporation of genomic research into buffalo farming will aid in fulfilling their potential as livestock. Physiological studies are developing greater milk production (Catillo *et al.*, 2002; Salari, Altomonte and Martini, 2013; Valsalan *et al.*, 2014; Costa *et al.*, 2021; Eldawy *et al.*, 2021), meat production (Ekiz *et al.*, 2018; Guerrero Legarreta *et al.*, 2020), feed

efficiency (Subhashchandra bose *et al.*, 2014; Negesse, Datt and Kundu, 2016; Sharma *et al.*, 2018), reproductive success (Barile, 2005; Baruselli *et al.*, 2023; Nava-Trujillo *et al.*, no date), disease resilience (Locatelli *et al.*, 2013; Kaur *et al.*, 2016; Catozzi *et al.*, 2017), and heat stress (Petrocchi Jasinski *et al.*, 2023) among others. Genomic studies in buffaloes will unearth the gene-phenotype interactions behind these observations.

As climate change enacts its effects, average global temperatures will increase, freshwater sources fluctuate, greater exposure to pathogenic organisms, and the reliable growth of crops become unfeasible due to extremes in weather. These effects will threaten livestock productivity meaning that livestock will either need to adapt, or farmers import new breeds and species that already feature useful adaptations. Additionally, with a growing human population and increasing demand for livestock derived products, farms will need to become more efficient to increase output. Water buffalo may offer a timely contribution to future food security. Their environmental resilience is a great asset in the face of climatic instability with heat tolerance, disease resistance, and nutrient conversion adaptations, particularly in those areas where crop growth declines. Farmers in the Philippines are already switching from crops to buffaloes due to extreme weather making crop farming unviable (Escarcha *et al.*, 2020). Any uptake in buffalo farming may require some adaptations or breeding to increase production. Fortunately, buffaloes are unlikely to have undergone any intensive selection and likely harbour high levels of genetic diversity. Initial evidence suggests that there has been continual uptake in genetic variation in river buffaloes since domestication through genetic exchange with wild buffaloes (Nagarajan, Nimisha and Kumar, 2015). An abundance of diversity provides an ideal foundation to generate selection programs and target breeding towards more productive buffaloes to make commercial farming of buffaloes more competitive with already established species. Additional traits such as lower methane emissions from buffaloes compared to cattle may contribute to minimizing GHG contributions (Mendoza *et al.*, 2020). Recent generation of genomic resources for domestic buffaloes means that the potential of buffaloes can now begin to be fully realized.

1.7. Thesis Aims

This thesis contributes to the growing body of genomic research surrounding domestic water buffaloes. The aims surround two general themes. Firstly, the demography of various

domestic water buffalo populations is reconstructed to understand the processes that shaped current patterns of genetic variation. The second objective is to assess the adaptive potential of buffaloes. This involves identifying genes relevant for buffalo adaptation and may be useful for targets of selection for future buffalo function. These aims were both completed using a commercially available Axiom^{TM} Buffalo Genotyping Array. This array is currently the most accessible means to large scale population genomic data for buffaloes and is relevant to developing new breeding management plans utilizing genomic selection. The output of this thesis provides characterization of genetic resources for an understudied livestock species that is crucial for people across Southern Asia.

1.7.1. Demographic Processes

Quantifying genetic diversity within a unit and understanding variation across a spatial structure is imperative to forming management plans for conservation and breed development. Intricate genetic, environmental, and biological processes have shaped current genetic variation, and for livestock, this is further complicated by artificial processes under human control. Elucidating patterns of divergence and gene flow within a livestock species aid in understanding how patterns of genetic variation were developed. Chapters Two and Three identify the pattern of genetic variation across two commercially relevant breeds, the murrah and Mediterranean riverine buffaloes, due to their great milk yields (Borghese, 2005; Borghese, Chiariotta & Barile, 2022). Here, differences in genetic variation are evaluated in the context of recent formation of new populations. Chapter Two quantifies the levels of genetic diversity of Mediterranean buffalo retained by two farms in United Kingdom since importation from continental Europe, whilst Chapter Three identifies this over global populations of murrah buffalo. This will give indications into the levels of overall genetic diversity retained and as such, give insights into their genetic health and potential adaptability. Population structure analysis is then completed to determine the extent of genetic differentiation across Mediterranean (Chapter Two) and murrah (Chapter Three) populations before calculation of N_e is used to reconstruct recent trends in diversity through time. In Chapter Two, we know that UK farmers typically import buffaloes from two sources, Romania and Italy. Therefore, we further explore their recent history through identity by state, and ancestry modelling for each farm, determining genetic proportions descended

from each reference population. This work will help guide future management plans of buffaloes in guiding where to import new individuals from.

Chapters Four and Five expand analyses across a global dataset of domestic buffaloes comprising both riverine and swamp species. SNP genotyping arrays suffer from two sources of ascertainment bias; i) selected SNPs are typically highly polymorphic in order to differentiate between closely related individuals of the same population, thus distorting allele frequency spectrums that do not represent normal standing genetic variation, and ii) sub-populations not included in SNP selection will be underrepresented in genetic diversities as specific polymorphic SNPs to those populations are not included. In the buffalo genotyping array used here, this ascertainment bias is exacerbated as swamp buffalo were not included in SNP selection, which up until genetic evidence, was unclear whether river and swamp buffalo were of the same species or not. Chapter Four therefore investigates how to handle ascertainment bias, and more specifically, whether differences in genetic variation between river and swamp buffaloes can be overcome using linkage disequilibrium pruning. Using the outcome of Chapter Four, deeper evolutionary relationships between river populations and swamp buffalo are explored using Approximate Bayesian Computation (ABC). Divergence times, gene flow, and N_e backwards through time are calculated using ABC modelling to decipher the extent of differentiation between riverine and swamp buffaloes. Additionally, times of migrations out of India were tested for riverine buffaloes that established core historical populations in the Middle East and Europe.

1.7.2. Selection Signatures

Genomic selection is integral to maximizing productivity in livestock and conserving important regions of the genome. The majority of selection studies in livestock identify candidate genes for production traits, however with rapidly changing climatic conditions, finding alleles linked to local adaptivity of environments is vital for ensuring livestock populations remain phenotypically plastic, healthy, and productive. In Chapter Two, regions under positive selection are identified using runs of homozygosity (ROHs) in each UK population. Genes found within these regions were identified for gene function, protein interactions, prevalence within biological processes, and known quantitative trait loci (QTL) associated. This gives indication into whether UK buffalo are genetically responding to

modern farming systems. Adaptation to the UK was further assessed by identifying any genes under positive selection in both populations using XP-EHH. To assist with identifying hitchhiking regions, average LD was calculated across the genome for each UK population, giving additional insights into potential genomic selection efficiency. Furthermore, LD distributions were compared between UK populations to determine whether future potential selection mechanisms were easily transferrable between populations. The same was done for murrah populations in Chapter Three however, results were further filtered. The goal of this chapter was to find evidence of rapid recent selection and adaptation to new environments. Therefore, regions under selection were only studied if they were unique to specific murrah populations (established <100 years). Long haplotypic methods such as ROHs and XP-EHH were used to identify hard selective sweeps, meanwhile softer selective sweeps were identified using methods such as PCAdapt and R-Samβada, linking selection to population structure and environmental adaptation, respectively. This work will add new information into genomic adaptability of buffaloes which would back up physiological studies. Finally, as part of Chapter Five, regions of divergence or balancing selection were identified that may provide functional reasons as to how domestic buffaloes historically diverged to become different species yet are fully capable of interbreeding.

Chapter Two

Demography & Selection in UK Water Buffalo

2.1. Abstract

A few thousand Mediterranean water buffalo occupy a small sector of the livestock industry within the United Kingdom (UK). Despite living in the UK for around 60 years, public awareness about the species remains low. It is imperative to support small industries and conserve genetic diversity across livestock, aiding in breed improvement whilst preserving their genetic health and adaptive potential. This study uncovers the UK buffaloes' genetic diversity, their European ancestry, and any genes under selection. Water buffalo were sampled across two farms in the UK (RV_UK1 and RV_UK2) and genotyped using the Axiom™ 90K Buffalo SNP Genotyping Array, generating 63,603 high-quality SNP markers after quality control. High levels of genetic diversity were found in both UK populations with no significant decline of genetic diversity in UK buffaloes since importation. In accordance, there was no inbreeding in the UK populations and their divergence from Europe was low to moderate. Population structure was driven by Eastern (e.g., Romania) or Western (e.g., Italy) European origins with UK buffaloes found closely associated to Italy. Discrepancies in ancestry between the two UK populations appears due to genetic drift in RV_UK2. Runs of homozygosity (ROHs) under selection in each population suggests alternative management of populations, likely reflecting different intensities of each farm. Quantitative trait loci (QTLs) revealed an excess of production traits in RV_UK1 in line with an expanding modern farm needing to generate high yielding buffalo. Meanwhile RV_UK2 feature an excess of milk associated QTLs under selection which would be the classical expectation of dairy buffaloes. Selective sweeps present in both UK populations found several genes related to immunity, milk production, and thermoregulation which may indicate important regions to increasing production in buffaloes or adaptation from exposure to a new environment and pathogenic community.

2.2. Introduction

Domestic water buffalo are famed for their nutrient rich milk. However, they are a little-known domestic species across the United Kingdom (UK). This is down to their local rarity as the population size of water buffaloes in the UK is dwarfed by traditional British livestock. Since 2013, the water buffalo population in Great Britain has declined 21.5% from around 4,000 individuals to 3,000 (Figure 2.1). In comparison, there were 9.6 million cattle recorded in 2022 (Department for Environment Food & Rural Affairs, 2022). Despite a decline in the population size, a handful of ambitious British farmers aim to establish an economically viable buffalo farming industry amongst the UK livestock landscape. To achieve this, water buffaloes will be pushed towards modern large-scale farming (greater than 300 individuals) to increase productivity, a farming intensity that is uncommon across global traditional buffalo farming.

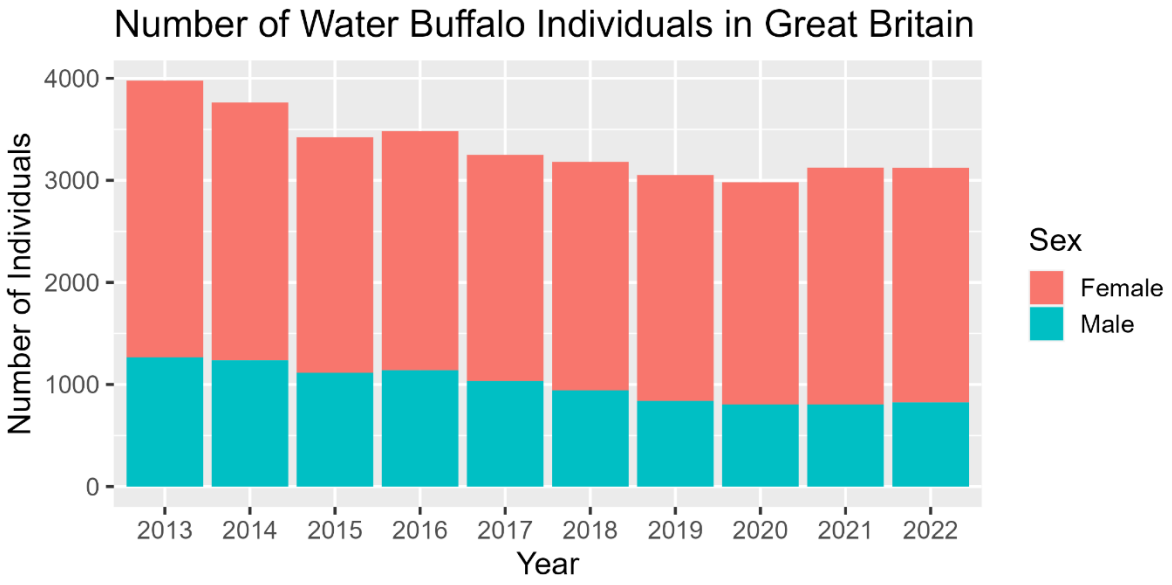


Figure 2.1: Water buffalo population size in Great Britain from 2013 – 2022 for males (blue) and females (red). Data provided by the British Cattle Movement Service and are correct as of 02/05/2023. Data does not include Northern Ireland.

The Mediterranean water buffalo is the sole buffalo breed currently found in the UK. This breed falls under the river buffalo (*Bubalus bubalis bubalis*) species domesticated from the wild Asian water buffalo (*Bubalus arnee*) in India (Satish Kumar *et al.*, 2007; Colli, Milanesi, Vajana, *et al.*, 2018). Mediterranean buffalo were historically found in the Campania region

of Italy and Romanian areas surrounding the Danube River and Carpathian Mountains. River buffalo from India likely reached the ancient Mesopotamia region across present day Middle East around the 5th Century by way of settlement of the Indian Al-Zutt community (Wordsworth *et al.*, 2021). Migration through Saracen trading routes from the 7th Century likely facilitated establishment of buffalo populations in Italy and Romania to become the Mediterranean breed known today (Colli, Milanesi, Vajana, *et al.*, 2018; Wordsworth *et al.*, 2021). It does not appear that Mediterranean buffalo have undergone any historical admixture, however there are reports that some crusaders returned to Europe with buffaloes in the 12th and 13th centuries (Colli, Milanesi, Vajana, *et al.*, 2018). Expansion and trade of water buffaloes from their European origins into modern farming across Europe and UK has mostly occurred since the second half of the 20th Century, albeit slowly and in low numbers. With more than 400,000 buffalo, Italy features by far the strongest European industry as no other country in Europe harbours more than 15,000 individuals (Minervino *et al.*, 2020; Cesarani *et al.*, 2021).

Mediterranean buffaloes are among the most productive diary buffalo breeds alongside the murrah and nili-ravi of India and Pakistan, respectively. These breeds produce, on average, more than 2,000kg of milk per 270 days of lactation with many elite individuals producing over 5,000kg (Borghese, 2005). Most buffalo outside of intensively managed herds produce less than 2,000kg of milk per 270 days of lactation (Borghese, 2005). Buffalo milk is valued for the higher concentration of solids than cattle milk, particularly the great fat and protein content (Guo and Hendricks, 2010). This results in buffalo milk being a source of high energy content understating why river buffalo are popular and vital among the rural communities in India and Pakistan (Guo and Hendricks, 2010). In Europe, these qualities provide ideal coagulation properties for the production of high end luxury mozzarella cheese (Costa *et al.*, 2020). To a lesser extent, European buffaloes are used as a source of meat. Buffalo meat can provide greater health benefits than cattle owing to lower fat content, cholesterol, myristic and palmitic acids whilst retaining similar protein levels (Guerrero Legarreta *et al.*, 2020; Di Stasio and Brugiapaglia, 2021).

The nutrient content of buffalo milk alone has not been enough for Mediterranean buffaloes to become a widespread domestic species in Europe. Mediterranean buffalo persist as the only breed in Europe (Bulgarian murrah being a crossbreed between Mediterranean

and imported murrah) and likely haven't undergone the intense selection that other livestock species have (e.g., cattle and pigs). Mitochondrial analysis of river buffaloes shows a large variation in haplotypes and lack of bottleneck (Satish Kumar *et al.*, 2007; Nagarajan, Nimisha and Kumar, 2015). Comparatively, taurine cattle have resided in Europe since shortly after their domestication 10,000 years ago, and as such have become integrated into society across Europe diversifying into more than 400 breeds (Beja-Pereira *et al.*, 2006; Pitt, Sevane, *et al.*, 2019). The result of this is that cattle can be successfully farmed across all environments in Europe and commercial breeds have undergone intense selection to maximise yields (Consortium, 2006; Soini *et al.*, 2012). The Holstein Friesian cattle is widely regarded as the most productive cattle breed. The milk producing capacity of Holstein cattle vastly exceeds that of Mediterranean buffalo at an average 10,000kg of milk per 305 days of lactation, enabling an economically cheaper production of milk (Hansen, 1999; Breider, 2019; Piwczyński, Brzozowski and Sitkowska, 2020; Radwan, El Qaliouby and Elfadl, 2020). At these quantities of produce, it is of no surprise that Holstein cattle are globally farmed at an industrial scale.

From a meat perspective, buffalo meat has struggled to become popular as poor sensory characteristics make for unpleasant experience. Traditionally, buffalo meat was harvested when culling old buffaloes which generates low quality meat that is already tougher and darker than cattle meat due to its healthier properties cattle (Guerrero Legarreta *et al.*, 2020; Di Stasio and Brugiapaglia, 2021). The use of optimized diets and low slaughter age can generate meat more comparable to cattle, however lower growth rates and dressing percentages makes an economically viable buffalo meat production difficult to maintain (Guerrero Legarreta *et al.*, 2020; Di Stasio and Brugiapaglia, 2021).

The strength of the buffalo industry in Italy is therefore somewhat of an outlier in the farming landscape of Mediterranean buffalo. The productivity of Italian buffaloes is not maintained elsewhere as, for example, Mediterranean buffalo in Eastern Europe often produce approximately 1,500kg or less milk (Coroian *et al.*, 2011; Matiuti *et al.*, 2020; Popa *et al.*, no date). Italy's success in buffalo farming has been generated through multiple reasons including investment, organised breeding systems, availability of appropriate pastures, surplus of milk quotas, and value of product, and is now being further improved through genomic selection (Borghese, 2005). Italian buffalo milk is valued for its use in producing high

quality Mozzarella di Bufala Campana cheese, which has been recognised under the Protected Designation of Origin (PDO) certification (Costa *et al.*, 2020; Cesarani *et al.*, 2021). This investment into quality buffalo milk in Italy means the milk is two to three times more expensive than many other buffalo farming countries (Pisanu *et al.*, 2019). Therefore, to establish a thriving buffalo industry in the United Kingdom, British farmers will need to replicate the success of Italian buffalo industry, producing a highly productive herd stock and generate a valued product.

This chapter comprises of a genetic analysis of Mediterranean water buffaloes sampled from two farms in the United Kingdom using the Axiom™ 90K Buffalo SNP Genotyping Array (Iamartino *et al.*, 2017). The study here aims to genetically characterise and evaluate the current status of British buffaloes. To do so, the sampled farms were compared to a range of European populations, including countries of historical origin in Europe (represented by Italy and Romania), from two previous buffalo studies (Colli, Milanesi, Vajana, *et al.*, 2018; Noce *et al.*, 2021). The specific goals of this chapter are outlined as follows.

The first goal is to determine whether British buffaloes have lost any genetic diversity since importation to the UK. Small populations are particularly vulnerable to loss of genetic variation with following effects typically associated with reduced fitness and increased in health issues (Lande, 1988; Pekkala *et al.*, 2012, 2014). The contraction of a large population under decline, or in the example of British buffaloes, migration of a new sub-population can increase the chances of deleterious alleles becoming dominant (Kyriazis, Wayne and Lohmueller, 2021). Within a large population, these alleles are often hidden in heterozygous states at low frequencies with individuals showing no health risks (Pekkala *et al.*, 2014; Kyriazis, Wayne and Lohmueller, 2021). In small populations, rare deleterious alleles may persist at greater frequency and individuals carrying these alleles are more likely to reproduce. Shifts in allele frequencies that enable this phenomenon fall under the term of genetic drift which can be exacerbated through genetic bottlenecks (Lande, 1988; Broquet *et al.*, 2010; Angst *et al.*, 2022). Small populations such as the establishment of British buffaloes may be vulnerable to a genetic bottleneck or detrimental drift if they failed to capture enough diversity from the original host population. Therefore, this study firstly aims to quantify the amount of genetic diversity present within British buffaloes and the extent of genetic

differentiation from continental origin populations. The outcome of this allows inference into the genetic health and potential adaptability of British buffaloes.

Secondly, the ancestral origins of British buffaloes were determined. This goal seeks to map out their recent history uncovering any shifts in historical population sizes and any admixture from differing populations of Mediterranean buffalo. Analysing trends in effective population sizes back through time contributes to understanding if any major events have impacted a species current genetic variation (Lande, 1988; Wang, 2005; Hare *et al.*, 2011). This knowledge provides important contributions to thorough management plans for future development of breeds (Lande, 1988; Wang, 2005; Hare *et al.*, 2011). Farmers are likely to import new individuals from continental Europe to restock, expand, or develop their herds. Importation without genetic knowledge has the potential to incorporate damaging admixture into the population that can lead to genetic erosion (Rege and Gibson, 2003; Leroy *et al.*, 2018).

British buffaloes are potentially composed of a mixture of Italian and Romanian buffaloes. Although buffaloes from these countries are defined under the Mediterranean breed, populations within these countries have possibly been separated for much of their time in Europe and thus experienced different recent evolutionary histories. Italian Mediterranean buffaloes have undergone large scale development to increase productivity to become among the best buffalo milk producers, whilst Romanian buffalo may exhibit environmental adaptations to the cold climate of the Carpathian Mountain range (Borghese, 2013b; Yáñez *et al.*, 2020; Cesarani *et al.*, 2021). UK farmers initially imported cheap buffaloes from Eastern Europe to set up farms before importing from Italy to increase productivity (Borghese, 2013b). Substantial admixture between buffaloes of these two countries may lead to intermediate phenotypes that are neither as productive as Italian buffalo, or climate resilient like Romanian buffalo (Martínez *et al.*, 2012; Leroy *et al.*, 2018). This admixture is therefore harmful to successful development of farms and genetic analysis can identify any areas of genetic erosion that should be mitigated.

Finally, the third goal is to identify any regions across the genome under selection in each UK population that may be of importance for livestock function and preservation. Selection may be subjected upon British buffaloes in relation to a new environmental or a developmental pressure by humans. Identification of selection here is achieved through

analyses of long continuous homozygous genomic segments called runs of homozygosity (ROHs) and signatures of recent selective sweeps via cross population extended haplotype homozygosity (XP-EHH). Directional selection over time such as artificial selection leads to the preference of a limited number of superior individuals for breeding (Peripolli *et al.*, 2017). The continual use of these individuals leads to the inheritance of large identity-by-descent regions and a reduction in the diversity of haplotypes surrounding the target locus (Peripolli *et al.*, 2017). When ROHs are present at the same location across many individuals within a population, this suggests that the locus harbours important genes (Pemberton *et al.*, 2012). Therefore, significant genes found within ROHs were identified in British buffaloes and functional analysis conducted. ROH analysis targets longer tracts of the genome. Shorter targets of selection were identified via signatures of recent positive selection using an extended haplotype method. Strong positive selection on a locus can cause a sudden shift in frequency of the preferred allele towards fixation (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). The speed at which this happens causes flanking DNA to also be selected for, leaving a highly linked detectable segment (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). This form of signature infers recent selection as variation in the flanking DNA that is not under selection will be generated via recombination over time (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Therefore, detection of positive selection here is used to infer recent adaptation of buffaloes to the UK.

2.3. Materials and Methods

2.3.1. Sample Collection & Data Generation

Water buffalo were sampled at two farms from the UK. A total of 69 water buffalo were sampled using nasal swabs in compliance with home office laws. Farm 1 (RV_UK1 hereafter) had 41 water buffalo sampled while 28 were sampled from farm 2 (RV_UK2 hereafter). The nasal swabs were transferred to Neogene UK for sample preparation and genotyping using the Axiom™ Buffalo Genotyping Array featuring approximately 90,000 SNP markers (Iamartino *et al.*, 2017).

The raw genotyped data was analysed in Axiom™ Analysis Suite Software v4.03. 68 out of the 69 buffalo samples passed quality control measures for calling genotypes. 75,679

SNP genotypes were successfully called using quality control thresholds outlined in Table S2.1, and exported in the PLINK v1.9 PED/MAP format (Chang *et al.*, 2015). The dataset underwent further quality control in PLINK v1.9, tailoring the dataset to the requirements of each analytical test. SNPs were aligned to the water buffalo reference genome UOA_WB_1 (RefSeq: GCF_003121395.1) with non-autosomal SNPs and those with no known chromosomal locations being removed. SNPs and individuals with a call rate greater than 0.95 were retained, with additional removal of founders, giving an initial PLINK v1.9 QC dataset of 63,603 SNPs across 65 individuals. This dataset was used for runs of homozygosity analysis (Section 2.3.5) as these tests are less sensitive than association-based tests (e.g. genome-wide associations) to sporadic incorrect genotypes that are cautiously removed through applying a minor allele frequency (MAF) filter (Meyermans *et al.*, 2020). A MAF filter of 0.01 was then applied producing a dataset of 60,990 SNPs. This dataset was used for linkage disequilibrium (Section 2.3.3) and identity-by-descent (Section 2.3.4) analysis of UK water buffalo.

Publicly available datasets containing European water buffalo populations were taken from two published papers for comparison against UK buffaloes (Colli, Milanesi, Vajana, *et al.*, 2018; Noce *et al.*, 2021). These populations cover several European countries featuring multiple Mediterranean water buffalo populations and one Bulgarian murrah (Table 2.1). Populations from Colli *et al.*, (2018) and Noce *et al.*, (2021) underwent the same quality control process in PLINK v1.9 producing 46,888 and 61,766 SNPs respectively. Noce *et al.*, (2021) was merged with the UK buffalo dataset first, retaining 53,274 SNPs present in both datasets. Due to losing increasingly more SNPs as more datasets were combined, this dataset was used for detecting signatures of positive selection (Section 2.3.4) before being merged with Colli *et al.*, (2018). 40,695 SNPs were shared across all three datasets. This shared dataset was used for calculating genetic diversities, population structure, and gene flow (Section 2.3.2). Linkage disequilibrium pruning was conducted to generate independent markers to avoid overestimation of statistical values through multicollinearity effects via highly correlated SNPs (Malomane *et al.*, 2018). The first SNP of every linked pair of markers was removed using a sliding window approach of 50 SNPs, step size of 10 SNPs and an R^2 threshold of 0.1. This gave a dataset comprising of 7,222 SNPs for demographic analysis across European buffaloes. To ensure robust results, the dataset was randomly split in two with the second dataset used to results obtained in the first dataset.

Table 2.1: European water buffalo populations used and display of locations of each population, breed, sample sizes, and referenced studies.

Population ID	Country	Region	Breed	Sample Size	Reference
RV_UK1	UK	Somerset	Mediterranean	38	
RV_UK2	UK	Somerset		27	
RV_ITA	Italy		Mediterranean	15	(Colli, Milanesi, Vajana, *et al.*, 2018)
RV_ROM_CL	Romania			13	
RV_BUL_VAR	Bulgaria	Varna	Bulgarian Murrah	58	(Noce *et al.*, 2021)
RV_GER_BOR	Germany	Born	Mediterranean	28	
RV_GER_JUT		Jüterbog		27	
RV_GER_STA		Stadland		26	
RV_GER_WIE		Wiesenburg		28	
RV_HUN_CSA	Hungary	Csákvar		17	
RV_HUN_FOL		Földes		19	
RV_HUN_TIZ		Tiszataj		19	
RV_ROM_MER	Romania	Mera		16	
RV_ROM_SEC		Sercaia		47	

2.3.2. Genetic Diversities & Population Structure

Genetic diversity was summarised using inbreeding coefficient (F), observed (H_O), and expected (H_E) heterozygosity for each population in PLINK v1.9. Summaries (e.g., mean) and analysis of genetic diversities was completed in R v4.0.0 (R Core Team, 2018) using custom R scripts. A Welch's t-test was used to test for significant differences between H_O and H_E within each population because of non-normal data identified using the Shapiro-Wilk test. A Kruskal-Wallis was used to test for significant differences in H_O between all populations.

Population structure across European buffaloes was assessed using several methods. Pairwise population F_{ST} values were calculated across the dataset using Arlequin v3.5.2.2 (Excoffier and Lischer, 2010), with a neighbour-net network created in SplitsTree v4.14.4 (Huson and Bryant, 2006). Major divisions between individuals were assessed via a multidimensional scaling (MDS) analysis via PLINK using raw Hamming's distance to reduce the dataset to 20 dimensions. Genetic clustering of individuals was done using ADMIXTURE v1.3.0 testing for the ideal number of unique ancestral genetic clusters (K) across the dataset (Alexander, Novembre and Lange, 2009). Values of K from 1 to 20 were tested with five

repeats for each value of K. The preferred value of K was identified as the value with the lowest estimated cross-validation output. An AMOVA was used to further understand the distribution of variation observed across individuals or populations within the dataset via Arlequin. Treemix was used to detect the presence of gene flow between populations by incrementally adding migrations between populations (Pickrell and Pritchard, 2012). Like ADMIXTURE, this tests between a defined number of migrations (K) however, the ideal value of K was selected when variation explained across the dataset exceeded 99.8% in accordance with (Pickrell and Pritchard, 2012). Values of K were tested from 1 to 20.

2.3.3. Linkage Disequilibrium & Persistence of Phase

Linkage disequilibrium in the form of Pearson's r and r^2 was calculated between all pairwise SNPs separated no more than 10Mbp from each other in each UK population. This allowed calculation of the average linkage disequilibrium across the genome for identifying candidates of selection and visualising the recent demographic history of UK water buffalo. All linkage disequilibrium analyses post-PLINK were completed in R with custom scripts. Average genome linkage was calculated by grouping SNP pairs according to distance apart into 50kbp bins up to 1Mbp and taking the mean r^2 of each bin. This was then visualised in the form of a decay curve showing mean and variance of linkage as the mean distance between SNPs increased. The consistency of linkage disequilibrium in each UK population was compared using persistence of phase. Phase refers to the assignment of alleles to maternal and paternal chromosomes, while persistence of phase evaluates the order of alleles between populations using linkage disequilibrium. Closely related populations will likely have higher persistence of phase making transferability of genomic predictions between these populations easier (de Roos *et al.*, 2008; Wang *et al.*, 2013). Therefore, to calculate persistence of phase, Pearson's correlation was calculated on r values between SNPs (in PLINK) between both UK populations using custom R scripts. SNPs were grouped into 50kbp bins, and the mean and variance plotted.

2.3.4. Ancestry

Recent history of UK water buffalo was extracted with the following methods: i) Identity-by-descent (IBD) evaluating the relationships between all UK individuals, ii) r^2 linkage

values were used to calculate effective population size (N_e) of UK populations over recent history, and iii) the proportion of the genome relating to European ancestral populations was computed for both UK populations.

IBD between all UK individuals was calculated in PLINK and summarised in R using custom scripts. Welch's t-tests (due to non-normal data) were used to determine if there was a statistical difference in IBD between UK populations, as well as testing IBD between vs within populations. A network of IBD relationships between individuals greater than 10% was produced using network R package (Butts, 2008).

Population trends over time were estimated using N_e following Mokhber *et al.*, (2019). This was calculated using Sved (1971) N_e equation (1) implemented through custom R scripts that evaluates the average r^2 linkage values and distance, c, between SNPs in Morgans. Genetic distance was converted to physical distance assuming 1cM ~ 1Mb.

$$Ne = \left(\frac{1}{4c}\right)\left(\frac{1}{r^2} - 1\right) \quad (1)$$

A sample size correction was applied to r^2 using equation (2) below where n represents the number of haplotypes in the sample, and r^2 must be greater than 0, and less than 1.

$$Corrected\ r^2 = \frac{Computed\ r^2 - \frac{1}{n}}{1 - \frac{1}{n}} \quad (2)$$

Lastly, generation (t) corresponding to distance (c) between SNPs was calculated using equation (3).

$$t = \frac{1}{2c} \quad (3)$$

Water buffaloes from Italy and Romania have been imported to the UK. Local Ancestry in adMixed Populations (LAMP) was used to determine the proportions of the UK populations genomes deriving from the two ancestral origins, represented by RV_ITA (Italy), and RV_ROM_CL (Romania) (Baran *et al.*, 2012). Since ancestral populations were available, the

LAMP-ANC algorithm was employed. As UK water buffalo were imported later in the second half of the 20th century, the time since admixture was set to 8 generations, where 1 generation = 6 years (Mintoo *et al.*, 2019). Admixture proportions required were estimated by running ADMIXTURE for each UK population with RV_ITA and RV_ROM_CL when K = 2.

2.3.5. Runs of Homozygosity

Runs of Homozygosity (ROH) were identified in each UK population in PLINK under consideration of Meyermans *et al* (2020). Parameters used to identify ROHs are as follows: minimum SNP density of 80kbp, maximum gap between SNPs being 1000kbp, maximum of one heterozygous SNP, maximum of one missing SNP, and a ROH window threshold (defined as hit rate of a SNP in all scanning windows) of 0.05. The minimum number of SNPs per ROH (L) in each population was calculated using equation (4) below, where α represents the percentage of false positive ROH (0.05 here), n_s represents the number of SNPs, n_i the number of individuals within the population, and het equals the mean average heterozygosity across all SNPs (Lencz *et al.*, 2007; Purfield *et al.*, 2012).

$$L = \frac{log_e\left(\frac{\alpha}{n_s n_i}\right)}{log_e(1 - het)} \quad (4)$$

The minimum number of SNPs per ROH for each UK population resulted in 38 for RV_UK1, and 36 for RV_UK2. ROH PLINK outputs were analysed using detectRUNs package in R (Biscarini et al, 2019). Detected ROHs in each UK population were summarised using length, total number, ROH inbreeding coefficient (F_{ROH}, equation 5), number of ROH per class.

$$F_{ROH} = \frac{\sum ROH\ Length}{Genome\ Length} \quad (5)$$

As ROHs can be large and contain many genes, identifying the exact genes under selection is difficult. Therefore, a series of functional analyses were conducted to infer biological relevance. The top 0.01 SNPs occurring within ROHs across each UK population were retained as candidates of selection. Genes found close to these significant SNPs underwent the following analyses. i) Gene ontology using GOrilla (Eden *et al.*, 2009) was used to identify any enriched biological pathways. ii) Protein interactions between genes were

quantified using STRING (available at string-db.org) to detect genes that had numerous and strong interactions with other genes within ROHs. Lastly, quantitative trait loci (QTLs) that overlapped significant SNPs were identified. QTLs have not yet been extensively identified for water buffalo so instead, cattle QTLs that have been mapped to the water buffalo genome were used (Nascimento *et al.*, 2021). These QTLs were identified from *Bos taurus* UMD3.1.1 reference genome (RefSeq: GCF_000003055.6). A Fisher's exact test was used to test for significant over- and underrepresentation of each QTL class overlapping ROHs. The test evaluated the relationship between QTLs in ROHs and total number of QTLs.

2.3.6. Signatures of Selection

Cross-population extended haplotype homozygosity (XP-EHH) implemented in selscan (Szpiech and Hernandez, 2014) was used to identify recent positive selective sweeps that have occurred in UK buffaloes. XP-EHH is a bidirectional test that compares the strength of selection between two populations. RV_GER_STA was chosen as the reference population here due to its similarity to UK buffaloes (see Section 2.4.2). RV_UK1, RV_UK2, and RV_GER_STA were all phased in Beagle v3.3 (Browning and Browning, 2007). Each of the two UK populations were then tested against RV_GER_STA using XP-EHH under default settings in selscan. The results of XP-EHH were then standardized across the genome according to allele frequency bins using the norm package in selscan. In R, SNPs were ranked in descending order according to the selection score and a p-value generated. This p-value was then converted into negative $\log_{10}$ values, and a Manhattan plot was generated showing the strength of positive selection across each UK population. The top 0.01 SNPs with the greatest negative $\log_{10}$ values in each test were chosen as candidates for selection for each UK buffalo population. To reduce the risk of false positives, only selection signatures that occur in both UK populations were retained and inferred as adaptations to the UK. Genes linked to SNPs were reviewed to infer adaptations and livestock function.

2.4. Results

2.4.1. Genetic Diversities

Observed heterozygosity across European river buffalo populations ranged between 0.368 (RV_HUN_TIZ) and 0.442 (RV_ROM_CL) with a mean of 0.398 (±0.019), as seen in Table 2.2. All populations featured a higher H_O than H_E (ranging from 0.339 – 0.409; mean = 0.379 ±0.019), though no significant differences were found. The greater H_O values were reflected in the inbreeding coefficient, F, which was negative across all populations (-0.168 – -0.013), thus indicating an absence of inbreeding. As expected, both UK populations were among the lowest H_O with 0.392 (±0.157) and 0.385 (±0.174) for RV_UK1 and RV_UK2, respectively. Both H_O (χ^2 = 488.08, df = 13, p < 0.0001) and H_E (χ^2 = 799.74, df = 13, p < 0.0001) significantly differed across populations.

Table 2.2: Genetic diversities across European water buffalo populations. H_O = Observed Heterozygosity, H_E = Expected Heterozygosity, F = Inbreeding Coefficient.

Population	H_O (±SD)	H_E (±SD)	F (±SD)
RV_UK1	0.392 ± 0.157	0.387 ± 0.140	-0.013 ± 0.049
RV_UK2	0.385 ± 0.174	0.371 ± 0.149	-0.038 ± 0.086
RV_ITA	0.389 ± 0.182	0.378 ± 0.144	-0.028 ± 0.056
RV_ROM_CL	0.442 ± 0.214	0.383 ± 0.145	-0.168 ± 0.184
RV_BUL_VAR	0.420 ± 0.135	0.409 ± 0.117	-0.026 ± 0.046
RV_GER_BOR	0.404 ± 0.162	0.398 ± 0.136	-0.016 ± 0.058
RV_GER_JUT	0.400 ± 0.184	0.358 ± 0.148	-0.120 ± 0.099
RV_GER_STA	0.392 ± 0.167	0.387 ± 0.140	-0.015 ± 0.082
RV_GER_WIE	0.416 ± 0.172	0.394 ± 0.140	-0.058 ± 0.054
RV_HUN_CSA	0.373 ± 0.186	0.364 ± 0.155	-0.027 ± 0.073
RV_HUN_FOL	0.390 ± 0.191	0.361 ± 0.156	-0.081 ± 0.063
RV_HUN_TIZ	0.368 ± 0.191	0.339 ± 0.166	-0.087 ± 0.078
RV_ROM_MER	0.402 ± 0.172	0.395 ± 0.135	-0.017 ± 0.080
RV_ROM_SEC	0.396 ± 0.166	0.377 ± 0.144	-0.051 ± 0.046

2.4.2. Population Structure & Gene Flow

Relationships between populations was observed using a variety of methods. First, pairwise F_{ST} was calculated between each population. The mean F_{ST} across all pairwise combinations was 0.087 (±0.035) and ranged between 0.008 (RV_HUN_CSA – RV_HUN_FOL) to 0.180 (RV_HUN_TIZ –RV_GER_JUT). The neighbour-net network summarising F_{ST} results

can be seen in Figure 2.2. The network shows a clear East-West trend with Hungary and Romanian populations featuring on an opposing end to the Western Italian population. Both UK populations are placed closed to the Italian population, with RV_UK1 (F_{ST} = 0.009) sitting more tightly than RV_UK2 (F_{ST} = 0.061) to RV_ITA.

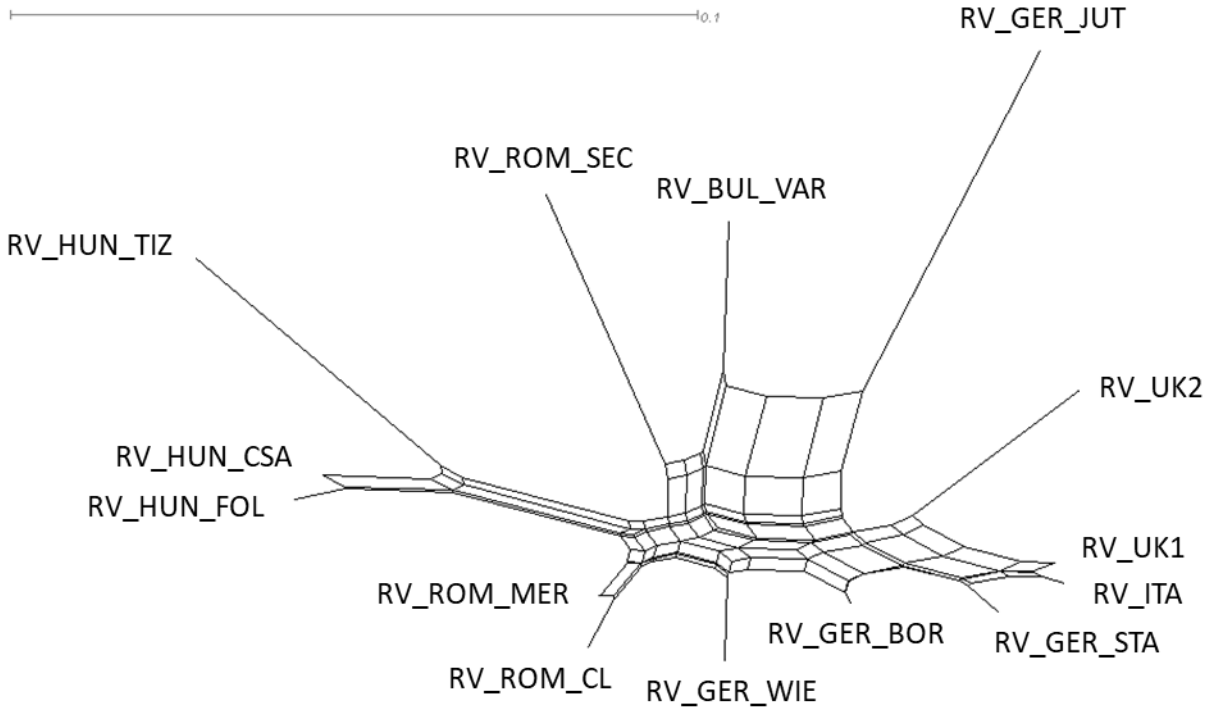


Figure 2.2: F_{ST} network showing relationships between European buffalo population. Network produced using neighbour-net method in SplitsTree4.

Following on from this, major variation across the dataset was extracted and lineated using an MDS. The first three components are plotted in Figure 2.3. Component 1 captured 19.9% of variation across the dataset and featured an East-West trend like F_{ST} results. Components 2 (14.1%) and 3 (10.3%) separated the Eastern populations in Bulgaria, Hungary, and Romania. Again, like the F_{ST} results, RV_UK1 is almost indistinguishable to RV_ITA, with RV_UK2 sitting nearby. Full breakdown of eigenvalues by each component can be found in Table S2.2.

Attempts to define genetic clusters were done using ADMIXTURE (Figure 2.4). At K = 2, an East-West split is also observed with both UK populations associated with RV_ITA. At K=3, the Hungarian populations split from Eastern populations. Both UK populations remain clustered with RV_ITA until K = 6 where RV_UK2 splits to form its own unique cluster. Cross-

validation revealed that the most efficient and preferred model was K = 20 whereby a lot of intra-population variation is observed (e.g., RV_BUL_VAR, RV_ROM_SEC). At K = 20, RV_UK2 still presents a unique genetic cluster while RV_UK1 remains clustered with RV_ITA. AMOVA results confirm that a large amount of variation is present within individuals as 93.57% of variation is observed at this level (Table S2.3).

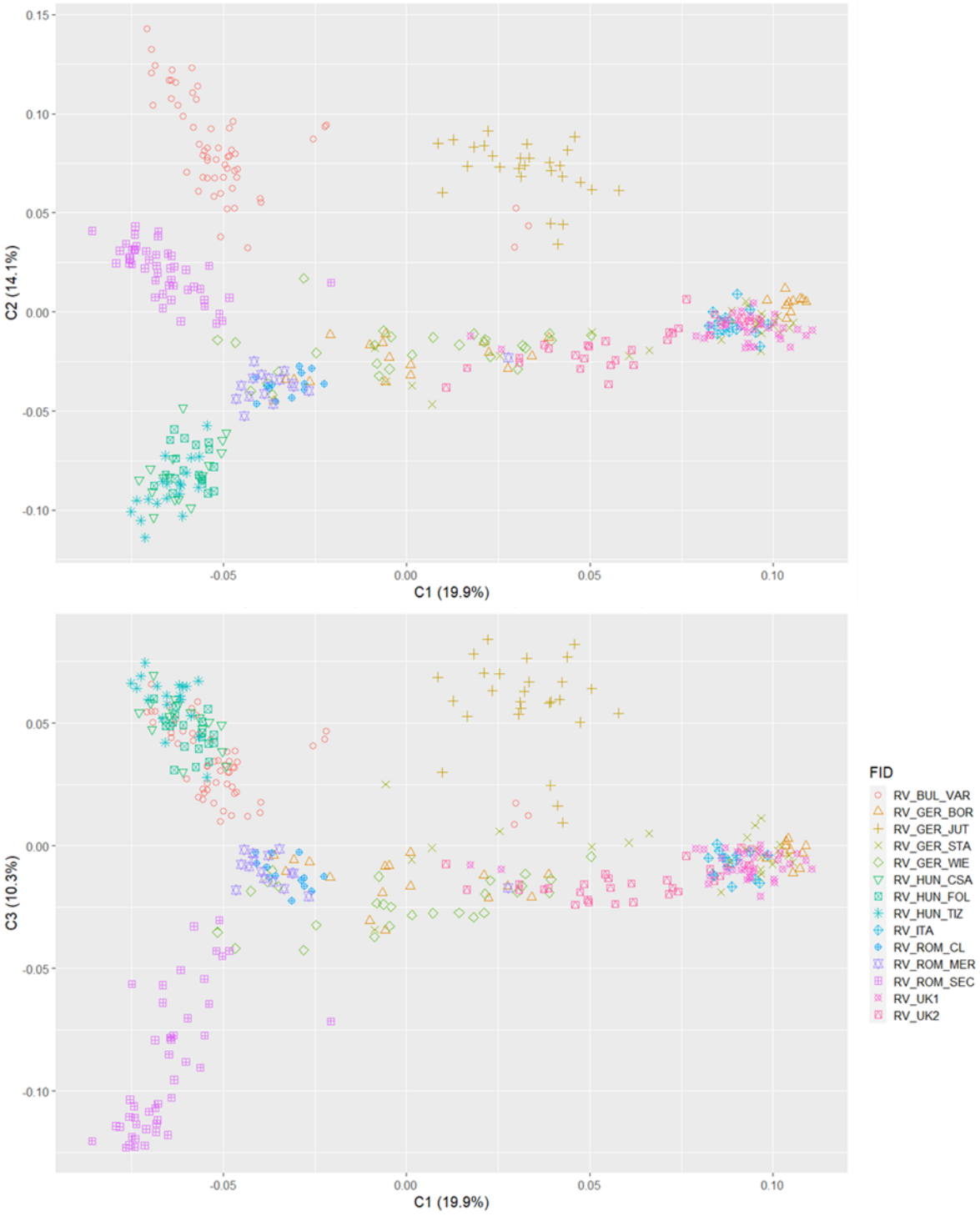


Figure 2.3: Multidimensional scaling plot showing genetic structure across European buffalo populations for axis 1-3. Component 1 captured 19.9% of total variation and separates populations by an East—West split. Component 2 (14.1%) and component 3 (10.3%) separates Eastern European populations. FID = Family ID (i.e. population).

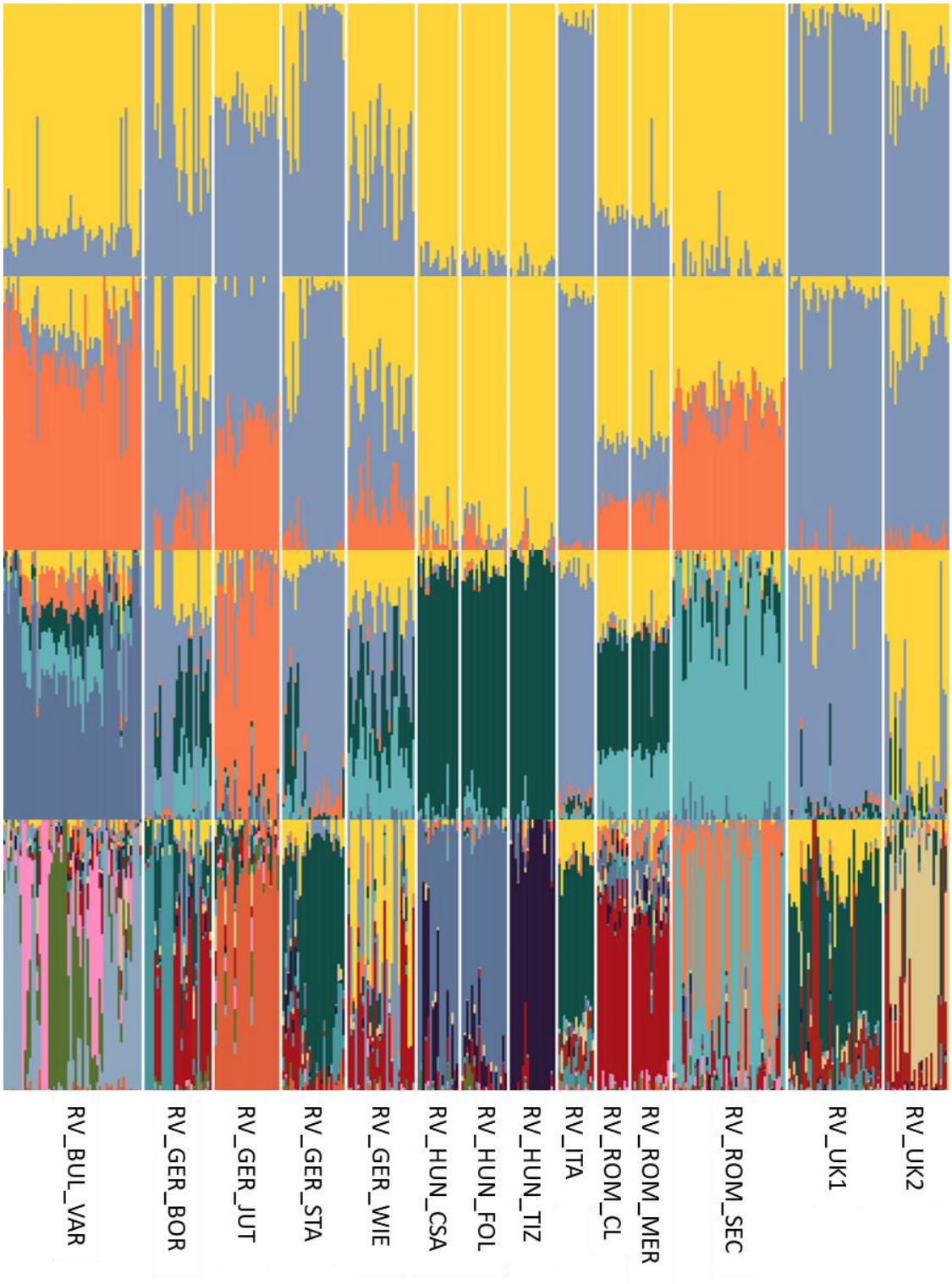


Figure 2.4: Admixture plots showing European water buffalo population structure across different values of K (2, 3, 6, and 20, respectively). K = 2 shows the East-West European split between populations, K = 3 shows splitting of Hungarian populations from Eastern cluster. At K = 6, RV_UK2 splits from Italian cluster, remaining unique in the most appropriate model of K = 20 whereas RV_UK1 remains clustered with Italy throughout.

Finally, the presence of migrations between populations was tested by adding nine migration vectors to the tree involving animal movement between the majority of the European populations. 99.8% of the variation across the dataset could be explained (Figure 2.5). UK populations had connections with Romanian and German populations.

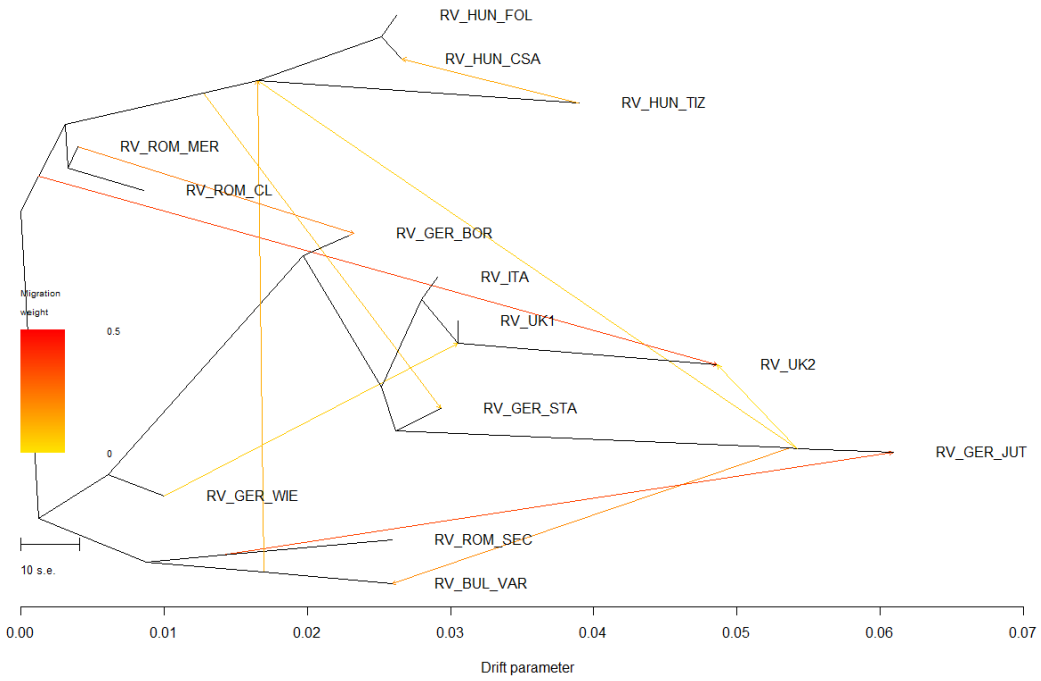


Figure 2.5: Treemix model explaining 99.8% of variation captured across the dataset after the addition of migration edges. Nine migrations were found to be the most appropriate model involving almost all European populations.

2.4.3. Linkage Disequilibrium & Persistence of Phase

Linkage disequilibrium across the genome was summarised for both UK populations and a decay curve plotted from up to 1Mb away from a SNP can be seen in Figure 2.6. Both UK populations begin with approximately an R^2 value of 0.37 for markers separated by 50kb. The LD in both populations then declines with RV_UK1 declining at faster rate with lower R^2 values at all points after 50kb indicating that RV_UK1 has a greater level of genetic diversity than RV_UK2. Average R^2 values appeared to asymptote at 0.079 and 0.119 for SNPs separated by 1Mb for RV_UK1 and RV_UK2 respectively.

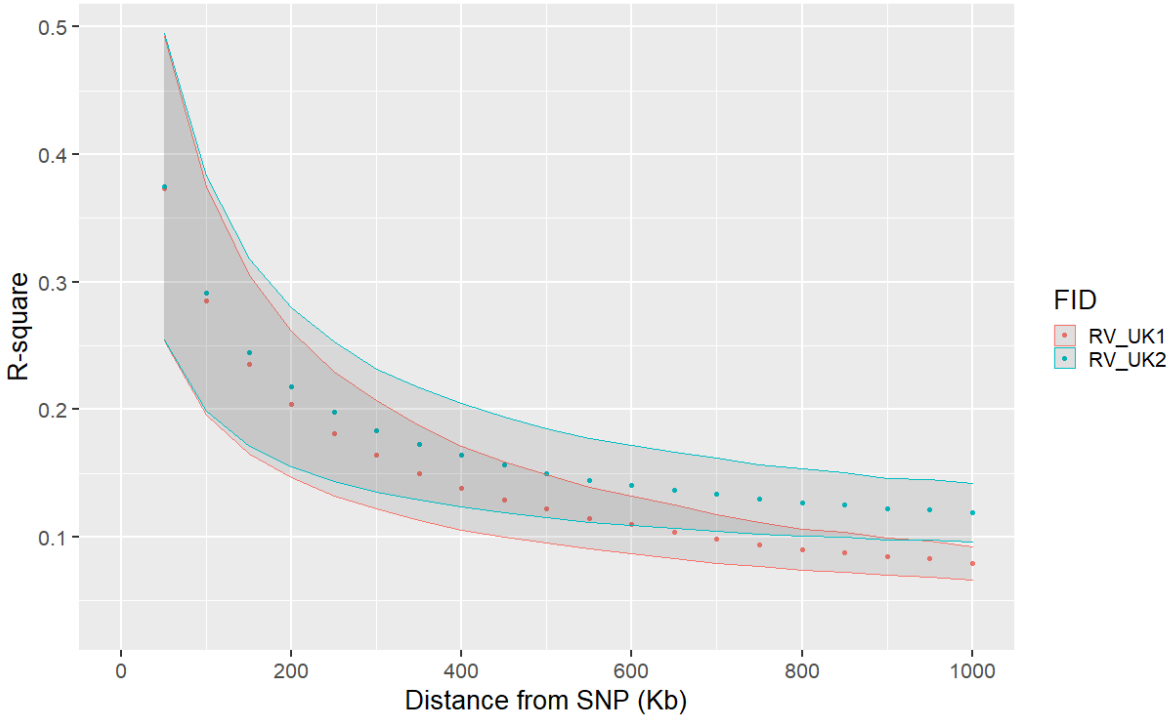


Figure 2.6: Linkage disequilibrium decay curve for UK populations. Both populations show similar levels of average linkage away from a SNP with RV_UK1 showing less linkage at greater distances than RV_UK2. FID = Family ID (i.e. population).

Since the average linkage in both UK populations is similar, persistence of phase was used to determine if the pattern of linkage disequilibrium was consistent across both populations (Figure 2.7). High values of R^2 at all distances would indicate that two populations are near identical and therefore future genomic selection programmes would be highly transferrable between populations, while lower values imply genetic separation of populations due to different demographic histories. Persistence of phase r^2 begins at 0.482 at 50kb away from a SNP and finishes at 0.145 at 1Mb. These values are in line with results from comparisons of different buffalo breeds (Deng *et al.*, 2019; Mokhber *et al.*, 2019). Therefore, RV_UK1 and RV_UK2 are kept as separate populations for future analysis here.

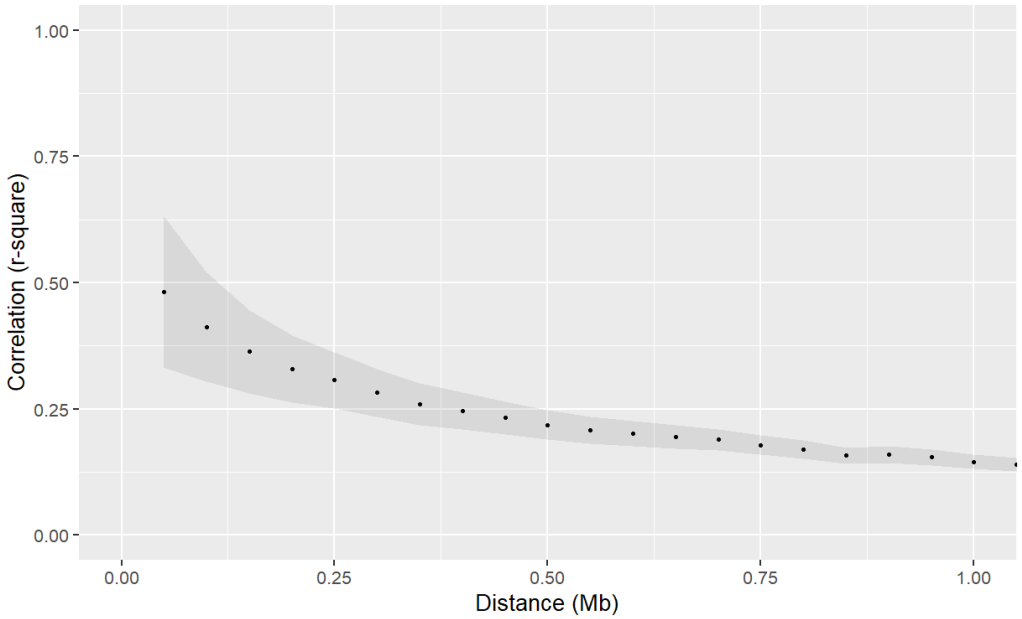


Figure 2.7: Persistence of phase between UK buffalo populations. UK populations do not show high persistence of phase levels and instead the extent of differences of linkage patterns across the genome are similar to that of between Iranian buffalo breeds.

2.4.4. Ancestry

The differences observed so far between RV_UK1 and RV_UK2 suggest that both populations have not had identical recent histories. Therefore, genetic ancestry was assessed using several methods. Mean proportional IBD was calculated for all pairs of UK buffaloes. IBD was significantly higher ($W = 101743$, $p < 0.001$) in RV_UK2 (mean = 0.103 ±0.124) than RV_UK1 (mean = 0.035 ±0.058). In addition to this, IBD was significantly greater ($W = 299902$, $p < 0.001$) within farms than between farms. These results are reflected in the network showing IBD > 0.1 between individuals produced in Figure 2.8. Each population mostly separates from one another, and the relationships between individuals in RV_UK2 are typically greater than between individuals in RV_UK1.

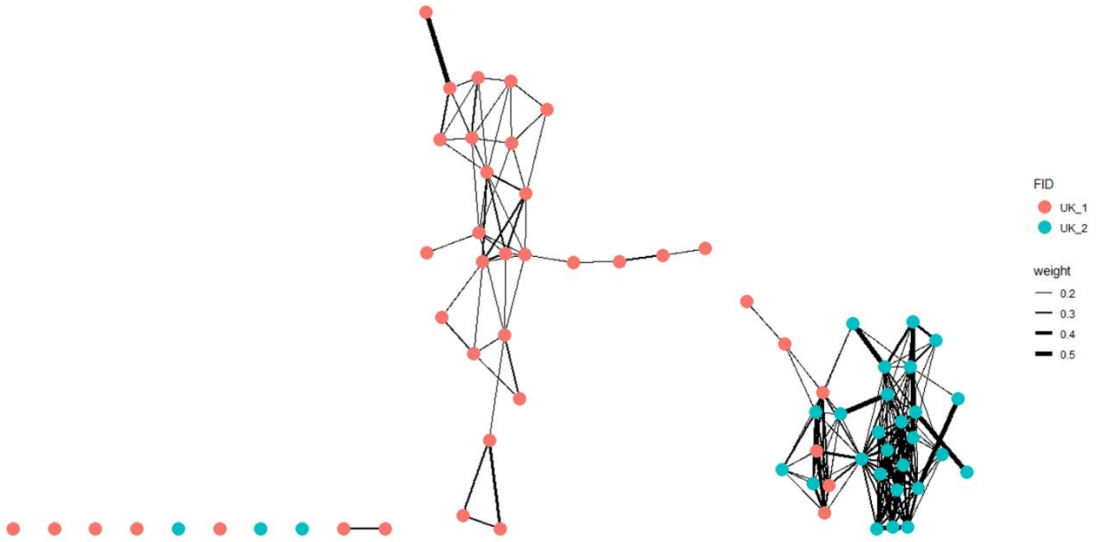


Figure 2.8: IBD network showing relationships (IBD > 10%) between UK buffalo individuals. IBD is significantly greater in RV_UK2 than RV_UK1 and most individuals show relationships within their respective farm. Points with no line attached feature an IBD < 0.1 to all other individuals and appear unrelated. FID = Family ID (i.e. population).

N_e (Figure 2.9) follows the trend of differing recent histories between UK populations as from generation 5 to 250, RV_UK1 features a greater N_e at all time points. RV_UK1 starts at an N_e of 215 compared to an Ne of 55 for RV_UK2. After 250 generations ago, both UK populations share similar N_e, probably as they converge into the same ancestral population. RV_UK2 features an often seen ever-decreasing slope up until the present day whereas RV_UK1's curve plateaus between generations 18 and 100, suggesting some influx of new genetic diversity present in its history.

LAMP-ANC was used to determine the ancestral origins across each UK populations genome. Each UK population was compared to RV_ITA and RV_ROM_CL, representing the initial founding populations of European buffaloes in the East and West. Results of ancestry across each chromosome for each UK population can be found in Figure 2.10. RV_UK1 displayed 80.5% RV_ITA ancestry, and 19.5% RV_ROM_CL, whereas RV_UK2 showed an ancestry of 93.0% RV_ITA, and 7.0% RV_ROM_CL. Despite the differences in ancestry proportions, regions of RV_ROM_CL in RV_UK2 occur in the same places as in RV_UK1 suggesting some form of shared historic ancestry. Interestingly, the ancestry here contrast with prior results that suggest RV_UK1 is more closely related to RV_ITA than RV_UK2.

Therefore, due to the greater levels of IBD seen in RV_UK2, it is speculated that the greater RV_ITA ancestry in RV_UK2 is due to genetic drift and not true ancestry.

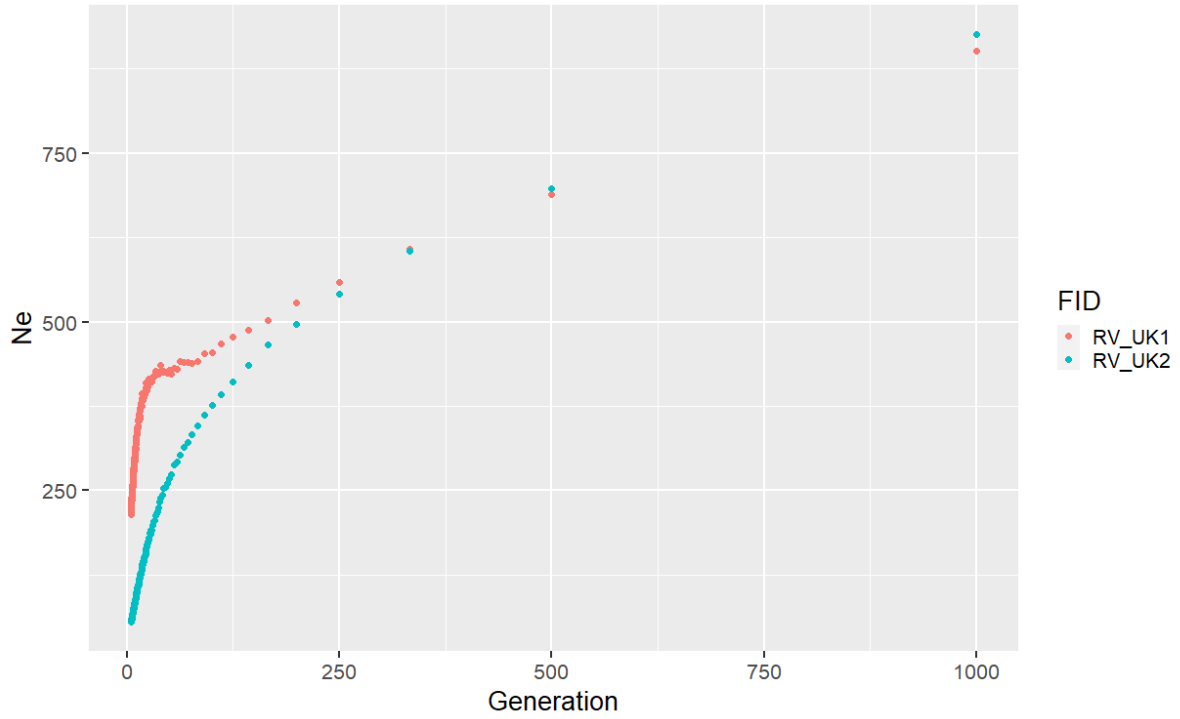


Figure 2.9: N_e trend of UK populations over recent history. N_e was calculated using Sved (1971) equation as used in Mokhber *et al.*, (2019). FID = Family ID (i.e. population).

To back up this claim of genetic drift on RV_UK2, 1,000 bootstrap replicates of F_{ST} (compared to RV_ITA and RV_ROM_CL) and average F_{ROH} metrics were conducted (Figures S2.4 – S2.6). RV_UK2 F_{ST} (mean = 0.050; 0.040 – 0.062 95% CI) compared to RV_ITA was greater than RV_UK1 vs RV_ITA (mean = 0.024; 0.021 – 0.029 95% CI) with no overlap of distributions. Meanwhile there is large overlap in F_{ST} distribution for RV_UK1 (mean = 0.070; 0.062 – 0.077 95% CI) and RV_UK2 (mean = 0.067; 0.059 – 0.077 95% CI) vs RV_ROM_CL. Replicates of F_{ROH} reveal that RV_UK2 F_{ROH} (mean = 0.159 ± 0.018) is significantly higher ($W = 209330$, $p < 0.001$) than RV_UK1 (mean = 0.142 ± 0.010). Therefore, due to i) greater IBD in RV_UK2, ii) smaller N_e in RV_UK2, iii) greater F_{ROH}, and iv) greater F_{ST} vs RV_ITA, it appears the difference in ancestry results in the UK populations is potentially due to genetic drift in RV_UK2.

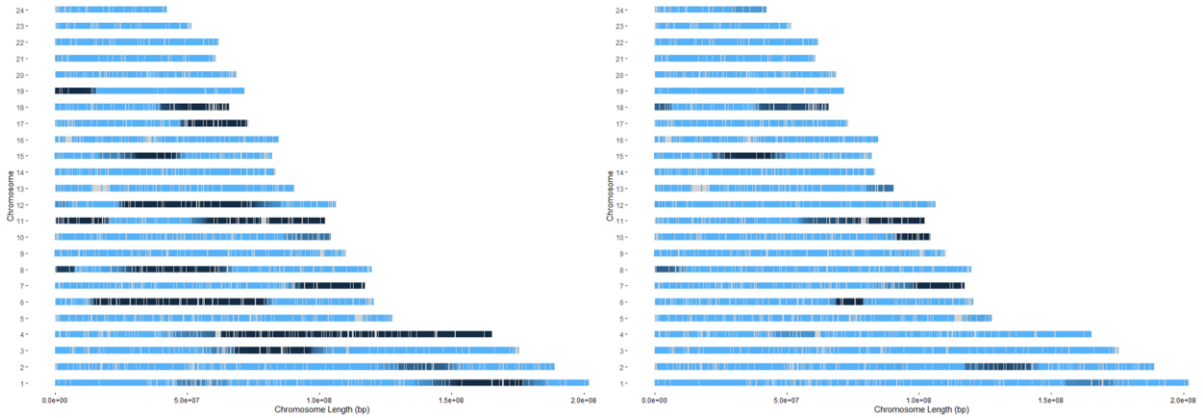


Figure 2.10: Karyotype plots showing ancestral proportions calculated in LAMP-ANC across chromosomes for RV_UK1 (left) and RV_UK2 (right). Dark blue areas indicate Romanian ancestry while light blue indicates Italian ancestry. Grey patches indicate chromosomal regions with missing data, i.e., no SNP markers present.

2.4.5. Runs of Homozygosity

Numerous ROHs were found across the genomes of both UK populations (Table 2.3). RV_UK2 featured a non-significantly higher ($W = 506$, $p = 0.796$) F_{ROH} with a mean of 0.159 (± 0.070) compared to 0.142 (± 0.041) for RV_UK1. F_{ROH} was significantly correlated to inbreeding (F) coefficient (Section 2.4.1) for both RV_UK1 ($r^2 = 0.900$, $t_{39, 37} = 12.548$, $p < 0.001$) and RV_UK2 ($r^2 = 0.981$, $t_{27,25} = 24.984$, $p < 0.001$), showing that both statistics reflected one another despite differing values and SNP subsets. The average length of ROHs were also non-significantly higher in RV_UK2 at 3.58Mbp (± 5.15) compared to RV_UK1 average of 2.77Mbp (± 3.13). RV_UK1 typically featured more shorter runs while RV_UK2 had an increased number of long ROHs. Though a chi-squared test across average number of ROHs per class between UK populations showed no significant over- or underrepresentation ($\chi^2 = 3.432$, $df = 4$, $p = 0.488$). Comparing the sum of ROHs versus the total number of ROHs per individual, it is seen that both UK populations lean towards a smaller effective population size and show a presence of consanguineous mating (Figure S2.7).

Table 2.3: ROH metrics across both UK populations.

Population	F_{ROH}	ROH Mean Length	Average number of ROHs per class (Mb) per individual				
			1-2	**2-4**	**4-8**	**8-16**	**>16**
RV_UK1	0.142 (± 0.041)	2.77 (± 3.13)	62.0 (± 11.3)	50.3 (± 9.6)	11.0 (± 4.4)	2.7 (± 2.7)	1.1 (± 1.9)
RV_UK2	0.159 (± 0.070)	3.58 (± 5.15)	53.1 (± 9.0)	37.0 (± 10.3)	12.1 (± 6.4)	4.3 (± 4.6)	3.7 (± 3.6)

The persistence of reoccurring ROHs across a population indicates consistent regions under selection that are eliminating genetic diversity. The top 0.01 SNPs occurring in ROHs across each UK population were identified as outliers and therefore potential candidates of selection (Figure 2.11). The thresholds (proportion of individuals with ROH in a population) for each population were 43.6% and 51.9% for RV_UK1 and RV_UK2 respectively. 101 genes associated with outlier SNPs were identified in RV_UK1 across 17 regions, whilst 128 genes across 15 regions were found in RV_UK2. Full list of genes inferred as candidates of selection can be found in Table S2.8.

Gene ontology revealed several enriched biological processes in each UK population (Table S2.9). Three enriched biological pathways were found in RV_UK1 which were GO:2000564 Regulation of CD8-positive alpha-beta T-cell proliferation, GO:0046328 Regulation of JNK cascade, and GO:0015824 Proline transport. Meanwhile, RV_UK2 gave 15 enriched pathways surrounding metabolic activity (GO:0005986, GO:0006543, GO:0005985, GO:0009312, GO:0006537, GO:0046351, GO:0042132, GO:0004359), muscular (GO:0016459, GO:0003774, GO:0030898), and signalling activity (GO:0007259, GO:0097696, GO:0014037).

Highly connected genes were identified using STRING to reveal the interactions between gene proteins in each UK population. 54 (53.5%) genes interacted with at least one other gene found in outlier ROHs (between 1-8 interactions) in RV_UK1. Among these, a highly connected cluster on chromosome 6 is observed that largely surrounds immune responses and reproductive function. *AP4B1* (Adaptor Related Protein Complex 4 Subunit Beta 1; BBU6: 29,303,283 – 29,313,751) encodes a subunit that contributes to targeting proteins from the trans-Golgi network to the endosomal-lysosomal system (Ebrahimi-Fakhari

et al., 2018); *PHTF1* (Putative Homeodomain Transcription Factor 1; BBU6: 29,448,577 – 29,530,219) is highly expressed in the testis and operates in the cis-Golgi network and endoplasmic reticulum membrane (Oyhenart *et al.*, 2003); *RSBN1* (Round Spermatid Basic Protein 1; BBU6: 29,403,113 – 29,446,217) is a testis-specific histone demethylase that modulates chromosome architecture (Y. Wang *et al.*, 2021); *MAGI3* (Membrane Associated Guanylate Kinase, WW and PDZ Domain Containing 3; BBU6: 29,525,900 – 29,788,113) protein acts as a scaffolding protein at cell-cell junctions (Kotelevets and Chastre, 2021); *OLFML3* (Olfactomedin Like 3; BBU6: 41,072,042 – 41,293,983) encodes an extracellular matrix glycoprotein facilitating protein-protein interactions, cell adhesions, and intercellular interactions (Tomarev and Nakaya, 2009). *BCL2L15* (BCL2 Like 15; BBU6: 29,5321,726 – 29,328,426) meanwhile participates in regulating cell death by controlling mitochondrial outer membrane permeabilization (Than *et al.*, 2019; Miyai, Hendawy and Sato, 2021). *SYT6* (Synaptotagmin 6; BBU6: 29,054,894 – 29,121,528) acts in synapse regulating neuronal transmission (De León *et al.*, 2021).

93 (72.7%) genes had at least one interaction (between 1-8 interactions) in RV_UK2. Of the most connected genes, three small clusters of genes were found. The first of the three clusters included genes of varying associations. *FBXW4* (F-Box and WD Repeat Domain Containing 4; BBU23: 22,171,598 – 22,254,838) induces protein degradation and is involved in normal limb development (Q. Han *et al.*, 2020); *POLL* (DNA Polymerase Lambda; BBU23: 22,140,608 – 22,149,066) functions in base excision repair and non-homologous end-joining and is involved in DNA damage tolerance (Garcia-Diaz *et al.*, 2005; Nemec *et al.*, 2016); *DPCD* (Deleted in Primary Ciliary Dyskinesia Homolog; BBU23: 22,149,003 – 22,171,083) is involved in the generation and maintenance of ciliated cells (Lee, 2013); *OGA* (O-GlcNAcase; BBU23: 22,330,809 – 22,359,629) modifies cytoplasmic and nuclear proteins associated with homeostasis (Zhang *et al.*, 2014); *BTRC* (Beta-Transducin Repeat Containing E3 Ubiquitin Protein Ligase; BBU23: 21,946,315 – 22,120,070) is an F-box protein involved in the Wnt/Beta-Catenin signalling pathway that is relevant to reproductive pathways (Marete *et al.*, 2018).

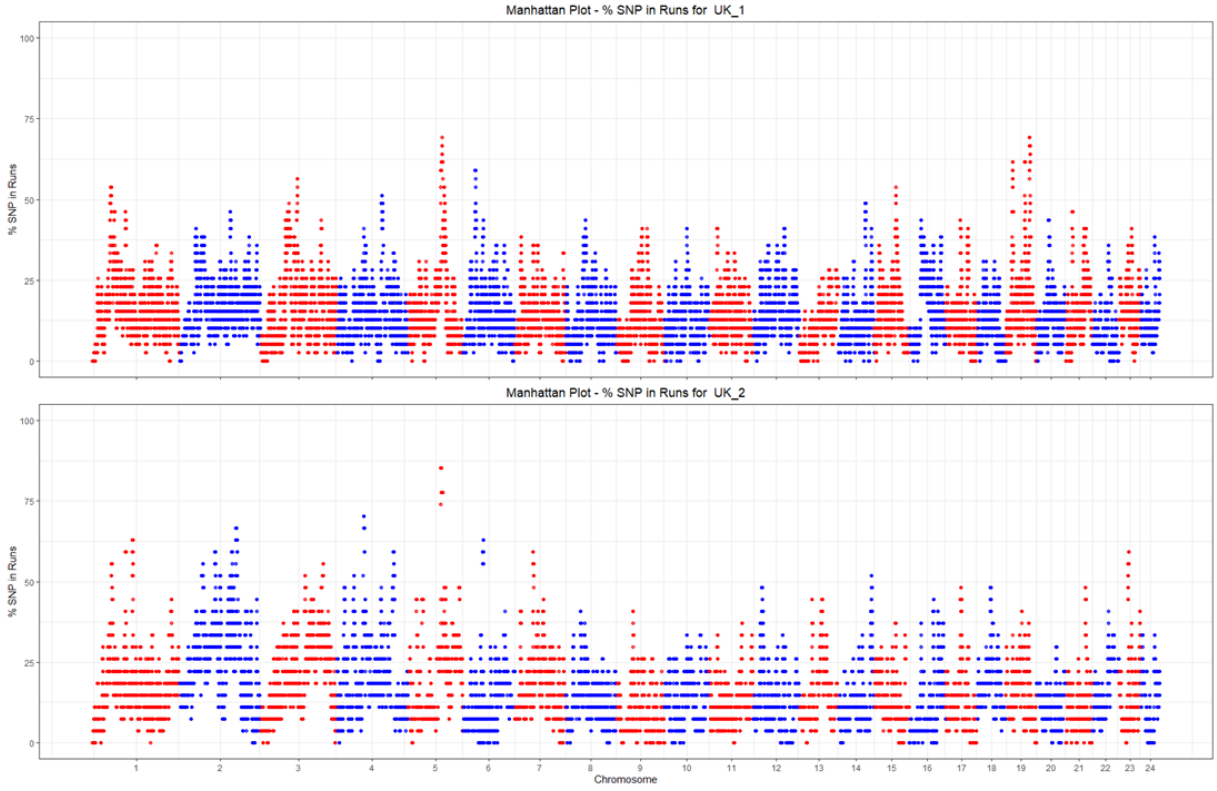


Figure 2.11: Manhattan plots showing the proportion that a SNP is found in a ROH for each UK buffalo population. 0.01 thresholds were 43.6% for RV_UK1 and 51.9% for RV_UK2. 101 genes associated with significant SNPs were identified in RV_UK1 across 17 regions, whilst 128 genes across 15 regions were found in RV_UK2.

Second cluster included *ERBB3* (Erb-B2 Receptor Tyrosine Kinase 3; BBU4: 62,971,471 – 62,992,692) is an essential regulator of cell growth (Nguyen *et al.*, 2018). Also in this cluster is multiple STAT (Signal Transducer and Activator of Transcription) genes that are a family of transcription activators that are all involved in immune responses through the Jak-STAT pathway (Villarino, Kanno and O'Shea, 2017). STAT genes found in RV_UK2 ROHs are *STAT1* (BBU2: 132,434,013 – 132,474,613), *STAT2* (BBU4: 63,197,550 – 63,212,979), *STAT4* (BBU2: 132,486,904 – 132,623,372), *STAT6* (BBU4: 63,946,070 – 63,960,143).

The remaining cluster is linked to the second cluster via *PA2G4* (Proliferation-Associated 2G4; BBU4: 62,993,661 – 63,001,142). This gene encodes an RNA-binding protein that is involved in growth regulation and has been shown to interact with *ERBB3* (Stevenson *et al.*, 2020); *SUOX* (Sulfite Oxidase; BBU4: 62,889,065 – 62,893,453) protein is localised to the intermembrane space of mitochondria and catalyzes the oxidation of sulfite to sulfate (Du *et al.*, 2021); *MARS1* (Methionyl-tRNA Synthetase 1; BBU4: 64,301,863 – 64,329,569) encodes

a component of multi-tRNA synthetase complex that catalyses the ligation of methionine to tRNA molecules (Syed *et al.*, 2023), *RPS26* (Ribosomal Protein S26; BBU4: 62,925,726 – 62,928,063) is a ribosomal protein component with roles in ribosome assembly and protein translation (Chen *et al.*, 2021); and two unknown genes of XP_006054002.1, and XP_006069518.1.

Numerous cattle QTLs overlapped with the candidate regions identified here. 249 QTLs were identified across RV_UK1 while 414 were identified in RV_UK2. The distribution of QTL classes across significant ROHs for each population can be found in (Figures S2.10 & S2.11). Fisher's exact test was used to determine over-, and underrepresentation of QTLs found in each population (Table 2.4). Four of the six QTL classes were found to be significantly over- or underrepresented in RV_UK1. Milk and reproduction traits were underrepresented with odds ratios (OR) of 0.538 and 0.522 respectively, while exterior (OR = 1.875) and production (OR = 2.265) traits were overrepresented. In RV_UK2, milk (OR = 1.231) traits were overrepresented while production (OR = 0.513) QTLs were underrepresented.

Table 2.4: Fisher's Exact test results of QTL classes found in ROHs across UK populations. Significant results, after Bonferroni's correction ($p < 0.083$) are in bold. QTLs were obtained from *Bos taurus* UMD3.1.1 reference genome (RefSeq: GCF_000003055.6) that have been mapped to water buffalo reference genome UOA_WB_1 (RefSeq: GCF_003121395.1) by Nascimento *et al* (2021). Definitions and traits for each QTL class can be found at Cattle QTL Database (animalgenome.org).

QTL Class	Population	Total Number of QTLs	QTLs in ROHs	Odds Ratio	P-Value
Exterior	**RV_UK1**	5329	49	**1.875**	**0.000**
	RV_UK2		51	1.071	0.644
Health	RV_UK1	1509	12	1.495	0.207
	RV_UK2		10	0.727	0.404
Meat & Carcass	RV_UK1	1376	9	1.216	0.572
	RV_UK2		16	1.305	0.307
Milk	**RV_UK1**	20830	77	**0.538**	**0.000**
	RV_UK2		209	**1.231**	**0.037**
Production	**RV_UK1**	6796	70	**2.265**	**0.000**
	RV_UK2		34	**0.513**	**0.000**
Reproduction	**RV_UK1**	10097	32	**0.522**	**0.000**
	RV_UK2		94	1.043	0.721

2.4.6. Signatures of Positive Selection

Recent adaptation to the UK by buffaloes was inferred through the detection of positive selection. Each UK population was compared to the closely related RV_GER_STA population and significant regions under positive selection in both UK populations was considered as potential adaptations to UK. Six significant regions under positive selection were present in both UK populations amounting to 24 genes (Table 2.5). 43 significant regions comprising of 213 genes were found for RV_UK1 in total while 47 regions with 255 genes were found in RV_UK2 (Table S2.12).

Multiple genes were found on the region under selection on BBU1 (187578102 – 187861398). Of these, two genes (*SLC37A1* and *PDE9A*) have been associated with milk traits (Yang *et al.*, 2015; Raven *et al.*, 2016; Sanchez *et al.*, 2021). *SLC37A1* (Solute Carrier Family 37 Member 1) is a glucose-6-phosphate antiporter while *PDE9A* (Phosphodiesterase 9A) hydrolyses the secondary messenger cGMP (Kotera *et al.*, 2010; Pan *et al.*, 2011). The remaining genes appear to be associated with cellular processes and stress response. *WDR4* forms a complex with *METTL1* to exert its effects on a variety of cellular processes (Ruiz-Arroyo *et al.*, 2023). *NDUFV3* encodes an accessory subunit of the mitochondrial respiratory chain NADH dehydrogenase (Complex 1) which is involved in ATP production (Guerrero-Castillo *et al.*, 2017). *PKNOX1* (PBX/Knotted 1 Homeobox 1) is a transcription activator and is a regulator of oxidative phosphorylation components (Kanzleiter *et al.*, 2014). *CBS* (Cystathionine Beta-Synthase) is active in production of taurine which has cellular roles of electrolyte balance, immune response, and antioxidant function (Zhou *et al.*, 2017; Darang *et al.*, 2022). *U2AF1* (U2 Small Nuclear RNA Auxiliary Factor 1) is involved in mediation of RNA splicing (Dutta *et al.*, 2021). *CRYAA* (Crystallin Alpha A) is a members of heat shock protein (HSP20) family and is a major structural protein in eye lenses (Chang *et al.*, 2022).

TUT7 (Terminal Uridylyl Transferase 7) and *GAS1* (Growth Arrest Specific 1) were found under selection on BBU3 (142275429 – 142532051). TUT7 which is expressed in macrophages impacts innate immune responses and shows antiviral functions (Kozlowski *et al.*, 2017). *GAS1* is involved in growth suppression by blocking entry of the cell cycle into S phase (Martinelli and Fan, 2007). *CYLC2* (Cylicin 2), *TRABD2B* (TraB Domain Containing 2B), and *CSNK1G3* (Casein Kinase 1 Gamma 3) were under selection on BBU3, BBU6, and BBU9

respectively. *CYLC2* is specifically expressed in testis and is part of the cytoskeletal calyx of mammalian sperm heads (Hou *et al.*, 2019). *TRABD2B* enables Wnt-protein binding activity and metalloendopeptidase activity and therefore is involved in various regulatory pathways (Grainger and Willert, 2018). *CSNK1G3* phosphorylates and conducts post translational modifications of caseins in cattle (Buitenhuis *et al.*, 2016).

Table 2.5: Regions of genome under selection in both UK populations when compared to RV_GER_STA, and associated genes. Start and end positions relate to positions on each chromosome where SNPs within (identified using XP-EHH) are under positive selection. Genes listed are those that are found within or overlapping the start and end positions. SNPs were mapped to water buffalo reference genome UOA_WB_1 (RefSeq: GCF_003121395.1).

Chromosome (BBU)	Chromosome Start Position	Chromosome End Position	Genes
1	187578102	187861398	SLC37A1, PDE9A, WDR4, NDUFV3, PKNOX1, CBS, U2AF1, CRYAA
3	142275429	142532051	TUT7, GAS1
	154753980	155514138	CYLC2
6	97852006	-	TRABD2B
9	81392762	81472090	CSNK1G3
11	90784888	90819545	CLIP, PDCD7, UBAP1L, KBTBD13, RASL12, SLC51B, MTFMT, SPG21, ANKDD1A, PTGER2, TXNDC16

BBU11 (90784888 – 90819545) provided a region under selection featuring multiple genes akin to BBU1. Genes found in this region were often associated with immunity, inflammation among other body functions. *CLIP* (Cartilage Intermediate Layer Protein) is an abundant protein in cartilage and is a mediator of extracellular matrix remodelling (van Nieuwenhoven *et al.*, 2017). *PDCD7* (Programmed Cell Death 7) promotes apoptosis when overexpressed (Ghafouri-Fard *et al.*, 2022). *UBAP1L* (Ubiquitin Associated Protein 1 Like) enables ubiquitin binding activity and mutations are linked to neurodegeneration mediated by apoptosis (Lin *et al.*, 2019). *KBTBD13* (Kelch Repeat and BTB Domain Containing 13) has cytoskeletal regulation with implications in skeletal muscle morphology and cardiomyopathy (Gao *et al.*, 2020; de Winter *et al.*, no date). *RASL12* (RAS Like Family 12) is a cytoplasmic GTPase recruited to microtubules with varying cellular effects (Dhanaraman *et al.*, 2020). *SLC51B* (Solute Carrier Family 51 Subunit Beta) is essential for intestinal bile acid absorption and dietary lipid absorption (Ballatori *et al.*, 2013). *MTFMT* (Mitochondrial Methionyl-tRNA

Formyltransferase) formylates methionyl-tRNA in for mitochondrial protein synthesis (La Piana *et al.*, 2017). *SPG21* (SPG21 Abhydrolase Domain Containing Maspardin) is involved in repression of T cell activation (Zeitlmann *et al.*, 2001). *ANKDD1A* (Ankyrin Repeat and Death Domain Containing 1A) plays a role in tumour suppressor in glioblastoma multiforme and is tightly associated with T cells (Zhao *et al.*, 2021). *PTGER2* (Prostaglandin E Receptor 2) is expressed in the endometrium and is involved in tissue repair (Gao *et al.*, 2017). *TXNDC16* (Thioredoxin Domain Containing 16) is an endoplasmic reticulum luminal glycoprotein (Harz *et al.*, 2014).

2.5. Discussion

The access of genomic data is greatly assisting the development of all major livestock species through targeted artificial selection (Rothschild and Plastow, 2014; Georges, Charlier and Hayes, 2019). Husbandry programmes are able to precisely target genomic regions around desired traits while monitoring the genetic health of populations (Hayes *et al.*, 2009; Jonas and Koning, 2015). Many commercial livestock breeds have already undergone rapid genomic improvement in pursuit of greater livestock yields and as a result, they have lost much of their genetic diversity (Meuwissen, 2009; Eusebi, Martinez and Cortes, 2020). However, those breeds and species that have yet to be subjected to intense selection may have wealth of valuable genetic diversity (Olschewsky and Hinrichs, 2021). Correct utilisation of this genetic diversity is key for retaining the adaptive potential of livestock whilst improving productivity and resilience (Meuwissen, 2009; Boettcher *et al.*, 2015; Eusebi, Martinez and Cortes, 2020). The study here evaluates the current standing genetic variation in UK water buffaloes to aid initiation of incorporating genetic data into buffalo development.

2.5.1. Genetic Diversity, Population Structure & Gene Flow

Water buffaloes occur in low numbers in the UK compared to other livestock species that may be problematic if their genetic effective population size also reflects this. Small populations are susceptible to genetic bottlenecks, genetic drift and inbreeding depression leading to an increased chance of fixation of harmful deleterious alleles (Kyriazis, Wayne and Lohmueller, 2021). Positively, there was a high level of H_O at 0.392 and 0.385 in RV_UK1 and RV_UK2 respectively. Both UK farms largely reflected their continental counterparts in their level of genetic diversity with only a marginal decline in H_O. H_O was greater than H_E and as a

result there was no evidence to suggest any presence of inbreeding among the farms. Due to their lack of inbreeding and a greater observed heterozygosity than expected, this suggests an isolate breaking effect caused by recent influx of new genetic diversity, likely through the importation of buffaloes from different sub-populations (Wahlund, 1928; Šnjegota *et al.*, 2021; Kanaka *et al.*, 2023).

The extent of allelic differentiation between UK and European buffaloes was examined through a series of population-based test. Despite the high and similar levels of genetic diversity across buffaloes, there was a moderate amount of differentiation across the dataset. F_{ST} results gave a mean difference of 0.087 (Range of 0.008 to 0.180). This level of F_{ST} is within the range of breed differences in other European livestock species which range between 0.026 – 0.068 in Spanish cattle (Cañas-Álvarez *et al.*, 2015), 0.014 – 0.168 in Northern European cattle (Schmidtmann *et al.*, 2021), 0.013 – 0.164 in Italian goats (Nicoloso *et al.*, 2015), 0.020 – 0.201 in Welsh sheep (Beynon *et al.*, 2015), and 0.085 – 0.161 and 0.088 – 0.202 in European pigs (Munoz *et al.,* 2019; Bovo *et al.*, 2020). Therefore, it is likely that buffalo populations across Europe have been separated for some time and divergence between these sub-populations is possibly at or nearing breed level. An East-West structuring is observed within European water buffaloes with Hungarian and Italian populations taking up polar ends. The remaining populations take up positions between these. Noce *et al.*, (2021) notes that the Hungarian populations present a uniquely genetic group that may have had less contact with outside populations. Therefore, isolation of these populations may have allowed greater genetic drift leading to an increased F_{ST}. Nevertheless, the result of the network appears to show a gradient of genetic difference across Europe.

The split in buffalo populations in Europe likely represents their origins. Water buffaloes reached Europe via Saracen trade routes 7th – 9th centuries established between the Middle East and Europe (Wordsworth *et al.*, 2021). Due to water buffaloes' adaptations to wetlands such as wallowing and surviving of low-quality vegetation, wetland areas in Italy and Romania facilitated buffalo rearing in Europe in areas where cattle could not persist (De la Cruz-Cruz *et al.*, 2014; Yáñez *et al.*, 2020; Wordsworth *et al.*, 2021). River buffaloes have not been used for long distance trade like cattle or horses and as such, Eastern and Western buffaloes probably did not share much historical interaction after their separation. Both UK water buffalo populations sit tightly to the Italian population with low F_{ST} scores (0.009 &

0.061) showing that both populations descend from Italian stocks. This result was similarly shown in the MDS analysis as Eastern and Western populations were separated with German populations falling between. Again, UK buffaloes sit tightly with Italian individuals, although in both F_{ST} and MDS, RV_UK2 shows greater deviance.

Following on from this, clustering of genetically similar populations and the presence of gene flow was tested for. Admixture clustering analysis again displayed the East-West split at K = 2. RV_UK1 remains clustered with RV_ITA throughout the analysis while unexpectedly RV_UK2 separates to form its own cluster that is retained to the final model (K = 20). The final admixture model featured the identification of clusters within populations and an abundance of variation across the dataset. AMOVA results show that this is due to the overwhelming majority of variation being present within individuals which is expected due to the use of commonly polymorphic SNPs in arrays skewing the allele frequency distribution towards variants with average higher minimum allele frequencies (Lachance and Tishkoff, 2013; Malomane *et al.*, 2018; Quinto-Cortés *et al.*, 2018).

The abundance of variation across individuals can also be explained by migration between populations preventing a clear distinguished structuring. An unrooted tree for Treemix was used here as it is known that all European buffalo originally descend from India, therefore no outgroup is needed (Colli, Milanesi, Vajana, *et al.*, 2018). Whilst the unrooted tree doesn't determine evolutionary order of populations, it does still reveal information surrounding the relationships between the populations. Treemix results showed numerous migration connections (K = 9) between populations with the majority involving populations featuring excess observed heterozygosity. UK populations again sit close to Italy in the topology of the outputted tree. Migration edges between UK and European populations occur with German and Romanian populations fitting with prior importation of Romanian buffalo (Borghese, 2013b). This abundance of variation across European buffaloes and numerous migrations could potentially be explained by the lack of a breed system in European buffaloes. Other European domestic species feature a large number of breeds with little crossbreeding enabling the retention of unique characteristics (Community Based Management of Animal Genetic Resources, 2001). Meanwhile, Mediterranean water buffaloes are largely the only breed available in Europe, except for the Bulgarian murrah which is a crossbreed between

Mediterranean and murrah breeds. Therefore, farmers may obtain new individuals from various populations as phenotypically, buffaloes across Europe appear largely the same.

2.5.2. Linkage Disequilibrium & Ancestry

The patterns of non-random association between alleles at different loci across the genome can be used to deduce the evolutionary history of populations via linkage disequilibrium (Slatkin, 2008). Both UK populations showed similar levels of linkage with RV_UK1 descending to a slightly lower level at 1Mb than RV_UK2 indicating a marginally greater amount of genetic diversity in line with heterozygosity results. Average linkage in UK buffaloes was similar to other buffalo populations, again indicating that there has been no dramatic change in genetic diversity (Deng *et al.*, 2019; Mokhber *et al.*, 2019). Despite the similarity in linkage decay between both populations, persistence of phase indicated substantial differences in the linkage pattern across the genome of both populations. Results showed a similar level of difference as comparisons between Iranian breeds indicating that genomic selection schemes upon one UK farm may not transferrable to others (Mokhber *et al.*, 2019). Starting at $r^2 < 0.5$ at 50kbp between UK buffalo (within breed) is lower than between breed estimates for other livestock species. For example, correlations of r (at 50kbp) between Holstein-Friesian, Jersey, and Angus cattle has been found $r^2 > 0.6$ (de Roos *et al.*, 2008), $r^2 > 0.8$ between Hereford and Bradford cattle (Biegelmeyer *et al.*, 2016), $r^2 > 0.6$ between Spanish breeds (Cañas-Álvarez *et al.*, 2016), $r^2 > 0.7$ between Yorkshire, Landrace, and Duroc pig breeds (Grossi *et al.*, 2017), and $r^2 > 0.5$ between Sunite, German mutton merino, and dorper sheep (Zhao *et al.*, 2014).

The recent N_e in UK buffaloes was estimated from their linkage disequilibrium. N_e between UK buffaloes diverge from 250 generations ago to present, suggesting shared ancestry before this point. Reproductive age remains highly variable in buffaloes due to issues such as delayed maturity, therefore the majority of calving occurs around 5/6 years (Ingawale & Dhoble 2004; Ibrahim 2012; Borghese, Chiariotti and Barile, 2022). Using a generation time of six years, this suggests a divergence starting ~1,250 years ago (~750 CE) (Mintoo *et al.*, 2019). This falls in line with expansion of buffalo across Middle East and Europe via the Saracens (Wordsworth *et al.*, 2021). From 250 generations ago to the present, RV_UK2 presents a continually diminishing trend suggesting isolation of the population which over

time will lead to greater vulnerability of genetic drift and inbreeding depression (Pekkala *et al.*, 2012; Angst *et al.*, 2022). At the most recent estimate of three generations ago, RV_UK2 has an N_e of 55. An N_e below 50 is typically considered as endangered and can lead to declining fitness within the population (Taberlet *et al.*, 2008; Kristensen *et al.*, 2015). However, RV_UK2 is a small farm and the ratio between N_e and census size (N_c) greatly favours N_e suggesting high variation within the population in line with genetic diversity results. Any long-term isolation of this population can be easily managed with addition of unrelated individuals from other farms as the breed-wide N_e of Mediterranean buffalo is likely much greater (Angst *et al.*, 2022). Meanwhile, RV_UK1 N_e did not seem to decline until about 40 generations ago suggesting a maintenance of genetic diversity or potentially some influx near the point for the N_e estimate showing a bump in the trend. It is known that RV_UK1 is an expanding farm and that is regularly obtaining new individuals from Italy. The high genetic diversity and increased N_e suggests that RV_UK1 may be a suitable population where selection intensity can increase without threatening putting at risk the populations genetic diversity and adaptive potential (Hayes *et al.*, 2008; Rexroad and Vallejo, 2009). Following implementation of selection programme and stability of breeding herd, it would be expected that N_e declines in the near future as fewer lineages with favourable traits are selected for (Biegelmeyer *et al.*, 2016; Makanjuola *et al.*, 2020). The differences in N_e were reflected in IBD as RV_UK2 featured significantly higher IBD than RV_UK1. Like population structure results, LAMP-ANC showed strong Italian ancestry within UK buffaloes, however differences in ancestry proportions, unique clustering of RV_UK2, and its smaller N_e suggest genetic drift is present within this population.

2.5.3. Runs of Homozygosity and Candidate Genes of Selection

RV_UK2 featured a greater F_{ROH} and longer ROHs on average compared to RV_UK1 following expectations from linkage disequilibrium results, however these statistics were non-significant. F_{ROH} values of 0.142 and 0.159 for UK buffaloes were similar to F_{ROH} values across European buffaloes fitting that UK buffaloes have not lost genetic diversity (Macciotta *et al.*, 2021; Noce *et al.*, 2021). In comparing 'sum of ROHs' against 'total ROH per individual', both populations sat to the upper right of x = y line indicating some presence of consanguineous breeding and leaning towards a smaller effective population size. This occurs in livestock

species as breeding of individuals is biased towards the small number individuals with the greatest preferred traits (Kirin *et al.*, 2010; Purfield *et al.*, 2012; Ferenčaković *et al.*, 2013; Peripolli *et al.*, 2017).

Generation by generation, genetic diversity is generated through new mutations and distributed in a population via recombination during meiosis as individuals reproduce (Wright, 2005; Webster and Hurst, 2012; Arbel-Eden and Simchen, 2019). The expectation from this observation would be that genetic variation is randomly distributed across the genome. ROHs go against this as positive selection preserves long stretches of the genome without any genetic variation (Quilez *et al.*, 2011; Bosse *et al.*, 2012; Keller *et al.*, 2012; Pemberton *et al.*, 2012). The presence of selection on these regions implies that the underlying functions are important for that population or species in question (Peripolli *et al.*, 2017; Rebelato and Caetano, 2018). Candidates of selection within significant ROHs were identified in both UK populations, giving 17 and 15 ROH regions in RV_UK1 and RV_UK2 featuring 101 and 128 genes. Functional analysis of these genes showed that there was greater connectivity among genes in RV_UK2 than RV_UK1 in both protein-protein interactions, and gene ontology. From N_e analysis, RV_UK2 appears to have a more isolated population and therefore a possibly more stable genomic landscape. The absence of new genetic diversity will lead to the retention of ROHs within the population, which is beneficial when ROHs surround targets of selection.

With new genetic diversity in the population, ROHs in regions without shared ancestry will be broken down due to the presence of different alleles. Admixed populations typically feature low levels of ROHs through influx of genetic variation from the new population in regions that were previously homozygous, which may disrupt targets of selection (Ceballos *et al.*, 2018; Szpiech *et al.*, 2019; Yoshida *et al.*, 2020). Although not an admixed population in terms of crossbreeding, RV_UK1 has obtained individuals from different farms which to some extent may be from different sub-populations. As such, different sources of genetic diversity may disrupt interacting ROHs. Over time, more targets of selection may be revealed as unneeded genetic variation is filtered out. QTL results surprisingly show that milk QTLs are underrepresented in RV_UK1 whilst production and exterior based traits are overrepresented. Mediterranean buffaloes are farmed for the highly nutritious milk that is rich in fat and protein for mozzarella and as such, the quality of the milk is already ideal. Water

buffaloes' downfall is that their milk yields are lower than that of commercial cattle. Thus, increased selection on production-based traits may increase growth rates of buffaloes and yields in order to become economically viable in comparison to commercial cattle farms.

In addition to identifying candidate genes in ROHs, genes under recent positive selection were detected. Populations introduced to new environmental conditions can be subjected to new selective pressures and undergo rapid adaptation (Qanbari and Simianer, 2014; Pitt, Bruford, *et al.*, 2019). This often occurs through positive selection whereby the favourable allele rapidly nears fixation across the population and in doing so, causes near fixation of flanking genomic regions, leaving a detectable signature through long haplotypes (Sabeti *et al.*, 2006; Qanbari and Simianer, 2014). XP-EHH was used to detect recent positive selection in UK buffaloes by testing against a Mediterranean buffalo population from outside of the UK. The logic here being that the variation across the genome will be consistent except for those regions under selection where allele frequencies will near fixation in one (Sabeti *et al.*, 2007). Signatures of positive selection found in both UK populations were inferred as potential candidates of UK adaptation. Six regions totalling 24 genes (Table 2.5) were found under selection in both populations. Functions behind these genes often fell into a number of reoccurring categories of which included milk production, immune pathways and antioxidant functions among others. These functions are biologically relevant to buffaloes due to their dairy and immunity traits.

Several candidate genes under selection were associated with milk traits, such as milk production or associated with mammary tissues. As primarily a dairy livestock species, this is to be expected as farmers select for the most productive individuals which in turn leads to an underlying selective pressure upon genes involved in milk production although these candidate genes may indicate recent changes in milk production as you would expect historic selection of milk associated genes would be close to fixation and therefore, ability to detect selection using long range methods have reduced power (Sabeti *et al.*, 2007; Hayes *et al.*, 2008). In the ROH gene cluster in RV_UK1, *BCL2L15* has been found to increase in expression in mammary tissue during lactation in sheep compared to pregnancy suggesting a role in maintaining milk secretory cells and milk production, whilst in goats, depletion of *BCL2L15* reduced endometrial receptivity through activation of *STAT1* and *STAT3* pathways (Paten *et al.*, 2015; Yang *et al.*, 2020).

The *BTRC* candidate of selection in RV_UK2 is associated with milk production across a number of cattle breeds. Across Holstein and Jersey dairy cattle, this gene was associated with milk production traits and within a highly significant fat kg QTL (Raven *et al.*, 2016). *BTRC* functions within the Wnt/-catenin signalling pathway and indirectly activates nuclear factor kappa-B (NF-kB) which is relevant in mammary development and pregnancy (Raven *et al.*, 2016). This gene was also found to be significantly associated with udder morphology in Holstein, Montbeliarde, and Normande dairy cattle, and present in ROHs across Holstein-Friesian (Red-and-White variety), White-Backed, Polish Red-and-White and Polish black and white and attributed to mammary gland development (Marete *et al.*, 2018; Szmatoła *et al.*, 2019). *BTRC* has also been found associated with body size and leg conformation in Brown Swiss cattle and is known to be behind split-hand and foot malformations (van den Berg *et al.*, 2014; Fang and Pausch, 2019). In *ERBB3* we see a gene involved in regulation of cell proliferation and differentiation, therefore has a wide range of biological effects. *ERBB3* has been found to be a crucial in proliferation of the alveolar mammary epithelial cells in clinical mice trials and its function through the PI3K/Akt pathway supports and regulates milk production (Ollier *et al.*, 2007; Li *et al.*, 2016; M. M. Williams *et al.*, 2017). *PA2G4* similarly regulates cell proliferation and differentiation and interacts with *ERBB3* to enact its function (Ollier *et al.*, 2007; Stevenson *et al.*, 2020; Alan Harris *et al.*, 2023). *RPS26* has been associated with fat yield in Brazilian buffaloes milk and decreased expression during lactation in sheep (de Camargo *et al.*, 2015; Wang *et al.*, 2020).

Further milk associated genes were found in the XP-EHH results shared in both UK populations. *SLC37A1* features increased expression in the bovine mammary glands and associated with milk yield, composition, mineral content, somatic cell score (Raven *et al.*, 2016; M. M. Williams *et al.*, 2017; Iung *et al.*, 2019; Sanchez *et al.*, 2021). This gene along with *PDE9A* in the same selection signature has been shown to be under positive selection in murrah buffalo (Tyagi *et al.*, 2021). This region may be retained in Mediterranean buffalo since domestication in India or these genes are important for greater milk production as murrah, and Mediterranean buffalo are the two most productive river breeds. *GAS1* (Growth Arrest Specific 1), is involved with growth suppression by blocking entry of the cell cycle into S phase. *GAS1* has been shown to be linked to mammary gland development, alpha casein synthesis, and is a candidate gene for mastitis (Naderi *et al.*, 2018; N. Song *et al.*, 2022).

CSNK1G3 phosphorylates and conducts post translational modifications of caseins in cattle (Buitenhuis *et al.*, 2016). Phosphorylation of such caseins is important for calcium in milk and *CSNK1G3* has been suggested as a candidate gene behind economic focus on protein yields in cattle milk (Bijl *et al.*, 2014; Kim, Sonstegard and Rothschild, 2015). *SLC51B* (Solute Carrier Family 51 Subunit Beta) is essential for intestinal bile acid absorption and dietary lipid absorption (Ballatori *et al.*, 2013). Expression of this gene is significantly associated with lipid droplet area and inflammation it is highly likely linked to bovine milk fatty acid composition (Gebreyesus *et al.*, 2019; Herrera-Marcos *et al.*, 2022). *PTGER2* (Prostaglandin E Receptor 2) is a metabolite of arachidonic acid and has different biological activities in a wide range of tissues. *PTGER2* is expressed in the endometrium and contributes to endometrial growth in cattle, therefore important in pregnancy (Atli et al., 2010; Gao et al., 2017).

A host of genes relating to immune and inflammatory responses were found under selection in UK buffaloes. In RV_UK1, the ROH found on chromosome 6 was additionally found under positive selection in river buffalo that largely comprised of Italian Mediterranean (Sun, Shen, *et al.*, 2020). Notably the overlapping genes were *AP4B1*, *PHTF1* and *BCL2L15*. Whilst *PHTF1* and *BCL2L15* have been linked with general immunity and apoptotic functions, *AP4B1* is associated with Bovine Leukaemia Virus in Holstein cows. This gene has been found to be involved in activation of nuclear factor kappa B (NFKB) and becoming highly expressed in B lymphocytes upon stimulation (Brym *et al.*, 2016; Lobo-Alves, de Oliveira and Petzl-Erler, 2019). This selection signature in RV_UK1 is therefore most likely derived from its original Italian population and has so far been maintained. Mediterranean buffalo directly contributed approximately 700 million Euros to the Italian economy and any prevalence of infectious diseases in buffalo can contribute to significant economic losses through reduced yields, culling, and a decrease in market value of milk (Villanueva *et al.*, 2018; Chirone *et al.*, 2022; Vecchio *et al.*, no date). The Italian buffalo industry has invested in improving Mediterranean buffaloes' health and milk yields through, for example, improved farming conditions, feed optimization, and genomic selection (Biffani, Gómez and Cesarani, 2021; Cesarani *et al.*, 2021; Salzano *et al.*, 2021, 2023; Gowane and Vohra, 2022; Vecchio *et al.*, no date). Underlying selection in immunity genes contribute to resistance of disease outbreaks that occur on large farms. In the case of ovine herpesvirus 2, which Italian buffalo can contract, there is an absence of rapid disease spread to all individuals appear to suggest some form of innate

immunity within herds (Rozins and Day, 2016; Amoroso, Galiero and Fusco, 2017; Ferrara *et al.*, 2023). Therefore, maintenance of selection signatures like this may be of importance in UK buffalo in developing resistant large-scale farms. *RSBN1*, *MAGI3*, and *OLFM3* are further found within this selection signature and associated with immune functions. Chromatin remodelling gene, *RSBN1* is upregulated upon infection with Type 1 PRRSV strains and is a target of HIV-1 (Kaminski *et al.*, 2016; Crisci *et al.*, 2020). *MAGI3* features links to immunity through ATK3 gene and interactions with Human T-cell leukaemia virus type 1 Tax oncoprotein (Ohashi *et al.*, 2004; Farmanullah *et al.*, 2020). While *OLFM3* is associated with microglia functions (Neidert *et al.*, 2018; Chen *et al.*, 2020; Toedebusch *et al.*, 2021).

Regarding RV_UK2 ROHs, *POLL* appears linked to reduced B-cell diversity and prevalence in deficient mice and *DPCD* is upregulated during ciliogenesis (Lee, 2013; Ovsyannikova *et al.*, 2016). Four STAT genes were found within RV_UK2 ROHs. STAT genes are a family of transcription activators that operate through the cytokine stimulated JAK/STAT pathway. This pathway can regulate cellular processes such as growth, proliferation, differentiation, and apoptosis but is prominently involved in immune response (Villarino, Kanno and O'Shea, 2017; Owen, Brockwell and Parker, 2019; Hu *et al.*, 2021). *STAT1* is involved in regulation of cell growth, tumour inhibition and regulation of the immune system (Owen, Brockwell and Parker, 2019; Hu *et al.*, 2021). Reduced or loss of function of *STAT1* increases susceptibility to pathogenic diseases of all kinds (Hu *et al.*, 2021). In buffaloes, *STAT1* has been found to be 99% identical to *Bos taurus* and is expressed mostly in the lymph tissues. Further expression is found in the mammary tissues leading to *STAT1* being described as a candidate gene for milk production traits (Deng *et al.*, 2015, 2016). This follows the same trend seen in cattle as variants surrounding *STAT1* have been associated with a variety of milk traits in Holstein, Jersey, Brown Swiss, and Czech Fleckvieh cattle (Cobanoglu *et al.*, 2006; Askari *et al.*, 2013; Rychtářová *et al.*, 2014; Ozden Cobanoglu *et al.*, 2016; Pegolo *et al.*, 2016). Additionally, *STAT1* may regulate the endometrium during pregnancy in cattle (Dickson *et al.*, 2022). Interferon pathways involving *STAT1* are upregulated during pregnancy and associated with lower conception rates in cattle under heat stress (Catozzi *et al.*, 2020). One study in buffaloes showed that almost all buffalo cows that stopped reproducing, and therefore, stopped producing milk showed moderate to severe endometriosis, a disease which increases growth of the endometrium (Salzano *et al.*, 2020). Although unclear as to how, *STAT1* related

pathways may be involved in endometriosis (Hou *et al.*, 2022; Park and Han, 2022). The other STAT genes found within RV_UK2 ROHs were *STAT2*, *STAT4*, and *STAT6*. *STAT2* forms a complex with *STAT1* and operates within its immune response, pro-inflammatory, and anti-tumour pathways while *STAT4* immune response involves T cell activation (Villarino, Kanno and O'Shea, 2017; Owen, Brockwell and Parker, 2019; Hu *et al.*, 2021). Although involved in immune responses, *STAT6* has been found within ROHs of Iranian buffalo and under selection within Brazilian murrah buffaloes and is thought to increase resilience of buffalo cows to negative energy balance (Ghoreishifar *et al.*, 2020; de Araujo Neto *et al.*, 2022). This occurs when energy expenditure of an individual is greater than its intake and dairy cows are susceptible after giving birth causes detrimental effects to their health and milk production (Mozduri, Bakhtiarizadeh and Salehi, 2018).

Several genes involved in immune response were found under positive selection in both UK populations. *TUT7* is expressed in macrophages and impacts innate immune responses, showing antiviral functions (Kozlowski *et al.*, 2017; Le Pen *et al.*, 2018). *PDCD7* promotes apoptosis when overexpressed. This gene is induced by glucocorticoids and is involved in regulation of T-cells whilst a target of influenza and staphylococci (Guan *et al.*, 2019; Shaohua Wang *et al.*, 2021; Ghafouri-Fard *et al.*, 2022). *ANKDD1A* meanwhile is tightly associated with T cells that function in immune responses to tumour growth (Zhao *et al.*, 2021).

As seen, many of the genes found in ROH and XP-EHH analysis are associated with milk production and immunity in UK buffaloes which follows two of the key traits of keeping buffaloes as livestock. Genes surrounding antioxidant and thermal regulation were additionally found that may support the health and productivity of water buffaloes. Under selection in both populations, *CBS* is active in the production of taurine that produces an antioxidant function by preventing the build-up of reactive oxygen substances in the respiratory chain and participates in energy homeostasis (Zhou *et al.*, 2017; Lopreiato *et al.*, 2020; Darang *et al.*, 2022). Found alongside *CBS*, *NDUFV3* is directly involved in ATP production due to being an accessory subunit of mitochondrial Complex I. This gene has been found under selection in fine-wool sheep in cold climates suggesting adaptive role in thermogenesis (T. Guo *et al.*, 2021). *PKNOX1* meanwhile is a regulator of oxidative phosphorylation components (Kanzleiter *et al.*, 2014). *CRYAA* in the same region is a heat

shock protein that is a major component within the eye lens that maintains transparency and refractive power (Chang *et al.*, 2022; Chakraborty, De and Saha, 2023). *CRYAA* along with *CBS*, *NDUFV3*, and *PKNOX1* are found in the same region under selection as *PDE9A* which is seen under selection in murrah buffalo (Tyagi *et al.*, 2021). Therefore, it may be automatically assumed that this region is key to milk production in tropical climate however, *CRYAA* is has been shown to be robust across a variety of temperatures and is adaptable across organisms in different environmental temperatures (Posner *et al.*, 2012; Malik *et al.*, 2021; Chang *et al.*, 2022). This region may present an opportunity to identify environmental adaptation in river buffaloes.

The genes highlighted above show associations with relevant biological functions for livestock that can be used to maintain productive and healthy stock. Water buffaloes originate from topical/sub-tropical Asia and as such are adapted to these climates. Therefore, since residing in Europe, water buffaloes likely largely still possess these tropical adaptations. They may, however, be evolving to European conditions if natural and artificial selective pressures are present or buffaloes are resilient to environmental change through phenotypic plasticity (Fox *et al.*, 2019). For example, water buffalo calves are born with fur all over however into adulthood they lose this fur to help with heat regulation in India (Yáñez *et al.*, 2020). Meanwhile in colder European climates, this hair is not completely lost (Yáñez *et al.*, 2020). Due to Italy's organised buffalo industry, research is actively being conducted on Mediterranean buffaloes for improvement of the local industry. Recently, it was found that Mediterranean buffaloes are productive at a temperature-humidity index (THI) above 60 (Matera *et al.*, 2022). Below this, general declines in productivity are observable as milk yield declines and somatic cell score increases (Matera *et al.*, 2022). Somatic cell score is a major milk indicator for dairy livestock as the majority of cells comprising this statistic are leukocytes (Bradley and Green, 2005; Lindmark-Månsson *et al.*, 2006; Sant'Anna and Paranhos da Costa, 2011). An increase in somatic cell score in milk suggests greater vulnerabilities to infections, particularly mastitis, and other implications such as greater level of inflammation (Bradley and Green, 2005; Lindmark-Månsson *et al.*, 2006; Sant'Anna and Paranhos da Costa, 2011). The UK has an average THI of 59 and as such, buffaloes in the UK face conditions at times of the year that are not ideal for productivity (Dunn *et al.*, 2014). Therefore, the genes found under selection may suggest the UK buffaloes are adapting to the UK through immunity

related genes to overcome greater susceptibility to infection, therefore minimizing any potential loss in productivity. Lastly, a region under selection that features linked milk production and thermoregulatory genes may indicate additional adaption to a new environment, however further analysis is required to decipher whether this region has been maintained from India or if divergence is present.

2.6. Conclusion

The results found in this study reveal the extent of genetic variation among UK buffalo for the first time. The generation of novel genomic data can uncover the evolutionary history behind UK buffaloes' genetic diversity, identifying any detrimental effects or important regions for preservation. As such, these can aid in initiating development of genomic data into management plans to ensure adaptable, robust, and productive buffaloes for the future. High levels of genetic diversity in UK buffaloes show that despite their low numbers of individuals, they have avoided any potential harmful genetic effects such a bottlenecks and inbreeding depression. This is likely due to the lack of intense artificial selection in the history of river buffaloes that other commercial livestock breeds have faced. Additionally, despite UK buffaloes being almost entirely of Italian descent, imported individuals have likely come from a range of lineages maintaining variation across the genome. This variation likely means that there is a great potential to shape UK buffaloes to their surrounding environment and to improve their productivity. The advent of genomic data means that this can be planned thoroughly as to minimize loss of genetic diversity and risk of genetic inbreeding leading to health problems. A range of candidate genes were identified that may be of importance to buffaloes in the UK. Several genes were found relating to milk production, immunity, antioxidant, and thermoregulation, all of which would be expected given river buffalo are known for their milk, disease resistance, and adaptations to tropical climate. Interestingly though, UK farms displayed little consistency in their underlying regions of selection. Functional analysis of QTLs showed an excess of production associated regions in RV_UK1 which is the production-orientated farm, whilst more classical milk associated regions were found in RV_UK2, the hobby-orientated farm. This implies that despite the relatively little amount of time of water buffalo in the UK, they are able to genetically respond to differing

selective pressure and this evidence of adaptability provides encouragement for future food security and changing climate in the UK.

Chapter Three

Recent Selection and Adaptation in Murrah Water Buffalo

3.1. Abstract

The murrah buffalo breed comprises approximately 25% of the global population and is the most important riverine breed. The productivity of this dairy breed has led to its global transportation in the past century, therefore exposing the breed to new biota, environments, and differing demands from people. This chapter genotyped a novel Indian murrah population and performed a comparative analysis to global murrah buffalo populations to understand changes in genetic variation and identify genomic regions of recent adaptation. All populations featured high levels of heterozygosity as expected due to originating from Indian stock where genetic diversity is greatest for river buffalo. No incidences of increased inbreeding were found in newly established populations following exportation out if India. A population expansion was observed within Indian murrah buffaloes, possibly representing a potential historical admixture event ~250 years ago that led to breeding with local buffalo breeds. Sub-population structuring found within the novel Indian murrah population may indicate development of the breed with those genes under selection relating to milk and reproductive functions. Genes linked with energy regulation were found in Brazilian murrah, a population farmed in modern, commercial farming set-ups. Selection for alternative livestock traits were found in Colombian buffalo that have ancestry in the draught buffalypso breed, and in Bulgarian murrah as coat colour related genes that may be linked to energy demands for meat and draught. Lastly, evidence of adaptations to new environments were identified through genes relating to neuronal, vascular, and metabolic functions that may enable buffaloes to alter their behaviour and homeostatic processes in response. Results from this chapter indicated that buffaloes can adapt to differing selective pressures.

3.2. Introduction

Approximately 97% of the global water buffalo population is found across tropical and sub-tropical Asia with greater than 50%, totalling 110 million heads, residing in India alone (FAO STAT; DAHD 2023). The extensive use of buffaloes in Southern Asia has been underpinned by their adaptations to a tropical and sub-tropical climate. Buffaloes feature both physical, and behavioural adaptations that enable them to reduce stress and remain productive under harsh conditions (Yáñez *et al.*, 2020; Bertoni *et al.*, 2021; Mota-Rojas *et al.*, 2021). For example, buffalo breeds in India feature high melatonin levels to cope with high temperatures and exposure to sunlight, whilst always losing fur in adulthood (Yáñez *et al.*, 2020). Behaviourally, buffaloes thrive alongside water as wallowing is needed to overcome overheating due to their inability to sweat whilst mud helps to resist diseases and parasites (Yáñez *et al.*, 2020; Mota-Rojas *et al.*, 2021). Due to the need for buffaloes to remain near water bodies, there has been little historical movement out of their natural domestic range in Southern Asia. Other livestock species such as cattle or equids are more capable of moving large distances across dry lands at quicker speeds (National Research Council 1981; Siddiky and Faruque, 2017). Historical migrations of domestic buffaloes only led to the founding of sporadic populations across wetlands areas of the Middle East and Mediterranean (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020; Wordsworth *et al.*, 2021). Development of the modern world and globalization over the past 100 years has enabled easy transfer of goods and services around the world with livestock being no exception.

Water buffaloes have been transported across the globe, albeit in very small numbers, reaching countries such as Brazil and Colombia in South America, Cuba in the Caribbean, Mozambique in Africa, Australia, and Central Europe (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). These countries have imported buffalo largely for their milk production, but also for beef and draught. Countries importing livestock in modern times have had the luxury of importing the most suitable and productive breeds, and with technology such as assisted reproductive technologies (ART), there is greater access to highly valuable individuals (Hansen, 2014, 2020; Gelayenew and Asebe, 2016). The ability to specifically choose the best breeds and individuals typically leads to a small number of international breeds within a livestock species (Hoffmann, 2011; Biscarini *et al.*, 2015; Sponenberg *et al.*, 2018). Well known examples of this are the Holstein-Friesian cattle and Large White pigs

(Marshall, 2014). In water buffalo, the murrah breed from India is the most frequent and globally traded breed, whilst the Mediterranean buffalo from Italy, and Nili-Ravi breed from Pakistan are used to a lesser extent (Maheswarappa *et al.*, 2022). The murrah buffalo as a breed falls under the riverine buffalo species (*Bubalus bubalis bubalis*), deriving from the Jind, Hisar, and Rohtak districts of the Haryana province in North-western India (Pawar, 2012; Singh, 2014; Kumar *et al.*, 2019).

Murrah buffalo account for 44% (48 million) of the Indian buffalo population (Table 3.1) and their dominance as a breed has been due to their greater milk yields and productivity (DHAD 2013). No other breed in India makes up greater than 4% of the buffalo population, with only non-descript buffaloes matching a similar share at 43% of the population (DHAD, 2013). Murrah buffalo on average exceed 2,000kg per 305 days of lactation with elite milkers producing up to 5,000kg (Borghese, 2013b; Thiruvenkadan *et al.*, 2014; Verma *et al.*, 2016; Kumar *et al.*, 2019; Zhang, Colli and Barker, 2020; Bharadwaj *et al.*, no date; Jakhar, Vinayak and Singh, no date). Buffalo in India are effective milk producers, contributing approximately 55% of the total milk production, despite comprising of 35% of total cattle (Geetha, Chakravarty and Kumar, 2006; Cruz, 2007; Valsalan *et al.*, 2014). Their milk is rich in protein and fat, therefore providing an important energy source particularly for rural communities (Nanda and Nakao, 2003; Kumar *et al.*, 2019). Buffalo in India are integral for smallholders as more than a source of food. Buffalo are known as a living bank, and are generally considered as an important asset for family wealth (Nanda and Nakao, 2003). The majority of buffaloes are kept by small holdings that mostly keep a few individuals in low external input systems (Deb *et al.*, 2016). Larger commercial operations such as those seen in developed countries are rare (Kumar *et al.*, 2019). Therefore, despite the development of murrah buffalo, this breed still does not match the outputs of industrial cattle breeds that have access to modern optimized farming systems and sophisticated breeding management schemes including genomic selection (Zhang, Colli and Barker, 2020). With the movement of buffalo today and ease of access to genetically improve individuals through ART, farmers and countries can invest into murrah buffalo and therefore continue to develop the breed. In their native India, murrah buffalo are frequently bred with local buffalo to produce a cross termed graded murrah (Kumar *et al.,* 2017). This increases the productivity of the local buffaloes, but also helps to retain some local adaptations and disease resistance. This theme and purpose of

crossbreeding is common among all countries of importation. Brazil primarily import murrah buffaloes and, to a lesser extent, other breeds such as Mediterranean and swamp carabao to establish an organised farming system (da Silva *et al.*, 2021). The majority of murrah importations are for the purpose of improving local breeds and populations through crossbreeding.

Table 3.1: Number of buffaloes per breed in India as of livestock census 2012 (Taken from DHAD 2013). Graded buffaloes are crossbred individuals of mixed ancestry.

Breed	Number of Purebred Individuals	Number of Graded Individuals	Total Number of Individuals per Breed (Pure + Graded)	Proportion of Total Indian Buffaloes (%)
Murrah	11,686,198	36,568,676	48,254,874	44.39
Surti	1,886,280	2,006,614	3,892,894	3.58
Mehsana	2,676,699	948,426	3,625,125	3.33
Jaffarabadi	571,077	1,200,421	1,771,498	1.63
Bhadawari	583,599	1,170,188	1,753,787	1.61
Nili Ravi	129,411	547,834	677,245	0.62
Pandharpuri	287,751	195,987	483,738	0.45
Banni	239,572	142,550	382,122	0.35
Marathwadi	278,502	98,093	376,595	0.35
Nagpuri	73,584	117,410	190,994	0.18
Khalahandi	115,213	26,802	142,015	0.13
Toda	3,003	2,533	5,536	0.01
Chilika	2,599	787	3,386	0.00
Non-Descript			47,142,313	43.37
Total	18,533,185	43,023,385	108,702,122	100.00

Bulgaria imported murrah in 1962 to cross with their indigenous Mediterranean buffaloes, forming the Bulgarian murrah breed that has since been maintained (Borghese, 2013b). With development of this new breed, milk yields increased proving the programme

to be successful (Borghese, 2013a). Similarly, the Philippines imported a large number of Indian and Bulgarian murrah to cross with their swamp carabao buffalo (Cruz, 2007). Swamp buffaloes were historically domesticated for draught however recent mechanisation across South East Asia has led to a continuing decline in population size (Cruz, 2007; Zhang, Colli and Barker, 2020) Therefore, swamp buffaloes have recently been repurposed for greater food production. At an adult size of 250-300kg, they are far smaller than river buffaloes (450-600kg), producing less beef (Cruz 2009). Additionally, swamp buffaloes produce little milk at 400-800kg per lactation cycle, far less than 1500kg of undeveloped river buffalo (Cruz 2009). Crossbreeding with murrah buffalo greatly increases the productivity of swamp buffalo for food production, with hybrids typically producing an intermediate phenotype.

Newly established murrah populations will encounter new environmental and farming pressures. Therefore, this chapter looks to identify recent adaptations occurring across global murrah buffaloes. This includes the addition of a newly genotyped Indian murrah population to the growing data generated using the Axiom^{TM} Buffalo Genotyping Array. This novel Indian population was compared to five populations from South America, Philippines and Bulgaria depicting the most prominent murrah utilizing countries. The study had two aims. The first aim was to genetically characterize murrah populations, assessing differences in levels of genetic diversity and inbreeding. Strong selection for increased yields is typically found in tandem with declines in genetic diversity as individuals regarded to be of less merit are removed from the breeding pool (Blackburn, 2012; Leroy, 2014). Reduced diversity can lead to reduced adaptability or an increase in inbreeding depression (Doublet *et al.*, 2019). Examples of this has been observed in other livestock species through increased incidence of reproductive issues that decreases productivity (Malhado *et al.*, 2013; Manoj *et al.*, 2017)

The second aim of the study was to identify unique regions under selection in each population. Recent adaptations can occur through a variety of selective pressures between a livestock population, its environment, and production system. Identifying the underlying genetic changes behind adaptations will aid in conservation of genetic resources in livestock and improve breed management for future economic goals and environmental resilience (Biscarini *et al.*, 2015). Selection is expected to leave distinct footprints on the genome, however detecting these can be challenging due to a variety of factors including recombination, selection intensity, demography, and admixture (Pritchard, Pickrell and Coop,

2010; Hermisson and Pennings, 2017). A range of methods have been produced to study selection, of which several are employed here. Hard selective sweeps can be identified through long haplotypic methods that detect alleles nearing fixation across a population due to a rapid shift in the frequency of a beneficial allele (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). These sweeps typically cause linked flanking regions to also appear under selection (Qanbari and Simianer, 2014; Saravanan *et al.*, 2020). Contrastingly, soft sweeps affect shorter regions, and typically leave more standing genetic variation across the population, making them harder to detect (Saravanan *et al.*, 2020). Several studies have focused on selection in buffaloes using runs of homozygosity (Davoudi *et al.*, 2020; Fallahi *et al.*, 2020; Ghoreishifar *et al.*, 2020; Macciotta *et al.*, 2021; Nascimento *et al.*, 2021; Noce *et al.*, 2021; S. Liu *et al.*, 2022), long haplotypic methods (Mokhber *et al.*, 2018; Lu *et al.*, 2020), and association studies with livestock traits (de Camargo *et al.*, 2015; Cesarani *et al.*, 2021). This study attempts to include finer scale selection occurring in buffaloes which may represent local adaptations. The use of spatial and landscape genomics will be critical for maintaining productivity of domesticated species in differing environments to avoid reduced or low-quality yields from maladapted individuals (Hayes, Lewin and Goddard, 2013).

3.3. Materials and Methods

3.3.1. Sample Collection & Data Generation

A total of 156 Indian murrah buffalo were sampled across central India. Samples were processed and genotyped at United Biologicals using the Axiom™ Buffalo Genotyping Array (Iamartino *et al.*, 2017). The raw genotyped data was analysed in Axiom™ Analysis Suite Software v4.03 along with the UK buffaloes studied in Chapter One. 148 samples successfully passed genotyping to give 75,679 SNPs, and 130 samples for 63,603 SNPs after quality control in PLINK (Chang *et al.*, 2015). All quality control parameters used in genotyping and PLINK can be found in Section 2.3.1 of Chapter One.

Likewise with Chapter Two, Indian murrah buffalo dataset was merged with public datasets of Colli *et al.*, (2018) and Noce *et al.*, (2021) to obtain five additional murrah derived populations from South America, Europe, and Philippines (Table 3.2). The combined datasets resulted in a total of 40,695 SNPs genotyped for 232 individuals. This dataset was used for Sections 3.3.3 – 3.3.6. An independent set of SNPs was generated via linkage disequilibrium

pruning in PLINK for analysis of genetic diversity and population structure (Section 3.3.2). The first SNP of every linked pair of markers was removed using a sliding window approach of 50 SNPs, step size of 10 SNPs and an R^2 threshold of 0.05 above which markers were considered in LD; this pruning resulted in 6,510 SNPs being retained. To determine that the results were repeatable, the dataset was randomly split in two and analyses were run on each dataset to check that results were consistent across both.

Table 3.2: Murrah derived buffalo populations analysed in this chapter. Locations of origin, breed, sample size, and referenced studies for each population are given.

Population ID	Country	Region	Breed	Sample Size	Reference
RV_IND	India	Karnataka, Andhra Pradesh, Tamil Nadu	Murrah	130	
RV_BRA	Brazil		Murrah	15	(Colli, Milanesi, Vajana, *et al.*, 2018)
RV_COL	Colombia		No Description	12	
RV_PH_BUL	Philippines		Bulgarian Murrah	11	
RV_PH_IND	Philippines		Murrah	6	
RV_BUL_VAR	Bulgaria	Varna	Bulgarian Murrah	58	(Noce *et al.*, 2021)

3.3.2. Genetic Diversity & Population Structure

Observed (H_O) and expected (H_E) heterozygosity along with inbreeding coefficient (F) were calculated for each population in PLINK v1.9 and analysed in R v4.0.0 (R Core Team, 2018) using custom R scripts. An ANOVA was used to test for significant differences in H_O across populations while a Welch's t-test (non-normality determined using Shapiro-Wilk test) tested for significant differences between H_O and H_E within each population.

Genetic differentiation between murrah populations was assessed using pairwise F_{ST}. Pairwise population F_{ST} was calculated using Arlequin v3.5.2.2 (Excoffier and Lischer, 2010). A neighbour-net network to illustrate the relationships between groups of murrah buffaloes was created in SplitsTree v4.14.4 (Huson and Bryant, 2006). Genetic clustering of individuals was done using ADMIXTURE v1.3.0 (Alexander, Novembre and Lange, 2009), testing for the ideal number of unique ancestral genetic clusters (K) across the dataset. Values of K from 1

to 20 were tested. A cross-validation method, implemented within ADMIXTURE, was used to determine the most suitable value of K in the data.

3.3.3. Linkage Disequilibrium & Effective Population Size

Pearson's correlation coefficient, r^2, between pairs of SNPs was calculated in PLINK for each population. SNP comparisons were calculated between adjacent SNPs and SNPs separated by up to 10Mbp to estimate average linkage across the genome. Linkage decay was estimated up to 1Mbp using bins of 50kbp intervals. The distance when r^2 equalled 0.2 for each population was taken as the linked flanking distance away from a SNP to detect genes under selection (Mokhber *et al.*, 2019).

Effective population size (N_e) was calculated following the method of Sved (1971) used in Chapter Two (Section 2.3.4, Equations 1, 2 & 3) (Mokhber *et al.*, 2019). In summary, Sved, 1971 N_e equations evaluates the average r^2 linkage values (adjusted for unequal sample sizes) and distance, c, between SNPs in Morgans.

3.3.4. Runs of Homozygosity

Runs of Homozygosity (ROH) were calculated in PLINK. Parameters used to identify ROHs can be found in Chapter Two Section 2.3.5, briefly: the minimum number of SNPs per ROH in each population was calculated using Equation 4 in Section 2.3.5 (Lencz et al., 2007; Purfield et al., 2012). The minimum number of SNPs per ROH for each murrah population resulted in counts between 23 – 34 (RV_BRA = 31, RV_BUL_VAR = 34, RV_COL = 29, RV_IND_1 = 29, RV_IND_2 = 30, RV_IND_3 = 34, RV_PH_IND = 23, RV_PH_BUL = 27). ROH PLINK outputs were analysed and summarised using detectRUNs package in R (Biscarini et al, 2019). An ANOVA was conducted to test for significant differences in both F_{ROH} and length of ROHs between populations in R.

To identify areas of divergence in murrah buffaloes that may relate to recent selection, pairwise population F_{ST} values were calculated for all population comparisons. F_{ST} was calculated in a windowed manner using VCFtools (Danecek et al 2011). Window sizes were 1Mbp with a step size of 100kbp. Windows with high F^{ST} values occurring in ROHs within populations may indicate areas of divergence and underlying adaptive regions. Genes within ROHs featuring high F_{ST} were identified as candidate genes of selection, and their functions were identified.

3.3.5. Environmental & Spatial Selection

PCAdapt (Luu, Bazin and Blum, 2017) was used to find SNPs under selection while accounting for population structure among murrah buffaloes, and R-Samβada (Stucki *et al.*, 2017; Duruz *et al.*, 2019) was used to identify SNPs associated with differences in environmental variables between the countries. Using PCAdapt, the dataset was summarised into 20 principal components (PC) before identifying the ideal number of PCs that reflected population structure using Cattell's rule. After retaining meaningful PCs, the SNP loadings of each PC were checked for even distribution. A q-value was generated from the p-values for each SNP by applying a false discovery rate (FDR) of 10% (Luu, Bazin and Blum, 2017). Retained SNPs had their associated PC identified and genes within 120,000bp either side were identified as being linked and potentially under selection.

Selection occurring in buffaloes relating to environmental variables was conducted using R-Samβada. Coordinates for murrah populations taken from Colli *et al.*, (2018) and Noce *et al.*, (2021) were approximated by taking primary regions of buffalo industry in said countries as follows; RV_BUL_VAR (43.236073, 27.84434), RV_BRA (-23.563483, -46.730364), RV_COL (2.576514, -77.870692), RV_PH_BUL and RV_PH_IND (15.73717, 120.936379). Environmental variables were obtained from WorldClim v2.1 (https://worldclim.org/data/worldclim21.html), at a spatial resolution of 5 minutes to allow for a regional average of the surrounding area. Various metrics measuring temperature and precipitation were obtained from the WorldClim dataset. Collinearity was reduced from the dataset of environmental variables by removing the second variable of pairs with an $r^2 > 0.8$, leaving nine variables (Longitude, Latitude, bio1, bio2, bio3, bio13, bio14, tmax7, prec11). Population structure was accounted for in the analysis via a PCA method implemented within R-Samβada. Six PCs were used in line with population structure results from ADMIXTURE and PCAdapt analysis here (Sections 2.4.1 & 2.4.4). After applying an FDR of 10%, the most significant SNPs for each environmental variable were analysed for their nearest gene to identify potential adaptive properties (Luu, Bazin and Blum, 2017; Duruz *et al.*, 2019).

3.3.6. Signatures of Positive Selection

Cross-population extended haplotype homozygosity (XP-EHH) implemented in selscan (Szpiech and Hernandez, 2014) was used to identify recent positive selective sweeps unique

to each murrah population. All populations were phased in Beagle v3.3 (Browning and Browning, 2007), before calculation of pairwise XP-EHH scores using standard settings. Pairwise comparisons for all population combinations were computed. XP-EHH scores were normalised using the norm package of selscan. The median value was calculated for each SNP of all pairwise comparisons for each population to identify SNPs persistently under selection. Median XP-EHH scores were then converted into p-values and negative $\log_{10}$ values. SNPs featuring a p-value below 0.001 (extreme outliers), and with an XP-EHH score greater than 2 were considered as candidates of selection and associated genes were identified.

3.4. Results

3.4.1. Genetic Diversity & Population Structure

Genetic diversity was high across all murrah populations. H_O ranged from 0.448 (RV_BRA) to 0.506 (RV_PH_IND) with a mean of 0.469 (±0.020) across all populations (Table 3.3). H_O was non-significantly higher than H_E for all populations. There was no presence of inbreeding in any population as F was negative across all populations (-0.242 – -0.007). H_O (χ^2 = 189.77, df = 5, $p < 0.0001$) and H_E (χ^2 = 786.43, df = 5, $p < 0.0001$) significantly differed across the six murrah populations.

Table 3.3: Genetic diversities across murrah buffalo populations. H_O = Observed Heterozygosity, H_E = Expected Heterozygosity, F = Inbreeding Coefficient. Bold H_O & H_E values are significantly different ($P < 0.05$). An ANOVA of H_O results showed that RV_PH_BUL and RV_PH_IND had significantly greater H_O compared to the other populations.

Population	H_O (±SD)	H_E (±SD)	F (±SD)
RV_IND	0.465 ± 0.090	0.462 ± 0.078	-0.007 ± 0.032
RV_BRA	0.448 ± 0.145	0.437 ± 0.089	-0.025 ± 0.067
RV_COL	0.462 ± 0.160	0.442 ± 0.087	-0.045 ± 0.034
RV_PH_BUL	0.474 ± 0.178	0.432 ± 0.097	-0.120 ± 0.176
RV_PH_IND	0.506 ± 0.239	0.432 ± 0.097	-0.242 ± 0.158
RV_BUL_VAR	0.459 ± 0.121	0.441 ± 0.097	-0.039 ± 0.039

Average F_{ST} values across all pairwise combinations were low at an average of 0.022 (±0.011), ranging from 0.007 (RV_PH_IND – RV_IND) to 0.049 (RV_BUL_VAR – RV_BRA). A neighbour-net network generated from pairwise F_{ST} scores is shown below (Figure 3.1). The

network shows a simple pattern as Indian murrah are found at centrally while South American and Bulgarian populations are found at opposing ends.

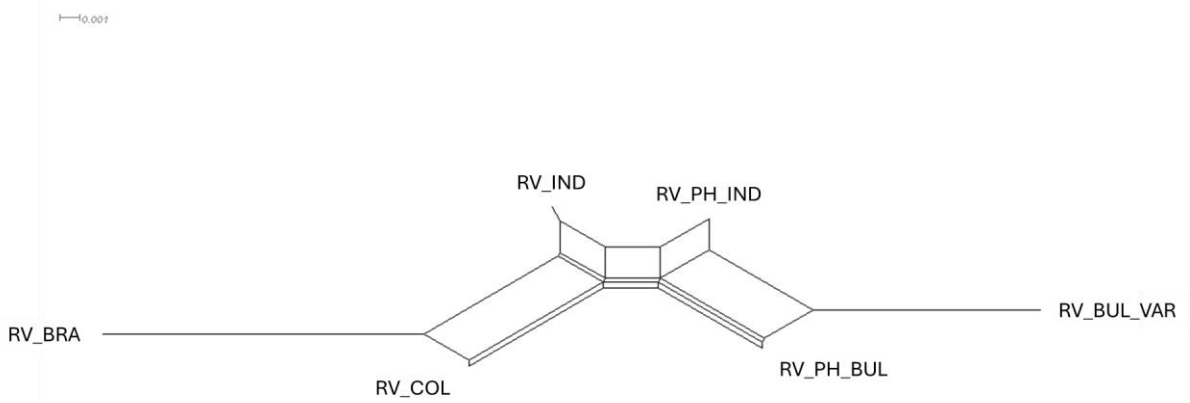


Figure 3.1: F_{ST} network of murrah buffalo populations. Indian murrah sits at the centre of the network with Bulgarian murrah and South American murrah sitting at opposing ends.

Genetic clustering (Figure 3.2) using ADMIXTURE revealed that K=6 was the most preferred solution (CV values found in Figure S3.1). However, only sub structuring of RV_BUL_VAR and RV_IND was found. Populations of RV_BRA, RV_COL, RV_PH_BUL, and RV_PH_IND remained genetically clustered with the main group of RV_IND. Sub structuring of RV_BUL_VAR occurred at K=2 and K=3. RV_IND sub structuring occurred at K= 4 and K=5. Further sub structuring of RV_BUL_VAR was captured at K=6. Sub structuring of RV_IND was accounted for by separation of its three genetic clusters for selection studies. RV_BUL_VAR was not split up due to mosaicism rather than underlying genetic structure.

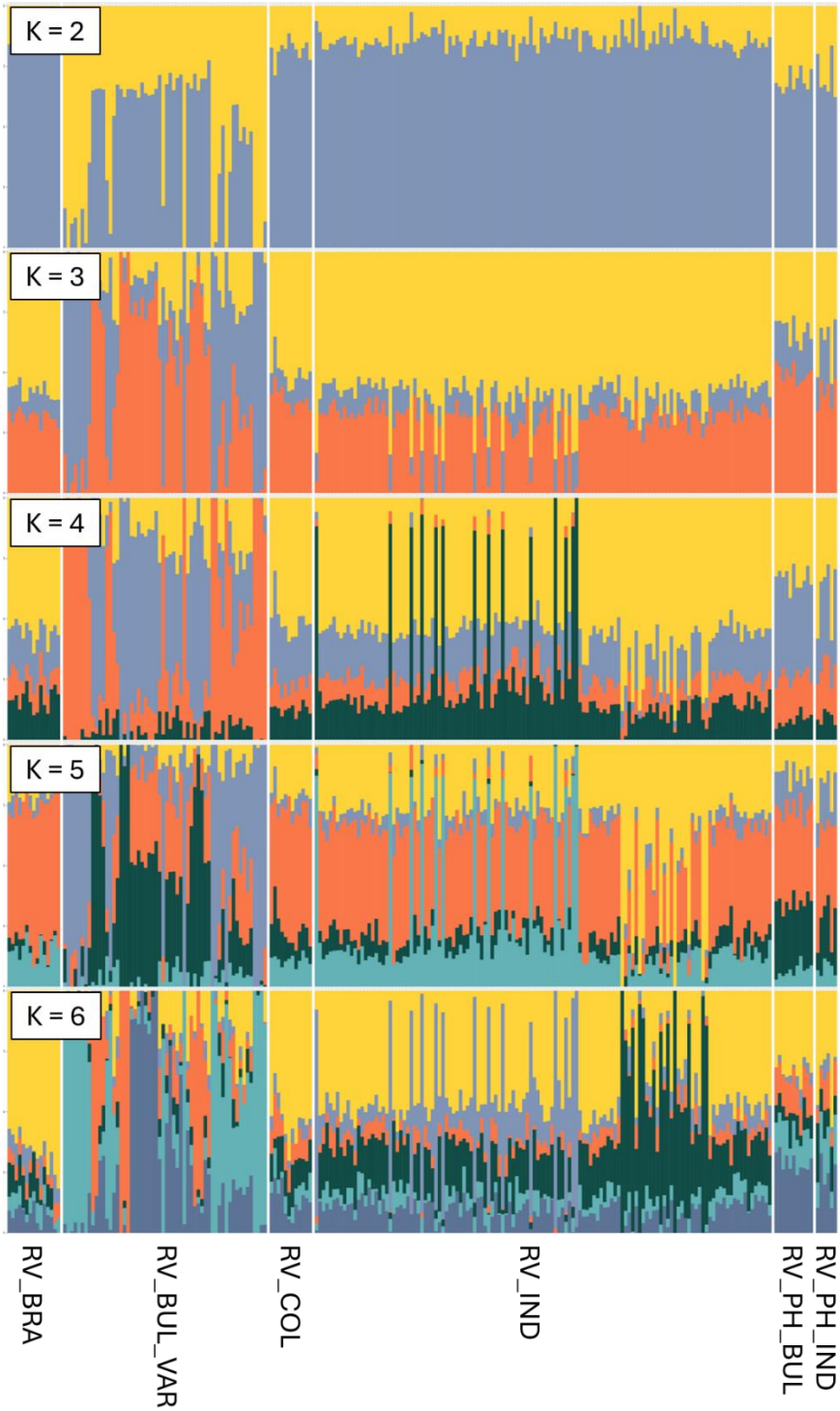


Figure 3.2: ADMIXTURE plots showing genetic clustering of murrah buffalo populations for K=2 to K=6. K=6 provided the most efficient solution according to cross-validation results.

3.4.2. Linkage Disequilibrium & Effective Population Size

Average linkage disequilibrium ranged between 0.240 (RV_IND_3) to 0.415 (RV_PH_IND) at 50kbp from the SNP in question (Figure 3.3). All populations LD descended towards a plateau at 1Mbp. RV_IND_3 was found to have the lowest levels of r^2, declining to 0.018, as RV_PH_IND was found with the highest r^2 levels of 0.270 at 1Mbp. RV_PH_IND features far greater levels of linkage between SNPs than the other populations owing to its small sample size (n = 6). When adjusted for sample size as used in Ne, the r^2 values are 0.298 (50kbp) and 0.123 (1Mbp). The majority of murrah populations showed similar levels of linkage disequilibrium across the genome however, the crossbred RV_BUL_VAR and sub-population of RV_IND_3 showed lower levels. Linkage disequilibrium was used to estimate the length of flanking regions either side of SNPs under selection for each population at r^2 = 0.2. Lengths ranged from approximately 75kbp (RV_IND_3) to 175kbp (RV_PH_BUL), excluding RV_PH_IND, with a mean of approximately 120kbp across murrah buffalo.

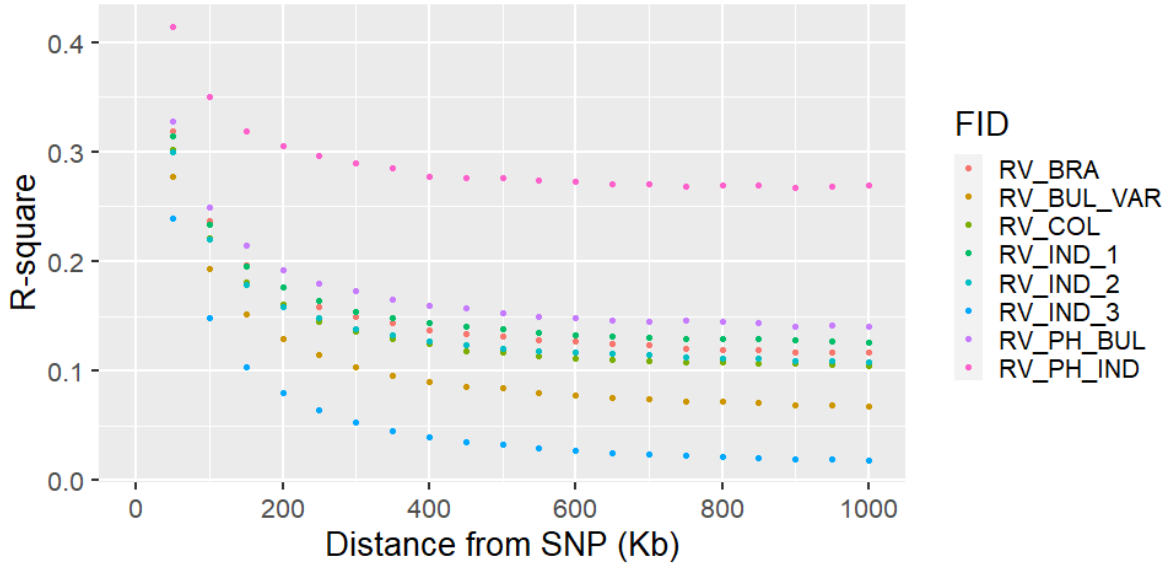


Figure 3.3: Average linkage disequilibrium between SNPs over 1Mbp away from a locus across each murrah buffalo population. Average linkage disequilibrium ranges from 0.240 (RV_IND_3) to 0.415 (RV_PH_IND) at 50kbp from a SNP declining to 0.018 (RV_IND_3) to 0.270 (RV_PH_IND) at 1Mbp. RV_PH_IND higher linkages is likely due to small sample size (n = 6). FID = Family ID (i.e. population).

N_e (Figure 3.4) ranged from 19 (RV_PH_IND) to 1,744 (RV_IND_3) 5 generations ago. The majority of murrah populations showed similar trends of declining N_e from past to

present except RV_IND_3. This Indian population showed a spike peaking at 32 generations ago reaching an N_e of 3,619, indicative of a population expansion before declining until the present day. All populations converge after 250 generations ago towards an N_e of approximately 1,500 at 1,000 generations ago.

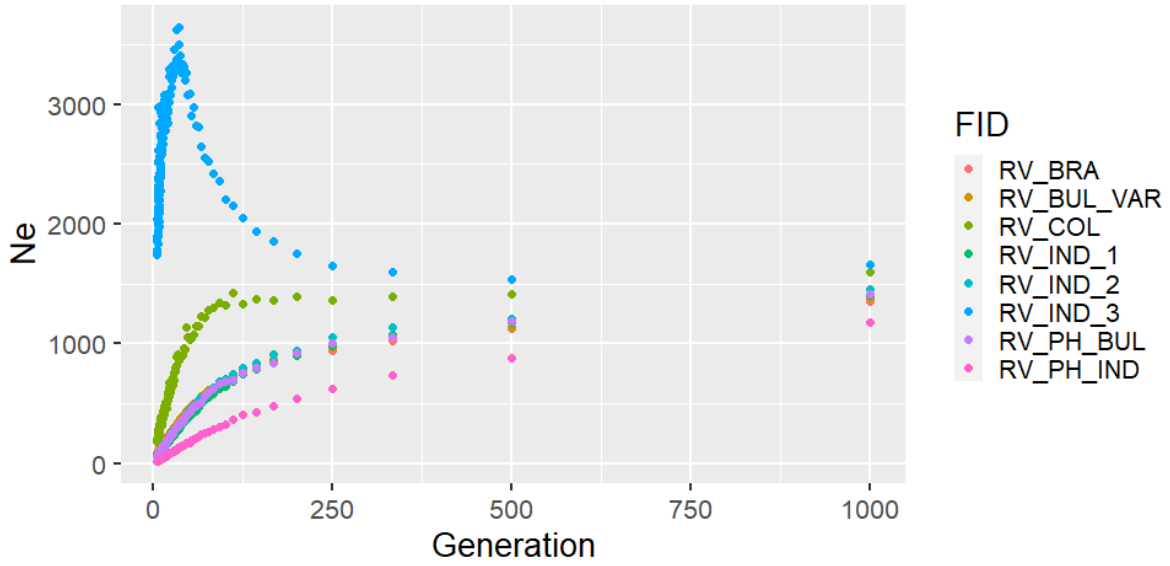


Figure 3.4: Effective population sizes of each murrah buffalo population over the last 1,000 generations. Most populations displayed the typical declining trend from past to present except RV_IND_3 that presented a peak at 32 generations ago indicating a population expansion. FID = Family ID (i.e. population).

3.4.3. Runs of Homozygosity

The lengths and frequencies of ROHs differed between the populations. An overview is shown in Table 3.4. F_{ROH} was low across all populations although significant differences were found between populations for both F_{ROH} ($F_{222,7} = 11.55$, $p < 0.001$) and length of ROHs ($F_{3412,7} = 30.88$, $p < 0.001$). F_{ROH} featured an average of 0.029 (± 0.035) across all buffaloes, ranging from 0.013 (RV_IND_1) to 0.071 (RV_BRA). Tukey's HSD (Table S3.2) showed that RV_BRA and RV_BUL_VAR both had a significantly greater proportion of their genomes as ROHs, and significantly longer ROHs than other populations. Murrah populations outside of RV_BRA and RV_BUL_VAR were dominated by short ROHs with a near absence of long ROHs.

Table 3.4: ROH metrics across murrah populations. F_{ROH} being the inbreeding coefficient for ROHs.

Population	F_{ROH}	Mean Length	Average Number of ROHs Per Class (Mb) Per Individual				
			1-2	2-4	4-8	8-16	>16
RV_BRA	0.071 (± 0.058)	7.50 (± 10.02)	5.07 (± 2.49)	8.80 (± 3.00)	3.40 (± 2.10)	3.33 (± 2.97)	2.73 (± 3.03)
RV_BUL_VAR	0.049 (± 0.036)	6.47 (± 7.11)	1.93 (± 1.30)	8.62 (± 2.57)	3.86 (± 3.03)	2.53 (± 2.13)	1.74 (± 2.19)
RV_COL	0.027 (± 0.018)	3.38 (± 3.48)	8.25 (± 2.34)	8.42 (± 4.27)	1.33 (± 1.50)	1.58 (± 2.11)	0.42 (± 0.79)
RV_IND_1	0.013 (± 0.002)	2.38 (± 1.15)	6.36 (± 2.24)	6.36 (± 1.98)	0.93 (± 1.00)	0.14 (± 0.36)	0.00 (± 0.00)
RV_IND_2	0.022 (± 0.034)	3.68 (± 7.57)	5.59 (± 2.76)	7.00 (± 2.15)	1.41 (± 1.46)	0.41 (± 1.22)	0.29 (± 1.21)
RV_IND_3	0.014 (± 0.024)	3.73 (± 5.57)	2.19 (± 1.44)	5.66 (± 2.22)	0.90 (± 1.01)	0.34 (± 0.72)	0.24 (± 1.48)
RV_PH_BUL	0.039 (± 0.034)	3.84 (± 8.92)	11.20 (± 5.92)	10.20 (± 5.61)	1.30 (± 1.34)	1.70 (± 1.89)	0.50 (± 0.97)
RV_PH_IND	0.019 (± 0.015)	2.00 (± 1.65)	17.67 (± 10.91)	4.33 (± 3.20)	0.50 (± 0.84)	0.50 (± 1.22)	0.00 (± 0.00)

Extreme (top 1% SNPs) ROHs were identified in all populations. The number of ROH outliers rapidly increased as sample size decreased. The population with the most samples, RV_IND_3 (n = 99), featured 13 ROHs meanwhile RV_PH_IND (n = 6) presented 113 ROHs. This is due to threshold for a ROH to be considered significant being increasingly easier to exceed at small samples. For example, RV_PH_IND only required a ROH to occur in one individual. Therefore, it is important to take results cautiously at small sample sizes that may not have strong population-based evidence. Following the application of F_{ST} scores, the number of ROHs showing high F_{ST} scores (indicating potential divergence) was far lower in line with low divergence between populations. Diverging ROHs may infer regions of adaptations in murrah populations. Proportion of divergent significant ROHs for each population can be found in Table 3.5. All populations showed similar numbers of diverging ROHs (0 – 10). Only RV_IND_3 featured zero diverging ROHs.

Table 3.5: Number of divergent significant ROHs per population. Divergent ROHs were determined as regions where high F_{ST} SNPs overlapped with ROH outliers.

Population	Number of Significant ROHs	Number of Divergent ROHs (ROH & F_{ST} outliers)	Percentage of Divergent ROHs (%)
RV_BRA	23	5	21.74
RV_BUL_VAR	15	3	20.00
RV_COL	49	3	6.12
RV_IND_1	35	4	11.43
RV_IND_2	14	4	28.57
RV_IND_3	13	0	0.00
RV_PH_BUL	16	4	25.00
RV_PH_IND	113	10	8.85

The full table of diverging ROHs and the genes found within these can be seen in Table S3.3. One region on chromosome 9 appeared in three populations that were RV_BRA, RV_IND_2, and RV_PH_BUL, whilst one region on chromosome 2 appeared in both RV_IND_2 and RV_COL. Genes found within diverging ROHs were studied through scientific literature to identify putative environmental or livestock adaptive properties. RV_BRA showed gene functions relating to reproduction (BBU9: 58046218 – 62175428), brain and CNS development (BBU1: 4713280 – 49099346), and growth (BBU12: 46848107 - 48349500) functions, whilst South American counterpart RV_COL ROH (BBU2: 45662839 – 48841313) featured several muscle and growth traits. RV_IND_2 ROHs overlapped with RV_BRA and RV_COL, however further ROHs featured functions in reproduction (BBU14: 77753047 – 80039684; BBU17: 28910234 – 30395257). Whilst additional Indian population RV_IND_3 showed no divergent ROHs, RV_IND_1 featured genes relating to immunity, reproduction, and milk production. Bulgarian murrah populations of RV_BUL_VAR and RV_PH_BUL meanwhile featured genes frequently relating to growth.

QTLs overlapping divergent ROHs were identified across all populations (Table S3.3). The number of QTL traits found across populations ranged from 0 to 113. QTL traits were typically spread across Exterior, Milk, Production, and Reproduction QTL classes, while Health, and Meat & Carcass rarely featured. Contrastingly, RV_IND_3 and RV_PH_IND featured no QTLs, despite being on opposing ends of the spectrum for number of divergent ROHs. Due to featuring no QTLs for RV_IND_1 and RV_PH_IND, a chi-square was run for the remaining six

populations. Chi-square test returned a non-significant result ($\chi = 22.92_{25}$, $p = 0.582$). Counts of QTL traits for each population can be found in Table S3.4.

3.4.4. Environmental & Spatial Selection

PCA results reflected those of ADMIXTURE. Of 20 principal components initially computed, 6 were carried forward for identifying significant SNPs relating to population structure. These 6 PCs captured a total of 80.6% of variation across the dataset. PC1 (18.2%) and PC2 (14.4%) largely separated RV_BUL_VAR from the remaining populations (Figure 3.5). PC3 (13.3%) split the sub-populations within RV_IND, whilst PC4 (12.3%) separated South American populations from Indian. PC5 (11.7%) showed further variation within RV_BUL_VAR and PC6 (10.7%) again showing separation of RV_IND and RV_BRA.

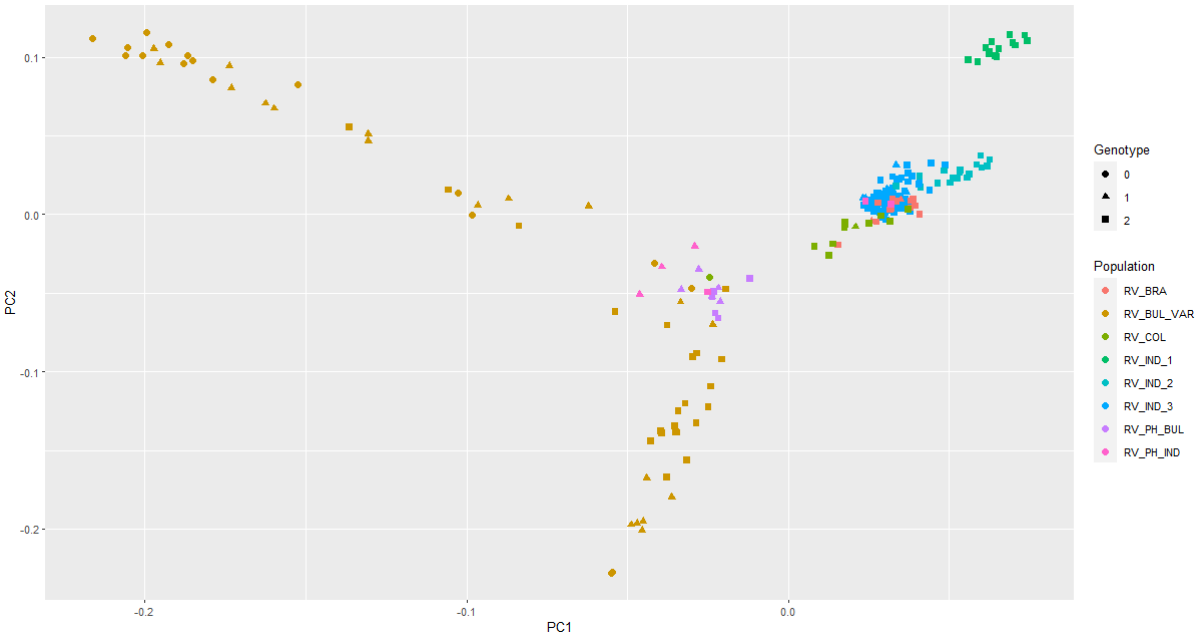


Figure 3.5: PCA plot displaying population structure across murrah buffalo populations. PC1 captured 18.2% of variation across the dataset, while PC2 captured 14.4%. Genotypes for significant SNP AX-85052722 are displayed via shapes with 0 being homozygote allele A, 1 being heterozygote, and 2 meaning homozygote allele B.

PC loadings were evenly distributed across all SNPs with no particular genomic regions dominating. Therefore, the full SNP dataset was retained. 405 SNPs were retained as the top 1% most significant SNPs associated with the PCs. After generating q-values via application of a false discovery rate of 10%, 8 SNPs remained. 6 of the 8 SNPs were significantly associated with PC1 that separate Bulgarian murrah from pure murrah. One SNP was associated with

PC4 and PC6 each. Significant SNPs and associated genes are found in Table 3.6. PCA plots showing all distribution of genotypes for significant SNPs can be found in Figures (3.5 & S3.5 – S3.11).

Table 3.6: Significant SNPs related to population structure across murrah buffaloes. SNPs are associated with different retained principal components after filtering via q-values applying a false discovery rate of 10%. Associated genes were identified within 120,000bp flanking regions of the SNP in question.

SNP	Chromosome	Position	Principal Component	Genes
AX-85115728	1	61409049	4	TRNAC-GCA
AX-85126364	11	23558419	1	RAD51B, ZFYVE26
AX-85042861	11	48810245	1	PRTG, NEDD4
AX-85052722	11	97245872	1	HRH2, CPLX2, THOC3
AX-85136739	12	2572267	6	ARID5A, KANSL3, FER1L5, LMAN2L, CNNM4, CNNM3, ANKRD23, ANKRD39, SEMA4C, FAM178B
AX-85112690	16	10093932	1	TSPAN18, CD82
AX-85114321	16	14825721	1	-
AX-85092744	19	62289241	1	DAP, ANKRD33B

A total of 11 known protein coding genes were associated with PC1 across chromosomes BBU11, BBU16, and BBU19 covering a range of functions. These genes featured several functions such as DNA repair (RAD51B), embryonic CNS development (PRTG), protein degradation (NEDD4), transcription export complex (CPLX2, THOC3), and cell death (DAP). No genes were found associated with AX-85114321 on BBU16 at position 14825721. The closest gene LRRC4C (Leucine Rich Repeat Containing 4C) is a member of netrin family of axon guidance molecules. One gene was found linked to the SNP at BBU1 61409049 associated with PC4, appearing linked with South American populations of RV_BRA and RV_COL. Only TRNAC-GCA was found in this region however the nearest gene was NCAM2 (Neural Cell Adhesion Molecule 2), an immunoglobulin membrane protein, that has previously been associated with milk yields in Brazilian buffalo (de Camargo *et al.*, 2015). Several genes were found associated with PC6 on BBU12 (position 2572267) with these functioning across energy metabolism and reproduction.

R-Samβada analysis revealed a multitude of SNPs associated with the environmental variables used (Correlations between environmental variables shown in Figures S3.12 and

S3.13). The top outlier SNPs for each environmental variable were analysed for their nearest gene. An example of this is shown in Figure 3.6 showing q-values of SNPs in relation to average precipitation in November (mm). Investigated SNPs and genotypes for each environmental variable and their nearest gene can be found in Table S3.14.

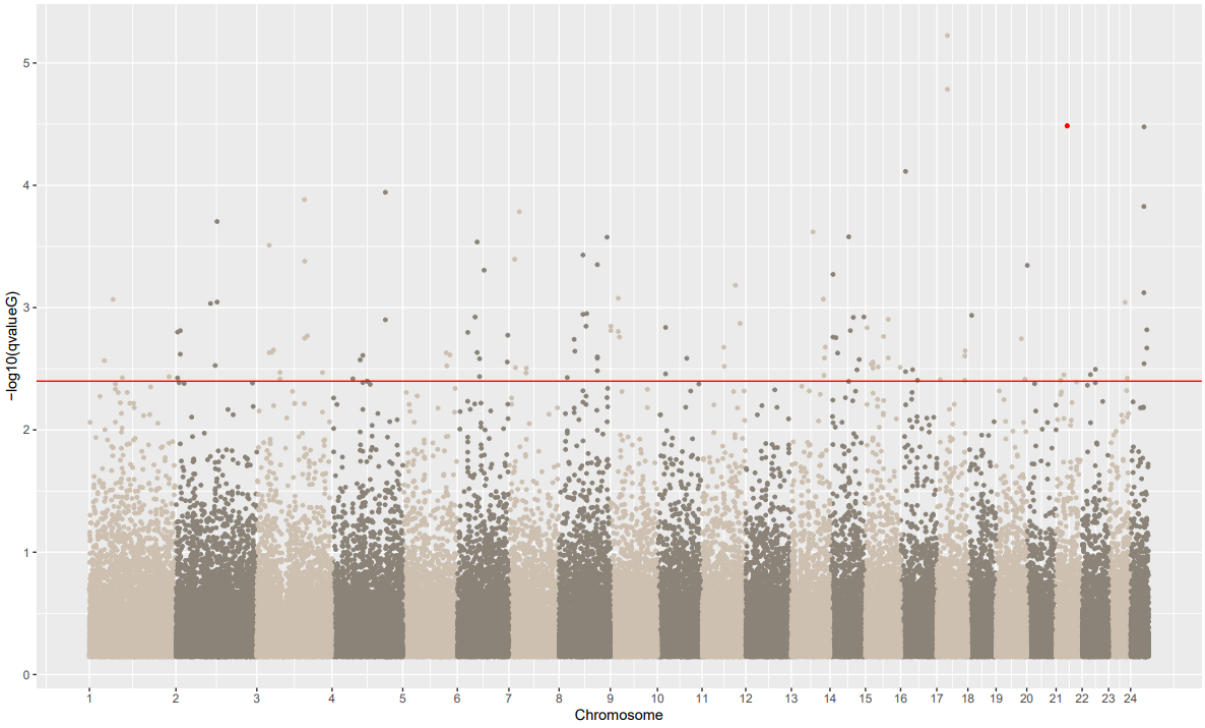


Figure 3.6: Manhattan plot showing SNP q-values calculated from G-score for association with average precipitation in November (mm). The red line indicates the threshold for top 0.1% SNPs, and the red dot shows SNP AX-85126469 where the genotype GG is highly associated with precipitation. The SNPs is nearby the gene CHL1 which has thermotolerance functions. Density of plot includes all three genotypes for each locus.

Functions relating to regulation of lipid metabolism involved TBC1D4 (GG at AX-85059527; Longitude), LYPLAL1 (TT at AX-85142269; Latitude), RAB18 (AA at AX-85104855; bio3), IRS1 (GG at AX-85105029; bio3). Several genes found have associations with functions surrounding high altitude or cold temperature. These were TENM2 (GG at AX-85106286; Longitude), NBEAL1 (TT at AX-85061534; bio2), HS3ST4 (CC at AX-85098066; bio2), and CHL1 (GG at AX-85126469; prec11). An example of the distribution of the genotype of GG for SNP AX-85126469 across murrah buffaloes is plotted in Figure 3.7. This SNP was nearby the CHL1 gene and associated with average precipitation in November (mm). 31.5% of Indian murrah (RV_IND) featured the GG genotype that was rare amongst other populations (9.8%).

Additionally, different environments host a variety of alternative diseases and pathogens, and genes with functions relating to immune response were found such as TENT5A (AA at AX-85094684; bio2), MFSD1 (CC at AX-85130232; bio14), IGSF21 (CC at AX-85045997; tmax7).

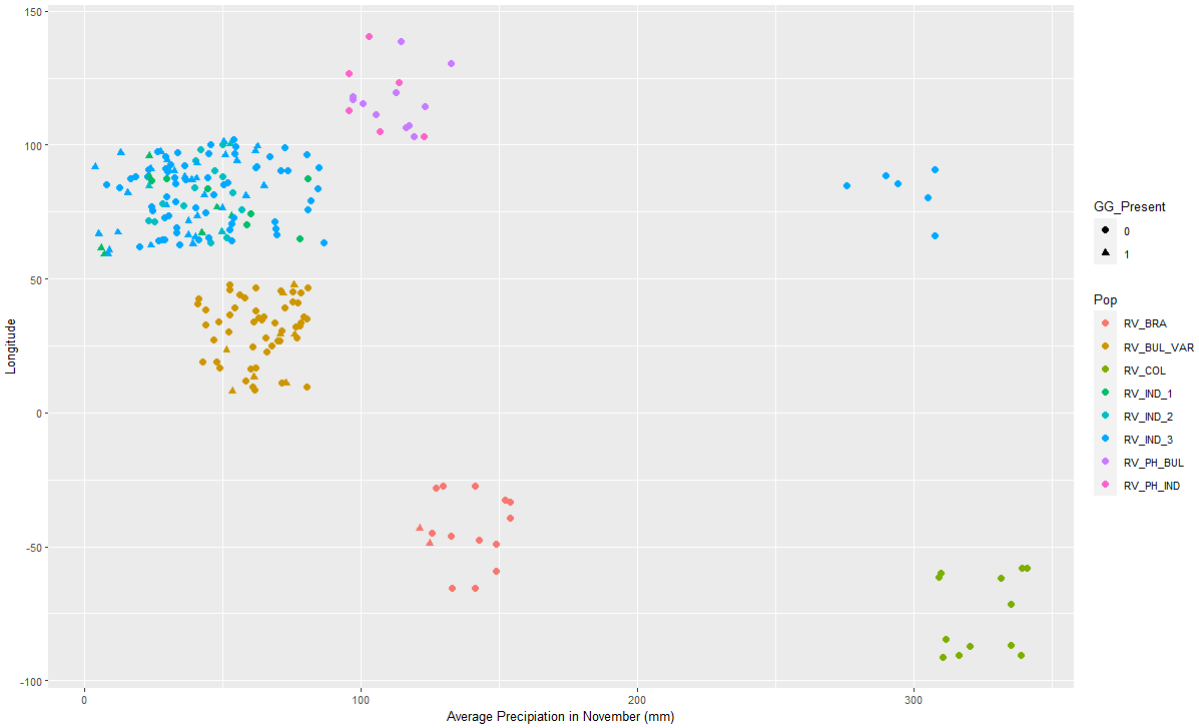


Figure 3.7: Plot showing presence/absence of GG genotype at SNP AX-85126469 in murrah buffalo that was associated with Average Precipitation in November (mm) plotted against longitude. GG genotype is frequent among Indian (RV_IND) populations. Populations displays as clusters using geom_jitter in GGPlot2 to reduce overlapping of points.

3.4.5. Signatures of Positive Selection

Regions under positive selection were identified in all populations. The distribution of these regions can be seen in Figure 3.8. The full list of regions under selection and the genes associated can be found in Table S3.15. RV_IND_3 featured the fewest number of SNPs at 17, whilst RV_IND_1 featured the most at 26 SNPs. Gene ontology analysis was carried out and enriched pathways can be found in Table S3.16.

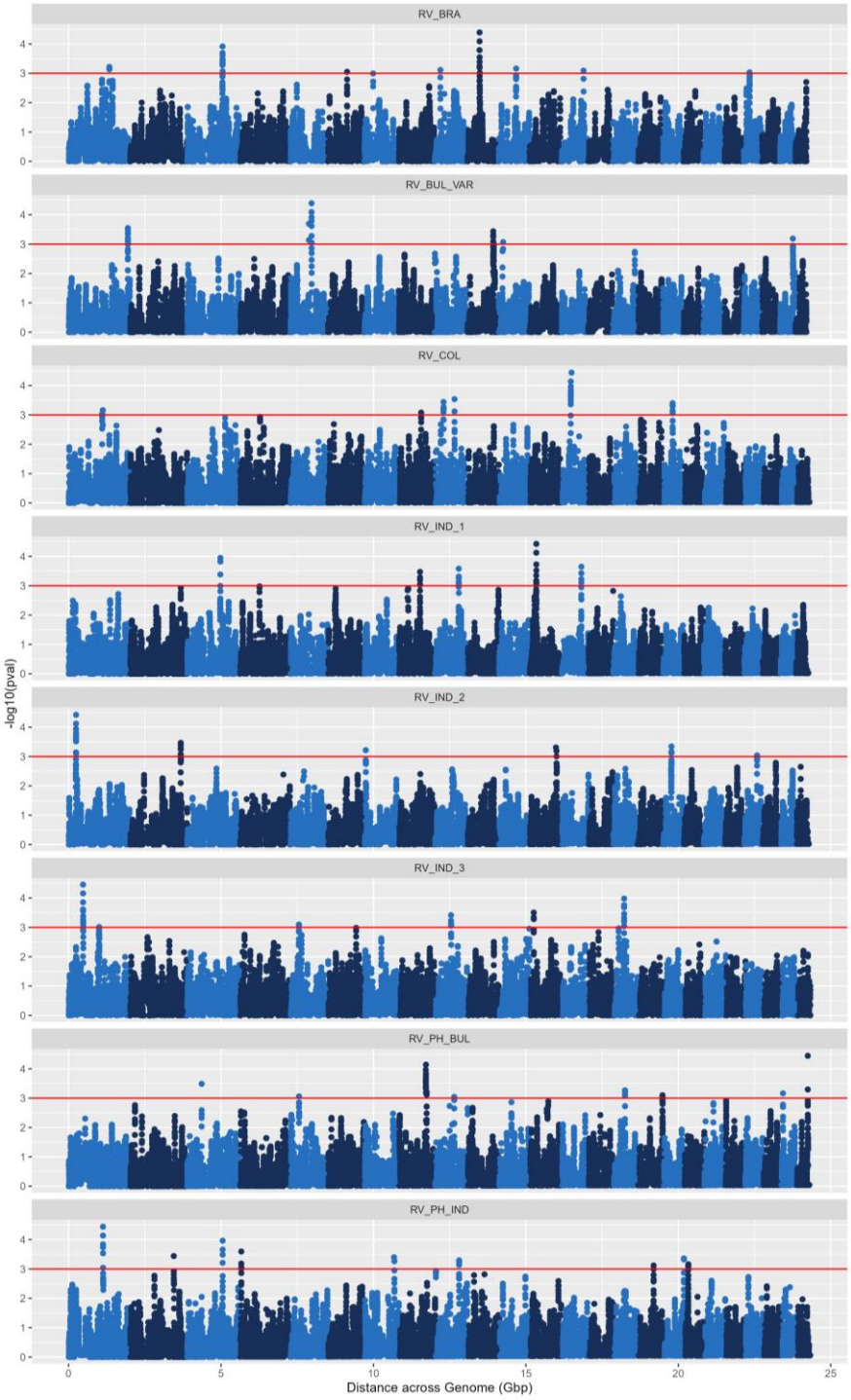


Figure 3.8: Significant SNPs under selection in each murrah population. Median was taken from XP-EHH scores computed across all pairwise combinations before p-values were generated for each SNP. The top 0.1% (red line) and SNPs with an XP-EHH score greater than 2 were taken as significant candidates of selection.

3.5. Discussion

Intense selection upon productive breeds often leads to a reduction in genetic diversity (FAO, 2007; Groeneveld *et al.*, 2010). Murrah buffalo are the most productive and popular buffalo breed (Kumar *et al.*, 2019; Zhang, Colli and Barker, 2020). Because of this, they have been transported across Asia, South America, and Europe to establish new populations or improve local buffaloes through crossbreeding. Having unlikely undergone intense selection for commercial purposes, murrah buffalo may still possess high levels of genetic diversity that may be associated with local adaptions (Groeneveld *et al.*, 2010; Naskar *et al.*, 2012; Hoffmann, 2013; Biscarini *et al.*, 2015; Velado-Alonso, Morales-Castilla and Gómez-Sal, 2020). This study analysed murrah buffalo from five countries to assess the levels of genetic diversity and identify any unique regions of genome that could suggest recent adaptations within the breed.

3.5.1. Genetic Variation and Population Structure

Genetic diversity was high across all populations with an average H_O ranging from 0.448 (RV_BRA) to 0.506 (RV_PH_IND). The Philippine populations of RV_PH_BUL and RV_PH_IND featured unusually high H_O, however this may be due to uneven sampling or small sample sizes as their H_E were the two lowest. Nevertheless, H_E was also high (0.412 – 0.462) across all populations. Murrah originate from the Haryana region of North-West India, nearby the putative domestication centre (Indus valley) for river buffalo (Satish Kumar *et al.*, 2007; Colli, Milanesi, Vajana, *et al.*, 2018). Genetic diversity is typically highest at the domestication centre, therefore the murrah buffalo here likely capture a large proportion of this original standing variation (Bruford, Bradley and Luikart, 2003; Gepts and Papa, 2003; Colli, Milanesi, Vajana, *et al.*, 2018). Genetic diversity in river buffalo has also been supplemented via gene flow with the wild Asian water buffalo (*Bubalus arnee*) since domestication (Satish Kumar *et al.*, 2007; Kandel *et al.*, 2019). Only Philippine populations significantly differed in H_O to other populations indicating that exported populations have generally captured enough genetic diversity from their stock population in India. In tandem with high levels of H_O, there was an absence of inbreeding in all populations.

Expectedly, there was low divergence between the populations with an average F_{ST} of approximately 2%. Divergence among river buffalo breeds is typically low (Kumar *et al.*, 2006;

Colli, Milanesi, Vajana, *et al.*, 2018; Thakor *et al.*, 2021). The greatest difference found between populations was between RV_BRA and RV_BUL_VAR at 4.9%. As both populations share no relation to each other since being established, this increase may be due to several reasons. For example, i) each population likely captured a marginally different subset of diversity from stock populations (Bruford, Bradley and Luikart, 2003), ii) RV_BUL_VAR features unrelated Mediterranean ancestry (Borghese, 2013a), iii) enacting evolutionary forces such as drift and selection (Willi *et al.*, 2007). This translates to the pattern revealed in the F_{ST} network as Indian populations sit at the centre whilst South American and Bulgarian murrah populations oppose each other. The ADMIXTURE results further explain this pattern (Figure 3.9). RV_BUL_VAR produces intrapopulation variation, showing ancestry from three genetic sources. This is likely splitting the individuals by the proportions of murrah and Mediterranean ancestry that features in the genome. Homogenous clustering would be expected over time as the key genetic diversity for the Bulgarian murrah breed is retained, and any unwanted murrah and Mediterranean ancestry is removed. The majority of RV_IND showed genetic similarity; however, two small clusters of individuals became distinguishable. Individuals within these clusters occurred at limited sites at the edge of the sampling range. These cluster may relate to underlying population structure, isolation of breeding individuals, or potential selection and breed development. South American and Philippine populations did not become distinguishable from the majority of Indian murrah buffalo highlighting low levels of divergence in line with F_{ST} results. Therefore, RV_PH_BUL is likely dominated by murrah ancestry as opposed to Mediterranean.

The absence of inbreeding and high levels of genetic diversity were reflected in the linkage disequilibrium results. All populations bar RV_PH_IND started with an r^2 of approximately 0.3, descending to <0.15 by 1Mb. RV_IND_3 features the lowest r^2 values starting at 0.240 and descending to 0.018 indicating the great variety of individuals within this population. This translates to a far greater effective population size for RV_IND_3 starting at 1,744 (5 generations ago). Interestingly, a peak is observed in RV_IND_3 at 32 generations ago reaching a maximum N_e of 3,619, indicating an increase in genetic diversity at this time before declining. Since, underlying population structure has been accounted for according to ADMIXTURE results here by way of separation of RV_IND, this expansion event is potentially due to an admixture event in the history of murrah buffalo. The expansion observed

corresponds to a time point around 1800CE (generation time = 6 years). This corresponds to a period of power transfer from the Mughal and Maratha Empires to colonisation by the British in the 19th century (Guha, 2019). This transfer of power coincided with a period of drought in India leading to the decline in agricultural output (Clingingsmith and Williamson, 2008). The drought led people to move from rural areas to urban defensive centres (Clingingsmith and Williamson, 2008). A lack of crops and labourers in rural areas, facilitates farmers selling livestock to raise cash for food (Clingingsmith and Williamson, 2008; Venot, Reddy and Umapathy, 2010). This theme still occurs today as livestock are used as a source of wealth and a buffer to crisis in India (Venot, Reddy and Umapathy, 2010). Perhaps admixture within murrah buffalo has been triggered by societal and environmental factors that led to the dispersal of murrah and subsequent contact with other breeds or populations.

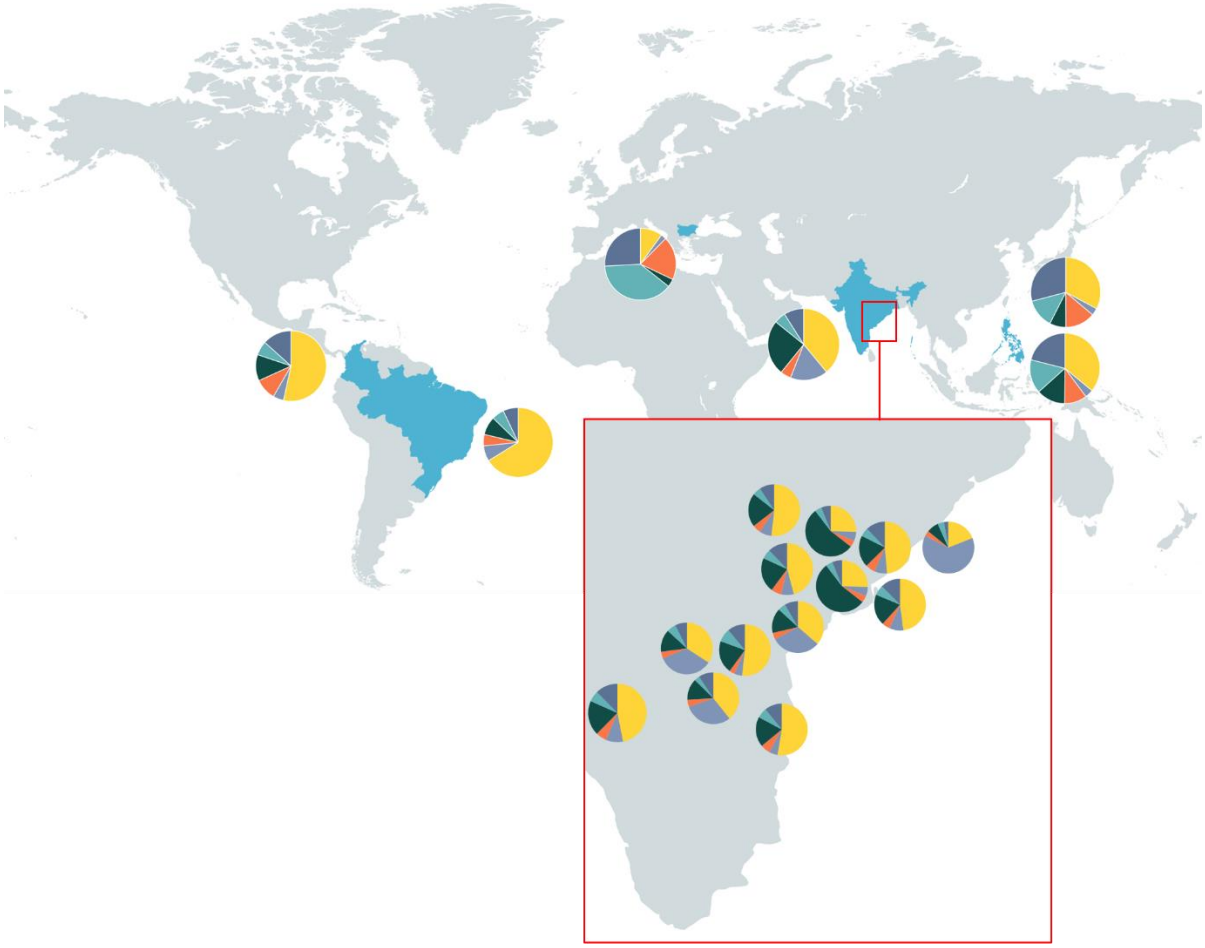


Figure 3.9: Map of average ancestry (calculated from ADMIXTURE results; K = 6) across murrah populations. Blue countries indicate countries possessing murrah derived populations used in this study. Charts are placed in their approximate geographic locations.

Murrah buffaloes are frequently crossed with local buffaloes across India, producing graded buffaloes. The purpose of these crosses is to upgrade yields of local buffaloes whilst retaining some local adaptations and disease resilience (Kumar *et al.*, 2017). From Table 3.1, there is a 1:3 proportion of pure to graded murrah. It's likely that the murrah sampled here have previously mixed with local buffaloes which would explain why the population expansion is only observed in RV_IND_3, and not all murrah populations here. Although RV_IND_3 appears homogenous, incorporation of variation from populations outside of original murrah lineages will inflate estimates of Ne (Orozco-terWengel, 2016). Following this logic, it may be expected to observe this signal in RV_BUL_VAR that is a known crossbreed. In this case, not enough time has passed to observe an expansion in N_e. Recent crossing between two populations results in long tracts of ancestry pertaining to each source (Harris and Nielsen, 2013; Loh *et al.*, 2013; Jin *et al.*, 2014; Korunes and Goldberg, 2021). Over time, recombination will break down and mixed these ancestral tracts and an expansion event will become visible (Harris and Nielsen, 2013; Loh *et al.*, 2013; Jin *et al.*, 2014; Korunes and Goldberg, 2021). Remaining populations meanwhile are probably founded from purebred murrah and therefore will not show an expansion event in N_e.

3.5.2. Signatures of Selection

India

River buffaloes have persisted in India for the past 6,000 years since domestication in the Indus valley region (Satish Kumar *et al.*, 2007). Murrah buffaloes rose to prominence due to their greater milk production capabilities, becoming the most popular buffalo breed (Kumar *et al.*, 2019). As such, murrah buffaloes are inherently adapted to being productive in tropical and sub-tropical climates (Yáñez *et al.*, 2020). In the Indian murrah (RV_IND) populations here, particularly that of RV_IND_3, low levels of linkage disequilibrium were observed along with high levels of heterozygosity translating into the absence of long homozygous tracts such as ROHs. These observations can be explained by frequent uptake of new genetic diversity through further breeding with wild buffaloes and crosses with local buffaloes across India. Adaptations of river buffalo to India are likely older, enabling recombination to break down any selective sweeps to shorter regions surrounding the beneficial locus (Stephan, 2019). RV_IND populations were among the shortest lengths for ROHs, and no unique ROHs were found in RV_IND_3 in comparison with other murrah

buffaloes. This is likely because RV_IND_3 shares genetic variation with all populations making it unlikely to harbour large unique genomic regions and is probably the closest population in this analysis to a de facto stock population of which other murrah populations are founded from. Furthermore, significant SNPs in the PCAdapt analysis appear driven by other isolated populations.

This trend continued in XP-EHH results as RV_IND_3 featured the fewest SNPs under selection. However, these regions found followed traits that may contribute to murrah's popularity. On BBU1 (47237918 – 47596180; 47715670 – 47750793), OLIG2, PAXBP1, SYNJ1, CFAP298, and EVAC1 are all involved in the development of the central nervous system (CNS), and this region has been associated with the polled horn phenotype in cattle and yaks (Meijer *et al.*, 2012; Liang *et al.*, 2016; Stafuzza *et al.*, 2018; Wang *et al.*, 2019; H. Jiang *et al.*, 2022). OLIG2 generates motor neurons and oligodendrocytes that form myelin sheaths, increase neuronal impulse speeds through saltatory conduction of action potential (Meijer *et al.*, 2012; Komatsu *et al.*, 2020). OLIG2 is involved in horn ontogenesis of taurine cattle, and horn bud differentiation and frontal skin in cattle (Allais-Bonnet *et al.*, 2013; Wiedemar *et al.*, 2014; Stafuzza *et al.*, 2018). PAXBP1 (PAX3 and PAX7 Binding Protein 1) is expressed in the cerebellar hemisphere and cerebellum and muscle precursor cells and potentially has a role in facial bone development (Mohammadparast *et al.*, 2014; Alharby *et al.*, 2017; Aldersey *et al.*, 2020; Zhou *et al.*, 2021). SYNJ1 (Synaptojanin 1) meanwhile plays a role in phosphorylation and recycling of synaptic vesicles (Al Zaabi, Al Menhali and Al-Jasmi, 2018). CFAP298 (Cilia and Flagella Associated Protein 298) controls spinal cord formation and functionality of cilia in central canal in zebrafish (Bearce *et al.*, 2022). These cilia ensure circulation of cerebrospinal fluid in the lumen of brain and spinal cord cavities, and mutations result in scoliosis-like deformities of the spine in zebrafish (Marie-Hardy *et al.*, 2021). EVAC1 (Eva-1 Homolog C) is involved with neuron survival and growth and development of olfactory and optic sensory axons and neural structures (James *et al.*, 2013; Casto-Rebollo *et al.*, 2020). The polled phenotype is not present in buffaloes unlike cattle and yak, however murrah feature altered horn morphology with small, curled horns. Selection on horn size and the central nervous system typically occurs in livestock to increase docility and reduce risk of harm to humans (Haskell, Simm and Turner, 2014; Norris *et al.,* 2014; Simon, Drögemüller and Lühken, 2022;

Algra *et al.*, 2023). This region may contribute to the ease of handling of murrah resulting in their popularity across India.

A host of genes found under selection in RV_IND_3 appeared to be involved in metabolic activities which may relate to murrah buffaloes' ability to remain productive under stressful conditions. GRPEL2 (GrpE Like 2, Mitochondrial; BBU9) regulates the mitochondrial import process to maintain mitochondrial homeostasis (Yang *et al.*, 2023). The encoded protein acts as a sensor of oxidative stress and interacts with mtHSP70 in states of high oxidative stress to prevent misfolding of mitochondrial imported proteins (Srivastava *et al.*, 2017; Konovalova *et al.*, 2018; Heras-Molina *et al.*, 2022). Increased expression has been found coinciding with times of greater lipid oxidation in pigs (Heras-Molina *et al.*, 2022). AUP1 (AUP1 Lipid Droplet Regulating VLDL Assembly Factor) meanwhile regulates Rtg3 required for stationary phase in mitophagy and may function to transduce a signal from specific hyperoxidized proteins to activate the retrograde response that impact metabolic states (Journo, Mor and Abeliovich, 2009; Chen *et al.*, 2022). MRPL53 (Mitochondrial Ribosomal Protein L53) encodes a large subunit of mitochondrial ribosome 39S and part of the machinery for the mitochondrial oxidative phosphorylation system whilst INO80B (INO80 Complex Subunit B) is a DNA damage repair gene (Serber *et al.*, 2016; Li *et al.*, 2022). DCTN1 (Dynactin Subunit 1) is an axonal transport gene for maintaining motor neurons and has been associated with milk yield in murrah buffalo further highlighting the importance of this region in the breed (Kumar *et al.*, 2023). Increased stress on livestock reduces productivity, and for dairy buffalo, a reduction in milk yields (De Rosa *et al.*, 2009; Caroprese *et al.*, 2010; Pawar, 2012; de la Cruz-Cruz *et al.*, 2019; Saqib *et al.*, 2022). Oxidative stress occurs when reactive oxygen species are not effectively removed following metabolic processes, causing cellular damage, and activating apoptotic pathways (Puppel, Kapusta and Kuczyńska, 2015). This stress can be exacerbated through environmental conditions such as heat or poor feed (Sordillo and Aitken, 2009; Pedernera *et al.*, 2010; Z. Guo *et al.*, 2021). Genes such as these may underline murrah buffaloes' ability to remain productive in low quality environments.

The genetic separation of RV_IND_1 and RV_IND_2 may indicate murrah buffalo lineages and individuals that are being selected for and developed. Unlike RV_IND_3, these two populations did feature unique ROHs. For RV_IND_1, genes found within divergent ROHs surrounded immunity and production. ALCAM (Activated Leukocyte Cell Adhesion Molecule)

and CBLB (Cbl Proto-Oncogene B) are markers for mastitis and are important for cell adhesion and signalling in T-cells respectively (Lkhagvadorj *et al.*, 2010; Ju *et al.*, 2018; S. Han *et al.*, 2020; Dysin, Barkova and Pozovnikova, 2021). CD47 (CD47 Molecule) is an immune response gene that becomes upregulated in lactating mammary gland (Suchyta *et al.*, 2003). TNFRSF13C (TNF Receptor Superfamily Member 13C) encodes a receptor for B cell activating factor (Demasius *et al.*, 2013). Increased BAFF expression was found in infected dairy goats spleen (Fu *et al.*, 2022). Nearby, SHISA8 (Shisa Family Member 8) is associated with somatic cell score in Brazilian murrah buffalo (Lázaro *et al.*, 2021). In the same region, GRAP2 (GRB2 Related Adaptor Protein 2) is upregulated in buffalo milk somatic cells and functions in the T cell receptor signalling pathway while immune-related gene MTMR11 (Myotubularin Related Protein 11) has additionally been observed under selection in river buffaloes (Ahlawat *et al.*, 2021; Ravi Kumar *et al.*, 2023). The ROH on BBU6 features TENT5C, VTCN1, TRIM45, CD101, and CD2, all of which have immune functions (Brym *et al.*, 2016; Dai *et al.*, 2019; Simons *et al.*, 2019; Liudkovska *et al.*, 2022; Etchevers *et al.*, 2023). Many genes across the ROHs were linked to milk production in livestock. SREBF2 (Sterol Regulatory Element Binding Transcription Factor 2) regulates expression of lipogenic genes and is involved in milk fat synthesis in buffaloes (Piantoni *et al.*, 2010; Ye *et al.*, 2022). Further lipid metabolic genes for dairy cattle include POLR3H, EP300, ANXA9, CERS2 (Kulig *et al.*, 2010, 2013; Martínez-Royo *et al.*, 2010; Romao *et al.*, 2014; Puig-Oliveras *et al.*, 2016; Rico *et al.*, 2016; Palombo *et al.*, 2018; McFadden and Rico, 2019; Hu *et al.*, 2021; Pecka-Kiełb *et al.*, 2021). CTSK on BBU6 has been found under increased expression in mammary glands during lactation in buffaloes (Sodhi *et al.*, 2023).

The balance between immunity and milk production is important for economic output as both processes are energetically expensive to dairy livestock. High yielding dairy cattle have been shown to feature defective inflammatory responses and reduced immunity owing to intense metabolic changes around calving (Bronzo *et al.*, 2020). Infection in cows can lead to great economic loss through reduced milk yields, veterinary costs, and loss of livestock (Bronzo *et al.*, 2020). Milk yields are reduced as greater energy and nutrient demands are placed on producing immunity-related cells (Rauw, 2012). The calorific requirements are increased in infected cows which are met by utilising energy stores such as fat reserves (Rauw, 2012). Additionally, integrity of the blood-milk barrier is impaired, preventing the required

nutrients being transferred to milk secreting cells (Wellnitz and Bruckmaier, 2021). Effective immune systems leads to lower somatic cell counts in milk, meaning higher quality and longer lasting products (Alhussien and Dang, 2018). Selection of immunity and milk production genes may explain why this genetic cluster is found across several locations as farmers may be identifying the most productive and resilient murrah buffaloes. Further evidence of these results occurs in XP-EHH regions in the form of PTAR1 (virus resistance), MAMDC2 (Antiviral response in microglia), MORN2 (facilitates phagocytosis), RFXAP (regulates MHC II genes), and SKAP2 (modulates immunity) (Wu *et al.*, 2008; Butte *et al.*, 2012; Abnave *et al.*, 2014; Cho *et al.*, 2014; Hanna and Etzioni, 2014; Riblett *et al.*, 2016; Tumasian *et al.*, 2021; Yiliang Wang *et al.*, 2022). These immunity genes can be found alongside production associated genes such as ALG5. This gene is differentially expressed at different stages of lactation and participates in lipid biosynthesis and glycerolipid metabolic processes (Gebreyesus *et al.*, 2019).

RV_IND_2 has two ROHs that overlapped with RV_COL and RV_BRA populations, while additional ROHs further highlight reproductive functions. On BBU14, JAG1 (Jagged Canonical Notch Ligand 1) operates via the Notch pathway, a conserved pathway crucial to embryonic development. This gene has shown signalling function in embryogenesis and has been associated to reproductive function in Nellore cattle, and promotion of mammary cell proliferation during early lactation in cows (Tripurani *et al.*, 2011; Wang *et al.*, 2012; Sbardella *et al.*, 2021). Genes on BBU17 include ZNF84 (Zinc Finger Protein 84) that is a key regulator in ovaries of Brahman cattle with expression changes in uterine tissues associated with puberty (Nguyen *et al.*, 2017; Fortes *et al.*, 2018). Associations with puberty have also been observed in the adjacent ZNF605 (Zinc Finger Protein 605) (Fortes *et al.*, 2018). Furthermore, GUCY1A1 is associated with litter size in sheep and ASIC5 variants linked to pregnancy loss (Al Qahtani *et al.*, 2021). ASIC5 has previously been found under selection in river buffalo (Saravanan, Rajawat, *et al.*, 2022). Further reproductive genes were found in XP-EHH results as TUSC3 (Tumour Suppressor Candidate 3) is expressed in spermatocytes, inducing sperm differentiation and maturation (Zhou and Clapham, 2009; Yu *et al.*, 2017; Sun *et al.*, 2023). GRHL1 (Grainyhead Like Transcription Factor 1) functions in placenta development and has been associated with litter traits in pigs while the absence of TAF1B (TATA-box binding protein associated factor, RNA polymerase I Subunit B) causes the accumulation of late stage egg chambers in ovaries (Ding *et al.*, 2021; S.-Y. Chen *et al.*, 2022; X. Wang *et al.*, 2022). Analysis

of mRNA expression of this gene corresponds with success of AI-pregnant heifers (Dickinson *et al.*, 2018; Moorey and Biase, 2020). TAF1B also is associated with body weight of yaks at the time weaning (Jia *et al.*, 2020). Similarly, TDO2 (Tryptophan 2,3-Dioxygenase) is also linked with weaning weight in Charolais cattle (Garza-Brenner *et al.*, 2020). These genes may indicate potential development of murrah buffaloes in RV_IND_2 for increased reproductive and infantile developmental success.

India has been upgrading buffaloes through national programmes to increase milk yields and avoid unproductive buffaloes via genetic erosion. Uninformed admixture (or outbreeding) is frequent in rural areas as buffaloes are grazed in common lands allowing free admixture of breeds. Uninformed admixture results in reduced productivity through erosion of beneficial haplotypes maintained by breed selection. In a bid to reduce the impact of genetic erosion (and improve milk yields via more efficient selection) of buffaloes in India, the government began implementing ART schemes since the 1950s (Singh and Balhara, 2016). The first strategic artificial insemination programme (1951 – 1956) targeted 150 key villages to improve cattle and buffaloes (Singh and Balhara, 2016). The current National Dairy Plan now incorporates 14 major milk producing states accounting for more than 80% of the country's cattle and buffalo population, and greater than 90% of the country's milk production (Singh and Balhara, 2016). The scheme uses a strict number of breeding buffalo bulls produced through progeny testing and pedigree selection programmes (Singh and Balhara, 2016). Through modernisation and greater use of ART in India, highly productive buffaloes can be developed more easily via selection of elite individuals. The two unique clusters within India murrah buffaloes may be evidence of lineages being selected for improved milk production and reproduction success.

South America

Numerous livestock species have been imported into the Americas and have successfully adapted to its climate. The evolution of Creole cattle is an example of this. Iberian cattle were imported to central America late 15th century, and these adapted to various ecotypes forming the creole breeds (Pitt *et al.*, 2019). Adaptations of creole cattle include the thermoregulatory slick hair phenotype (Pitt *et al.*, 2019). Buffalo arrived in South America later, being imported to Brazil from the 1890s onwards (da Silva *et al.*, 2021). The majority of

these were either Mediterranean buffalo from Italy or murrah of Indian origin. Swamp buffaloes have also been imported to a lesser extent. Unlike previous livestock species, buffaloes were imported for their natural tropical adaptations. Adaptations to hot climates, pathogenic resistance, and productivity in harsh conditions are ideal for use in South America. Buffaloes were therefore imported to replace locally adapted but low productive breeds such as creole cattle (da Silva *et al.*, 2021). Although both RV_BRA and RV_COL are genetically similar to their Indian ancestors, regions under selections suggest that some recent adaptation to South America.

Divergent ROHs in RV_BRA were found associated with neural development, energy metabolism, reproduction, milk production, and immune response. The ROH on BBU1 overlapped with XP-EHH results in RV_IND_3 including genes OLIG2, PAXBP1, SYNJ1, CFAP298, and EVAC1. Additional genes captured in the longer ROH include URB1 (URB1 Ribosome Biogenesis Homolog), MRAP (Melanocortin 2 Receptor Accessory Protein), MIS18A (MIS18 Kinetochore Protein A), HUNK (Hormonally Up-Regulated Neu-Associated Kinase), and SCAF4 (SR-Related CTD Associated Factor 4) all of which are expressed in the brain and CNS. MRAP and SCAF4 have been shown to be involved with regulation of metabolism, appetite, and food intake (Webb and Clark, 2010; Zhang *et al.*, 2021; Novo *et al.*, 2022). SOD1 (Superoxide Dismutase 1) and TIAM1 (TIAM Rac1 Associated GEF 1) were also found in this region and both genes have been linked with heat stress (Flori *et al.*, 2012; Lopreiato *et al.*, 2020; Khan *et al.*, 2021; Zeng *et al.*, 2022). Therefore, this cluster of genes under selection may allude to behavioural adaptations in murrah buffalo through improved management of buffaloes. In this case, for example, stress may be reduced through increased docility (Canario *et al.*, 2013; Friedrich, Brand and Schwerin, 2015). Lazarov *et al.,* (2021) found SNPs surrounding URB1 and MIS18A were linked with somatic cell score in Brazilian buffaloes. Reduced stress through docility and thermotolerance would reduce the release of stress hormones leading to a reduction in inflammatory signalling and immune response therefore reducing number of white blood cells in milk (Rauw, 2012; Alhussien and Dang, 2018; Matera *et al.*, 2022).

BBU9 revealed a range of genes associated with thermoregulation and reproduction. DELE1 (DAP3 Binding Cell Death Enhancer 1) responds to mitochondrial stress and is able upregulate heat shock proteins (Fessler *et al.*, 2020; Guo *et al.*, 2020; Girardin *et al.*, 2021;

Huynh *et al.*, 2023). TAF7 (TAT-Box Binding Protein Associated Factor 7) can regulate heat shock proteins and enhance efficient recovery of cells to thermal stress (Nagashimada *et al.*, 2018; Y. Liu *et al.*, 2022). Expression of CD14 (CD14 Molecule) has been linked with tropical thermal stress (Alhussien *et al.*, 2018). TMEM173 (Also known as STING1) has been found under selection in Fuzhong swamp buffalo, a place that faces hot and humid summers (Sun, Huang, *et al.*, 2020). Selection of this gene has additionally been found in Southern Chinese cattle (Y. Liu *et al.*, 2022). ECSCR (Endothelial Cell Surface Expressed Chemotaxis and Apoptosis Regulator) encodes a protein primarily found in endothelial cells and blood vessels, and therefore may have a role in tolerance to heat stress in cattle (Peripolli *et al.*, 2018). Reproductive genes include GNPDA1 (Glucosamine-6-Phosphate Deaminase 1) which is upregulated in pLH blastocysts under ART (Dyck *et al.*, 2014). HDAC3 (Histone Deacetylase 3) has been shown related to competence of bovine oocytes, and higher transcripts in cows suggest importance for preparation of oocytes (Silva *et al.*, 2019; Kawamoto *et al.*, 2022). Furthermore, expression of this gene is linked with improvement of bovine embryo production and preimplantation development (Silva *et al.*, 2019). HB-EGF (Heparin Binding EGF Like Growth Factor) has shown evidence for pregnancy success (Sá Filho *et al.*, 2017; Dolebo *et al.*, 2019). Other genes such as EGR1 and KDM3B have been linked to fertility (Mota *et al.*, 2017; Dolebo *et al.*, 2019; Kang *et al.*, 2022). Brazil are among the world's largest users of ART to support their meat and dairy production that ranks 2nd and 5th in the world, respectively (Sartori *et al.*, 2016). Year on year increases in artificial inseminations, embryonic transfers, and somatic cell nuclear transfers have been observed (Sartori *et al.*, 2016). Therefore, it is of no surprise that genes relating to success of pregnancy are observed under selection. The additional occurrence of thermotolerance genes in this region may help mitigate additional stress on cows that can occur from increase metabolic rates during energetically expensive processes such as reproduction (Cartwright *et al.*, 2023). This ROH was additionally found in RV_IND_2 and RV_PH_BUL, likely being used in a similar context, particularly that of RV_PH_BUL which also resides in hot and humid conditions. On BBU12, EIF2AK3 and FABP1 may have roles in energy balance while FOXI3 regulates hair follicle development (McCarthy *et al.*, 2010; Shahzad *et al.*, 2015; Guan *et al.*, 2016; Diniz *et al.*, 2020; Bolormaa *et al.*, 2021; García-Roche *et al.*, 2021).

Regions under recent selection in RV_BRA from XP-EHH results may further add to these themes. VPS13A (Vacuolar Protein Sorting 13 Homolog A) encodes the protein chorein. This protein is chief moderator of secretion and density of blood platelets. Heat stress raises platelet numbers of blood density which raises danger of cerebral and coronary thrombosis. Therefore, this gene may act to mitigate risk of damage from increased heat (Amiri Ghanatsaman *et al.*, 2023; Zhong *et al.*, 2023). CSNK1G1 (Casein Kinase 1 Gamma 1) is associated with cattle temperament whilst GRM7 (Glutamate Metabotropic Receptor 7) operates in excitatory glutamatergic synapses and is associated with behavioural neurological disorders such as autism and ADHD (Alvarenga *et al.*, 2023). PCAdapt revealed a SNP associated with RV_BRA. NCAM2 (Neural Cell Adhesion Molecule 2) was the closest gene and has previously been associated with dairy cattle and goats (Ding *et al.*, 2022; Amiri Ghanatsaman *et al.*, 2023). This gene has also been found associated with milk, fat, and protein yield in Brazilian buffaloes indicating that selection on livestock traits is affecting genetic structuring of murrah populations (Venturini *et al.*, 2014).

In the Colombian population, the few genes found within ROHs appeared to revolve mostly around growth. On BBU2, genes were found associated with body and carcass traits relating to fat (LRRC1, HCRTR2, HMGCLL1, BMP5), and muscle (MLIP, FAM83B, COL21A1) functions (Ahmady *et al.*, 2011; Shao *et al.*, 2011; Tetens *et al.*, 2015; Falker-Gieske *et al.*, 2019; Ghosh *et al.*, 2019; Tang *et al.*, 2019; Yuan *et al.*, 2019; Chen *et al.*, 2020; Sigdel *et al.*, 2020; Park *et al.*, 2021; Tumasian *et al.*, 2021; Ramos *et al.*, 2023). Meanwhile CACYBP (Calcyclin Binding Protein) on BBU5 relates to skeletal muscle function in Holstein cattle (Yougbaré *et al.*, 2021). Although not defined as a breed in Colli *et al.*, (2018), RV_COL is genetically very murrah-like. Water buffalo were originally imported to Colombia from the 1970s (Zava, 2009). Buffaloes were obtained from nearby countries such as Trinidad & Tobago, before defined breeds such as murrah were imported from neighbouring Brazil (Zava, 2009). Here in the Caribbean, river buffalo of Indian origin (e.g., murrah and Jafferabadi) were developed for draft power and greater meat production leading to buffaloes with increased muscle mass, coined the buffalypso (Bennett, Garcia & Lampkin, 2007). N_e results in section 3.4.2 show elevated N_e for RV_COL, though not a distinctive population expansion like RV_IND_3, which could be explain by breeding between closely related gene pools, e.g.,

closely related Indian breeds (Thakor *et al.*, 2021). Therefore, these genes relating to growth functions may be traces of alternative use from dairy to draught in river buffaloes.

Colombian buffalo revealed little information on potential adaptations to South America in lieu of XP-EHH results. Of the genes found under selection, STXBP5L (Syntaxin Binding Protein 5L) is associated with skin photoaging, while its paralog STXB5 promotes platelet secretion (Le Clerc *et al.*, 2013; Gorski *et al.*, 2019). Adjacent to this is POLQ (DNA Polymerase Theta), which participates in DNA repair. Loss of POLQ in humans and mouse cells causes sensitivity to ionizing radiation (Yousefzadeh and Wood, 2013). GSDME (Gasdermin E) is key for induction of pyroptosis, a form of cell death regulated via inflammation (Hikima and Morimoto, 2023). On BBU9, EDIL3 (EGF Like Repeats And Discoidin Domain 3) mediates angiogenesis and participates in inflammatory pathways (Gasca *et al.*, 2020; Becker *et al.*, 2022; Sokolova *et al.*, 2023). GLRB (Glycine Receptor Beta) and GRIA2 (Glutamate Ionotropic Receptor AMPA Type Subunit 2) are neuronal genes both of which are associated with flight speed and temperament in cattle, respectively (Santos *et al.*, 2017; Ruiz-De-La-Cruz *et al.*, 2023). These genes indicate increased exposure to solar radiation. Several buffalo farms can now be found in states (e.g., Antioquia & Santander) at higher altitudes, where cooler temperatures mean that buffaloes may not need to wallow, and as such, are more exposed to solar radiation. Alternatively, in tandem with neuronal genes and growth traits from ROHs, these genes may further contribute to adaptation to draught usage.

Bulgaria

Buffalo in Europe almost entirely comprise of the Mediterranean breed that adapted to Europe over the past 1,500 years (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). Although Italy has bred Mediterranean buffaloes to be among the most productive breeds of river buffalo, those in Eastern Europe lacked significant improvement in milk production. Farmers in Bulgaria proceeded to import murrah buffaloes to improve milk production and increase disease resistance within the breed (Borghese, 2013a). The result of this crossbreed formation has produced a higher frequency of longer ROHs compared to other murrah populations. This may be due to regions of either Mediterranean or murrah buffalo being selected for, or not enough time has passed to break down these regions from recombination. Most significant SNPs identified from PCAdapt analysis likely relate to

Mediterranean ancestry as PC1 distinguishes between pure murrah and the crossbred murrah. Several SNPs relate to genes associated to growth traits. Local Bulgarian buffaloes prior to crossbreeding were additionally used for draught and meat purposes and are among the fastest growing and largest of riverine buffaloes (Antonio, 2010; Borghese, 2013a; Noce *et al.*, 2021). RAD51 (RAD51 Recombinase) and ZFYVE26 (Zinc Finger FYVE-Type Containing 26) have been associated with carcass fatness in Nellore cattle and residual feed intake in pigs, respectively (Do *et al.*, 2014; Martins *et al.*, 2021). PRTG is associated with back fatness in Nellore cattle and average daily gain in Italian beef cattle (Júnior *et al.*, 2016; Mancin *et al.*, 2022). NEDD4 has been found associated with meat traits in Nellore cattle and sheep, and is involved in promotion of bone development, cellular growth (Barnes *et al.*, 2019; Bakoev *et al.*, 2020; Frezarim *et al.*, 2022; Krivoruchko *et al.*, 2023). Other genes found in this analysis include TSPAN18, CD82, ANKRD33B all may play roles in inflammation and immune response, while CPLX2 is a encodes a presynaptic protein and this gene is found under selection Bashan cattle, known for their gentle temperament (Sun *et al.*, 2023).

Genes associated with growth traits and immunity are further observed in RV_BUL_VAR within divergent ROHs and under selection in XP-EHH results. Within the ROH on BBU4, RBFOX2 (RNA Binding Fox-1 Homolog 2) regulates MEF2D-V4 which promotes differentiation of chicken myoblasts, while knockouts in mice disrupt SLC25A4 involved in energy production, leading to heart and skeletal muscle defects (Ouyang *et al.*, 2020; Wierzbicka *et al.*, 2023). This gene has been under selection in Boer goats that excel on growth performance over other breeds (Yuan *et al.*, 2022). Interestingly, on BBU14, a QTL for coat colour pigmentation in cattle is found. This corresponds with the gene ATRN (Attractin) that has been shown to influence coat colour in mice, and linked to pigment switching in Holstein and Hanwoo cattle (Seo *et al.*, 2007; Pausch *et al.*, 2012). River buffaloes typically feature near black coat phenotypes, however brown coats have previously been developed in Bulgarian buffaloes (Borghese, 2013b). ATRN is a receptor for the ASIP protein that is well known for coat colour phenotypes in animals (Y. Liu *et al.*, 2018). Also, ATRN can promote degradation of melanocortin receptors which influences intracellular cAMP levels. Increased cAMP is able to influence lipolysis in adipose tissue and energy metabolism in liver and muscle (Y. Liu *et al.*, 2018). Therefore, development of the brown coat phenotype may be potentially linked to selection of any growth or energy related traits.

The remaining genes observed were frequently associations with the following: SIGLEC1, MAVS, DMBT1, CPXM2, ABRAXAS2 with immunity; HSPA12B, PANK2, PCNA, SLC23A2 for thermotolerance; and BMP2, PROKR2, CDS2, PRDX3, GRK5 BAG3, ACADSB, CTBP2, ADAM12 associated with various lipid and metabolic functions. Notably, CRLS1 (Cardiolipin Synthase 1) is involved in mitochondrial membrane function and is predominantly expressed by tissues with high levels of energy metabolism and has been linked with cold adaptation (Trovato *et al.*, 2015; You *et al.*, 2020). The SNP under selection at 43545229 on BBU23 overlapped with ROH results for RV_BUL_VAR. Within this, we find NKX1-2 (NK1 Homeobox 2) that promotes adipogenesis and inhibits osteoblastogenic differentiation, and differential methylation of this gene can be observed at lower temperatures (0.8 – 13C) in humans (Chen *et al.*, 2019; Xu *et al.*, 2020; Wang *et al.*, 2023). Adjacent to this, LHPP (Phospholysine Phosphohistidine Inorganic Pyrophosphate Phosphatase) catalyses hydrolysis of imidodiphosphate which is essential for maintaining mitochondrial membrane potential (Conte *et al.*, 2021). This gene is expressed in mammary glands of Holstein cows and ewes where metabolic changes are highly regulated at a transcriptional level (Mach *et al.*, 2012; Conte *et al.*, 2021). Genes under selection in Bulgarian murrah buffaloes appear related to energy homeostasis. Whether this is due to adaptation to a colder environment, greater milk production from crossbreeding Mediterranean and murrah buffaloes, or additional selection on growth traits can be further investigated.

Philippines

The Philippines is within the historic range of the swamp buffalo having been introduced to the islands post domestication 4,000 years ago (Barker, 2014; Colli, Milanesi, Vajana, *et al.*, 2018). With the decline in swamp buffalo population owing to mechanization of farming and low production (milk and meat) capabilities, the Philippines have established river buffalo populations to hybridise with swamp buffaloes. Due to increasing frequency of typhoons, flooding, and droughts ruining crop production, smallholders in Philippines are increasingly swapping crop production to livestock (Escarcha *et al.*, 2020). With the help of CDP, river buffalo are used to increase growth speeds (70 – 100% increase) and milk production (200 – 300% increase) of local swamp buffaloes to increase income for families across Philippines (Cruz, 2015). Murrah buffaloes from India and Bulgaria are the main source of river buffalo genetic variation used for crossbreeding. Although one Philippine murrah

(RV_PH_IND) populations is of small sample size, the Philippine Bulgarian murrah population still revealed regions under selection. Across ROH on BBU8 in RV_PH_BUL, several genes relate to growth traits. PPP1R3A (Protein Phosphatase 1 Regulatory Subunit 3A) encodes a subunit of PP1 involved in insulin signalling pathway promoting glycogen synthesis and has shown to be downregulated in cattle with restricted feed (Keogh *et al.*, 2015). LSMEM1 (Leucine Rich Single-Pass Membrane Protein 1), IFRD1 (Interferon Related Developmental Regulator 1), ZNF277 (Zinc Finger Protein 277), DOCK4 (Dedicator of Cytokinesis 4), LRRN3 (Leucine Rich Repeat Neuronal 3) all have roles in cellular growth, proliferation, differentiation, and apoptosis (Cheng *et al.*, 2014; Wang, Chen and Wang, 2019; Y. Song *et al.*, 2022; Téteau *et al.*, 2023). IFRD1 in particular has been associated with average daily gain in Nellore and Marchigiana cattle whilst expression differed in rapid growth Dorset sheep compared to slow growth small tail han sheep (Miao, Luo and Qin, 2015; Sorbolini *et al.*, 2017; Peripolli *et al.*, 2018). This gene may also potentially be involved in the double muscle phenotype in Charolais cattle (Jahuey-Martínez *et al.*, 2019). Therefore, this region could be a useful candidate region for increased growth of buffalo in the Philippines. The ROH on BBU20 meanwhile provided MPHOSPH10 (M-Phase Phosphoprotein 10) and MCEE (Methymalonyl-CoA Epimerase), both of which have been attributed to adipose functions in cows and pigs (Moisá *et al.*, 2017; Y. Jiang *et al.*, 2022).

Again, a gene relating to coat colour was found under selection. This time, in RV_PH_BUL, FOXP2 (Forkhead Box P2) was found within the ROH on BBU8. FOXP2 is a well-known language and speech related gene in humans that is integral to brain development (Dediu and Christiansen, 2016). However, this gene has additionally been linked to coat colour in Vrindavani cattle (Chhotaray *et al.*, 2021). TRPM1 (Transient Receptor Potential Cation Channel Subfamily M Member 1) was additionally found, this time in the ROH on BBU20. TRPM1 is a calcium channel where knockdowns in the gene causes an influx of calcium ions into melanocytes (Devi *et al.*, 2013; Shenyuan Wang *et al.*, 2021). Signalling through Ca^{2+}, such as level of exposure to solar UV radiation, controls cellular melanogenesis (Bellono and Oancea, 2014; Dumbuya, Hafez and Oancea, 2020). Variants in TRPM1 are responsible for spotting phenotype in horses while high expression in goats corresponded with brown skin (Peng *et al.*, 2017; Derks and Steensma, 2021). Along with the gene in RV_BUL_VAR, these

genes can be explored in the future to determine if they feature any further functionality or purely visual.

XP-EHH results in RV_PH_BUL revealed further metabolic and growth-related genes. SLC13A1 (Solute Carrier Family 13 Member 1) that encodes a renal sodium/sulphate cotransporter and is involved in sulphate homeostasis (Boegheim *et al.*, 2017). Sulphate is essential for several processes such as cellular metabolism and growth. SLC13A1 has been associated with bone disorders in Simmental cattle, and sulphated proteoglycans are essential for maintaining normal structure during bone and cartilage formation (Bordbar *et al.*, 2019). Within the same region, IQUB (IQ Motif and Ubiquitin Domain) was found and knockdowns of this gene can prevent c-myc expression, a muscle regulator gene (Li *et al.*, 2018; Bordbar *et al.*, 2019). NDUFA5 (NADH: Ubiquinone Oxidoreductase Subunit A5) participates in transportation of electrons in mitochondria and therefore energy production while GRIA4 (Glutamate Ionotropic Receptor AMPA Typer Subunit 4) may additionally be involved in metabolic homeostasis as this gene has been linked with cold adaptation in Siberian cattle (Bordbar *et al.*, 2019; Igoshin *et al.*, 2019, 2021; Swartz *et al.*, 2021). ASB15, LMOD2, and WASL all have roles in muscle development (McDaneld, Hancock and Moody, 2004; Yu *et al.*, 2007; McDaneld and Spurlock, 2008; Yamashiro *et al.*, 2012; Bordbar *et al.*, 2019; Lange *et al.*, 2021). RV_PH_IND revealed various genes related to livestock productivity however with no clear pattern.

Environmental Adaptations

Selection in livestock is largely dictated by pressures on production traits. Declines in productivity are realised when livestock encounter environments that they are not adapted to as energy is required to maintain bodily functions under stressful conditions (Niyas *et al.*, 2015; Bernabucci, 2019; Passamonti *et al.*, 2021). In response, individuals with beneficial alleles may display greater productivity due to being more resilient and better adapted, i.e. a higher fitness (Kawecki and Ebert, 2004; Savolainen, Lascoux and Merilä, 2013). Over time, this will lead to a population or breed becoming locally adapted to the new environment (Passamonti *et al.*, 2021; Velado-Alonso, Morales-Castilla and Gómez-Sal, 2022). The top SNPs associated with environmental variables tested here elucidated potential environmental

adaptability capabilities in domestic buffalo for the first time. Associated genes appear to mostly relate to thermoregulation and altitude, according to current literature.

TENM2 (Teneurin Transmembrane Protein 2), associated with temperature and longitude, is a nerve system gene found under selection in high altitude ranging feral Andean horses (Hendrickson, 2013). This gene is involved in embryogenesis and neuronal development. Changes surrounding TENM2 may yield altered behavioural responses as differences in methylation were found in algae eating rock dweller cichlids compared to other species (Vernaz *et al.*, 2021). Under colder conditions, buffaloes increase food consumption to increase metabolism, alter resting behaviours and maintain close body contact with other individuals (Yáñez *et al.*, 2020). NBEAL1 (Neurobeachin Like 1) has previously been linked to high altitude adaptation in several livestock species and features in thermal adaptation via regulation of synaptic transmission (Serranito *et al.*, 2021). NBEAL1 is abundantly expressed in arteries and involved in cholesterol metabolism which may explain links to thermal adaptation (Bindesbøll *et al.*, 2020). CHL1 (Cell Adhesion Molecule L1 Like) is vital for brain and neural development, but is also found expressed in the carotid body (Huang *et al.*, 2013; Fischer and Drago, 2017). Results show this genotype frequent in RV_IND individuals and associated with low precipitation. CHL1 recruits heat shock protein Hsc70 (HSP70) to the synaptic membrane and vesicles (Leshchyns'ka *et al.*, 2006). A study on buffaloes at different altitudes showed that low altitude populations are vulnerable to chronic heat stress, and this would be exacerbated in areas of low precipitation (Lan *et al.*, 2022). HSP70 has been associated with vascular diseases such as strokes (Mehta *et al.*, 2005; Allende *et al.*, 2016; Kim *et al.*, 2018). Therefore, changes in CHL1 function may assist against neuronal and vascular diseases under heat stress. TBC1D4 (TBC1 Domain Family Member 4) and HS3ST4 (Heparan Sulfate-Glucosamine 3-Sulfotransferase 4) have both been associated with cold adaptation. TBC1D4 found under selection in Beringia humans from the arctic region and is likely involved in fat metabolism and changes due to low carbohydrate diet (McGarrah, 2017). HS3ST4 meanwhile, produces heparan sulfate that affects blood thickness and is also under selection in arctic inuit population (Reynolds *et al.*, 2019). Another gene relating to coat colour was found dominated in Bulgarian murrah. ADAMTS20 (ADAM Metallopeptidase With Thrombospondin Type 1 Motif 20) regulates melanocyte colonization of skin and is associated with coat colour variation in goats (Oget, Servin and Palhière, 2019).

Although recent changes mostly link to production traits in buffaloes, these genes indicate that buffaloes can adapt to different climates. Changes in neuronal, metabolic, and vascular genes may help buffaloes alter their behaviour and maintain homeostatic body function to mitigate stress. This capability makes them potentially ideal livestock species for changing future climatic conditions in addition to their current traits of being disease resistant and maintaining productivity in low quality conditions.

3.6. Conclusion

The findings presented here show the level genetic variation across murrah buffaloes and reveal potential adaptations occurring across the globe. All populations featured high levels of heterozygosity with low levels of genetic differentiation between populations as expected. As such, the majority of murrah buffaloes clustered together except for the crossbred RV_BUL_VAR, and two small groups of individuals within RV_IND. Genes identified under selection may relate to how each population has been used. The few regions under selection in RV_IND_3 were associated with oxidative stress in addition to brain and CNS development, which may relate to temperament and adaptation to low quality environments. RV_IND_1 and RV_IND_2 differed in that their genes under selection appear associated with milk production, reproduction, and immunity, suggesting potential development of the breed. RV_BRA shared similarities in terms of genes geared towards productivity with buffalo farming in Brazil taking place in a larger scale operation. Colombian buffalo featured loci under selection associated with muscle development which may relate to ancestry from the buffalypso breed that was used in draught. Meanwhile genes affecting coat colour were identified across all selection methods in Bulgarian murrah that uniquely features the brown coat colour. Evidence of adaptations to the environment were found in relation to neuronal, vascular, and metabolic functions that may help buffaloes regulate homeostasis under different conditions. In summary, a range of potential population specific candidate genes and genomic loci have been identified aiding in understanding the adaptive potential of buffaloes.

Chapter Four

Mitigating the Effects of Ascertainment Bias using Linkage Disequilibrium Pruning

---

4.1. Abstract

SNP genotyping arrays are comprised of highly polymorphic markers that can discern between closely related individuals at an affordable cost. This is beneficial for linking differences in SNPs with phenotypic variation. Though useful for genomic selection within the commercial livestock sector, evolutionary studies face bias as natural genetic variation often features an abundance of lowly frequency alleles. Furthermore, for the Axiom™ Buffalo Genotyping Array, swamp buffalo were not included in SNP selection. This creates further bias in the array as the genetic variation in swamp buffalo is underrepresented. This chapter evaluates ascertainment bias within the Axiom™ Buffalo Genotyping Array in comparison to whole genome sequencing data. This chapter additionally attempts to reduce bias in the array for swamp buffalo by using linkage disequilibrium to prune overrepresented low frequency SNPs. MAF distributions revealed the disparity in allele frequencies between the two species and between data types as an abundance of high frequency alleles were found in the river buffalo array data. Ascertainment bias was evident across most statistical analysis as river buffalo array data was unaffected by LD pruning while swamp and WGS data fluctuated. LD pruning did reduce ascertainment bias as differences between river and swamp buffalo could be minimized with swamp diversity increasing. This trend was not shared by WGS data where results are mirrored depending on the LD pruning target. Therefore, LD pruning in WGS data captured unique genetic variation specific to either species. Without unique swamp variation, LD pruning in the SNP array retained ancestral SNPs that have remained in both species via processes such as balancing selection. ABC results showed that the consequences were that genetic relationships between river and swamp buffalo appeared closer. Studies using the Axiom™ Buffalo Genotyping Array should only focus on river buffalo as swamp buffalo require a new array that incorporates the species unique variation.

4.2. Introduction

The ability to sequence genomes rapidly expanded our understanding of genetic variation and its effects on biological processes across a range of species (Cunningham *et al.*, 2019; Giani *et al.*, 2020). The greatest impact in humans, the has been in medicine where genetics has improved diagnostic services and treatments of many diseases (Manolio *et al.*, 2019; Green *et al.*, 2020). Elsewhere, access to genomic information for the livestock industry provides the raw material for formation of precision breeding plans, targeting alleles that express a desired trait (Meuwissen, Hayes and Goddard, 2013; Georges, Charlier and Hayes, 2019; Burrow and Goddard, 2023). Whole genome sequencing for livestock species provides the full complement of genetic variation, however sequencing as, data storage, and processing infrastructure is costly for large numbers. Utilising genomics without substantial monetary backing is therefore challenging, though large-scale sequencing projects do exist (Daetwyler *et al.*, 2014). The production of single nucleotide polymorphism (SNP) genotyping arrays (henceforth referred to as SNP arrays) offers a cost effective and reproducible vector of genomic data that maintains a high resolution (Syvänen, 2005; Gurgul *et al.*, 2014; Georges, Charlier and Hayes, 2019). As of 2016, over 1.2 million dairy cattle have been genotyped using SNP arrays in the US alone (Wiggans *et al.*, 2017).

SNP array panels may feature upwards of hundreds of thousands of nucleotide probes that hybridise to complementary regions of the tested DNA. Multiple microarray technologies have been developed with Affymetrix and Illumina being the major producers. Affymetrix microarray technology functions uses probe-pairs that compose of a perfect match probe and a mismatch probe differing in a single nucleotide (LaFramboise, 2009). Alternative binding of the complementary DNA to the different probes produces detectable difference in signal intensities. In comparison, Illumina microarrays use extensions of fluorescently labelled nucleotides, conveying SNP genotypes through colour ratios (LaFramboise, 2009). The allocated probes are predefined based on informative polymorphic SNPs selected through the SNP discovery process (Figure 4.1).

Selection of SNPs for microarrays are often based upon a small number of sequenced DNA samples. In the livestock sector, these samples will usually cover breeds of interest. Following sequencing and SNP detection, several stages of filtering and validation, whilst also considering practical factors such as cost, occur to obtain to final SNP set for microarray

production (Geibel *et al.*, 2021). Importantly, the resulting set of SNPs from this process captures tens of thousands of evenly distributed polymorphic SNPs across the genome providing a high resolution for which differences in variation between individuals can be quantified and attributed to functional traits (Gurgul *et al.*, 2014; Georges, Charlier and Hayes, 2019; Berry and Spangler, 2023). The first commercial high density SNP array for livestock produced was the Illumina Bovine SNP50 BeadChip for cattle (Matukumalli *et al.*, 2009).This array contained 54,001 SNPs that enabled the selection of highly desirable individuals on a commercial scale through generation of genomic estimated breeding values (EBVs) developed by Meuwissen *et al,* (2001). Production of various SNP array platforms have been extended to a variety of livestock species, e.g., sheep (Kijas *et al.*, 2009), goats (Tosser-Klopp *et al.*, 2014), and pigs (Ramos *et al.*, 2009), facilitating the genetic improvement of livestock and intensification of the industry (Hayes *et al.*, 2009; Hayes, Lewin and Goddard, 2013; Stock and Reents, 2013; Gorjanc *et al.*, 2015; Brito *et al.*, 2021).

The Axiom^{TM} Buffalo Genotyping Array became commercially available in 2017, featuring 89,988 SNP markers (Iamartino *et al.*, 2017). This marked the first opportunity for large-scale genomic testing of domestic buffaloes, leading to the formation and implementation of genomic selection programmes. Outside of commercial interest, the cost effective and reproducible state of SNP arrays makes the buffalo array an attractive source of genomic information for studying the global patterns of genetic variation and investigation of their evolutionary history (Auton *et al.*, 2009; Kijas *et al.*, 2009; Orozco-terWengel *et al.*, 2015; Iamartino *et al.*, 2017; Colli, Milanesi, Vajana, *et al.*, 2018; Rougemont and Bernatchez, 2018; Muñoz *et al.*, 2019; Pitt *et al.*, 2019; Eusebi, Martinez and Cortes, 2020; Olschewsky and Hinrichs, 2021). Both river and swamp buffalo derive from the same wild ancestor and are able to hybridise to produce fertile offspring, yet there are substantial genomic, phenotypic, and livestock trait (i.e., milk or draught) differences (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020). Understanding the genetic variation within domestic buffaloes will greatly contribute to the effective management of genetic resources across the two species. In turn, this will guide breeding of genetically healthy and productive livestock while avoiding any detrimental effects such as inbreeding or genetic erosion (Plastow, 2016; Rauw, 2016; Lopes *et al.*, 2017; Georges, Charlier and Hayes, 2019; Mrode *et al.*, 2019; Wu and Zhao, 2021).

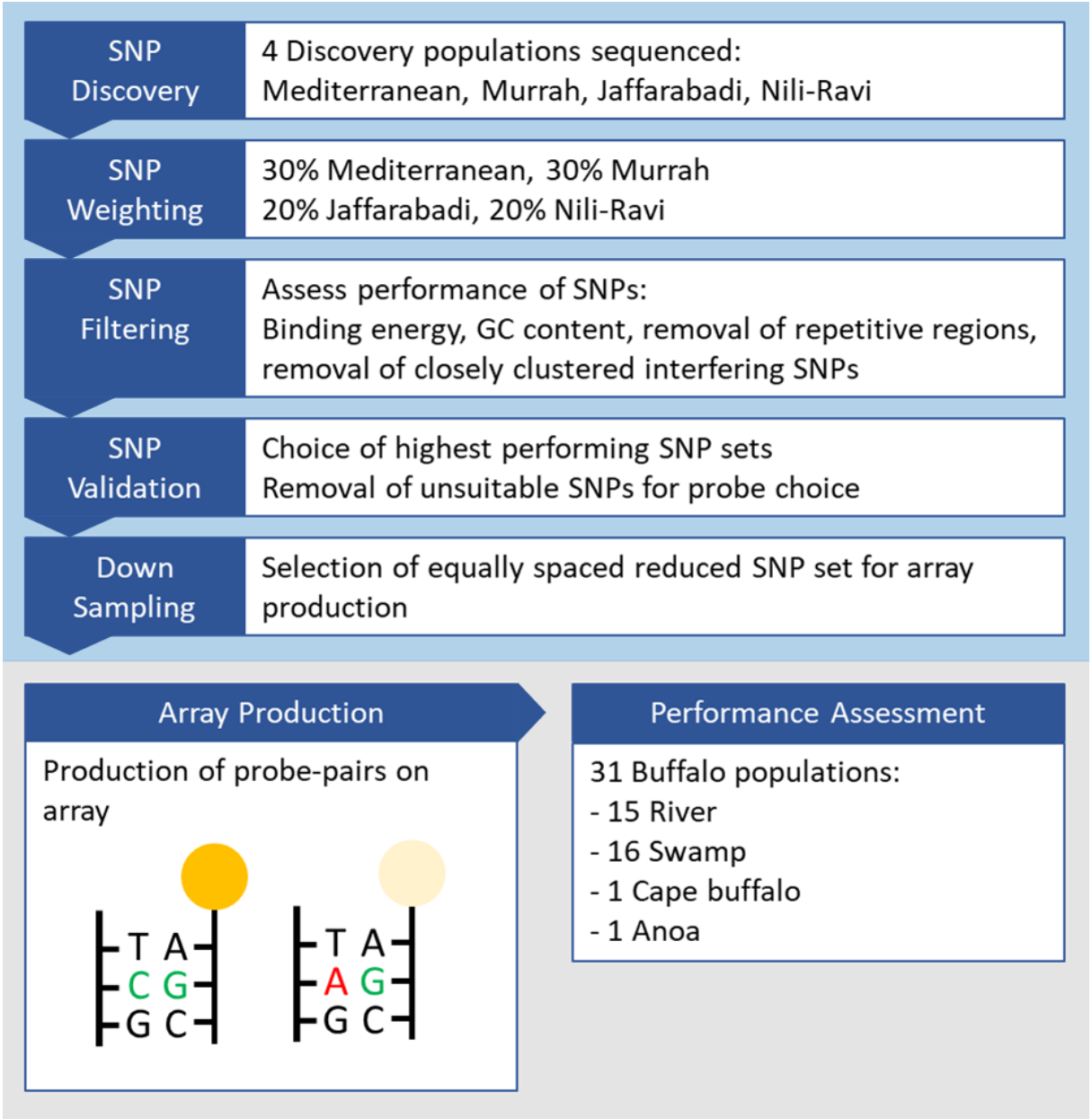


Figure 4.1: Flowchart of SNP genotyping array design for the Axiom™ Buffalo Genotyping Array. Information in figure on array design has been obtained from Iamartino *et al.*, (2017). Four river buffalo populations were used to identify potential SNPs for the buffalo array. A series of filtering, validation, and down sampling processes (array design indicated by blue background) take place to select final SNP subset. Array was produced (array production & assessment in grey background) using Affymetrix probe-pair technology and 31 buffalo populations were used to assess the performance of the array. Stages are in order and direction is shown via arrows. Graphic adapted from Geibel *et al.*, (2021).

Representative samples are required to obtain accurate statistical outputs and interpretations of those populations studied. Genetic studies require a representative sample of the standing genetic variation. While SNP genotyping arrays provide ample benefits, they however do not typically represent standard levels of genetic variation (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Benjelloun *et al.*, 2019; Geibel *et al.*, 2021). SNP arrays are deliberately produced with an abundance of highly polymorphic SNPs , with rarer SNPs avoided (Nielsen, 2004; Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Geibel *et al.*, 2021). This form of ascertainment bias leads to distortions in allele frequency and overinflation of statistics (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018; Benjelloun *et al.*, 2019; Geibel *et al.*, 2021). For example, allele frequency distributions produce a signal more akin to genetic bottlenecks due to the absence of rare alleles, while estimates of heterozygosity and divergence are frequently over or under represented in comparison to WGS (Nielsen, 2004; Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018; Benjelloun *et al.*, 2019).

Statistics such as principal component analysis (PCA) are more robust against ascertainment bias, therefore providing possibilities to reduce its impact on results (McTavish and Hillis, 2015). Linkage disequilibrium is used over frequency-based methods for identifying regions under selection in SNP arrays due to the bias in allelic frequencies(Qanbari and Simianer, 2014). However, the skewed allele frequencies in SNP arrays is not the only bias. The SNPs in the Axiom™ Buffalo Genotyping Array were selected on Murrah (30%), Mediterranean (30%), Jaffarabadi (20%), and Nili-Ravi (20%), thus, the genetic variation present in the SNP array represents polymorphisms found in these breeds (Iamartino *et al.*, 2017). When genotyping breeds and populations outside of those in the discovery panel, no further variation unique to those breeds can be captured (Pérez-Enciso, Rincón and Legarra, 2015; Eusebi, Martinez and Cortes, 2020; Olschewsky and Hinrichs, 2021). Recently diverged and isolated populations will share the majority of their genetic variation, and therefore will be minimally affected in statistical analysis. Populations that are less related (i.e. older divergence) will share less genetic variation with respect to that captured by the SNP array, and the SNPs will largely appear as monomorphic or rare alleles. These genotyped populations will consequently appear as though they possess little genetic diversity when, in

reality, their diversity is simply not captured by the array (Nielsen, 2004; Albrechtsen, Nielsen and Nielsen, 2010; Malomane *et al.*, 2018). The four breeds that founded the buffalo array are all riverine breeds thus a bias is created against swamp buffalo.

This chapter evaluates ascertainment bias between river and swamp buffaloes present in the Axiom™ Buffalo Genotyping Array and attempts to minimize its effects through alternative selection of independent SNPs. Several strategies have been suggested to minimize bias in arrays. Some studies use quality control filtering such as minor allele frequencies to remove rare alleles or SNPs out of Hardy-Weinberg's Equilibrium while others use statistics resistant to bias (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018). Other studies specifically model ascertainment bias in their analysis, thus matching neutral models to observed data (Quinto-Cortés *et al.*, 2018). The use of linkage disequilibrium (LD) pruning is used within this chapter and has been previously suggested to reduce the effects of ascertainment bias across populations of wild and domestic chicken (Malomane *et al.*, 2018). The rationale behind this is that LD pruning removes nearby SNPs that reflect the same genealogical history, i.e. are linked. This method may aid in removal of the excess of monomorphic and rare alleles that are not reflective of their natural genetic variation. Here, LD pruning is alternatively conducted on the two domestic buffalo species and to understand if ascertainment bias can be removed (extending using LD to mitigate ascertainment bias to two different species). This strategy was replicated in WGS data to compare to the array data.

4.3. Materials and Methods

4.3.1. Sample Collection & Data Generation

15 Indian murrah buffaloes generated in Chapter Three were chosen and added to Colli *et al,* (2018) global buffalo SNP array dataset. In line with previous Chapters, the SNP array dataset featured 40,695 SNPs. Quality control filters to obtain this number of SNPs can be found in Chapter Two Section 2.3.1. Populations that were present in Luo *et al,* (2021) whole genome sequencing dataset were kept for comparative analysis between arrays and WGS. The full breakdown of populations used in this chapter and their sample sizes can be found in Table 4.1. Seven river buffalo populations and six swamp buffalo populations were available across both datasets. Both species featured populations covering their core historic

range. I.e., river buffalo populations represented India, Pakistan, Middle East, and Europe, while swamp populations represented China, Thailand, Philippines, and Indonesian islands.

The full Axiom™ Buffalo Genotyping Array features up to 89,988 SNPs. Analysis was carried out to check that the distribution of SNPs between the full probe map and the reduced genotyped data were similar across chromosomes. The full annotation was acquired from ThermoFisher Scientific (available at: https://www.thermofisher.com/order/catalog/product/550431). SNPs were removed if they were missing chromosomal locations, present on non-autosomal chromosomes, or were duplicated quality control probes so to match the genotyped data. The distances between SNPs on the probe map and genotyped data were computed and summarised using R (R Core Team 2018). A Pearson's rank correlation test was conducted between the ThermoFisher probe map and present genotyped data for the total number of SNP markers present on each chromosome.

Table 4.1: Buffalo population details and samples sizes available across both SNP array and WGS datasets. RV_IND was produced within this thesis (see Chapter Three Section 3.3.1 for further details), all other populations were obtained from Colli *et al,* (2018) and Luo *et al,* (2021).

Population	Species	Breed	ID	SNP Array Sample Size	WGS Sample Size
India	River	Murrah	RV_IND	15	3
Pakistan		Aza Kheli	RV_AZA	3	5
		Kudhi	RV_KUN	10	5
		Nili-Ravi	RV_NR	15	8
Iran		Khuzestani	RV_KHU	10	5
		Mazandari	RV_MAZ	8	4
Italy		Mediterranean	RV_MED	15	13
River Sub-Total				76	43
China	Swamp	Guizhou	SW_GZ	11	12
		Hunan	SW_HUN	15	5
Philippines			SW_PHI	15	5
Indonesia			SW_SUL	11	5
			SW_SUM	13	5
Thailand			SW_THA	6	5
Swamp Sub-Total				71	37

WGS data obtained from Luo *et al*, (2021) contained 33,516,506 SNP markers. To remove the excessive difference in resolutions between WGS and the Axiom™ Buffalo

Genotyping Array, a pseudo-SNP array dataset was generated using WGS data. A gamma distribution was fitted to the distribution of distances between SNPs in the probe map using the fitdistplus R package (Delignette-Muller & Dutang, 2015). Using this gamma distribution, a vector of 72,434 randomly generated SNP distances were sampled. These distances were used to extract SNPs from the WGS data, producing a pseudo-array that with SNP markers distributed approximately equally to the Axiom™ Buffalo Genotyping Array.

To evaluate ascertainment bias within the SNP array, subsets of SNPs obtained from different populations were generated using linkage disequilibrium (LD) pruning using PLINK (Chang *et al,* 2015). LD pruning was conducted using the settings of window size of 50 SNPs, window step size of 10 SNPs, and r^2 thresholds set at 0.01, 0.05, 0.1, 0.2, 0.5 and 0.8. Multiple datasets were generated at these r^2 thresholds targeting different species and populations (Figure 4.2). The following LD pruning sets were formed: No LD pruning (referred to as all no pruning - ALL NP/r^2 = 1.00), LD pruning across all samples (ALL), LD pruning across all river samples (RIVER), LD pruning across all swamp samples (SWAMP), LD pruning across a population representative of river domestication centre (RIVPK_NIL), LD pruning across a population representative of river non-domestication centre (RIVIT_MED), LD pruning across a population representative of swamp domestication centre (SWATH_THS), and LD pruning across a population representative of swamp non-domestication centre (SWACN_GUI). The number of SNP markers retained across all datasets can be found in Supplementary Table S4.1. Due to the number of repetitive datasets formed, this report henceforth refers to i) **data** as being either SNP array or WGS, ii) **dataset** being how the data was LD pruned (e.g., ALL, RIVER, SWAMP), and iii) **r^2 threshold** as the r^2 (0.8, 0.5, 0.2 etc) used in LD pruning.

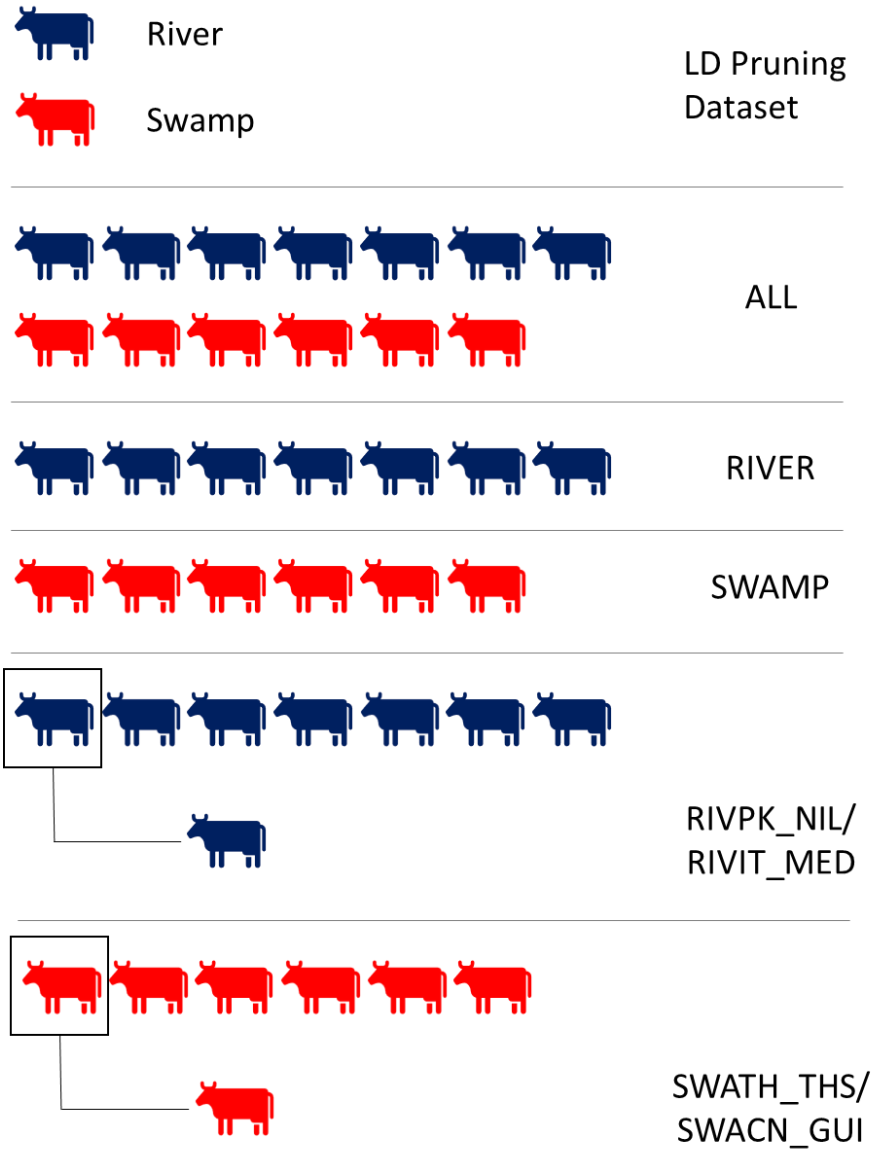


Figure 4.2: Species and populations targeted for generating each LD dataset. Same populations were used for SNP array and WGS data. River (blue) and swamp (red) are defined by the different colours. Each buffalo represents a single population. The number and colour of populations shown for each dataset indicates which populations were used in selecting SNPs by linkage disequilibrium (LD) pruning. Dataset names are linked to the populations that LD pruning was used on, i.e. ALL = LD pruning on all buffalo populations, RIVIT_MED = LD pruning only on the river population from Italy.

4.3.2. Genetic Diversity & Population Structure

Genetic diversities across all LD datasets were measured using a variety of metrics. PLINK was used to calculate minor allelic frequencies (--freq) and both observed (H_O) and

expected (H_E) heterozygosity (--hardy) across all SNPs per population per dataset. The inbreeding coefficient (F) was calculated using moments of method calculation (--het, Equation 1) across all polymorphic SNPs per population in PLINK. All outputs were analysed in R using custom scripts. Choice of T-test was determined through use of Shapiro-Wilks test for normality ($P < 0.05$ being non-normally distributed) and equal variances test ($P < 0.05$ being unequal variances between groups) were conducted for each comparison. ANOVAs were used to test for significant differences across r^2 thresholds within a dataset, and across datasets of the same r^2 threshold.

$$F = \frac{Observed\ Homozygote\ Count - Expected\ Homozygote\ Count}{Total\ Number\ of\ Markers - Expected\ Homozygote\ Count} \quad (1)$$

Pairwise population F_{ST} comparisons were calculated within each dataset using VCFtools (Danecek *et al,* 2011). Parametric and non-parametric T-tests were used to determine significant differences between species and data. Parametric and non-parametric T-tests and ANOVAs were used to analyse differences across species, data, datasets and r^2 thresholds. F_{ST} results were further analysed via Pearson's correlation tests to determine if the same patterns in genetic variation were present between datasets. P-values from correlation test outputs were adjusted using the false discovery rate (FDR) to account for multiple testing. Consistency of population structure was further analysed using a non-multidimensional scaling (NMDS) analysis in PLINK to support F_{ST} results. 20 synthetic components were computed in PLINK and NMDS outputs were analysed using custom R scripts. Correlation tests were completed to identify significant correlation across major axis between datasets, with P-values adjusted for multiple testing using FDR.

4.3.3. Runs of Homozygosity

Runs of homozygosity (ROH) provided a statistic that is calculated per individual instead of being population based. ROHs were calculated in PLINK (--homozyg) using settings found in (Macciotta *et al.*, 2021) that were done using the Colli *et al,* (2018) SNP array dataset. Outputs were analysed in R using the package detectRUNS (Biscarini *et al,* 2019). The settings used to calculate ROHs were 1Mb minimum ROH length, 15 SNPs minimum number of SNPs in ROH, 0 heterozygous or missing SNPs in ROH. ROHs were summarised by three metrics

being i) Number of ROHs, ii) Length of ROHs, and iii) Distribution of ROH classes. Parametric and non-metric T-tests were used to determine significant differences in these measures between species and between datasets, again accounting for normality and equal variances.

4.3.4. Demographic Modelling

Approximate Bayesian Computation (ABC) was used to observe how different LD datasets affected demographic modelling. No evolutionary model has currently been produced for domestic water buffalo, therefore the simplest of models was generated for this analysis. One river and one swamp population were used that joined at some point back in time. The two populations used were a Mediterranean river buffalo (RIVIT_MED) population composing of 13 individuals and a Chinese swamp buffalo population (SWACN_GUI) of 11 individuals. Populations were selected by way of the largest WGS populations available, and datasets were randomly sampled for 2,500 SNP markers to reduce computational time for ABC simulations. Three models were simulated being i) no migration, ii) river to swamp migration, and iii) swamp to river migration. The ABCtoolbox pipeline was used (Wegmann *et al.*, 2010). ABCsampler implemented within ABCtooolbox was used to randomly sample parameter values from defined prior ranges for effective population sizes (N_e), divergence time, and migration rates (Table 4.2). One million backward coalescent simulations were generated using FastSimCoal v2.6 with summary statistics of the simulated genetic variation calculated using Arlequin v2.6.2.2 (Excoffier and Lischer, 2010; Excoffier *et al.*, 2013, 2021). The closest 1,000 simulations by distance to observed summary statistics were retained for generating posterior distributions using ABCestimator within ABCtoolbox.

Table 4.2: Prior ranges and distributions for parameter sampling

Parameter	Abbreviation	Parameter Sampling Distribution	Minimum	Maximum
Effective Population Size River	NR	Log_{10}	2	7.5
Effective Population Size Swamp	NS	Log_{10}	2	7.5
Effective Population Size Wild Ancestor	NA	Log_{10}	2	8
Divergence Time	T-Split	Log_{10}	1	6
Migration Rate River to Swamp	MRS	Log_{10}	-6	0
Migration Rate Swamp to River	MSR	Log_{10}	-6	0

4.4. Results

4.4.1. Dataset Generation

123,040 SNP markers were found in the full Axiom™ Buffalo Genotyping Array probe map. After removal of markers such as duplicated quality control markers, and those with missing locations, 72,434 autosomal SNP markers remained for further filtering. 40,695 SNPs remained in the observed SNP array dataset following quality control filtering and merging of datasets (Chapter Two, Section 2.3.1). The distribution of SNPs across chromosomes for the full probe map and genotyped dataset can be seen in Figure 4.3. SNP distributions between the probe map and observed SNP data significantly correlated (r^2 = 0.999, $p < 0.0001$). Therefore, despite the loss of SNPs in the observed genotyped data, the distribution is still representative of the full probe map.

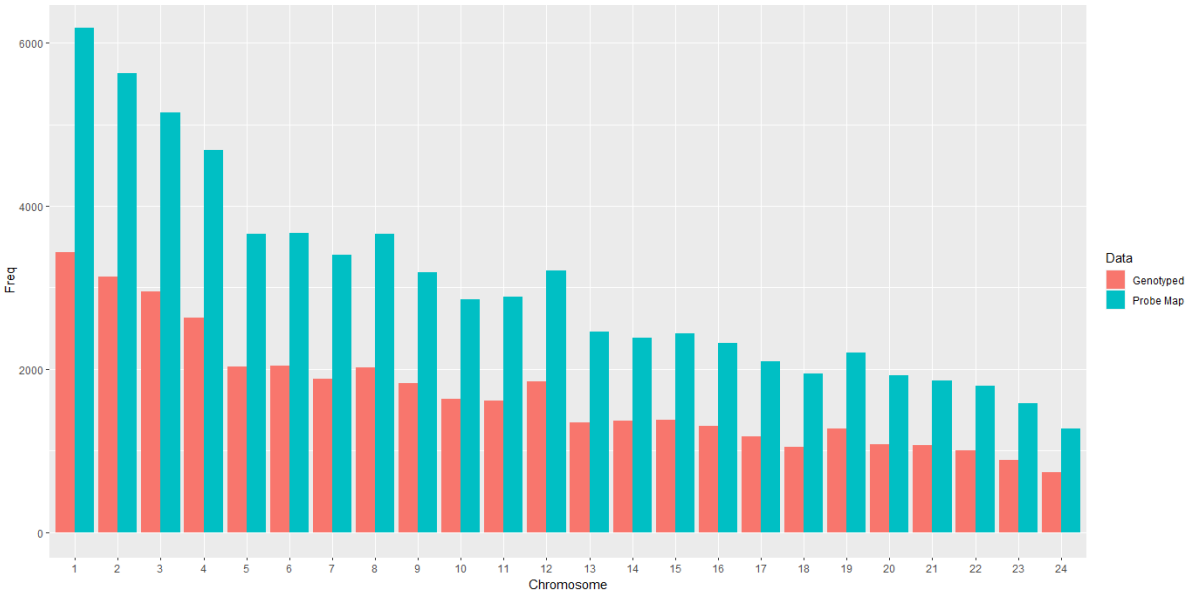


Figure 4.3: Total number of SNP markers found across each chromosome in the full Axiom™ Buffalo Genotyping Array probe map (blue) and present genotyped buffalo populations (red).

Reduced numbers of SNPs in the observed data led to larger distances between SNPs (Figure 4.4). Observed genotyped data featured an average distance of 60.7kbp (±57.3) between SNPs, compared to 34.2kbp (±20.4) in the full probe map. A gamma distribution was fitted to the probe map, giving a shape 5.59 and rate of 0.000169 for which 72,434 SNPs were randomly extracted from WGS dataset according to distances similar to that of the SNP array. Datasets for SNP array and WGS array then underwent linkage disequilibrium pruning. The

number of SNP markers retained in each LD dataset can be found in Supplementary Table S4.1.

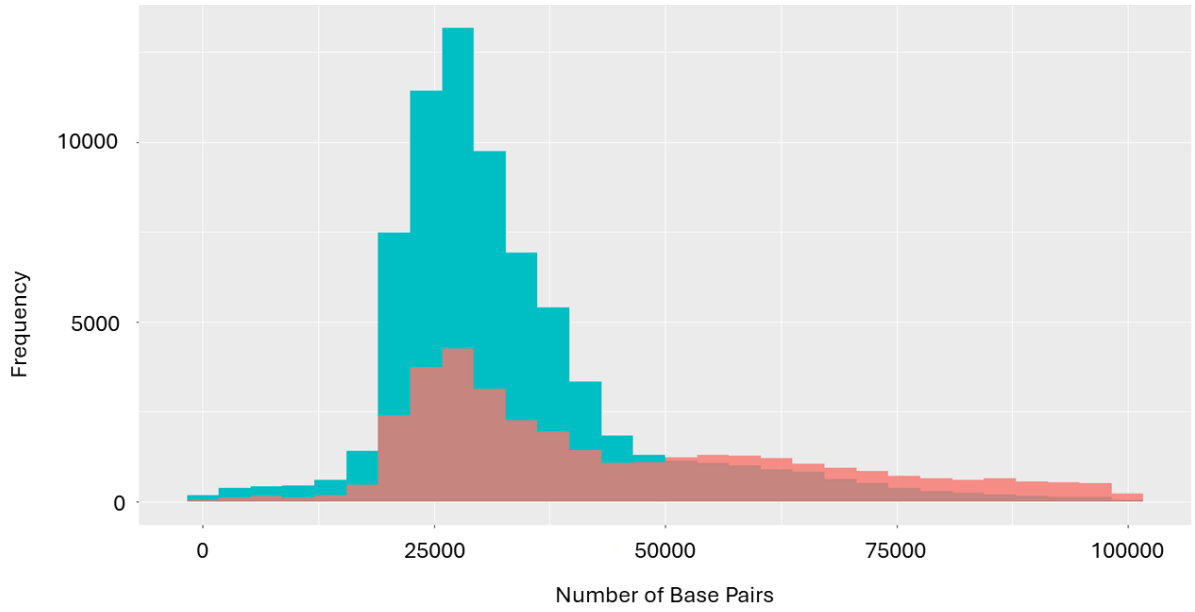


Figure 4.4: Distribution of distances between SNPs in the full Axiom™ Buffalo Genotyping Array probe map (blue) and present genotyped dataset (red). Reduced SNP markers led to larger SNP distances in the genotyped dataset increasing from approximately 34.2 kbp to 60.7 kbp.

4.4.2. Genetic Diversities

It is well known that allele frequencies in SNP arrays typically do not represent that of natural genetic variation owing to the selection of SNP markers with higher frequencies (Nielsen, 2004; Clark *et al.*, 2005; Lachance and Tishkoff, 2013; Benjelloun *et al.*, 2019). Minor allele frequencies (MAF) were calculated across all populations for all datasets for both SNP array and WGS data. MAF distributions for river and swamp buffalo in all SNP markers for SNP array and WGS data, and LD pruned ($r^2 = 0.2$) data are shown in Figure 4.5. River buffalo MAFs in the SNP array are uncorrelated with the majority of WGS comparisons (Supplementary Table S4.2). SNP array and WGS MAF distributions become significantly correlated when LD pruning is carried out on specific populations (RIVIT_MED & RIVPK_NIL) and low r^2 thresholds (≤ 0.1) within RIVER LD dataset. Swamp buffalo MAFs between SNP array and WGS data are significantly correlated for all LD datasets bar $r^2 = 0.01$ and $r^2 = 0.05$ in the ALL LD dataset (Supplementary Table S4.2). Natural genetic variation typically shows an abundance of low

frequency alleles as seen across WGS. Instead in SNP arrays, an abundance of higher frequency alleles was observed as seen across all SNP array dataset for river buffaloes (Figure 4.5). A large majority of low frequency alleles are observed across most WGS datasets. MAF distributions shift towards higher frequencies in a species as the r^2 threshold becomes closer to 0 when LD pruning is conducted on that same species (i.e. higher allele frequencies found in river buffalo when LD pruning targets river buffalo). When LD pruning does not include a species, the MAF distribution remains shifted towards an abundance of low frequency alleles. Therefore, MAF distributions between species within WGS data are only correlated in the ALL LD dataset that incorporates both river and swamp buffalo (Supplementary Table S4.2). River buffalo MAF distributions are largely unaffected by LD pruning in the SNP array and remain uncorrelated with swamp buffalo MAF distributions until LD pruning is specifically conducted of swamp buffalo (SWAMP, SWACN_GUI, SWATH_THS; Supplementary Table S4.2).

A similar pattern was found in heterozygosity results (Supplementary Tables S4.3, S4.4, S4.7, and S4.8). Both H_O and H_E remained high across all LD datasets and r^2 thresholds for river buffalo in the SNP array dataset (H_O: 0.373 – 0.472; H_E: 0.352 – 0.444). In contrast H_O and H_E greatly fluctuated in the swamp buffalo SNP data (H_O: 0.115 – 0.429; H_E: 0.112 – 0.421) and both species in WGS data (River H_O: 0.125 – 0.473; River H_E: 0.112 – 0.416; Swamp H_O: 0.094 – 0.423; Swamp H_E: 0.089 – 0.408) depending upon the LD dataset and r^2 threshold. H_O and H_E are shown for both species across SNP array and WGS data for no LD pruning and r^2 = 0.2 for ALL, RIVER, and SWAMP LD datasets in Figures 4.6 and Figure 4.7, respectively. H_O and H_E was significantly greater ($P < 0.05$) in SNP array data for river buffaloes than WGS data for the majority of LD datasets bar lower r^2 thresholds in RIVER, RIVIT_MED, and RIVPK_NIL LD datasets (Supplementary Tables S4.11). In SNP array data, H_O and H_E became significantly greater as r^2 threshold decrease in RIVER, RIVIT_MED, and RIVPK_NIL LD datasets, while there was no effect in other datasets (Supplementary Tables S4.11). H_O and H_E for river buffalo became significantly greater in all LD datasets in WGS data as r^2 decreases, as a wider spectrum of allele frequencies was included. This effect was far larger in ALL, RIVER, RIVIT_MED, and RIVPK_NIL, compared to SWAMP, SWATH_THS, and SWACN_GUI LD datasets. This trend was not consistent in the SNP array probably because the SNPs in the array are selected for alleles occurring at medium to high frequencies. Swamp buffalo H_O and H_E typically only increased when included in LD pruning in both SNP array and WGS data.

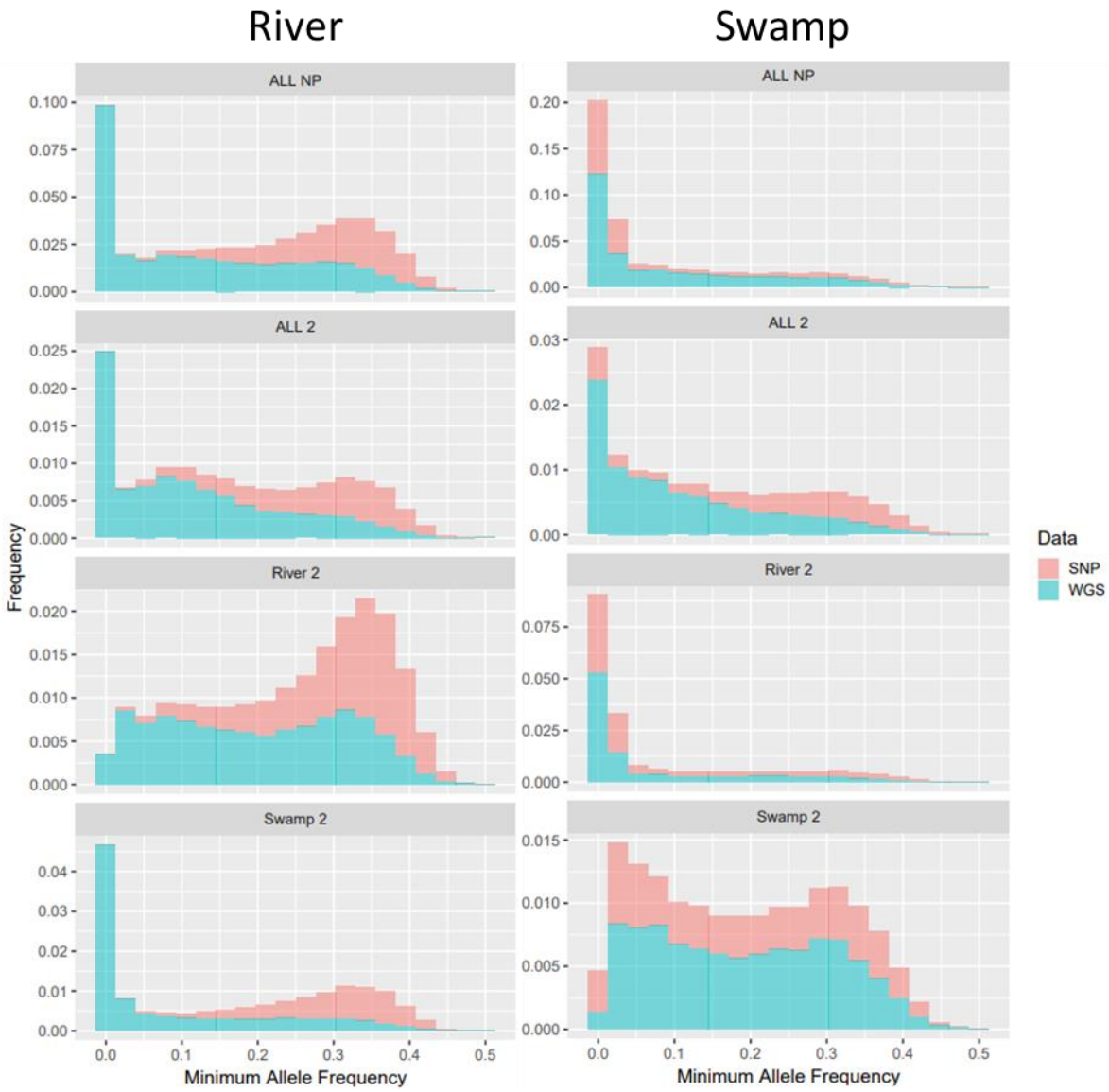


Figure 4.5: MAF frequency distributions across river (left) and swamp (right) buffalo for SNP array (red) and WGS (blue) data. First row features all SNPs with no LD pruning, second row features LD pruning across all buffalo individuals, third row features SNP obtained from LD pruning across river buffalo, and fourth row from LD pruning across swamp buffalo. Rows two to four are LD pruning with an r^2 threshold of 0.2.

H_O and H_E was significantly greater in river buffaloes than swamp buffaloes in SNP array data across most LD datasets and r^2 thresholds (Supplementary Tables S4.3 & S4.4), although at low r^2 thresholds in swamp pruned LD datasets, this difference rapidly decreased or was even overturned. For example, SWAMP $r^2 = 0.01$: Swamp $H_O = 0.429$, River $H_O = 0.396$, while in comparison H_O across the original non-pruned SNP dataset is 0.397 for river buffalo, and 0.118 for swamp buffalo. The same did not occur in WGS data where river buffalo H_O and

H_E was greater across all r^2 thresholds in ALL, RIVER, RIVIT_MED, and RIVPK_NIL LD datasets, whereas swamp H_O and H_E was greater in SWAMP, SWATH_THS, and SWACN_GUI LD datasets (Supplementary Tables S4.7 & S4.8). Significant differences in H_O and H_E were found between SNP array and WGS data in swamp buffaloes except for when LD pruning was conducted solely on swamp buffalo (SWAMP, SWACN_GUI, SWATH_THS LD datasets) (Supplementary Tables S4.11). Swamp buffalo H_O and H_E was highest when river buffalo were excluded from LD pruning.

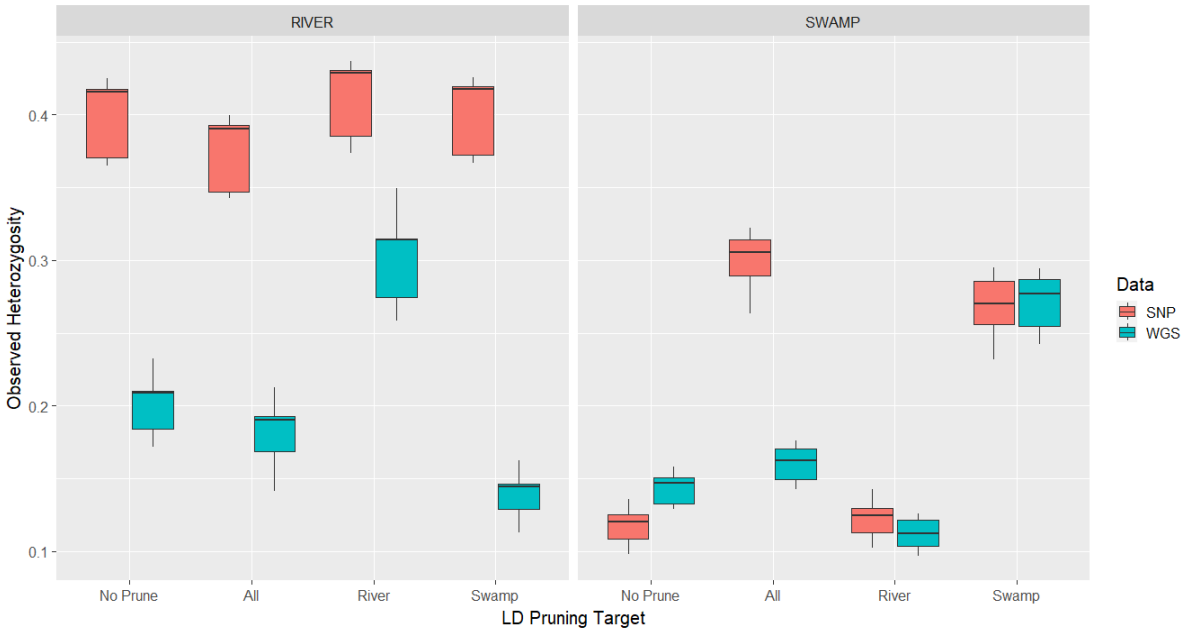


Figure 4.6: Observed heterozygosity for river (left) and swamp (right) buffaloes in SNP array (red) and WGS (blue) data across ALL (NP & r^2 = 0.2), RIVER, and SWAMP (both r^2 = 0.2). River buffalo H_O remains high across all SNP array datasets whilst swamp buffalo H_O increases when incorporated in LD pruning. H_O usually increases in the species the LD pruning is used on for WGS data.

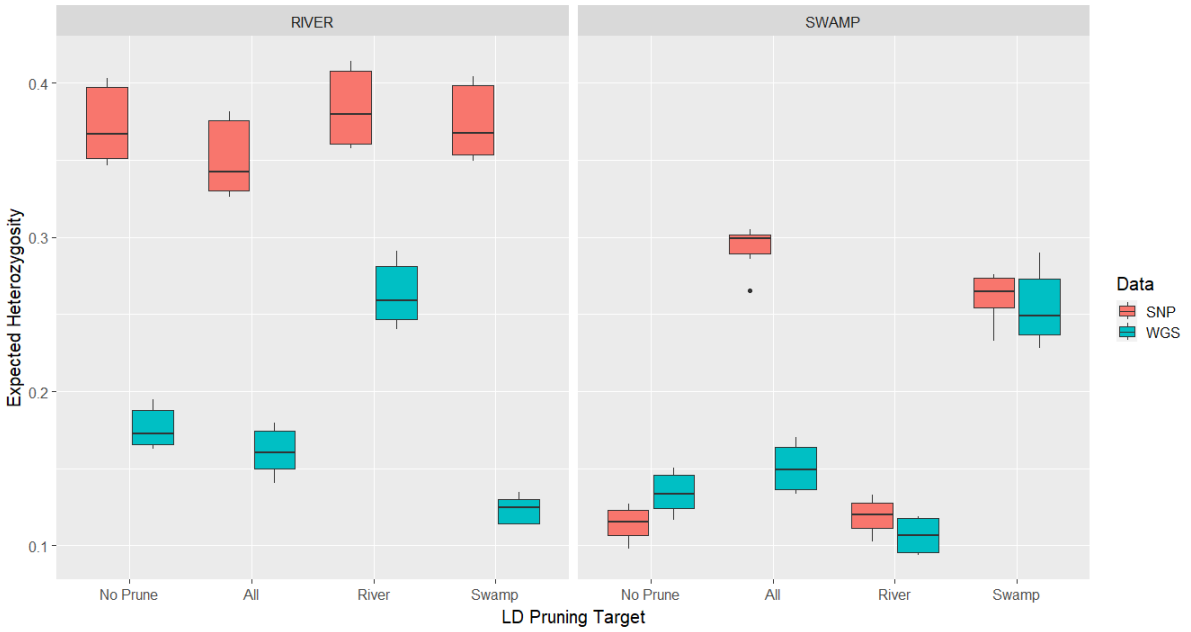


Figure 4.7: Expected heterozygosity for river (left) and swamp (right) buffaloes in SNP array (red) and WGS (blue) data across ALL (NP & $r^2 = 0.2$), RIVER, and SWAMP (both $r^2 = 0.2$). River buffalo H_E remains high across all SNP array datasets whilst swamp buffalo H_E increases when incorporated in LD pruning. H_E usually increases in the species the LD pruning is used on for WGS data.

In contrast to the fluctuations found in H_O and H_E across data and species, F remained stable across datasets and r^2 thresholds (Reduced plot in Figure 4.8). Inbreeding was effectively near absent across both river and swamp buffaloes across all data, datasets and r^2 thresholds, on average being found approximately 0 or negative. F was on average higher for both river and swamp buffalo in SNP array data compared to WGS data (Supplementary Tables S4.6 & S4.10). Sporadic significant differences were found in F when comparing river and swamp buffaloes in both SNP array and WGS data though the general trends were i) swamp buffalo featured a marginally higher F than river, ii) SNP array data featured higher F values than WGS data, and iii) F was largely unaffected by LD pruning and r^2 thresholds (Supplementary Table S4.6, S4.10 & S4.11).

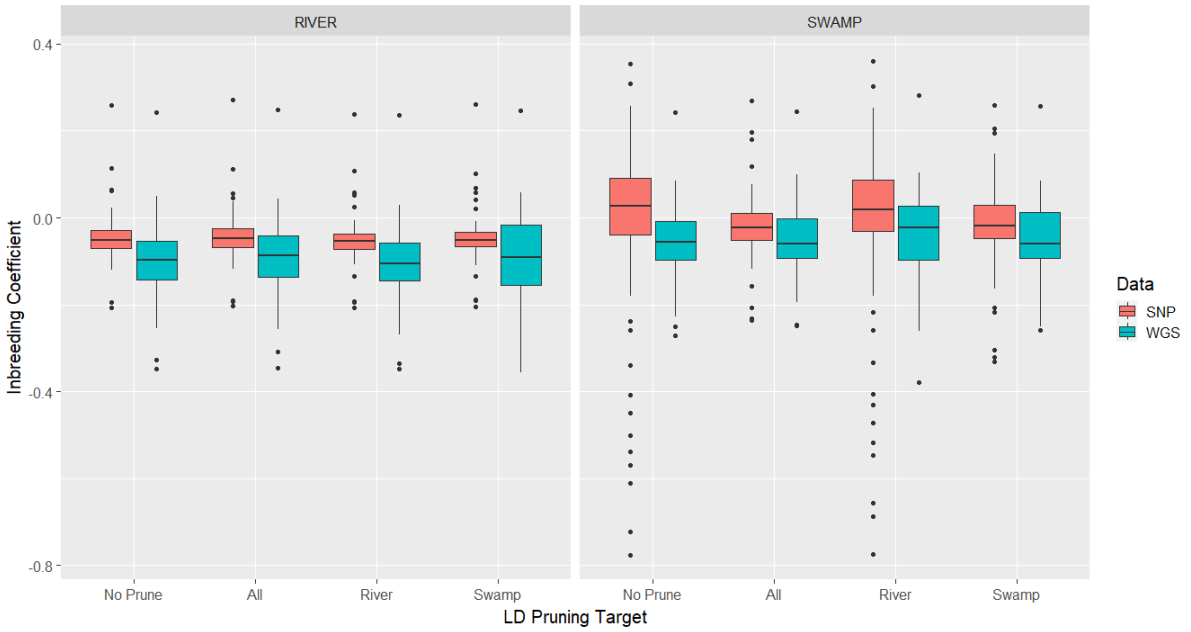


Figure 4.8: Inbreeding coefficient for river (left) and swamp (right) buffaloes in SNP array (red) and WGS (blue) data across ALL (NP & r^2 = 0.2), RIVER, and SWAMP (both r^2 = 0.2). F was largely unaffected by LD pruning, and most LD datasets for both river and swamp found that F ≤ 0, indicating an absence of inbreeding across both species.

F_{ST} values were minimally affected within species, however interspecies comparisons fluctuated greatly (reduced plot Figure 4.9). No significant differences were found for river-river comparisons between SNP array and WGS data, LD dataset or r^2 thresholds (Supplementary Table S4.12). Swamp-swamp pairwise F_{ST} results were additionally unaffected by LD pruning though r^2 thresholds of 0.01 and 0.05 did produce significant differences across LD datasets (Supplementary Table S4.13). River-swamp F_{ST} values were frequently found to be significantly lower in WGS data compared to SNP array data (Supplementary Table S4.14) for ALL, RIVER, RIVIT_MED, and RIVPK_NIL LD datasets. However, in SWAMP, SWACN_GUI, SWATH_THS LD datasets, F_{ST} for river-swamp comparisons were significantly greater in WGS data than SNP array data. Patterns of F_{ST} between data and LD datasets became uncorrelated as LD pruning specifically targeted species and at lower r^2 thresholds.

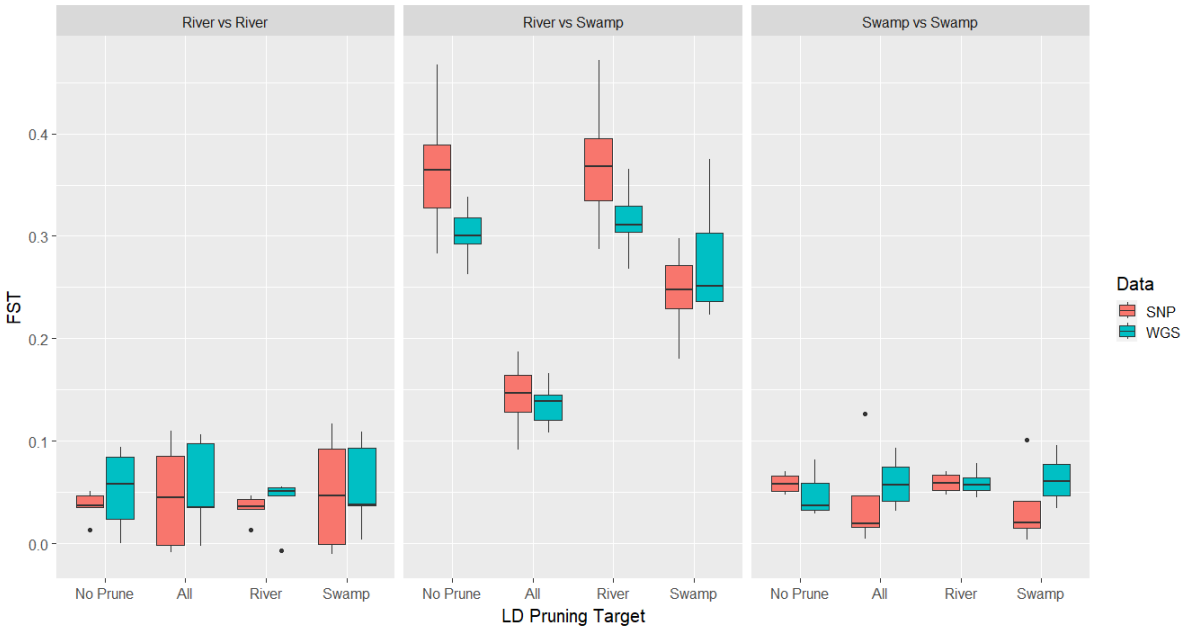


Figure 4.9: Average pairwise F_{ST} values for river-river (left), river-swamp (central), and swamp-swamp (right) buffaloes in SNP array (red) and WGS (blue) data across ALL (NP & r^2 = 0.2), RIVER, and SWAMP (both r^2 = 0.2). F_{ST} was unaffected by LD pruning for both within species pairwise comparisons however fluctuated greatly in river-swamp pairwise comparisons.

4.4.3. Population Structure

NMDS analysis evaluating differences in population structure was affected by data and LD pruning. In all data and datasets, component 1 (reduced plot in Figure 4.10) distinguished river and swamp buffalo populations with all pairwise comparisons between datasets being significantly correlated (3655 total pairwise comparisons covering all datasets, 100% significant correlations). Component 2 (reduced plot in Figure 4.10) distinguished European and non-European river buffalo across most SNP array datasets, and all WGS datasets including river buffalo in SNP selection. WGS datasets that conducted LD pruning on swamp buffalo led to component 2 differentiating between continental and Southeast Asian swamp buffalo populations. As a result, 63.9% of pairwise comparisons (covering all datasets) were significantly correlated for component 2. The majority of differences were led by WGS SWAMP, WGS SWACN_GUI, and WGS SWATH_THS. Low r^2 thresholds (0.01, 0.05) for SNP array data for LD pruning on swamp buffalo also followed this pattern. Component 3 (reduced plot in Figure 4.11) again showed divisions between datasets capturing variation across river or swamp buffaloes. From this component onwards, the targeting of LD pruning dictated the population structuring. For datasets under river buffalo LD pruning, component 3

differentiated between Middle Eastern and non-Middle Eastern river buffalo. While in swamp buffalo LD pruning datasets, component 3 continued to divide swamp buffalo populations. Notably though, original non-LD pruned datasets in both SNP array and WGS datasets captured river buffalo variation across components 2 and 3, while conventional LD pruning captured swamp buffalo variation from component 3. The greater division between variation captured in component 3 meant that 38.7% of pairwise F_{ST} comparisons across all datasets significantly correlated with each other. No further components were analysed after component 3 due to small amounts of variation captured.

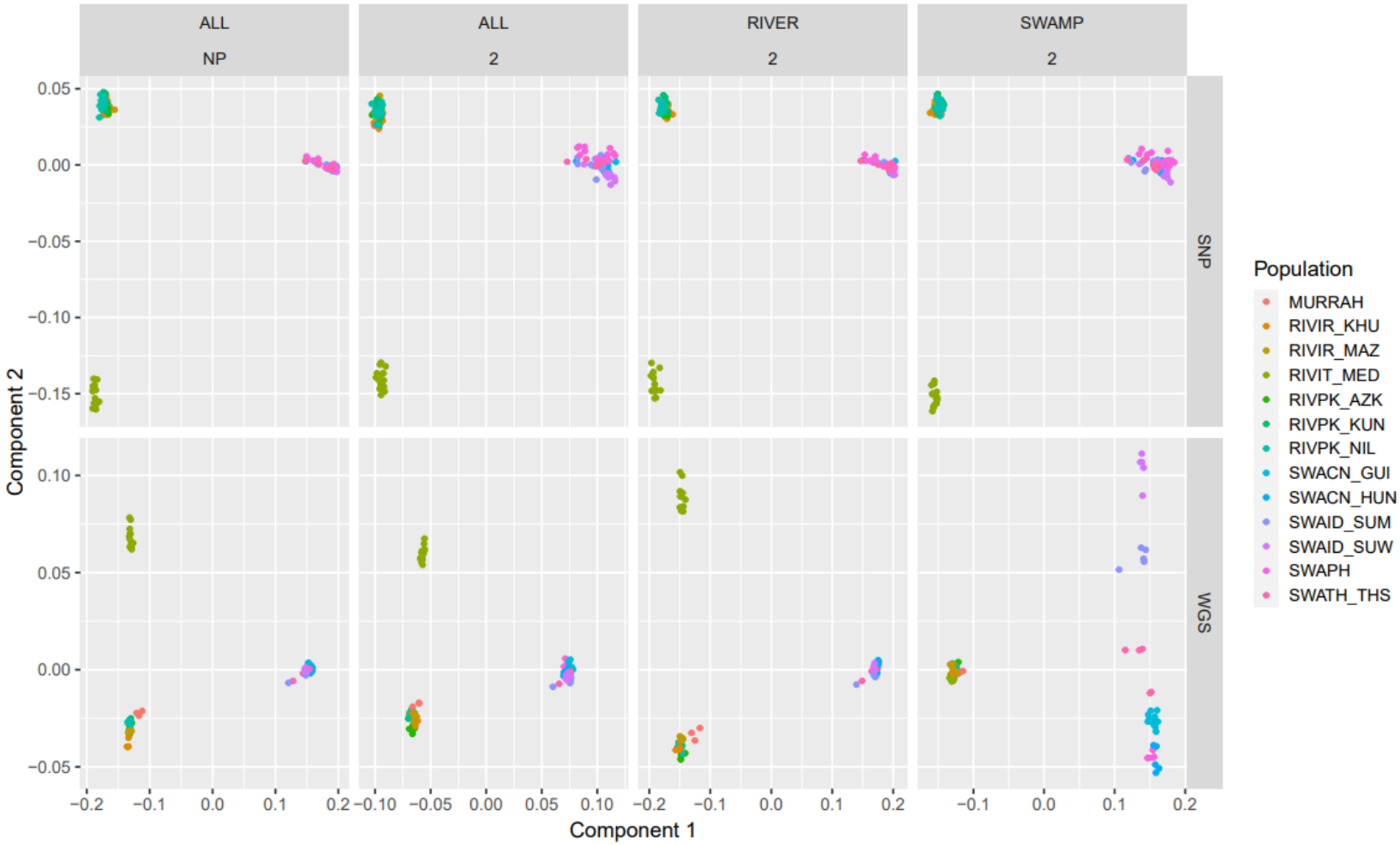


Figure 4.10: NMDS plot showing components 1 and 2 for all buffalo populations across SNP array (Top) and WGS (Bottom) data. LD datasets ALL (NP & r^2 = 0.2), RIVER, and SWAMP (Both r^2 = 0.2) are shown. Component 1 remains constant across all LD datasets, however component 2 capture swamp buffalo variation when LD pruning targets swamp buffalo (bottom right).

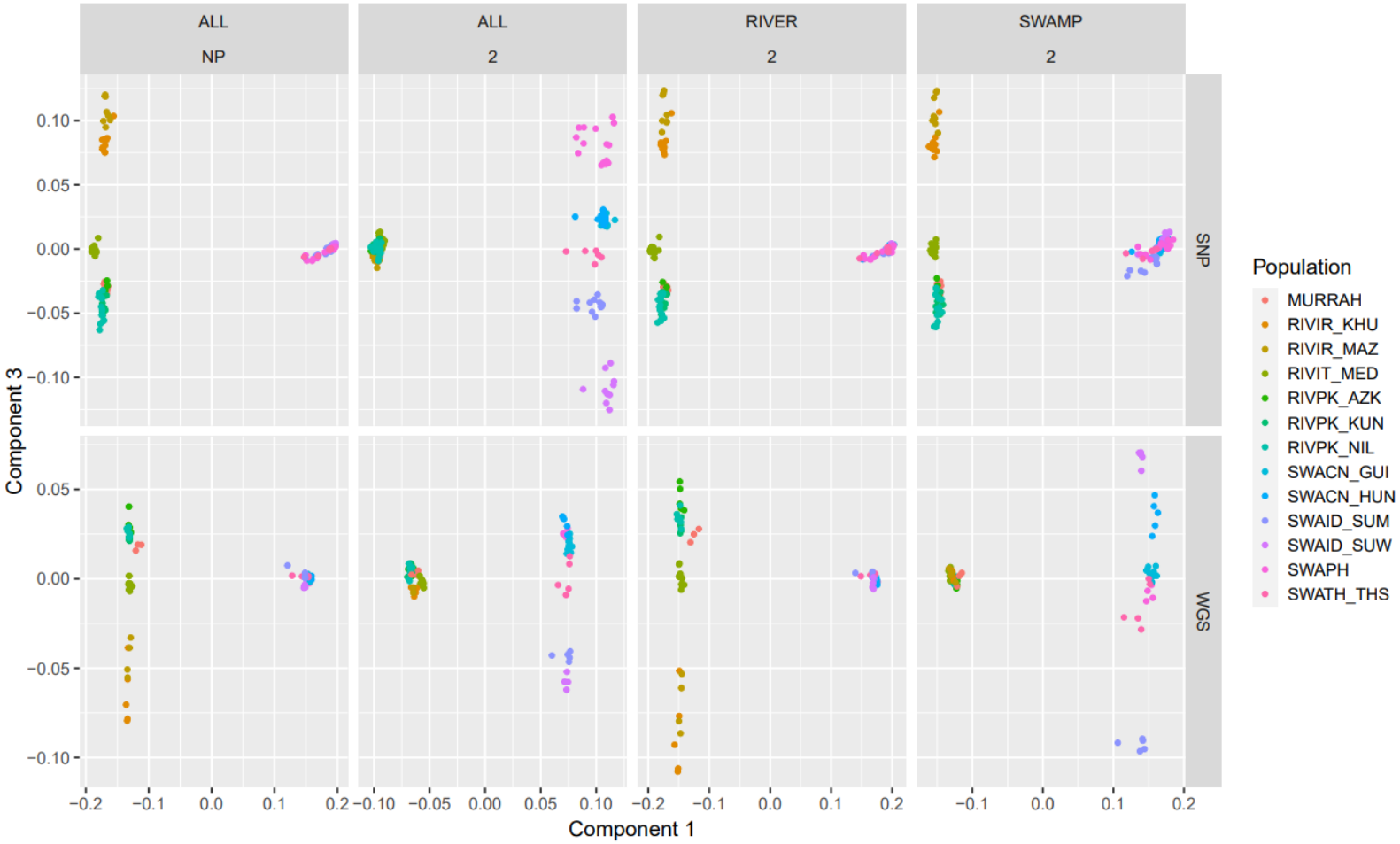


Figure 4.11: NMDS plot showing components 1 and 3 for all buffalo populations across SNP array (Top) and WGS (Bottom) data. LD datasets ALL (NP & r^2 = 0.2), RIVER, and SWAMP (Both r^2 = 0.2) are shown. Component 3 captures river or swamp buffalo variation depending upon the LD pruning target.

4.4.4. Runs of Homozygosity

ROHs (>1 Mb) were measured using three metrics: i) number of runs, ii) length of runs, and iii) ROHs per class. Across ALL, SWAMP, SWACN_GUI, and SWATH_THS, river buffalo possessed fewer ROHs in SNP array data compared to WGS (fewer in RIVER, RIVIT_MED, and RIVPK_NIL) (Supplementary Table S4.15). The number of ROHs was frequently higher in swamp buffalo SNP array data than WGS in all LD datasets (Supplementary Table S4.16). In WGS data, the species that LD pruning was conducted on, featured fewer number of ROHs (Figure 4.12). However, the same pattern is not repeated in SNP array data where swamp buffalo either possessed substantially more ROHs than river buffalo (e.g., ALL, RIVER, RIVIT_MED & RIVPK_NIL datasets) or produced a similar number of ROHs (e.g. SWAMP, SWACN_GUI & SWATH_THS). In the full dataset, 11.750 ROHs per individuals were found across river buffaloes compared to 19.973 in WGS, whilst the difference in swamp buffalo was far greater at 41.042 ROHs in SNP array compared to 26.674 in WGS data.

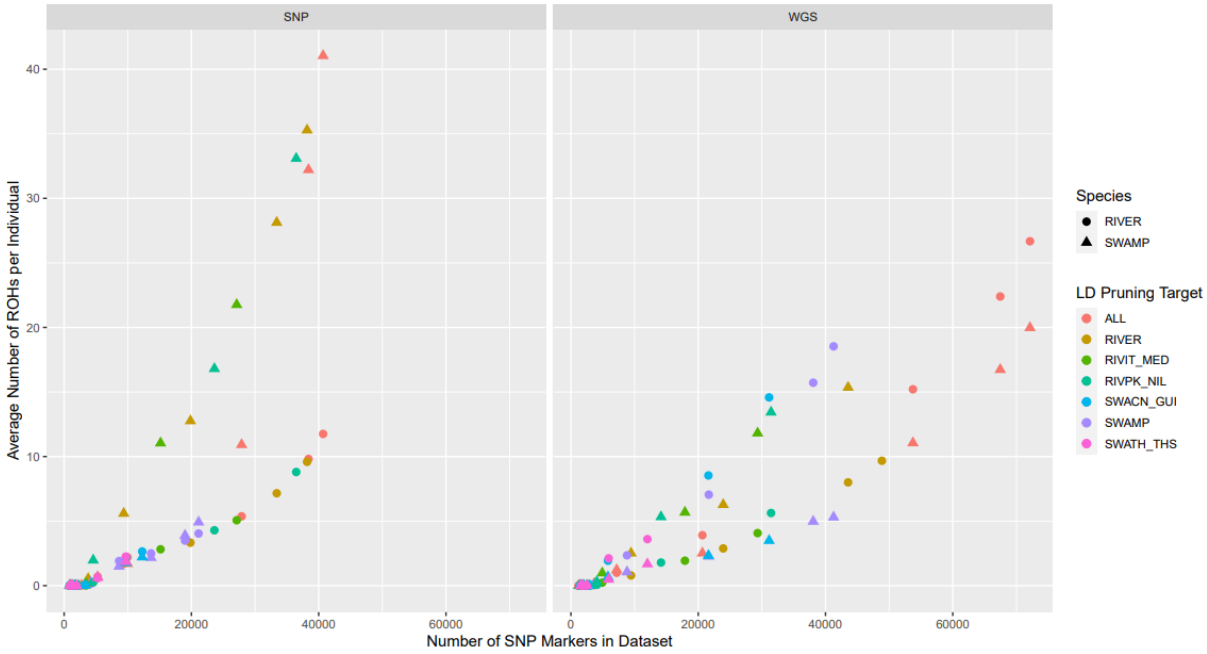


Figure 4.12: Average number of ROHs across river (circle) and swamp (triangle) compared to number of markers present in each LD dataset.

The greater number of ROHs detected, on average, translated into significantly longer ROHs in SNP array data in both river and swamp buffalo species for the majority of LD datasets compared to WGS data (Figure 4.13, Supplementary Tables S4.17-S4.19). ROHs were significantly shorter in swamp buffalo compared to river buffalo in both SNP array and WGS

data at higher r^2 thresholds. The length of ROHs became similar in SNP array as LD pruning increasingly targeted swamp buffalo (SWAMP, SWACN_GUI, SWATH_THS). ROH lengths in WGS became similar as r^2 thresholds decreased in all LD datasets.

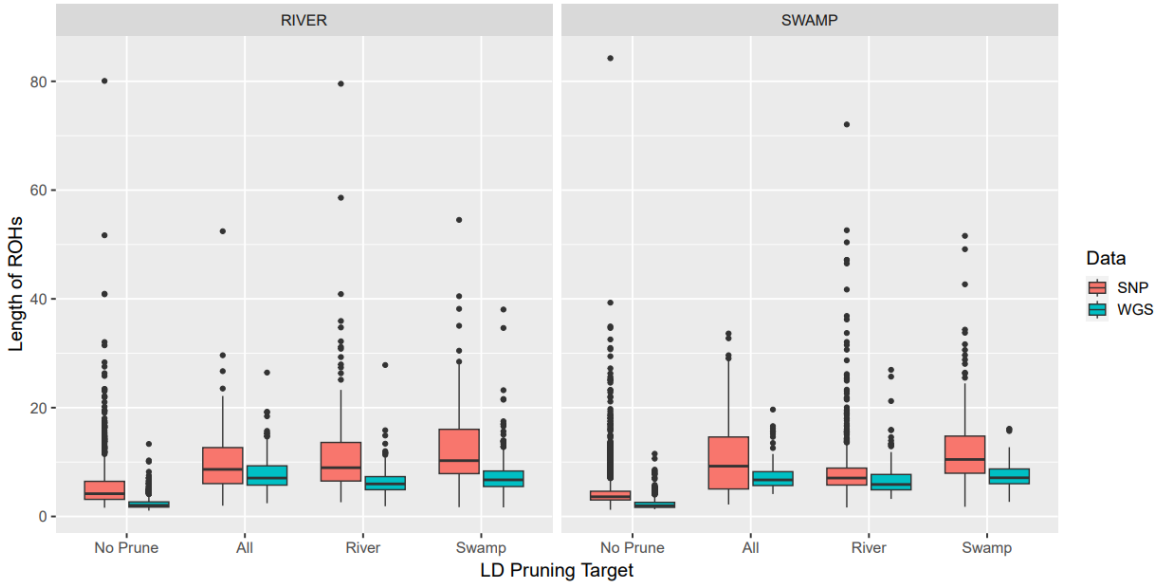


Figure 4.13: Average length of ROHs across river (left) and swamp (right) buffaloes for SNP array (red) and WGS (blue) data for non-LD pruned and r^2 = 0.2 in ALL, RIVER, and SWAMP LD datasets.

Shorter ROHs in WGS data resulted in an abundance of ROHs in smaller ROH class sizes. Few ROHs were detected in both river and swamp buffaloes in SNP array data for class 1 (<1 per individual), while WGS data produced averages of 12.6 and 10.5 1-2Mb ROHs for river and swamp buffalo, respectively (Table 4.3; Supplementary Table S4.20). Shorter ROHs were largely undetectable upon LD pruning in both SNP array and WGS data. ROHs were numerous in Class 2 (2-4Mb) for both SNP array and WGS data in the absence of LD pruning. River buffalo SNP array data featured the lowest number of average class 2 ROHs at 5.4 per individual (compared to 11.9 in river WGS), while swamp SNP array data featured far more than WGS at 25.3 compared to 8 (Table 4.4; Supplementary Table S4.21). The number of ROHs in larger class sizes (3,4,5) declines rapidly in WGS data whereas SNP array data decreases to approximately 1 ROH per individual in classes 4 and 5. Swamp buffalo SNP array data is more greatly affected by LD pruning compared to river buffaloes. Number of ROHs per class for classes 3, 4, and 5 can be found in Supplementary Tables S4.22 – S4.24).

Table 4.3: Number of Class 1 sized ROHS (1-2Mb). Full table found in Supplementary Table S4.20.

Data	Species	LD Pruned Dataset			
		ALL (NP)	ALL (0.2)	River	Swamp
SNP	River	0.039	0.013	0.000	0.013
	Swamp	0.268	0.000	0.056	0.014
WGS	River	12.605	0.000	0.057	0.023
	Swamp	10.459	0.000	0.000	0.000

Table 4.4; Number of Class 2 sized ROHS (2-4Mb). Full table found in Supplementary Table S4.21.

Data	Species	LD Pruned Dataset			
		ALL (NP)	ALL (0.2)	River	Swamp
SNP	River	5.395	0.237	0.092	0.158
	Swamp	25.282	0.268	0.521	0.113
WGS	River	11.860	0.070	0.163	0.028
	Swamp	8.000	0.000	0.351	0.000

Table 4.5; Number of Class 3 sized ROHS (4-8Mb). Full table found in Supplementary Table S4.22.

Data	Species	LD Pruned Dataset			
		ALL (NP)	ALL (0.2)	River	Swamp
SNP	River	4.263	0.789	1.224	0.474
	Swamp	12.986	0.465	7.563	0.437
WGS	River	2.093	2.442	2.116	4.977
	Swamp	1.378	1.784	4.568	1.432

4.4.5. Demographic Modelling

The outcome of LD pruning on previous statistics was observed in evolutionary modelling using ABC (Figure 4.14; Supplementary Figures S4.1 & S4.2). Standard deviation for river, swamp, and total H_O, mean H_O, total H_O, and pairwise F_{ST} were retained for posterior estimation. N_e for river and swamp populations, time of divergence, and migration were most affected by LD pruning. River buffalo N_e increased from between 2.5-3.5 to 4.5-5 under LD pruning across all individuals in the SNP array and WGS data. The same trend was observed for N_e for swamp buffalo however LD pruning on swamp buffalo in the SNP array also produced an increase. Ancestral N_e sizes were largely unaffected by LD pruning, as was

mutation rate. Time of divergence decreased from approximately 3.7 to below 3 when LD pruning across all individuals for SNP array and WGS, and across swamp buffalo in the SNP array. Although previous migration between domestic buffalo species seems unlikely to have occurred, increased migration rates were found in these datasets (Supplementary Figures S4.1 & S4.2).

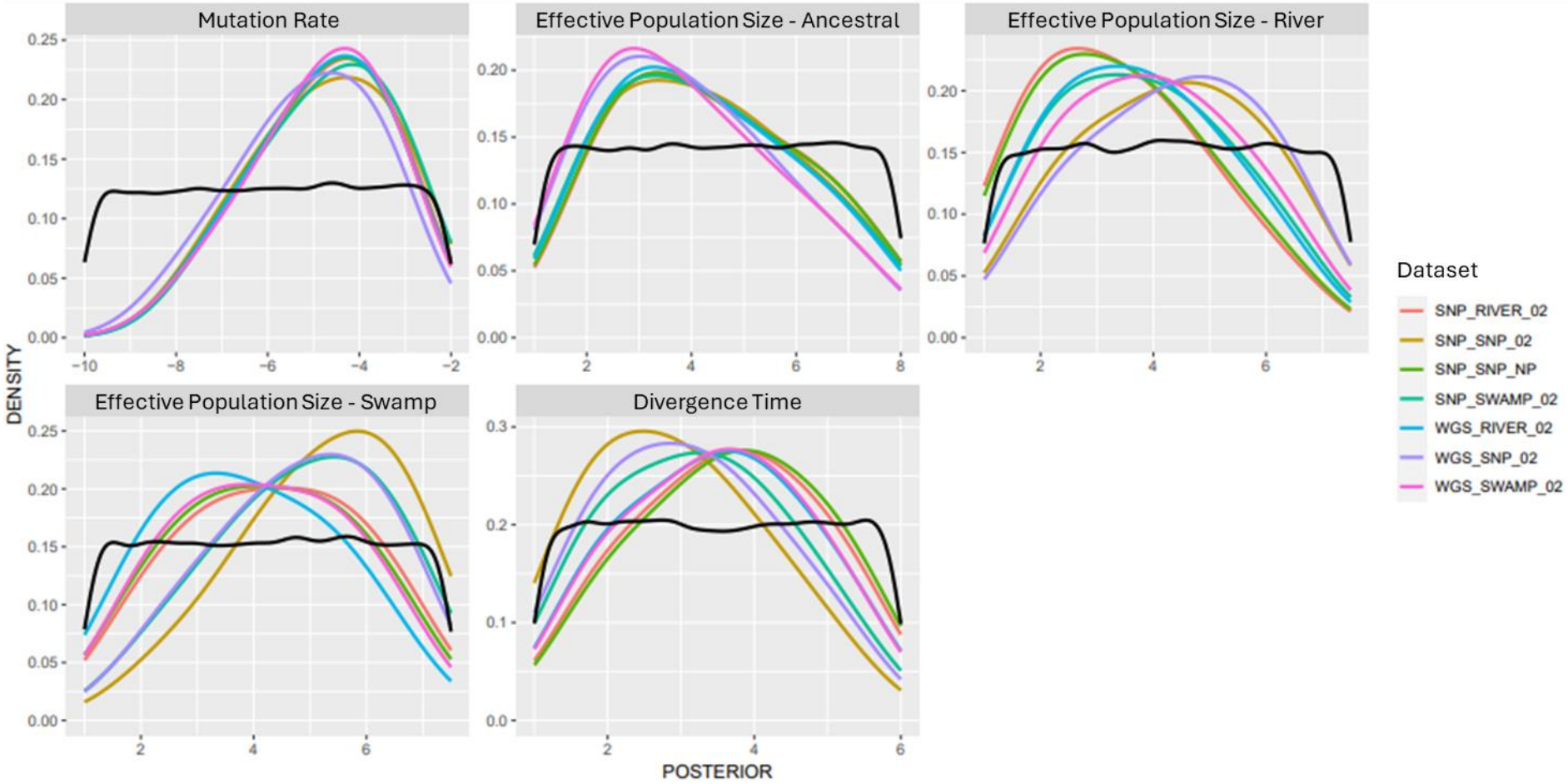


Figure 4.14: Posterior distributions estimated in a no migration model between river and swamp buffalo for WGS and SNP data and LD datasets. The black line displays the prior distribution for each parameter.

4.5. Discussion

SNP arrays provide accessibility to genomic resources in species that would otherwise be challenging and costly to obtain. However, minimizing ascertainment bias is important for ensuring accurate, and robust evolutionary studies within SNP arrays. The Axiom™ Buffalo Genotyping Array is the only affordable and reproducible source of genomic data for analysis of large-scale domestic buffalo populations (Iamartino *et al.*, 2017). Furthermore, this array is commercially valuable, enabling the establishment of genomic selection programmes monitoring genetic health of buffaloes and identifying key genes for increased production yields or greater adaptability (Venturini *et al.*, 2014; Mokhber, 2017; Colli, Milanesi, Vajana, *et al.*, 2018; Du *et al.*, 2019; Lázaro *et al.*, 2021; Macciotta *et al.*, 2021; Nascimento *et al.*, 2021; Noce *et al.*, 2021; Rahimmadar *et al.*, 2021; S. Liu *et al.*, 2022). However, as SNP markers selected for the buffalo array were chosen on polymorphic states solely within river buffalo species, this creates two forms of ascertainment bias, as i) SNPs do not represent natural standing genetic variation, and ii) swamp diversity will be underrepresented in comparison to river buffalo (Iamartino *et al.*, 2017; Colli, Milanesi, Vajana, *et al.*, 2018). Though ascertainment bias could be overcome by other sequencing technologies, WGS remains costly for large scale analysis and methods such as ddRAD-seq tends to result in losses in the number of SNP markers in both pre- and post-sequencing procedures (Cumer *et al.*, 2021; Ros-Freixedes *et al.*, 2022). This chapter has studied the extent of the bias found within the Axiom™ Buffalo Genotyping Array and provides advice of how to manage this in analysis.

Genetic variation is usually observed in the form of an abundance of low frequency that are population or geographically unique with far fewer high frequency alleles that are common across populations or species (Nielsen, 2004; Albrechtsen, Nielsen and Nielsen, 2010). This natural order of genetic variation was observed in the WGS datasets sampled here for both river and swamp buffaloes as expected (Section 4.4.2, Figure 4.5). In comparison with the two species for SNP array data, a very different result was found. Swamp buffalo SNP array data significantly reflected that of WGS data in its entirety, showing similar patterns of MAF distributions across all datasets. Consequently, river buffalo SNP data revealed the opposite with numerous high frequency alleles, and few low frequency alleles. Fewer rare alleles occur in SNP arrays as SNP discovery is based on a small subset of individuals and therefore loci where rare alleles are found in the wider population will be perceived as

monomorphic, and thus avoided in SNP selection (Geibel *et al.*, 2021). This pattern of MAF distributions is common across SNP platforms for various species and is the main form of ascertainment bias in arrays (Nielsen, 2004; Clark *et al.*, 2005; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018; Geibel *et al.*, 2021; Olschewsky and Hinrichs, 2021). Therefore, river buffalo SNP array data was largely found to not significantly correlate with WGS data nor swamp buffalo data (Supplementary Table S4.2). This was until selection of higher allelic frequencies occurred in WGS and swamp buffalo through lower r^2 thresholds or targeted LD pruning.

Interestingly, when WGS data was LD pruned specifically for either river or swamp buffalo, only high frequency alleles found in the targeted species were retained. Lower frequency or monomorphic SNPs that are highly correlated with each are more likely to be removed (VanLiere and Rosenberg, 2008; Malomane *et al.*, 2018). However, the same did not occur for the other species, suggesting that retained SNPs are unique to the target species. This result indicated that river and swamp buffalo are considerably different genetically and that variation is not consistent between the two species. This pattern was not replicated across the SNP array data as river buffalo MAF distributions remained consistent across all datasets, and swamp buffalo MAF distributions only shifted towards high frequencies when LD pruning is targeted. These patterns displayed the bias between river and swamp buffalo in the array as the markers almost represent a near random selection for swamp buffalo, aside from exclusion of unique swamp variation. Meanwhile genetic variation in river buffalo is unlikely to be affected by data processing techniques. This observation may not be as harsh in other livestock SNP arrays as greater diversities of breeds are accounted for in SNP discovery (Vaysse *et al.*, 2011; Ahmad *et al.*, 2020).

The distribution of allele frequencies shown above translated into profound effects on diversity estimates, notably heterozygosity measured here (Section 4.4.2). H_O and H_E produced the same results. When LD pruning included a species, both H_O and H_E increased, further increasing as r^2 thresholds lowered (Supplementary Tables S4.3, S4.4, S4.7 & S4.8). Removal of lower frequency alleles will result in higher heterozygosity values. When LD pruning was carried out in the opposing species, H_O and H_E remained fairly consistent with few significant differences between r2 thresholds. Without LD pruning, there was a large significant difference between river and swamp buffalo H_O and H_E. Average H_O in river buffalo

was high at 0.397 +- 0.048 (He = 0.373 +- 0.022) while swamp H_O is 0.118 +- 0.011 (He = 0.114 +- 0.004). Genetic diversity metrics are frequently overestimated in SNP arrays therefore the outcome of swamp buffalo is likely similar to the real value, while the biased river diversity results is likely overestimated (Nielsen, 2004; Geibel *et al.*, 2021). This difference only increased when LD pruning is targeted at river buffaloes.

The differences in H_O and H_E can be reduced or even removed when LD pruning targeted swamp buffalo due to the greater emphasis on swamp diversity. When LD pruning was undertaken on both species, river buffalo H_O and H_E remained higher than swamp buffaloes, but this difference became non-significant from an r^2 of 0.1 or less. When LD pruning focused on swamp buffalo, this difference was overcome at low r^2 thresholds as significantly higher H_O and H_E estimates in swamp buffalo were found. For example, at r^2 = 0.01, swamp H_O was 0.421 +- 0.016 (in SWAMP LD dataset) while river H_O was 0.373 +- 0.021. Although the difference between river and swamp diversity estimates can be removed, it is important to question whether this is correct. Though the difference between river and swamp diversity is overestimated in the SNP array data, it does appear that river buffalo do possess higher levels of genetic diversity across WGS studies (Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020). Therefore, any data handling or statistical analysis in the SNP array should probably be done with caution to ensure that results are correct, and that misinterpretations generated from the influence of ascertainment bias are avoided.

Similar observations were found in WGS H_O and H_E results except for that river buffalo results now fluctuated. When LD pruning was conducted across both species, river buffalo H_O and H_E consistently remained significantly higher than swamp buffalo across most r^2 thresholds. When LD pruning focused on swamp buffalo, river buffalo H_O and H_E dropped to between 0.1 – 0.2 similar to that of swamp buffalo in the SNP array data. The lack of shared SNP variation between river and swamp buffaloes is a common theme across WGS results that is not reflected in SNP array data. The retention of high H_O and H_E values in river buffalo SNP array data when LD pruning on swamp buffalo occurred indicates that the retained SNPs are frequent in both species despite divergence. Therefore, these SNPs may be shared ancestral variation and diversity may have been kept high through balancing selection, preventing processes such as genetic drift diverging the markers in the absence of migration between species (Asthana, Schmidt and Sunyaev, 2005; Fijarczyk and Babik, 2015; Gao,

Przeworski and Sella, 2015). It is also noteworthy that at low r^2 thresholds, the datasets consisted of a small number of markers (e.g. < 1000 SNPs), therefore results are at greater risk of showing more extremes values (Benjelloun *et al.*, 2019).

While H_O and H_E are greatly affected by SNP selection, other statistics are more robust to ascertainment bias. The inbreeding coefficient, F, evaluates the difference between H_O and H_E, therefore is only dependent upon what is expected of that dataset, and thus decoupled from differences in SNP selection. As a result, the different datasets and r^2 thresholds had little difference, and significant results were unlikely overly meaningful as majority of F results were negative indicating an absence of inbreeding within buffaloes (Colli *et al.*, 2018). F values for river buffalo were typically lower in both SNP array and WGS data despite any LD pruning and often no significant differences were found between species or between data within species. A caveat of the calculation of F here is that it is calculated per population. In this case, swamp F would likely be far higher as monomorphic SNPs would be included, perceiving swamp buffalo as more inbred, a result that appears false.

The correct estimation of statistics such as H_O and H_E may not be as important in biased data if patterns of genetic variation remain the same. I.e., there is acknowledgement that river buffalo H_O is inflated, however the important message is that river buffaloes feature greater levels of diversity compared to swamp buffalo. Patterns of genetic variation were further explored using F_{ST} and NMDS in population structure analysis. F_{ST} results found here revealed that within species comparisons were unaffected by LD pruning or array/WGS data. F_{ST} is a product of variation between populations, therefore the similarity of WGS and SNP array F_{ST} values indicates that ascertainment bias is not affecting within species comparisons (Albrechtsen, Nielsen and Nielsen, 2010). However, between river and swamp populations, F_{ST} was drastically affected by LD pruning. F_{ST} was in excess of 0.3 in both WGS and SNP array data without any LD pruning and was maintained at this level when LD pruning focused solely on river buffalo (Section 4.4.2). The inclusion of swamp buffalo within LD pruning led to a reduction in F_{ST} values in the SNP array (Supplementary Table S4.14). Within WGS data, F_{ST} values would reduce until $r^2 = 0.2$, and then begin increasing again. Selection of common, higher frequency SNPs by LD pruning across river and swamp populations leads to less variation across the dataset, and lower values of F_{ST} (Albrechtsen, Nielsen and Nielsen, 2010). However, due to the dissimilarity between river and swamp buffalo, inflated estimates of F_{ST}

at high r^2 thresholds were observed. This result was likely due to drift on older SNPs, while selection of those SNPs shared across both species led to underestimated F_{ST} values at lower r^2 thresholds compared to WGS data, making the two species appear more similar (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013). Inconsistencies in F_{ST} under LD pruning between distantly related populations was also identified in Malomane *et al,* (2018).

While overall F_{ST} values may change, patterns of variation also changed between datasets. Correlation analysis revealed that F_{ST} values across population comparisons became uncorrelated as LD pruning focused on swamp buffalo, therefore indicating that alternative LD pruning, and attempts to mitigate ascertainment bias, can impact overall patterns of variation and therefore impact interpretations. This is evident from NMDS analysis. From component 2 onwards, different LD pruned datasets captured different proportions of variation between the species (Figures 4.10 & 4.11). Focus on river buffalo populations distinguished between river populations initially, while the opposite largely occurred for swamp buffalo. The original datasets or LD pruning on both species appeared to show the same results. PCA-based methods are considered to be resilient against ascertainment bias between SNP array and WGS data (McTavish and Hillis, 2015).

Runs of homozygosity provided an alternative statistic to assess ascertainment bias as this is a statistic calculated per individual, rather than per population. Although as the number of markers decreased and allele frequencies increased down r^2 thresholds, detectability of ROHs became increasingly difficult, this method provided the means to assess captured variation across the LD datasets (Hillestad *et al.*, 2017). The number of ROHs captured in river and swamp buffalo in the ALL dataset in WGS remained fairly similar for each with river buffalo (~5 ROHs more per individual at an r^2 threshold >= 0.5). SNP array data goes against this as swamp buffalo possessed approximately 2-4 times as many ROHs as river buffalo, evidencing the disparity of low frequency/monomorphic markers in SNP array data between species. Furthermore, WGS data revealed a reciprocal pattern as LD pruning is focused on each species. River buffalo possessed more ROHs in swamp pruned datasets and vice versa, further displaying the dissimilarity is SNPs in WGS data. Meanwhile in SNP array data, swamp buffalo ROHs were either higher or equal to river buffaloes. This supported indications that

when LD pruning is carried out on swamp buffalo, shared diversity between species was retained and therefore similar levels of ROHs were captured in each species.

Increased number of ROHs did not translate into longer ROHs in swamp buffalo SNP array data. Instead, river buffalo ROHs were often significantly greater than swamp buffalo ROHs, however this disparity was removed at lower r^2 thresholds. Significantly shorter ROHs were found in both buffalo species compared to SNP array data. This may be due to the increased number of SNPs present in WGS datasets capturing genetic diversity that is not present in the SNP array data. WGS data has been found to detect shorter ROHs than SNP array due to the presence of heterozygous regions missed in lower density SNP arrays (Szmatoła *et al.*, 2019).

ABC offers a flexible and efficient mechanism of studying evolutionary histories of species. However, its use of summary statistics may leave it vulnerable to the effects of ascertainment bias. Within this study, posterior estimation used heterozygosity and F_{ST} as summary statistics. As observed above, these were affected by ascertainment bias within the SNP array and knock on effects were observed in ABC modelling. Posteriors were largely similar between SNP array and WGS data, however LD pruning across all buffaloes resulted in shifts in river and swamp N_e, time of divergence, and migration. Increases in river and swamp N_e were observed, while time of divergence decreased (Figure 4.14, Supplementary Figures S4.1 & S4.2). Greater selection of higher frequency SNPs leads to increases in heterozygosity and therefore overestimation of effective population sizes, while the absence of rare alleles leads to lower F_{ST} values and more recent times of divergence (Nielsen, 2004). Migration rate consistently tended towards the lower boundary of the parameter, though the addition LD pruning across all individuals had some effect. The retention of ancestral SNPs makes the two species appear as though gene flow has occurred. However, these SNPs were retained by both species upon divergence (Wakeley *et al.,* 2001; Nielsen, 2004). Greater selection of SNPs shared across river and swamp buffalo would play into this overestimation. The only other LD dataset to share the results with ALL datasets is SWAMP in the SNP array data, therefore emphasising the impact of ancestrally shared SNPs or those under balancing selection on the results. This outcome is problematic as although these models are the most simplified for buffaloes, posterior estimates would go against current literature of buffalo domestication.

It is increasingly clear that river and swamp buffalo are separate species with an older divergence than expected (Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Zhang, Colli and Barker, 2020; Curaudeau, Rozzi and Hassanin, 2021). Posterior estimates of < 1,000 generations (~6,000 years) for divergence time as seen in some LD datasets would suggest a post-domestication divergence which is unlikely. The lack of shared mitochondrial and Y-chromosome haplotypes, along with nuclear F_{ST} estimates greater than approximately 0.3 suggest that river and swamp are distinct species with little presence migration (Wang *et al.*, 2017; Colli, Milanesi, Vajana, *et al.*, 2018; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Zhang, Colli and Barker, 2020; Curaudeau, Rozzi and Hassanin, 2021). Previous studies have identified mechanisms of reducing impact of ascertainment bias on ABC models. Quinto-Cortes *et al,* (2018) biased their coalescent simulations to match that of SNP array dataset by providing quality control filters to simulated genetic variation.

4.6. Conclusion

This study evaluated ascertainment bias found within the Axiom™ Buffalo Genotyping Array. Considerable differences between raw values were found across statistics between SNP array and WGS data for river buffaloes which is explained by the choice of high frequency markers found in SNP arrays. Swamp buffalo results for SNP array data were underestimated in comparison of river buffalo SNP array data as a result of ascertainment bias derived from array design. Swamp buffalo results for the SNP array frequently replicated WGS data. This result indicates that the SNPs in the array are reflective of natural genetic variation for swamp buffalo as the divergence between river and swamp buffalo is great enough that the specifically selected polymorphic SNPs for river buffalo appear to mimic a random subset in swamp buffalo. Reciprocal patterns of genetic variation captured by LD pruning that targeted each species in WGS LD datasets further showed that river and swamp buffalo are genetically dissimilar. This pattern was not shared in SNP array data as river buffalo variation was largely unaffected by LD pruning. Within the SNP array data, the bias between river and swamp buffalo could be reduced by LD pruning, although extensive pruning altered patterns of variation that is likely not representative across buffaloes. While LD pruning on WGS data appears to capture unique variation in each species, increased selection of higher frequency markers in SNP array data likely retains alleles that are polymorphic in both species and thus under balancing selection. The SNP array is unsuitable for analysis in swamp buffalo due to

the extent of differentiation between river and swamp buffalo. It would therefore be recommended for future studies that evolutionary studies comparing river and swamp buffalo should utilise less biased methods such as ddRAD sequencing to capture variation across both species. The lack of polymorphic SNPs in the array for swamp buffalo makes the array unsuitable for genomic selection programmes, therefore a second array either unique to swamp buffalo or containing both species would be needed.

Chapter Five

Evolutionary Modelling of Domestic Water Buffaloes using Approximate Bayesian Computation

---

5.1. Abstract

Modelling evolutionary histories of species can provide valuable knowledge on how populations have changed throughout time. The knock-on effect presents opportunities to infer impacts from external factors such as fluctuations in environmental conditions and human societies. Furthermore, these models provide a neutral comparison that may aid future studies in identifying important genomic regions under selection. This chapter quantified and modelled genetic variation across global domestic water buffalo populations of both river and swamp species. Genetic diversity was found to be highest in India, the putative domestication centre for river buffalo. The assemblage of an extensive European dataset here revealed greater variability between Eastern and Western populations than found in Colli *et al.,* (2018). Evolutionary modelling results indicated that divergence times of river buffalo matched migration of humans out of India. Translocation of river buffaloes from the India to the Middle East may have begun early in the common era with the settlement of 'al-Zuṭṭ' across marginal wetlands of the Middle East. An influx of new genetic diversity likely continued from India to the Middle East maintaining genetic similarity between these regions as observed by a more recent divergence between Indian and Middle Eastern populations. The greater divergence found in European buffaloes compared to non-European populations suggests isolation of Mediterranean populations. Despite river and swamp both likely deriving from *Bubalus arnee*, modelling results indicate that their wild progenitors diverged in the Pleistocene. This result supports that river and swamp buffalo were independently domesticated. Results revealing loci under balancing selection found genes associated with survivability and cell viability. The lack of genetic divergence at key biological pathways may

be the reason why river and swamp buffalo are able to hybridise and successfully produce fertile offspring. Regions of divergence between river and swamp genomes were found to be linked with livestock function. Results of this chapter found that genetic variation across river and swamp buffaloes has been greatly shaped by humans.

5.2. Introduction

The subtribe *Bubalina,* comprising of all buffalo species, arose in the Late Miocene following rapid diversification of *Bovini* (Hassanin and Ropiquet, 2004; MacEachern, McEwan and Goddard, 2009; Hassanin *et al.*, 2012; Hassanin, 2014). African and Asian groups can be found within *Bubalina* that are denoted by the genera *Syncerus* and *Bubalus*, respectively. The divergence between these groups dates back to between 5 – 8.5 Ma, although some estimates indicate divergence times as recent as 2.6 Ma (Hassanin and Ropiquet, 2004; MacEachern, McEwan and Goddard, 2009; Hassanin *et al.*, 2012; Hassanin, 2014; Curaudeau, Rozzi and Hassanin, 2021). It appears that the transition from tropical woodland to tropical grasslands in Arabia between 11.6 – 7.25 Ma facilitated the spread of *Syncerus and Bubalus* predecessor, perhaps *Pachyportax*, from Southern Asia into Africa (Pound *et al.*, 2012; Hassanin, 2014). Precise splitting of African and Asian lineages is not clear, though desertification of Arabia post 7 Ma may be the underlying cause (Pound *et al.*, 2012; Hassanin, 2014). Paleontological data finds separation within *Bubalina* during the Pliocene. In Africa, *Ugandax* found in eastern and southern Africa between 2 – 5.7 Ma and is thought to have given rise to *Syncerus* (Hassanin 2014). Meanwhile, *Proamphibos* (2.2 – 3.5 Ma), and later *Hemibos* in the Early Pleistocene, are potential predecessors to the *Bubalus* lineage across Asia (Hassanin 2014).

The dynamic Pleistocene offered a plethora of opportunities for *Bubalus*. Southern Asia presented a stronghold for the group with fossil records showing species ranging into Southeast Asia, Northeastern China, and Europe. Five species of *Hemibos* have been identified across Eurasia all dating to the Early and Middle Pleistocene. Three (*H. acuticornis, H. triquetricornis,* and *H. antelopinus*) of these have been found in Upper Siwaliks deposits of the Pinjor Formation, while *H. gracilis* was found in Longdan, China, and *H. galerianus* in Spain and Italy (Martínez-Navarro and Palombo, 2004; Wang, 2006; Martinez-Navarro & Palombo, 2007; Martínez-Navarro *et al.*, 2011; Siddiq *et al.*, 2019). Transitioning into the Middle-Upper Pleistocene, an emergence of *Bubalus* species is observed. Fossils of *Bubalis murrensis*, the rare European water buffalo, have been found across continental Europe (Koenigswald *et al.*, 2019). This species appeared to have struggled with periods of cold temperatures, and likely fluctuated between Northern and Southern latitudes in sync with glaciations (Koenigswald *et*

al., 2019; Vislobokova *et al.*, 2021). The most recently dated fossil specimen corresponds to 12.7 kya, prior to the Younger Dryas (Vislobokova, Tarasenko and Lopatin, 2020; Vislobokova *et al.*, 2021). In China, a series of shorthorn buffalo persisted in Northeastern regions through the Pleistocene and into the Holocene. Cladistic analysis suggests that as many as six *Bubalus* species may have been present in China with the oldest species, *B. brevicornis*, found in Queque cave deposits dating to 1 Ma (Dong *et al.*, 2014). The most well-known shorthorn species, *B. mephistopheles*, was briefly touted as a potential progenitor to domestic swamp buffalo before mitochondrial analysis showed otherwise (Yang *et al.*, 2008). This species survived into the Holocene, and remains from human hunting can be dated up until 3.6 kya (Yang *et al.*, 2008). The most important Pleistocene *Bubalus* species are likely that of *B. paleoindicus* and *B. paleokerbau* whose distributions likely shaped the current extant species.

Four wild extant species of *Bubalus* can be found in Southeast Asia. Two species of anoa (*Bubalus depressicornis* – the lowland anoa, *Bubalus quarlesi* – the mountain anoa) are both found in tropical rainforests on Sulawesi and Buton Island. The third species, the tamaraw (*Bubalus mindorensis*), persists on Mindoro Island of the Philippines. How these three species got to occupy these islands is not fully clear. No wild relatives currently live on nearby islands such as Borneo, Java, and Sumatra. In the case of Borneo, no *Bubalus* fossils have been found, as opposed to fossils of *B. paleokerbau* on Sumatra and Java (Rozzi, 2017). In the case of anoa, genetic studies suggest a divergence time of anoa and water buffalo of between 0.5 – 3 Ma depending on the use of mitochondrial or nuclear DNA (Priyono *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021; Schreiber and Seibold, no date). The current accepted divergence hypothesis suggests migration of *B. paleokerbau* during times of low sea levels from continental Southeast Asia through Sumatra and into Java, before reaching Sulawesi (Rozzi, 2017). Changes in sea level over the Pleistocene caused the fragmentation of Sulawesi into smaller islands, possibly facilitating the evolution of two distinct anoa species as seen in other species on Sulawesi (Burton, Hedges and Mustari, 2005). The tamaraw provides a more difficult case to explain due to the lack of fossil records hinting at a dispersal route through Southeast Asia. However, the tamaraw may represent an alternative dispersal to anoa as another *Bubalus* species (the extinct *B. cebuensis*) was found on the neighboring island of Cebu, and a third much larger species more comparable to *B. arnee* was identified on the island of Luzon dating back to 65 kya (Croft *et al.*, 2006; Amano *et al.*, 2013).

The Asian wild water buffalo (*B. arnee*) is the fourth extant species. Less than 5,000 individuals remain across wetlands in Northern India and remnant populations in continental Southeast Asia (Kaul *et al,* 2019). A recent emergence of *B. arnee* from *B. palaoindicus* has been hypothesized, with *B. palaeindicus* fossils having been found in central India dating to the Late Pleistocene (Badam, Sathe and Kajale, 1996). *B. palaeindicus* remains have been found alongside extinct megafauna such as the Indian hippopotamus indicating a previous wetter landscape (Badam, Sathe and Kajale, 1996). Aridification of the landscape may have pushed *Bubalus spp* northwards to where *B. arnee* is found today, consistent with claims that *B. arnee* has never been further south in India than the Godavari River (Zhang, Colli and Barker, 2020). The Asian wild water buffalo is the only buffalo species to have been domesticated, having given rise to two domestic species, the riverine (*Bubalus bubalis bubalus*) and the swamp (*Bubalus bubalis carabanensis*) buffalo (Zhang, Colli and Barker, 2020). River buffalo predominantly exist across the Indian subcontinent whilst swamp buffalo can be found in China and Southeast Asia. Initial work focused on identification of domestication origins of river and swamp species, deducing whether one or two domestication events had occurred and should the two forms of domestic species be considered the same species (Kumar *et al.*, 2006; Satish Kumar *et al.*, 2007; S. Kumar *et al.*, 2007; Yang *et al.*, 2008; Zhang *et al.*, 2011, 2016; Yue *et al.*, 2013; Wang *et al.*, 2017).

Mitochondrial analysis shows clear genetic distinctions between river and swamp buffaloes. Estimates of divergence from mitochondrial DNA between domestic buffaloes range between 0.1 – 2 Ma (Tanaka *et al.*, 1996; Lau *et al.*, 1998; Wang *et al.*, 2017). However, the generation of additional nuclear markers presented new evidence with microsatellite and RFLP data supporting a Pleistocene divergence although estimate were more recent at 10,000 – 15,000 ya (Barker *et al.*, 1997). NGS data has since suggested at an intermediate divergence between mitochondrial and microsatellite data between 0.2 and 0.5 Ma (Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021). Supporting genetic data, river and swamp buffalo feature different chromosomal numbers at 24 and 23 pairs, respectively (Ulbrich & Fischer 1967; Fischer & Ulbrich 1968; Iannuzzi 1998). Swamp buffalo feature a chromosomal fusion between what are chromosomes 2 and 3 in the river buffalo (Ulbrich & Fischer 1967; Fischer & Ulbrich 1968; Iannuzzi 1998). Domestic buffaloes are also morphologically distinct. River buffaloes feature large bodies with a near black coloration all

over, along with small, curled horns while the swamp buffaloes feature white legs and chevron under neck with large horns like its wild ancestor, although with a smaller stature (Zhang, Colli and Barker, 2020).

Evidence overwhelmingly shows that river and swamp buffaloes are different species and genetically distinct (Barker *et al.*, 1997; S. Kumar *et al.*, 2007; Yindee *et al.*, 2010; Zhang *et al.*, 2011; Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Zhang, Colli and Barker, 2020; Curaudeau, Rozzi and Hassanin, 2021). However, both species are fully capable of interbreeding to produce fertile offspring which will remain possible until biological processes between species become incompatible (Ottenburghs *et al.*, 2020). (Zhang, Colli and Barker, 2020). Our understanding of the evolutionary relationship between river and swamp buffalo is unclear with a large variation in estimations. Furthermore, how domestic buffaloes have diverged but remain capable of hybridizing has not been investigated. Previous studies have inferred the demographic history of river and swamp buffaloes, however this is yet to have been explicitly modelled(Mintoo *et al.*, 2019; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020).

This chapter quantifies genetic variation across river and swamp buffalo populations and models their evolutionary history, estimating changes in N_e, divergence times and presence of gene flow post-split. The models further investigate recent divergence between river breeds through modelling the number of dispersals out of India. Evolutionary modelling is conducted using Approximate Bayesian Computation (ABC) as this enables efficient testing of multiple specific scenarios (Csilléry *et al.*, 2010). Following modelling, loci under divergent or balancing selection are tested for to identify genes associated with behind the genetic divergence of domestic buffalo species.

5.3. Materials and Methods

5.3.1. Sample Collection & Data Generation

The full buffalo genotyping dataset was used here consisting of UK buffalo (Chapter Two), Indian murrah (Chapter Three), European populations from Noce *et al,* (2021), and global population data from Colli *et al,* (2018). Full details of population used can be found in Table 5.1 below. Two populations (RIVPK_AZK & SWACN_WEN from Colli *et al.,* 2018) were excluded from analysis due to small sample sizes (both n=3). Dataset generation was the same as Chapter Two and Chapter Three, giving a dataset of 821 individuals covering 40,695 SNPs

from the Axiom™ Buffalo Genotyping Array. Due to the ascertainment bias found within the array between river and swamp buffalo, polymorphic markers (MAF > 0.05) found across swamp buffalo were retained for analysis. LD pruning was not used to generate a dataset due to potentially emphasising areas of balancing selection that may make river and swamp buffalo appear evolutionary closer (Chapter Four).

5.3.2. Genetic Diversities & Population Structure

The observed (H_O) and expected (H_E) heterozygosity (--hardy), and inbreeding coefficient (F) (--het) were calculated for each breed in PLINK v1.9 (Chang *et al,* 2015). Summaries (e.g., mean) and analysis of genetic diversities was completed in R v4.0.0 using custom R scripts (R Core Team, 2018). An ANOVA was used to test for significant differences in heterozygosity across populations. Following Shapiro-Wilkes tests of normality, a Welsh's t-test was used to test for significant differences between H_O and H_E within each population.

Population structure across the global dataset buffaloes was assessed using several methods. Pairwise population F_{ST} was calculated using Arlequin v3.5.2.2 (Excoffier and Lischer, 2010), and summarised with a neighbour-net network created in SplitsTree v4.14.4 (Huson and Bryant, 2006). Major axes of variation were generated via a multidimensional scaling (MDS) analysis using raw Hamming's distance to reduce the dataset to 20 dimensions via PLINK, observing major divisions in populations. Genetic clustering of individuals was achieved using ADMIXTURE v1.3.0 (Alexander, Novembre and Lange, 2009). This tests for the ideal number of unique ancestral genetic clusters (K) across the dataset. Values of K from 1 to 40 were tested along with a cross-validation method implemented within ADMIXTURE that repeated each value of K five times to identify the most efficient value of K that explained the data. An AMOVA was used to further understand the distribution of variation across the dataset in Arlequin. Treemix was used to detect the presence of gene flow between populations by incrementally adding migration events between populations (Pickrell and Pritchard, 2012). Like ADMIXTURE, this tests between a defined number of migrations (K) however, the ideal value of K was selected when variation explained across the dataset exceeded 99.8% in accordance with Pickrell and Pritchard (2012). Values of K were tested from 1 to 15.

5.3.3. Approximate Bayesian Computation

Evolutionary modelling was conducted using Approximate Bayesian Computation (ABC) pipeline implemented in ABCtoolbox (Wegmann *et al.,* 2010). A series of models were defined depicting a split between river and swamp buffalo followed by two further splits within river buffalo (Figure 5.1). These additional splits represent dispersal of river buffalo out of India into the Middle East and Europe. Parameters were established to randomly sample values for divergence times, effective population sizes, and migration rates. Details of these can be found in Supplementary Table S5.1. FastSimCoal v2.6 was used to generate backward coalescent simulations (Excoffier *et al.*, 2021). 1,000,000 simulations were generated for each scenario simulating 2,500 unlinked SNPs, and summary statistics (observed data results across summary statistics found in Supplementary Table S5.2) were calculated in arlsumstat v3.5.2.2 (Excoffier and Lischer, 2010). The closest 1,000 simulations by distance to the observed data were retained to calculate posterior distributions. Generation of posterior distributions was completed using ABCestimator within ABCtoolbox. Highly linked summary statistics identified by Spearman's ranked correlation including Bonferroni's correction for multiple comparisons were removed. Model fit was assessed by ensuring observed summary statistics fell within the 95% quantiles of the simulated distribution for each statistic. A GLM computing the likelihood of observed data compared to likelihoods of each retained simulation was carried out to further assess model fit (Leuenberger and Wegmann, 2010). A posterior probability p-value was generated that represented the proportion of retained simulations with an equal or smaller likelihood than the observed data. Model discrimination was assessed by calculating the Bayes Factor (BF) by taking the quotient marginal densities of two models. A BF greater than 3 suggests strong support of the first model.

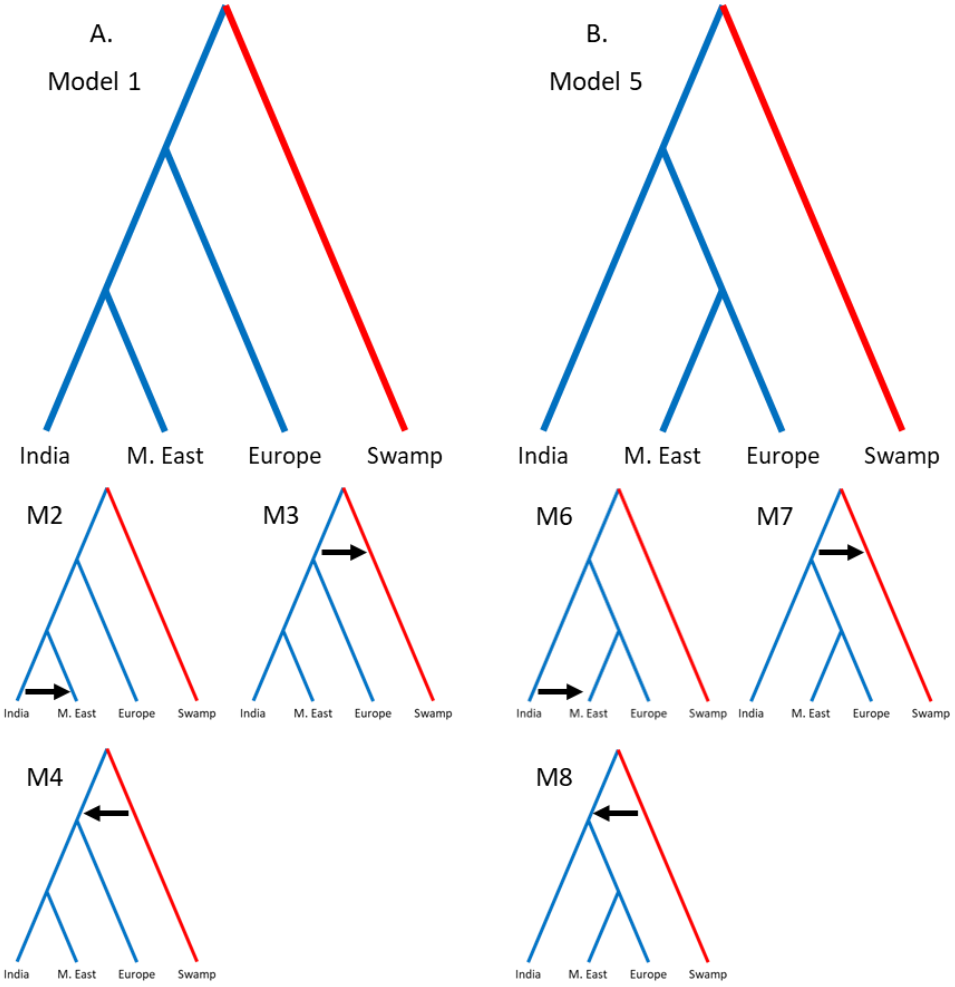


Figure 5.1: Model scenarios tested in ABC analysis. A. Model 1 shows two river (blue) buffalo dispersals out of India whereas B. Model 5 features one dispersal out of India followed by a split in this population as buffaloes reach Europe. Both models additionally estimate the divergence time between river and swamp buffaloes. Subsequent models tested presence of migration between river and swamp species, and from India to M. East post-split as indicated by the arrows.

5.3.4. Diverging Loci & Balancing Selection

River and swamp buffalo are genetically distinct, however are successfully able to hybridise to produce fertile offspring (ref). Loci under selection were identified between river and swamp buffalo using neutrality tests within Arlequin v3.5.2.2, via the command line executable arlsumstat (Excoffier and Lischer, 2010). The identification of diverging loci and loci that remain polymorphic in both species (balancing selection) was achieved by comparing F_{ST} and H_O values in the observed data against 1,000 bootstrap replicates of 50 individuals randomly sampled for river (25 samples) and swamp (25 samples) data. This enabled

detection of loci persistently under selection despite any underlying population within each species. H_O and F_{ST} of each locus were compared to coalescent simulations generated using a hierarchical island model to obtain a p-value. For each of the 1,000 bootstrap replicates, 20,000 coalescent simulations were generated to estimate p-values, simulating 2 groups of 25 demes each to match that of the observed data. No limits were set for expected heterozygosity or derived allele frequency for simulations. Proportion of differences was selected as the distance method for AMOVA computations. Average median p-value across all replicates was calculated for each locus and loci with a p-value <0.05 were deemed to be under selection.

5.4. Results

5.4.1. Genetic Diversity

Values of H_O, H_E, and F can be found in Table 5.1. H_O ranged from 0.244 (SWA_ID_NUT) to 0.449 (RV_PH_IMR). River buffaloes featured an average H_O of 0.387 (±0.029) that was significantly greater than swamp (H_O = 0.301 ±0.042; W = 380, p < 0.0001). River H_O ranged from 0.338 to 0.449 with those populations originating from India featuring the highest diversity closely followed by Middle Eastern populations. Populations of Mediterranean buffaloes originating from Europe typically possessed the lowest H_O of river buffaloes. Swamp buffalo H_O ranged from 0.244 (SWA_ID_NUT) to 0.430 (SWA_PH_ADM). Again, populations surrounding its domestication origin (e.g. SWA_TH_THS) featured higher H_O whereas those further afield (e.g. SWA_ID_NUT) featured lower H_O. H_O significantly differed across all (χ = 41195, df = 41, p < 0.0001), river (χ = 7425, df = 26, p < 0.0001), and swamp (χ = 8080, df = 14, p < 0.0001) populations. Similar patterns of variation were found across H_E that ranges from 0.236 (SWA_ID_NUT) to 0.417 (RV_IND). River buffalo H_E ranged from 0.298 (RV_MZ) to 0.417 (RV_IND), while swamp ranged from 0.236 (SWA_ID_NUT) to 0.397 (SWA_PH_ADM). River H_E was significantly greater than swamp (H_E = 0.290 ±0.035; W = 381, p < 0.0001). H_E significantly differed across all (χ = 44525, df = 41, p < 0.0001), river (χ = 10513, df = 26, p < 0.0001), and swamp (χ = 7382, df = 14, p < 0.0001). The majority of populations did not show significant differences between H_O and H_E. F was largely negative across all populations, ranging from -0.222 (RV_PH_IMR) to 0.004 (SWA_ID_SUW), indicating an absence of inbreeding within buffaloes.

Population ID	Species	Breed	Country	Sample Size	H_O	(±SD)	H_E	(±SD)	F	(±SD)	Reference
RV_UK1	River	Mediterranean	UK	39	0.367	±0.154	0.362	±0.139	-0.014	±0.049	
RV_UK2	River	Mediterranean	UK	27	0.362	±0.173	0.348	±0.148	-0.042	±0.083	
RV_IND	River	Murrah	India	130	0.420	±0.101	0.417	±0.092	-0.005	±0.029	
RV_BUL_VAR	River	Bulgarian Murrah	Bulgaria	58	0.406	±0.129	0.397	±0.112	-0.024	±0.043	(Noce *et al.,* 2021)
RV_GER_BOR	River	Mediterranean	Germany	28	0.380	±0.155	0.374	±0.132	-0.016	±0.055	
RV_GET_JUT	River	Mediterranean	Germany	27	0.382	±0.186	0.341	±0.150	-0.120	±0.104	
RV_GER_STAD	River	Mediterranean	Germany	26	0.364	±0.162	0.362	±0.139	-0.006	±0.074	
RV_GER_WIES	River	Mediterranean	Germany	28	0.390	±0.164	0.371	±0.136	-0.051	±0.058	
RV_HUN_CSAK	River	Mediterranean	Hungary	17	0.346	±0.184	0.338	±0.157	-0.025	±0.072	
RV_HUN_HT	River	Mediterranean	Hungary	19	0.364	±0.194	0.336	±0.160	-0.082	±0.057	
RV_HUN_TISZ	River	Mediterranean	Hungary	19	0.340	±0.209	0.311	±0.173	-0.092	±0.080	
RV_ROM_MERA	River	Mediterranean	Romania	16	0.372	±0.171	0.366	±0.138	-0.016	±0.083	
RV_ROM_SERC	River	Mediterranean	Romania	47	0.375	±0.165	0.355	±0.145	-0.054	±0.043	
RV_BR_MUR	River	Murrah	Brazil	15	0.410	±0.155	0.401	±0.112	-0.022	±0.067	(Colli *et al,* 2018)
RV_COL	River	-	Colombia	12	0.422	±0.165	0.405	±0.106	-0.042	±0.025	
RV_EG	River	Egyptian	Egypt	16	0.392	±0.166	0.380	±0.127	-0.037	±0.084	
RV_IR_AZA	River	Azari	Iran	9	0.395	±0.195	0.374	±0.134	-0.057	±0.036	
RV_IR_KHU	River	Khuzestani	Iran	10	0.373	±0.189	0.368	±0.138	-0.013	±0.106	
RV_IR_MAZ	River	Mazandarani	Iran	8	0.373	±0.213	0.350	±0.153	-0.068	±0.056	
RV_IT_MED	River	Mediterranean	Italy	15	0.364	±0.179	0.355	±0.145	-0.026	±0.043	
RV_MZ	River	Mediterranean	Mozambique	7	0.338	±0.243	0.298	±0.179	-0.135	±0.082	

RV_PH_BMR	River	Bulgarian Murrah	Philippines	11	0.431	±0.187	0.392	±0.119	-0.121	±0.169
RV_PH_IMR	River	Murrah	Philippines	6	0.449	±0.246	0.371	±0.141	-0.222	±0.125
RV_PK_KUN	River	Kundhi	Pakistan	10	0.420	±0.183	0.394	±0.118	-0.066	±0.021
RV_PK_NIL	River	Nili-Ravi	Pakistan	15	0.426	±0.153	0.404	±0.106	-0.055	±0.035
RV_ROM	River	Mediterranean	Romania	13	0.405	±0.216	0.353	±0.152	-0.159	±0.158
RV_TR_ANA	River	Anatolian	Turkey	15	0.395	±0.160	0.392	±0.119	-0.008	±0.037
SWA_BR_MUR	Swamp	Carabao	Brazil	10	0.324	±0.228	0.289	±0.177	-0.123	±0.081
SWA_CN_ENS	Swamp	-	China	15	0.292	±0.198	0.287	±0.173	-0.017	±0.038
SWA_CN_FUL	Swamp	-	China	15	0.296	±0.199	0.286	±0.173	-0.033	±0.013
SWA_CN_GUI	Swamp	-	China	11	0.291	±0.202	0.288	±0.173	-0.008	±0.095
SWA_CN_HUN	Swamp	-	China	15	0.305	±0.194	0.295	±0.167	-0.035	±0.048
SWA_CN_YAN	Swamp	-	China	14	0.314	±0.205	0.294	±0.168	-0.069	±0.107
SWA_CN_YIB	Swamp	-	China	15	0.293	±0.197	0.287	±0.173	-0.022	±0.022
SWA_ID_JAV	Swamp	-	Indonesia	13	0.265	±0.212	0.256	±0.186	-0.034	±0.121
SWA_ID_NUT	Swamp	-	Indonesia	7	0.244	±0.234	0.236	±0.197	-0.037	±0.089
SWA_ID_SUM	Swamp	-	Indonesia	13	0.307	±0.203	0.295	±0.169	-0.044	±0.114
SWA_ID_SUW	Swamp	-	Indonesia	11	0.255	±0.212	0.257	±0.188	0.004	±0.105
SWA_PH	Swamp	-	Philippines	15	0.281	±0.193	0.280	±0.172	-0.003	±0.116
SWA_PH_ADM	Swamp	-	Philippines	10	0.430	±0.193	0.397	±0.114	-0.084	±0.098
SWA_TH_THS	Swamp	-	Thailand	6	0.321	±0.226	0.300	±0.168	-0.069	±0.092
SWA_TH_THT	Swamp	-	Thailand	8	0.298	±0.205	0.298	±0.168	0.000	±0.122

Table 5.1: Domestic water buffalo populations used in study including metadata and calculated summary statistics.

5.4.2. Population Structure & Gene Flow

F_{ST} values across the dataset ranged from 0.004 (SWA_CN_GUI vs SWA_CN_YIB) to 0.432 (SWA_ID_NUT vs RV_MZ). Within river buffaloes these ranged from 0.009 (RV_HUN_CSAK vs RV_HUN_HT) to 0.219 (RV_HUN_TISZ vs RV_MZ), averaging 0.094 (±0.043). Comparisons within swamp buffalo showed similar levels of differentiation, averaging 0.095 (±0.058) and ranging from 0.004 (SWA_CN_GUI vs SWA_CN_YIB) to 0.243 (SWA_BR_MUR vs SWA_ID_NUT). F_{ST} values were greatest between river and swamp buffaloes ranging from 0.100 (SWA_PH_ADM vs RV_CO) to 0.432 (SWA_ID_NUT vs RV_MZ) with an average of 0.309 (±0.056). A network summarising an F_{ST} distance matrix can be found in Figure 5.2. River and swamp buffaloes can be found an opposing ends of the network with admixed swamp populations of SWA_BR_MUR and SWA_PH_ADM found centrally. Within the river buffalo cluster, RV_IND population generated within this thesis was found at the core of the cluster. Populations of Indian origin and Pakistan populations were found nearby, while Middle Eastern populations formed their own cluster. European populations were found to be more distantly related to Indian populations than Middle Eastern populations with greater subdivision as Eastern (RV_HUN & RV_ROM populations) and Western clusters (RV_IT & RV_UK populations) were visible. Thai swamp populations were found at the core of the swamp cluster with Chinese populations clustering together whereas Southernly Indonesian populations (SWA_ID_NUT, SWA_ID_JAV, SWA_ID_SUW) were found more distantly related. All pairwise F_{ST} comparisons were significant bar SWA_TH_THS vs SWA_TH_THT.

ADMIXTURE results revealed populations structuring across river and swamp buffaloes (Figure 5.3 & Figure 5.4). K = 2 separated river and swamp buffaloes with admixed populations of SWA_BR_MUR and SWA_PH_ADM evident from partial river ancestry. The following values of K distinguished various European populations (K = 4 split East and West European populations) until K = 7 where the first within-swamp clustering occurred via separation of three Indonesian populations. At K = 10, Middle Eastern river populations clustered together away from Indian and Pakistan populations. At K = 11 and K = 12, Brazilian carabao of SWA_BR_MU and Philippine swamp SWA_PH became distinct, respectively. Although K = 24 is the most efficient clustering solution (Supplementary Figure S5.3), CV values do not improve much following K = 12, with ADMIXTURE plots becoming biologically

obscure (Supplementary Figure S5.4). MDS analysis (Supplementary Figure S5.6) produced three major components of variations (eigenvalue > 1) with splits similarly captured by F_{ST} and ADMIXTURE. Component 1 captured 56.6% of variation across the dataset distinguishing between river and swamp species. Admixed populations of SWA_BR_MUR and SWA_PH_ADM were found with intermediatory values. Component two captured 11.3% of variation, separating European from Asian populations. Component three (6.6%) further separated European populations into East and West.

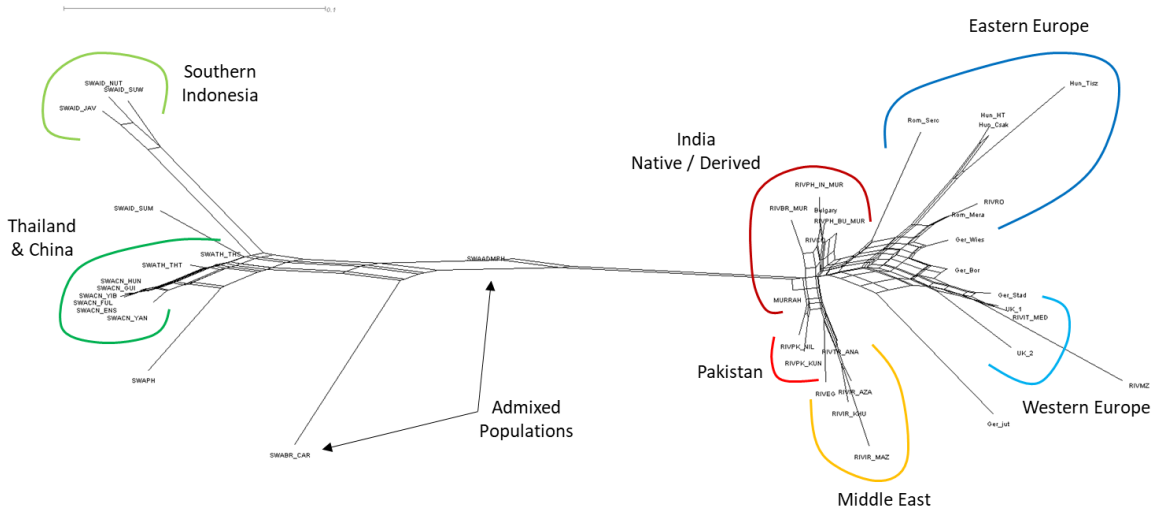


Figure 5.2: Neighbour-net network based upon pairwise population F_{ST} scores. River (right) and swamp (left) show substantial genetic differences, with known hybrid populations occurring centrally. Coloured lines indicate the main genetic clusters.

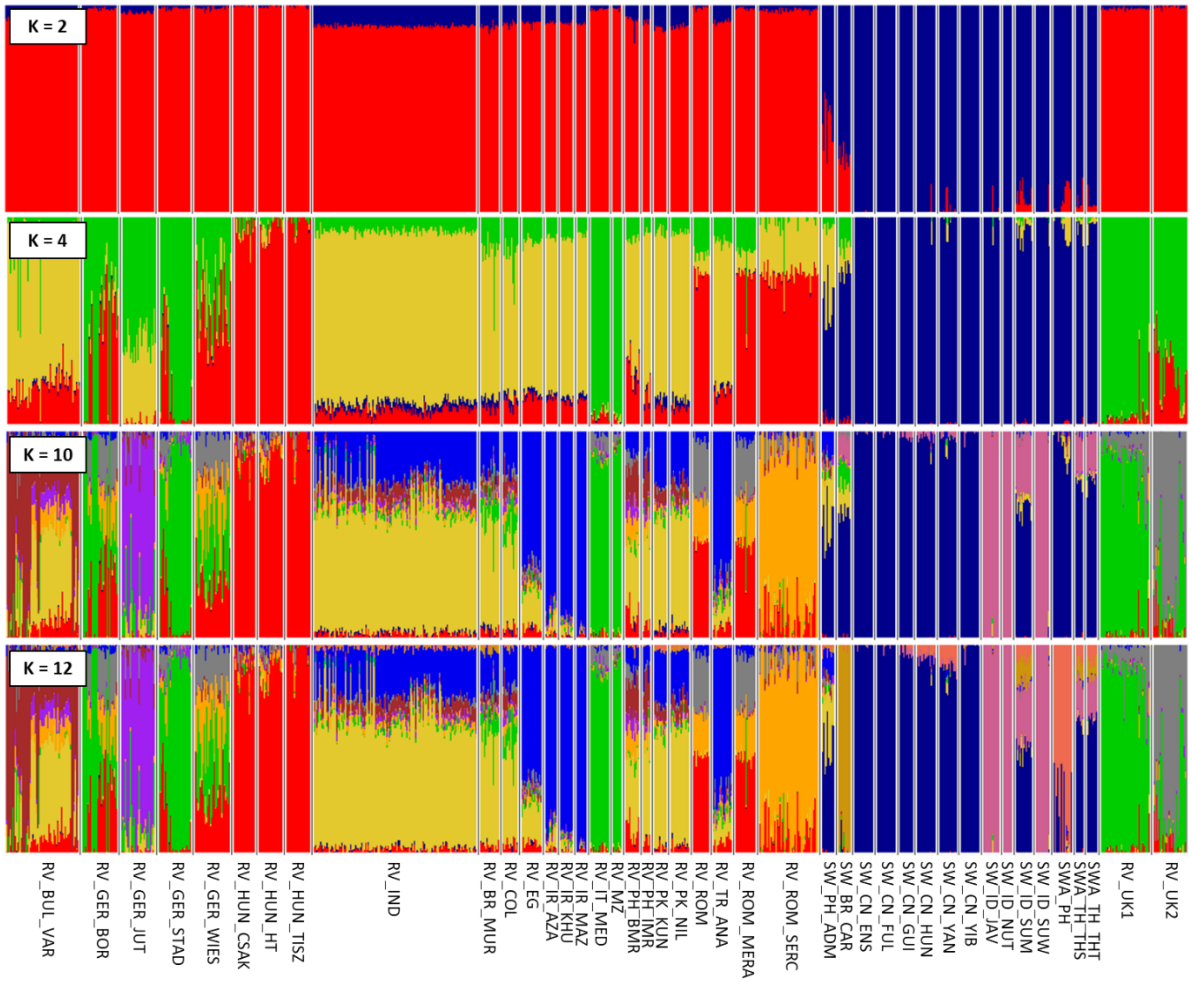


Figure 5.3: Population clustering analysis from ADMIXTURE. K = 2 split river and swamp buffalo. K = 4 showed separation into East and West within Mediterranean populations. Further separation occurred within river buffalo until K = 10 where swamp buffalo began separating via Indonesian populations. CV values and biological interpretation did not radically improve after K = 12. K = 24 was the best fitting value of K in shown in Supplementary Figure S5.3. River buffalo indicated by prefix 'RV', and swamp buffalo prefix is 'SWA'.

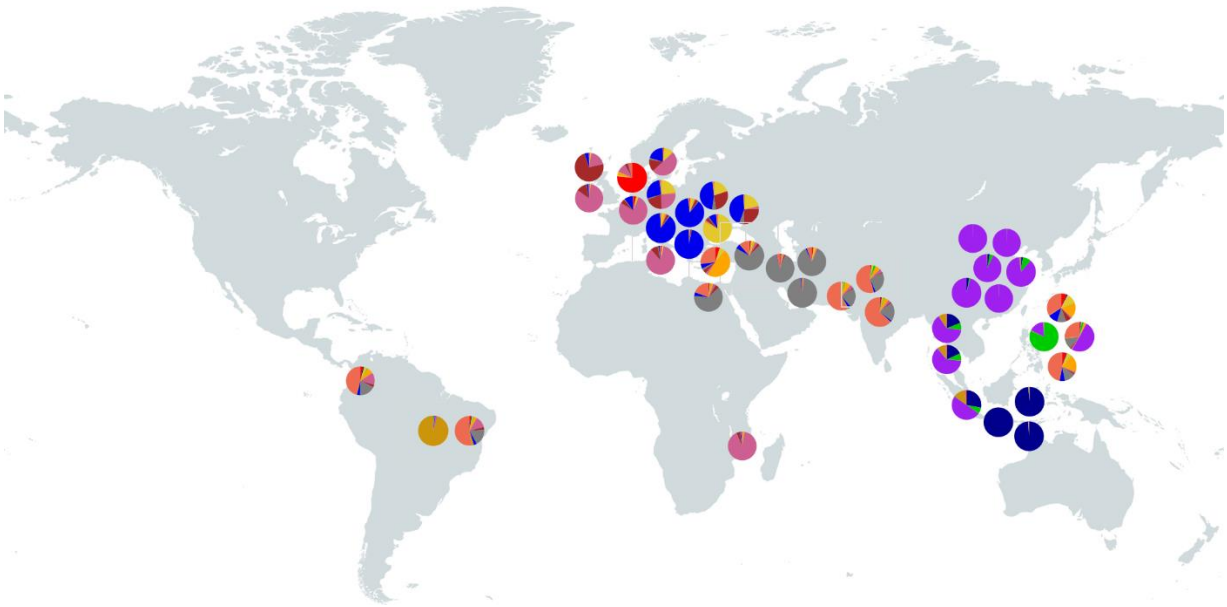


Figure 5.4: Map of average ancestry (ADMIXTURE results at K = 12; colours not matching Figure 5.3) across global buffalo populations. Charts are in their approximate geographic locations.

Treemix analysis elucidated pathways of migration within the dataset. 99.8% of variation was accounted for at K = 2 (Figure 5.5). Migration axes were found between from RV_PK_NIL to the admixed swamp population SWA_PH_ADM, and between German populations of RV_GER_JUT and RV_GER_BOR. Admixture detected in SWA_BR_CAR in MDS and ADMIXTURE analysis was identifiable at K = 3 in Treemix. Variance explained and log likelihoods for all values of K are shown in Supplementary Table S5.5. AMOVA results indicated that the majority of variation is held within individuals (70.7%) followed by across groups (e.g. river and swamp) (23.2%) as shown in Supplementary Table S5.7.

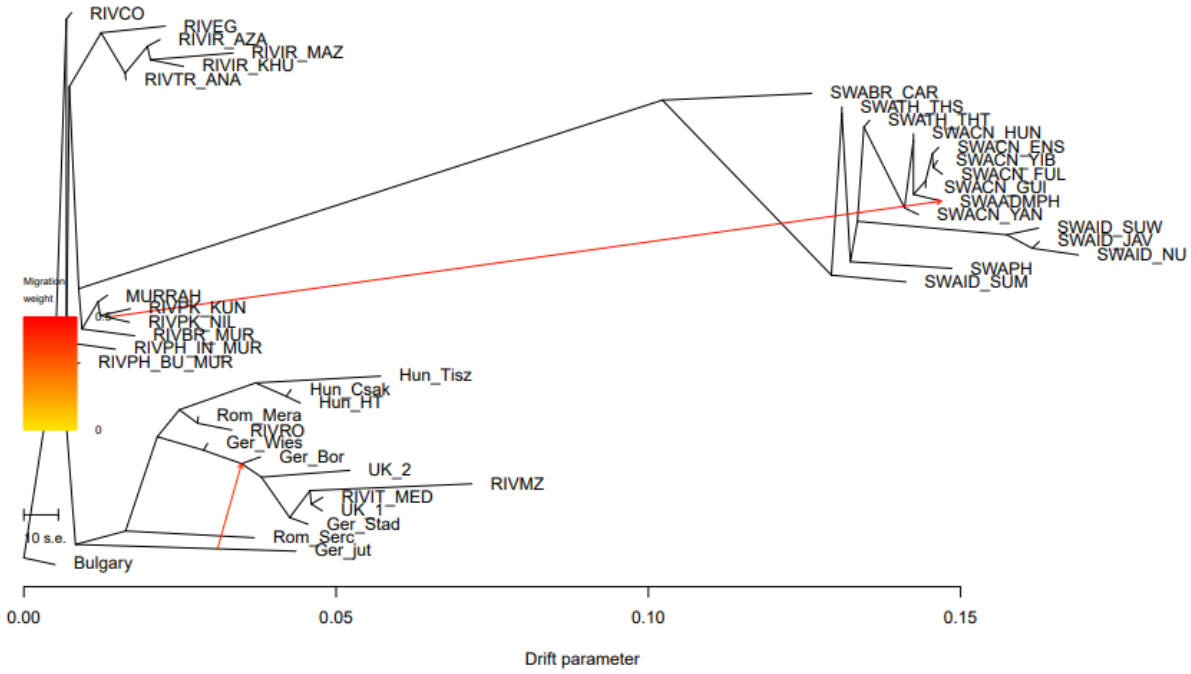


Figure 5.5: Treemix model that accounted for 99.8% of variation. Two migration edges were found in the preferred model. The only migration found between river and swamp involved the hybrid swamp population.

5.4.3. Evolutionary Modelling

Reconstruction of buffalo demographic history was achieved using an ABC pipeline incorporating backwards coalescent simulations. 17 summary statistics were retained, with those being heterozygosity (total and per population), and pairwise F_{ST}. All retained observed statistics fell within the 95% quantiles of simulated statistics across all scenarios. Model 4 consistently outperformed all scenarios in terms of Bayes Factor (>3). Posterior distributions for model 4 are shown in Figure 5.6. Indian murrah buffalo featured the largest N_e at $10^{5.76}$ while Iran ($10^{3.63}$), Italy ($10^{3.31}$), and swamp ($10^{3.21}$) buffalo populations were all far smaller. Moving backwards from India to ancestral populations, N_e declined prior to river and swamp divergence ($10^{5.98} > 10^{3.39} > 10^{2.42}$). Iranian and Italian populations dispersed separately out of India. Italian ancestors diverged from India approximately 337 generations ago, while Iranian divergence from India was more recent at 194 generations. River and swamp buffalo featured an older divergence at 2451 generations ago, with detection of swamp migration into river buffalo.

Table 5.2: Resulting p-values and marginal densities of each simulated model. The best performing model was identified using Bayes Factor (MD(A)/MD(B)) greater than 3 (in bold). N.D = Number of Dispersals, Mig = Migration Direction, P = P-value, MD = Marginal Density

N.D	Mig	Model	P	MD	Bayes Factor							
					1	2	3	4	5	6	7	8
Two	None	1	0.495	0.194	-	0.039	0.024	0.000	0.119	0.024	0.006	0.421
	India > Iran	2	0.293	4.939	**25.425**	-	0.610	0.007	**3.024**	0.607	0.140	**10.692**
	River > Swamp	3	0.518	8.103	**41.707**	1.640	-	0.011	**4.961**	0.996	0.229	**17.539**
	Swamp > River	4	0.411	735.032	**41.707**	**148.807**	**90.713**	-	**450.048**	**90.327**	**20.816**	**1590.978**
One	None	5	0.386	1.633	**8.407**	0.331	0.202	0.002	-	0.201	0.046	**3.535**
	India > Iran	6	0.418	8.137	**41.885**	1.647	1.004	0.011	**4.982**	-	0.230	**17.613**
	River > Swamp	7	0.330	35.311	**181.755**	**7.149**	**4.358**	0.048	**21.620**	**4.339**	-	**76.431**
	Swamp > River	8	0.396	0.462	2.378	0.094	0.057	0.001	0.283	0.057	0.013	-

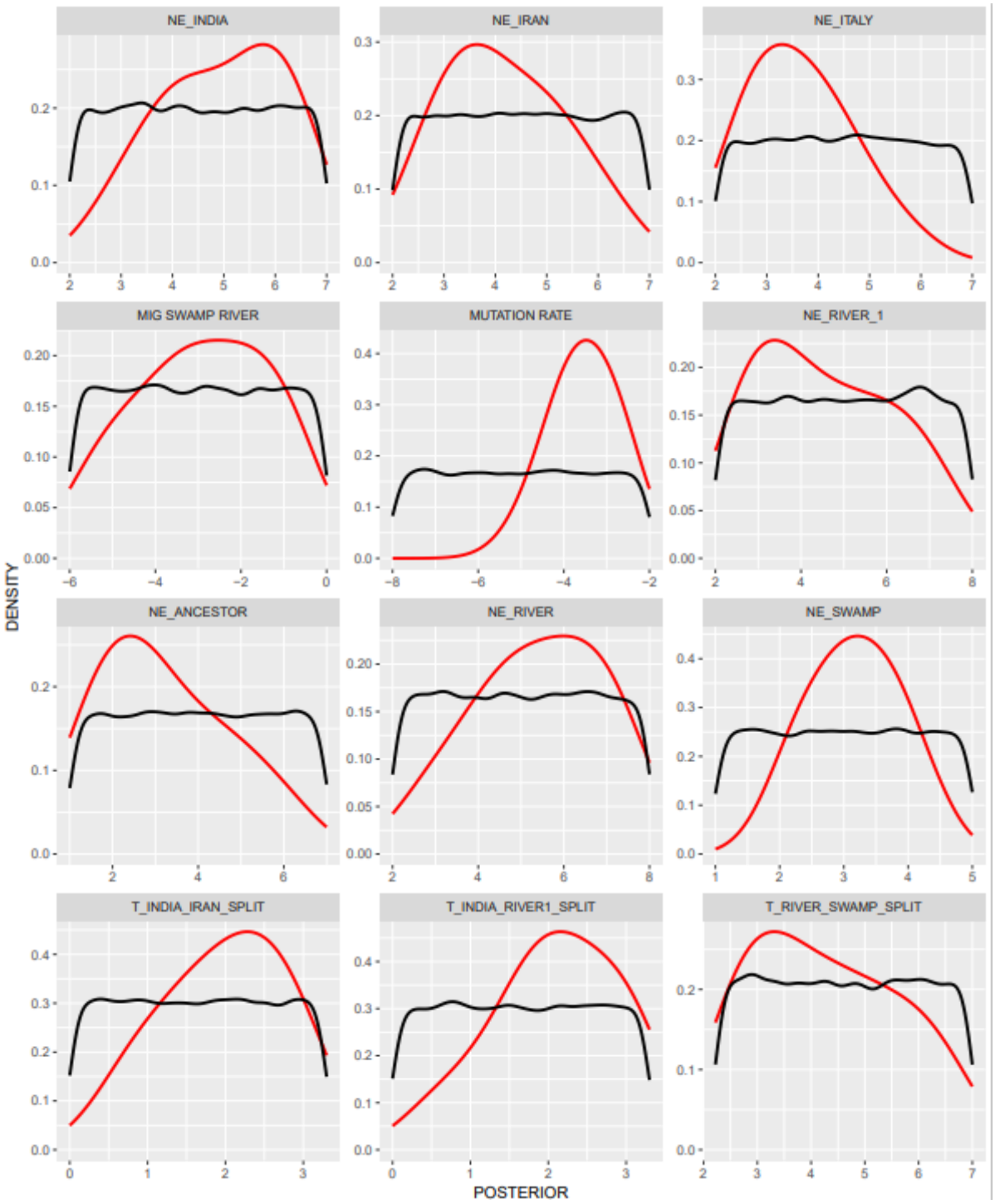


Figure 5.6: Posterior distributions (red) for the best performing model (4) that simulated divergence of river and swamp buffalo followed by two dispersals of river buffaloes out of India, with additional migration occurring from swamp to river buffalo. Black lines indicate the prior distribution. NE = Effective population size, MIG = Migration Rate, and T = Divergence time.

5.4.4. Selection between Domestic Buffaloes

Testing for selection between river and swamp contributed to unearthing the genetic mechanisms behind an older divergence but maintaining two compatible species for interbreeding. 88 SNPs proved significant ($p < 0.05$) with 25 showing high F_{ST} scores suggesting divergent selection, and 63 with low F_{ST} under balancing selection (Supplementary Table S5.8). Approximately half of SNPs were found within genes with 14/25 (56%) SNPs within genes for divergent selection, and 30/63 (48%) for balancing selection. Figure 5.7 shows the distribution of SNPs for F_{ST} and H_O values. Genes linked with SNPs with high F_{ST} values can be attributed to livestock function such as milk production (NCKAP5, CAMK2D, PDE4D) and sperm quality (MARCH1, PCSK6) in line with river buffalo function. Whereas a number of genes related to neuronal (NLGN1, TMEM156) and body growth (DLG2, FGF3, BMP2, RBMS3, SNX29) fitted swamp buffalo draught function. In contrast, genes under balancing selection related to survival notably in immune genes (NCAM2, TREM1, SPPL2A, OCLN), sensory (OTOL1, TMEM132E, RGS7BP), reproductive viability (WIF1, CRISPDL1, PDLIM1), and examples of genes where heterozygote advantage prevents lethality (RNASEH2B, MAJIN).

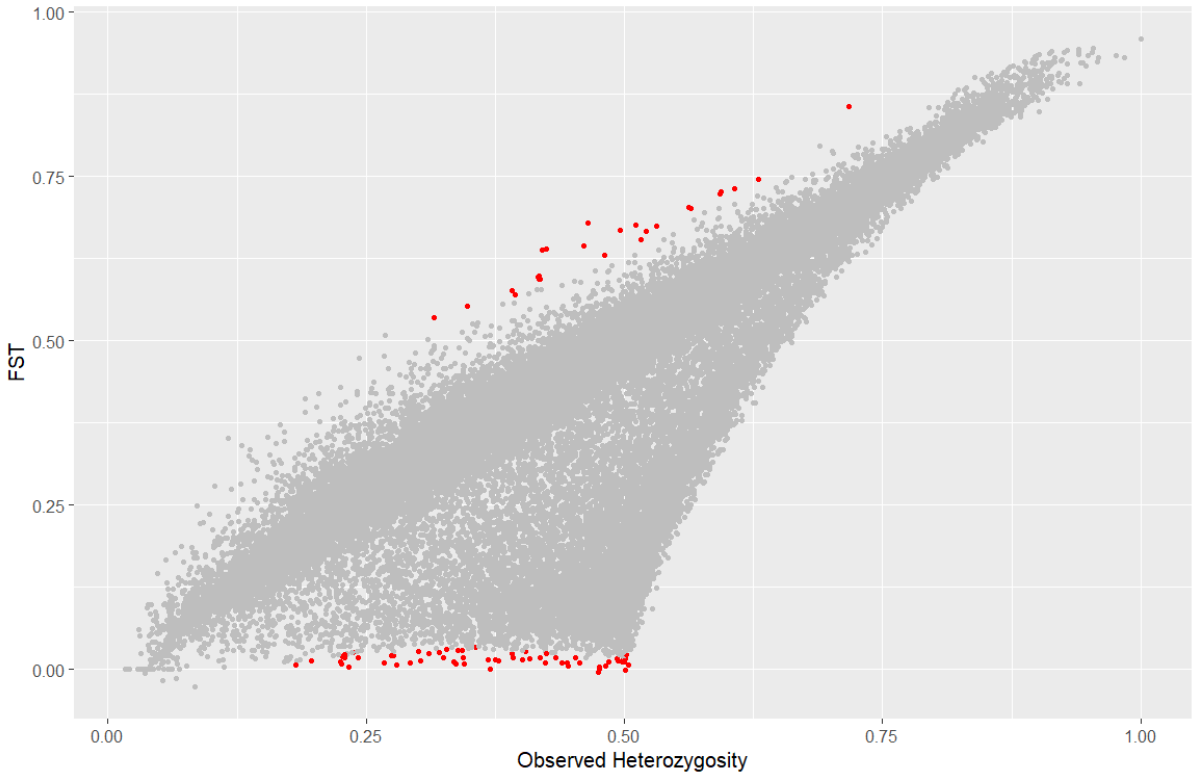


Figure 5.7: Significant loci under selection between river and swamp buffalo via comparison of observed F_{ST} and H_O to neutral coalescent simulations. Red points indicate SNP markers significantly under selection ($P<0.05$).

5.5. Discussion

Uncovering domestic water buffaloes' evolutionary history will provide a variety of benefits for farming of this livestock species. Quantifying genomic resources across the agricultural industry is crucial for safeguarding genetic variability for future use (Taberlet *et al.*, 2011; Biscarini *et al.*, 2015; Bruford *et al.*, 2015; Eusebi, Martinez and Cortes, 2020; Gicquel *et al.*, 2020). Increasing intensification of the livestock industry favoring few beneficial commercial breeds is leading to genetic erosion through extinction of less productive individuals and breeds (FAO, 2007; Yaro *et al.*, 2017; Eusebi, Martinez and Cortes, 2020). Reduced genetic diversity often leads to reduced adaptability in species, and with increasing consequences of climate change, variability in livestock will be essential to breeding resilient stock (Naskar *et al.*, 2012; Eusebi, Martinez and Cortes, 2020; Rovelli *et al.*, 2020). Modelling domestic buffaloes' evolutionary history and understanding genetic variation across the species will aid in identifying key regions of the genome that may be of adaptive importance, be it environmental or productivity. This knowledge will help manage each species correctly and advise on beneficial crossbreeding that increases productivity while maintaining adaptive traits (Eusebi, Martinez and Cortes, 2020). Additional knowledge may also be revealed such as elucidating societal and cultural dynamics in historic human populations (Felius *et al.*, 2014; Scheu *et al.*, 2015; Almathen *et al.*, 2016; Colli *et al.*, 2018). This chapter quantified genetic variation across global buffalo populations, modelled river and swamp buffalo evolution, and provided functional evidence behind the genomic discrepancies between buffalo species.

5.5.1. Demography

The two species of domestic water buffaloes provide important livestock functions to the communities they serve. The dairy producing river buffalo provides a vital source of food security in marginal lands where crop growth is not reliable, meanwhile swamp buffaloes are an excellent means of draught across agricultural wetlands such as rice paddies (Zhang, Colli and Barker, 2020). The two species are physiologically, genetically, and geographically distinct, though the ability to successfully hybridise leaves their evolutionary history unclear

(Zhang, Colli and Barker, 2020). The addition of Indian river buffalo to the global buffalo array dataset provided important representation for the putative domestication centre for river buffalo. Trends in genetic diversity in livestock species typically decreases away from domestication centres (Bruford, Bradley and Luikart, 2003; Groeneveld *et al.*, 2010). This is due to subsets of variation being captured as new populations are often established on small founding populations. Domestic buffaloes follow this trend. Colli *et al,* (2018) showed that river buffaloes genetic diversity decreased away from India and into Europe. The same is found in mitochondrial data as Northwest India possesses the greatest diversity of haplotypes (Kumar *et al.*, 2006; S. Kumar *et al.*, 2007). Here, Indian and Pakistan populations all featured the highest H_O (> 0.4). H_O can be affected by sample size, and although the large RV_IND population does not feature the highest H_O, it does possess the highest H_E (0.417). Thus, the newly sampled Indian murrah population fulfills expectation that genetic diversity is greatest where domestication originated. The addition of UK buffaloes provided populations at the very edge of river buffaloes historic range, and again, in line with expectations UK buffaloes were among the lowest H_O & H_E with only Hungarian and Mozambique populations found with lower. Historic populations of Italy, Romania, and Middle East were found with intermediate levels of genetic diversity between Indian and fringe European populations.

Ascertainment bias within the array means that river buffalo diversity is likely inflated compared to sequencing data, whilst swamp diversity is underrepresented in comparison to river results (Iamartino *et al.*, 2017; Colli *et al.*, 2018). Chapter Four revealed that selection of increasingly more polymorphic markers can remove this difference, although this may distort the true pattern of genetic diversity across domestic buffaloes. A MAF cutoff was implemented in swamp buffalo that increased their genetic diversity comparatively compared to the original dataset in Chapter Four. Importantly this maintained the trend that river buffaloes possess greater levels of diversity than swamp buffalo as observed in previous studies (Colli, Milanesi, Talenti, *et al.*, 2018; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020). The dataset generated here maintained the trends in genetic diversity of swamp buffalo found in Colli *et al,* (2018) where diversity is greatest in continental Southeast Asian, with decreasing diversity in more isolated island populations of Philippines and Southern Indonesia. Despite lower estimates of diversity, no populations appear to suffer from any inbreeding with negative inbreeding coefficients found across the majority of populations. This may instead

indicate an isolated breaking effect driven by some admixture (Wahlund, 1928; Šnjegota *et al.*, 2021; Kanaka *et al.*, 2023). Domestic buffaloes have largely been farmed in low external input systems with little intensification and selection on populations. Although defined breeds do exist, buffaloes are often allowed to mix with other individuals leading to high numbers of non-descript and graded buffaloes (DAHD, 2023). Additionally, the overlap in domestic and wild buffalo ranges means that genetic exchange between these species is likely to have occurred far beyond initial domestication times (Satish Kumar *et al.*, 2007; Choudhury, 2014; Nagarajan, Nimisha and Kumar, 2015; Kandel *et al.*, 2019). Such opportunities for admixture within buffaloes and an absence of intensification may explain the little evidence of inbreeding.

The addition of Indian buffalo along with UK and other European populations revealed further variability across river water buffalo that was previously undetected(Noce *et al.*, 2021). Across the dataset, four main genetic clusters were found. Those of Indian and Pakistan origin clustered together in $F_{ST,}$ ADMIXTURE and MDS analysis. Populations that were recently exported e.g., South America and Philippine Indian originating populations showed little evidence of substantial divergence (average $F_{ST} < 0.03$ in comparison to RV_IND and RV_PK). The Bulgarian murrah (RV_BUL_VAR) cross revealed more variation owing to mixed ancestry of Mediterranean and murrah breeds as seen in MDS and ADMIXUTRE results (Borghese, 2013a). Middle Eastern populations of Iran, Egypt, and Turkey formed their own cluster away from European populations suggesting that these populations are unrelated despite being geographical neighbours. Domestic buffaloes depend upon nearby water sources for thermoregulation as they lack sweat glands (Marai and Haeeb, 2010; Yáñez *et al.*, 2020). The arid landscape west of the Indus valley is unlikely to be traversable for river buffaloes, unlike oxen. Instead, Colli *et al,* (2018) hypothesized a maritime pathway between India and Middle East. The dissimilarity found between Middle Eastern and European populations may be explained through replenishment of genetic diversity from new trade from India. Documentation of buffalo trade is rare with little apparent cultural importance outside of India. Until the mid-20th century, the Southern wetlands of Iraq was an important place of historic buffalo farming, with buffaloes maintaining waterways for transport, controlling vegetation through grazing, and economic gain through milk production (Fazaa, Dunn and Whittingham, 2018). The Basra region was economically crucial throughout Late Antiquity

and a key port for Indian Ocean trade (Simpson, n.d.). These active trade routes may have led to translocation of domestic buffaloes. It is known that Indian traders were able to access the Middle East via Basra and formed trade routes up the Euphrates and Tigris rivers (Simpson, n.d.). If Middle Eastern populations represent a new influx of diversity, European populations may indicate an original historic dispersal out of India. The division of European populations into East and West indicates separate transport routes across the Mediterranean to the wetlands of Italy and Danube delta. Despite all European buffaloes falling under the Mediterranean breed, these East and West populations appear genetically distinct with little genetic evidence of interaction post-split. Europe was already dominated by productive livestock such as taurine cattle, and resilient species such as sheep and goats for less productive environments. Water buffalo were likely inferior to these established species and of little use outside of wetlands. German populations showed admixture between East and West Mediterranean populations that may be due to establishment of farms using cheap Eastern buffalo before uptake of more expensive and productive Italian buffalo.

Population structure results detected the same structuring pattern across swamp buffalo as Colli *et al,* (2018), however the variation contributions from swamp buffalo were far smaller (e.g. clustering occurred at higher values of K in ADMIXTURE). This may be due to the greatly increased number of river buffaloes added to the dataset (Colli river n = 155, here river n = 643), along with the variability of the buffalo array being dedicated to river buffalo. Patterns of variation across swamp buffalo are largely geographic. Zhang *et al,* (2016) found that mitochondrial haplotypes were highly structured across Southeast Asia. Dispersal of swamp buffalo may have only occurred in tandem with rice cultivation (Zhang *et al.*, 2016). Thai populations are found close to the center of swamp buffalo clusters in F_{ST} network, and along with higher H_O, this suggests a potential location for swamp domestication. Analysis of swamp mitochondrial and Y-chromosome haplotypes across Southeast Asia found that diversity was greatest in surrounding Northern Thailand and Southern China (Zhang *et al.*, 2016; Wang *et al.*, 2017). Here, results revealed that Chinese populations cluster closely together, while Southern Indonesian populations are more distantly related which is likely due to small, isolated populations being susceptible to genetic drift (Colli, Milanesi, Vajana, *et al.*, 2018).

Patterns of genetic variation across domestic water buffalo shows several trends such as genetic diversity being greatest in river than swamp, and greatest in India and Northern Thailand, river and swamps respective domestication origins (Satish Kumar *et al.*, 2007; S. Kumar *et al.*, 2007; Zhang *et al.*, 2011, 2016; Colli, *et al.*, 2018; Luo *et al.*, 2020; Sun *et al.*, 2020). Questions remain as to timings of divergence and levels of gene flow between species and populations. Scenarios modelled tested the divergence times of river and swamp species, and level of migration between species. The scenarios further modeled river buffaloes' historic dispersal out of India, and into the Middle East and Europe. Due to ascertainment bias of the buffalo array and potential effects of over parameterization in ABC, swamp buffalo dispersals were not further modelled (Csilléry *et al.*, 2010; Quinto-Cortés *et al.*, 2018).

Model 4 was the most favorable outcome of ABC analysis. This model encompassed a river-swamp divergence with subsequent migration from swamp to river. River buffalo were found to have been exported out of India to the Middle East twice, led first by the ancestors of European populations. Replacement of this genetic diversity within the Middle East likely occurred via new supply of buffaloes from India to the Middle East. As stated previously, domestic buffaloes are greatly productive in tropical and sub-tropical environments, being resilient to higher temperatures, diseases and parasites, and effective on low-quality feed (Marai and Haeeb, 2010; Yáñez *et al.*, 2020). However, due to the inability to sweat, buffaloes require a source of water for thermoregulation (Marai and Haeeb, 2010; Yáñez *et al.*, 2020). This means that water buffaloes are more geographically restricted to lowland wetlands. The inability to venture long distances in drier environments means that separate buffalo populations are unlikely to interact without environmental change or translocation from humans, hence explaining genetic structuring between Europe, Middle East, and Asia (plus river between swamp). Although modelling of further swamp populations does not occur here, current literature suggests that there is geographical sub structuring in swamp buffalo that may be due to isolation in swamp buffalo populations by river basins or local geographical factors (Zhang *et al.*, 2016). Within this chapter, this is observable from Southeast Asian island populations. River buffalo meanwhile were historically isolated to wetland environments across India (Indus valley), Middle East (Fertile Crescent), and Europe (Pontine Marshes and Danube), all of which are accessible through known maritime trade routes.

The dispersal of river buffalo from India occurred within the past 6,000 years, the domestication point for river buffalo (Zeuner, 1963; Patel & Meadow, 1998; Satish Kumar *et al.*, 2007). From mitochondrial analysis, it is known that all river buffaloes originally derive from India (Coroian *et al.,* 2015; Zhang, Colli and Barker, 2020). The latter divergence between river and swamp species is expected to date prior to the Holocene although precise estimates are yet to be determined (Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021). ABC results estimated divergence times for within river buffaloes to be ~2,000 YA for initially dispersal out of India to the Middle East of which European buffaloes are derived from. This genetic distinction between European and Middle Eastern populations is likely due to replacement of genetic diversity within the Middle East. Estimates from ABC results suggest Middle East populations diverged from India more recently at ~1,150 YA. A few hypotheses occur as to river buffalo dispersal through the Middle East and into Europe involving Crusaders, gifts to the King of Lombardy, and Saracen traders (Colli *et al.*, 2018). Wordsworth *et al,* (2021) analysed historic buffalo remains in the Middle East and dated the occurrence of buffaloes to the region and alluded to mechanisms of arrival. The arrival of Indian people called 'al Zutt' into Southern Mesopotamia was identified. These people reached Antakya (ancient Antioch) under the Umayyad Caliphate by the end of Late Antiquity (Simpson, n.d.). These settlers likely brought native livestock, probably domestic river buffalo. Following initial arrival in Southern Mesopotamia early in the Common Era, river buffalo likely dispersed along the Euphrates and Tigris rivers towards Europe. Reaching Antakya by the onset of the 9th century, river buffaloes likely crossed the Mediterranean along Saracen trade routes (Colli *et al.*, 2018). Estimates of N_e trends found in both UK buffalo using linkage disequilibrium inferred a divergence time of ~1,500 YA, further supporting ABC results (Chapter Two). European buffaloes may well derive from the original dispersal of domesticated river buffalo out of India. Current Middle Eastern buffalo appear genetically more similar to Indian buffalo, indicating that an influx of genetic diversity occurred within the region displacing original European ancestors (Colli *et al.*, 2018). Maritime trade routes between India and Middle East were a significant source of commerce between the regions, and water buffalo may have been continually traded following initial arrival (Simpson, n.d.). The port of Basra was a prominent gateway for Indian Ocean trade into Mesopotamia.

Current NGS estimates of river and swamp buffalo divergence range up to 800,000 YA (Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021). The extensive genetic differentiation between the two species suggests an absence of interaction post divergence. The preferred model here supports some swamp to river migration. The time of divergence from ABC results here estimated ~14,700 YA, prior to the onset of the Younger Dryas and Holocene. This result maintains two independent domestication events for river and swamp (splitting time pre-domestication). However, this presents the most recent divergence time in current literature. Recency of the estimate may be down to the data used within the study. Chapter Four showed that the use of SNPs that were polymorphic across both species resulted in more recent estimates of divergence between river and swamp. Here, SNPs that were monomorphic/lowly polymorphic in swamp were removed to overcome some bias towards river buffaloes. Consequently, this likely selected ancestral SNPs shared by both species and those retained under balancing selection, thus the two species appeared more closely related. Faster mutation rates permitted here may also reduce splitting times. With a p-value of 0.411, there is greater scope to refine the ABC model. This data type may also underlie migration found as all the swamp variation is present in river buffalo. Mimicking SNP discovery (and therefore ascertainment bias)in the coalescent simulations like in Quinto-Cortés *et al.*, (2018) may aid in producing more realistic model outputs for river and swamp comparison. Regardless of the timing of divergence,, it appears that river and swamp buffalo separation is linked to glacial periods within the Pleistocene (Mintoo *et al.*, 2019; Luo *et al.*, 2020). Lower sea levels surrounding Southeast Asia led to an increase in landmass as previous marine environments became occupied with humid and marshy vegetation (Wang *et al.*, 2009; Hamilton *et al.*, 2024). These new landmasses may have facilitated the dispersal of *Bubalus arnee* (or ancestral species) from Northern India into Southeast Asia by providing a bypass around the Indo-Burma Mountain ranges. This migratory route is hypothesised for the evolution of the anoa (Rozzi, 2017).

Estimates of N_e in the ABC analysis largely followed genetic diversity results. India featured the largest N_e at $10^{5.76}$ while Middle East, Europe, and swamp populations all feature smaller values of N_e at $10^{3.63}$, $10^{3.31}$, and $10^{3.21}$, respectively. Middle East and European populations likely captured a subset of genetic diversity from India, therefore smaller N_e are logical. Swamp N_e may be underestimated due to ascertainment bias in the array, or

conversely, river buffalo N_e being inflated. N_e typically declines back in time with estimates of river buffalo N_e changing from $10^{5.76}$ to, $10^{5.98}$, $10^{3.39}$ and $10^{2.42}$. Analysis of mitochondrial haplotypes in river buffalo shows that there has been continuous introgression from wild buffalo, and as such, predicting precise estimates of domestication in river buffalo is difficult (Nagarajan, Nimisha and Kumar, 2015). The resulting pattern is that N_e continually increases over time in river buffalo.

5.5.2. Genetic Divergence

Isolated populations will naturally diverge to become separate species. In practice, this requires sufficient genetic change between populations to become biologically incompatible. Genetic evidence indicates that domestic riverine and swamp buffalo are separate species (Wang *et al.*, 2017; Curaudeau, Rozzi and Hassanin, 2021). Within pure populations, there is an absence of shared mitochondrial and Y-chromosome haplotypes, chromosomal numbers differ, and they present high level of genomic differentiation (Barker *et al.*, 1997; S. Kumar *et al.*, 2007; Yindee *et al.*, 2010; Zhang *et al.*, 2011; Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Zhang, Colli and Barker, 2020; Curaudeau, Rozzi and Hassanin, 2021). In this study here, F_{ST} values (Section 5.4.1) between river and swamp populations similarly result in average values > 0.3 and the first division in clustering solutions consistently resolve riverine and swamp individuals (Section 5.4.2). Furthermore, ABC evolutionary modelling produces a probable divergence time before the Holocene, agreeing with independent domestication events for riverine and swamp buffaloes. Despite this however, riverine and swamp buffaloes are fully capable of hybridising indicating that on a cellular level, they remain similar.

More than twice as many SNPs were identified relating to balancing selection between buffalo species than divergent selection. This in part may be due to the ascertainment bias of the Axiom™ Buffalo Genotyping Array as all SNPs included in the panel are targeted at river buffalo. Therefore, without unique swamp buffalo variation, the array will not be able to detect further regions of divergence. Of the SNPs that were identified under divergent selection associated genes appear to relate to livestock function of each buffalo species. River buffalo were primarily domesticated for their milk production. PDE4D (Phosphodiesterase 4D) and CAMK2D (Calcium/Calmodulin Dependent Protein Kinase II Delta) were all found in divergent SNPs. PDE4D may impact fat yield in milk via cAMP signalling (Kim *et al.*, 2021).

CAMK2D takes part in calcium signalling and this gene has previously been linked to differences between thin and fat tailed sheep bred for milk or meat production (Luo *et al.*, 2022). NCKAP5 (NCK Associated Protein 5) was additionally found nearby a divergent SNP. This gene has been found under selection in river buffaloes and has been attributed to temperament in cattle (Valente *et al.*, 2016; Saravanan *et al.*, 2023)

Alternatively, swamp buffalo as draught livestock require adaptations for docility and endurance. NLGN1 (Neuroligin 1) and TMEM156 (Transmembrane Protein 156) are two neuronal genes that may signify selection upon draught ability of swamp buffalo. Both genes have been linked with mental disorders or differences in neuronal function but more specifically, NLGN1 functions in synaptic transmission and effects behaviour via differential expression altering excitatory/inhibitory balance in hippocampus (Kilaru *et al.*, 2016; Nakanishi *et al.*, 2017; Katzman and Alberini, 2018; Gupta *et al.*, 2022; Chalkiadaki *et al.*, 2023). NLGN1 has previously been found in enriched neuronal pathways in swamp buffalo (Sun, Shen, *et al.*, 2020). DLG2 (Discs Large MAGUK Scaffold Protein 2) shows further neuronal function as knockdown study in mice suggests this gene effects striatal connectivity and frequency of excitatory spontaneous postsynaptic currents, and this gene has been shown under differential selection between herding dogs and livestock guardians (Yoo *et al.*, 2020). In cattle, DLG2 has been associated with digital cushion thickness and destructive lesions (Stambuk, Staiger, Heins, *et al.*, 2020; Stambuk, Staiger, Nazari-Ghadikolaei, *et al.*, 2020; Li *et al.*, 2023). Swamp buffalo are known for being more resilient as draught animals than cattle, showing reduced hoof lesions and injuries, whilst also recovering at a faster rate. FGF3 (Fibroblast Growth Factor 3), BMP2 (Bone Morphogenetic Protein 2), RBMS3 (RNA Binding Motif Single Stranded Interacting Protein 3), SNX29 (Sorting Nexin 29) all appear to have function relating to growth or muscle function. FGF3 is an oncogene and growth factor and has been associated with average daily gain (Seong *et al.*, 2022). BMP2 has been identified under selection Bulgarian river buffalo in Chapter Three, an example of river buffalo population bred for increased growth and meat traits. RBMS3 has been previously found under selection in swamp buffalo and belongs to the family of c-Myc genes and may effect bone density while (Ravi Kumar *et al.*, 2020).

As river and swamp buffalo can interbreed, the expectation may be that reproductive functions are near identical. Genes associated with reproduction were found for both

divergent and balancing selection SNPs. For divergent SNPs, MARCH1 (Membrane Associated Ring-CH-Type Finger 1) and PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6) were found. MARCH1 functions in immunological pathways however has been linked to various bull semen traits in cattle (Liu *et al.*, 2017; Modiba *et al.*, 2022; Mukherjee *et al.*, 2023). PCSK6 processes proteins in secretory pathway but has been linked with reproductive function, while also occurring in a CNV region in Iranian river buffalo (Strillacci *et al.*, 2021). Increasing intensification and genetic improvement of river buffaloes has led to widespread use of assisted reproductive technologies (ART). Therefore, genes under selection for semen quality traits may influence to success of ART but may not interfere with essential reproductive pathways.

SNPs under balancing selection with reproductive gene links may function in pivotal roles. ATF2 (Activating Transcription Factor 2) features oncogenic function and responds to genotoxic stress. Studies have linked this gene to fertility with knockouts of this gene result in lethality and expression is crucial for survival of bovine embryos (Lau and Ronai, 2012; Orozco-Lucero, Dufort and Sirard, 2017; Tarekegn *et al.*, 2021). WIF1 (WNT Inhibitory Factor 1) participates in WNT signalling, a known critical pathway for embryonic development. Expression is important for cell proliferation and differentiation across a wide variety of cell and tissue types, but notably is essential for embryo implantation and development in cattle (Nascimento *et al.*, 2018). KAZN (Kazrin, Periplakin Interacting Protein) is a diagnostic marker for ovarian cancer with further links to pregnancy complications and poly cystic ovary syndrome (Barbitoff *et al.*, 2020; Egashira *et al.*, 2022; Zhu *et al.*, 2022). CRISPLD1 (Cysteine Rich Secretory Protein LCCL Domain Containing 1) was found to influence calcium regulation via a binding site similar to that of helothermine toxin that blocks cardiac ryanodine receptor channels, thus leading to heart failure (Khadjeh *et al.*, 2020). Calcium is important in fertilization and Sujit *et al,* (2018) elucidated CRISPLD1 as a seminal plasma biomarker. PDLIM1 (PDZ And LIM Domain 1) is a cytoskeletal protein that is essential for spermatid differentiation and spermatogenesis and as such, was found to have strong predictive power for uterine embryo presence (Shang *et al.*, 2016; Kusama *et al.*, 2021). Divergence of reproductive genes under balancing selection may be impossible due to causing cancers or infertility and therefore current viable genetic variation is maintained. Adding to this RNASEH2B (Ribonuclease H2 Subunit B) is an example of a gene under balancing selection in

livestock as a homozygous deletion was identified that causes embryonic lethality however, carriers are maintained in Nordic Red cattle due to positive effects on milk yield and composition (Kadri *et al.*, 2014).

A variety of immune genes were found under balancing selection. NCAM2 (Neural Cell Adhesion Molecule 2) is part of the immunoglobulin superfamily and a cell adhesion molecule that is regulated by STAT5 and cytokines for natural killer cell survival and differentiation (Nelson *et al.*, 2006). Although NCAM2 appears under balancing selection across buffaloes, this gene shows value to Brazilian murrah buffaloes from Chapter Three and GWAS studies (de Camargo *et al.*, 2015). TREM1 (Triggering Receptor Expressed On Myeloid Cells 1) is an immunoglobulin receptor present on macrophages and neutrophils and is involved in triggering inflammatory responses (Ornatowska *et al.*, 2007). SPPL2A (Signal Peptide Peptidase Like 2A) is crucial for B cell development by maintaining integrity of MHCII containing endosomes (Schneppenheim *et al.*, 2012). SNP nearby NEU1 (Neuraminidase 1) was found within the MHC region whereby it is known that alleles are shared across both river and swamp buffaloes (Sena *et al.*, 2003; Mishra *et al.*, 2020). Maintaining genetic variation across immune genes is crucial for adaptable, responsive and effective immune system vital for survival of organisms in face of disease and parasites. Similarly, genetic variation may be beneficial for sensory traits enabling species to adapt to changes in the environment and faunal composition. OTOL1 (Otolin 1), TMEM132E (Transmembrane Protein 132E), and RGS7BP (Regulator Of G Protein Signalling 7 Binding Protein) have all been linked to expression within and development of inner ears (Seim *et al.*, 2013; Megdiche *et al.*, 2019; Yuan Wang *et al.*, 2022). Domestic buffaloes are predominantly kept in low input extensive systems and are allowed to roam freely. Living across Southern Asia, their range entirely overlaps with tigers that are the apex predator in the region where water buffaloes fulfil a primary source of prey. Studies on livestock predation show that water buffaloes are predated far less than cattle (up to 25 times less likely) which may be due to their greater ability to detect and defend against predators compared to cattle whose reactions are dispersal (Hoogesteijn and Hoogesteijn, 2008; Miller, Jhala and Jena, 2016; Kolipaka *et al.*, 2017). As water buffalo are farmed more intensively, there may be a decoupling of genes under balancing selection that are useful within a natural context.

5.6. Conclusion

This chapter further studies the genetic variation present across global domestic water buffalo; quantifying and explaining the patterns and differences between species and populations. The valuable addition of an Indian murrah population provides further knowledge surrounding river buffalo ancestry and available genetic diversity. This population originates from the putative river buffalo domestication location (India), and as such, was found to possess the highest genetic diversity in line with domestication expectations. Additional European populations from Noce *et al,* (2021), and UK buffalo genotyped here shows revealed greater variability in Europe, previously not detected, mostly through an East and West split. Modelling of domestic buffalo evolution produced greater insights into the dispersal of buffaloes. Mediterranean buffaloes likely derived from the first dispersal out of India, via settlement of the 'al-Zuṭṭ' people in the Middle East early in the Common Era. Translocation of river buffaloes likely continued from India to the Middle East, replacing original genetic diversity creating the observed genetic differences to European populations. River and swamp species were independently domesticated and diverged during the Pleistocene, although accurate estimates may not be achievable using buffalo array data due to ascertainment bias. Although river and swamp buffalo are genetically, they are successfully able to hybridise. Divergent loci between the species were found associated with livestock function (suggesting artificial selection), while loci under balancing selection appeared important for survivability and cell viability. Therefore, important regions for biological compatibility between buffalo species may have remained similar, thus allowing hybridisation.

Chapter Six

General Discussion

6.1. Background

Agriculture and livestock industries are integral for producing enough food to support the large human populations seen today. An increase in human population coupled with development of countries has seen the demand for livestock products rapidly increase in the past 50 years (Steinfeld, Wassenaar and Jutzi, 2006; FAO, 2009; Godde *et al.*, 2018). Anthropomorphic change on the planet is having severe consequences in environmental effects. Rising global temperatures, demands on freshwater resources, and a spread in diseases are placing human society under pressure (Tomley and Shirley, 2009; Thornton, 2010; Baumgard *et al.*, 2012; Leng and Hall, 2021; Cheng, McCarl and Fei, 2022). The consequences threaten food production systems as the degradation of seasonal weather patterns effect reliability of crop production, and extreme weather conditions, e.g., heat waves and drought, cause serious health conditions in livestock that drastically reduce yields (Thornton, 2010; Cheng, McCarl and Fei, 2022).

To mitigate such conditions whilst retaining productive livestock, breeding programs must be optimized and complex genotype-phenotype interactions understood (Goddard, 2012; Hayes, Lewin and Goddard, 2013; Meuwissen, Hayes and Goddard, 2013). The availability of large-scale genomic data has provided ample opportunities for guiding livestock breeding (Hayes, Lewin and Goddard, 2013; Meuwissen, Hayes and Goddard, 2013; Gorjanc *et al.*, 2015). While genomics has improved production gains, the same must be achieved with environmental adaptivity. Local livestock breeds host a wealth of genetic diversity relating to environmental adaptation (Hoffmann, 2013; Savolainen, Lascoux and Merilä, 2013; Biscarini *et al.*, 2015). These breeds are typically farmed in extensive systems and as such, have become inherently adapted to their immediate surroundings. However, due to their small distributions, and lower productivity, local livestock breeds have typically not been genetically characterized and monitored. 16% of livestock breeds are at risk of extinction, and a further

35% are uncharacterized presenting a serious threat to future food security if these genetic resources are lost (FAO, 2015).

The domestic water buffalo is an example of this. Water buffalo have been among the most important livestock species to Southern Asia with 95% of the global population found here but gains little attention elsewhere. Adaptations to tropical conditions, e.g., thermotolerance, feed efficiency, and resistance to diseases and parasites, may be of great use to northernly latitudes as climates become more tropical (Marai and Haeeb, 2010; Yáñez *et al.*, 2020). Genomic resources have become easily accessible for large scale population analysis in the form of the Axiom^{TM} Buffalo Genotyping Array, presenting the opportunity to understand the genetic variation found across domestic buffaloes and evaluate their adaptive potential.

6.2. Completion of Aims

The primary goal of a livestock species is the production of a valuable commodity. In the case of domestic water buffalo, riverine breeds are dairy producers while swamp buffalo provide draught (Zhang, Colli and Barker, 2020). Both domestic species provide meat as a secondary product. **Chapters Two** and **Three** studied breeds of river buffalo that are becoming commercially relevant and beginning to be subjected to the pressures of modern farming systems. **Chapter Two** analyzed 69 individuals of Mediterranean buffaloes sampled across two farms from the United Kingdom and compared differences in genetic variation to populations across Europe. The aims of this chapter were to genetically characterize these populations and identify any notable changes in genetic variation. Such changes may take the form of significant declines in genetic diversity since importation, or rapid shifts in allelic frequencies suggesting recent selection.

Although observed heterozygosity of UK buffaloes was lower than continental counterparts, differences were non-significant. Declines in genetic diversity can often be seen in newly formed populations as a limited amount of genetic variation is captured from the original population in founding individuals (Bruford, Bradley and Luikart, 2003). It is important to monitor genetic diversity over time in livestock populations. The consequences of reduced diversity include an increased risk in inbreeding that may result in adverse health effects. This is especially important under commercial scale farming as genomic selection programs often

see an increase in inbreeding (Meuwissen, Hayes and Goddard, 2013; Meuwissen *et al.*, 2020). Despite a marginal decline, comparative levels of observed heterozygosity across UK buffaloes suggests that a variety of individuals have been imported. Furthermore, H_O was found to be higher than H_E across both populations, a signal observed in the presence of recent gene flow, indicating that recent imports are bringing new variation into UK populations. This result was also found across other populations in **Chapters Two**, **Three**, and **Five**. As a result of this observation, there is an absence of inbreeding found within UK buffaloes. The high levels of variation identified indicate that an abundance of variation is available for genomic gain of production yields through selective breeding.

Successful breeding programs rely upon breeding individuals that share the desired trait so that underlying genotypes are maintained (Meuwissen, Hayes and Goddard, 2001). Breeding of closely related individuals will result in increased inbreeding down generations (Hayes *et al.*, 2009; Meuwissen *et al.*, 2020). An influx of new genetic variation obtained by breeding more distantly related individuals together will overcome any inbreeding; however, this may disrupt the desired phenotype and regions under selection (Toosi, Fernando and Dekkers, 2010). It is essential to identify the genetic origins of novel populations so that future imports can be targeted for minimal disruption. Using F_{ST}, MDS, and ADMIXTURE analysis, UK populations are shown to tightly cluster with Italian populations. Some deviation was observed between UK populations, particularly that of RV_UK2 looking more unique. Further exploration of differences between UK populations was conducted using IBS and LAMP-ANC analysis. Mediterranean buffaloes in the UK are known to originate from two sources, Italy and Romania (Borghese, 2013a). Farmers initially import less productive, and therefore cheaper Romanian buffaloes to establish herds before increasing production capabilities by importing more valuable Italian individuals (Borghese, 2013a). The similarity of UK buffaloes to Italian buffaloes was further observed in LAMP-ANC analysis that identified proportions of the genome relating to either Italian or Romanian ancestry. Both UK populations featured greater than 85% of the genome as Italian ancestry, indicating any potential admixture from Romanian populations was low. Notably, Romanian ancestry was largely located in the same genomic regions of both populations suggesting these regions may have been maintained since pre-importation to UK. A greater proportion of Romanian ancestry was identified in RV_UK2 than RV_UK1, however population structure results (e.g., F_{ST}, ADMIXTURE) show

both UK populations to be dissimilar to Romanian populations. Therefore, it was suspected that RV_UK2 may be undergoing some genetic drift due to the smaller size and less commercial activity of the farm. Conversely, it is known that RV_UK1 is continually importing further Italian individuals. IBS results show that individuals in RV_UK2 are significantly more related to each other than RV_UK1. Therefore, RV_UK2 may be more susceptible to genetic drift as allelic frequencies can change more rapidly and become fixated in smaller populations. Genetic drift and any potential threat of future inbreeding can be easily mitigated through incorporation of new individuals.

Genomic selection using genotyping arrays is frequently carried out by i) close associations between alleles and observed phenotypes (e.g., GWAS), or ii) identification of genomic regions under higher linkage disequilibrium encapsulating functional genes (e.g., XP-EHH). In the absence of phenotypic data for either UK population, LD was explored. Both populations featured reasonably low levels of LD compared to other livestock species, reflecting the high levels of genetic diversity found. Additionally, patterns of LD were not consistent between UK populations, and as such, transfer of genomic selection programs would have low confidence of success. Recent trends in genetic diversity through time were reconstructed by calculation of N_e from LD. Both populations displayed decreasing estimates of N_e as time became more recent with RV_UK1 possessing higher N_e until approximately 250 generations ago. Higher estimates of N_e in RV_UK1 indicate that implementation of genomic selection programs would be more powerful than in RV_UK2. The translation of lower N_e in RV_UK2 meant that a greater proportion of the genome possessed ROHs compared to RV_UK1, although this difference was non-significant. The maintenance of ROHs within each population revealed clusters of interconnected genes functioning in relevant livestock traits (e.g., milk production) that may be under selection. Interestingly, representation of QTL classes differed between UK populations. RV_UK1 featured a significant overrepresentation of exterior and production QTL classes that incorporate traits such as udder morphology, while RV_UK2 featured a more classical expectation of an overrepresentation of milk QTLs. Genes relating to morphology may therefore be of greater importance in increasing milk yields in domestic buffaloes. The differences in UK populations were further observed with little overlap in genes under selection from XP-EHH results. Though, having only been present

in the UK for around 50 years, and with little environmental change to continental Europe, there may not be extensive selection pressures on UK buffalo.

Likewise with Mediterranean buffalo, the Indian murrah buffalo is another commercially relevant breed. In **Chapter Three**, global populations of murrah buffaloes were compared to an Indian population to evaluate the adaptability of domestic buffaloes. Unique adaptations found in newly formed populations would suggest that domestic buffaloes are adaptable to changing climatic conditions. Similarly to **Chapter Two**, all murrah populations showed high levels of genetic diversity and an absence of genetic diversity. Again, it is likely due to the lack of intensive selection on domestic buffalo populations. Despite continental separation of studied murrah populations (Asia, Europe & South America), divergence was low with average F_{ST} scores of 2.2%, though this is marginally inflated due to incorporation of Mediterranean ancestry found within the crossbred population of RV_BUL_VAR (Bulgarian murrah). ADMIXTURE results were unable to distinguish between populations bar detecting different ancestry within RV_BUL_VAR and some potential underlying structure within the Indian murrah population.

Assessment of LD across all populations revealed rapid rates of decay within Indian murrah buffalo. Interestingly, a population expansion event was found in Indian murrah buffalo in N_e results (after controlling for any population structure). This expansion peaked at 32 generations ago that relates to approximately 192 years ago. Indian murrah individuals here were sampled from eastern and central India whilst the murrah breed originated from the Haryana province in Northwest India (Kumar *et al.*, 2019). This timepoint corresponds to a period of drought and power transfer in India that may have led to dispersals of people and livestock (Clingingsmith and Williamson, 2008; Guha, 2019). Alternatively, murrah buffalo are frequently bred with local breeds to improve milk production whilst maintaining local adaptivity (Kumar and Singh, 2010; Kumar, 2015; Kumar *et al.*, 2019). Graded murrah buffaloes outnumber purebred murrah by a ratio of 3:1, therefore this expansion signal may be due to persistent admixture with local breeds. All other murrah populations featured lower N_e as expected due to each capturing a subset of variation of the original population.

The adaptive potential of murrah buffaloes was evaluated through detecting unique regions of the genome under selection in each population. Due to the very recent exportation of murrah buffalo out of India (<100 years), unique regions under selection may indicate a

rapid shift in allelic frequencies in response to new selection pressures. This was evaluated using multiple methods: i) ROHs, ii) XP-EHH, iii) PCAdapt, and iv) R-Samβada. The former two methods provided selection signatures based on elevated levels of LD, meanwhile the latter targeted fine scale selection that may be a response to wider variation in environmental changes. What was found was that ROHs and XP-EHH analysis revealed genes that relate to alternative functions between murrah populations. For example, genes relating to increased growth or muscle function were found under selection in RV_BUL_VAR and RV_CO (Colombia). Bulgarian murrah have been reared for increased muscle mass for greater meat production whilst Colombian buffalo have been imported from the Caribbean where the buffalypso (derived from murrah and other Indian breeds) was turned into a draught breed. The Brazilian murrah population is subject to greater intensity of milk production, and in this population an array of genes was linked to energy metabolism, thermoregulation, and reproduction. Milk production in dairy livestock is energetically expensive and physiologically demanding, even more so under tropical climatic conditions (Collier, Renquist and Xiao, 2017; Cheng, McCarl and Fei, 2022). These genes may represent candidates for buffalo farming, particularly in stock health and maintenance over generations. Indian murrah featured fewer unique regions under selection, most probably because similar signals are also found in other murrah populations due to little divergence. However, genes under selection did link to relevant livestock processes such as oxidative stress and immune responses. Though this study lacked fine scale sampling, R-Samβada results may provide some inferences into environmental adaptation of domestic buffaloes. CHL1 was found linked with precipitation with alternative genotypes found in Indian murrah to other murrah populations. These buffaloes reside in the driest areas studied here. CHL1 functions by recruiting Hsc70 that helps against neuronal and vascular diseases under heat stress, of which these buffaloes will more likely be subjected to with reduced access to water.

The Axiom^{TM} Buffalo Genotyping Array provides the most affordable and reproducible opportunity for large scale genomic studies of domestic water buffalo. Genotyping microarrays have provided easy access for farmers to genetically analyse their livestock to a reasonably high genomic resolution. Livestock microarrays are purpose built for the ability to generate genomic selection programs in their target species (Hayes *et al.*, 2009; Hayes, Lewin and Goddard, 2013; Meuwissen, Hayes and Goddard, 2013, 2016; Kemper *et al.*, 2016;

Georges, Charlier and Hayes, 2019). This goal is achieved by utilizing linkage disequilibrium to attribute changes in SNP markers to functional genes. SNPs of higher frequencies that are common across host breeds are deliberately chosen to differentiate between closely related individuals. Whilst useful for microarray function, these SNPs no longer represent natural variation of populations that normally possess an abundance of low frequency alleles (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; McTavish and Hillis, 2015; Benjelloun *et al.*, 2019). Bias in allelic frequencies will likely distort statistical analysis within evolutionary studies (Albrechtsen, Nielsen and Nielsen, 2010; Lachance and Tishkoff, 2013; Malomane *et al.*, 2018). **Chapter Four** evaluated and analysed ascertainment bias within the Axiom^{TM} Buffalo Genotyping Array in comparison to WGS data obtained from Luo *et al,* (2020). In addition to biased allele frequencies in comparison to natural variation, SNP selection for the Axiom^{TM} Buffalo Genotyping Array only included river breeds. Therefore, further bias is present as swamp buffalo genetic variation is underrepresented when genotyped with this array.

Chapter Four revealed that river buffalo genetic diversity is significantly inflated compared to natural variation found in WGS data. Shifts in allelic frequencies towards higher frequencies are observed and MAF distributions did not correlate with WGS data. This outcome is common across microarrays as observed in cattle (Pitt, Sevane, *et al.*, 2019), sheep (Benjelloun *et al.*, 2019), and chickens (Malomane *et al.*, 2018). Despite inflated diversity estimates, patterns of variation across river buffalo populations were minimally affected by ascertainment bias within the array. This result suggests that interpretations are likely correct bar differences in raw values. Swamp buffalo meanwhile were found to have significantly lower genetic diversity in comparison to river buffalo in both array and WGS data. However, the difference in diversity between river and swamp buffalo was far greater in array data, clearly visualizing the effect of ascertainment bias between the species. Swamp buffalo diversity largely mimicked WGS data. This suggests that the SNP array is effectively a random subset of markers to swamp buffalo, though lacking in swamp specific variation. The knock-on effects of the differences between river and swamp buffalo means that comparative analysis is distorted (e.g., F_{ST} overestimated in comparison to WGS).

Ascertainment bias within the array between river and swamp buffalo could be overcome through LD pruning techniques by targeting SNP selection in swamp buffalo.

Differences in genetic diversities were minimised, and F_{ST} scores reduced. However, LD pruning only increased diversity within swamp buffalo, with minimal effect on river buffalo diversity. This result was not replicated in WGS data. Therefore, the SNPs retained were likely targets of shared ancestral genetic variation, or those markers under balancing selection. This pattern was most easily observed in ROH analysis. LD pruning in each species in WGS data resulted in lower detection of ROHs as removal of monomorphic linked SNP markers were removed from the dataset. In the other species, ROHs presence increased as genetic variation is not shared between river and swamp buffalo. In the array, swamp buffalo always featured more ROHs unless LD pruning targeted swamp buffalo, in which case ROHs became similar to river buffalo. The outcome of ascertainment bias meant that estimates of evolutionary history may be distorted as SNP datasets that targeted the markers of shared variation resulted in more recent estimates of divergence. Future evolutionary studies across river and swamp buffalo would benefit from using less biased methods such as ddRAD-sequencing, and future SNP arrays should either focus solely on swamp buffalo or incorporate both species in SNP discovery.

Evolutionary studies on domestic buffalo have elucidated relationships between populations and quantified levels of genetic variation (Satish Kumar *et al.*, 2007; Zhang *et al.*, 2016; Colli, Milanesi, Vajana, *et al.*, 2018; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021). **Chapter Five** aimed to establish a cohesive model of evolution for domestic buffaloes, performing complex modelling to estimate divergence between river and swamp species. The modelling additionally investigated the cause of observed genetic structuring across river buffalo. Several theories hypothesize how buffaloes reached Europe including gifts to the King of Lombardy, returning crusaders, and Arab traders (Colli, Milanesi, Vajana, *et al.*, 2018). Results of the final model scenarios consistently favoured two dispersals of river buffalo from India. Indian Ocean trade has been an incredibly important source of commerce throughout human society; therefore, it is unsurprising that water buffaloes may have been translocated by Indian people. The preferred model estimated an initial dispersal of river buffalo from India approximately 2,000 YA. This time aligns with Indians settlers, called 'al-Zutt' reaching the Middle East early in the Common Era (Wordsworth *et al.*, 2021; Simpson, n.d.). The initial river buffalo founding population likely spread up through Mesopotamia along the Euphrates and Tigris rivers reaching ancient Antioch around the 9th

century (Wordsworth *et al.*, 2021; Simpson, n.d.). Current European and Middle Eastern buffalo populations appear genetic dissimilar. The model estimated a more recent divergence of Middle Eastern buffalo from India approximately 1,100 YA. This new influx of genetic diversity likely replaced the original diversity from European ancestors in the area. Buffaloes reaching Europe, likely via Arab trade routes across the Mediterranean, likely became isolated, becoming more distantly related and possessing lower genetic diversity (Colli, Milanesi, Vajana, *et al.*, 2018; Zhang, Colli and Barker, 2020; Wordsworth *et al.*, 2021). N_e trend results from **Chapter One** additionally supported this timing of river buffalo divergence.

Precise estimates of river-swamp species divergence have been challenging to obtain. Current literature produces a range of estimates over the course of the Pleistocene (Wang *et al.*, 2017; Luo *et al.*, 2020; Sun, Shen, *et al.*, 2020; Curaudeau, Rozzi and Hassanin, 2021). The preferred model estimated a divergence time of ~14,700 YA, before the Holocene. This estimate is notably the most recent estimate of divergence in comparison to previous studies. However, there may be several caveats throwing caution to interpretation of the result. ABC attempts to characterize complex events spanning long time periods in a single model. Thus, outcomes of the models rely heavily upon some assumptions that include that i) populations are sufficiently sampled, ii) model parameters chosen are capable of depicting the species evolutionary history, and that iii) genetic variation is capable of being condensed into chosen summary statistics (Beaumont, 2010; Csilléry *et al.*, 2010). The Axiom^{TM} Buffalo Genotyping Array provides significant bias within the data as shown in **Chapter Four.** River buffalo diversity is inflated compared to natural genetic variation due to a deliberate selection of higher frequency alleles in probe production. Comparatively, swamp buffalo diversity is underrepresented due to being absent during SNP selection. Polymorphic SNPs present in both species are likely shared ancestral variation which will present a more recent divergence time. Furthermore, the model here permits a higher substitution rate that additionally would reduce divergence times. Substitution rates across the genome can vary, and with the added complexity of distorted allelic frequencies, this makes defining an exact rate troublesome (Goldman and Yang, 1994; Excoffier, Foll and Petit, 2009). Nevertheless, river and swamp buffalo were certainly domesticated independently, and shared a Pleistocene divergence. Both species show distinct genetic differences yet viably hybridize. Analysis of selection signatures relating to areas of divergence or similarity provided inferences into functional

underlying reasons. Within divergent markers, genes were frequently attributed to livestock functions such as milk production (e.g., CAMK2D) or increased growth (e.g., FGF3). Markers under balancing selection were associated with survival, such as immune response (e.g., NCAM2) or cell lethality (e.g., RNASEH2B). The validity of these genes would need to be further explored as it is difficult to distinguish between spurious outlier loci of chance, or genuine selection signatures (Kelley *et al.*, 2006).

6.3. Future Directions

This thesis was largely successful in achieving its aims in characterizing genetic resources for domestic water buffalo and uncovering their adaptive potential. Future work should therefore focus on genetic improvement and implementation of genetic breeding strategies. Samples from the two most productive buffalo breeds were analyzed in **Chapter Two** and **Chapter Three**, with those being the Mediterranean and murrah breeds, respectively. A range of candidate genes under positive selection that may be important for domestic buffaloes were identified. Inferences could only be made using prior literature. Phenotypic data should be collected on all individuals to assign greater functionality to SNPs for targeting increased production yields. Using a combination of phenotypic and genotypic data would enable the use of genome-wide association studies (GWAS) and genomic breeding values (GEBVs) (Meuwissen, Hayes and Goddard, 2013, 2016). These tests would allow for quantification of variation in phenotypic traits being accounted for by specific genotypes. With this data, genomic breeding programs could be established. Reliable results for quantitative genetics require hundreds of individuals and such studies are often out of scope for research projects or smallholders. Therefore, genomic selection needs to be led by more commercialized farms. In buffaloes, current GWAS analysis is predominantly found associated with Italian and Brazilian where large scale buffalo farms occur (Venturini *et al.*, 2014; de Camargo *et al.*, 2015; Iamartino *et al.*, 2017; Cesarani *et al.*, 2021). Within commercial environments, any genetic improvement is likely to be focused on productive gains leading to a decoupling of environmental fitness. The public availability of genetic data would help the involvement of researchers to analyze environmental traits in livestock for future adaptability.

A new database of buffalo specific QTLs should be formed and generation of novel genomic resources and knowledge for domestic buffaloes improves. Current studies, likewise,

with **Chapter Two**, use transferred cattle QTLs to the buffalo genome, which may not share functionality. Kumar *et al,* (2023) established QTLs for murrah buffaloes using double digestion RAD sequencing, however the replicability of these results will be difficult as the generation of exact SNP markers in ddRAD-seq each time is not guaranteed. With the availability of the Axiom^{TM} Buffalo Genotyping Array, a dedicated QTL database can be established from SNP markers here. Such improvements in buffalo genomics would be greatly beneficial in modernizing buffalo farming in line with industrialized livestock breeds.

The exacerbation of climate change threatens the reliable production of food as environmental conditions become altered and the frequency of extreme weather patterns increase. In the absence of successful crop growth, examples of farmers in Southeast Asia switching to buffalo farming is already occurring (Escarcha *et al.*, 2018, 2020). For livestock, species must remain adaptable to changing environmental conditions whilst still producing viable quantities of milk and meat (or other products). Genomic studies on livestock are often limited to small regions such as single farms. To link genetic variation to environmental factors, individuals need to be sampled in a sufficient manner to capture a wealth of environmental variation. Likewise, in **Chapter Three**, our sampling is limited to country level resolution. Therefore, future sampling strategies of buffalo studies should focus on identifying individuals across a large range to capture fine-scale changes in environments. Additional efforts need to ensure sampled individuals originate from those areas and are not translocated frequently so that genotypes reflect their environment.

With continual genomic studies on buffaloes, fine scale environmental studies will eventually be possible, however, results will likely be biased towards buffaloes in larger farms with greater financial resources. Most buffaloes are farmed by small holders across southern Asia. These buffaloes are farmed in low input external systems, and likely hold a large amount of genetic variation useful for environmental adaptivity. Therefore, efforts should be made to study buffalo in areas without great financial resources. This will aid small-scale farmers in breeding more productive stock, but also enable evaluation of genetic variation of local breeds, something that is currently lacking (FAO, 2015). With the rarity of the wild Asian water buffalo, obtaining new genetic variation associated with environmental fitness would be incredibly difficult. This is likely impossible for swamp buffalo as their wild ancestral populations are almost certainly extinct (Kaul *et al,* 2019). The extinction of wild ancestors to

livestock (e.g., aurochs) is a common trend, and one with negative consequences of lost genetic variation that simply cannot be recovered. Rural populations of domestic buffaloes may hold a wealth of genetic variation that needs to be characterized and conserved.

The Axiom™ Buffalo Genotyping Array was founded solely from river buffalo breeds. It is now abundantly clear that river buffalo and swamp buffalo are considerably different genetically. River buffaloes are commercially more relevant with their global population size growing each year due to their milk production (Zhang, Colli and Barker, 2020). However, for swamp buffalo, population sizes are in decline (Zhang, Colli and Barker, 2020). The loss of swamp buffalo would result in a loss of unique genetic variation (genetic erosion) that may be valuable in the future for adaptivity. Therefore, future work should commit more resources towards swamp buffalo to conserve the unique genetic variation that they hold. Problematically, without swamp buffalo being a productive livestock species and generating a commercial interest, this goal is difficult to achieve. While commercial interest will continually develop river buffalo resources, future research focusing on swamp buffalo will be valuable in filling those knowledge gaps left by industry. Farms across Southeast Asia and China are already repurposing swamp buffalo for greater meat and milk production which may provide appetite for more swamp buffalo research (Escarcha *et al.*, 2018; Deng *et al.*, 2019; Shaari *et al.*, 2019). Generation of swamp specific genomic resources such as a dedicated genotyping array may help improve swamp buffalo breeding plans at commercial farms. Any creation of such an array should focus on commercial farms to maintain viability of the livestock product and have the financial backing to generate a commercial breed of swamp buffalo so that swamp buffalo have a role in future farming.

Genetic improvement of swamp buffalo is currently achieved via crossbreeding with river buffalo. **Chapter Five** showed that divergence of river and swamp buffalo genomes is greatest where genes are attributed to production qualities and livestock traits. Persistent hybridization may lead to genetic erosion of swamp buffalo. Future research surrounding river and swamp buffalo hybridization would be of great value to identify those genotypes that improve production traits for swamp buffalo while minimizing the genetic erosion of other regions of the genome that may be useful for, e.g., local environmental adaptations.

The implications of **Chapter Four** and **Chapter Five's** research, is that river and swamp buffalo should be treated as two separate species, as proposed in Curaudeau, Rozzi and

Hassanin, (2021). Therefore, genetic variation should be conserved and managed separately, not relying on the persistence of one species over the other, particularly as swamp buffaloes are in decline. Furthermore, major splits in river buffalo populations were tracked inferring that dispersal out of India is linked with movements of Indian and Arabian traders. The resulting model should be further refined and provide opportunities for identification of important genes under selection during key events in buffalo evolution. Further investigation into domestic buffaloes' history in the Middle East may reveal new information into the cultural history of people in the region and reveal how people lived and moved in historic times (Wordsworth *et al.*, 2021).

Bibliography

Abnave, P. *et al.* (2014) 'Screening in Planarians Identifies MORN2 as a Key Component in LC3-Associated Phagocytosis and Resistance to Bacterial Infection', *Cell Host & Microbe*, 16(3), pp. 338–350. Available at: https://doi.org/10.1016/j.chom.2014.08.002.

Ahlawat, S. *et al.* (2021) 'Comparative gene expression profiling of milk somatic cells of Sahiwal cattle and Murrah buffaloes', *Gene*, 764, p. 145101. Available at: https://doi.org/10.1016/j.gene.2020.145101.

Ahmad, S.F. *et al.* (2020) 'Evaluation of two bovine SNP genotyping arrays for breed clustering and stratification analysis in well-known taurine and indicine breeds', *Animal Biotechnology*, 31(3), pp. 268–275. Available at: https://doi.org/10.1080/10495398.2019.1578227.

Ahmady, E. *et al.* (2011) 'Identification of a Novel Muscle A-type Lamin-interacting Protein (MLIP) *', *Journal of Biological Chemistry*, 286(22), pp. 19702–19713. Available at: https://doi.org/10.1074/jbc.M110.165548.

Al Qahtani, N.H. *et al.* (2021) 'Whole-Genome Sequencing Reveals Exonic Variation of ASIC5 Gene Results in Recurrent Pregnancy Loss', *Frontiers in Medicine*, 8. Available at: https://www.frontiersin.org/articles/10.3389/fmed.2021.699672 (Accessed: 4 September 2023).

Al Zaabi, N., Al Menhali, N. and Al-Jasmi, F. (2018) 'SYNJ1 gene associated with neonatal onset of neurodegenerative disorder and intractable seizure', *Molecular Genetics & Genomic Medicine*, 6(1), pp. 109–113. Available at: https://doi.org/10.1002/mgg3.341.

Alan Harris, R. *et al.* (2023) 'Loci on chromosome 12q13.2 encompassing ERBB3, PA2G4 and RAB5B are associated with polycystic ovary syndrome', *Gene*, 852, p. 147062. Available at: https://doi.org/10.1016/j.gene.2022.147062.

Albrechtsen, A., Nielsen, F.C. and Nielsen, R. (2010) 'Ascertainment Biases in SNP Chips Affect Measures of Population Divergence', *Molecular Biology and Evolution*, 27(11), pp. 2534–2547. Available at: https://doi.org/10.1093/molbev/msq148.

Aldersey, J.E. *et al.* (2020) 'Understanding the effects of the bovine POLLED variants', *Animal Genetics*, 51(2), pp. 166–176. Available at: https://doi.org/10.1111/age.12915.

Alexander, D.H., Novembre, J. and Lange, K. (2009) 'Fast model-based estimation of ancestry in unrelated individuals', *Genome Research*, 19(9), pp. 1655–1664. Available at: https://doi.org/10.1101/gr.094052.109.

Algra, M. *et al.* (2023) 'Evaluation of the Thermal Response of the Horns in Dairy Cattle', *Animals*, 13(3), p. 500. Available at: https://doi.org/10.3390/ani13030500.

Alharby, E. *et al.* (2017) 'A homozygous potentially pathogenic variant in the PAXBP1 gene in a large family with global developmental delay and myopathic hypotonia', *Clinical Genetics*, 92(6), pp. 579–586. Available at: https://doi.org/10.1111/cge.13051.

Alhussien, M.N. *et al.* (2018) 'Effect of tropical thermal stress on peri-implantation immune responses in cows', *Theriogenology*, 114, pp. 149–158. Available at: https://doi.org/10.1016/j.theriogenology.2018.03.036.

Alhussien, M.N. and Dang, A.K. (2018) 'Milk somatic cells, factors influencing their release, future prospects, and practical utility in dairy animals: An overview', *Veterinary World*, 11(5), pp. 562–577. Available at: https://doi.org/10.14202/vetworld.2018.562-577.

Allais-Bonnet, A. *et al.* (2013) 'Novel Insights into the Bovine Polled Phenotype and Horn Ontogenesis in Bovidae', *PLOS ONE*, 8(5), p. e63512. Available at: https://doi.org/10.1371/journal.pone.0063512.

Allende, M. *et al.* (2016) 'Hsp70 protects from stroke in atrial fibrillation patients by preventing thrombosis without increased bleeding risk†', *Cardiovascular Research*, 110(3), pp. 309–318. Available at: https://doi.org/10.1093/cvr/cvw049.

Almathen, F. *et al.* (2016) 'Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary', *Proceedings of the National Academy of Sciences*, 113(24), pp. 6707–6712. Available at: https://doi.org/10.1073/pnas.1519508113.

Alvarenga, A.B. *et al.* (2023) 'Unraveling the phenotypic and genomic background of behavioral plasticity and temperament in North American Angus cattle', *Genetics Selection Evolution*, 55(1), p. 3. Available at: https://doi.org/10.1186/s12711-023-00777-3.

Amano, N. *et al.* (2013) 'Introduced Domestic Animals in the Neolithic and Metal Age of the Philippines: Evidence From Nagsabaran, Northern Luzon', *The Journal of Island and Coastal Archaeology*, 8(3), pp. 317–335. Available at: https://doi.org/10.1080/15564894.2013.781084.

Amiri Ghanatsaman, Z. *et al.* (2023) 'Comparative genomic analysis uncovers candidate genes related with milk production and adaptive traits in goat breeds', *Scientific Reports*, 13(1), p. 8722. Available at: https://doi.org/10.1038/s41598-023-35973-0.

Amoroso, M.G., Galiero, G. and Fusco, G. (2017) 'Genetic characterization of ovine herpesvirus 2 strains involved in water buffaloes malignant catarrhal fever outbreaks in Southern Italy', *Veterinary Microbiology*, 199, pp. 31–35. Available at: https://doi.org/10.1016/j.vetmic.2016.12.020.

Angst, P. *et al.* (2022) 'Genetic Drift Shapes the Evolution of a Highly Dynamic Metapopulation', *Molecular Biology and Evolution*, 39(12), p. msac264. Available at: https://doi.org/10.1093/molbev/msac264.

Antonio, B. (2010) 'Development and perspective of Buffalo and Buffalo market in Europe and Near East', *Revista Veterinaria*, 21.

de Araujo Neto, F.R. *et al.* (2022) 'Genomic study of the resilience of buffalo cows to a negative energy balance', *Journal of Applied Genetics*, 63(2), pp. 379–388. Available at: https://doi.org/10.1007/s13353-021-00680-x.

Arbel-Eden, A. and Simchen, G. (2019) 'Elevated Mutagenicity in Meiosis and Its Mechanism', *BioEssays*, 41(4), p. 1800235. Available at: https://doi.org/10.1002/bies.201800235.

van Arendonk, J.A.M. and Liinamo, A.-E. (2003) 'Dairy cattle production in Europe', *Theriogenology*, 59(2), pp. 563–569. Available at: https://doi.org/10.1016/S0093-691X(02)01240-2.

Askari, G.H. *et al.* (2013) 'Effect of STAT1 variants on milk production traits in Esfahan Holstein cows.', *International Journal of Advanced Biological and Biomedical Research*, 1(8), pp. 851–857.

Asner, G.P. *et al.* (2004) 'Grazing Systems, Ecosystem Responses, and Global Change', *Annual Review of Environment and Resources*, 29(1), pp. 261–299. Available at: https://doi.org/10.1146/annurev.energy.29.062403.102142.

Asthana, S., Schmidt, S. and Sunyaev, S. (2005) 'A limited role for balancing selection', *Trends in Genetics*, 21(1), pp. 30–32. Available at: https://doi.org/10.1016/j.tig.2004.11.001.

Auton, A. *et al.* (2009) 'Global distribution of genomic diversity underscores rich complex history of continental human populations', *Genome Research*, 19(5), pp. 795–803. Available at: https://doi.org/10.1101/gr.088898.108.

Badam, G.L., Sathe, W. and Kajale, M.D. (1996) 'On a Partial Skull of Bubalus Sp. from the Late Pleistocene Alluvial Deposits Around Walki, Bhima Valley, Maharashtra', *Bulletin of the Deccan College Research Institute*, 56/57, pp. 27–35.

Bagath, M. *et al.* (2019) 'The impact of heat stress on the immune system in dairy cattle: A review', *Research in Veterinary Science*, 126, pp. 94–102. Available at: https://doi.org/10.1016/j.rvsc.2019.08.011.

Bakoev, S. *et al.* (2020) 'Survey of SNPs Associated with Total Number Born and Total Number Born Alive in Pig', *Genes*, 11(5), p. 491. Available at: https://doi.org/10.3390/genes11050491.

Ballatori, N. *et al.* (2013) 'The heteromeric organic solute transporter, OSTα–OSTβ/SLC51: A transporter for steroid-derived molecules', *Molecular Aspects of Medicine*, 34(2), pp. 683–692. Available at: https://doi.org/10.1016/j.mam.2012.11.005.

Baran, Y. *et al.* (2012) 'Fast and accurate inference of local ancestry in Latino populations', *Bioinformatics*, 28(10), pp. 1359–1367. Available at: https://doi.org/10.1093/bioinformatics/bts144.

Barbitoff, Y.A. *et al.* (2020) 'A Data-Driven Review of the Genetic Factors of Pregnancy Complications', *International Journal of Molecular Sciences*, 21(9), p. 3384. Available at: https://doi.org/10.3390/ijms21093384.

Barile, V.L. (2005) 'Improving reproductive efficiency in female buffaloes', *Livestock Production Science*, 92(3), pp. 183–194. Available at: https://doi.org/10.1016/j.livprodsci.2004.06.014.

Barker, J. (2014) 'Water Buffalo: Domestication', *Encyclopedia of Global Archaeology* [Preprint]. Available at: https://doi.org/10.1007/978-1-4419-0465-2_2217.

Barker, J.S. (2001) 'Conservation and management of genetic diversity: a domestic animal perspective', *Canadian Journal of Forest Research*, 31(4), pp. 588–595. Available at: https://doi.org/10.1139/x00-180.

Barker, J.S.F. *et al.* (1997) 'Genetic diversity of Asian water buffalo (Bubalus bubalis): microsatellite variation and a comparison with protein-coding loci', *Animal Genetics*, 28(2), pp. 103–115. Available at: https://doi.org/10.1111/j.1365-2052.1997.00085.x.

Barnes, N.E. *et al.* (2019) 'Thermal challenge alters the transcriptional profile of the breast muscle in turkey poults', *Poultry Science*, 98(1), pp. 74–91. Available at: https://doi.org/10.3382/ps/pey401.

Bar-On, Y.M., Phillips, R. and Milo, R. (2018) 'The biomass distribution on Earth', *Proceedings of the National Academy of Sciences*, 115(25), pp. 6506–6511. Available at: https://doi.org/10.1073/pnas.1711842115.

Barton, N.H. (2022) 'The "New Synthesis"', *Proceedings of the National Academy of Sciences*, 119(30), p. e2122147119. Available at: https://doi.org/10.1073/pnas.2122147119.

Baruselli, P.S. *et al.* (2023) 'Review: Development, adoption, and impact of assisted reproduction in domestic buffaloes', *animal*, 17, p. 100764. Available at: https://doi.org/10.1016/j.animal.2023.100764.

Baumgard, L.H. *et al.* (2012) 'Impact of Climate Change on Livestock Production', in V. Sejian et al. (eds) *Environmental Stress and Amelioration in Livestock Production*. Berlin, Heidelberg: Springer, pp. 413–468. Available at: https://doi.org/10.1007/978-3-642-29205-7_15.

Bearce, E.A. *et al.* (2022) 'Urotensin II-related peptides, Urp1 and Urp2, control zebrafish spine morphology', *eLife*. Edited by M.E. Bronner, 11, p. e83883. Available at: https://doi.org/10.7554/eLife.83883.

Beaumont, M.A. (1999) 'Detecting Population Expansion and Decline Using Microsatellites', *Genetics*, 153(4), pp. 2013–2029. Available at: https://doi.org/10.1093/genetics/153.4.2013.

Beaumont, M.A. (2010) 'Approximate Bayesian Computation in Evolution and Ecology', *Annual Review of Ecology, Evolution, and Systematics*, 41(1), pp. 379–406. Available at: https://doi.org/10.1146/annurev-ecolsys-102209-144621.

Becker, G.M. *et al.* (2022) 'Variants Within Genes EDIL3 and ADGRB3 are Associated With Divergent Fecal Egg Counts in Katahdin Sheep at Weaning', *Frontiers in Genetics*, 13. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2022.817319 (Accessed: 13 October 2023).

Beilharz, R.G., Luxford, B.G. and Wilkinson, J.L. (1993) 'Quantitative genetics and evolution: Is our understanding of genetics sufficient to explain evolution?', *Journal of Animal Breeding and Genetics*, 110(1–6), pp. 161–170. Available at: https://doi.org/10.1111/j.1439-0388.1993.tb00728.x.

Beja-Pereira, A. *et al.* (2006) 'The origin of European cattle: Evidence from modern and ancient DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 103(21), pp. 8113–8118. Available at: https://doi.org/10.1073/pnas.0509210103.

Belhadj Slimen, I. *et al.* (2016) 'Heat stress effects on livestock: molecular, cellular and metabolic aspects, a review', *Journal of Animal Physiology and Animal Nutrition*, 100(3), pp. 401–412. Available at: https://doi.org/10.1111/jpn.12379.

Bellono, N.W. and Oancea, E.V. (2014) 'Ion transport in pigmentation', *Archives of Biochemistry and Biophysics*, 563, pp. 35–41. Available at: https://doi.org/10.1016/j.abb.2014.06.020.

Benjelloun, B. *et al.* (2019) 'An evaluation of sequencing coverage and genotyping strategies to assess neutral and adaptive diversity', *Molecular Ecology Resources*, 19(6), pp. 1497–1515. Available at: https://doi.org/10.1111/1755-0998.13070.

van den Berg, I. *et al.* (2014) 'Concordance analysis for QTL detection in dairy cattle: a case study of leg morphology', *Genetics Selection Evolution*, 46(1), p. 31. Available at: https://doi.org/10.1186/1297-9686-46-31.

Bergman, J. *et al.* (2023) 'Worldwide Late Pleistocene and Early Holocene population declines in extant megafauna are associated with Homo sapiens expansion rather than climate change', *Nature Communications*, 14(1), p. 7679. Available at: https://doi.org/10.1038/s41467-023-43426-5.

Bergström, A. *et al.* (2020) 'Origins and genetic legacy of prehistoric dogs', *Science*, 370(6516), pp. 557–564. Available at: https://doi.org/10.1126/science.aba9572.

Bernabucci, U. (2019) 'Climate change: impact on livestock and how can we adapt', *Animal Frontiers*, 9(1), pp. 3–5. Available at: https://doi.org/10.1093/af/vfy039.

Berry, D.P. and Spangler, M.L. (2023) 'Animal board invited review: Practical applications of genomic information in livestock', *animal*, 17(11), p. 100996. Available at: https://doi.org/10.1016/j.animal.2023.100996.

Bertoni, A. *et al.* (2021) 'Dual-Purpose Water Buffalo Production Systems in Tropical Latin America: Bases for a Sustainable Model', *Animals*, 11(10), p. 2910. Available at: https://doi.org/10.3390/ani11102910.

Beynon, S.E. *et al.* (2015) 'Population structure and history of the Welsh sheep breeds determined by whole genome genotyping', *BMC Genetics*, 16(1), p. 65. Available at: https://doi.org/10.1186/s12863-015-0216-x.

Bharadwaj, A. *et al.* (no date) 'Association of breed characteristics with milk production in Murrah buffaloes'.

Bibi, F. (2013) 'A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics', *BMC Evolutionary Biology*, 13(1), p. 166. Available at: https://doi.org/10.1186/1471-2148-13-166.

Biegelmeyer, P. *et al.* (2016) 'Linkage disequilibrium, persistence of phase and effective population size estimates in Hereford and Braford cattle', *BMC Genetics*, 17(1), p. 32. Available at: https://doi.org/10.1186/s12863-016-0339-8.

Biffani, S., Gómez, M. and Cesarani, A. (2021) 'Pitfalls and opportunities of genetic and genomic evaluation in the Buffalo species: experiences from Italy', *Interbull Bulletin*, (56), pp. 94–101.

Bijl, E. *et al.* (2014) 'Phosphorylation of αS1-casein is regulated by different genes', *Journal of Dairy Science*, 97(11), pp. 7240–7246. Available at: https://doi.org/10.3168/jds.2014-8061.

Bilal, M., Suleman, M. and Kakar, Dr.A.R. (2006) 'Buffalo: Black gold of Pakistan', *Livestock Res Rural Dev*, 8.

Bindesbøll, C. *et al.* (2020) 'NBEAL1 controls SREBP2 processing and cholesterol metabolism and is a susceptibility locus for coronary artery disease', *Scientific Reports*, 10(1), p. 4528. Available at: https://doi.org/10.1038/s41598-020-61352-0.

Biscarini, F. *et al.* (2015) 'Challenges and opportunities in genetic improvement of local livestock breeds', *Frontiers in Genetics*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00033 (Accessed: 19 April 2023).

Blackburn, H. (2012) 'Genetic Selection and Conservation of Genetic Diversity*', *Reproduction in Domestic Animals*, 47(s4), pp. 249–254. Available at: https://doi.org/10.1111/j.1439-0531.2012.02083.x.

Boegheim, I.J.M. *et al.* (2017) 'Current insights into the molecular genetic basis of dwarfism in livestock', *The Veterinary Journal*, 224, pp. 64–75. Available at: https://doi.org/10.1016/j.tvjl.2017.05.014.

Boettcher, P.J. *et al.* (2015) 'Genetic resources and genomics for adaptation of livestock to climate change', *Frontiers in Genetics*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2014.00461 (Accessed: 21 March 2023).

Bolormaa, S. *et al.* (2021) 'A conditional multi-trait sequence GWAS discovers pleiotropic candidate genes and variants for sheep wool, skin wrinkle and breech cover traits', *Genetics Selection Evolution*, 53(1), p. 58. Available at: https://doi.org/10.1186/s12711-021-00651-0.

Bordbar, F. *et al.* (2019) 'Identification of muscle-specific candidate genes in Simmental beef cattle using imputed next generation sequencing', *PLOS ONE*, 14(10), p. e0223671. Available at: https://doi.org/10.1371/journal.pone.0223671.

Borghese, A. (2013a) 'Buffalo Livestock and Products in Europe', *Buffalo Bulletin* [Preprint], (1).

Borghese, A. (2013b) 'BUFFALO LIVESTOCK AND PRODUCTS IN EUROPE', 7.

Borghese, A. and Moioli, B. (2016) 'Buffalo: Mediterranean Region', *Reference Module in Food Science*, pp. B978-0-08-100596-5.21232-8. Available at: https://doi.org/10.1016/B978-0-08-100596-5.21232-8.

Bosse, M. *et al.* (2012) 'Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape', *PLOS Genetics*, 8(11), p. e1003100. Available at: https://doi.org/10.1371/journal.pgen.1003100.

Bovo, S. *et al.* (2020) 'Whole-genome sequencing of European autochthonous and commercial pig breeds allows the detection of signatures of selection for adaptation of genetic resources to different breeding and production systems', *Genetics Selection Evolution*, 52(1), p. 33. Available at: https://doi.org/10.1186/s12711-020-00553-7.

Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) 'An Expanded View of Complex Traits: From Polygenic to Omnigenic', *Cell*, 169(7), pp. 1177–1186. Available at: https://doi.org/10.1016/j.cell.2017.05.038.

Bradley, A. and Green, M. (2005) 'Use and interpretation of somatic cell count data in dairy cows', *In Practice*, 27(6), pp. 310–315. Available at: https://doi.org/10.1136/inpract.27.6.310.

Breider (2019) *Short communication: Heritability of methane production and genetic correlations with milk yield and body weight in Holstein-Friesian dairy cows | Elsevier Enhanced Reader*. Available at: https://doi.org/10.3168/jds.2018-15909.

Brito, L.F. *et al.* (2021) 'Review: Genetic selection of high-yielding dairy cattle toward sustainable farming systems in a rapidly changing world', *Animal*, 15, p. 100292. Available at: https://doi.org/10.1016/j.animal.2021.100292.

Bronzo, V. *et al.* (2020) 'The Role of Innate Immune Response and Microbiome in Resilience of Dairy Cattle to Disease: The Mastitis Model', *Animals*, 10(8), p. 1397. Available at: https://doi.org/10.3390/ani10081397.

Broquet, T. *et al.* (2010) 'Genetic Bottlenecks Driven by Population Disconnection', *Conservation Biology*, 24(6), pp. 1596–1605. Available at: https://doi.org/10.1111/j.1523-1739.2010.01556.x.

Browning, S.R. and Browning, B.L. (2007) 'Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering', *The*

American Journal of Human Genetics, 81(5), pp. 1084–1097. Available at: https://doi.org/10.1086/521987.

Bruford, M.W. *et al.* (2015) 'Prospects and challenges for the conservation of farm animal genomic resources, 2015-2025', *Frontiers in Genetics*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00314 (Accessed: 18 April 2023).

Bruford, M.W., Bradley, D.G. and Luikart, G. (2003) 'DNA markers reveal the complexity of livestock domestication', *Nature Reviews Genetics*, 4(11), pp. 900–910. Available at: https://doi.org/10.1038/nrg1203.

Bruford, M.W. and Wayne, R.K. (1993) 'Microsatellites and their application to population genetic studies', *Current Opinion in Genetics & Development*, 3(6), pp. 939–943. Available at: https://doi.org/10.1016/0959-437X(93)90017-J.

Brym, P. *et al.* (2016) 'Genome-wide association study for host response to bovine leukemia virus in Holstein cows', *Veterinary Immunology and Immunopathology*, 175, pp. 24–35. Available at: https://doi.org/10.1016/j.vetimm.2016.04.012.

Buitenhuis, B. *et al.* (2016) 'Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle', *BMC Genetics*, 17(1), p. 114. Available at: https://doi.org/10.1186/s12863-016-0421-2.

Burrow, H. and Goddard, M. (2023) 'Application of Genetics and Genomics in Livestock Production', *Agriculture*, 13(2), p. 386. Available at: https://doi.org/10.3390/agriculture13020386.

Burton, J.A., Hedges, S. and Mustari, A.H. (2005) 'The taxonomic status, distribution and conservation of the lowland anoa Bubalus depressicornis and mountain anoa Bubalus quarlesi', *Mammal Review*, 35(1), pp. 25–50. Available at: https://doi.org/10.1111/j.1365-2907.2005.00048.x.

Butte, M.J. *et al.* (2012) 'CD28 Costimulation Regulates Genome-Wide Effects on Alternative Splicing', *PLOS ONE*, 7(6), p. e40032. Available at: https://doi.org/10.1371/journal.pone.0040032.

Butts, C.T. (2008) 'network: A Package for Managing Relational Data in R', *Journal of Statistical Software*, 24, pp. 1–36. Available at: https://doi.org/10.18637/jss.v024.i02.

de Camargo, G. *et al.* (2015) 'Prospecting major genes in dairy buffaloes', *BMC Genomics*, 16(1), p. 872. Available at: https://doi.org/10.1186/s12864-015-1986-2.

Canario, L. *et al.* (2013) 'Genetics of behavioural adaptation of livestock to farming conditions', *animal*, 7(3), pp. 357–377. Available at: https://doi.org/10.1017/S1751731112001978.

Cañas-Álvarez, J.J. *et al.* (2015) 'Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip1', *Journal of Animal Science*, 93(11), pp. 5164–5174. Available at: https://doi.org/10.2527/jas.2015-9271.

Cañas-Álvarez, J.J. *et al.* (2016) 'Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip1', *Journal of Animal Science*, 94(7), pp. 2779–2788. Available at: https://doi.org/10.2527/jas.2016-0425.

Caroprese, M. *et al.* (2010) 'Relationship between cortisol response to stress and behavior, immune profile, and production performance of dairy ewes', *Journal of Dairy Science*, 93(6), pp. 2395–2403. Available at: https://doi.org/10.3168/jds.2009-2604.

Cartwright, S.L. *et al.* (2023) 'Impact of heat stress on dairy cattle and selection strategies for thermotolerance: a review', *Frontiers in Veterinary Science*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2023.1198697 (Accessed: 13 November 2023).

Casto-Rebollo, C. *et al.* (2020) 'Identification of functional mutations associated with environmental variance of litter size in rabbits', *Genetics Selection Evolution*, 52(1), p. 22. Available at: https://doi.org/10.1186/s12711-020-00542-w.

Catillo, G. *et al.* (2002) 'Effects of Age and Calving Season on Lactation Curves of Milk Production Traits in Italian Water Buffaloes', *Journal of Dairy Science*, 85(5), pp. 1298–1306. Available at: https://doi.org/10.3168/jds.S0022-0302(02)74194-5.

Catozzi, C. *et al.* (2017) 'The microbiota of water buffalo milk during mastitis', *PLOS ONE*, 12(9), p. e0184710. Available at: https://doi.org/10.1371/journal.pone.0184710.

Catozzi, C. *et al.* (2020) 'In-vitro effect of heat stress on bovine monocytes lifespan and polarization', *Immunobiology*, 225(2), p. 151888. Available at: https://doi.org/10.1016/j.imbio.2019.11.023.

Ceballos, F.C. *et al.* (2018) 'Runs of homozygosity: windows into population history and trait architecture', *Nature Reviews Genetics*, 19(4), pp. 220–234. Available at: https://doi.org/10.1038/nrg.2017.109.

Cesarani, A. *et al.* (2021) 'Genomic investigation of milk production in Italian buffalo', *Italian Journal of Animal Science*, 20(1), pp. 539–547. Available at: https://doi.org/10.1080/1828051X.2021.1902404.

Chakraborty, A., De, P. and Saha, S. (2023) 'Structure–function relationship of α-crystallin in the context of vertebrate lens evolution and its role in eye disorders', *Journal of Proteins and Proteomics*, 14(1), pp. 25–41. Available at: https://doi.org/10.1007/s42485-022-00101-5.

Chalkiadaki, K. *et al.* (2023) 'Reversal of memory and autism-related phenotypes in Tsc2+/− mice via inhibition of Nlgn1', *Frontiers in Cell and Developmental Biology*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fcell.2023.1205112 (Accessed: 26 December 2023).

Chang, C.C. *et al.* (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*, 4(1), pp. s13742-015-0047–8. Available at: https://doi.org/10.1186/s13742-015-0047-8.

Chang, Y.-Y. *et al.* (2022) 'Conformational Changes of α-Crystallin Proteins Induced by Heat Stress', *International Journal of Molecular Sciences*, 23(16), p. 9347. Available at: https://doi.org/10.3390/ijms23169347.

Chauhan, S.S. *et al.* (2021) 'Impacts of heat stress on immune responses and oxidative stress in farm animals and nutritional strategies for amelioration', *International Journal of Biometeorology*, 65(7), pp. 1231–1244. Available at: https://doi.org/10.1007/s00484-021-02083-3.

Chen, C. *et al.* (2021) 'Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner', *Cellular & Molecular Immunology*, 18(7), pp. 1844–1846. Available at: https://doi.org/10.1038/s41423-021-00699-4.

Chen, C. *et al.* (2022) 'AUP1 regulates lipid metabolism and induces lipid accumulation to accelerate the progression of renal clear cell carcinoma', *Cancer Science*, 113(8), pp. 2600–2615. Available at: https://doi.org/10.1111/cas.15445.

Chen, N. *et al.* (2019) 'The transcription factor NKX1-2 promotes adipogenesis and may contribute to a balance between adipocyte and osteoblast differentiation', *Journal of Biological Chemistry*, 294(48), pp. 18408–18420. Available at: https://doi.org/10.1074/jbc.RA119.007967.

Chen, P. *et al.* (2020) 'Circadian Regulator CLOCK Recruits Immune-Suppressive Microglia into the GBM Tumor Microenvironment', *Cancer Discovery*, 10(3), pp. 371–381. Available at: https://doi.org/10.1158/2159-8290.CD-19-0400.

Chen, S.-Y. *et al.* (2022) 'Identifying pleiotropic variants and candidate genes for fertility and reproduction traits in Holstein cattle via association studies based on imputed whole-genome sequence genotypes', *BMC Genomics*, 23(1), p. 331. Available at: https://doi.org/10.1186/s12864-022-08555-z.

Cheng, K. *et al.* (2014) 'Divergent effects of muscarinic receptor subtype gene ablation on murine colon tumorigenesis reveals association of M3R and zinc finger protein 277 expression in colon neoplasia', *Molecular Cancer*, 13(1), p. 77. Available at: https://doi.org/10.1186/1476-4598-13-77.

Cheng, M., McCarl, B. and Fei, C. (2022) 'Climate Change and Livestock Production: A Literature Review', *Atmosphere*, 13(1), p. 140. Available at: https://doi.org/10.3390/atmos13010140.

Chhotaray, S. *et al.* (2021) 'Genome-wide association study reveals genes crucial for coat color production in Vrindavani cattle', *Livestock Science*, 247, p. 104476. Available at: https://doi.org/10.1016/j.livsci.2021.104476.

Chirone, R. *et al.* (2022) 'Life Cycle Assessment of buffalo milk: A case study of three farms in southern Italy', *Journal of Cleaner Production*, 365, p. 132816. Available at: https://doi.org/10.1016/j.jclepro.2022.132816.

Cho, V. *et al.* (2014) 'The RNA-binding protein hnRNPLL induces a T cell alternative splicing program delineated by differential intron retention in polyadenylated RNA', *Genome Biology*, 15(1), p. R26. Available at: https://doi.org/10.1186/gb-2014-15-1-r26.

Choudhury, A. (2014) 'Wild water buffalo Bubalus arnee (Kerr, 1792)', in M. Melletti and J. Burton (eds) *Ecology, Evolution and Behaviour of Wild Cattle*. 1st edn. Cambridge University Press, pp. 255–301. Available at: https://doi.org/10.1017/CBO9781139568098.018.

Clark, A.G. *et al.* (2005) 'Ascertainment bias in studies of human genome-wide polymorphism', *Genome Research*, 15(11), pp. 1496–1502. Available at: https://doi.org/10.1101/gr.4107905.

Clark, P.U. *et al.* (2009) 'The Last Glacial Maximum', *Science*, 325(5941), pp. 710–714. Available at: https://doi.org/10.1126/science.1172873.

Clark, P.U. and Mix, A.C. (2002) 'Ice sheets and sea level of the Last Glacial Maximum', *Quaternary Science Reviews*, 21(1), pp. 1–7. Available at: https://doi.org/10.1016/S0277-3791(01)00118-4.

Clingingsmith, D. and Williamson, J.G. (2008) 'Deindustrialization in 18th and 19th century India: Mughal decline, climate shocks and British industrial ascent', *Explorations in Economic History*, 45(3), pp. 209–234. Available at: https://doi.org/10.1016/j.eeh.2007.11.002.

Cobanoglu, O. *et al.* (2006) 'Effects of the Signal Transducer and Activator of Transcription 1 (STAT1) Gene on Milk Production Traits in Holstein Dairy Cattle', *Journal of Dairy Science*, 89(11), pp. 4433–4437. Available at: https://doi.org/10.3168/jds.S0022-0302(06)72491-2.

Cockrill, W.R. (1981) 'The Water Buffalo: A Review', *British Veterinary Journal*, 137(1), pp. 8–16. Available at: https://doi.org/10.1016/S0007-1935(17)31782-7.

Colli, L., Milanesi, M., Talenti, A., *et al.* (2018) 'Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes', *Genetics Selection Evolution*, 50(1), p. 58. Available at: https://doi.org/10.1186/s12711-018-0422-x.

Colli, L., Milanesi, M., Vajana, E., *et al.* (2018) 'New Insights on Water Buffalo Genomic Diversity and Post-Domestication Migration Routes From Medium Density SNP Chip Data', *Frontiers in Genetics*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00053 (Accessed: 17 October 2022).

Collier, R.J., Renquist, B.J. and Xiao, Y. (2017) 'A 100-Year Review: Stress physiology including heat stress', *Journal of Dairy Science*, 100(12), pp. 10367–10380. Available at: https://doi.org/10.3168/jds.2017-13676.

Consortium, E.C.G.D. (2006) 'Marker-assisted conservation of European cattle breeds: an evaluation', *Animal Genetics*, 37(5), pp. 475–481. Available at: https://doi.org/10.1111/j.1365-2052.2006.01511.x.

Conte, G. *et al.* (2021) 'Transcriptome Adaptation of the Ovine Mammary Gland to Dietary Supplementation of Extruded Linseed', *Animals*, 11(9), p. 2707. Available at: https://doi.org/10.3390/ani11092707.

Corbin, L. j. *et al.* (2012) 'Estimation of historical effective population size using linkage disequilibria with marker data', *Journal of Animal Breeding and Genetics*, 129(4), pp. 257–270. Available at: https://doi.org/10.1111/j.1439-0388.2012.01003.x.

Coroian, A. *et al.* (2011) 'Study on some milk production indices of Romanian buffalo', 3(1).

Cortes, O., Cañon, J. and Gama, L.T. (2022) 'Applications of Microsatellites and Single Nucleotide Polymorphisms for the Genetic Characterization of Cattle and Small Ruminants: An Overview', *Ruminants*, 2(4), pp. 456–470. Available at: https://doi.org/10.3390/ruminants2040032.

Costa, A. *et al.* (2020) 'Phenotypic Characterization of Milk Yield and Quality Traits in a Large Population of Water Buffaloes', *Animals*, 10(2), p. 327. Available at: https://doi.org/10.3390/ani10020327.

Costa, A. *et al.* (2021) 'Milk somatic cell count-derived traits as new indicators to monitor udder health in dairy buffaloes', *Italian Journal of Animal Science*, 20(1), pp. 548–558. Available at: https://doi.org/10.1080/1828051X.2021.1899856.

Crisci, E. *et al.* (2020) 'Distinctive Cellular and Metabolic Reprogramming in Porcine Lung Mononuclear Phagocytes Infected With Type 1 PRRSV Strains', *Frontiers in Immunology*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.588411 (Accessed: 21 November 2022).

Croft, D.A. *et al.* (2006) 'Fossil Remains of a New, Diminutive Bubalus (Artiodactyla: Bovidae: Bovini) from Cebu Island, Philippines', *Journal of Mammalogy*, 87(5), pp. 1037–1051. Available at: https://doi.org/10.1644/06-MAMM-A-018R.1.

Cruz, L.C. (2007) 'Trends in buffalo production in Asia', *Italian Journal of Animal Science*, 6(sup2), pp. 9–24. Available at: https://doi.org/10.4081/ijas.2007.s2.9.

de la Cruz-Cruz, L.A. *et al.* (2019) 'Effects of weaning on the stress responses and productivity of water buffalo in different breeding systems: A review', *Livestock Science*, 226, pp. 73–81. Available at: https://doi.org/10.1016/j.livsci.2019.05.020.

Csilléry, K. *et al.* (2010) 'Approximate Bayesian Computation (ABC) in practice', *Trends in Ecology & Evolution*, 25(7), pp. 410–418. Available at: https://doi.org/10.1016/j.tree.2010.04.001.

Cumer, T. *et al.* (2021) 'Double-digest RAD-sequencing: do pre- and post-sequencing protocol parameters impact biological results?', *Molecular Genetics and Genomics*, 296(2), pp. 457–471. Available at: https://doi.org/10.1007/s00438-020-01756-9.

Cunningham, F. *et al.* (2019) 'Ensembl 2019', *Nucleic Acids Research*, 47(D1), pp. D745–D751. Available at: https://doi.org/10.1093/nar/gky1113.

Curaudeau, M., Rozzi, R. and Hassanin, A. (2021) 'The genome of the lowland anoa (Bubalus depressicornis) illuminates the origin of river and swamp buffalo', *Molecular Phylogenetics and Evolution*, 161, p. 107170. Available at: https://doi.org/10.1016/j.ympev.2021.107170.

Curik, I., Ferenčaković, M. and Sölkner, J. (2014) 'Inbreeding and runs of homozygosity: A possible solution to an old problem', *Livestock Science*, 166, pp. 26–34. Available at: https://doi.org/10.1016/j.livsci.2014.05.034.

Dadousis, C. *et al.* (2022) 'Genomic inbreeding coefficients using imputed genotypes: Assessing different estimators in Holstein-Friesian dairy cows', *Journal of Dairy Science*, 105(7), pp. 5926–5945. Available at: https://doi.org/10.3168/jds.2021-21125.

Daetwyler, H.D. *et al.* (2014) 'Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle', *Nature Genetics*, 46(8), pp. 858–865. Available at: https://doi.org/10.1038/ng.3034.

Dai, P. *et al.* (2019) 'The roles of Nrf2 and autophagy in modulating inflammation mediated by TLR4 - NFκB in A549 cell exposed to layer house particulate matter 2.5 (PM2.5)', *Chemosphere*, 235, pp. 1134–1145. Available at: https://doi.org/10.1016/j.chemosphere.2019.07.002.

Darang, E. *et al.* (2022) 'Identification of Key Genes and Potential Pathways Associated with Mastitis Induced by E. coli', *Biochemical Genetics* [Preprint]. Available at: https://doi.org/10.1007/s10528-022-10254-z.

Davoudi, P. *et al.* (2020) 'EXPLORING THE STRUCTURE OF HAPLOTYPE BLOCKS, RUNS OF HOMOZYGOSITY AND EFFECTIVE POPULATION SIZE IN KHUZESTANI RIVER BUFFALO', *Slovak Journal of Animal Science*, 53(02), pp. 67–77.

De la Cruz-Cruz, L. *et al.* (2014) 'The behaviour and productivity of water buffalo in different breeding systems: a review', *Veterinární Medicína*, 59(No. 4), pp. 181–193. Available at: https://doi.org/10.17221/7479-VETMED.

De León, C. *et al.* (2021) 'Research Article Selection of genomic regions and genes associated with adaptation and fertility traits in two Colombian creole cattle breeds', *Genetics and Molecular Research*, 20(3). Available at: https://doi.org/10.4238/gmr18882.

De Rosa, G. *et al.* (2009) 'Behavior and milk production of buffalo cows as affected by housing system', *Journal of Dairy Science*, 92(3), pp. 907–912. Available at: https://doi.org/10.3168/jds.2008-1157.

Deb, G.K. *et al.* (2016) 'Safe and Sustainable Traditional Production: The Water Buffalo in Asia', *Frontiers in Environmental Science*, 4. Available at: https://www.frontiersin.org/articles/10.3389/fenvs.2016.00038 (Accessed: 23 October 2023).

Dediu, D. and Christiansen, M.H. (2016) 'Language Evolution: Constraints and Opportunities From Modern Genetics', *Topics in Cognitive Science*, 8(2), pp. 361–370. Available at: https://doi.org/10.1111/tops.12195.

Demasius, W. *et al.* (2013) 'Monitoring the immune response to vaccination with an inactivated vaccine associated to bovine neonatal pancytopenia by deep sequencing transcriptome analysis in cattle', *Veterinary Research*, 44(1), p. 93. Available at: https://doi.org/10.1186/1297-9716-44-93.

Deng, T. *et al.* (2015) 'Molecular cloning and expression analysis of the STAT1 gene in the water buffalo (Bubalus bubalis)', *Tropical Animal Health and Production*, 47(1), pp. 53–59. Available at: https://doi.org/10.1007/s11250-014-0682-6.

Deng, T. *et al.* (2019) 'Genome-Wide SNP Data Revealed the Extent of Linkage Disequilibrium, Persistence of Phase and Effective Population Size in Purebred and Crossbred Buffalo Populations', *Frontiers in Genetics*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00688 (Accessed: 26 October 2022).

Deng, T.X. *et al.* (2016) 'Associations between polymorphisms of the STAT1 gene and milk production traits in water buffaloes1', *Journal of Animal Science*, 94(3), pp. 927–935. Available at: https://doi.org/10.2527/jas.2015-0139.

Derks, M.F.L. and Steensma, M. (2021) 'Review: Balancing Selection for Deleterious Alleles in Livestock', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.761728 (Accessed: 1 September 2023).

Devi, S. *et al.* (2013) 'Metabotropic glutamate receptor 6 signaling enhances TRPM1 calcium channel function and increases melanin content in human melanocytes', *Pigment Cell & Melanoma Research*, 26(3), pp. 348–356. Available at: https://doi.org/10.1111/pcmr.12083.

Dhanaraman, T. *et al.* (2020) 'RASSF effectors couple diverse RAS subfamily GTPases to the Hippo pathway', *Science Signaling*, 13(653), p. eabb4778. Available at: https://doi.org/10.1126/scisignal.abb4778.

Di Stasio, L. and Brugiapaglia, A. (2021) 'Current Knowledge on River Buffalo Meat: A Critical Analysis', *Animals*, 11(7), p. 2111. Available at: https://doi.org/10.3390/ani11072111.

Diamond, J. (2002) 'Evolution, consequences and future of plant and animal domestication', *Nature*, 418(6898), pp. 700–707. Available at: https://doi.org/10.1038/nature01019.

Dickinson, S.E. *et al.* (2018) 'Transcriptome profiles in peripheral white blood cells at the time of artificial insemination discriminate beef heifers with different fertility potential', *BMC Genomics*, 19(1), p. 129. Available at: https://doi.org/10.1186/s12864-018-4505-4.

Dickson, M.J. *et al.* (2022) 'The endometrial transcriptomic response to pregnancy is altered in cows after uterine infection', *PLOS ONE*, 17(3), p. e0265062. Available at: https://doi.org/10.1371/journal.pone.0265062.

Ding, R. *et al.* (2021) 'Genome-wide association studies reveals polygenic genetic architecture of litter traits in Duroc pigs', *Theriogenology*, 173, pp. 269–278. Available at: https://doi.org/10.1016/j.theriogenology.2021.08.012.

Ding, X. *et al.* (2022) 'A molecular marker of milk composition traits in NCAM2 gene of Chinese Holstein', *Animal Biotechnology*, 33(1), pp. 79–84. Available at: https://doi.org/10.1080/10495398.2020.1772802.

Ding, Z.-L. *et al.* (2012) 'Origins of domestic dog in Southern East Asia is supported by analysis of Y-chromosome DNA', *Heredity*, 108(5), pp. 507–514. Available at: https://doi.org/10.1038/hdy.2011.114.

Diniz, W.J.S. *et al.* (2020) 'FABP1 and SLC2A5 expression levels affect feed efficiency-related traits', *Agri Gene*, 15, p. 100100. Available at: https://doi.org/10.1016/j.aggene.2019.100100.

Do, D.N. *et al.* (2014) 'Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake', *Frontiers in Genetics*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2014.00307 (Accessed: 11 September 2023).

Doekes, H.P., Bijma, P. and Windig, J.J. (2021) 'How Depressing Is Inbreeding? A Meta-Analysis of 30 Years of Research on the Effects of Inbreeding in Livestock', *Genes*, 12(6), p. 926. Available at: https://doi.org/10.3390/genes12060926.

Dolebo, A.T. *et al.* (2019) 'Genome-wide scans identify known and novel regions associated with prolificacy and reproduction traits in a sub-Saharan African indigenous sheep (Ovis aries)', *Mammalian Genome*, 30(11), pp. 339–352. Available at: https://doi.org/10.1007/s00335-019-09820-5.

Dong, W. *et al.* (2014) 'The Early Pleistocene water buffalo associated with Gigantopithecus from Chongzuo in southern China', *Quaternary International*, 354, pp. 86–93. Available at: https://doi.org/10.1016/j.quaint.2013.12.054.

Doublet, A.-C. *et al.* (2019) 'The impact of genomic selection on genetic diversity and genetic gain in three French dairy cattle breeds', *Genetics Selection Evolution*, 51(1), p. 52. Available at: https://doi.org/10.1186/s12711-019-0495-1.

Du, C. *et al.* (2019) 'Systematic analyses for candidate genes of milk production traits in water buffalo (Bubalus Bubalis)', *Animal Genetics*, 50(3), pp. 207–216. Available at: https://doi.org/10.1111/age.12739.

Du, P. *et al.* (2021) 'Identification of a novel SUOX pathogenic variants as the cause of isolated sulfite oxidase deficiency in a Chinese pedigree', *Molecular Genetics & Genomic Medicine*, 9(2), p. e1590. Available at: https://doi.org/10.1002/mgg3.1590.

Dudley, N. and Alexander, S. (2017) 'Agriculture and biodiversity: a review', *Biodiversity*, 18(2–3), pp. 45–49. Available at: https://doi.org/10.1080/14888386.2017.1351892.

Dumbuya, H., Hafez, S.Y. and Oancea, E. (2020) 'Cross talk between calcium and ROS regulate the UVA-induced melanin response in human melanocytes', *The FASEB Journal*, 34(9), pp. 11605–11623. Available at: https://doi.org/10.1096/fj.201903024R.

Dunn, R.J.H. *et al.* (2014) 'Analysis of heat stress in UK dairy cattle and impact on milk yields', *Environmental Research Letters*, 9(6), p. 064006. Available at: https://doi.org/10.1088/1748-9326/9/6/064006.

Duruz, S. *et al.* (2019) 'Rapid identification and interpretation of gene–environment associations using the new R.SamBada landscape genomics pipeline', *Molecular Ecology Resources*, 19(5), pp. 1355–1365. Available at: https://doi.org/10.1111/1755-0998.13044.

Dutta, A. *et al.* (2021) 'U2af1 is required for survival and function of hematopoietic stem/progenitor cells', *Leukemia*, 35(8), pp. 2382–2398. Available at: https://doi.org/10.1038/s41375-020-01116-x.

Dyck, M.K. *et al.* (2014) 'Reproductive technologies and the porcine embryonic transcriptome', *Animal Reproduction Science*, 149(1), pp. 11–18. Available at: https://doi.org/10.1016/j.anireprosci.2014.05.013.

Dysin, A.P., Barkova, O.Y. and Pozovnikova, M.V. (2021) 'The Role of microRNAs in the Mammary Gland Development, Health, and Function of Cattle, Goats, and Sheep', *Non-Coding RNA*, 7(4), p. 78. Available at: https://doi.org/10.3390/ncrna7040078.

Ebrahimi-Fakhari, D. *et al.* (2018) 'Clinical and genetic characterization of AP4B1-associated SPG47', *American Journal of Medical Genetics Part A*, 176(2), pp. 311–318. Available at: https://doi.org/10.1002/ajmg.a.38561.

Eden, E. *et al.* (2009) 'GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists', *BMC Bioinformatics*, 10, p. 48. Available at: https://doi.org/10.1186/1471-2105-10-48.

Egashira, E.M. *et al.* (2022) 'Investigation of biomarkers in Endometriosis-associated infertility: Systematic Review', *Anais da Academia Brasileira de Ciências*, 94, p. e20211572. Available at: https://doi.org/10.1590/0001-3765202220211572.

Ekiz, B. *et al.* (2018) 'Carcass and Meat Quality of Male and Female Water Buffaloes Finished Under an Intensive Production System', *Annals of Animal Science*, 18(2), pp. 557–574. Available at: https://doi.org/10.1515/aoas-2017-0036.

Eldawy, M.H. *et al.* (2021) 'Milk production potential and reproductive performance of Egyptian buffalo cows', *Tropical Animal Health and Production*, 53(2), p. 282. Available at: https://doi.org/10.1007/s11250-021-02722-2.

Escarcha, J.F. *et al.* (2018) 'Understanding climate change impacts on water buffalo production through farmers' perceptions', *Climate Risk Management*, 20, pp. 50–63. Available at: https://doi.org/10.1016/j.crm.2018.03.003.

Escarcha, J.F. *et al.* (2020) 'Livelihoods transformation and climate change adaptation: The case of smallholder water buffalo farmers in the Philippines', *Environmental Development*, 33, p. 100468. Available at: https://doi.org/10.1016/j.envdev.2019.100468.

Etchevers, L. *et al.* (2023) 'Exogenous ACTH stimulus during the preovulatory period alters patterns of leukocyte recruitment in the ovary of dairy cows', *Theriogenology*, 195, pp. 176–186. Available at: https://doi.org/10.1016/j.theriogenology.2022.10.028.

Eusebi, P.G., Martinez, A. and Cortes, O. (2020) 'Genomic Tools for Effective Conservation of Livestock Breed Diversity', *Diversity*, 12(1), p. 8. Available at: https://doi.org/10.3390/d12010008.

Excoffier, L. *et al.* (2013) 'Robust Demographic Inference from Genomic and SNP Data', *PLOS Genetics*, 9(10), p. e1003905. Available at: https://doi.org/10.1371/journal.pgen.1003905.

Excoffier, L. *et al.* (2021) 'fastsimcoal2: demographic inference under complex evolutionary scenarios', *Bioinformatics (Oxford, England)*, 37(24), pp. 4882–4885. Available at: https://doi.org/10.1093/bioinformatics/btab468.

Excoffier, L., Foll, M. and Petit, R.J. (2009) 'Genetic Consequences of Range Expansions', *Annual Review of Ecology, Evolution, and Systematics*, 40(1), pp. 481–501. Available at: https://doi.org/10.1146/annurev.ecolsys.39.110707.173414.

Excoffier, L. and Lischer, H.E.L. (2010) 'Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology Resources*, 10(3), pp. 564–567. Available at: https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Falker-Gieske, C. *et al.* (2019) 'GWAS for Meat and Carcass Traits Using Imputed Sequence Level Genotypes in Pooled F2-Designs in Pigs', *G3 Genes|Genomes|Genetics*, 9(9), pp. 2823–2834. Available at: https://doi.org/10.1534/g3.119.400452.

Fallahi, M.H. *et al.* (2020) 'Assessment of Genetic Diversity in Azerbaijani Buffalo Population in Iran Based on Runs of Homozygosity Stretches', *Russian Journal of Genetics*, 56(10), pp. 1224–1231. Available at: https://doi.org/10.1134/S102279542010004X.

Fan, B. *et al.* (2010) 'Development and Application of High-density SNP Arrays in Genomic Studies of Domestic Animals', *Asian-Australasian Journal of Animal Sciences*, 23(7), pp. 833–847. Available at: https://doi.org/10.5713/ajas.2010.r.03.

Fang, Z.-H. and Pausch, H. (2019) 'Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle', *BMC Genomics*, 20(1), p. 695. Available at: https://doi.org/10.1186/s12864-019-6066-6.

Farmanullah, U. *et al.* (2020) 'Adaptive Molecular Evolution of *AKT3* Gene for Positive Diversifying Selection in Mammals', *BioMed Research International*, 2020, p. e2584627. Available at: https://doi.org/10.1155/2020/2584627.

Fazaa, N.A., Dunn, J.C. and Whittingham, M.J. (2018) 'Evaluation of the Ecosystem Services of the Central Marsh in Southern Iraq', *Baghdad Science Journal*, 15(4), p. 0369. Available at: https://doi.org/10.21123/bsj.2018.15.4.0369.

Felius, M. *et al.* (2014) 'On the History of Cattle Genetic Resources', *Diversity*, 6(4), pp. 705–750. Available at: https://doi.org/10.3390/d6040705.

Ferenčaković, M. *et al.* (2013) 'Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations', *Journal of Animal Breeding and Genetics*, 130(4), pp. 286–293. Available at: https://doi.org/10.1111/jbg.12012.

Fernández, M.E. *et al.* (2013) 'Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd', *Genetics and Molecular Biology*, 36, pp. 185–191. Available at: https://doi.org/10.1590/S1415-47572013000200008.

Ferrara, G. *et al.* (2023) 'Seroprevalence and Risk Factors for Bovine Coronavirus Infection among Dairy Cattle and Water Buffalo in Campania Region, Southern Italy', *Animals*, 13(5), p. 772. Available at: https://doi.org/10.3390/ani13050772.

Fessler, E. *et al.* (2020) 'A pathway coordinated by DELE1 relays mitochondrial stress to the cytosol', *Nature*, 579(7799), pp. 433–437. Available at: https://doi.org/10.1038/s41586-020-2076-4.

Fijarczyk, A. and Babik, W. (2015) 'Detecting balancing selection in genomes: limits and prospects', *Molecular Ecology*, 24(14), pp. 3529–3545. Available at: https://doi.org/10.1111/mec.13226.

Fischer, E.K. and Drago, A. (2017) 'A molecular pathway analysis stresses the role of inflammation and oxidative stress towards cognition in schizophrenia', *Journal of Neural Transmission*, 124(7), pp. 765–774. Available at: https://doi.org/10.1007/s00702-017-1730-y.

Flori, L. *et al.* (2012) 'A Quasi-Exclusive European Ancestry in the Senepol Tropical Cattle Breed Highlights the Importance of the slick Locus in Tropical Adaptation', *PLOS ONE*, 7(5), p. e36133. Available at: https://doi.org/10.1371/journal.pone.0036133.

Fortes, M.R.S. *et al.* (2018) 'Pre- and post-puberty expression of genes and proteins in the uterus of Bos indicus heifers: the luteal phase effect post-puberty', *Animal Genetics*, 49(6), pp. 539–549. Available at: https://doi.org/10.1111/age.12721.

Fox, R.J. *et al.* (2019) 'Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1768), p. 20180174. Available at: https://doi.org/10.1098/rstb.2018.0174.

Frantz, L.A. *et al.* (2013) 'Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus', *Genome Biology*, 14(9), p. R107. Available at: https://doi.org/10.1186/gb-2013-14-9-r107.

Frantz, L.A.F. *et al.* (2016) 'Genomic and archaeological evidence suggest a dual origin of domestic dogs', *Science*, 352(6290), pp. 1228–1231. Available at: https://doi.org/10.1126/science.aaf3161.

Freedman, A.H. *et al.* (2014) 'Genome Sequencing Highlights the Dynamic Early History of Dogs', *PLOS Genetics*, 10(1), p. e1004016. Available at: https://doi.org/10.1371/journal.pgen.1004016.

Frezarim, G.B. *et al.* (2022) 'Genes and proteins associated with ribeye area and meat tenderness in a commercial Nellore cattle population', *Genome*, 65(4), pp. 229–240. Available at: https://doi.org/10.1139/gen-2020-0163.

Friedrich, J., Brand, B. and Schwerin, M. (2015) 'Genetics of cattle temperament and its impact on livestock production and breeding – a review', *Archives Animal Breeding*, 58(1), pp. 13–21. Available at: https://doi.org/10.5194/aab-58-13-2015.

Fu, M. *et al.* (2022) 'Combined Transcriptomic and Proteomic of Corynebacterium pseudotuberculosis Infection in the Spleen of Dairy Goats', *Animals*, 12(23), p. 3270. Available at: https://doi.org/10.3390/ani12233270.

Galibert, F. *et al.* (2011) 'Toward understanding dog evolutionary and domestication history', *Comptes Rendus Biologies*, 334(3), pp. 190–196. Available at: https://doi.org/10.1016/j.crvi.2010.12.011.

Gandini, G.C. and Villa, E. (2003) 'Analysis of the cultural value of local livestock breeds: a methodology', *Journal of Animal Breeding and Genetics*, 120(1), pp. 1–11. Available at: https://doi.org/10.1046/j.1439-0388.2003.00365.x.

Gao, J. *et al.* (2020) 'Genomic Characteristics and Selection Signatures in Indigenous Chongming White Goat (Capra hircus)', *Frontiers in Genetics*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00901 (Accessed: 19 November 2022).

Gao, L. *et al.* (2017) 'PTGER2 activation induces PTGS-2 and growth factor gene expression in endometrial epithelial cells of cattle', *Animal Reproduction Science*, 187, pp. 54–63. Available at: https://doi.org/10.1016/j.anireprosci.2017.10.005.

Gao, Z., Przeworski, M. and Sella, G. (2015) 'Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species', *Evolution*, 69(2), pp. 431–446. Available at: https://doi.org/10.1111/evo.12567.

Garcia-Diaz, M. *et al.* (2005) 'Structure–function studies of DNA polymerase lambda', *DNA Repair*, 4(12), pp. 1358–1367. Available at: https://doi.org/10.1016/j.dnarep.2005.09.001.

García-Roche, M. *et al.* (2021) 'Glucose and Fatty Acid Metabolism of Dairy Cows in a Total Mixed Ration or Pasture-Based System During Lactation', *Frontiers in Animal Science*, 2. Available at: https://www.frontiersin.org/articles/10.3389/fanim.2021.622500 (Accessed: 30 August 2023).

Garza-Brenner, E. *et al.* (2020) 'Influence of temperament-related genes on live weight traits of Charolais cows', *Revista Brasileira de Zootecnia*, 49, p. e20180121. Available at: https://doi.org/10.37496/rbz4920180121.

Gasca, J. *et al.* (2020) 'EDIL3 promotes epithelial–mesenchymal transition and paclitaxel resistance through its interaction with integrin αVβ3 in cancer cells', *Cell Death Discovery*, 6(1), pp. 1–14. Available at: https://doi.org/10.1038/s41420-020-00322-x.

Gautier, M. *et al.* (2016) 'Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences', *Molecular Biology and Evolution*, 33(11), pp. 2801–2814. Available at: https://doi.org/10.1093/molbev/msw144.

Gebreyesus, G. *et al.* (2019) 'Multi-population GWAS and enrichment analyses reveal novel genomic regions and promising candidate genes underlying bovine milk fatty acid composition', *BMC Genomics*, 20(1), p. 178. Available at: https://doi.org/10.1186/s12864-019-5573-9.

Geetha, E., Chakravarty, A.K. and Kumar, K.V. (2006) 'Genetic Persistency of First Lactation Milk Yield Estimated Using Random Regression Model for Indian Murrah Buffaloes', *Asian-Australasian Journal of Animal Sciences*, 19(12), pp. 1696–1701. Available at: https://doi.org/10.5713/ajas.2006.1696.

Geibel, J. *et al.* (2021) 'How array design creates SNP ascertainment bias', *PLOS ONE*, 16(3), p. e0245178. Available at: https://doi.org/10.1371/journal.pone.0245178.

Gelayenew, B. and Asebe, G. (2016) 'Review on Major Assisted Reproductive Technologies Global Journal of Science Frontier Research: D Agriculture and Veterinary Review on Major Assisted

Reproductive Technologies', *Global Journal of Science Frontier Research: D Agriculture and Veterinary*, 16, pp. 51–55.

Georges, M., Charlier, C. and Hayes, B. (2019) 'Harnessing genomic information for livestock improvement', *Nature Reviews Genetics*, 20(3), pp. 135–156. Available at: https://doi.org/10.1038/s41576-018-0082-2.

Gepts, P. and Papa, R. (2003) 'Evolution during Domestication', in *Encycl Life Sci*. Available at: https://doi.org/10.1038/npg.els.0003071.

Germonpré, M. *et al.* (2009) 'Fossil dogs and wolves from Palaeolithic sites in Belgium, the Ukraine and Russia: osteometry, ancient DNA and stable isotopes', *Journal of Archaeological Science*, 36(2), pp. 473–490. Available at: https://doi.org/10.1016/j.jas.2008.09.033.

Ghafouri-Fard, S. *et al.* (2022) 'Interplay Between Non-Coding RNAs and Programmed Cell Death Proteins', *Frontiers in Oncology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fonc.2022.808475 (Accessed: 19 November 2022).

Ghoreishifar, S.M. *et al.* (2020) 'Genomic measures of inbreeding coefficients and genome-wide scan for runs of homozygosity islands in Iranian river buffalo, Bubalus bubalis', *BMC Genetics*, 21(1), p. 16. Available at: https://doi.org/10.1186/s12863-020-0824-y.

Ghosh, M. *et al.* (2019) 'The first comprehensive description of the expression profile of genes involved in differential body growth and the immune system of the Jeju Native Pig and miniature pig', *Amino Acids*, 51(3), pp. 495–511. Available at: https://doi.org/10.1007/s00726-018-2685-5.

Giani, A.M. *et al.* (2020) 'Long walk to genomics: History and current approaches to genome sequencing and assembly', *Computational and Structural Biotechnology Journal*, 18, pp. 9–19. Available at: https://doi.org/10.1016/j.csbj.2019.11.002.

Gicquel, E. *et al.* (2020) 'Impact of conservation measures on demography and genetic variability of livestock breeds', *animal*, 14(4), pp. 670–680. Available at: https://doi.org/10.1017/S1751731119002672.

Girardin, S.E. *et al.* (2021) 'The eIF2α kinase HRI in innate immunity, proteostasis, and mitochondrial stress', *The FEBS Journal*, 288(10), pp. 3094–3107. Available at: https://doi.org/10.1111/febs.15553.

Goddard, M.E. (2012) 'Uses of genomics in livestock agriculture', *Animal Production Science*, 52(3), pp. 73–77. Available at: https://doi.org/10.1071/AN11180.

Godde, C.M. *et al.* (2018) 'Grazing systems expansion and intensification: Drivers, dynamics, and trade-offs', *Global Food Security*, 16, pp. 93–105. Available at: https://doi.org/10.1016/j.gfs.2017.11.003.

Goldman, N. and Yang, Z. (1994) 'A codon-based model of nucleotide substitution for protein-coding DNA sequences.', *Molecular Biology and Evolution*, 11(5), pp. 725–736. Available at: https://doi.org/10.1093/oxfordjournals.molbev.a040153.

Gonzalez-Rivas, P.A. *et al.* (2020) 'Effects of heat stress on animal physiology, metabolism, and meat quality: A review', *Meat Science*, 162, p. 108025. Available at: https://doi.org/10.1016/j.meatsci.2019.108025.

Gorjanc, G. *et al.* (2015) 'Potential of genotyping-by-sequencing for genomic selection in livestock populations', *Genetics Selection Evolution*, 47(1), pp. 1–14. Available at: https://doi.org/10.1186/s12711-015-0102-z.

Gorski, M.M. *et al.* (2019) 'Complications of whole-exome sequencing for causal gene discovery in primary platelet secretion defects', *Haematologica*, 104(10), pp. 2084–2090. Available at: https://doi.org/10.3324/haematol.2018.204990.

Gouveia, J.J. de S. *et al.* (2014) 'Identification of selection signatures in livestock species', *Genetics and Molecular Biology*, 37, pp. 330–342. Available at: https://doi.org/10.1590/S1415-47572014000300004.

Gowane, G.R. and Vohra, V. (2022) 'Advances in Buffalo Breeding: A Journey from Classical Breeding to Genomic Selection', in M.S. Chauhan and N. Selokar (eds) *Biotechnological Applications in Buffalo Research*. Singapore: Springer, pp. 57–77. Available at: https://doi.org/10.1007/978-981-16-7531-7_3.

Grainger, S. and Willert, K. (2018) 'Mechanisms of Wnt signaling and control', *WIREs Systems Biology and Medicine*, 10(5), p. e1422. Available at: https://doi.org/10.1002/wsbm.1422.

Green, E.D. *et al.* (2020) 'Strategic vision for improving human health at The Forefront of Genomics', *Nature*, 586(7831), pp. 683–692. Available at: https://doi.org/10.1038/s41586-020-2817-4.

Groeneveld, L.F. *et al.* (2010) 'Genetic diversity in farm animals – a review', *Animal Genetics*, 41(s1), pp. 6–31. Available at: https://doi.org/10.1111/j.1365-2052.2010.02038.x.

Grossi, D.A. *et al.* (2017) 'Genetic diversity, extent of linkage disequilibrium and persistence of gametic phase in Canadian pigs', *BMC Genetics*, 18(1), p. 6. Available at: https://doi.org/10.1186/s12863-017-0473-y.

Guan, D. *et al.* (2016) 'Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus)', *Scientific Reports*, 6(1), p. 36372. Available at: https://doi.org/10.1038/srep36372.

Guan, X. *et al.* (2019) 'The Molecular Evolution and Functional Divergence of Lamprey Programmed Cell Death Genes', *Frontiers in Immunology*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2019.01382 (Accessed: 19 November 2022).

Guerrero Legarreta, I. *et al.* (2020) 'River Buffalo Meat Production and Quality: Sustainability, Productivity, Nutritional and Sensory Properties', *Journal of Buffalo Science*, pp. 159–169.

Guerrero-Castillo, S. *et al.* (2017) 'Identification and evolutionary analysis of tissue-specific isoforms of mitochondrial complex I subunit NDUFV3', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1858(3), pp. 208–217. Available at: https://doi.org/10.1016/j.bbabio.2016.12.004.

Guha, S. (2019) 'The Maratha Empire', in *Oxford Research Encyclopedia of Asian History*. Available at: https://doi.org/10.1093/acrefore/9780190277727.013.356.

Gummow, B. (2010) 'Challenges posed by new and re-emerging infectious diseases in livestock production, wildlife and humans', *Livestock Science*, 130(1), pp. 41–46. Available at: https://doi.org/10.1016/j.livsci.2010.02.009.

Guo, M. and Hendricks, G. (2010) '14 - Improving buffalo milk', in M.W. Griffiths (ed.) *Improving the Safety and Quality of Milk*. Woodhead Publishing (Woodhead Publishing Series in Food Science, Technology and Nutrition), pp. 402–416. Available at: https://doi.org/10.1533/9781845699437.3.402.

Guo, T. *et al.* (2021) 'Selective Sweeps Uncovering the Genetic Basis of Horn and Adaptability Traits on Fine-Wool Sheep in China', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.604235 (Accessed: 11 April 2023).

Guo, X. *et al.* (2020) 'Mitochondrial stress is relayed to the cytosol by an OMA1–DELE1–HRI pathway', *Nature*, 579(7799), pp. 427–432. Available at: https://doi.org/10.1038/s41586-020-2078-2.

Guo, Z. *et al.* (2021) 'Impacts of Heat Stress-Induced Oxidative Stress on the Milk Protein Biosynthesis of Dairy Cows', *Animals*, 11(3), p. 726. Available at: https://doi.org/10.3390/ani11030726.

Gupta, G. *et al.* (2022) 'Genome-wide association study for suicide in high–risk isolated historical population from North East India', *Journal of Affective Disorders Reports*, 8, p. 100327. Available at: https://doi.org/10.1016/j.jadr.2022.100327.

Gurgul, A. *et al.* (2014) 'The application of genome-wide SNP genotyping methods in studies on livestock genomes', *Journal of Applied Genetics*, 55(2), pp. 197–208. Available at: https://doi.org/10.1007/s13353-014-0202-4.

Hamilton, R. *et al.* (2024) 'Forest mosaics, not savanna corridors, dominated in Southeast Asia during the Last Glacial Maximum', *Proceedings of the National Academy of Sciences*, 121(1), p. e2311280120. Available at: https://doi.org/10.1073/pnas.2311280120.

Han, Q. *et al.* (2020) 'FBXW4 Is Highly Expressed and Associated With Poor Survival in Acute Myeloid Leukemia', *Frontiers in Oncology*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fonc.2020.00149 (Accessed: 21 March 2023).

Han, S. *et al.* (2020) 'Bta-miR-223 Targeting CBLB Contributes to Resistance to Staphylococcus aureus Mastitis Through the PI3K/AKT/NF-κB Pathway', *Frontiers in Veterinary Science*, 7. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2020.00529 (Accessed: 13 September 2023).

Hanna, S. and Etzioni, A. (2014) 'MHC class I and II deficiencies', *Journal of Allergy and Clinical Immunology*, 134(2), pp. 269–275. Available at: https://doi.org/10.1016/j.jaci.2014.06.001.

Hanotte, O. and Jianlin, H. (2005) 'GENETIC CHARACTERIZATION OF LIVESTOCK POPULATIONS AND ITS USE IN CONSERVATION DECISION-MAKING'.

Hansen, L.B. (1999) *Consequences of Selection for Milk Yield from a Geneticist's Viewpoint | Elsevier Enhanced Reader*. Available at: https://doi.org/10.3168/jds.S0022-0302(00)74980-0.

Hansen, P.J. (2014) 'Current and Future Assisted Reproductive Technologies for Mammalian Farm Animals', in G.C. Lamb and N. DiLorenzo (eds) *Current and Future Reproductive Technologies and World Food Production*. New York, NY: Springer (Advances in Experimental Medicine and Biology), pp. 1–22. Available at: https://doi.org/10.1007/978-1-4614-8887-3_1.

Hansen, P.J. (2020) 'Implications of Assisted Reproductive Technologies for Pregnancy Outcomes in Mammals', *Annual Review of Animal Biosciences*, 8(1), pp. 395–413. Available at: https://doi.org/10.1146/annurev-animal-021419-084010.

Hare, M.P. *et al.* (2011) 'Understanding and Estimating Effective Population Size for Practical Application in Marine Species Management', *Conservation Biology*, 25(3), pp. 438–449. Available at: https://doi.org/10.1111/j.1523-1739.2010.01637.x.

Harris, K. and Nielsen, R. (2013) 'Inferring Demographic History from a Spectrum of Shared Haplotype Lengths', *PLOS Genetics*, 9(6), p. e1003521. Available at: https://doi.org/10.1371/journal.pgen.1003521.

Harz, C. *et al.* (2014) 'Secretion and Immunogenicity of the Meningioma-Associated Antigen TXNDC16', *The Journal of Immunology*, 193(6), pp. 3146–3154. Available at: https://doi.org/10.4049/jimmunol.1303098.

Haskell, M.J., Simm, G. and Turner, S.P. (2014) 'Genetic selection for temperament traits in dairy and beef cattle', *Frontiers in Genetics*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2014.00368 (Accessed: 30 October 2023).

Hassanin, A. *et al.* (2012) 'Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes', *Comptes Rendus Biologies*, 335(1), pp. 32–50. Available at: https://doi.org/10.1016/j.crvi.2011.11.002.

Hassanin, A. and Ropiquet, A. (2004) 'Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, Bos sauveli Urbain 1937', *Molecular Phylogenetics and Evolution*, 33(3), pp. 896–907. Available at: https://doi.org/10.1016/j.ympev.2004.08.009.

Hayes, B.J. *et al.* (2008) 'The origin of selection signatures on bovine chromosome 6', *Animal Genetics*, 39(2), pp. 105–111. Available at: https://doi.org/10.1111/j.1365-2052.2007.01683.x.

Hayes, B.J. *et al.* (2009) 'Invited review: Genomic selection in dairy cattle: Progress and challenges', *Journal of Dairy Science*, 92(2), pp. 433–443. Available at: https://doi.org/10.3168/jds.2008-1646.

Hayes, B.J., Lewin, H.A. and Goddard, M.E. (2013) 'The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation', *Trends in Genetics*, 29(4), pp. 206–214. Available at: https://doi.org/10.1016/j.tig.2012.11.009.

Hegde, N. (2019) 'Livestock Development for Sustainable Livelihood of Small Farmers', pp. 1–17.

Hendrickson, S.L. (2013) 'A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo', *BMC Evolutionary Biology*, 13(1), p. 273. Available at: https://doi.org/10.1186/1471-2148-13-273.

Heras-Molina, A. *et al.* (2022) 'Hypothalamic transcriptome analysis reveals male-specific differences in molecular pathways related to oxidative phosphorylation between Iberian pig genotypes', *PLOS ONE*, 17(8), p. e0272775. Available at: https://doi.org/10.1371/journal.pone.0272775.

Hermisson, J. and Pennings, P.S. (2017) 'Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation', *Methods in Ecology and Evolution*, 8(6), pp. 700–716. Available at: https://doi.org/10.1111/2041-210X.12808.

Herrera-Marcos, L.V. *et al.* (2022) 'Hepatic galectin-3 is associated with lipid droplet area in non-alcoholic steatohepatitis in a new swine model', *Scientific Reports*, 12(1), p. 1024. Available at: https://doi.org/10.1038/s41598-022-04971-z.

Herrero, M. *et al.* (2013) 'The roles of livestock in developing countries', *animal*, 7(s1), pp. 3–18. Available at: https://doi.org/10.1017/S1751731112001954.

Hikima, J. and Morimoto, N. (2023) 'Chapter 15 - Evolutive aspects of inflammasomes', in P. Pelegrin (ed.) *Inflammasome Biology*. Academic Press, pp. 235–245. Available at: https://doi.org/10.1016/B978-0-323-91802-2.00013-X.

Hillestad, B. *et al.* (2017) 'Detection of runs of homozygosity in Norwegian Red: Density, criteria and genotyping quality control', *Acta Agriculturae Scandinavica, Section A — Animal Science*, 67(3–4), pp. 107–116. Available at: https://doi.org/10.1080/09064702.2018.1501088.

Hoffmann, I. (2011) 'Livestock biodiversity and sustainability', *Livestock Science*, 139(1), pp. 69–79. Available at: https://doi.org/10.1016/j.livsci.2011.03.016.

Hoffmann, I. (2013) 'Adaptation to climate change – exploring the potential of locally adapted breeds', *animal*, 7(s2), pp. 346–362. Available at: https://doi.org/10.1017/S1751731113000815.

Hoogesteijn, R. and Hoogesteijn, A. (2008) 'Conflicts between cattle ranching and large predators in Venezuela: could use of water buffalo facilitate felid conservation?', *Oryx*, 42(1), pp. 132–138. Available at: https://doi.org/10.1017/S0030605308001105.

Hou, S. *et al.* (2022) 'Downregulating HK2 inhibits proliferation of endometrial stromal cells through a noncanonical pathway involving phosphorylation of signal transducer and activator of transcription 1 in endometriosis', *Biology of Reproduction*, 107(2), pp. 488–499. Available at: https://doi.org/10.1093/biolre/ioac081.

Hou, Z. *et al.* (2019) 'Comparative proteomic identification buffalo spermatozoa during in vitro capacitation', *Theriogenology*, 126, pp. 303–309. Available at: https://doi.org/10.1016/j.theriogenology.2018.12.025.

Hu, X. *et al.* (2021) 'The JAK/STAT signaling pathway: from bench to clinic', *Signal Transduction and Targeted Therapy*, 6(1), pp. 1–33. Available at: https://doi.org/10.1038/s41392-021-00791-1.

Huang, X. *et al.* (2013) 'Loss of cell adhesion molecule CHL1 improves homeostatic adaptation and survival in hypoxic stress', *Cell Death & Disease*, 4(8), pp. e768–e768. Available at: https://doi.org/10.1038/cddis.2013.284.

Hughes, P.D., Gibbard, P.L. and Ehlers, J. (2013) 'Timing of glaciation during the last glacial cycle: evaluating the concept of a global "Last Glacial Maximum" (LGM)', *Earth-Science Reviews*, 125, pp. 171–198. Available at: https://doi.org/10.1016/j.earscirev.2013.07.003.

Huson, D.H. and Bryant, D. (2006) 'Application of Phylogenetic Networks in Evolutionary Studies', *Molecular Biology and Evolution*, 23(2), pp. 254–267. Available at: https://doi.org/10.1093/molbev/msj030.

Huynh, H. *et al.* (2023) 'DELE1 is protective for mitochondrial cardiomyopathy', *Journal of Molecular and Cellular Cardiology*, 175, pp. 44–48. Available at: https://doi.org/10.1016/j.yjmcc.2022.12.003.

Iamartino, D. *et al.* (2017) 'Design and validation of a 90K SNP genotyping assay for the water buffalo (Bubalus bubalis)', *PLOS ONE*, 12(10), p. e0185220. Available at: https://doi.org/10.1371/journal.pone.0185220.

Igoshin, A. *et al.* (2021) 'Whole-Genome Resequencing Points to Candidate DNA Loci Affecting Body Temperature under Cold Stress in Siberian Cattle Populations', *Life*, 11(9), p. 959. Available at: https://doi.org/10.3390/life11090959.

Igoshin, A.V. *et al.* (2019) 'Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in Siberian cattle populations', *BMC Genetics*, 20(1), p. 26. Available at: https://doi.org/10.1186/s12863-019-0725-0.

Iung, L.H.S. *et al.* (2019) 'Genome-wide association study for milk production traits in a Brazilian Holstein population', *Journal of Dairy Science*, 102(6), pp. 5305–5314. Available at: https://doi.org/10.3168/jds.2018-14811.

Jahuey-Martínez, F.J. *et al.* (2019) 'Signatures of selection in Charolais beef cattle identified by genome-wide analysis', *Journal of Animal Breeding and Genetics*, 136(5), pp. 378–389. Available at: https://doi.org/10.1111/jbg.12399.

Jakhar, V., Vinayak, A.K. and Singh, K.P. (no date) 'GENETIC EVALUATION OF PERFORMANCE ATTRIBUTES IN MURRAH BUFFALOES'.

James, G. *et al.* (2013) 'The Expression Pattern of EVA1C, a Novel Slit Receptor, Is Consistent with an Axon Guidance Role in the Mouse Nervous System', *PLOS ONE*, 8(9), p. e74115. Available at: https://doi.org/10.1371/journal.pone.0074115.

Jia, C. *et al.* (2020) 'Identification of genetic loci associated with growth traits at weaning in yak through a genome-wide association study', *Animal Genetics*, 51(2), pp. 300–305. Available at: https://doi.org/10.1111/age.12897.

Jiang, H. *et al.* (2022) 'Genome-wide identification of SNPs associated with body weight in yak', *BMC Genomics*, 23(1), p. 833. Available at: https://doi.org/10.1186/s12864-022-09077-4.

Jiang, Y. *et al.* (2022) 'Genome-wide detection of genetic structure and runs of homozygosity analysis in Anhui indigenous and Western commercial pig breeds using PorcineSNP80k data', *BMC Genomics*, 23(1), p. 373. Available at: https://doi.org/10.1186/s12864-022-08583-9.

Jin, W. *et al.* (2014) 'Distribution of ancestral chromosomal segments in admixed genomes and its implications for inferring population history and admixture mapping', *European Journal of Human Genetics*, 22(7), pp. 930–937. Available at: https://doi.org/10.1038/ejhg.2013.265.

Jonas, E. and Koning, D.-J. de (2015) 'Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs', *Frontiers in Genetics*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00049 (Accessed: 21 March 2023).

Journo, D., Mor, A. and Abeliovich, H. (2009) 'Aup1-mediated Regulation of Rtg3 during Mitophagy *', *Journal of Biological Chemistry*, 284(51), pp. 35885–35895. Available at: https://doi.org/10.1074/jbc.M109.048140.

Ju, Z. *et al.* (2018) 'Solexa sequencing and custom microRNA chip reveal repertoire of microRNAs in mammary gland of bovine suffering from natural infectious mastitis', *Animal Genetics*, 49(1), pp. 3–18. Available at: https://doi.org/10.1111/age.12628.

Júnior, G.A.F. *et al.* (2016) 'Genome scan for postmortem carcass traits in Nellore cattle1', *Journal of Animal Science*, 94(10), pp. 4087–4095. Available at: https://doi.org/10.2527/jas.2016-0632.

Kadri, N.K. *et al.* (2014) 'A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock', *PLOS Genetics*, 10(1), p. e1004049. Available at: https://doi.org/10.1371/journal.pgen.1004049.

Kadzere, C.T. *et al.* (2002) 'Heat stress in lactating dairy cows: a review', *Livestock Production Science*, 77(1), pp. 59–91. Available at: https://doi.org/10.1016/S0301-6226(01)00330-X.

Kamaruddin, N.C. *et al.* (2021) 'The First Report of Ruminant Fascioliasis in Sabah, East Malaysia', *Journal of Parasitology Research*, 2021, p. e6691483. Available at: https://doi.org/10.1155/2021/6691483.

Kaminski, R. *et al.* (2016) 'Elimination of HIV-1 Genomes from Human T-lymphoid Cells by CRISPR/Cas9 Gene Editing', *Scientific Reports*, 6(1), p. 22555. Available at: https://doi.org/10.1038/srep22555.

Kanaka, K.K. *et al.* (2023) 'On the concepts and measures of diversity in the genomics era', *Current Plant Biology*, 33, p. 100278. Available at: https://doi.org/10.1016/j.cpb.2023.100278.

Kandel, R.C. *et al.* (2019) 'Revisiting genetic structure of Wild Buffaloes Bubalus arnee Kerr, 1792 (Mammalia: Artiodactyla: Bovidae) in Koshi Tappu Wildlife Reserve, Nepal: an assessment for translocation programs', *Journal of Threatened Taxa*, 11(15), pp. 14942–14954. Available at: https://doi.org/10.11609/jott.4940.11.15.14942-14954.

Kang, Y. *et al.* (2022) 'A 7-nt nucleotide sequence variant within the sheep KDM3B gene affects female reproduction traits', *Animal Biotechnology*, 33(7), pp. 1661–1667. Available at: https://doi.org/10.1080/10495398.2021.1929270.

Kanzleiter, T. *et al.* (2014) 'Pknox1/Prep1 Regulates Mitochondrial Oxidative Phosphorylation Components in Skeletal Muscle', *Molecular and Cellular Biology*, 34(2), pp. 290–298. Available at: https://doi.org/10.1128/MCB.01232-13.

Katzman, A. and Alberini, C.M. (2018) 'NLGN1 and NLGN2 in the prefrontal cortex: their role in memory consolidation and strengthening', *Current Opinion in Neurobiology*, 48, pp. 122–130. Available at: https://doi.org/10.1016/j.conb.2017.12.003.

Kaur, S. *et al.* (2016) 'The genetic and regulatory architecture of ERBB3-type 1 diabetes susceptibility locus', *Molecular and Cellular Endocrinology*, 419, pp. 83–91. Available at: https://doi.org/10.1016/j.mce.2015.10.002.

Kawamoto, T.S. *et al.* (2022) 'Dynamics of the Reproductive Changes and Acquisition of Oocyte Competence in Nelore (Bos taurus indicus) Calves during the Early and Intermediate Prepubertal Periods', *Animals*, 12(16), p. 2137. Available at: https://doi.org/10.3390/ani12162137.

Kawecki, T.J. and Ebert, D. (2004) 'Conceptual issues in local adaptation', *Ecology Letters*, 7(12), pp. 1225–1241. Available at: https://doi.org/10.1111/j.1461-0248.2004.00684.x.

Keller, M.C. *et al.* (2012) 'Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor', *PLOS Genetics*, 8(4), p. e1002656. Available at: https://doi.org/10.1371/journal.pgen.1002656.

Kelley, J.L. *et al.* (2006) 'Genomic signatures of positive selection in humans and the limits of outlier approaches', *Genome Research*, 16(8), pp. 980–989. Available at: https://doi.org/10.1101/gr.5157306.

Kemper, K.E. *et al.* (2016) 'Leveraging genetically simple traits to identify small-effect variants for complex phenotypes', *BMC Genomics*, 17(1), p. 858. Available at: https://doi.org/10.1186/s12864-016-3175-3.

Keogh, K. *et al.* (2015) 'Insulin secretion and signaling in response to dietary restriction and subsequent re-alimentation in cattle', *Physiological Genomics*, 47(8), pp. 344–354. Available at: https://doi.org/10.1152/physiolgenomics.00002.2015.

Khadjeh, S. *et al.* (2020) 'CRISPLD1: a novel conserved target in the transition to human heart failure', *Basic Research in Cardiology*, 115(3), p. 27. Available at: https://doi.org/10.1007/s00395-020-0784-4.

Khan, A. *et al.* (2021) 'SOD1 Gene Silencing Promotes Apoptosis and Suppresses Proliferation of Heat-Stressed Bovine Granulosa Cells via Induction of Oxidative Stress', *Veterinary Sciences*, 8(12), p. 326. Available at: https://doi.org/10.3390/vetsci8120326.

Kierstein, G. *et al.* (2004) 'Analysis of mitochondrial D-loop region casts new light on domestic water buffalo (Bubalus bubalis) phylogeny', *Molecular Phylogenetics and Evolution*, 30(2), pp. 308–324. Available at: https://doi.org/10.1016/S1055-7903(03)00221-5.

Kijas, J.W. *et al.* (2009) 'A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds', *PLOS ONE*, 4(3), p. e4668. Available at: https://doi.org/10.1371/journal.pone.0004668.

Kilaru, V. *et al.* (2016) 'Genome-wide gene-based analysis suggests an association between Neuroligin 1 (NLGN1) and post-traumatic stress disorder', *Translational Psychiatry*, 6(5), pp. e820–e820. Available at: https://doi.org/10.1038/tp.2016.69.

Kim, E.-S., Sonstegard, T.S. and Rothschild, M.F. (2015) 'Recent artificial selection in U.S. Jersey cattle impacts autozygosity levels of specific genomic regions', *BMC Genomics*, 16(1), p. 302. Available at: https://doi.org/10.1186/s12864-015-1500-x.

Kim, J.Y. *et al.* (2018) 'The 70-kDa heat shock protein (Hsp70) as a therapeutic target for stroke', *Expert Opinion on Therapeutic Targets*, 22(3), pp. 191–199. Available at: https://doi.org/10.1080/14728222.2018.1439477.

Kim, S. *et al.* (2021) 'Genome-Wide Identification of Candidate Genes for Milk Production Traits in Korean Holstein Cattle', *Animals*, 11(5), p. 1392. Available at: https://doi.org/10.3390/ani11051392.

Kirin, M. *et al.* (2010) 'Genomic Runs of Homozygosity Record Population History and Consanguinity', *PLOS ONE*, 5(11), p. e13996. Available at: https://doi.org/10.1371/journal.pone.0013996.

Kleppel, G.S. and Frank, D.A. (2022) 'Structure and functioning of wild and agricultural grazing ecosystems: A comparative review', *Frontiers in Sustainable Food Systems*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fsufs.2022.945514 (Accessed: 30 December 2023).

Koch, B.J., Hungate, B.A. and Price, L.B. (2017) 'Food-animal production and the spread of antibiotic resistance: the role of ecology', *Frontiers in Ecology and the Environment*, 15(6), pp. 309–318. Available at: https://doi.org/10.1002/fee.1505.

Koenigswald, W. v. *et al.* (2019) 'First evidence of Pleistocene Bubalus murrensis in France and the stratigraphic occurrences of Bubalus in Europe', *Quaternary International*, 522, pp. 85–93. Available at: https://doi.org/10.1016/j.quaint.2019.06.019.

Kolipaka, S.S. *et al.* (2017) 'Wild versus domestic prey in the diet of reintroduced tigers (Panthera tigris) in the livestock-dominated multiple-use forests of Panna Tiger Reserve, India', *PLOS ONE*, 12(4), p. e0174844. Available at: https://doi.org/10.1371/journal.pone.0174844.

Komatsu, H. *et al.* (2020) 'Ethnicity-Dependent Effects of Schizophrenia Risk Variants of the OLIG2 Gene on OLIG2 Transcription and White Matter Integrity', *Schizophrenia Bulletin*, 46(6), pp. 1619–1628. Available at: https://doi.org/10.1093/schbul/sbaa049.

Konovalova, S. *et al.* (2018) 'Redox regulation of GRPEL2 nucleotide exchange factor for mitochondrial HSP70 chaperone', *Redox Biology*, 19, pp. 37–45. Available at: https://doi.org/10.1016/j.redox.2018.07.024.

Korunes, K.L. and Goldberg, A. (2021) 'Human genetic admixture', *PLOS Genetics*, 17(3), p. e1009374. Available at: https://doi.org/10.1371/journal.pgen.1009374.

Kotelevets, L. and Chastre, E. (2021) 'A New Story of the Three Magi: Scaffolding Proteins and lncRNA Suppressors of Cancer', *Cancers*, 13(17), p. 4264. Available at: https://doi.org/10.3390/cancers13174264.

Kotera, J. *et al.* (2010) 'Chapter 196 - Phosphodiesterase-9A: A cGMP-Specific Enzyme', in R.A. Bradshaw and E.A. Dennis (eds) *Handbook of Cell Signaling (Second Edition)*. San Diego: Academic Press, pp. 1589–1593. Available at: https://doi.org/10.1016/B978-0-12-374145-5.00196-0.

Kozlowski, E. *et al.* (2017) 'The RNA uridyltransferase Zcchc6 is expressed in macrophages and impacts innate immune responses', *PLOS ONE*, 12(6), p. e0179797. Available at: https://doi.org/10.1371/journal.pone.0179797.

Kristensen, T.N. *et al.* (2015) 'What can livestock breeders learn from conservation genetics and vice versa?', *Frontiers in Genetics*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00038 (Accessed: 12 April 2023).

Krivoruchko, A. *et al.* (2023) 'Genome-Wide Search for Associations with Meat Production Parameters in Karachaevsky Sheep Breed Using the Illumina BeadChip 600 K', *Genes*, 14(6), p. 1288. Available at: https://doi.org/10.3390/genes14061288.

Kulig, H. *et al.* (2010) 'ANXA9, SLC27A3, FABP3 and FABP4 single nucleotide polymorphisms in relation to milk production traits in Jersey cows', *Czech Journal of Animal Science*, 55, pp. 463–467. Available at: https://doi.org/10.17221/1712-CJAS.

Kulig, H. *et al.* (2013) 'FABP3, FABP4 and ANXA9 SNP genotypes in relation to breeding values for milk production traits in Polish Holstein-Friesian cows', *Russian Journal of Genetics*, 49(8), pp. 852–856. Available at: https://doi.org/10.1134/S1022795413080085.

Kumar, H. *et al.* (2023) 'Detection of genome-wide copy number variation in Murrah buffaloes', *Animal Biotechnology*, 0(0), pp. 1–13. Available at: https://doi.org/10.1080/10495398.2023.2227670.

Kumar, M. *et al.* (2017) 'Influence of Environmental Factors on Performance Traits in Murrah Buffaloes: A Review', *Research & Reviews: Journal of Veterinary Science and Technology*, 6, pp. 6–16.

Kumar, M. *et al.* (2019) 'Status, constraints and future prospects of Murrah buffaloes in India', *The Indian journal of animal sciences*, 12, pp. 1291–1302. Available at: https://doi.org/10.56093/ijans.v89i12.96616.

Kumar, R. and Singh, R. (2010) 'Buffalo production system in India', 21, pp. 32–37.

Kumar, S. *et al.* (2006) 'Genetic variation and relationships among eight Indian riverine buffalo breeds', *Molecular Ecology*, 15(3), pp. 593–600. Available at: https://doi.org/10.1111/j.1365-294X.2006.02837.x.

Kumar, S. *et al.* (2007) 'Mitochondrial DNA analyses of Indian water buffalo support a distinct genetic origin of river and swamp buffalo', *Animal Genetics*, 38(3), pp. 227–232. Available at: https://doi.org/10.1111/j.1365-2052.2007.01602.x.

Kumar, Satish *et al.* (2007) 'Phylogeography and domestication of Indian river buffalo', *BMC Evolutionary Biology*, 7(1), p. 186. Available at: https://doi.org/10.1186/1471-2148-7-186.

Kumar, V. (2015) 'Factors Affecting Performance of Indian Murrah Buffalo: A Review', *Journal of Buffalo Science*, 4, pp. 21–27. Available at: https://doi.org/10.6000/1927-520X.2015.04.01.4.

Kusama, K. *et al.* (2021) 'Day 7 Embryos Change the Proteomics and Exosomal Micro-RNAs Content of Bovine Uterine Fluid: Involvement of Innate Immune Functions', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.676791 (Accessed: 28 December 2023).

Kyriazis, C.C., Wayne, R.K. and Lohmueller, K.E. (2021) 'Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression', *Evolution Letters*, 5(1), pp. 33–47. Available at: https://doi.org/10.1002/evl3.209.

La Piana, R. *et al.* (2017) 'Identification and functional characterization of a novel MTFMT mutation associated with selective vulnerability of the visual pathway and a mild neurological phenotype', *neurogenetics*, 18(2), pp. 97–103. Available at: https://doi.org/10.1007/s10048-016-0506-0.

Lachance, J. and Tishkoff, S.A. (2013) 'SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it', *BioEssays*, 35(9), pp. 780–786. Available at: https://doi.org/10.1002/bies.201300014.

LaFramboise, T. (2009) 'Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances', *Nucleic Acids Research*, 37(13), pp. 4181–4193. Available at: https://doi.org/10.1093/nar/gkp552.

Lan, Q. *et al.* (2022) 'Physiological and Proteomic Responses of Dairy Buffalo to Heat Stress Induced by Different Altitudes', *Metabolites*, 12(10), p. 909. Available at: https://doi.org/10.3390/metabo12100909.

Lande, R. (1988) 'Genetics and Demography in Biological Conservation', *Science*, 241(4872), pp. 1455–1460. Available at: https://doi.org/10.1126/science.3420403.

Lange, S. *et al.* (2021) 'MicroRNAs for Virus Pathogenicity and Host Responses, Identified in SARS-CoV-2 Genomes, May Play Roles in Viral-Host Co-Evolution in Putative Zoonotic Host Species', *Viruses*, 13(1), p. 117. Available at: https://doi.org/10.3390/v13010117.

Larson, G. and Fuller, D.Q. (2014) 'The Evolution of Animal Domestication', *Annual Review of Ecology, Evolution, and Systematics*, 45(1), pp. 115–136. Available at: https://doi.org/10.1146/annurev-ecolsys-110512-135813.

Lau, C.H. *et al.* (1998) 'Genetic diversity of Asian water buffalo (Bubalus bubalis): mitochondrial DNA D-loop and cytochrome b sequence variation', *Animal Genetics*, 29(4), pp. 253–264. Available at: https://doi.org/10.1046/j.1365-2052.1998.00309.x.

Lau, E. and Ronai, Z.A. (2012) 'ATF2 – at the crossroad of nuclear and cytosolic functions', *Journal of Cell Science*, 125(12), pp. 2815–2824. Available at: https://doi.org/10.1242/jcs.095000.

Lázaro, S.F. *et al.* (2021) 'Genomic studies of milk-related traits in water buffalo (Bubalus bubalis) based on single-step genomic best linear unbiased prediction and random regression models', *Journal of Dairy Science*, 104(5), pp. 5768–5793. Available at: https://doi.org/10.3168/jds.2020-19534.

Le Clerc, S. *et al.* (2013) 'A Genome-Wide Association Study in Caucasian Women Points Out a Putative Role of the STXBP5L Gene in Facial Photoaging', *Journal of Investigative Dermatology*, 133(4), pp. 929–935. Available at: https://doi.org/10.1038/jid.2012.458.

Le Pen, J. *et al.* (2018) 'Terminal uridylyltransferases target RNA viruses as part of the innate immune system', *Nature Structural & Molecular Biology*, 25(9), pp. 778–786. Available at: https://doi.org/10.1038/s41594-018-0106-9.

Lee, L. (2013) 'Riding the wave of ependymal cilia: Genetic susceptibility to hydrocephalus in primary ciliary dyskinesia', *Journal of Neuroscience Research*, 91(9), pp. 1117–1132. Available at: https://doi.org/10.1002/jnr.23238.

Lei, C.Z. *et al.* (2007) 'Independent maternal origin of Chinese swamp buffalo (Bubalus bubalis)', *Animal Genetics*, 38(2), pp. 97–102. Available at: https://doi.org/10.1111/j.1365-2052.2007.01567.x.

Lencz, T. *et al.* (2007) 'Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia', *Proceedings of the National Academy of Sciences of the United States of America*, 104(50), pp. 19942–19947. Available at: https://doi.org/10.1073/pnas.0710021104.

Leng, G. and Hall, J.W. (2021) 'Where is the Planetary Boundary for freshwater being exceeded because of livestock farming?', *Science of The Total Environment*, 760, p. 144035. Available at: https://doi.org/10.1016/j.scitotenv.2020.144035.

Leroy, G. (2014) 'Inbreeding depression in livestock species: review and meta-analysis', *Animal Genetics*, 45(5), pp. 618–628. Available at: https://doi.org/10.1111/age.12178.

Leroy, G. *et al.* (2018) 'Next-generation metrics for monitoring genetic erosion within populations of conservation concern', *Evolutionary Applications*, 11(7), pp. 1066–1083. Available at: https://doi.org/10.1111/eva.12564.

Leshchyns'ka, I. *et al.* (2006) 'The Adhesion Molecule CHL1 Regulates Uncoating of Clathrin-Coated Synaptic Vesicles', *Neuron*, 52(6), pp. 1011–1025. Available at: https://doi.org/10.1016/j.neuron.2006.10.020.

Leuenberger, C. and Wegmann, D. (2010) 'Bayesian Computation and Model Selection Without Likelihoods', *Genetics*, 184(1), pp. 243–252. Available at: https://doi.org/10.1534/genetics.109.109058.

Li, C. *et al.* (2016) 'RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population', *Scientific Reports*, 6(1), p. 26813. Available at: https://doi.org/10.1038/srep26813.

Li, H. and Durbin, R. (2011) 'Inference of human population history from individual whole-genome sequences', *Nature*, 475(7357), pp. 493–496. Available at: https://doi.org/10.1038/nature10231.

Li, K. *et al.* (2018) 'Upregulated IQUB promotes cell proliferation and migration via activating Akt/GSK3β/β-catenin signaling pathway in breast cancer', *Cancer Medicine*, 7(8), pp. 3875–3888. Available at: https://doi.org/10.1002/cam4.1568.

Li, M. *et al.* (2022) 'Variant analyses of candidate genes in orofacial clefts in multi-ethnic populations', *Oral Diseases*, 28(7), pp. 1921–1935. Available at: https://doi.org/10.1111/odi.13932.

Li, R. *et al.* (2023) 'Genome-Wide Signal Selection Analysis Revealing Genes Potentially Related to Sheep-Milk-Production Traits', *Animals*, 13(10), p. 1654. Available at: https://doi.org/10.3390/ani13101654.

Liang, C. *et al.* (2016) 'Genome-wide Association Study Identifies Loci for the Polled Phenotype in Yak', *PLOS ONE*, 11(7), p. e0158642. Available at: https://doi.org/10.1371/journal.pone.0158642.

Lin, X. *et al.* (2019) 'Stop-gain mutations in UBAP1 cause pure autosomal-dominant spastic paraplegia', *Brain*, 142(8), pp. 2238–2252. Available at: https://doi.org/10.1093/brain/awz158.

Lindmark-Månsson, H. *et al.* (2006) 'Relationship between somatic cell count, individual leukocyte populations and milk components in bovine udder quarter milk', *International Dairy Journal*, 16(7), pp. 717–727. Available at: https://doi.org/10.1016/j.idairyj.2005.07.003.

Liu, J.J. *et al.* (2018) 'Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo', *Journal of Dairy Science*, 101(1), pp. 433–444. Available at: https://doi.org/10.3168/jds.2017-13246.

Liu, S. *et al.* (2017) 'Genetic effects of PDGFRB and MARCH1 identified in GWAS revealing strong associations with semen production traits in Chinese Holstein bulls', *BMC Genetics*, 18(1), p. 63. Available at: https://doi.org/10.1186/s12863-017-0527-1.

Liu, S. *et al.* (2022) 'Genome-wide analysis of runs of homozygosity in Italian Mediterranean buffalo', *Journal of Dairy Science*, 105(5), pp. 4324–4334. Available at: https://doi.org/10.3168/jds.2021-21543.

Liu, Y. *et al.* (2018) 'Agouti Signaling Protein and Its Receptors as Potential Molecular Markers for Intramuscular and Body Fat Deposition in Cattle', *Frontiers in Physiology*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fphys.2018.00172 (Accessed: 18 September 2023).

Liu, Y. *et al.* (2022) 'Genomic inbreeding and runs of homozygosity analysis of indigenous cattle populations in southern China', *PLOS ONE*, 17(8), p. e0271718. Available at: https://doi.org/10.1371/journal.pone.0271718.

Liudkovska, V. *et al.* (2022) 'TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals', *Science Advances*, 8(46), p. eadd9468. Available at: https://doi.org/10.1126/sciadv.add9468.

Lkhagvadorj, S. *et al.* (2010) 'Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 298(2), pp. R494–R507. Available at: https://doi.org/10.1152/ajpregu.00632.2009.

Lobo-Alves, S.C., de Oliveira, L.A. and Petzl-Erler, M.L. (2019) 'Region 1p13.2 including the RSBN1, PTPN22, AP4B1 and long non-coding RNA genes does not bear risk factors for endemic pemphigus foliaceus (fogo selvagem)', *International Journal of Immunogenetics*, 46(3), pp. 139–145. Available at: https://doi.org/10.1111/iji.12423.

Locatelli, C. *et al.* (2013) 'Effect on quarter milk somatic cell count and antimicrobial susceptibility of Staphylococcus rostri causing intramammary infection in dairy water buffaloes', *Journal of Dairy Science*, 96(6), pp. 3799–3805. Available at: https://doi.org/10.3168/jds.2012-6275.

Loh, P.-R. *et al.* (2013) 'Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium', *Genetics*, 193(4), pp. 1233–1254. Available at: https://doi.org/10.1534/genetics.112.147330.

Lopes, M.S. *et al.* (2017) 'Genomic selection for crossbred performance accounting for breed-specific effects', *Genetics Selection Evolution*, 49(1), p. 51. Available at: https://doi.org/10.1186/s12711-017-0328-z.

Lopreiato, V. *et al.* (2020) 'Methyl donor supply to heat stress-challenged polymorphonuclear leukocytes from lactating Holstein cows enhances 1-carbon metabolism, immune response, and cytoprotective gene network abundance', *Journal of Dairy Science*, 103(11), pp. 10477–10493. Available at: https://doi.org/10.3168/jds.2020-18638.

Low, W.Y. *et al.* (2019) 'Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity', *Nature Communications*, 10(1), p. 260. Available at: https://doi.org/10.1038/s41467-018-08260-0.

Lu, X.R. *et al.* (2020) 'Genome-wide analysis reveals genetic diversity, linkage disequilibrium, and selection for milk production traits in Chinese buffalo breeds', *Journal of Dairy Science*, 103(5), pp. 4545–4556. Available at: https://doi.org/10.3168/jds.2019-17364.

Luo, R. *et al.* (2022) 'Genome-Wide DNA Methylation Patterns of Muscle and Tail-Fat in DairyMeade Sheep and Mongolian Sheep', *Animals*, 12(11), p. 1399. Available at: https://doi.org/10.3390/ani12111399.

Luo, X. *et al.* (2020) 'Understanding divergent domestication traits from the whole-genome sequencing of swamp- and river-buffalo populations', *National Science Review*, 7(3), pp. 686–701. Available at: https://doi.org/10.1093/nsr/nwaa024.

Luskin, M.S. *et al.* (2021) 'African Swine Fever threatens Southeast Asia's 11 endemic wild pig species', *Conservation Letters*, 14(3), p. e12784. Available at: https://doi.org/10.1111/conl.12784.

Luu, K., Bazin, E. and Blum, M.G.B. (2017) 'pcadapt: an R package to perform genome scans for selection based on principal component analysis', *Molecular Ecology Resources*, 17(1), pp. 67–77. Available at: https://doi.org/10.1111/1755-0998.12592.

Macciotta, N.P.P. *et al.* (2021) 'The distribution of runs of homozygosity in the genome of river and swamp buffaloes reveals a history of adaptation, migration and crossbred events', *Genetics Selection Evolution*, 53(1), p. 20. Available at: https://doi.org/10.1186/s12711-021-00616-3.

MacEachern, S., McEwan, J. and Goddard, M. (2009a) 'Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae)', *BMC Genomics*, 10(1), p. 177. Available at: https://doi.org/10.1186/1471-2164-10-177.

MacEachern, S., McEwan, J. and Goddard, M. (2009b) 'Phylogenetic reconstruction and the identification of ancient polymorphism in the Bovini tribe (Bovidae, Bovinae)', *BMC Genomics*, 10(1), p. 177. Available at: https://doi.org/10.1186/1471-2164-10-177.

Mach, N. *et al.* (2012) 'Pleiotropic effects of polymorphism of the gene diacylglycerol-O-transferase 1 (DGAT1) in the mammary gland tissue of dairy cows', *Journal of Dairy Science*, 95(9), pp. 4989–5000. Available at: https://doi.org/10.3168/jds.2012-5348.

Maheswarappa, N.B. *et al.* (2022) 'Water Buffalo: Origin, Emergence, and Domestication', in N.B. Maheswarappa et al. (eds) *Asiatic Water Buffalo: A Sustainable and Healthy Red Meat Source*. Singapore: Springer Nature, pp. 1–10. Available at: https://doi.org/10.1007/978-981-19-2619-8_1.

Makanjuola, B.O. *et al.* (2020) 'Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations', *Journal of Dairy Science*, 103(6), pp. 5183–5199. Available at: https://doi.org/10.3168/jds.2019-18013.

Malhado, C.H.M. *et al.* (2013) 'Inbreeding depression on production and reproduction traits of buffaloes from Brazil', *Animal Science Journal*, 84(4), pp. 289–295. Available at: https://doi.org/10.1111/asj.12006.

Malik, A. *et al.* (2021) 'Protection of ζ-crystallin by α-crystallin under thermal stress', *International Journal of Biological Macromolecules*, 167, pp. 289–298. Available at: https://doi.org/10.1016/j.ijbiomac.2020.11.183.

Malomane, D.K. *et al.* (2018) 'Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies', *BMC Genomics*, 19(1), p. 22. Available at: https://doi.org/10.1186/s12864-017-4416-9.

Mancin, E. *et al.* (2022) 'Genome Wide Association Study of Beef Traits in Local Alpine Breed Reveals the Diversity of the Pathways Involved and the Role of Time Stratification', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.746665 (Accessed: 11 September 2023).

Manoj, M. *et al.* (2017) 'Effect of functional traits on subsequent reproduction performance of Murrah buffaloes in India', *Journal of Applied Animal Research*, 45(1), pp. 22–28. Available at: https://doi.org/10.1080/09712119.2015.1102727.

Manolio, T.A. *et al.* (2019) 'Opportunities, resources, and techniques for implementing genomics in clinical care', *The Lancet*, 394(10197), pp. 511–520. Available at: https://doi.org/10.1016/S0140-6736(19)31140-7.

Mapiye, O. *et al.* (2020) 'Livelihood, Food and Nutrition Security in Southern Africa: What Role Do Indigenous Cattle Genetic Resources Play?', *Diversity*, 12(2), p. 74. Available at: https://doi.org/10.3390/d12020074.

Marai, I.F.M. and Haeeb, A.A.M. (2010) 'Buffalo's biological functions as affected by heat stress — A review', *Livestock Science*, 127(2), pp. 89–109. Available at: https://doi.org/10.1016/j.livsci.2009.08.001.

Marete, A. *et al.* (2018) 'A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds', *PLOS ONE*, 13(7), p. e0199931. Available at: https://doi.org/10.1371/journal.pone.0199931.

Marie-Hardy, L. *et al.* (2021) 'The orthopedic characterization of cfap298tm304 mutants validate zebrafish to faithfully model human AIS', *Scientific Reports*, 11(1), p. 7392. Available at: https://doi.org/10.1038/s41598-021-86856-1.

Marjoram, P. and Tavaré, S. (2006) 'Modern computational approaches for analysing molecular genetic variation data', *Nature Reviews Genetics*, 7(10), pp. 759–770. Available at: https://doi.org/10.1038/nrg1961.

Marom, N. and Bar-Oz, G. (2013) 'The Prey Pathway: A Regional History of Cattle (Bos taurus) and Pig (Sus scrofa) Domestication in the Northern Jordan Valley, Israel', *PLOS ONE*, 8(2), p. e55958. Available at: https://doi.org/10.1371/journal.pone.0055958.

Marshall, K. (2014) 'Optimizing the use of breed types in developing country livestock production systems: a neglected research area', *Journal of Animal Breeding and Genetics*, 131(5), pp. 329–340. Available at: https://doi.org/10.1111/jbg.12080.

Martinelli, D.C. and Fan, C.-M. (2007) 'The Role of Gas1 in Embryonic Development and its Implications for Human Disease', *Cell Cycle*, 6(21), pp. 2650–2655. Available at: https://doi.org/10.4161/cc.6.21.4877.

Martínez, A.M. *et al.* (2012) 'Genetic Footprints of Iberian Cattle in America 500 Years after the Arrival of Columbus', *PLOS ONE*, 7(11), p. e49066. Available at: https://doi.org/10.1371/journal.pone.0049066.

Martínez-Navarro, B. *et al.* (2011) 'Presence of the Asian origin Bovini, Hemibos sp. aff. Hemibos gracilis and Bison sp., at the early Pleistocene site of Venta Micena (Orce, Spain)', *Quaternary International*, 243(1), pp. 54–60. Available at: https://doi.org/10.1016/j.quaint.2011.05.016.

Martínez-Royo, A. *et al.* (2010) 'The bovine annexin 9 gene (ANXA9) is significantly associated with milk-fat yield in a Spanish Holstein–Friesian population', *Research in Veterinary Science*, 88(3), pp. 452–455. Available at: https://doi.org/10.1016/j.rvsc.2009.12.009.

Martins, R. *et al.* (2021) 'Genome-wide association study and pathway analysis for carcass fatness in Nellore cattle measured by ultrasound', *Animal Genetics*, 52(5), pp. 730–733. Available at: https://doi.org/10.1111/age.13129.

Martínez-Navarro, B. and Palombo, M.R. (2004) 'Occurrence of the Indian genus Hemibos (Bovini, Bovidae, Mammalia) at the Early–Middle Pleistocene transition in Italy', *Quaternary Research*, 61(3), pp. 314–317. Available at: https://doi.org/10.1016/j.yqres.2004.02.002.

Matera, R. *et al.* (2022) 'Relationship among production traits, somatic cell score and temperature–humidity index in the Italian Mediterranean Buffalo', *Italian Journal of Animal Science*, 21(1), pp. 551–561. Available at: https://doi.org/10.1080/1828051X.2022.2042407.

Matiuti, M. *et al.* (2020) *Particularities of Buffalo breeding in Romania*. Available at: https://doi.org/10.13140/RG.2.2.23304.24323.

Matukumalli, L.K. *et al.* (2009) 'Development and Characterization of a High Density SNP Genotyping Assay for Cattle', *PLOS ONE*, 4(4), p. e5350. Available at: https://doi.org/10.1371/journal.pone.0005350.

McCarthy, S.D. *et al.* (2010) 'Negative energy balance and hepatic gene expression patterns in high-yielding dairy cows during the early postpartum period: a global approach', *Physiological Genomics*, 42A(3), pp. 188–199. Available at: https://doi.org/10.1152/physiolgenomics.00118.2010.

McDaneld, T.G., Hancock, D.L. and Moody, D.E. (2004) 'Altered mRNA abundance of ASB15 and four other genes in skeletal muscle following administration of β-adrenergic receptor agonists', *Physiological Genomics*, 16(2), pp. 275–283. Available at: https://doi.org/10.1152/physiolgenomics.00127.2003.

McDaneld, T.G. and Spurlock, D.M. (2008) 'Ankyrin repeat and suppressor of cytokine signaling (SOCS) box-containing protein (ASB) 15 alters differentiation of mouse C2C12 myoblasts and phosphorylation of mitogen-activated protein kinase and Akt1', *Journal of Animal Science*, 86(11), pp. 2897–2902. Available at: https://doi.org/10.2527/jas.2008-1076.

McFadden, J.W. and Rico, J.E. (2019) 'Invited review: Sphingolipid biology in the dairy cow: The emerging role of ceramide', *Journal of Dairy Science*, 102(9), pp. 7619–7639. Available at: https://doi.org/10.3168/jds.2018-16095.

McTavish, E.J. and Hillis, D.M. (2015) 'How do SNP ascertainment schemes and population demographics affect inferences about population history?', *BMC Genomics*, 16(1), p. 266. Available at: https://doi.org/10.1186/s12864-015-1469-5.

Megdiche, S. *et al.* (2019) 'A Combined Multi-Cohort Approach Reveals Novel and Known Genome-Wide Selection Signatures for Wool Traits in Merino and Merino-Derived Sheep Breeds', *Frontiers in Genetics*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01025 (Accessed: 28 December 2023).

Mehta, T.A. *et al.* (2005) 'Heat Shock Proteins in Vascular Disease—A Review', *European Journal of Vascular and Endovascular Surgery*, 29(4), pp. 395–402. Available at: https://doi.org/10.1016/j.ejvs.2005.01.005.

Meijer, D.H. *et al.* (2012) 'Separated at birth? The functional and molecular divergence of OLIG1 and OLIG2', *Nature Reviews Neuroscience*, 13(12), pp. 819–831. Available at: https://doi.org/10.1038/nrn3386.

Mendoza, A. *et al.* (2020) 'Similarities and Differences between River Buffaloes and Cattle: Health, Physiological, Behavioral and Productivity Aspects', *Journal of Buffalo Science*, 9, pp. 92–109. Available at: https://doi.org/10.6000/1927-520X.2020.09.12.

Meuwissen, T. (2009) 'Genetic management of small populations: A review', *Acta Agriculturae Scandinavica, Section A — Animal Science*, 59(2), pp. 71–79. Available at: https://doi.org/10.1080/09064700903118148.

Meuwissen, T., Hayes, B. and Goddard, M. (2013) 'Accelerating Improvement of Livestock with Genomic Selection', *Annual Review of Animal Biosciences*, 1(1), pp. 221–237. Available at: https://doi.org/10.1146/annurev-animal-031412-103705.

Meuwissen, T., Hayes, B. and Goddard, M. (2016) 'Genomic selection: A paradigm shift in animal breeding', *Animal Frontiers*, 6(1), pp. 6–14. Available at: https://doi.org/10.2527/af.2016-0002.

Meuwissen, T.H.E. *et al.* (2020) 'Management of Genetic Diversity in the Era of Genomics', *Frontiers in Genetics*, 11. Available at: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2020.00880 (Accessed: 29 February 2024).

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) 'Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps', *Genetics*, 157(4), pp. 1819–1829. Available at: https://doi.org/10.1093/genetics/157.4.1819.

Meyermans, R. *et al.* (2020) 'How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species', *BMC Genomics*, 21(1), p. 94. Available at: https://doi.org/10.1186/s12864-020-6463-x.

Miao, X., Luo, Q. and Qin, X. (2015) 'Genome-wide analysis reveals the differential regulations of mRNAs and miRNAs in Dorset and Small Tail Han sheep muscles', *Gene*, 562(2), pp. 188–196. Available at: https://doi.org/10.1016/j.gene.2015.02.070.

Mignon-Grasteau, S. *et al.* (2005) 'Genetics of adaptation and domestication in livestock', *Livestock Production Science*, 93(1), pp. 3–14. Available at: https://doi.org/10.1016/j.livprodsci.2004.11.001.

Miller, J.R.B., Jhala, Y.V. and Jena, J. (2016) 'Livestock losses and hotspots of attack from tigers and leopards in Kanha Tiger Reserve, Central India', *Regional Environmental Change*, 16(1), pp. 17–29. Available at: https://doi.org/10.1007/s10113-015-0871-5.

Minervino, A.H.H. *et al.* (2020) 'Bubalus bubalis: A Short Story', *Frontiers in Veterinary Science*, 7. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2020.570413 (Accessed: 12 April 2023).

Mintoo, A.A. *et al.* (2019) 'Draft genome of the river water buffalo', *Ecology and Evolution*, 9(6), pp. 3378–3388. Available at: https://doi.org/10.1002/ece3.4965.

Mirkena, T. *et al.* (2010) 'Genetics of adaptation in domestic farm animals: A review', *Livestock Science*, 132(1), pp. 1–12. Available at: https://doi.org/10.1016/j.livsci.2010.05.003.

Mishra, B. p. *et al.* (2015) 'Genetic analysis of river, swamp and hybrid buffaloes of north-east India throw new light on phylogeography of water buffalo (Bubalus bubalis)', *Journal of Animal Breeding and Genetics*, 132(6), pp. 454–466. Available at: https://doi.org/10.1111/jbg.12141.

Mishra, S.K. *et al.* (2020) 'Diversity analysis at MHC class II *DQA* locus in buffalo (*Bubalus bubalis*) indicates extensive duplication and trans-species evolution', *Genomics*, 112(6), pp. 4417–4426. Available at: https://doi.org/10.1016/j.ygeno.2020.07.041.

Miyai, S., Hendawy, A.O. and Sato, K. (2021) 'Gene expression profile of peripheral blood mononuclear cells in mild to moderate obesity in dogs', *Veterinary and Animal Science*, 13, p. 100183. Available at: https://doi.org/10.1016/j.vas.2021.100183.

Modiba, M.C. *et al.* (2022) 'Runs of Homozygosity and Quantitative Trait Locus/Association for Semen Parameters in Selected Chinese and South African Beef Cattle', *Animals*, 12(12), p. 1546. Available at: https://doi.org/10.3390/ani12121546.

Mohammadparast, S. *et al.* (2014) 'Exceptional expansion and conservation of a CT-repeat complex in the core promoter of PAXBP1 in primates', *American Journal of Primatology*, 76(8), pp. 747–756. Available at: https://doi.org/10.1002/ajp.22266.

Moisá, S.J. *et al.* (2017) 'Transcriptional changes in mesenteric and subcutaneous adipose tissue from Holstein cows in response to plane of dietary energy', *Journal of Animal Science and Biotechnology*, 8(1), p. 85. Available at: https://doi.org/10.1186/s40104-017-0215-z.

Mokhber, M. (2017) *GENOME-WIDE ASSOCIATION STUDY FOR MILK PRODUCTION IN IRANIAN BUFFALO*.

Mokhber, M. *et al.* (2018) 'A genome-wide scan for signatures of selection in Azeri and Khuzestani buffalo breeds', *BMC Genomics*, 19(1), p. 449. Available at: https://doi.org/10.1186/s12864-018-4759-x.

Mokhber, M. *et al.* (2019) 'Study of whole genome linkage disequilibrium patterns of Iranian water buffalo breeds using the Axiom Buffalo Genotyping 90K Array', *PLOS ONE*, 14(5), p. e0217687. Available at: https://doi.org/10.1371/journal.pone.0217687.

Moorey, S.E. and Biase, F.H. (2020) 'Beef heifer fertility: importance of management practices and technological advancements', *Journal of Animal Science and Biotechnology*, 11(1), p. 97. Available at: https://doi.org/10.1186/s40104-020-00503-9.

Mota, R. r. *et al.* (2017) 'Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle', *Journal of Animal Breeding and Genetics*, 134(6), pp. 484–492. Available at: https://doi.org/10.1111/jbg.12299.

Mota-Rojas, D. *et al.* (2021) 'Thermal biology in river buffalo in the humid tropics: neurophysiological and behavioral responses assessed by infrared thermography', *Journal of Animal Behaviour and Biometeorology*, 9(1), pp. 1–12. Available at: https://doi.org/10.31893/jabb.21003.

Mozduri, Z., Bakhtiarizadeh, M.R. and Salehi, A. (2018) 'Integrated regulatory network reveals novel candidate regulators in the development of negative energy balance in cattle', *animal*, 12(6), pp. 1196–1207. Available at: https://doi.org/10.1017/S1751731117003524.

Mrode, R. *et al.* (2019) 'Genomic Selection and Use of Molecular Tools in Breeding Programs for Indigenous and Crossbred Cattle in Developing Countries: Current Status and Future Prospects', *Frontiers in Genetics*, 9. Available at: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2018.00694 (Accessed: 13 February 2024).

Mukherjee, A. *et al.* (2023) 'Candidate genes and proteins regulating bull semen quality: a review', *Tropical Animal Health and Production*, 55(3), p. 212. Available at: https://doi.org/10.1007/s11250-023-03617-0.

Muñoz, M. *et al.* (2019) 'Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip', *Scientific Reports*, 9(1), p. 13546. Available at: https://doi.org/10.1038/s41598-019-49830-6.

Murtaza, M.A., Pandya, A.J. and Khan, M.M.H. (2017) 'Buffalo Milk', in *Handbook of Milk of Non-Bovine Mammals*. John Wiley & Sons, Ltd, pp. 261–283. Available at: https://doi.org/10.1002/9781119110316.ch4.

Naderi, S. *et al.* (2018) 'Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets', *Animal Genetics*, 49(3), pp. 178–192. Available at: https://doi.org/10.1111/age.12661.

Nagarajan, M., Nimisha, K. and Kumar, S. (2015) 'Mitochondrial DNA Variability of Domestic River Buffalo (Bubalus bubalis) Populations: Genetic Evidence for Domestication of River Buffalo in Indian Subcontinent', *Genome Biology and Evolution*, 7(5), pp. 1252–1259. Available at: https://doi.org/10.1093/gbe/evv067.

Nagashimada, M. *et al.* (2018) 'TAF7 is a heat-inducible unstable protein and is required for sustained expression of heat shock protein genes', *The FEBS Journal*, 285(17), pp. 3215–3224. Available at: https://doi.org/10.1111/febs.14604.

Nakanishi, M. *et al.* (2017) 'Functional significance of rare neuroligin 1 variants found in autism', *PLOS Genetics*, 13(8), p. e1006940. Available at: https://doi.org/10.1371/journal.pgen.1006940.

Nanda, A.S. and Nakao, T. (2003) 'Role of buffalo in the socioeconomic development of rural Asia: Current status and future prospectus', *Animal Science Journal*, 74(6), pp. 443–455. Available at: https://doi.org/10.1046/j.1344-3941.2003.00138.x.

Nascimento, A.V. *et al.* (2021) 'Inbreeding coefficients and runs of homozygosity islands in Brazilian water buffalo', *Journal of Dairy Science*, 104(2), pp. 1917–1927. Available at: https://doi.org/10.3168/jds.2020-18397.

Nascimento, A.V. do *et al.* (2018) 'Genome-wide association study using haplotype alleles for the evaluation of reproductive traits in Nelore cattle', *PLOS ONE*, 13(8), p. e0201876. Available at: https://doi.org/10.1371/journal.pone.0201876.

Naskar, S. *et al.* (2012) 'Genetic Adaptability of Livestock to Environmental Stresses', in V. Sejian et al. (eds) *Environmental Stress and Amelioration in Livestock Production*. Berlin, Heidelberg: Springer, pp. 317–378. Available at: https://doi.org/10.1007/978-3-642-29205-7_13.

Nava-Trujillo, H. *et al.* (no date) 'Reproductive Performance of Water Buffalo Cows: A Review of Affecting Factors'.

Negesse, T., Datt, C. and Kundu, S.S. (2016) 'Variability in residual feed intake and nutrient utilization in Murrah buffalo heifers', *Tropical Animal Health and Production*, 48(8), pp. 1577–1584. Available at: https://doi.org/10.1007/s11250-016-1130-6.

Neidert, N. *et al.* (2018) 'Microglia-Specific Expression of Olfml3 Is Directly Regulated by Transforming Growth Factor β1-Induced Smad2 Signaling', *Frontiers in Immunology*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fimmu.2018.01728 (Accessed: 8 December 2022).

Nelson, E.A. *et al.* (2006) 'Identification of Human STAT5-dependent Gene Regulatory Elements Based on Interspecies Homology *', *Journal of Biological Chemistry*, 281(36), pp. 26216–26224. Available at: https://doi.org/10.1074/jbc.M605001200.

Nemec, A.A. *et al.* (2016) 'Estrogen Drives Cellular Transformation and Mutagenesis in Cells Expressing the Breast Cancer–Associated R438W DNA Polymerase Lambda Protein', *Molecular Cancer Research*, 14(11), pp. 1068–1077. Available at: https://doi.org/10.1158/1541-7786.MCR-16-0209.

Nguyen, D.Q. *et al.* (2018) 'The role of ErbB3 binding protein 1 in cancer: Friend or foe?', *Journal of Cellular Physiology*, 233(12), pp. 9110–9120. Available at: https://doi.org/10.1002/jcp.26951.

Nguyen, L.T. *et al.* (2017) 'Global differential gene expression in the pituitary gland and the ovaries of pre- and postpubertal Brahman heifers1', *Journal of Animal Science*, 95(2), pp. 599–615. Available at: https://doi.org/10.2527/jas.2016.0921.

Nicoloso, L. *et al.* (2015) 'Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip', *Genetics Selection Evolution*, 47(1), p. 62. Available at: https://doi.org/10.1186/s12711-015-0140-6.

Nielsen, R. (2004) 'Population genetic analysis of ascertained SNP data', *Human Genomics*, 1(3), p. 218. Available at: https://doi.org/10.1186/1479-7364-1-3-218.

van Nieuwenhoven, F.A. *et al.* (2017) 'Cartilage intermediate layer protein 1 (CILP1): A novel mediator of cardiac extracellular matrix remodelling', *Scientific Reports*, 7(1), p. 16042. Available at: https://doi.org/10.1038/s41598-017-16201-y.

Niyas, P.A. *et al.* (2015) 'Adaptation of Livestock to Environmental Challenges', *Journal of Veterinary Science and Medical Diagnosis*, 4. Available at: https://doi.org/10.4172/2325-9590.1000162.

Noce, A. *et al.* (2021) 'Genetic Diversity of Bubalus bubalis in Germany and Global Relations of Its Genetic Background', *Frontiers in Genetics*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2020.610353 (Accessed: 17 October 2022).

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

Novo, L.C. *et al.* (2022) 'Genomic Analysis of Visceral Fat Accumulation in Holstein Cows', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.803216 (Accessed: 25 August 2023).

Oget, C., Servin, B. and Palhière, I. (2019) 'Genetic diversity analysis of French goat populations reveals selective sweeps involved in their differentiation', *Animal Genetics*, 50(1), pp. 54–63. Available at: https://doi.org/10.1111/age.12752.

Ohashi, M. *et al.* (2004) 'Human T-cell leukemia virus type 1 Tax oncoprotein induces and interacts with a multi-PDZ domain protein, MAGI-3', *Virology*, 320(1), pp. 52–62. Available at: https://doi.org/10.1016/j.virol.2003.11.014.

Ollier, S. *et al.* (2007) 'Mammary Transcriptome Analysis of Food-Deprived Lactating Goats Highlights Genes Involved in Milk Secretion and Programmed Cell Death', *The Journal of Nutrition*, 137(3), pp. 560–567. Available at: https://doi.org/10.1093/jn/137.3.560.

Ollivier, L. and Foulley, J.-L. (2009) 'Managing Genetic Diversity, Fitness and Adaptation of Farm Animal Genetic Resources', in J. van der Werf et al. (eds) *Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management*. Dordrecht: Springer Netherlands, pp. 201–227. Available at: https://doi.org/10.1007/978-1-4020-9005-9_14.

Olschewsky, A. and Hinrichs, D. (2021) 'An Overview of the Use of Genotyping Techniques for Assessing Genetic Diversity in Local Farm Animal Breeds', *Animals*, 11(7), p. 2016. Available at: https://doi.org/10.3390/ani11072016.

Oltenacu, P.A. and Broom, D.M. (2010) 'The impact of genetic selection for increased milk yield on the welfare of dairy cows', *Animal Welfare*, 19(S1), pp. 39–49. Available at: https://doi.org/10.1017/S0962728600002220.

O'Mara, F.P. (2012) 'The role of grasslands in food security and climate change', *Annals of Botany*, 110(6), pp. 1263–1270. Available at: https://doi.org/10.1093/aob/mcs209.

Ornatowska, M. *et al.* (2007) 'Functional genomics of silencing TREM-1 on TLR4 signaling in macrophages', *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 293(6), pp. L1377–L1384. Available at: https://doi.org/10.1152/ajplung.00140.2007.

Orozco-Lucero, E., Dufort, I. and Sirard, M.-A. (2017) 'Regulation of ATF1 and ATF2 transcripts by sequences in their 3' untranslated region in cleavage-stage cattle embryos', *Molecular Reproduction and Development*, 84(4), pp. 296–309. Available at: https://doi.org/10.1002/mrd.22785.

Orozco-terWengel, P. *et al.* (2015) 'Revisiting demographic processes in cattle with genome-wide population genetic analysis', *Frontiers in Genetics*, 6. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2015.00191 (Accessed: 19 March 2023).

Orozco-terWengel, P. (2016) 'The devil is in the details: the effect of population structure on demographic inference', *Heredity*, 116(4), pp. 349–350. Available at: https://doi.org/10.1038/hdy.2016.9.

Ottenburghs, J. *et al.* (2020) 'Recent introgression between Taiga Bean Goose and Tundra Bean Goose results in a largely homogeneous landscape of genetic differentiation', *Heredity*, 125(1), pp. 73–84. Available at: https://doi.org/10.1038/s41437-020-0322-z.

Ouyang, H. *et al.* (2020) 'A novel transcript of MEF2D promotes myoblast differentiation and its variations associated with growth traits in chicken', *PeerJ*, 8, p. e8351. Available at: https://doi.org/10.7717/peerj.8351.

Ovsyannikova, I.G. *et al.* (2016) 'Gene signatures associated with adaptive humoral immunity following seasonal influenza A/H1N1 vaccination', *Genes & Immunity*, 17(7), pp. 371–379. Available at: https://doi.org/10.1038/gene.2016.34.

Owen, K.L., Brockwell, N.K. and Parker, B.S. (2019) 'JAK-STAT Signaling: A Double-Edged Sword of Immune Regulation and Cancer Progression', *Cancers*, 11(12), p. 2002. Available at: https://doi.org/10.3390/cancers11122002.

Oyhenart, J. *et al.* (2003) 'Phtf1 Is an Integral Membrane Protein Localized in an Endoplasmic Reticulum Domain in Maturing Male Germ Cells1', *Biology of Reproduction*, 68(3), pp. 1044–1053. Available at: https://doi.org/10.1095/biolreprod.102.009787.

Ozden Cobanoglu *et al.* (2016) 'The Detection of STAT1 Gene Influencing Milk Related Traits in Turkish Holstein and Jersey Cows', *Journal of Agricultural Science and Technology A*, 6(4). Available at: https://doi.org/10.17265/2161-6256/2016.04.006.

Palombo, V. *et al.* (2018) 'Genome-wide association study of milk fatty acid composition in Italian Simmental and Italian Holstein cows using single nucleotide polymorphism arrays', *Journal of Dairy Science*, 101(12), pp. 11004–11019. Available at: https://doi.org/10.3168/jds.2018-14413.

Pan, C.-J. *et al.* (2011) 'SLC37A1 and SLC37A2 Are Phosphate-Linked, Glucose-6-Phosphate Antiporters', *PLOS ONE*, 6(9), p. e23157. Available at: https://doi.org/10.1371/journal.pone.0023157.

Park, J. *et al.* (2021) 'Growth factors improve the proliferation of Jeju black pig muscle cells by regulating myogenic differentiation 1 and growth-related genes', *Animal Bioscience*, 34(8), pp. 1392–1402. Available at: https://doi.org/10.5713/ab.20.0585.

Park, Y. and Han, S.J. (2022) 'Interferon Signaling in the Endometrium and in Endometriosis', *Biomolecules*, 12(11), p. 1554. Available at: https://doi.org/10.3390/biom12111554.

Passamonti, M.M. *et al.* (2021) 'The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock', *Animals*, 11(10), p. 2833. Available at: https://doi.org/10.3390/ani11102833.

Paten, A.M. *et al.* (2015) 'Functional development of the adult ovine mammary gland—insights from gene expression profiling', *BMC Genomics*, 16(1), p. 748. Available at: https://doi.org/10.1186/s12864-015-1947-9.

Pausch, H. *et al.* (2012) 'Identification of QTL for UV-Protective Eye Area Pigmentation in Cattle by Progeny Phenotyping and Genome-Wide Association Analysis', *PLOS ONE*, 7(5), p. e36346. Available at: https://doi.org/10.1371/journal.pone.0036346.

Pawar (2012) 'Effect of Year, Season and Parity on Milk Production Traits in Murrah Buffaloes', *Journal of Buffalo Science* [Preprint]. Available at: https://doi.org/10.6000/1927-520X.2012.01.01.22.

Pecka-Kiełb, E. *et al.* (2021) 'FASN, SCD1 and ANXA9 gene polymorphism as genetic predictors of the fatty acid profile of sheep milk', *Scientific Reports*, 11(1), p. 23761. Available at: https://doi.org/10.1038/s41598-021-03186-y.

Pedernera, M. *et al.* (2010) 'Effect of diet, energy balance and milk production on oxidative stress in early-lactating dairy cows grazing pasture', *The Veterinary Journal*, 186(3), pp. 352–357. Available at: https://doi.org/10.1016/j.tvjl.2009.09.003.

Pegolo, S. *et al.* (2016) 'Effects of candidate gene polymorphisms on the detailed fatty acids profile determined by gas chromatography in bovine milk', *Journal of Dairy Science*, 99(6), pp. 4558–4573. Available at: https://doi.org/10.3168/jds.2015-10420.

Pekkala, N. *et al.* (2012) 'Inbreeding rate modifies the dynamics of genetic load in small populations', *Ecology and Evolution*, 2(8), pp. 1791–1804. Available at: https://doi.org/10.1002/ece3.293.

Pekkala, N. *et al.* (2014) 'The effect of inbreeding rate on fitness, inbreeding depression and heterosis over a range of inbreeding coefficients', *Evolutionary Applications*, 7(9), pp. 1107–1119. Available at: https://doi.org/10.1111/eva.12145.

Pemberton, T.J. *et al.* (2012) 'Genomic Patterns of Homozygosity in Worldwide Human Populations', *American Journal of Human Genetics*, 91(2), pp. 275–292. Available at: https://doi.org/10.1016/j.ajhg.2012.06.014.

Peng, Y. *et al.* (2017) 'Illumina-sequencing based transcriptome study of coat color phenotypes in domestic goats', *Genes & Genomics*, 39(8), pp. 817–830. Available at: https://doi.org/10.1007/s13258-017-0543-6.

Pérez-Enciso, M., Rincón, J.C. and Legarra, A. (2015) 'Sequence- vs. chip-assisted genomic selection: accurate biological information is advised', *Genetics Selection Evolution*, 47(1), p. 43. Available at: https://doi.org/10.1186/s12711-015-0117-5.

Peripolli, E. *et al.* (2017) 'Runs of homozygosity: current knowledge and applications in livestock', *Animal Genetics*, 48(3), pp. 255–271. Available at: https://doi.org/10.1111/age.12526.

Peripolli, E. *et al.* (2018) 'Autozygosity islands and ROH patterns in Nellore lineages: evidence of selection for functionally important traits', *BMC Genomics*, 19(1), p. 680. Available at: https://doi.org/10.1186/s12864-018-5060-8.

Petrocchi Jasinski, F. *et al.* (2023) 'Responses of Dairy Buffalo to Heat Stress Conditions and Mitigation Strategies: A Review', *Animals*, 13(7), p. 1260. Available at: https://doi.org/10.3390/ani13071260.

Piantoni, P. *et al.* (2010) 'Expression of Metabolic, Tissue Remodeling, Oxidative Stress, and Inflammatory Pathways in Mammary Tissue during Involution in Lactating Dairy Cows', *Bioinformatics and Biology Insights*, 4, p. BBI.S5850. Available at: https://doi.org/10.4137/BBI.S5850.

Pickrell, J.K. and Pritchard, J.K. (2012) 'Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data', *PLOS Genetics*, 8(11), p. e1002967. Available at: https://doi.org/10.1371/journal.pgen.1002967.

Piry, S., Luikart, G. and Cornuet, J.-M. (1999) 'Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data', *Journal of Heredity*, 90(4), pp. 502–503. Available at: https://doi.org/10.1093/jhered/90.4.502.

Pisanu, S. *et al.* (2019) 'Proteomic changes in the milk of water buffaloes (Bubalus bubalis) with subclinical mastitis due to intramammary infection by Staphylococcus aureus and by non-aureus staphylococci', *Scientific Reports*, 9(1), p. 15850. Available at: https://doi.org/10.1038/s41598-019-52063-2.

Pitt, D., Bruford, M.W., *et al.* (2019) 'Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics', *Evolutionary Applications*, 12(1), pp. 105–122. Available at: https://doi.org/10.1111/eva.12641.

Pitt, D., Sevane, N., *et al.* (2019) 'Domestication of cattle: Two or three events?', *Evolutionary Applications*, 12(1), pp. 123–136. Available at: https://doi.org/10.1111/eva.12674.

Piwczyński, D., Brzozowski, M. and Sitkowska, B. (2020) 'The impact of the installation of an automatic milking system on female fertility traits in Holstein-Friesian cows', *Livestock Science*, 240, p. 104140. Available at: https://doi.org/10.1016/j.livsci.2020.104140.

Plastow, G.S. (2016) 'Genomics to benefit livestock production: improving animal health', *Revista Brasileira de Zootecnia*, 45, pp. 349–354. Available at: https://doi.org/10.1590/S1806-92902016000600010.

Popa, R. *et al.* (no date) 'ECONOMIC WEIGHT OF PRODUCTION TRAITS FOR ROMANIAN BUFFALO'.

Posner, M. *et al.* (2012) 'Functional Validation of Hydrophobic Adaptation to Physiological Temperature in the Small Heat Shock Protein αA-crystallin', *PLOS ONE*, 7(3), p. e34438. Available at: https://doi.org/10.1371/journal.pone.0034438.

Pound, M.J. *et al.* (2012) 'Global vegetation dynamics and latitudinal temperature gradients during the Mid to Late Miocene (15.97–5.33Ma)', *Earth-Science Reviews*, 112(1), pp. 1–22. Available at: https://doi.org/10.1016/j.earscirev.2012.02.005.

Pritchard, J.K., Pickrell, J.K. and Coop, G. (2010) 'The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation', *Current Biology*, 20(4), pp. R208–R215. Available at: https://doi.org/10.1016/j.cub.2009.11.055.

Priyono, D.S. *et al.* (2020) 'The first complete mitochondrial genome sequence of the endangered mountain anoa (Bubalus quarlesi) (Artiodactyla: Bovidae) and phylogenetic analysis', *Journal of Asia-Pacific Biodiversity*, 13(2), pp. 123–133. Available at: https://doi.org/10.1016/j.japb.2020.01.006.

Proudfoot, C. *et al.* (2020) 'Livestock breeding for the 21st century: the promise of the editing revolution', *Frontiers of Agricultural Science and Engineering*, 7(2), p. 129. Available at: https://doi.org/10.15302/J-FASE-2019304.

Puig-Oliveras, A. *et al.* (2016) 'Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat', *Scientific Reports*, 6(1), p. 31803. Available at: https://doi.org/10.1038/srep31803.

Puppel, K., Kapusta, A. and Kuczyńska, B. (2015) 'The etiology of oxidative stress in the various species of animals, a review', *Journal of the Science of Food and Agriculture*, 95(11), pp. 2179–2184. Available at: https://doi.org/10.1002/jsfa.7015.

Purfield, D.C. *et al.* (2012) 'Runs of homozygosity and population history in cattle', *BMC Genetics*, 13(1), p. 70. Available at: https://doi.org/10.1186/1471-2156-13-70.

Qanbari, S. (2020) 'On the Extent of Linkage Disequilibrium in the Genome of Farm Animals', *Frontiers in Genetics*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01304 (Accessed: 25 April 2023).

Qanbari, S. and Simianer, H. (2014) 'Mapping signatures of positive selection in the genome of livestock', *Livestock Science*, 166, pp. 133–143. Available at: https://doi.org/10.1016/j.livsci.2014.05.003.

Qiu, Q. *et al.* (2015) 'Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions', *Nature Communications*, 6(1), p. 10283. Available at: https://doi.org/10.1038/ncomms10283.

Quilez, J. *et al.* (2011) 'A selective sweep of >8 Mb on chromosome 26 in the Boxer genome', *BMC Genomics*, 12(1), p. 339. Available at: https://doi.org/10.1186/1471-2164-12-339.

Quinto-Cortés, C.D. *et al.* (2018) 'Modeling SNP array ascertainment with Approximate Bayesian Computation for demographic inference', *Scientific Reports*, 8(1), p. 10209. Available at: https://doi.org/10.1038/s41598-018-28539-y.

Radwan, H., El Qaliouby, H. and Elfadl, E.A. (2020) 'Classification and prediction of milk yield level for Holstein Friesian cattle using parametric and non-parametric statistical classification models', *Journal of Advanced Veterinary and Animal Research*, 7(3), pp. 429–435. Available at: https://doi.org/10.5455/javar.2020.g438.

Rahimmadar, S. *et al.* (2021) 'Linkage Disequilibrium and Effective Population Size of Buffalo Populations of Iran, Turkey, Pakistan, and Egypt Using a Medium Density SNP Array', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.608186 (Accessed: 18 October 2022).

Ramos, A.M. *et al.* (2009) 'Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology', *PLOS ONE*, 4(8), p. e6524. Available at: https://doi.org/10.1371/journal.pone.0006524.

Ramos, Z. *et al.* (2023) 'Genomic Regions Associated with Wool, Growth and Reproduction Traits in Uruguayan Merino Sheep', *Genes*, 14(1), p. 167. Available at: https://doi.org/10.3390/genes14010167.

Rauw, W. (2012) 'Immune response from a resource allocation perspective', *Frontiers in Genetics*, 3. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2012.00267 (Accessed: 30 October 2023).

Rauw, W.M. (2016) 'Editorial: Improving Animal Welfare through Genetic Selection', *Frontiers in Genetics*, 7. Available at: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2016.00069 (Accessed: 13 February 2024).

Raven, L.-A. *et al.* (2016) 'Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle', *Mammalian Genome*, 27(1), pp. 81–97. Available at: https://doi.org/10.1007/s00335-015-9613-8.

Ravi Kumar, D. *et al.* (2020) 'Genomic diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes', *Genomics*, 112(3), pp. 2385–2392. Available at: https://doi.org/10.1016/j.ygeno.2020.01.010.

Ravi Kumar, D. *et al.* (2023) 'Genome-wide association study revealed suggestive QTLs for production and reproduction traits in Indian Murrah buffalo', *3 Biotech*, 13(3), p. 100. Available at: https://doi.org/10.1007/s13205-023-03505-2.

Rebelato, A.B. and Caetano, A.R. (2018) 'Runs of homozygosity for autozygosity estimation and genomic analysis in production animals', *Pesquisa Agropecuária Brasileira*, 53, pp. 975–984. Available at: https://doi.org/10.1590/S0100-204X2018000900001.

Rege, J.E.O. and Gibson, J.P. (2003) 'Animal genetic resources and economic development: issues in relation to economic valuation', *Ecological Economics*, 45(3), pp. 319–330. Available at: https://doi.org/10.1016/S0921-8009(03)00087-9.

Rexroad, C.E. and Vallejo, R.L. (2009) 'Estimates of linkage disequilibrium and effective population size in rainbow trout', *BMC Genetics*, 10(1), p. 83. Available at: https://doi.org/10.1186/1471-2156-10-83.

Reynolds, A.W. *et al.* (2019) 'Comparing signals of natural selection between three Indigenous North American populations', *Proceedings of the National Academy of Sciences*, 116(19), pp. 9312–9317. Available at: https://doi.org/10.1073/pnas.1819467116.

Riblett, A.M. *et al.* (2016) 'A Haploid Genetic Screen Identifies Heparan Sulfate Proteoglycans Supporting Rift Valley Fever Virus Infection', *Journal of Virology*, 90(3), pp. 1414–1423. Available at: https://doi.org/10.1128/jvi.02055-15.

Rico, J.E. *et al.* (2016) 'Palmitic acid feeding increases ceramide supply in association with increased milk yield, circulating nonesterified fatty acids, and adipose tissue responsiveness to a glucose

challenge', *Journal of Dairy Science*, 99(11), pp. 8817–8830. Available at: https://doi.org/10.3168/jds.2016-11296.

Robin, M. *et al.* (2022) 'Ancient mitochondrial and modern whole genomes unravel massive genetic diversity loss during near extinction of Alpine ibex', *Molecular Ecology*, 31(13), pp. 3548–3565. Available at: https://doi.org/10.1111/mec.16503.

Rojas-Downing, M.M. *et al.* (2017) 'Climate change and livestock: Impacts, adaptation, and mitigation', *Climate Risk Management*, 16, pp. 145–163. Available at: https://doi.org/10.1016/j.crm.2017.02.001.

Romao, J.M. *et al.* (2014) 'MicroRNAs in bovine adipogenesis: genomic context, expression and function', *BMC Genomics*, 15(1), p. 137. Available at: https://doi.org/10.1186/1471-2164-15-137.

de Roos, A.P.W. *et al.* (2008) 'Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle', *Genetics*, 179(3), pp. 1503–1512. Available at: https://doi.org/10.1534/genetics.107.084301.

Ros-Freixedes, R. *et al.* (2022) 'Genomic prediction with whole-genome sequence data in intensely selected pig lines', *Genetics Selection Evolution*, 54(1), p. 65. Available at: https://doi.org/10.1186/s12711-022-00756-0.

Rothschild, M.F. and Plastow, G.S. (2014) 'Applications of genomics to improve livestock in the developing world', *Livestock Science*, 166, pp. 76–83. Available at: https://doi.org/10.1016/j.livsci.2014.03.020.

Rougemont, Q. and Bernatchez, L. (2018) 'The demographic history of Atlantic salmon (Salmo salar) across its distribution range reconstructed from approximate Bayesian computations*', *Evolution*, 72(6), pp. 1261–1277. Available at: https://doi.org/10.1111/evo.13486.

Rovelli, G. *et al.* (2020) 'The genetics of phenotypic plasticity in livestock in the era of climate change: a review', *Italian Journal of Animal Science*, 19(1), pp. 997–1014. Available at: https://doi.org/10.1080/1828051X.2020.1809540.

Rozins, C. and Day, T. (2016) 'Disease eradication on large industrial farms', *Journal of Mathematical Biology*, 73(4), pp. 885–902. Available at: https://doi.org/10.1007/s00285-016-0973-9.

Rozzi, R. (2017) 'A new extinct dwarfed buffalo from Sulawesi and the evolution of the subgenus Anoa: An interdisciplinary perspective', *Quaternary Science Reviews*, 157, pp. 188–205. Available at: https://doi.org/10.1016/j.quascirev.2016.12.011.

Ruiz-Arroyo, V.M. *et al.* (2023) 'Structures and mechanisms of tRNA methylation by METTL1–WDR4', *Nature*, 613(7943), pp. 383–390. Available at: https://doi.org/10.1038/s41586-022-05565-5.

Ruiz-De-La-Cruz, G. *et al.* (2023) 'Characterization of intronic SNP located in candidate genes influencing cattle temperament', *Revista Brasileira de Zootecnia*, 52, p. e20220057. Available at: https://doi.org/10.37496/rbz5220220057.

Rychtářová, J. *et al.* (2014) 'Effect of DGAT1, BTN1A1, OLR1, and STAT1 genes on milk production and reproduction traits in the Czech Fleckvieh breed', *Czech Journal of Animal Science*, 59(No. 2), pp. 45–53. Available at: https://doi.org/10.17221/7228-CJAS.

Sá Filho, M.F. de *et al.* (2017) 'Impact of hormonal modulation at proestrus on ovarian responses and uterine gene expression of suckled anestrous beef cows', *Journal of Animal Science and Biotechnology*, 8(1), p. 79. Available at: https://doi.org/10.1186/s40104-017-0211-3.

Sabeti, P.C. *et al.* (2002) 'Detecting recent positive selection in the human genome from haplotype structure', *Nature*, 419(6909), pp. 832–837. Available at: https://doi.org/10.1038/nature01140.

Sabeti, P.C. *et al.* (2006) 'Positive Natural Selection in the Human Lineage', *Science*, 312(5780), pp. 1614–1620. Available at: https://doi.org/10.1126/science.1124309.

Sabeti, P.C. *et al.* (2007) 'Genome-wide detection and characterization of positive selection in human populations', *Nature*, 449(7164), pp. 913–918. Available at: https://doi.org/10.1038/nature06250.

Salari, F., Altomonte, I. and Martini, M. (2013) 'Buffalo milk: a case study of some parameters related to milk production.', *Large Animal Review*, 19(1), pp. 17–20.

Salzano, A. *et al.* (2020) 'Inflammatory response in repeat breeder buffaloes', *Theriogenology*, 145, pp. 31–38. Available at: https://doi.org/10.1016/j.theriogenology.2020.01.034.

Salzano, A. *et al.* (2021) 'Green feed increases antioxidant and antineoplastic activity of buffalo milk: A globally significant livestock', *Food Chemistry*, 344, p. 128669. Available at: https://doi.org/10.1016/j.foodchem.2020.128669.

Salzano, A. *et al.* (2023) 'Transcriptomic profiles of the ruminal wall in Italian Mediterranean dairy buffaloes fed green forage', *BMC Genomics*, 24(1), p. 133. Available at: https://doi.org/10.1186/s12864-023-09215-6.

Sanchez, M.-P. *et al.* (2021) 'Sequence-based GWAS and post-GWAS analyses reveal a key role of SLC37A1, ANKH, and regulatory regions on bovine milk mineral content', *Scientific Reports*, 11(1), p. 7537. Available at: https://doi.org/10.1038/s41598-021-87078-1.

Sánchez-Cordón, P.J. *et al.* (2018) 'African swine fever: A re-emerging viral disease threatening the global pig industry', *The Veterinary Journal*, 233, pp. 41–48. Available at: https://doi.org/10.1016/j.tvjl.2017.12.025.

Sant'Anna, A.C. and Paranhos da Costa, M.J.R. (2011) 'The relationship between dairy cow hygiene and somatic cell count in milk', *Journal of Dairy Science*, 94(8), pp. 3835–3844. Available at: https://doi.org/10.3168/jds.2010-3951.

Santos, F.C. dos *et al.* (2017) 'Identification of Candidate Genes for Reactivity in Guzerat (Bos indicus) Cattle: A Genome-Wide Association Study', *PLOS ONE*, 12(1), p. e0169163. Available at: https://doi.org/10.1371/journal.pone.0169163.

Saqib, M.N. *et al.* (2022) 'Association among metabolic status, oxidative stress, milk yield, body condition score and reproductive cyclicity in dairy buffaloes', *Reproduction in Domestic Animals*, 57(5), pp. 498–504. Available at: https://doi.org/10.1111/rda.14086.

Saravanan, K.A. *et al.* (2020) 'Selection signatures in livestock genome: A review of concepts, approaches and applications', *Livestock Science*, 241, p. 104257. Available at: https://doi.org/10.1016/j.livsci.2020.104257.

Saravanan, K.A., Panigrahi, M., *et al.* (2022) 'Advanced software programs for the analysis of genetic diversity in livestock genomics: a mini review', *Biological Rhythm Research*, 53(3), pp. 358–368. Available at: https://doi.org/10.1080/09291016.2019.1642650.

Saravanan, K.A., Rajawat, D., *et al.* (2022) 'Signatures of selection in riverine buffalo populations revealed by genome-wide SNP data', *Animal Biotechnology*, 0(0), pp. 1–12. Available at: https://doi.org/10.1080/10495398.2022.2145292.

Saravanan, K.A. *et al.* (2023) 'Signatures of selection in riverine buffalo populations revealed by genome-wide SNP data', *Animal Biotechnology*, 34(8), pp. 3343–3354. Available at: https://doi.org/10.1080/10495398.2022.2145292.

Sartori, R. *et al.* (2016) 'Update and overview on assisted reproductive technologies (ARTs) in Brazil', *Animal Reproduction*, 13(3), pp. 300–312. Available at: https://doi.org/10.21451/1984-3143-AR873.

Savolainen, O., Lascoux, M. and Merilä, J. (2013) 'Ecological genomics of local adaptation', *Nature Reviews Genetics*, 14(11), pp. 807–820. Available at: https://doi.org/10.1038/nrg3522.

Savolainen, P. *et al.* (2002) 'Genetic Evidence for an East Asian Origin of Domestic Dogs', *Science*, 298(5598), pp. 1610–1613. Available at: https://doi.org/10.1126/science.1073906.

Sbardella, A.P. *et al.* (2021) 'Genome-Wide Association Study Provides Insights into Important Genes for Reproductive Traits in Nelore Cattle', *Animals*, 11(5), p. 1386. Available at: https://doi.org/10.3390/ani11051386.

Scheu, A. *et al.* (2015) 'The genetic prehistory of domesticated cattle from their origin to the spread across Europe', *BMC Genetics*, 16(1), p. 54. Available at: https://doi.org/10.1186/s12863-015-0203-2.

Schiffels, S. and Durbin, R. (2014) 'Inferring human population size and separation history from multiple genome sequences', *Nature Genetics*, 46(8), pp. 919–925. Available at: https://doi.org/10.1038/ng.3015.

Schmidtmann, C. *et al.* (2021) 'Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe', *Genetics Selection Evolution*, 53(1), p. 23. Available at: https://doi.org/10.1186/s12711-021-00613-6.

Schneppenheim, J. *et al.* (2012) 'The intramembrane protease SPPL2a promotes B cell development and controls endosomal traffic by cleavage of the invariant chain', *Journal of Experimental Medicine*, 210(1), pp. 41–58. Available at: https://doi.org/10.1084/jem.20121069.

Schreiber, A. and Seibold, I. (no date) 'Cytochrome b Gene Haplotypes Characterize Chromosomal Lineages of Anoa, the Sulawesi Dwarf Buffalo (Bovidae: Bubalus sp.)'.

Seim, I. *et al.* (2013) 'Genome analysis reveals insights into physiology and longevity of the Brandt's bat Myotis brandtii', *Nature Communications*, 4(1), p. 2212. Available at: https://doi.org/10.1038/ncomms3212.

Sella, G. and Barton, N.H. (2019) 'Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies', *Annual Review of Genomics and Human Genetics*, 20(1), pp. 461–493. Available at: https://doi.org/10.1146/annurev-genom-083115-022316.

Sena, L. *et al.* (2003) 'Polymorphisms in MHC-DRA and -DRB alleles of water buffalo (Bubalus bubalis) reveal different features from cattle DR alleles', *Animal Genetics*, 34(1), pp. 1–10. Available at: https://doi.org/10.1046/j.1365-2052.2003.00920.x.

Seo, K. *et al.* (2007) 'Biology of epidermal and hair pigmentation in cattle: a mini-review', *Veterinary Dermatology*, 18(6), pp. 392–400. Available at: https://doi.org/10.1111/j.1365-3164.2007.00634.x.

Seong, H.-S. *et al.* (2022) 'Genetic Parameter Estimation and Genome-Wide Association Analysis of Social Genetic Effects on Average Daily Gain in Purebreds and Crossbreds', *Animals*, 12(17), p. 2300. Available at: https://doi.org/10.3390/ani12172300.

Serber, D.W. *et al.* (2016) 'The Mouse INO80 Chromatin-Remodeling Complex Is an Essential Meiotic Factor for Spermatogenesis1', *Biology of Reproduction*, 94(1), pp. 8, 1–9. Available at: https://doi.org/10.1095/biolreprod.115.135533.

Serranito, B. *et al.* (2021) 'Local adaptations of Mediterranean sheep and goats through an integrative approach', *Scientific Reports*, 11(1), p. 21363. Available at: https://doi.org/10.1038/s41598-021-00682-z.

Setyaningsih, C.A. *et al.* (2019) 'First palaeoecological evidence of buffalo husbandry and rice cultivation in the Kerinci Seblat National Park in Sumatra, Indonesia', *Vegetation History and Archaeobotany*, 28(6), pp. 591–606.

Shaari, N.' A.L. *et al.* (2019) 'Karyotypic and mtDNA based characterization of Malaysian water buffalo', *BMC Genetics*, 20(1), p. 37. Available at: https://doi.org/10.1186/s12863-019-0741-0.

Shahzad, K. *et al.* (2015) 'The effect of calving in the summer on the hepatic transcriptome of Holstein cows during the peripartal period', *Journal of Dairy Science*, 98(8), pp. 5401–5413. Available at: https://doi.org/10.3168/jds.2015-9409.

Shakun, J.D. and Carlson, A.E. (2010) 'A global perspective on Last Glacial Maximum to Holocene climate change', *Quaternary Science Reviews*, 29(15), pp. 1801–1816. Available at: https://doi.org/10.1016/j.quascirev.2010.03.016.

Shang, Y. *et al.* (2016) 'Autophagy regulates spermatid differentiation via degradation of PDLIM1', *Autophagy*, 12(9), pp. 1575–1592. Available at: https://doi.org/10.1080/15548627.2016.1192750.

Shao, G.C. *et al.* (2011) 'A C/T mutation in microRNA target sites in BMP5 gene is potentially associated with fatness in pigs', *Meat Science*, 87(3), pp. 299–303. Available at: https://doi.org/10.1016/j.meatsci.2010.09.013.

Sharma, A. *et al.* (2015) 'Stories and Challenges of Genome Wide Association Studies in Livestock — A Review', *Asian-Australasian Journal of Animal Sciences*, 28(10), pp. 1371–1379. Available at: https://doi.org/10.5713/ajas.14.0715.

Sharma, V.K. *et al.* (2018) 'Buffalo heifers selected for lower residual feed intake have lower feed intake, better dietary nitrogen utilisation and reduced enteric methane production', *Journal of Animal Physiology and Animal Nutrition*, 102(2), pp. e607–e614. Available at: https://doi.org/10.1111/jpn.12802.

Siddiq, M. *et al.* (2019) '17. Hemibos Sardhok BIOLOGIA 2016', *Biologia (Lahore, Pakistan)*, 62, pp. 305–311.

Sigdel, A. *et al.* (2020) 'Genetic dissection of reproductive performance of dairy cows under heat stress', *Animal Genetics*, 51(4), pp. 511–520. Available at: https://doi.org/10.1111/age.12943.

Silva, C.G. da *et al.* (2019) 'Use of trichostatin A alters the expression of HDAC3 and KAT2 and improves in vitro development of bovine embryos cloned using less methylated mesenchymal stem cells', *Reproduction in Domestic Animals*, 54(2), pp. 289–299. Available at: https://doi.org/10.1111/rda.13360.

da Silva, J.A.R. *et al.* (2021) 'Water buffalo production in the Brazilian Amazon Basin: a review', *Tropical Animal Health and Production*, 53(3), p. 343. Available at: https://doi.org/10.1007/s11250-021-02744-w.

Simon, R., Drögemüller, C. and Lühken, G. (2022) 'The Complex and Diverse Genetic Architecture of the Absence of Horns (Polledness) in Domestic Ruminants, including Goats and Sheep', *Genes*, 13(5), p. 832. Available at: https://doi.org/10.3390/genes13050832.

Simons, N.D. *et al.* (2019) 'Genome-Wide Patterns of Gene Expression in a Wild Primate Indicate Species-Specific Mechanisms Associated with Tolerance to Natural Simian Immunodeficiency Virus Infection', *Genome Biology and Evolution*, 11(6), pp. 1630–1643. Available at: https://doi.org/10.1093/gbe/evz099.

Singh, I. and Balhara, A.K. (2016) 'New approaches in buffalo artificial insemination programs with special reference to India', *Theriogenology*, 86(1), pp. 194–199. Available at: https://doi.org/10.1016/j.theriogenology.2016.04.031.

Singh, S.V. (2014) 'Co-incidence of Bovine Johne's disease and Bovine Brucellosis in Young Bulls of Murrah Breed in their Native Tract (Rohtak, Haryana, India)', *Advances in Animal and Veterinary Sciences*, 2(1S), pp. 23–25. Available at: https://doi.org/10.14737/journal.aavs/2014/2.1s.23.25.

Slatkin, M. (2008) 'Linkage disequilibrium — understanding the evolutionary past and mapping the medical future', *Nature Reviews Genetics*, 9(6), pp. 477–485. Available at: https://doi.org/10.1038/nrg2361.

Šnjegota, D. *et al.* (2021) 'Population genetic structure of wolves in the northwestern Dinaric-Balkan region', *Ecology and Evolution*, 11(24), pp. 18492–18504. Available at: https://doi.org/10.1002/ece3.8444.

Sodhi, M. *et al.* (2023) 'Expression profile of different classes of proteases in milk derived somatic cells across different lactation stages of indigenous cows (Bos indicus) and riverine buffaloes (Bubalus bubalis)', *Animal Biotechnology*, 34(1), pp. 15–24. Available at: https://doi.org/10.1080/10495398.2021.1930024.

Soini, K. *et al.* (2012) 'Developing a typology for local cattle breed farmers in Europe', *Journal of Animal Breeding and Genetics*, 129(6), pp. 436–447. Available at: https://doi.org/10.1111/j.1439-0388.2012.01009.x.

Sokolova, O.V. *et al.* (2023) 'Genetic Susceptibility to Ketosis in Cattle: Current State of Research', *Russian Journal of Genetics*, 59(3), pp. 251–262. Available at: https://doi.org/10.1134/S1022795423030110.

Song, N. *et al.* (2022) 'Mutation of Signal Transducer and Activator of Transcription 5 (STAT5) Binding Sites Decreases Milk Allergen αS1-Casein Content in Goat Mammary Epithelial Cells', *Foods*, 11(3), p. 346. Available at: https://doi.org/10.3390/foods11030346.

Song, Y. *et al.* (2022) 'Screening of Genes Related to Growth, Development and Meat Quality of Sahan Crossbred F1 Sheep Based on RNA-Seq Technology', *Frontiers in Veterinary Science*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fvets.2022.831519 (Accessed: 4 September 2023).

Sorbolini, S. *et al.* (2017) 'Genome wide association study on beef production traits in Marchigiana cattle breed', *Journal of Animal Breeding and Genetics*, 134(1), pp. 43–48. Available at: https://doi.org/10.1111/jbg.12227.

Sordillo, L.M. and Aitken, S.L. (2009) 'Impact of oxidative stress on the health and immune function of dairy cattle', *Veterinary Immunology and Immunopathology*, 128(1), pp. 104–109. Available at: https://doi.org/10.1016/j.vetimm.2008.10.305.

Sponenberg, D. *et al.* (2018) 'Conservation of rare and local breeds of livestock: -EN- -FR- La conservation des races rares ou locales d'animaux d'élevage -ES- Conservación de razas raras o autóctonas de ganado', *Revue Scientifique et Technique de l'OIE*, 37, pp. 259–267. Available at: https://doi.org/10.20506/rst.37.1.2756.

Srivastava, S. *et al.* (2017) 'Regulation of mitochondrial protein import by the nucleotide exchange factors GrpEL1 and GrpEL2 in human cells', *Journal of Biological Chemistry*, 292(44), pp. 18075–18090. Available at: https://doi.org/10.1074/jbc.M117.788463.

Stafuzza, N.B. *et al.* (2018) 'Genome-wide association study provides insights into genes related with horn development in Nelore beef cattle', *PLOS ONE*, 13(8), p. e0202978. Available at: https://doi.org/10.1371/journal.pone.0202978.

Stambuk, C.R., Staiger, E.A., Heins, B.J., *et al.* (2020) 'Exploring physiological and genetic variation of digital cushion thickness in Holstein and Jersey cows and bulls', *Journal of Dairy Science*, 103(10), pp. 9177–9194. Available at: https://doi.org/10.3168/jds.2020-18290.

Stambuk, C.R., Staiger, E.A., Nazari-Ghadikolaei, A., *et al.* (2020) 'Phenotypic characterization and genome-wide association studies of digital cushion thickness in Holstein cows', *Journal of Dairy Science*, 103(4), pp. 3289–3303. Available at: https://doi.org/10.3168/jds.2019-17409.

Steinfeld, H., Wassenaar, T. and Jutzi, S. (2006) 'Livestock production systems in developing countries: Status, drivers, trends', *Revue scientifique et technique (International Office of Epizootics)*, 25, pp. 505–16.

Stephan, W. (2019) 'Selective Sweeps', *Genetics*, 211(1), pp. 5–13. Available at: https://doi.org/10.1534/genetics.118.301319.

Stevenson, B.W. *et al.* (2020) 'A structural view of PA2G4 isoforms with opposing functions in cancer', *Journal of Biological Chemistry*, 295(47), pp. 16100–16112. Available at: https://doi.org/10.1074/jbc.REV120.014293.

Stock, K. and Reents, R. (2013) 'Genomic Selection: Status in Different Species and Challenges for Breeding', *Reproduction in Domestic Animals*, 48(s1), pp. 2–10. Available at: https://doi.org/10.1111/rda.12201.

St-Pierre, N.R., Cobanov, B. and Schnitkey, G. (2003) 'Economic Losses from Heat Stress by US Livestock Industries1', *Journal of Dairy Science*, 86, pp. E52–E77. Available at: https://doi.org/10.3168/jds.S0022-0302(03)74040-5.

Strillacci, M.G. *et al.* (2021) 'A genome-wide scan of copy number variants in three Iranian indigenous river buffaloes', *BMC Genomics*, 22(1), p. 305. Available at: https://doi.org/10.1186/s12864-021-07604-3.

Stucki, S. *et al.* (2017) 'High performance computation of landscape genomic models including local indicators of spatial association', *Molecular Ecology Resources*, 17(5), pp. 1072–1089. Available at: https://doi.org/10.1111/1755-0998.12629.

Subhashchandra bose, B.K. *et al.* (2014) 'Residual feed intake as a feed efficiency selection tool and its relationship with feed intake, performance and nutrient utilization in Murrah buffalo calves', *Tropical Animal Health and Production*, 46(4), pp. 615–621. Available at: https://doi.org/10.1007/s11250-014-0536-2.

Suchyta, S.P. *et al.* (2003) 'Bovine mammary gene expression profiling using a cDNA microarray enhanced for mammary-specific transcripts', *Physiological Genomics*, 16(1), pp. 8–18. Available at: https://doi.org/10.1152/physiolgenomics.00028.2003.

Sun, L. *et al.* (2023) 'Assessing genomic diversity and selective pressures in Bashan cattle by whole-genome sequencing data', *Animal Biotechnology*, 34(4), pp. 835–846. Available at: https://doi.org/10.1080/10495398.2021.1998094.

Sun, T., Shen, J., *et al.* (2020) 'Genomic analyses reveal distinct genetic architectures and selective pressures in buffaloes', *GigaScience*, 9(2), p. giz166. Available at: https://doi.org/10.1093/gigascience/giz166.

Sun, T., Wang, S., *et al.* (2020) 'Multiple domestication of swamp buffalo in China and South East Asia', *Journal of Animal Breeding and Genetics*, 137(3), pp. 331–340. Available at: https://doi.org/10.1111/jbg.12445.

Sun, T., Huang, G., *et al.* (2020) 'Selection signatures of Fuzhong Buffalo based on whole-genome sequences', *BMC Genomics*, 21(1), p. 674. Available at: https://doi.org/10.1186/s12864-020-07095-8.

Sved, J.A. (1971) 'Linkage disequilibrium and homozygosity of chromosome segments in finite populations', *Theoretical Population Biology*, 2(2), pp. 125–141. Available at: https://doi.org/10.1016/0040-5809(71)90011-6.

Swartz, T.H. *et al.* (2021) 'Characterization of the liver proteome in dairy cows experiencing negative energy balance at early lactation', *Journal of Proteomics*, 246, p. 104308. Available at: https://doi.org/10.1016/j.jprot.2021.104308.

Syed, A. *et al.* (2023) 'In Silico Analysis of MARS1 Gene to Elucidate Low- Frequency Variants Associated with Interstitial Lung and Liver Disease':, *Lahore Garrison University Journal of Life Sciences*, 7(01), pp. 68–93. Available at: https://doi.org/10.54692/lgujls.2023.0701240.

Syvänen, A.-C. (2005) 'Toward genome-wide SNP genotyping', *Nature Genetics*, 37(6), pp. S5–S10. Available at: https://doi.org/10.1038/ng1558.

Szmatoła, T. *et al.* (2019) 'A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types', *Animals*, 9(12), p. 1024. Available at: https://doi.org/10.3390/ani9121024.

Szpiech, Z.A. *et al.* (2019) 'Ancestry-Dependent Enrichment of Deleterious Homozygotes in Runs of Homozygosity', *The American Journal of Human Genetics*, 105(4), pp. 747–762. Available at: https://doi.org/10.1016/j.ajhg.2019.08.011.

Szpiech, Z.A. and Hernandez, R.D. (2014) 'selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection', *Molecular Biology and Evolution*, 31(10), pp. 2824–2827. Available at: https://doi.org/10.1093/molbev/msu211.

Taberlet, P. *et al.* (2008) 'Are cattle, sheep, and goats endangered species?', *Molecular Ecology*, 17(1), pp. 275–284. Available at: https://doi.org/10.1111/j.1365-294X.2007.03475.x.

Taberlet, P. *et al.* (2011) 'Conservation genetics of cattle, sheep, and goats', *Comptes Rendus Biologies*, 334(3), pp. 247–254. Available at: https://doi.org/10.1016/j.crvi.2010.12.007.

Tajima, F. (1989) 'Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism', *Genetics*, 123(3), pp. 585–595.

Tanaka, K. *et al.* (1996) 'Phylogenetic relationship among all living species of the genusBubalus based on DNA sequences of the cytochromeb gene', *Biochemical Genetics*, 34(11), pp. 443–452. Available at: https://doi.org/10.1007/BF00570125.

Tang, Z. *et al.* (2019) 'Genome-Wide Association Study Reveals Candidate Genes for Growth Relevant Traits in Pigs', *Frontiers in Genetics*, 10. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00302 (Accessed: 31 August 2023).

Tarekegn, G.M. *et al.* (2021) 'Single-step genome-wide association study uncovers known and novel candidate genomic regions for endocrine and classical fertility traits in Swedish Red and Holstein dairy cows', *Livestock Science*, 253, p. 104731. Available at: https://doi.org/10.1016/j.livsci.2021.104731.

Téteau, O. *et al.* (2023) 'Bisphenol A and bisphenol S both disrupt ovine granulosa cell steroidogenesis but through different molecular pathways', *Journal of Ovarian Research*, 16(1), p. 30. Available at: https://doi.org/10.1186/s13048-023-01114-4.

Tetens, J. *et al.* (2015) 'Polymorphisms within the APOBR gene are highly associated with milk levels of prognostic ketosis biomarkers in dairy cows', *Physiological Genomics*, 47(4), pp. 129–137. Available at: https://doi.org/10.1152/physiolgenomics.00126.2014.

Thakor, P.B. *et al.* (2021) 'High-throughput genotype-based population structure analysis of selected buffalo breeds', *Translational Animal Science*, 5(2), p. txab033. Available at: https://doi.org/10.1093/tas/txab033.

Than, V.T. *et al.* (2019) 'Bioinformatic identification and expression analysis of the chicken B cell lymphoma (BCL) gene', *Genes & Genomics*, 41(10), pp. 1195–1206. Available at: https://doi.org/10.1007/s13258-019-00849-z.

Thiruvenkadan, A.K. *et al.* (2014) 'Buffalo Bulletin (September 2014) Vol.33 No.3', *Buffalo Bulletin* [Preprint].

Thornton, P.K. (2010) 'Livestock production: recent trends, future prospects', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554), pp. 2853–2867. Available at: https://doi.org/10.1098/rstb.2010.0134.

Toedebusch, R.G. *et al.* (2021) 'Microglia-Derived Olfactomedin-like 3 Promotes Pro-Tumorigenic Microglial Function and Malignant Features of Glioma Cells', *International Journal of Molecular Sciences*, 22(23), p. 13052. Available at: https://doi.org/10.3390/ijms222313052.

Tomarev, S.I. and Nakaya, N. (2009) 'Olfactomedin Domain-Containing Proteins: Possible Mechanisms of Action and Functions in Normal Development and Pathology', *Molecular Neurobiology*, 40(2), pp. 122–138. Available at: https://doi.org/10.1007/s12035-009-8076-x.

Tomley, F.M. and Shirley, M.W. (2009) 'Livestock infectious diseases and zoonoses', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530), pp. 2637–2642. Available at: https://doi.org/10.1098/rstb.2009.0133.

Toosi, A., Fernando, R.L. and Dekkers, J.C.M. (2010) 'Genomic selection in admixed and crossbred populations1', *Journal of Animal Science*, 88(1), pp. 32–46. Available at: https://doi.org/10.2527/jas.2009-1975.

Tosser-Klopp, G. *et al.* (2014) 'Design and Characterization of a 52K SNP Chip for Goats', *PLOS ONE*, 9(1), p. e86227. Available at: https://doi.org/10.1371/journal.pone.0086227.

Tripurani, S.K. *et al.* (2011) 'A Novel Functional Role for the Oocyte-Specific Transcription Factor Newborn Ovary Homeobox (NOBOX) during Early Embryonic Development in Cattle', *Endocrinology*, 152(3), pp. 1013–1023. Available at: https://doi.org/10.1210/en.2010-1134.

Trovato, A. *et al.* (2015) 'Expression of genes involved in the T cell signalling pathway in circulating immune cells of cattle 24 months following oral challenge with Bovine Amyloidotic Spongiform Encephalopathy (BASE)', *BMC Veterinary Research*, 11(1), p. 105. Available at: https://doi.org/10.1186/s12917-015-0412-y.

Tully, K. and Ryals, R. (2017) 'Nutrient cycling in agroecosystems: Balancing food and environmental objectives', *Agroecology and Sustainable Food Systems*, 41(7), pp. 761–798. Available at: https://doi.org/10.1080/21683565.2017.1336149.

Tumasian, R.A. *et al.* (2021) 'Skeletal muscle transcriptome in healthy aging', *Nature Communications*, 12(1), p. 2014. Available at: https://doi.org/10.1038/s41467-021-22168-2.

Tyagi, S.K. *et al.* (2021) 'Comparative Signatures of Selection Analyses Identify Loci Under Positive Selection in the Murrah Buffalo of India', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.673697 (Accessed: 26 October 2022).

Udell, M.A.R., Dorey, N.R. and Wynne, C.D.L. (2010) 'What did domestication do to dogs? A new account of dogs' sensitivity to human actions', *Biological Reviews*, 85(2), pp. 327–345. Available at: https://doi.org/10.1111/j.1469-185X.2009.00104.x.

Upadhyay, M. *et al.* (2021) 'Whole genome sequencing reveals a complex introgression history and the basis of adaptation to subarctic climate in wild sheep', *Molecular Ecology*, 30(24), pp. 6701–6717. Available at: https://doi.org/10.1111/mec.16184.

Valente, T.S. *et al.* (2016) 'Genome-Wide Association Study between Single Nucleotide Polymorphisms and Flight Speed in Nellore Cattle', *PLOS ONE*, 11(6), p. e0156956. Available at: https://doi.org/10.1371/journal.pone.0156956.

Valsalan, J. *et al.* (2014) 'Enhancing milk and fertility performances using selection index developed for Indian Murrah buffaloes', *Tropical Animal Health and Production*, 46(6), pp. 967–974. Available at: https://doi.org/10.1007/s11250-014-0596-3.

VanLiere, J.M. and Rosenberg, N.A. (2008) 'Mathematical properties of the r2<math><msup is="true"><mrow is="true"><mi is="true">r</mi></mrow><mrow is="true"><mn is="true">2</mn></mrow></msup></math> measure of linkage disequilibrium', *Theoretical Population Biology*, 74(1), pp. 130–137. Available at: https://doi.org/10.1016/j.tpb.2008.05.006.

Vardhan, K.H., Kumar, P.S. and Panda, R.C. (2019) 'A review on heavy metal pollution, toxicity and remedial measures: Current trends and future perspectives', *Journal of Molecular Liquids*, 290, p. 111197. Available at: https://doi.org/10.1016/j.molliq.2019.111197.

Vaysse, A. *et al.* (2011) 'Identification of Genomic Regions Associated with Phenotypic Variation between Dog Breeds using Selection Mapping', *PLOS Genetics*, 7(10), p. e1002316. Available at: https://doi.org/10.1371/journal.pgen.1002316.

Vecchio, D. *et al.* (no date) 'Reproductive management of buffalo (Bubalus bubalis) in Italy'.

Velado-Alonso, E., Morales-Castilla, I. and Gómez-Sal, A. (2020) 'Recent land use and management changes decouple the adaptation of livestock diversity to the environment', *Scientific Reports*, 10(1), p. 21035. Available at: https://doi.org/10.1038/s41598-020-77878-2.

Velado-Alonso, E., Morales-Castilla, I. and Gómez-Sal, A. (2022) 'The landscapes of livestock diversity: grazing local breeds as a proxy for domesticated species adaptation to the environment', *Landscape Ecology*, 37(4), pp. 1035–1048. Available at: https://doi.org/10.1007/s10980-022-01429-5.

Venot, J.-P., Reddy, V.R. and Umapathy, D. (2010) 'Coping with drought in irrigated South India: Farmers' adjustments in Nagarjuna Sagar', *Agricultural Water Management*, 97(10), pp. 1434–1442. Available at: https://doi.org/10.1016/j.agwat.2010.04.009.

Venturini, G.C. *et al.* (2014) 'Association between single-nucleotide polymorphisms and milk production traits in buffalo', *Genetics and molecular research: GMR*, 13(4), pp. 10256–10268. Available at: https://doi.org/10.4238/2014.December.4.20.

Verlouw, J.A.M. *et al.* (2021) 'A comparison of genotyping arrays', *European Journal of Human Genetics*, 29(11), pp. 1611–1624. Available at: https://doi.org/10.1038/s41431-021-00917-7.

Verma, M.K. *et al.* (2016) 'Effect of genetic and non-genetic factors on milk yield and milk constituents in Murrah buffalo', *Indian Journal of Animal Research*, 51. Available at: https://doi.org/10.18805/ijar.9297.

Vernaz, G. *et al.* (2021) 'Mapping epigenetic divergence in the massive radiation of Lake Malawi cichlid fishes', *Nature Communications*, 12(1), p. 5870. Available at: https://doi.org/10.1038/s41467-021-26166-2.

Vignal, A. *et al.* (2002) 'A review on SNP and other types of molecular markers and their use in animal genetics', *Genetics Selection Evolution*, 34(3), p. 275. Available at: https://doi.org/10.1186/1297-9686-34-3-275.

Vigne, J.-D. (2011) 'The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere', *Comptes Rendus Biologies*, 334(3), pp. 171–181. Available at: https://doi.org/10.1016/j.crvi.2010.12.009.

Vilà, C. *et al.* (1997) 'Multiple and Ancient Origins of the Domestic Dog', *Science*, 276(5319), pp. 1687–1689. Available at: https://doi.org/10.1126/science.276.5319.1687.

Villanueva, M.A. *et al.* (2018) *Emerging Infectious Diseases in Water Buffalo: An Economic and Public Health Concern*, *Emerging Infectious Diseases in Water Buffalo - An Economic and Public Health Concern*. IntechOpen. Available at: https://doi.org/10.5772/intechopen.73395.

Villarino, A.V., Kanno, Y. and O'Shea, J.J. (2017) 'Mechanisms and consequences of Jak–STAT signaling in the immune system', *Nature Immunology*, 18(4), pp. 374–384. Available at: https://doi.org/10.1038/ni.3691.

Vislobokova, I.A. *et al.* (2021) 'An unexpected record of an extinct water buffalo Bubalus murrensis (Berckhemer, 1927) in the Last Glacial in Europe and its implication for dispersal pattern of this species', *Quaternary International*, 574, pp. 127–136. Available at: https://doi.org/10.1016/j.quaint.2020.12.020.

Vislobokova, I.A., Tarasenko, K.K. and Lopatin, A.V. (2020) 'First Discovery of the European Buffalo Bubalus murrensis (Artiodactyla, Bovidae) from the Pleistocene of the Russian Plain', *Doklady Biological Sciences*, 491(1), pp. 31–34. Available at: https://doi.org/10.1134/S001249662002012X.

Voight, B.F. *et al.* (2006) 'A Map of Recent Positive Selection in the Human Genome', *PLOS Biology*, 4(3), p. e72. Available at: https://doi.org/10.1371/journal.pbio.0040072.

vonHoldt, B.M. *et al.* (2010) 'Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication', *Nature*, 464(7290), pp. 898–902. Available at: https://doi.org/10.1038/nature08837.

Wahlund, S. (1928) 'Zusammensetzung Von Populationen Und Korrelationserscheinungen Vom Standpunkt Der Vererbungslehre Aus Betrachtet', *Hereditas*, 11(1), pp. 65–106. Available at: https://doi.org/10.1111/j.1601-5223.1928.tb02483.x.

Wang, B., Chen, L. and Wang, W. (2019) 'Genomic insights into ruminant evolution: from past to future prospects', *Zoological Research*, 40(6), pp. 476–487. Available at: https://doi.org/10.24272/j.issn.2095-8137.2019.061.

Wang, J. (2005) 'Estimation of effective population sizes from data on genetic markers', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1459), pp. 1395–1409. Available at: https://doi.org/10.1098/rstb.2005.1682.

Wang, J. *et al.* (2020) 'Comparison of the Transcriptome of the Ovine Mammary Gland in Lactating and Non-lactating Small-Tailed Han Sheep', *Frontiers in Genetics*, 11. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00472 (Accessed: 10 December 2022).

Wang, L. *et al.* (2013) 'Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds', *BMC Genetics*, 14(1), p. 115. Available at: https://doi.org/10.1186/1471-2156-14-115.

Wang, M. *et al.* (2012) 'MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation1', *Journal of Dairy Science*, 95(11), pp. 6529–6535. Available at: https://doi.org/10.3168/jds.2012-5748.

Wang, S. *et al.* (2017) 'Whole Mitogenomes Reveal the History of Swamp Buffalo: Initially Shaped by Glacial Periods and Eventually Modelled by Domestication', *Scientific Reports*, 7, p. 4708. Available at: https://doi.org/10.1038/s41598-017-04830-2.

Wang, Shaohua *et al.* (2021) 'Polymerase acidic subunit of H9N2 polymerase complex induces cell apoptosis by binding to PDCD 7 in A549 cells', *Virology Journal*, 18(1), p. 75. Available at: https://doi.org/10.1186/s12985-021-01547-7.

Wang, Shenyuan *et al.* (2021) 'Screening of genes coupled to heat response in Mongolian and Dorper sheep breeds', *Biologia*, 76(3), pp. 949–958. Available at: https://doi.org/10.2478/s11756-020-00616-6.

Wang, X. (2006) 'Zhanxiang, Tao, and Banyue: Early Pleistocene Mammalian Fauna from Longdan, Dongxiang, Gansu, China', *Journal of Vertebrate Paleontology*, 26(1), pp. 230–231. Available at: https://doi.org/10.1671/0272-4634(2006)26[230:R]2.0.CO;2.

Wang, X. *et al.* (2009) 'Vegetation on the Sunda Shelf, South China Sea, during the Last Glacial Maximum', *Palaeogeography, Palaeoclimatology, Palaeoecology*, 278(1), pp. 88–97. Available at: https://doi.org/10.1016/j.palaeo.2009.04.008.

Wang, X. *et al.* (2022) 'GWAS of Reproductive Traits in Large White Pigs on Chip and Imputed Whole-Genome Sequencing Data', *International Journal of Molecular Sciences*, 23(21), p. 13338. Available at: https://doi.org/10.3390/ijms232113338.

Wang, X. *et al.* (2023) 'A novel SNP in NKX1-2 gene is associated with carcass traits in Dezhou donkey', *BMC Genomic Data*, 24(1), p. 41. Available at: https://doi.org/10.1186/s12863-023-01145-2.

Wang, Y. *et al.* (2019) 'Genetic basis of ruminant headgear and rapid antler regeneration', *Science*, 364(6446), p. eaav6335. Available at: https://doi.org/10.1126/science.aav6335.

Wang, Y. *et al.* (2021) 'Comparative distributions of RSBN1 and methylated histone H4 Lysine 20 in the mouse spermatogenesis', *PLOS ONE*, 16(6), p. e0253897. Available at: https://doi.org/10.1371/journal.pone.0253897.

Wang, Yuan *et al.* (2022) 'Differential expression of the Tmem132 family genes in the developing mouse nervous system', *Gene Expression Patterns*, 45, p. 119257. Available at: https://doi.org/10.1016/j.gep.2022.119257.

Wang, Yiliang *et al.* (2022) 'MAMDC2, a gene highly expressed in microglia in experimental models of Alzheimers Disease, positively regulates the innate antiviral response during neurotropic virus infection', *Journal of Infection*, 84(2), pp. 187–204. Available at: https://doi.org/10.1016/j.jinf.2021.12.004.

Webb, T.R. and Clark, A.J.L. (2010) 'Minireview: The Melanocortin 2 Receptor Accessory Proteins', *Molecular Endocrinology*, 24(3), pp. 475–484. Available at: https://doi.org/10.1210/me.2009-0283.

Webster, M.T. and Hurst, L.D. (2012) 'Direct and indirect consequences of meiotic recombination: implications for genome evolution', *Trends in Genetics*, 28(3), pp. 101–109. Available at: https://doi.org/10.1016/j.tig.2011.11.002.

Wegmann, D. *et al.* (2010) 'ABCtoolbox: a versatile toolkit for approximate Bayesian computations', *BMC Bioinformatics*, 11(1), p. 116. Available at: https://doi.org/10.1186/1471-2105-11-116.

Weldenegodguad, M. *et al.* (2019) 'Whole-Genome Sequencing of Three Native Cattle Breeds Originating From the Northernmost Cattle Farming Regions', *Frontiers in Genetics*, 9. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00728 (Accessed: 5 January 2024).

Wellnitz, O. and Bruckmaier, R.M. (2021) 'Invited review: The role of the blood–milk barrier and its manipulation for the efficacy of the mammary immune response and milk production', *Journal of Dairy Science*, 104(6), pp. 6376–6388. Available at: https://doi.org/10.3168/jds.2020-20029.

White, N. *et al.* (2003) 'The Vulnerability of the Australian Beef Industry to Impacts of the Cattle Tick (Boophilus microplus) under Climate Change', *Climatic Change*, 61(1), pp. 157–190. Available at: https://doi.org/10.1023/A:1026354712890.

Wiedemar, N. *et al.* (2014) 'Independent Polled Mutations Leading to Complex Gene Expression Differences in Cattle', *PLOS ONE*, 9(3), p. e93435. Available at: https://doi.org/10.1371/journal.pone.0093435.

Wierzbicka, A. *et al.* (2023) 'Effect of different doses of cholecalciferol and calcidiol on meat quality parameters and skeletal muscle transcriptome profiles in swine', *Meat Science*, 197, p. 109071. Available at: https://doi.org/10.1016/j.meatsci.2022.109071.

Wiggans, G.R. *et al.* (2017) 'Genomic Selection in Dairy Cattle: The USDA Experience', *Annual Review of Animal Biosciences*, 5(1), pp. 309–327. Available at: https://doi.org/10.1146/annurev-animal-021815-111422.

Willi, Y. *et al.* (2007) 'Genetic isolation of fragmented populations is exacerbated by drift and selection', *Journal of Evolutionary Biology*, 20(2), pp. 534–542. Available at: https://doi.org/10.1111/j.1420-9101.2006.01263.x.

Williams, J.L. *et al.* (2017) 'Genome assembly and transcriptome resource for river buffalo, Bubalus bubalis (2n = 50)', *GigaScience*, 6(10), p. gix088. Available at: https://doi.org/10.1093/gigascience/gix088.

Williams, M.M. *et al.* (2017) 'ErbB3 drives mammary epithelial survival and differentiation during pregnancy and lactation', *Breast Cancer Research*, 19(1), p. 105. Available at: https://doi.org/10.1186/s13058-017-0893-7.

de Winter, J.M. *et al.* (no date) 'KBTBD13 is a novel cardiomyopathy gene', *Human Mutation*, n/a(n/a). Available at: https://doi.org/10.1002/humu.24499.

Woolhouse, M. *et al.* (2015) 'Antimicrobial resistance in humans, livestock and the wider environment', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1670), p. 20140083. Available at: https://doi.org/10.1098/rstb.2014.0083.

Wordsworth, P.D. *et al.* (2021) 'The earliest water buffalo in the Caucasus: shifting animals and people in the medieval Islamic world', *Antiquity*, 95(383), pp. 1231–1247. Available at: https://doi.org/10.15184/aqy.2021.108.

Wright, A.F. (2005) 'Genetic Variation: Polymorphisms and Mutations', in *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd. Available at: https://doi.org/10.1038/npg.els.0005005.

Wu, X.-L. and Zhao, S. (2021) 'Editorial: Advances in Genomics of Crossbred Farm Animals', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.709483 (Accessed: 13 February 2024).

Wu, Z. *et al.* (2008) 'Memory T Cell RNA Rearrangement Programmed by Heterogeneous Nuclear Ribonucleoprotein hnRNPLL', *Immunity*, 29(6), pp. 863–875. Available at: https://doi.org/10.1016/j.immuni.2008.11.004.

Wykes, D.L. (2004) 'Robert Bakewell (1725-1795) of Dishley: Farmer and Livestock Improver', *The Agricultural History Review*, 52(1), pp. 38–55.

Xu, R. *et al.* (2020) 'Environmental temperature and human epigenetic modifications: A systematic review', *Environmental Pollution*, 259, p. 113840. Available at: https://doi.org/10.1016/j.envpol.2019.113840.

Yamashiro, S. *et al.* (2012) 'Tropomodulins: Pointed-end capping proteins that regulate actin filament architecture in diverse cell types', *Cytoskeleton*, 69(6), pp. 337–370. Available at: https://doi.org/10.1002/cm.21031.

Yáñez, A. *et al.* (2020) 'Physiological and Behavioral Changes of Water Buffalo in Hot and Cold Systems: Review', *Journal of Buffalo Science*, 9, pp. 110–120. Available at: https://doi.org/10.6000/1927-520X.2020.09.13.

Yang, D. *et al.* (2020) 'BCL2L15 Depletion Inhibits Endometrial Receptivity via the STAT1 Signaling Pathway', *Genes*, 11(7), p. 816. Available at: https://doi.org/10.3390/genes11070816.

Yang, D.Y. *et al.* (2008) 'Wild or domesticated: DNA analysis of ancient water buffalo remains from north China', *Journal of Archaeological Science*, 35(10), pp. 2778–2785. Available at: https://doi.org/10.1016/j.jas.2008.05.010.

Yang, R. *et al.* (2023) 'Grpel2 maintains cardiomyocyte survival in diabetic cardiomyopathy through DLST-mediated mitochondrial dysfunction: a proof-of-concept study', *Journal of Translational Medicine*, 21(1), p. 200. Available at: https://doi.org/10.1186/s12967-023-04049-y.

Yang, S.-H. *et al.* (2015) 'Validation of PDE9A Gene Identified in GWAS Showing Strong Association with Milk Production Traits in Chinese Holstein', *International Journal of Molecular Sciences*, 16(11), pp. 26530–26542. Available at: https://doi.org/10.3390/ijms161125976.

Yaro, M. *et al.* (2017) 'Molecular identification of livestock breeds: a tool for modern conservation biology', *Biological Reviews*, 92(2), pp. 993–1010. Available at: https://doi.org/10.1111/brv.12265.

Ye, T. *et al.* (2022) 'Evolutionary analysis of buffalo sterol regulatory element-binding factor (SREBF) family genes and their affection on milk traits', *Animal Biotechnology*, 0(0), pp. 1–12. Available at: https://doi.org/10.1080/10495398.2022.2070185.

Yindee, M. *et al.* (2010) 'Y-chromosomal variation confirms independent domestications of swamp and river buffalo', *Animal Genetics*, 41(4), pp. 433–435. Available at: https://doi.org/10.1111/j.1365-2052.2010.02020.x.

Yoo, T. *et al.* (2020) 'A DLG2 deficiency in mice leads to reduced sociability and increased repetitive behavior accompanied by aberrant synaptic transmission in the dorsal striatum', *Molecular Autism*, 11(1), p. 19. Available at: https://doi.org/10.1186/s13229-020-00324-7.

Yoshida, G.M. *et al.* (2020) 'Estimates of Autozygosity Through Runs of Homozygosity in Farmed Coho Salmon', *Genes*, 11(5), p. 490. Available at: https://doi.org/10.3390/genes11050490.

You, M. *et al.* (2020) 'Alpha-Linolenic Acid-Enriched Butter Promotes Fatty Acid Remodeling and Thermogenic Activation in the Brown Adipose Tissue', *Nutrients*, 12(1), p. 136. Available at: https://doi.org/10.3390/nu12010136.

You, S. *et al.* (2021) 'African swine fever outbreaks in China led to gross domestic product and economic losses', *Nature Food*, 2(10), pp. 802–808. Available at: https://doi.org/10.1038/s43016-021-00362-1.

Yougbaré, B. *et al.* (2021) 'Genome-wide association study of trypanosome prevalence and morphometric traits in purebred and crossbred Baoulé cattle of Burkina Faso', *PLOS ONE*, 16(8), p. e0255089. Available at: https://doi.org/10.1371/journal.pone.0255089.

Yousefzadeh, M.J. and Wood, R.D. (2013) 'DNA polymerase POLQ and cellular defense against DNA damage', *DNA Repair*, 12(1), pp. 1–9. Available at: https://doi.org/10.1016/j.dnarep.2012.10.004.

Yu, S.L. *et al.* (2007) 'Identification of Differentially Expressed Genes in Distinct Skeletal Muscles in Cattle using cDNA Microarray', *Animal Biotechnology*, 18(4), pp. 275–285. Available at: https://doi.org/10.1080/10495390701413391.

Yu, X. *et al.* (2017) 'TUSC3: a novel tumour suppressor gene and its functional implications', *Journal of Cellular and Molecular Medicine*, 21(9), pp. 1711–1718. Available at: https://doi.org/10.1111/jcmm.13128.

Yuan, Y. *et al.* (2022) 'Genome-Wide Selective Analysis of Boer Goat to Investigate the Dynamic Heredity Evolution under Different Stages', *Animals*, 12(11), p. 1356. Available at: https://doi.org/10.3390/ani12111356.

Yuan, Z. *et al.* (2019) 'Transcriptomic analyses revealed common tailed and perirenal adipose differentially expressed genes in four Chinese indigenous sheep breeds', *Livestock Science*, 230, p. 103832. Available at: https://doi.org/10.1016/j.livsci.2019.103832.

Yue, X.-P. *et al.* (2013) 'Phylogeography and Domestication of Chinese Swamp Buffalo', *PLOS ONE*, 8(2), p. e56552. Available at: https://doi.org/10.1371/journal.pone.0056552.

Zava, M.A. (2009) 'PRESENT SITUATION AND FUTURE PERSPECTIVE OF BUFFALO PRODUCTION IN AMERICA', *J. Anim Plant Sci*, p. 8.

Zeder, M.A. (2012) 'The Domestication of Animals', *Journal of Anthropological Research*, 68(2), pp. 161–190. Available at: https://doi.org/10.3998/jar.0521004.0068.201.

Zeitlmann, L. *et al.* (2001) 'Cloning of ACP33 as a Novel Intracellular Ligand of CD4 *', *Journal of Biological Chemistry*, 276(12), pp. 9123–9132. Available at: https://doi.org/10.1074/jbc.M009270200.

Zeng, L. *et al.* (2022) 'Genes related to heat tolerance in cattle—a review', *Animal Biotechnology*, 0(0), pp. 1–9. Available at: https://doi.org/10.1080/10495398.2022.2047995.

Zhang, X. *et al.* (2021) 'A High-Density Genetic Linkage Map and Fine Mapping of QTL For Feed Conversion Efficiency in Common Carp (Cyprinus carpio)', *Frontiers in Genetics*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2021.778487 (Accessed: 25 August 2023).

Zhang, Y. *et al.* (2007) 'Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers', *Animal Genetics*, 38(6), pp. 569–575. Available at: https://doi.org/10.1111/j.1365-2052.2007.01648.x.

Zhang, Y. *et al.* (2011) 'Genetic differentiation of water buffalo (Bubalus bubalis) populations in China, Nepal and south-east Asia: inferences on the region of domestication of the swamp buffalo', *Animal Genetics*, 42(4), pp. 366–377. Available at: https://doi.org/10.1111/j.1365-2052.2010.02166.x.

Zhang, Y., Colli, L. and Barker, J.S.F. (2020) 'Asian water buffalo: domestication, history and genetics', *Animal Genetics*, 51(2), pp. 177–191. Available at: https://doi.org/10.1111/age.12911.

Zhang, Yi *et al.* (2016) 'Strong and stable geographic differentiation of swamp buffalo maternal and paternal lineages indicates domestication in the China/Indochina border region', *Molecular Ecology*, 25(7), pp. 1530–1550. Available at: https://doi.org/10.1111/mec.13518.

Zhang, Z. *et al.* (2014) 'O-GlcNAcase Expression is Sensitive to Changes in O-GlcNAc Homeostasis', *Frontiers in Endocrinology*, 5. Available at: https://www.frontiersin.org/articles/10.3389/fendo.2014.00206 (Accessed: 8 December 2022).

Zhao, F. *et al.* (2014) 'Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations', *Livestock Science*, 170, pp. 22–29. Available at: https://doi.org/10.1016/j.livsci.2014.10.015.

Zhao, P. *et al.* (2021) 'ANKDD1A may serve as a critical gene in the immune microenvironment of breast cancer', *Translational Cancer Research*, 10(3), pp. 1358–1367. Available at: https://doi.org/10.21037/tcr-20-2685.

Zhong, Z.Q. *et al.* (2023) 'Genome-wide scans for selection signatures in indigenous pigs revealed candidate genes relating to heat tolerance', *animal*, 17(7), p. 100882. Available at: https://doi.org/10.1016/j.animal.2023.100882.

Zhou, H. and Clapham, D.E. (2009) 'Mammalian MagT1 and TUSC3 are required for cellular magnesium uptake and vertebrate embryonic development', *Proceedings of the National Academy of Sciences*, 106(37), pp. 15750–15755. Available at: https://doi.org/10.1073/pnas.0908332106.

Zhou, S. *et al.* (2021) 'Paxbp1 controls a key checkpoint for cell growth and survival during early activation of quiescent muscle satellite cells', *Proceedings of the National Academy of Sciences*, 118(13), p. e2021093118. Available at: https://doi.org/10.1073/pnas.2021093118.

Zhou, Z. *et al.* (2017) 'Hepatic Activity and Transcription of Betaine-Homocysteine Methyltransferase, Methionine Synthase, and Cystathionine Synthase in Periparturient Dairy Cows Are Altered to Different Extents by Supply of Methionine and Choline', *The Journal of Nutrition*, 147(1), pp. 11–19. Available at: https://doi.org/10.3945/jn.116.240234.

Zhu, S. *et al.* (2022) 'KAZN as a diagnostic marker in ovarian cancer: a comprehensive analysis based on microarray, mRNA-sequencing, and methylation data', *BMC Cancer*, 22(1), p. 662. Available at: https://doi.org/10.1186/s12885-022-09747-2.

Supplementary Materials

Table S2.1: Quality control thresholds used to call genotypes from raw data in Axiom™ Analysis Suite Software v4.03.

Threshold Type	Metric	Threshold Value
Sample Quality Control	DQC	≥ 0.82
	Quality Control Call Rate	≥ 97
	Percent of Passing Samples	≥ 95
	Average Call Rate for Passing Samples	≥ 97
SNP Quality Control	Species Type	Diploid
	Call Rate Cutoff	≥ 97
	FLD Cutoff	≥ 3.6
	Het So Cutoff	≥ -0.1
	Het so X Chr Cutoff	≥ -0.1
	Het So OTV Cutoff	≥ -0.3
	Hom RO 1 Cutoff	≥ 0.6
	Hom Ro 2 Cutoff	≥ 0.3
	Hom Ro 3 Cutoff	≥ -0.9
	Hom Ro	True
	Hom Het	True
	Num Minor Allele Cutoff	≥ 2
	Hom Ro Hap 1 X Chr Cutoff	≥ 0.1
	Hom Ro Hap 1 MT Chr Cutoff	≥ 0.4
	Hom Ro Hap 2 X Chr Cutoff	≥ 0.05
	Hom Ro Hap 2 MT Chr Cutoff	≥ 0.2
	AAF X Chr Cut	< 0.36
	FLD X Chr Cut	≥ 4
	HomFLD X Chr Cut	≥ 6.5
	HomFLD Y Chr Cut	≥ 6.5
	Min Y Chr Samples Cut	≥ 5
	Sign Diff Hom 1 Cutoff	≥ 0.5
	Sign Diff Hom 2 Cutoff	≥ 0.4
	Min Mean CP2 Cutoff	≥ 9
	Max Mean CP2 Cutoff	≤ 15
	Priority Order	PolyHighResolution, NoMinorHom, OTV, MonoHighResolution, CallRateBelowThreshold
	Recommended	PolyHighResolution, NoMinorHom, MonoHighResolution, Hemizygous
	Y Restrict	≤ 0.2

Table S2.2: Eigenvalues for each component captured by multidimensional scaling analysis.

Component	Eigenvalues
1	1.537370
2	1.087450
3	0.793862
4	0.606121
5	0.459875
6	0.390526
7	0.335176
8	0.321959
9	0.264270
10	0.222226
11	0.217258
12	0.205966
13	0.193472
14	0.180692
15	0.168650
16	0.160560
17	0.151979
18	0.148897
19	0.142883
20	0.140051

Table S2.3: AMOVA results showing partitioning of variation across individuals and populations.

Source of Variation	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among Populations within Groups	13	57241.897	69.96408 Va	9.11
Among Individuals Within Populations	365	247309.819	-20.56309 Vb	-2.68
Within Individuals	379	272382.500	718.68734 Vc	93.57
Total	757	576934.215	768.08832	

Figure S2.4: FST distributions generated by bootstrap replicates of RV_UK1 vs RV_ITA and RV_UK2 vs RV_ITA

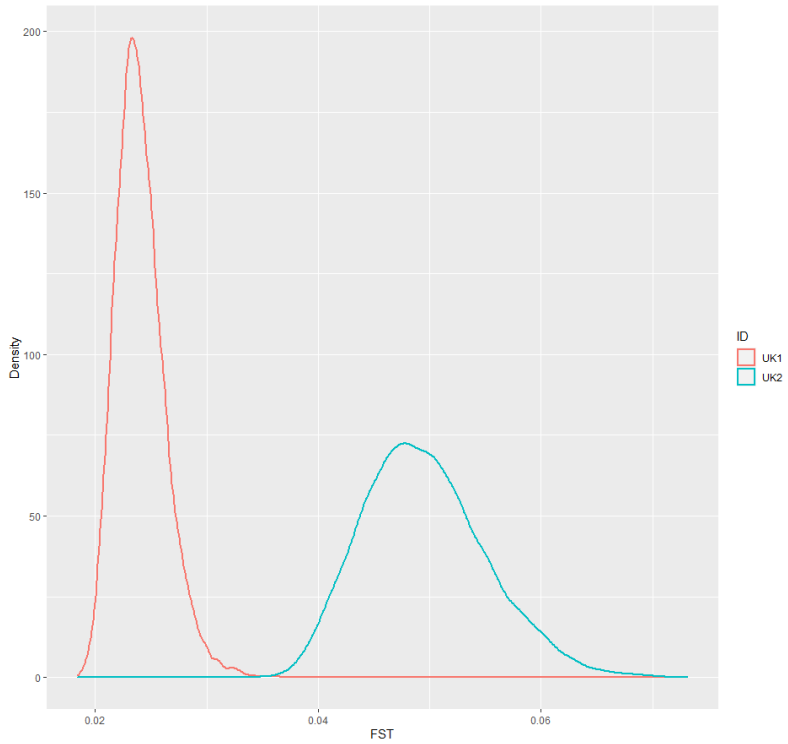


Figure S2.5: FST distributions generated by bootstrap replicates of RV_UK1 vs RV_ROM_CL and RV_UK2 vs RV_ROM_CL

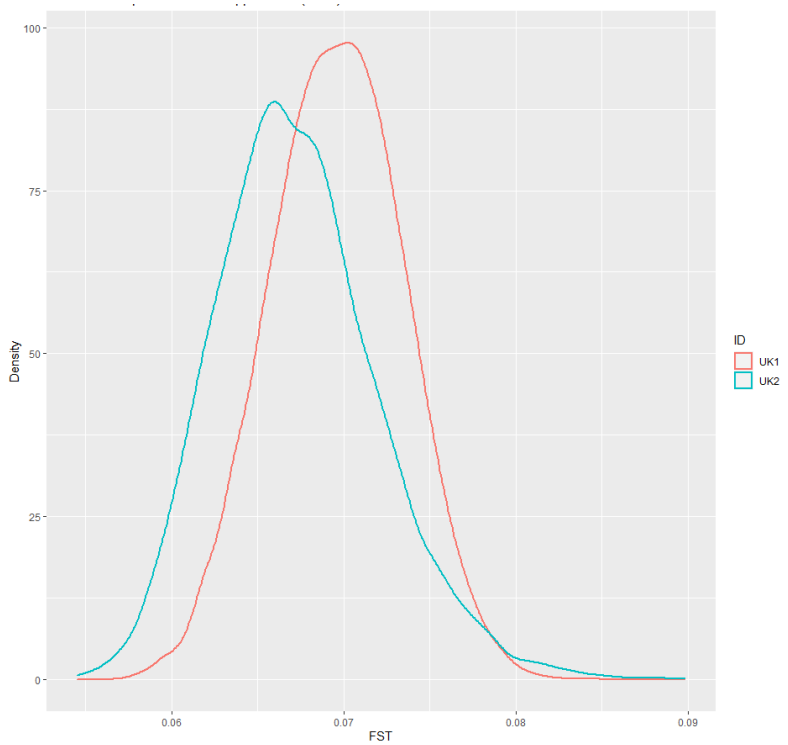


Figure S2.6: FROH distributions generated in RV_UK1 and RV_UK2

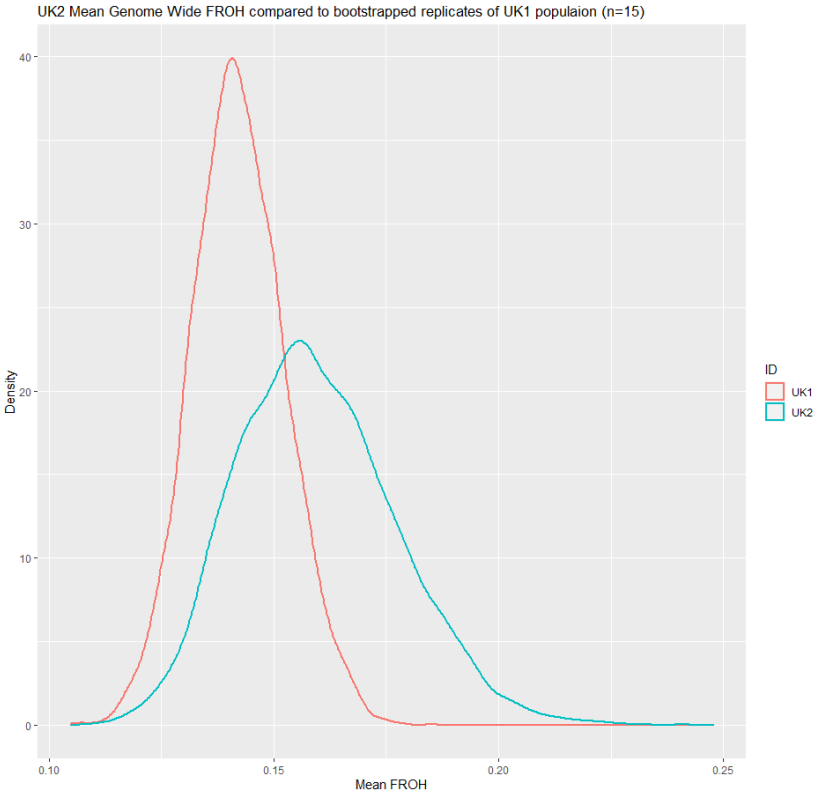


Figure S2.7: Plot showing sum ROHs vs total number of ROHs for all individuals of RV_UK1 and RV_UK2. X = y shown by dashed black line.

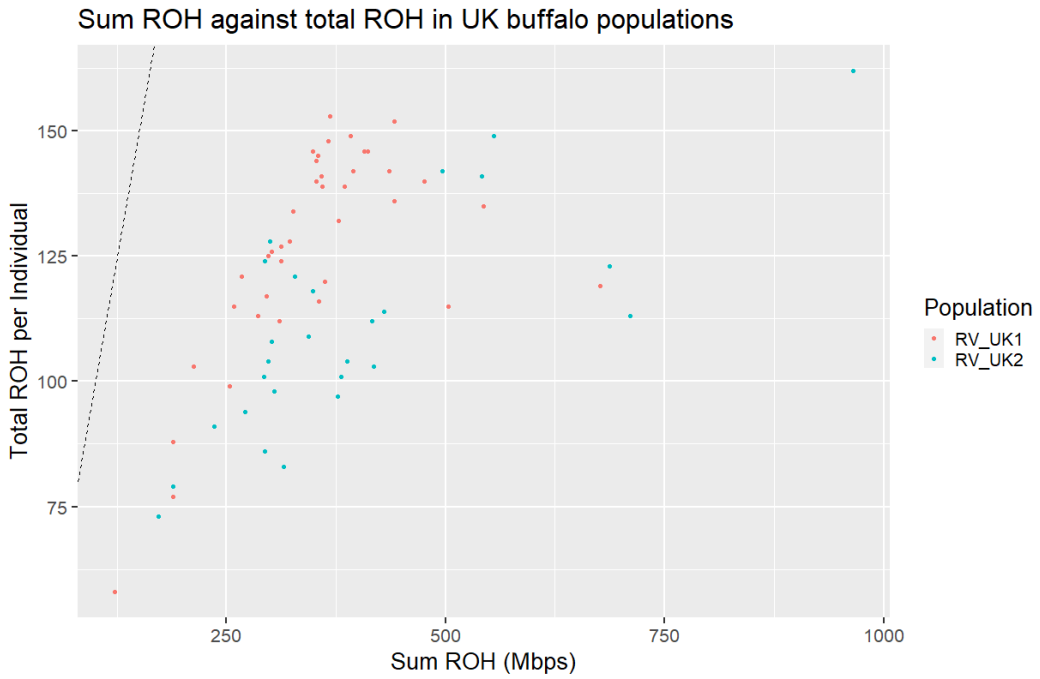


Table S2.8: Significant ROHs and genes found within regions for both RV_UK1 and RV_UK2.

Population	Chromosome (BBU)	Start	End	Genes
RV_UK1	1	40618376	44032773	CSMD1, MYOM2, KBTBD11, ARHGEF10, CLN8, TRNAW-CCA, DLGAP2
		50404478	51196161	CLDN17, GRIK1
		76018494	76518042	
	2	117129833	117517993	NCKAP5
	3	61862936	61969410	-
		64021694	64797055	CBR4, SH3RF1, NEK1, CLCN3, TRNAH-AUG, MFAP3L
		82344868	83657693	IZUMO3, TRNAW-CCA, ELAVL2
	4	103159550	105807905	MGAT4C, NTS, RASSF9, ALX1, LRRIQ1, TSPAN19, SLC6A15
	5	75486279	79733876	TRNAC-ACA, PANX1, HEPHL1, VSTM5, MED17, TAF1D, CEP295, DEUP1, SLC36A4, MTNR1B, FAT3, TRNAC-GCA
		83413928	85500121	TMEM135, FZD4, PRSS23, ME3, CCDC81, HIKESHI, EED, PICALM, CCDC83, SYTL2
	6	28106618	29855359	TSPAN2, TSHB, SYCP1, SIKE1, CSDE1, NRAS, AMPD1, DENND2C, BCAS2, TRIM33, SYT6, OLFML3, HIPK1, DCLRE1B, AP4B1, BCL2L15, PTPN22, RSBN1, PHTF1, MAGI3
	14	62533451	63775086	MALRD1, ARL5B, CCDC7
	15	50475003	50726444	CPA6
	19	15255940	17233155	HTR1A, IPO11, DIMT1, KIF2A
		44467270	45125124	TRNAY-GUA, TRNAC-GCA
		55683549	57425155	BASP1, MYO10, RETREG1, ZNF622, MARCH11, FBXL7
	21	13844195	14950320	ULK4, TRAK1, CCK, LYZL4, TRNAG-CCC, VIPR1, SEC22C, SS18L2, NKTR, ZBTB47, KLHL40, HHATL, CCDC13, HIGD1A, ZNF662, TRNAG-CCC
RV_UK2	1	43763283	45246299	DLGAP2, ERICH1
		76018494	77767873	CADM2
		92325535	93881377	-
	2	53232270	53595544	OCA2, HERC2, TRNAE-CUC
		81227949	82277332	B3GALT1, XIRP2
		120494471	121809491	DPP10
		130097357	130310554	-
		131521600	133260286	TRNAC-ACA, GYPC, TEX51, GLS, STAT1, STAT4, MYO1B, TRNAE-UUC, NABP1
	3	143189749	144021308	DAPK1, TRNAY-AUA, CTSL, FBP2, FBP1
	4	61937971	64525598	TRNAG-CCC, METTL7B, ITGA7, BLOC1S1, RDH5, CD63, GDF11, SARNP, ORMDL2, DNAJC14, MMP19, PYM1, DGKA, PMEL, CDK2, RAB5B, SUOX, IKZF4, RPS26, ERBB3, PA2G4, ZC3H10, ESYT1, MYL6B, MYL6, SMARCC2, TRNAS-CGA, RNF41, NABP2, SLC39A5, ANKRD52, COQ10A, CS, CNPY2, PAN2, TRNAE-UUC, IL23A, STAT2, APOF, TIMELESS, MIP, SPRYD4, GLS2, RBMS2, BAZ2A, ATP5F1B, PTGES3, NACA, PRIM1, TRNAG-CCC, SDR9C7, GPR182, ZBTB39, TAC3, MYO1A, NEMP1, NAB2, STAT6, LRP1, NXPH4, SHMT2, NDUFA4L2, STAC3, R3HDM2, INHBC, INHBE, GLI1, ARHGAP9, MARS, DDIT3, MBD6, DCTN2, KIF5A, PIP4K2C, DTX3, ARHGEF25, SLC26A10, B4GALNT1, OS9

	131296877	132073992	CHRM3
5	75857249	79733876	PANX1, HEPHL1, VSTM5, MED17, TAF1D, CEP295, DEUP1, SLC36A4, MTNR1B, FAT3, TRNAC-GCA
6	45760687	47489139	DPYD, PTBP2
7	42446383	43345211	-
23	21754196	22410305	TLX1, LBX1, BTRC, POLL, DPCD, FBXW4, FGF8, NPM3, OGA, KCNIP2, ARMH3

Table S2.9: Gene ontology results for genes found in significant ROHs for both RV_UK1 and RV_UK2.

Population	Gene Ontology Class	Biological Pathway	Description
RV_UK1	Process	GO:2000564	Regulation of CD8-positive, alpha-beta T cell proliferation
		GO:0046328	Regulation of JNK cascade
		GO:0015824	Proline transport
RV_UK2	Component	GO:0070876	SOSS complex
		GO:0016459	Myosin complex
	Function	GO:0042132	Fructose 1,6-bisphosphate 1-phosphatase activity
		GO:0004359	Glutaminase activity
		GO:0003774	Motor activity
		GO:0030898	Actin-dependent ATPase activity
	Process	GO:0005986	Sucrose biosynthetic process
		GO:0007259	JAK-STAT cascade
		GO:0097696	STAT cascade
		GO:0006543	Glutamine catabolic process
		GO:0005985	Sucrose metabolic process
		GO:0009312	Oligosaccharide biosynthetic process
		GO:0006537	Glutamate biosynthetic process
		GO:0046351	Disaccharide biosynthetic process
		GO:0014037	Schwann cell differentiation

Figure S2.10: QTL class distribution across RV_UK1 significant ROHs

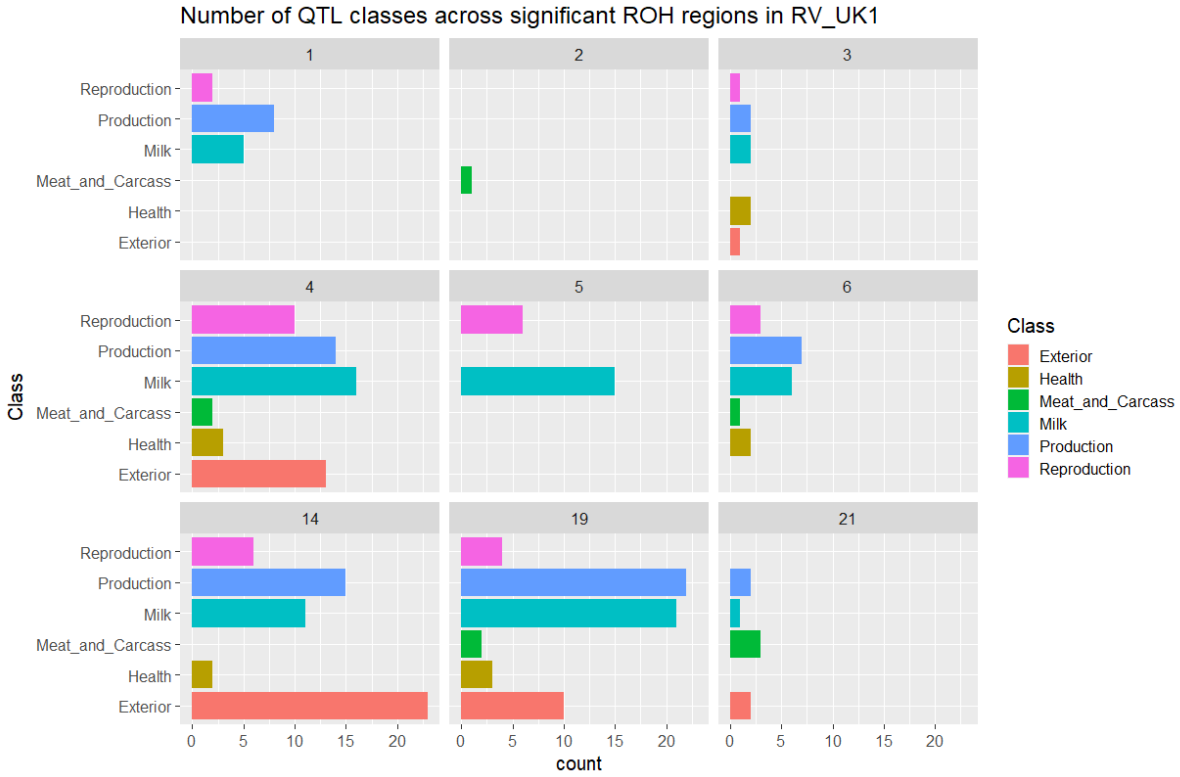


Figure S2.11: QTL class distribution across RV_UK2 significant ROHs

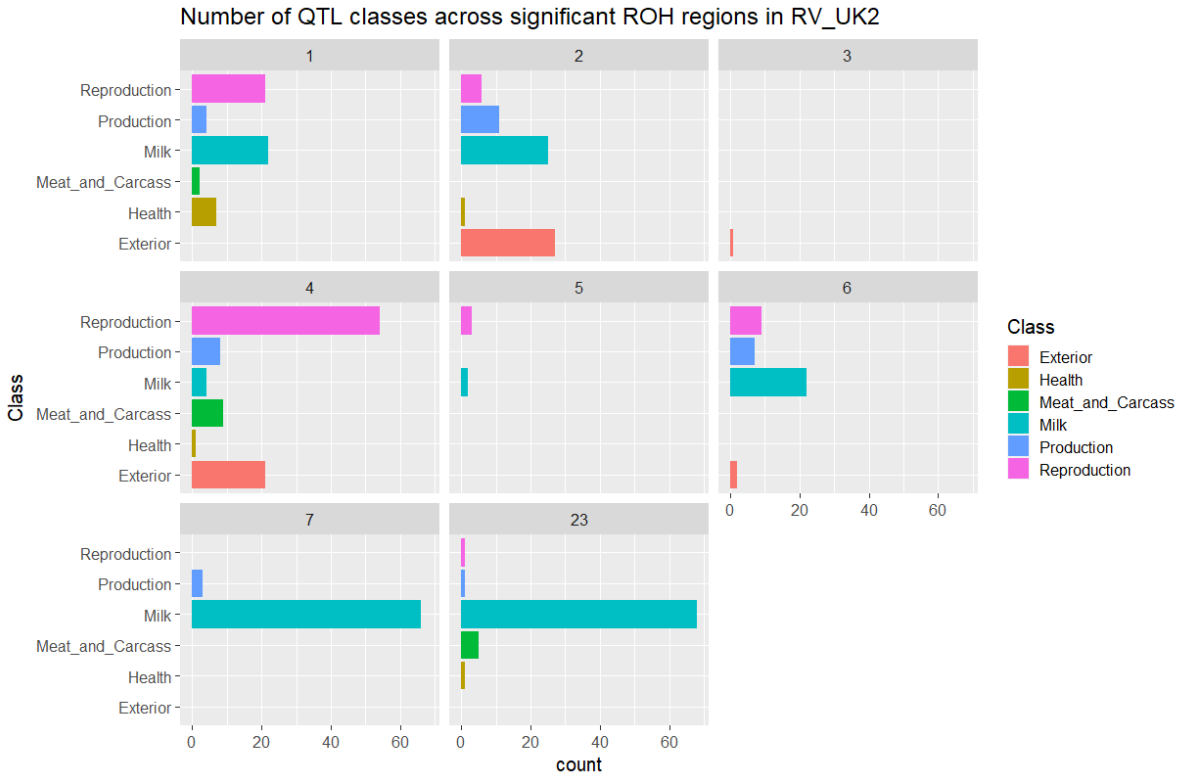


Table S2.12: Significant genes found in XP-EHH analysis for RV_UK1 vs RV_GER_STA and RV_UK2 vs RV_GER_STA

Population	Chromosome	Start	End	Genes
RV_UK1	1	27425486	28113771	TRNAC-ACA-4, TRIML1, TRIML2, ZFP42
		29984786	30123983	SORBS2, PDLIM3, CCDC110, C1H4orf47, UFSP2, ANKRD37, LRP2BP, SNX25, CFAP97, SLC25A4, HELT
		132745375	132868674	GNB4, MFN1, ZNF639, KCNMB3, PIK3CA, LOC112587440, ZMAT3, TRNAW-CCA-7, TRNAG-UCC-4, KCNMB2
		151230397	151336861	PPM1L, ARL14, KPNA4
		187354770	187861398	TFF2, TFF1, TMPRSS3, UBASH3A, RSPH1, SLC37A1, PDE9A, WDR4, NDUFV3, PKNOX1, CBS, U2AF1, CRYAA
	2	36017854	37531045	PEX6, GNMT, CNPY3, PTCRA, C2H6orf226, RPL7L1, BICRAL, TBCC, PRPH2, UBR2, TRNAG-UCC-7, TRERF1, MRPS10, GUCA1B, GUCA1A, TRNAW-CCA-13, C2H6orf132, TAF8, CCND3, BYSL, MED20, USP49, PRICKLE4, TOMM6, FRS3, PGC, TFEB, MDFI, FOXP4, NCR2, TREM1, TREML2, TREM2, TRNAG-UCC-8, TREML1, NFYA, OARD1, APOBEC2, TSPO2, UNC5CL
		117087325	118055365	NCKAP5, LYPD1, GPR39, SLC35F5
		152419212	152773351	ERBB4
	3	3930106	3955743	-
		15387323	15446777	SMARCD2, PSMC5, FTSJ3, DDX42, CCDC47, STRADA, LIMD2, MAP3K3, DCAF7, KCNH6, TANC2
		142275429	142464858	TUT7
		154753980	155449850	CYLC2
		169852072	169979708	TLR4
	4	8971634	9212951	TNRC6B, FAM83F, GRAP2, ENTHD1, CACNA1I, RPS19BP1, ATF4, MIEF1, TRNAG-CCC-33, MGAT3, TRNAG-CCC-34, TAB1, SYNGR1
		84453485	84625128	TMEM117, NELL2

	152915324	-	KCNMA1, DLG5, POLR3A, RPS24
5	123444264	123938178	TRNAR-CCU-11, LTO1, CCND1, FGF19, FGF4, FGF3, ANO1, FADD, PPFIA1, CTTN
6	81971515	82153482	ATG4C, DOCK7, ANGPTL3
	109602412	111177583	CLSPN, C6H1orf216, PSMB2, TFAP2E, NCDN, KIAA0319L, ZMYM4, SFPQ, ZMYM1, ZMYM6, TMEM35B, DLGAP3, SMIM12, GJA4, GJB3, GJB4, GJB5, C6H1orf94, CSMD2
7	78966472	79380792	TRNAC-GCA-116
	115977766	-	QRFPR, TNIP3, TRNAY-AUA-8, NDNF
8	100183355	100279260	CHRM2
9	56467422	56534117	-
	69973920	70459326	TRIM58, GCSAML, NLRP3, ZNF496
	81392762	81499515	CSNK1G3
	83180816	83312296	-
11	90660253	90691830	PARP16, TRNAG-CCC-98, CILP, CLPX, PDCD7, UBAP1L, KBTBD13, RASL12, SLC51B, MTFMT, SPG21, TRNAC-GCA-171, ANKDD1A, PTGER2
12	76869798	76933547	-
13	2064219	2932876	NAXD, RAB20, COL4A2, COL4A1, IRS2, MYO16, TRNAY-GUA-17
15	44970558	45102329	JPH1, TRNAE-UUC-100, LY96, TRNAC-GCA-211, TMEM70, ELOC, UBE2W, STAU2
	53620530	54056901	YTHDF3, TTPA
	61672639	-	SNTG1
16	26803089	27164245	LIN7C, LGR4, CCDC34, BBOX1, FIBIN
	44986300	-	TEAD1
17	23746675	24674935	TRNAC-GCA-224, TMEM132C, TRNAS-GGA-40, SLC15A4, GLT1D1, TRNAE-UUC-106
	50280821	50688764	-
18	41136026	41466603	ZNF536, TSHZ3
	65245471	65268241	ZNF671
19	56539547	57486561	RETREG1, ZNF622, MARCH11, FBXL7
20	30917827	-	-
	53507848	54572685	SV2B, SLCO3A1
22	20011293	20178300	PIEZO2, NAPG, APCDD1, VAPA, TXNDC2, RAB31

	24	17141512	17217979	XPO6, GSG1L, KIAA0556, TRNAG-UCC-70, TRNAG-UCC-71
RV_UK2	1	34633398	34884805	TENM3
		187578102	187861398	SLC37A1, PDE9A, WDR4, NDUFV3, PKNOX1, CBS, U2AF1, CRYAA
		189237181	189278514	TSPEAR, UBE2G2, SUMO3, PTTG1IP, ITGB2, FAM207A
	2	81785127	82330005	SCN7A, SCN9A
		108719654	-	-
		109201735	109844693	LRP1B
	3	50923062	-	USP32, C3H17orf64, APPBP2, PPM1D
		132330123	133177884	BMP1, PHYHIP, POLR3D, PIWIL2, SLC39A14, PPP3CC, SORBS3, PDLIM2, C3H8orf58, CCAR2, EGR3, PEBP4, RHOBTB2, CHMP7, R3HCC1, LOXL2, ENTPD4, SLC25A37
		142170136	142735671	C3H9orf153, ISCA1, TUT7, GAS1
		154753980	155514138	CYLC2
	4	25326050	25436574	MGP, ERP27, ARHGDIB, PDE6H, RERG, PTPRO
		96290081	96340164	CEP83, PLXNC1
		128531484	128885769	GPR137B, ERO1B, EDARADD, LGALS8, HEATR1, ACTN2, MTR, TRNAE-UUC-37
		129278440	130812166	RYR2, ZP4
	5	20720027	21220496	ABL2, TOR3A, FAM20B, RALGPS2, ANGPTL1, TEX35, RASAL2, TRNAC-GCA-91
		24155057	24746031	TNR, KIAA0040, TNN, MRPS14, CACYBP, RABGAP1L
	6	97691915	97975136	SLC5A9, TRABD2B
	7	12189331	12701213	CLNK, ZNF518B, WDR1, SLC2A9, DRD5, OTOP1, TMEM128, LYAR, ZBTB49, NSG1, STX18
		79432960	79558879	-
		102600603	102880560	STPG2
		109507742	109871388	CCSER1
	8	30432198	30553964	DNAH11, CDCA7L, RAPGEF5
		98863064	99047114	STRA8, CNOT4, TRNAR-GCG-2, NUP205, STMP1, SLC13A4, FAM180A
	9	2246697	2428110	PJA2, FER
		90516686	90546137	GNG7, DIRAS1, SLC39A3, SGTA, THOP1, ZNF554, TLE6, TLE2, AES, GNA11, TRNAG-

			CCC-90, GNA15, S1PR4, NCLN, CELF5
10	43016086	-	RNGTT
	67437756	-	-
	68087621	68110776	HDAC2, HS3ST5
	72440018	72687011	MCM9, FAM184A, MAN1A1
11	90784888	90819545	PDCD7, UBAP1L, KBTBD13, RASL12, SLC51B, MTFMT, SPG21, TRNAC-GCA-171, ANKDD1A, PTGER2, TXNDC16
	94642838	95043496	AGGF1, CRHBP, S100Z, F2RL1, F2R, IQGAP2, F2RL2, SV2C
	96429678	96798725	DRD1, SFXN1
12	18833396	18887222	CRIM1, TRNAC-ACA-69, FEZ2, VIT
	24154753	24358605	PKDCC, EML4
	24765361	24795830	KCNG3, MTA3
13	40601425	40625833	TBC1D4
	80309628	80336316	TRNAG-UCC-44
15	3353707	3976502	RALYL
16	28014281	28234371	ANO3, GDPD4, MYO7A, CAPN5, OMP, B3GNT6
17	19424207	19565039	RNF34, KDM2B, ORAI1, MORN3, TMEM120B, RHOF, SETD1B, HPD, PSMD9, WDR66, BCL7A, MLXIP
	20509691	20834314	CCDC62, HIP1R, VPS37B, TRNAC-ACA-93, ABCB9, OGFOD2, ARL6IP4, PITPNM2, MPHOSPH9, C17H12orf65, CDK2AP1, SBNO1, KMT5A, RILPL2, SNRNP35, RILPL1, TMED2, DDX55, EIF2B1, GTF2H3, TCTN2
19	3809158	4316279	STK10, EFCAB9, UBTD2, SH3PXD2B, NEURL1B, DUSP1, ERGIC1, RPL26L1
	40711638	40954154	NPR3, SUB1, ZFR, MTMR12
21	6275741	6782333	STT3B, OSBPL10, TRNAG-CCC-155, TRNAG-CCC-156, TRNAG-CCC-157, GPD1L, CMTM8, CMTM7, CMTM6
23	9415256	9562200	PAPSS2, TRNAC-ACA-119, ATAD1, PTEN
	20429959	21882369	GOT1, NKX2-3, SLC25A28, ENTPD7, TRNAG-GCC-26, CUTC, ABCC2, DNMBP, CPN1, ERLIN1, CHUK, CWF19L1, BLOC1S2, PKD2L1, SCD, TRNAC-ACA-121, WNT8B, SEC31B, NDUFB8, TRNAC-

			GCA-269, HIF1AN, PAX2, TRNAG-UCC-68, SLF2, SEMA4G, MRPL43, TWNK, LZTS2, PDZD7, SFXN3, KAZALD1, TLX1, LBX1, BTRC
24	4774595	4981253	RSPH10B, CCZ1, LMTK2, BHLHA15, TECPR1, BRI3, BAIAP2L1, NPTX2, TMEM130

Figure S3.1: Cross validation results from ADMIXTURE with K=6 being the most efficient model.

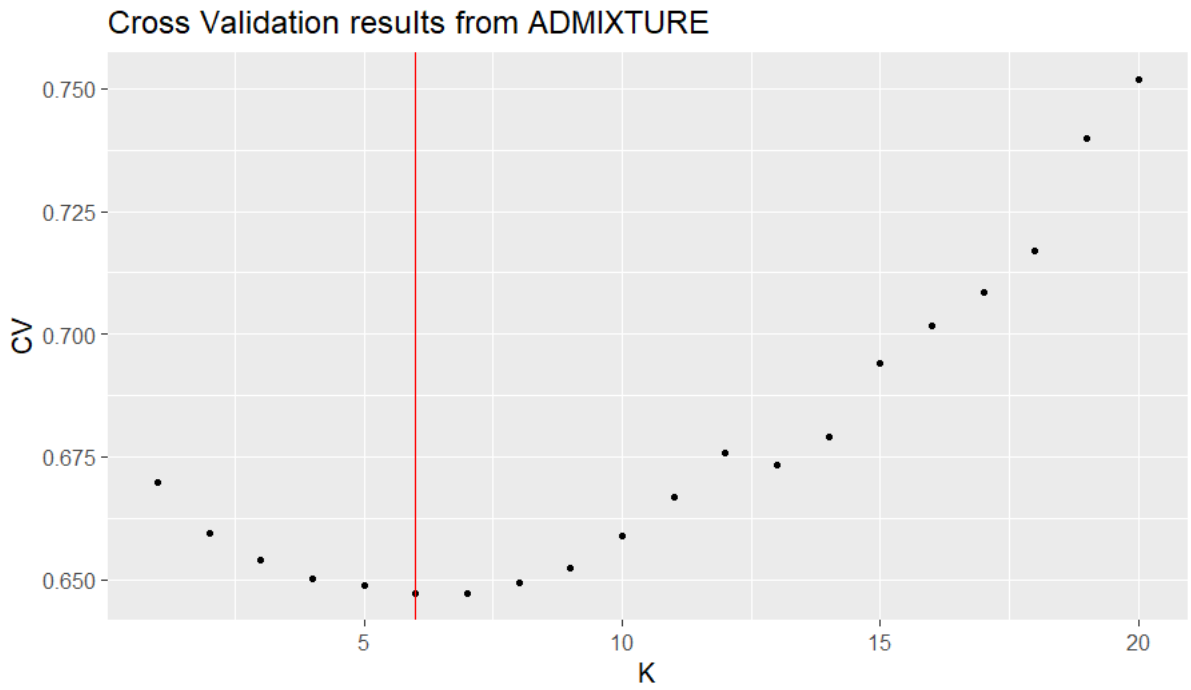


Table S3.2: Tukey's HSD post hoc test showing p-values for pairwise comparisons between murrah populations ROH lengths (above diagonal) and FROH (below diagonal). Significant results are shown in bold.

	RV_BRA	**RV_BUL_VAR**	**RV_COL**	**RV_IND**	**RV_PH_BUL**	**RV_PH_IND**
RV_BRA	-	0.199	**0.000**	**0.000**	**0.000**	**0.000**
RV_BUL_VAR	0.222	-	**0.000**	**0.000**	**0.000**	**0.000**
RV_COL	**0.008**	0.355	-	0.791	1.000	0.997
RV_IND	**0.000**	**0.003**	0.942	-	0.476	0.190
RV_PH_BUL	**0.000**	**0.036**	1.000	0.994	-	1.000
RV_PH_IND	**0.000**	**0.000**	0.853	1.000	0.979	-

Table S3.3: Locations of significant and divergent ROHs, the genes found within ROHs, and overlapping QTLs. Significance determined by SNPs exceeding the top 1% threshold, and divergent runs identified by regions of high FST scores.

Population	Chromosome	Start Position	End Position	Genes	QTL Traits
RV_BRA	1	47135280	49099346	OLIG2, PAXBP1, SYNJ1, CFAP298, EVA1C, URB1, MRAP, MIS18A, HUNK, SCAF4, SOD1, TIAM1	Luteal activity, Dry matter intake, Residual feed intake, Body weight (yearling), Carcass weight, Milk iron content,
	1	61373189	61984973	TRNAC-GCA	Body weight gain
	2	133666865	137040288	TMEFF2, TRNAS-GGA	Milk kappa-casein percentage, Bovine tuberculosis susceptibility
	9	58046218	62175428	NDFIP1, GNPDA1, RNF14, PCDH12, DELE1, PCDH1, TRNAC-GCA, ARAP3, FCHSD1, RELL2, HDAC3, DIAPH1, PCDHGC4, TAF7, PCDHB15, PCDHB10, PCDHB8, PCDHB7, PCDHB5, PCDHB4, PCDHB1, PCDHAC1, ZMAT2, HARS2, HARS, DND1, WDR55, IK, NDUFA2, TMCO6, CD14, SLC35A4, APBB3, SRA1, EIF4EBP3, SLC4A9, HBEGF, PFDN1, CYSTM1, IGIP, PURA, NRG2, PSD2, CXXC5, UBE2D2, TMEM173, SMIM33, ECSCR, DNAJC18, SPATA24, PROB1, MZB1, SLC23A1, PAIP2, MATR3, SIL1, CTNNA1, LRRTM2, HSPA9, ETF1, EGR1, REEP2, KDM3B, FAM53C, CDC25C, GFRA3, TRNAE-UUC, CDC23, KIF20A, BRD8, NME5, WNT8A, FAM13B, TRNAG-UCC, PKD2L2, MYOT, HNRNPA0, KLHL3	Milk kappa-casein percentage, Milk unglycosylated kappa-casein percentage, Milk protein percentage, Stillbirth, Calving ease, Foot angle, Udder attachment, Rear leg placement - rear view, PTA type, Calving ease (maternal), Stature, Body depth, Strength, Net merit, Milk fat percentage, Daughter pregnancy rate, Inseminations per conception, Milk protein yield, Milk yield, Milk fat yield, Calving to conception interval, Calving interval, Gestation length, Body weight (yearling), Shear force, Conception rate
	12	46848107	48349500	RPIA, TRNAG-UCC, EIF2AK3, TEX37, FOXI3, THNSL2, FABP1, SMYD1, KRCC1, CD8B, CD8A, RMND5A, RNF103, CHMP3, KDM3A, REEP1	Milk riboflavin content, Body weight gain, Age at puberty, Hoof and leg disorders, Body weight (mature), Somatic cell score
RV_BUL_VAR	4	45440372	47480462	MYH9, RBFOX2, MB, RASD2, MCM5, HMOX1, TOM1, HMGXB4, ISX, TRNAR-CCU	Body weight gain, Body weight (yearling), Body weight (birth), Udder depth, Somatic cell score, Stillbirth, Net merit, Milk protein yield, Rump width, Feet and leg conformation, Calving ease,

				Rear leg placement - rear view, Teat length, Rear leg placement - side view, Length of productive life, Calving ease (maternal), Daughter pregnancy rate, Foot angle, Teat placement – front, Strength, Inseminations per conception, PTA type, Udder attachment, Body depth, Stature, Milk fat percentage, Milk fat yield, Body weight (weaning), Milk protein percentage, Interval to first estrus after calving, Shear force
	14	32123680	37890053	ATRN, GFRA4, ADAM33, SIGLEC1, HSPA12B, SPEF1, CENPB, CDC25B, AP5S1, MAVS, PANK2, RNF24, TRNAE-UUC, SMOX, ADRA1D, TRNAC-ACA, HAO1, BMP2, TRNAC-GCA, FERMT1, LRRN4, CRLS1, MCM8, TRMT6, CHGB, SHLD1, GPCPD1, PROKR2, CDS2, PCNA, TMEM230, SLC23A2, RASSF2, PRND, PRNP, ZMYND11, DIP2C, LARP4B, TRNAG-CCC, GTPBP4, WDR37, ADARB2	Eye area pigmentation Milk yield Inhibin level, Average daily gain, Metabolic body weight, Interval to first estrus after calving, Net merit, Rear leg placement - side view, Milk fat yield, Rear leg placement - rear view, Stature, Body depth, Foot angle, Calving ease, Milk protein yield, Strength, Body weight gain, Milk rennet coagulation time, Milk potassium content, PTA type, Udder attachment, Udder height, Udder depth, Rump width, Milk iron content, Milk fat percentage, Milk protein percentage
	23	37731334	45481611	RAB11FIP2, FAM204A, PRLHR, CACUL1, TRNAG-CCC, NANOS1, EIF3A, FAM45A, SFXN4, PRDX3, GRK5, TRNAC-GCA, RGS10, TIAL1, TRNAG-CCC, BAG3, INPP5F, MCMBP, SEC23IP, PLPP4, WDR11, FGFR2, ATE1, NSMCE4A, TACC2, BTBD16, PLEKHA1, HTRA1, DMBT1, CUZD1, TRNAQ-UUG, FAM24A, PSTK, IKZF5, ACADSB, HMX3, HMX2, BUB3, GPR26, CPXM2, CHST15, OAT, NKX1-2, LHPP, FAM53B, EEF1AKMT2, ABRAXAS2, ZRANB1, CTBP2, TEX36, EDRF1, UROS, BCCIP, DHX32, TRNAE-CUC, FANK1, ADAM12	Milk yield, M. paratuberculosis susceptibility, Milk yield, Milk linoleic acid content, Twinning, Stature, Udder depth, Foot angle, Rear leg placement - side view, Calving ease, Calving ease (maternal), Rear leg placement - rear view, Stillbirth (maternal), Length of productive life, Stillbirth, Somatic cell score, Feet and leg conformation, Udder attachment, Strength, Rump width, Net

					merit, Udder height, PTA type, Body depth, Milk protein percentage, Milk stearic acid content, Milk protein yield, Milk fat yield, Milk fat percentage, Teat length, Body weight (yearling), Age at puberty, Milk C16 index, Milk palmitoleic acid content, Average daily gain, Lean meat yield, Milk protein percentage, Angularity, Milk protein yield, Milk kappa-casein percentage, Milk glycosylated kappa-casein percentage, Daughter pregnancy rate, Teat thickness, Udder structure
RV_COL	2	45662839	48841312	LRRC1, MLIP, TRNAW-CCA, TRNAC-GCA, TINAG, FAM83B, HCRTR2, GFRAL, HMGCLL1, BMP5, COL21A1, DST	Bovine tuberculosis susceptibility, Milk glycerophosphocholine content, Body weight (mature), Stillbirth (maternal), Feet and leg conformation, Length of productive life, Udder depth, Udder attachment, Somatic cell score, Daughter pregnancy rate, PTA type, Lean meat yield, Fertility treatments, Bovine leukemia virus susceptibility, Net merit, Body depth, Foot angle, Udder cleft, Stature, Teat placement – front, Strength, Rump width, Udder height, Calving ease (maternal), Milk kappa-casein percentage, Dairy form, Stillbirth, Rear leg placement - side view, Body weight gain, Teat length, Calving ease, Milk glycosylated kappa-casein percentage
	5	24129102	25636234	TNR, KIAA0040, TNN, MRPS14, CACYBP, RABGAP1L, TRNAG-ACC, TRNAC-ACA, GPR52	Milk glycosylated kappa-casein percentage, Milk potassium content, Milk alpha-lactalbumin percentage, Interval to first estrus after calving, Body weight (yearling)

	12	73764413	73764413	-	-
RV_IND_1	1	92570342	98163171	TRNAR-GCG, ALCAM, CBLB, CCDC54, BBX, CD47, IFT57, HHLA2, MYH15, CIP2A	Bovine tuberculosis susceptibility, Milk kappa-casein percentage, Milk glycosylated kappa-casein percentage, Age at puberty, Bovine respiratory disease susceptibility, Milk beta-lactoglobulin percentage, Milk yield, Milk protein yield, Milk fat yield, Lean meat yield, Average daily gain, Body weight (yearling), Body weight gain, Interval from first to last insemination, Inseminations per conception, Udder height, Feet and leg conformation, Stillbirth (maternal), Calving ease (maternal), Stature, Foot angle, Udder attachment, Net merit, PTA type, Length of productive life, Stillbirth, Udder depth, Average daily gain, Milk riboflavin content
	4	7280087	8951770	CENPM, TNFRSF13C, SHISA8, SREBF2, CCDC134, MEI1, SNU13, XRCC6, DESI1, PMM1, CSDC2, TRNAM-CAU, POLR3H, ACO2, PHF5A, TOB2, TEF, ZC3H7B, TRNAG-UCC, RANGAP1, CHADL, L3MBTL2, EP300, RBX1, XPNPEP3, DNAJB7, ST13, SLC25A17, MCHR1, MRTFA, SGSM3, ADSL, TNRC6B, FAM83F, GRAP2, ENTHD1	Length of productive life, Strength, Net merit, Foot angle, Rear leg placement - side view, Calving ease, Milk fat percentage, Dairy form, Feet and leg conformation, Daughter pregnancy rate, Milk protein percentage, Calving ease (maternal), Stillbirth, Teat length, Rear leg placement - rear view, Milk iron content, Body weight (yearling), Maturity rate, Sire conception rate, Milk protein yield, Milk yield, Milk fat yield, Marbling score, Fat thickness at the 12th rib
	6	19447997	21540090	BNIPL, PRUNE1, MINDY1, ANXA9, CERS2, SETDB1, ARNT, CTSK, CTSS, HORMAD1, GOLPH3L, ENSA, MCL1, ADAMTSL4, TRNAS-GGA, TRNAG-UCC, ECM1, TARS2, TRNAE-UUC, RPRD2, PRPF3, MRPS21, CIART, APH1A, CA14, ANP32E,	Lean meat yield, Rump conformation, Milk yield, Milk protein percentage, Pregnancy rate, Body weight gain

				PLEKHO1, VPS45, TRNAW-CCA, OTUD7B, MTMR11, SF3B4, SV2A, BOLA1, TRNAE-CUC, TRNAN-GUU, TRNAH-GUG, TRNAG-CCC, TRNAE-UUC, TRNAE-UUC, TRNAN-GUU, TRNAH-GUG, TRNAN-GUU, TRNAV-UAC, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAV-AAC, TRNAV-UAC, TRNAN-GUU, TRNAV-CAC, TRNAG-CCC, TRNAV-AAC, TRNAV-CAC, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAE-UUC, TRNAG-CCC, TRNAV-AAC, TRNAH-GUG, TRNAG-CCC, TRNAV-AAC, TRNAV-UAC, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAV-AAC, TRNAV-AAC, TRNAH-GUG, TRNAG-CCC, TRNAV-AAC, TRNAV-UAC, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAV-AAC, TRNAV-AAC, TRNAH-GUG, TRNAG-CCC, TRNAE-UUC, TRNAV-CAC, TRNAN-GUU, TRNAQ-CUG, TRNAV-AAC, TRNAV-UAC, TRNAQ-CUG, TRNAN-AUU, TRNAV-CAC, TRNAG-CCC, TRNAV-AAC, TRNAV-AAC, TRNAH-GUG, TRNAG-CCC, TRNAE-UUC, TRNAV-CAC, TRNAN-GUU, TRNAQ-CUG, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAE-UUC, TRNAG-CCC, TRNAH-GUG, TRNAV-AAC, TRNAV-AAC, TRNAG-CCC, TRNAV-CAC, TRNAN-GUU, TRNAQ-CUG, TRNAV-UAC, TRNAV-AAC, TRNAG-CCC, TRNAV-CAC, TRNAN-GUU, TRNAQ-CUG, TRNAV-UAC, TRNAN-GUU, TRNAV-AAC, TRNAN-GUU, TRNAH-GUG, TRNAN-GUU, TRNAV-AAC, TRNAV-CAC, TRNAN-GUU, TRNAH-GUG, TRNAG-CCC, TRNAE-UUC, TRNAH-GUG, TRNAN-GUU, TRNAH-GUG, TRNAN-GUU, TRNAV-CAC, TRNAV-AAC, TRNAN-GUU, TRNAH-GUG, TRNAQ-CUG, TRNAN-GUU, TRNAV-CAC, TRNAE-UUC, TRNAN-GUU, TRNAN-GUU, TRNAK-CUU, TRNAH-GUG, TRNAG-UCC, TRNAE-CUC, HJV, TXNIP, POLR3GL, ANKRD34A, LIX1L, RBM8A, PEX11B, ITGA10, ANKRD35, PIAS3, NUDT17, POLR3C, RNF115, CD160, PDZK1	
	6	24093261	26523748	SPAG17, WDR3, GDAP2, TENT5C, MAN1A2, TRNAE-UUC, VTCN1, TRIM45, TTF2, CD101, PTGFRN, CD2, IGSF3	Length of productive life, Dairy capacity composite index, Milk fat percentage, Milk protein yield, Milk casein percentage, Milk protein yield, Milk fat percentage, Milk alpha-S2-casein percentage, Milk phosphorylated alpha-

					S2-casein percentage, Bovine leukemia virus susceptibility, Tenderness score, Body weight (yearling), Bovine respiratory disease susceptibility
RV_IND_2	2	45919131	47487099	MLIP, TRNAW-CCA, TRNAC-GCA, TINAG, FAM83B, HCRTR2, GFRAL, HMGCLL1	Milk glycerophosphocholine content, Body weight (mature), Stillbirth (maternal), Feet and leg conformation, Length of productive life, Udder depth, Udder attachment, Somatic cell score, Daughter pregnancy rate, PTA type, Lean meat yield
	9	58362401	62412285	ARAP3, FCHSD1, RELL2, HDAC3, DIAPH1, PCDHGC4, TAF7, PCDHB15, PCDHB10, PCDHB8, PCDHB7, PCDHB5, PCDHB4, PCDHB1, PCDHAC1, ZMAT2, HARS2, HARS, DND1, WDR55, IK, NDUFA2, TMCO6, CD14, SLC35A4, APBB3, SRA1, EIF4EBP3, SLC4A9, HBEGF, PFDN1, CYSTM1, IGIP, PURA, NRG2, PSD2, CXXC5, UBE2D2, TMEM173, SMIM33, ECSCR, DNAJC18, SPATA24, PROB1, MZB1, SLC23A1, PAIP2, MATR3, SIL1, CTNNA1, LRRTM2, HSPA9, ETF1, EGR1, REEP2, KDM3B, FAM53C, CDC25C, GFRA3, TRNAE-UUC, CDC23, KIF20A, BRD8, NME5, WNT8A, FAM13B, TRNAG-UCC, PKD2L2, MYOT, HNRNPA0, KLHL3, SPOCK1	Milk kappa-casein percentage, Milk unglycosylated kappa-casein percentage, Milk protein percentage, Stillbirth, Calving ease, Foot angle, Udder attachment , Rear leg placement - rear view, PTA type, Calving ease (maternal), Stature, Body depth, Strength, Net merit, Milk fat percentage, Inseminations per conception, Milk protein yield, Milk yield, Milk fat yield, Calving to conception interval, Calving interval, Gestation length, Body weight (yearling), Shear force, Conception rate, Length of productive life, Rump width, Udder depth, Daughter pregnancy rate, Somatic cell score
	14	77753047	80039684	BTBD3, TRNAC-ACA, JAG1, SLX4IP, MKKS	Body depth, Udder cleft, Stature, Daughter pregnancy rate, Stillbirth (maternal), PTA type, Rear leg placement - rear view, Dairy form, Milk yield, Teat length, M. paratuberculosis susceptibility, Milk fat yield, Shear force, Length of productive life, Udder depth, Rump width, Feet and leg conformation, Calving ease (maternal), Teat placement

					– front, Foot angle, Udder attachment, Strength, Teat placement – rear, Udder height, Rear leg placement - side view, Net merit, Somatic cell score, Rump width, Milk protein yield, Stillbirth, Conception rate
	17	28910234	30395257	ZNF84, ZNF26, ZNF605, GUCY1A1, GUCY1B1, ASIC5, TDO2, CTSO	Milk myristoleic acid content, Milk C14 index, Tick resistance
RV_PH_BUL	8	517175	3385871	TRNAY-AUA	Metabolic body weight, Calving ease, Luteal activity, Fecal egg count
	8	53912684	57223933	FOXP2, PPP1R3A, SMIM30, GPR85, BMT2, TMEM168, LSMEM1, IFRD1, ZNF277, DOCK4, TRNAC-ACA, IMMP2L, LRRN3	Milk fat yield, Average daily gain, Bovine tuberculosis susceptibility ,Milk kappa-casein percentage
	9	61665681	62869428	BRD8, NME5, WNT8A, FAM13B, TRNAG-UCC, PKD2L2, MYOT, HNRNPA0, KLHL3, SPOCK1	Conception rate, Calving ease, Udder attachment, Milk fat percentage, Stature, Strength, Length of productive life, Stillbirth, Rump width, Foot angle, Milk fat yield, Udder depth, Daughter pregnancy rate, Body depth, Net merit, PTA type, Somatic cell score, Calving ease (maternal), Milk protein percentage, Rear leg placement - rear view, Milk protein yield, Milk yield, Feet and leg conformation, M. paratuberculosis susceptibility
	20	41869077	42150353	TRPM1, MTMR10, FAN1, MPHOSPH10, MCEE	Interval to first estrus after calving
RV_PH_IND	2	125608406	127008452	TRNAL-CAA, TFCP2L1, CLASP1, NIFK, TSN, TRNAY-GUA	-
	3	55724516	67567277	TRIM25, DGKE, NOG, ANKFN1, TMEM100, MMD, HLF, STXBP4, TOM1L1, TRNAC-GCA, KIF2B, CA10, TRNAS-GGA, MFSD14B, ANXA10, DDX60, PALLD, CBR4, SH3RF1, NEK1, CLCN3, TRNAH-AUG, MFAP3L, AADAT, GALNTL6	-
	3	81810890	85696041	IZUMO3, TRNAW-CCA, ELAVL2, DMRTA1, CDKN2B, MTAP, IFNE	-
	3	115661427	117152262	PRUNE2, FOXB2, VPS13A, GNA14, CEP78	-
	4	54379396	59331565	GNPTAB, SYCP3, CHPT1, MYBPC1, SPIC, ARL1, UTP20, SLC5A8, ANO4, TRNAW-CCA, TRNAC-ACA, GAS2L3, NR1H4,	-

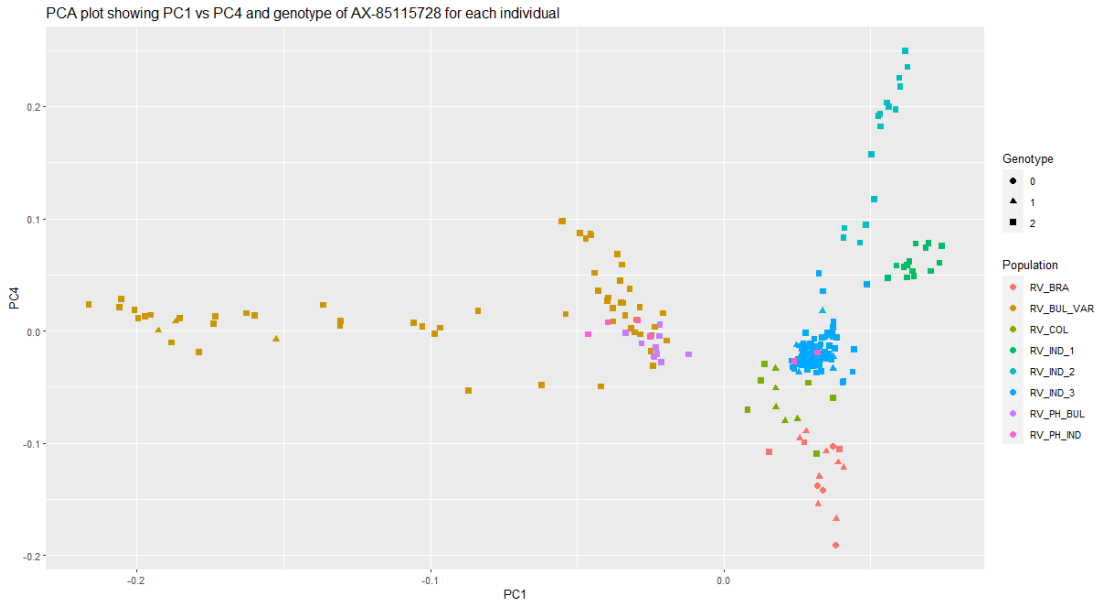
			SLC17A8, SCYL2, DEPDC4, ACTR6, UHRF1BP1L, ANKS1B, TRNAG-CCC, FAM71C, APAF1, IKBIP, SLC25A3, TMPO, TRNAW-CCA, TRNAD-GUC, NEDD1, CFAP54, CDK17	
	6	34302312	35846761	GPSM2, AKNAD1, STXBP3, FNDC7, PRPF38B, HENMT1, FAM102B, TRNAC-GCA, VAV3, NTNG1	-
	9	63761276	71039923	TGFBI, LECT2, IL9, SLC25A48, CXCL14, TRNAA-AGC, NEUROG1, TIFAB, PITX1, CATSPER3, PCBD2, TXNDC15, DDX46, CAMLG, SEC24A, SAR1B, JADE2, CDKN2AIPNL, UBE2B, CDKL3, PPP2CA, SKP1, TCF7, TRNAC-GCA, VDAC1, C9H5orf15, FSTL4, HSPA4, ZCCHC10, AFF4, LEAP2, GDF9, SHROOM1, SOWAHA, TRNAC-GCA, BTBD2, CSNK1G2, SCAMP4, ADAT3, ABHD17A, KLF16, REXO1, ATP8B3, ONECUT3, TCF3, MBD3, MEX3D, PLK5, ADAMTSL5, REEP6, PCSK4, APC2, RPS15, DAZAP1, GAMT, NDUFS7, TRNAN-GUU, TRNAF-GAA, EFNA2, CIRBP, MIDN, ATP5F1D, CBARP, STK11, SBNO2, GPX4, POLR2E, ARHGAP45, ABCA7, CNN2, TMEM259, GRIN3B, WDR18, ARID3A, KISS1R, TRNAW-CCA, R3HDM4, MED16, CFD, ELANE, PRTN3, AZU1, PLPPR3, PTBP1, MISP, PALM, PRSS57, FSTL3, RNF126, FGF22, POLRMT, HCN2, BSG, GZMM, CDC34, TPGS1, MADCAM1, ODF3L2, SHC2, C2CD4C, THEG, MIER2, PLPP2, PGBD2, TRNAE-CUC, TRNAL-CAA, ZNF692, ZNF672, SH3BP5L, LYPD8, TRIM58, GCSAML, NLRP3, ZNF496, TRIM52, RACK1, TRIM41, TRNAV-CAC, TRNAK-CUU, TRNAV-AAC, TRNAK-CUU, TRNAA-UGC, TRIM7, TRNAT-UGU, TRNAL-AAG	-
	15	29589045	30861497	CSMD3	-
	17	59617565	72540565	HHIP, ANAPC10, ABCE1, OTUD4, TRNAC-GCA, SMAD1, MMAA, ZNF827, LSM6, REELD1, SLC10A7, TTC29, POU4F2, EDNRA, TMEM184C, PRMT9, ARHGAP10, TRNAC-ACA, NR3C2, TRNAC-GCA, IQCM, TRNAE-UUC, DCLK2, LRBA, MAB21L2, RPS3A, SH3D19, PRSS48, FAM160A1, GATB, TRNAW-CCA, FBXW7, TMEM154, TIGD4, ARFIP1, FHDC1, TRIM2, MND1, TMEM131L, TLR2, RNF175, SFRP2, DCHS2, PLRG1, FGB, FGA, FGG, LRAT, RBM46, NPY2R, MAP9, TLL1, CPE, MSMO1, KLHL2	-

	18	38417846	39651090	PMFBP1, DHX38, TXNL4B, HP, DHODH, IST1, ZNF821, ATXN1L, AP1G1, PHLPP2, MARVELD3, TAT, CHST4, ZNF19, TRNAG-CCC, ZNF23, TRNAM-CAU, CALB2, CMTR2, HYDIN	-

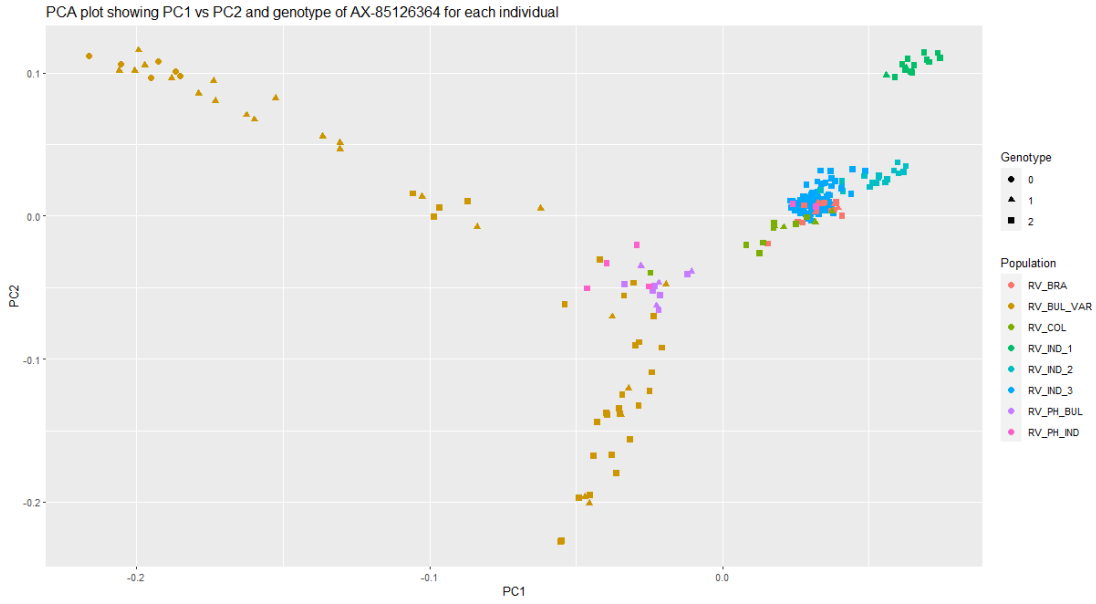
Table S3.4: Frequency Table of unique QTLs per murrah population

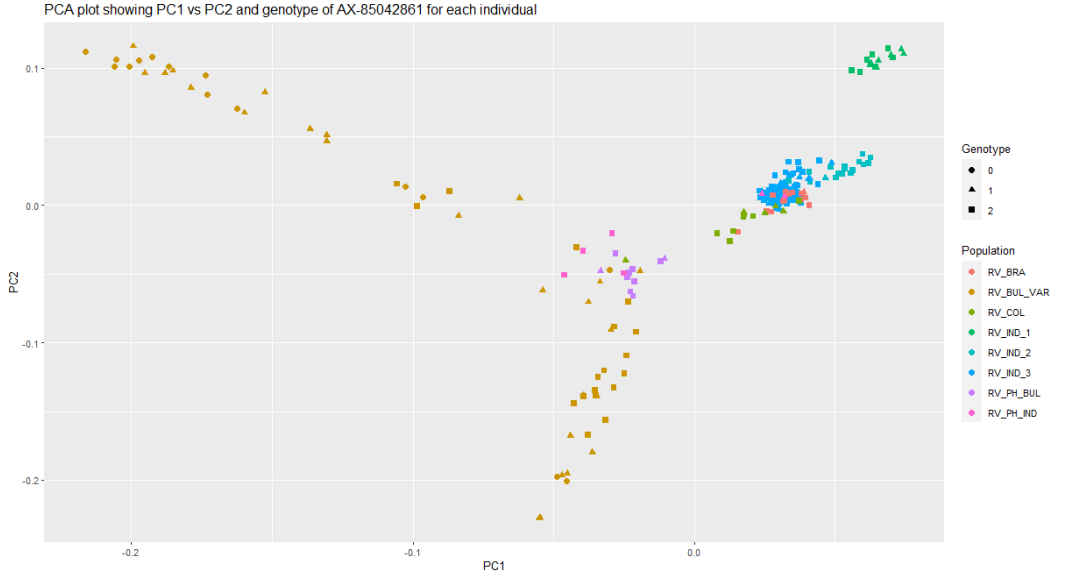
QTL Class	RV_BRA	RV_BUL_VAR	RV_COL	RV_IND_1	RV_IND_2	RV_IND_3	RV_PH_BUL	RV_PH_IND
Exterior	6	32	12	15	23	0	7	0
Health	2	3	3	4	6	0	4	0
Meat & Carcass	2	2	1	5	3	0	0	0
Milk	15	35	6	25	19	0	7	0
Production	11	24	8	14	15	0	7	0
Reproduction	12	17	7	13	17	0	8	0

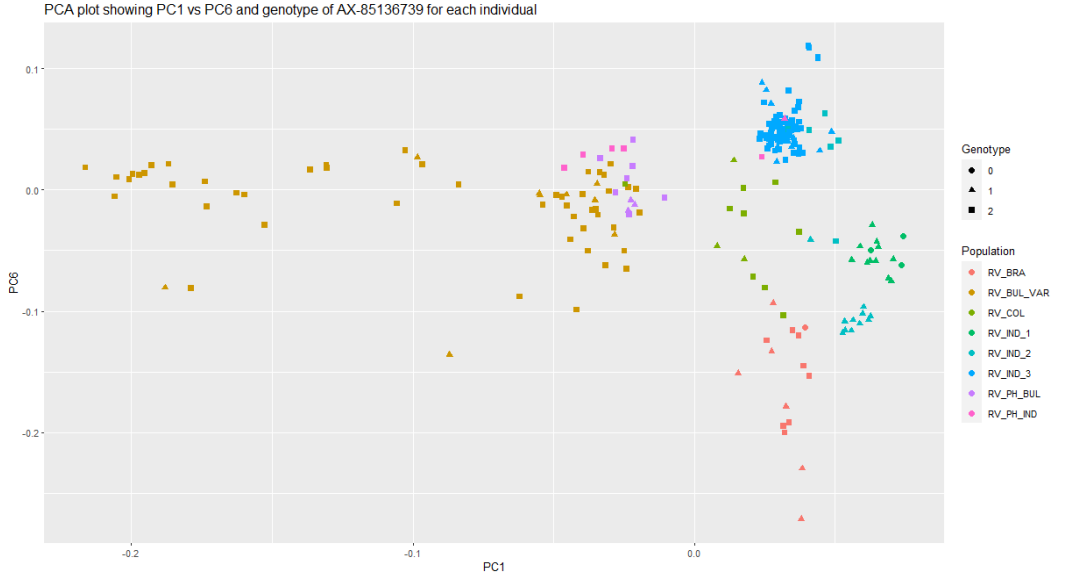
Figures S3.5 – S3.11: Distribution of genotypes for significant SNPs identified in PCAdapt analysis. In order: S2.6 AX-8511578 (BBU1: 61409049), S2.7 AX-85126364 (BBU11: 23558419), S2.8 AX-85042861 (BBU11: 48810245), S2.9 AX-85136739 (BBU12: 2572267), S2.10 AX-85112690 (BBU16: 10093932), S2.11 AX-85114321 (BBU16: 14825721), S2.12 AX-85092744 (BBU19: 62289241).



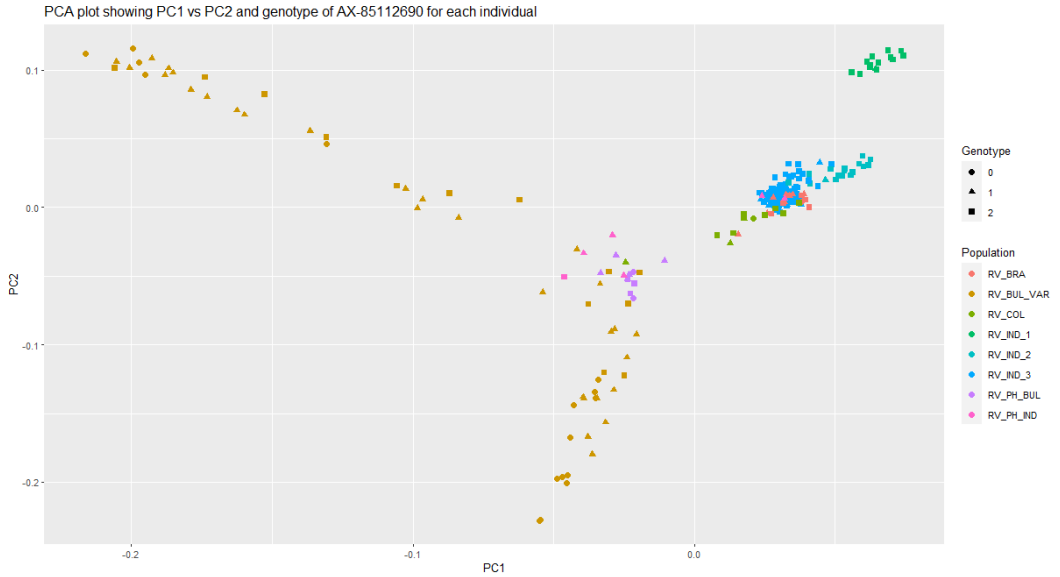
Supplementary Material

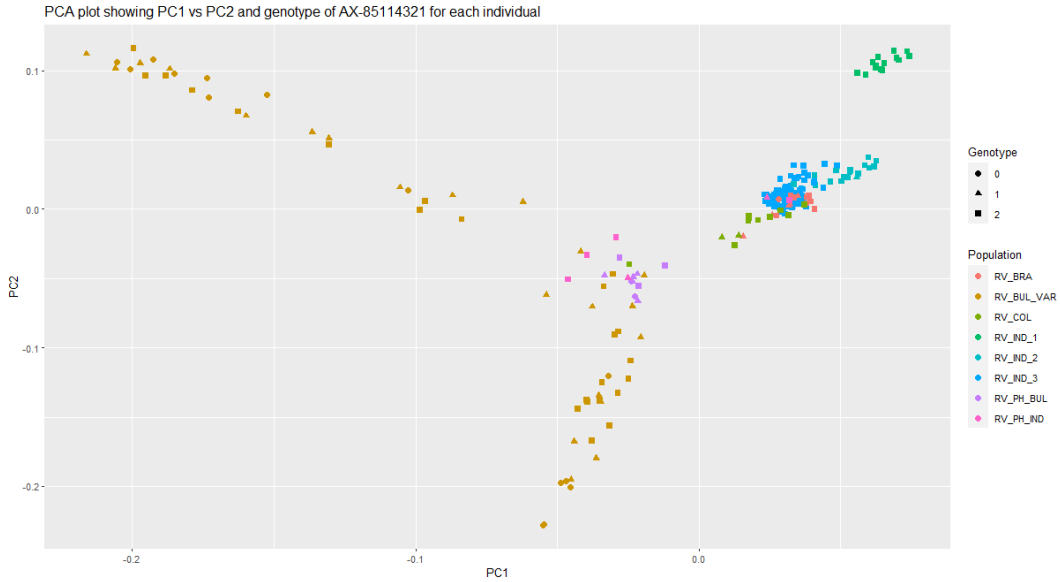


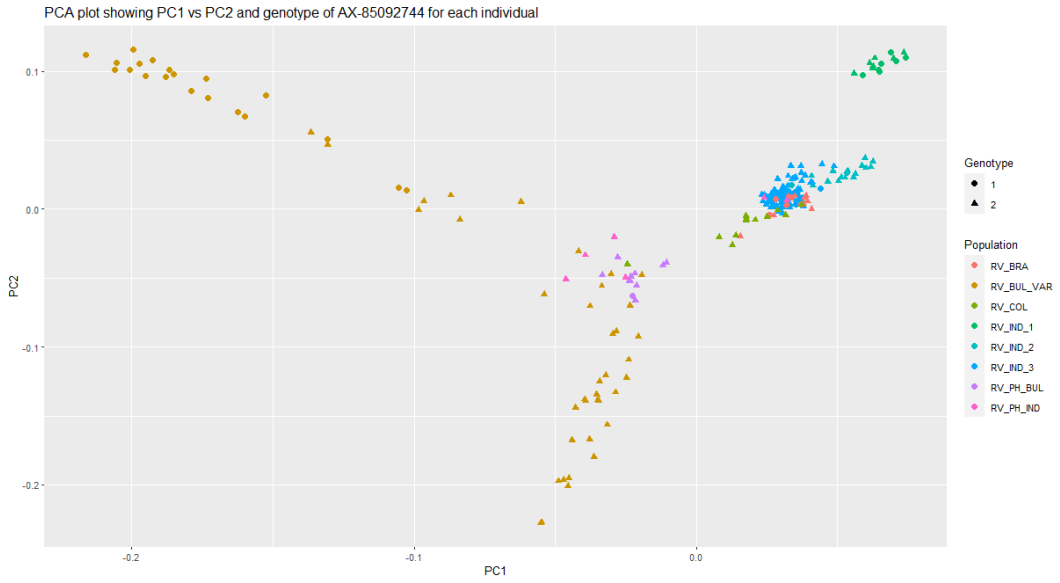




Supplementary Material

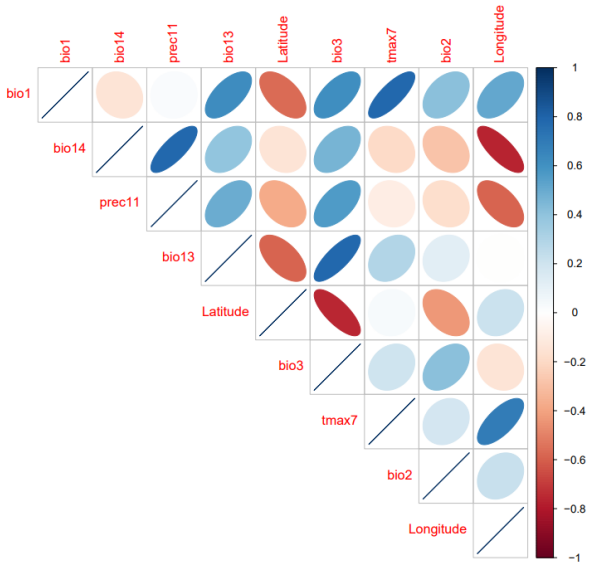






Supplementary Material

Figure S3.12 & S3.13: Correlation plots of r^2 values between environmental variables. S2.13 showing independent variables used.



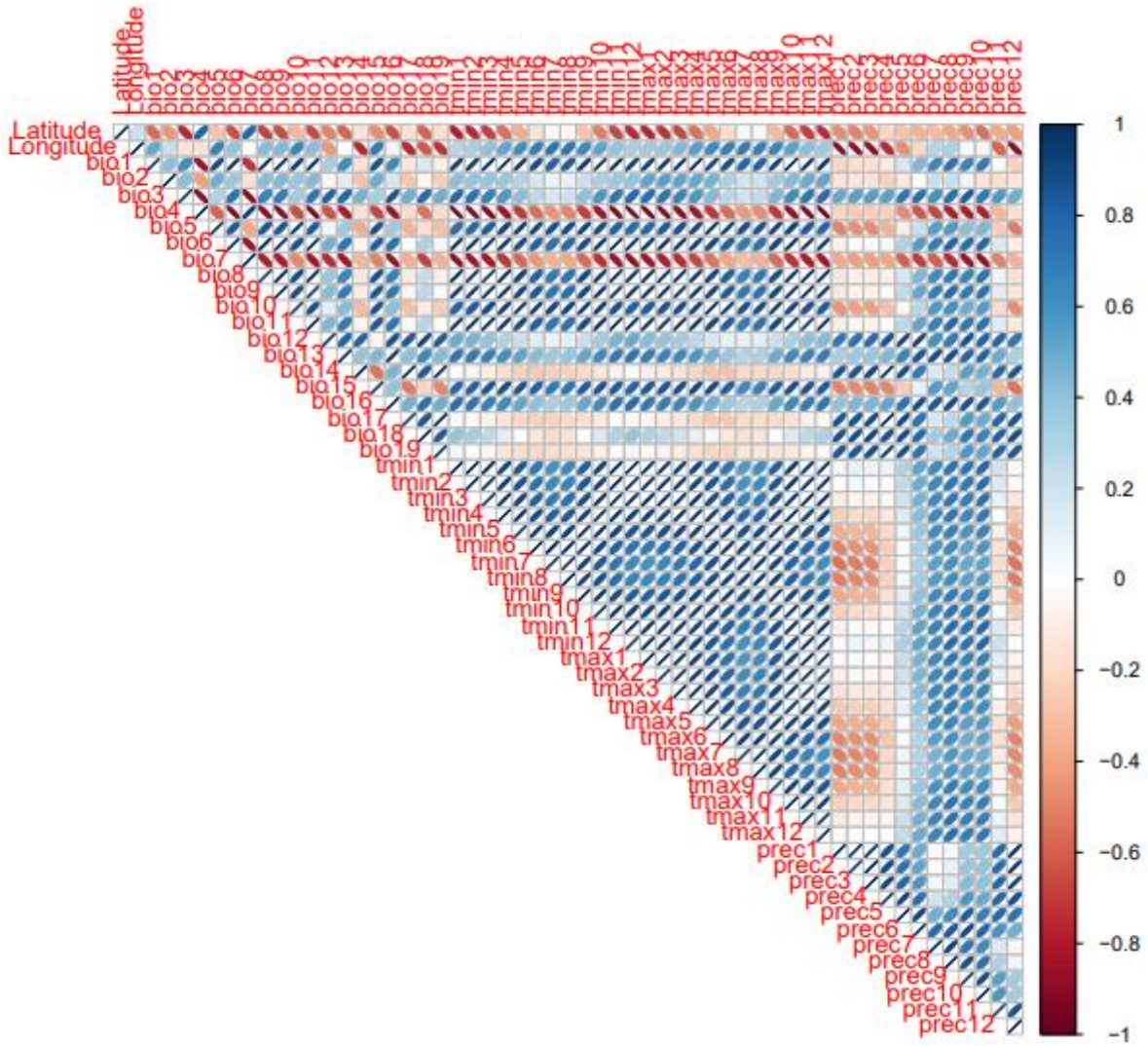


Table S3.14: R.Sambada Top SNPs per Environmental Variable:

SNP	Genotype	Environmental Variable	Q-value (G)	Chromosome	Position	Nearest Gene
AX-85106286	GG	Longitude	1.06E-06	9	31993994	TENM2
AX-85059527	GG		2.92E-06	13	40889619	TBC1D4
AX-85074562	CC		3.25E-06	6	14685275	MSTO1
AX-85094684	AA	Latitude	3.68E-11	10	83054152	TENT5A
AX-85142269	TT		3.68E-11	5	58652853	LYPLAL1
AX-85103603	TT		3.68E-11	4	82674961	ADAMTS20
AX-85092168	AA	bio1	2.29E-12	1	195528219	ERG*
AX-85094684	AA		5.98E-12	10	83054152	TENT5A
AX-85104855	AA		5.98E-12	14	46542199	RAB18
AX-85061534	TT	bio2	1.07E-05	2	144216220	NBEAL1*
AX-85098066	CC		1.34E-05	24	19052822	HS3ST4*
AX-85069096	GG		1.50E-05	2	144304554	NBEAL1*
AX-85104855	AA	bio3	7.30E-10	14	46542199	RAB18
AX-85105029	GG		1.40E-08	2	167810729	IRS1
AX-85094684	AA		1.40E-08	10	83054152	TENT5A
AX-85104855	AA	bio13	5.97E-11	14	46542199	RAB18
AX-85079551	TT		2.10E-08	19	31332072	ZNF131
AX-85066583	CC		5.31E-08	16	40734610	STK33
AX-85096756	CT	bio14	7.19E-08	15	59087609	MOS
AX-85056537	TT		8.35E-07	4	65746631	LRIG3
AX-85130232	CC		2.07E-06	1	153165367	MFSD1
AX-85045997	CC	tmax7	1.32E-09	2	187296850	IGSF21
AX-85106286	GG		1.40E-08	9	31993994	TENM2
AX-85103137	GG		5.55E-08	14	78164084	BTBD3
AX-85104083	CT	prec11	5.96E-06	17	24985973	TMEM132D*
AX-85104083	CC		1.64E-05	17	24985973	TMEM132D*
AX-85126469	GG		3.26E-05	21	25791681	CHL1

* SNP within gene

Table S3.15: XP-EHH selection signatures across each population

Population	Chromosome	Start Position	End Position	Genes
RV_BRA	1	133779976	133807071	-
	3	115919425	116229194	PRUNE2, FOXB2, VPS13A, GNA14
	6	63286893	-	-
	9	19058695	-	-
	10	39950115	40271533	NT5E, SNX14, SYNCRIP
		40858698	-	-
	11	57861601	-	CSNK1G1, PCLAF, TRIP4, ZNF609, OAZ2
	13	74125028	-	LRCH1, RUBCNL, LRRC63
	21	19523761	-	GRM7
RV_BUL_VAR	1	194336469	194565143	DYRK1A, KCNJ6
	5	70529743	70873701	-
	10	84863876	84958414	SH3BGRL2
		86179124	-	-
	23	43545229	-	OAT, NKX1-2, LHPP
RV_COL	1	110440719	110586482	STXBP5L, POLQ
	8	70999063	-	GSDME, MPP6
	9	26295138	-	EDIL3
		26460169	26672900	-
	13	29240287	29454906	-
	17	30963813	31151252	GLRB, GRIA2
RV_IND_1	3	108794711	-	PTAR1, TRNAY-AUA, C3H9orf135
		108913910	108966422	MAMDC2
	8	69083066	69226073	SNX10, SKAP2
	9	78347381	78367978	-
		78488459	78579231	-
	12	21145631	21300884	DHX57, MORN2, ARHGEF33, SOS1, CDKL4, MAP4K3
		21406441	21456965	-
		21569223	21649366	TMEM178A, THUMPD2
	13	65859076	66046219	SERTM1, RFXAP, SMAD9, ALG5, EXOSC8, SUPT20H
RV_IND_2	1	24166681	24505558	TUSC3, TRNAG-CCC, MSR1

	2	166439814	166710773	NYAP2
	7	6332726	-	TAPT1, PROM1
	12	87087448	-	RRM2, CYS1, KLF11, GRHL1, TAF1B
		87213093	-	-
	17	30161501	30215234	TDO2, CTSO
	21	42157322	-	-
RV_IND_3	1	47237918	47326479	OLIG2, C1H21orf62
		47447328	-	PAXBP1, SYNJ1
		47596180	-	CFAP298, EVA1C
		47715670	47750793	URB1, MRAP
	9	50452405	50525772	IL17B, PCYOX1L, GRPEL2, AFAP1L1, ABLIM3
	12	9926660	-	SEMA4F, M1AP, DOK1, LOXL3, HTRA2, AUP1, DQX1, TLX2, PCGF1
		10027482	10064161	LBX2, TTC31, CCDC142, MRPL53, MOGS, WBP1, INO80B, RTKN, WDR54, C12H2orf81, DCTN1
	15	31216570	31411859	CSMD3
RV_PH_BUL	5	28545216	-	KAZN
	8	87497483	88023496	SLC13A1, IQUB, NDUFA5, ASB15, LMOD2, WASL¸HYAL4, TMEM229A
		90305818	90381562	GRM8
	15	36559391	36700146	SAMD12
	16	82456798	82525348	GRIA4
	23	6014005	-	-
RV_PH_IND	1	112883396	113079871	HACD2, MYLK, CCDC14, ROPN1, KALRN
	3	116130568	116294369	VPS13A, GNA14
	4	3937165	4071981	WNT7B, ATXN10, FBLN1, RIBC2, SMC1B,
	9	77356325	77401670	-
	16	51482636	-	BSX, JHY, CRTAM, UBASH3B
	17	67333585	67363978	TRNAW-CCA, FBXW7
	18	9853920	10131682	CDH13, HSBP1, MLYCD, SLC38A8, MBTPS1, HSDL1, DNAAF1, TAF1C, ADAD2, KCNG4, WFDC1, ATP2C2, TLDC1

Table S3.16: XP-EHH Gene Ontology

Population	Gene Ontology Class	Biological Pathway	Description	Genes
RV_BRA	Process	GO:0030382	sperm mitochondrion organization	VPS13A
		GO:0046086	adenosine biosynthetic process	NT5E
	Function	GO:0010855	adenylate cyclase inhibitor activity	GRM7
		GO:0010854	adenylate cyclase regulator activity	GRM7
RV_BUL_VAR	Process	GO:0010121	arginine catabolic process to proline via ornithine	OAT
		GO:0019544	arginine catabolic process to glutamate	OAT
		GO:0019493	arginine catabolic process to proline	OAT
	Function	GO:0004587	ornithine-oxo-acid transaminase activity	OAT
		GO:0050155	ornithine(lysine) transaminase activity	OAT
RV_COL	Function	GO:0022835	transmitter-gated channel activity	GLRB, GRIA2
		GO:0022824	transmitter-gated ion channel activity	GLRB, GRIA2
		GO:0005230	extracellular ligand-gated ion channel activity	GLRB, GRIA2
		GO:0030594	neurotransmitter receptor activity	GLRB, GRIA2
	Component	GO:0016935	glycine-gated chloride channel complex	GLRB
RV_IND_2	Process	GO:0000083	regulation of transcription involved in G1/S transition of mitotic cell cycle	RRM2, KLF11
	Component	GO:0070860	RNA polymerase I core factor complex	TAF1B
RV_PH_BUL	Process	GO:0030041	actin filament polymerization	LMOD2, WASL
		GO:0007215	glutamate receptor signaling pathway	GRM8
		GO:0008154	actin polymerization or depolymerization	LMOD2, WASL
	Function	GO:0008066	glutamate receptor activity	GRM8, GRIA4
RV_PH_IND	Process	GO:0061180	mammary gland epithelium development	WNT7B, ATP2C2
		GO:0030324	lung development	WNT7B, FBXW7, DNAAF1

Supplementary Table S4.1: Number of markers retained by LD pruning in each dataset for SNP array and WGS.

LD Pruning Target	**Species**	**r^2 Threshold**						
		0.01	0.05	0.1	0.2	0.5	0.8	1
ALL	SNP	944	2739	5265	9952	27891	38397	40695
	WGS	1177	2826	7189	20656	53738	67457	72434
RIVER	SNP	1213	3780	9369	19847	33406	38178	
	WGS	1590	3668	9445	23929	43549	49959	
RIVIT_MED	SNP	852	1292	1764	3415	15153	27116	
	WGS	1358	1860	2502	4920	17912	29334	
RIVPK_NIL	SNP	952	1581	2259	4573	23613	36469	
	WGS	1657	1879	2580	4069	14157	31443	
SWAMP	SNP	1162	4203	8643	13663	19000	21135	
	WGS	1581	3681	8814	21663	38070	41271	
SWACN_GUI	SNP	871	1257	1785	3313	9741	12272	
	WGS	1395	2084	2979	5821	21602	31140	
SWATH_THS	SNP	931	1054	1281	1949	5259	9690	
	WGS	1627	1705	1875	2476	5929	12020	

Supplementary Table S4.2: MAF Correlation Tests between buffalo species (river MAF vs swamp MAF), and genomic data (SNP array MAF vs WGS MAF)

LD Pruning Target	Prune	Within Species between Data		Within Data between Species	
		River	Swamp	SNP	WGS
ALL	1.00	0.378	**< 0.001**	0.276	**< 0.001**
	0.80	0.400	**< 0.001**	0.278	**< 0.001**
	0.50	0.317	**< 0.001**	0.331	**< 0.001**
	0.20	0.175	**< 0.001**	**0.011**	**< 0.001**
	0.10	0.409	**0.016**	**< 0.001**	**< 0.001**
	0.05	0.247	0.810	**< 0.001**	**0.007**
	0.01	0.687	0.772	**< 0.001**	0.810
RIVER	0.80	0.585	**< 0.001**	0.275	0.284
	0.50	0.583	**< 0.001**	0.274	0.189
	0.20	0.308	**< 0.001**	0.308	0.136
	0.10	**0.003**	**< 0.001**	0.402	0.257
	0.05	**< 0.001**	**< 0.001**	0.491	0.717
	0.01	**< 0.001**	**< 0.001**	0.484	0.410
RIVIT_MED	0.80	**0.002**	**< 0.001**	0.294	0.641
	0.50	**0.002**	**< 0.001**	0.294	0.904
	0.20	**< 0.001**	**< 0.001**	0.334	0.748
	0.10	**< 0.001**	**< 0.001**	0.356	0.340
	0.05	**< 0.001**	**< 0.001**	0.361	0.255
	0.01	**< 0.001**	**< 0.001**	0.362	0.245
RIVPK_NIL	0.80	**0.003**	**< 0.001**	0.277	0.373
	0.50	**< 0.001**	**< 0.001**	0.296	0.343
	0.20	**< 0.001**	**< 0.001**	0.400	0.386
	0.10	**< 0.001**	**< 0.001**	0.429	0.401
	0.05	**< 0.001**	**< 0.001**	0.436	0.423
	0.01	**< 0.001**	**< 0.001**	0.457	0.433
SWAMP	0.80	0.299	**0.004**	0.451	0.482
	0.50	0.301	**0.001**	0.624	0.573
	0.20	0.299	**< 0.001**	0.373	0.527
	0.10	0.309	**< 0.001**	**0.020**	0.321
	0.05	0.326	**< 0.001**	**< 0.001**	0.105
	0.01	0.364	**< 0.001**	**< 0.001**	0.126
SWACN_GUI	0.80	0.297	**< 0.001**	**< 0.001**	0.993
	0.50	0.302	**< 0.001**	**< 0.001**	0.961
	0.20	0.326	**< 0.001**	**< 0.001**	0.513
	0.10	0.336	**< 0.001**	**< 0.001**	0.399
	0.05	0.336	**< 0.001**	**< 0.001**	0.419
	0.01	0.331	**< 0.001**	**< 0.001**	0.433
SWATH_THS	0.80	0.315	**< 0.001**	**< 0.001**	0.482
	0.50	0.320	**< 0.001**	**< 0.001**	0.417
	0.20	0.330	**< 0.001**	**< 0.001**	0.334
	0.10	0.322	**< 0.001**	**< 0.001**	0.373
	0.05	0.321	**< 0.001**	**< 0.001**	0.361
	0.01	0.320	**< 0.001**	**< 0.001**	0.366

Table S4.3: H_O across SNP array for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	0.402 (±0.045)	0.410 (±0.049)	0.396 (±0.047)	**0.373 (±0.046)**	**0.392 (±0.048)**	**0.397 (±0.048)**	**0.397 (±0.048)**	0.153
	Swamp	0.374 (±0.015)	0.387 (±0.018)	0.361 (±0.016)	**0.300 (±0.013)**	**0.157 (±0.011)**	**0.123 (±0.011)**	**0.118 (±0.011)**	**< 0.001**
RIVER	River	**0.472 (±0.052)**	**0.466 (±0.053)**	**0.436 (±0.050)**	**0.410 (±0.048)**	**0.399 (±0.048)**	**0.398 (±0.048)**		**< 0.001**
	Swamp	**0.124 (±0.011)**	**0.121 (±0.011)**	**0.124 (±0.012)**	**0.122 (±0.011)**	**0.121 (±0.011)**	**0.119 (±0.011)**		0.937
RIVIT_MED	River	**0.443 (±0.052)**	**0.442 (±0.054)**	**0.439 (±0.054)**	**0.430 (±0.050)**	**0.408 (±0.049)**	**0.405 (±0.050)**		**0.025**
	Swamp	**0.123 (±0.007)**	**0.120 (±0.009)**	**0.121 (±0.010)**	**0.122 (±0.010)**	**0.120 (±0.011)**	**0.120 (±0.011)**		0.987
RIVPK_NIL	River	**0.456 (±0.054)**	**0.456 (±0.054)**	**0.452 (±0.052)**	**0.445 (±0.051)**	**0.406 (±0.048)**	**0.399 (±0.048)**		0.155
	Swamp	**0.118 (±0.011)**	**0.115 (±0.009)**	**0.120 (±0.010)**	**0.120 (±0.013)**	**0.121 (±0.011)**	**0.119 (±0.011)**		0.940
SWAMP	River	0.396 (±0.049)	**0.398 (±0.047)**	**0.398 (±0.047)**	**0.399 (±0.048)**	**0.400 (±0.048)**	**0.400 (±0.048)**		0.836
	Swamp	0.429 (±0.021)	**0.330 (±0.013)**	**0.292 (±0.013)**	**0.268 (±0.017)**	**0.229 (±0.011)**	**0.214 (±0.012)**		**< 0.001**
SWACN_GUI	River	0.399 (±0.047)	0.403 (±0.050)	0.403 (±0.049)	0.399 (±0.049)	**0.398 (±0.048)**	**0.398 (±0.048)**		0.729
	Swamp	0.425 (±0.024)	0.416 (±0.024)	0.408 (±0.023)	0.375 (±0.017)	**0.324 (±0.016)**	**0.318 (±0.017)**		**< 0.001**
SWATH_THS	River	0.403 (±0.045)	0.403 (±0.046)	0.402 (±0.047)	0.401 (±0.046)	**0.402 (±0.047)**	**0.400 (±0.048)**		0.984
	Swamp	0.416 (±0.017)	0.408 (±0.018)	0.404 (±0.015)	0.391 (±0.010)	**0.349 (±0.008)**	**0.329 (±0.007)**		**0.027**
Significant	**River**	**0.002**	**0.001**	0.051	**0.012**	0.347	0.758		
	Swamp	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		

Table S4.4: H_E across SNP array for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	0.381 (±0.018)	0.387 (±0.021)	0.373 (±0.021)	**0.352 (±0.021)**	**0.370 (±0.022)**	**0.373 (±0.022)**	**0.373 (±0.022)**	0.258
	Swamp	0.366 (±0.010)	0.379 (±0.012)	0.354 (±0.011)	**0.293 (±0.005)**	**0.153 (±0.004)**	**0.120 (±0.004)**	**0.114 (±0.004)**	**< 0.001**
RIVER	River	**0.444 (±0.031)**	**0.438 (±0.025)**	**0.409 (±0.021)**	**0.384 (±0.021)**	**0.374 (±0.022)**	**0.373 (±0.022)**		**< 0.001**
	Swamp	**0.121 (±0.004)**	**0.118 (±0.004)**	**0.120 (±0.004)**	**0.119 (±0.004)**	**0.117 (±0.004)**	**0.116 (±0.004)**		0.915
RIVIT_MED	River	**0.414 (±0.040)**	**0.413 (±0.045)**	**0.412 (±0.038)**	**0.402 (±0.023)**	**0.382 (±0.022)**	**0.380 (±0.022)**		0.242
	Swamp	**0.121 (±0.003)**	**0.117 (±0.003)**	**0.117 (±0.003)**	**0.119 (±0.004)**	**0.117 (±0.004)**	**0.120 (±0.004)**		0.909
RIVPK_NIL	River	**0.427 (±0.004)**	**0.428 (±0.043)**	**0.425 (±0.042)**	**0.416 (±0.031)**	**0.380 (±0.022)**	**0.375 (±0.023)**		**0.048**
	Swamp	**0.114 (±0.005)**	**0.112 (±0.005)**	**0.118 (±0.005)**	**0.117 (±0.005)**	**0.118 (±0.004)**	**0.116 (±0.004)**		0.941
SWAMP	River	**0.373 (±0.021)**	**0.374 (±0.021)**	**0.374 (±0.022)**	**0.375 (±0.022)**	**0.376 (±0.023)**	**0.376 (±0.023)**		0.899
	Swamp	**0.421 (±0.016)**	**0.322 (±0.006)**	**0.284 (±0.006)**	**0.261 (±0.005)**	**0.222 (±0.004)**	**0.208 (±0.005)**		**< 0.001**
SWACN_GUI	River	0.375 (±0.024)	0.376 (±0.023)	0.377 (±0.023)	0.377 (±0.023)	**0.375 (±0.023)**	**0.375 (±0.022)**		0.994
	Swamp	0.413 (±0.039)	0.407 (±0.036)	0.396 (±0.029)	0.365 (±0.018)	**0.316 (±0.014)**	**0.311 (±0.015)**		**< 0.001**
SWATH_THS	River	0.377 (±0.021)	0.376 (±0.021)	0.377 (±0.022)	0.377 (±0.022)	**0.378 (±0.022)**	**0.377 (±0.022)**		0.999
	Swamp	0.404 (±0.040)	0.396 (±0.036)	0.399 (±0.034)	0.384 (±0.030)	**0.339 (±0.023)**	**0.322 (±0.022)**		**0.005**
Significant	**River**	**< 0.001**	**0.002**	**0.020**	**0.024**	0.830	0.871		
	Swamp	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		

Table S4.5: Number of Markers across SNP array for River and Swamp

LD Pruning Target	Species	r^2 Threshold 0.01		0.05		0.1		0.2		0.5		0.8		1	
		Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly
ALL	River	944	919 (±28.0)	2739	2682.5 (±75.4)	5265	5121.5 (±177.1)	9952	9599.5 (±430.2)	27891	27229 (±989.4)	38397	37582.5 (±1277.3)	40,695	39843 (±1355.8)
	Swamp		878 (±25.6)		2627 (±62.7)		4923 (±133.1)		8179 (±267.0)		14094 (±1509.2)		16357 (±2734.5)		16729 (±3014.2)
RIVER	River	1213	1208.5 (±10.7)	3780	3761 (±37.8)	9369	9247 (±185.1)	19847	19460 (±583.6)	33406	32692.5 (±1098.8)	38178	37376.5 (±1263.5)		
	Swamp		511 (±99.4)		1583 (±297.4)		4021 (±729.3)		8448.0 (±1489.0)		14035 (±2452.8)		15859 (±2804.5)		
RIVIT_MED	River	852	848 (±19.5)	1292	1287 (±32.7)	1764	1757.0 (±42.1)	3415	3402 (±88.5)	15153	15068 (±473.4)	27116	26955.5 (±871.1)		
	Swamp		368 (±62.3)		542 (±95.4)		740 (±136.8)		1447 (±267.2)		6376 (±1132.3)		11382 (±2023.7)		
RIVPK_NIL	River	952	947 (±14.3)	1581	1574.5 (±22.6)	2259	2252.5 (±34.4)	4573	451.5 (±78.3)	23613	23191 (±743.9)	36469	35762 (±1215.3)		
	Swamp		386 (±73.3)		641 (±124.4)		957 (±171.2)		1933 (±361.0)		9883 (±1743.3)		15158 (±2678.6)		
SWAMP	River	1162	1137 (±36.3)	4203	4106.5 (±133.7)	8643	8464.5 (±280)	13663	13405.5 (±439.0)	19000	18648.5 (±609.5)	21135	20736 (±676.0)		
	Swamp		1140 (±18.6)		3687 (±111.3)		7306 (±327.2)		11216 (±567.3)		13893 (±899.8)		14628 (±1226.3)		
SWACN_GUI	River	871	853 (±28.4)	1257	1235 (±39.0)	1785	1755.5 (±55.9)	3313	3251 (±105.5)	9741	9557 (±318.4)	12272	12038.5 (±397.9)		
	Swamp		859 (±20.4)		1234 (±31.9)		1743 (±51.1)		3139 (±136.3)		8795 (583.6)		11035 (±763.0)		
SWATH_THS	River	931	904 (±27.9)	1054	1024.5 (±31.7)	1281	1248.5 (±38.6)	1949	1904.5 (±59.6)	5259	5156.5 (±157.1)	9690	9506 (±304.5)		
	Swamp		914 (±24.8)		1018 (±30.2)		1250.5 (±35.6)		1878 (±59.8)		4832 (±237.1)		8700 (±488.3)		

Table S4.6: F across SNP array for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	**-0.043 (±0.066)**	-0.050 (±0.061)	**-0.050 (±0.062)**	**-0.048 (±0.060)**	**-0.046 (±0.060)**	**-0.053 (±0.060)**	**-0.052 (±0.060)**	0.988
	Swamp	**-0.017 (±0.085)**	-0.030 (±0.076)	**-0.021 (±0.074)**	**-0.023 (±0.081)**	**0.003 (±0.160)**	**0.025 (±0.220)**	**0.026 (±0.231)**	< 0.001
RIVER	River	**-0.044 (±0.068)**	**-0.052 (±0.059)**	**-0.051 (±0.059)**	**-0.055 (±0.057)**	**-0.054 (±0.059)**	**-0.053 (±0.060)**		0.999
	Swamp	**0.026 (±0.236)**	**0.025 (±0.240)**	**0.026 (±0.227)**	**0.018 (±0.227)**	**0.025 (±0.225)**	**0.025 (±0.228)**		0.999
RIVIT_MED	River	**-0.059 (±0.072)**	**-0.053 (±0.065)**	**-0.055 (±0.061)**	**-0.054 (±0.059)**	**-0.057 (±0.059)**	**-0.053 (±0.060)**		0.972
	Swamp	**0.039 (±0.223)**	**0.016 (±0.231)**	**0.020 (±0.225)**	**0.017 (±0.227)**	**0.030 (±0.227)**	**0.022 (±0.226)**		0.812
RIVPK_NIL	River	**-0.049 (±0.077)**	**-0.052 (±0.069)**	**-0.054 (±0.062)**	**-0.053 (±0.062)**	**-0.054 (±0.060)**	**-0.053 (±0.060)**		0.988
	Swamp	**0.020 (±0.244)**	**0.030 (±0.255)**	**0.034 (±0.237)**	**0.026 (±0.239)**	**0.027 (±0.226)**	**0.025 (±0.228)**		0.959
SWAMP	River	-0.045 (±0.065)	**-0.052 (±0.063)**	**-0.046 (±0.059)**	**-0.051 (±0.060)**	**-0.055 (±0.060)**	**-0.053 (±0.060)**		0.999
	Swamp	-0.025 (±0.074)	**-0.018 (±0.081)**	**-0.020 (±0.090)**	**-0.019 (±0.103)**	**0.007 (±0.138)**	**0.001 (±0.164)**		0.266
SWACN_GUI	River	-0.054 (±0.068)	**-0.052 (±0.065)**	-0.053 (±0.066)	-0.045 (±0.061)	-0.048 (±0.060)	-0.051 (±0.059)		0.921
	Swamp	-0.035 (±0.077)	**-0.018 (±0.076)**	-0.034 (±0.076)	-0.027 (±0.074)	-0.023 (±0.079)	-0.024 (±0.059)		0.648
SWATH_THS	River	-0.057 (±0.077)	-0.058 (±0.073)	-0.054 (±0.073)	**-0.050 (±0.061)**	-0.054 (±0.062)	**-0.050 (±0.060)**		0.997
	Swamp	-0.026 (±0.085)	-0.034 (±0.085)	-0.032 (±0.080)	**-0.022 (±0.077)**	-0.022 (±0.077)	**-0.027 (±0.078)**		0.601
River		0.946	0.978	0.985	0.961	0.966	0.998		
Swamp		**0.003**	**0.008**	**0.005**	**0.008**	**0.018**	**0.031**		

Table S4.7: H_O across WGS for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	**0.301 (±0.028)**	**0.300 (±0.029)**	0.232 (±0.029)	0.181 (±0.027)	**0.196 (±0.025)**	**0.204 (±0.025)**	**0.200 (±0.025)**	**< 0.001**
	Swamp	**0.209 (±0.015)**	**0.244 (±0.018)**	0.201 (±0.019)	0.160 (±0.018)	**0.153 (±0.016)**	**0.149 (±0.015)**	**0.144 (±0.014)**	**< 0.001**
RIVER	River	**0.473 (±0.047)**	**0.435 (±0.040)**	**0.345 (±0.032)**	**0.300 (±0.031)**	**0.284 (±0.030)**	**0.283 (±0.030)**		**< 0.001**
	Swamp	**0.111 (±0.012)**	**0.111 (±0.011)**	**0.118 (±0.012)**	**0.112 (±0.012)**	**0.108 (±0.012)**	**0.106 (±0.012)**		0.689
RIVIT_MED	River	**0.415 (±0.048)**	**0.413 (±0.051)**	**0.407 (±0.047)**	**0.371 (±0.035)**	**0.337 (±0.037)**	**0.334 (±0.040)**		**0.001**
	Swamp	**0.097 (±0.011)**	**0.094 (±0.011)**	**0.097 (±0.012)**	**0.104 (±0.012)**	**0.102 (±0.012)**	**0.098 (±0.011)**		0.594
RIVPK_NIL	River	**0.446 (±0.044)**	**0.440 (±0.042)**	**0.432 (±0.045)**	**0.420 (±0.042)**	**0.362 (±0.037)**	**0.333 (±0.039)**		**0.008**
	Swamp	**0.098 (±0.013)**	**0.097 (±0.013)**	**0.095 (±0.013)**	**0.095 (±0.012)**	**0.094 (±0.010)**	**0.096 (±0.011)**		0.986
SWAMP	River	**0.169 (±0.023)**	**0.166 (±0.025)**	**0.152 (±0.025)**	**0.139 (±0.024)**	**0.134 (±0.024)**	**0.132 (±0.023)**		**0.002**
	Swamp	**0.423 (±0.033)**	**0.393 (±0.030)**	**0.318 (±0.025)**	**0.271 (±0.026)**	**0.246 (±0.026)**	**0.243 (±0.026)**		**< 0.001**
SWACN_GUI	River	**0.169 (±0.027)**	**0.155 (±0.026)**	**0.148 (±0.025)**	**0.142 (±0.025)**	**0.130 (±0.024)**	**0.125 (±0.023)**		**0.007**
	Swamp	**0.403 (±0.038)**	**0.393 (±0.039)**	**0.381 (±0.038)**	**0.345 (±0.032)**	**0.285 (±0.033)**	**0.277 (±0.035)**		**< 0.001**
SWATH_THS	River	**0.149 (±0.028)**	**0.151 (±0.025)**	**0.153 (±0.025)**	**0.155 (±0.024)**	**0.141 (±0.024)**	**0.141 (±0.023)**		0.455
	Swamp	**0.381 (±0.036)**	**0.381 (±0.030)**	**0.379 (±0.025)**	**0.367 (±0.026)**	**0.341 (±0.026)**	**0.322 (±0.026)**		0.217
Significant	**River**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		
	Swamp	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		

Table S4.8: H_E across WGS for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	**0.265 (±0.010)**	**0.265 (±0.008)**	0.206 (±0.005)	0.161 (±0.006)	**0.173 (±0.003)**	**0.180 (±0.003)**	**0.177 (±0.003)**	**< 0.001**
	Swamp	**0.193 (±0.003)**	**0.230 (±0.004)**	0.190 (±0.004)	0.151 (±0.004)	**0.142 (±0.002)**	**0.139 (±0.001)**	**0.134 (±0.001)**	**< 0.001**
RIVER	River	**0.416 (±0.030)**	**0.382 (±0.017)**	**0.303 (±0.008)**	**0.264 (±0.008)**	**0.251 (±0.009)**	**0.250 (±0.009)**		**< 0.001**
	Swamp	**0.105 (±0.005)**	**0.104 (±0.004)**	**0.111 (±0.004)**	**0.107 (±0.004)**	**0.103 (±0.004)**	**0.101 (±0.004)**		0.339
RIVIT_MED	River	**0.364 (±0.054)**	**0.363 (±0.049)**	**0.357 (±0.036)**	**0.324 (±0.016)**	**0.296 (±0.016)**	**0.296 (±0.019)**		**0.007**
	Swamp	**0.091 (±0.005)**	**0.089 (±0.005)**	**0.092 (±0.005)**	**0.098 (±0.005)**	**0.097 (±0.004)**	**0.094 (±0.005)**		0.351
RIVPK_NIL	River	**0.390 (±0.048)**	**0.388 (±0.047)**	**0.382 (±0.045)**	**0.371 (±0.036)**	**0.319 (±0.022)**	**0.295 (±0.023)**		**0.001**
	Swamp	**0.095 (±0.004)**	**0.094 (±0.004)**	**0.093 (±0.004)**	**0.093 (±0.005)**	**0.091 (±0.005)**	**0.093 (±0.005)**		0.822
SWAMP	River	**0.148 (±0.003)**	**0.145 (±0.003)**	**0.135 (±0.004)**	**0.123 (±0.004)**	**0.119 (±0.004)**	**0.117 (±0.004)**		**< 0.001**
	Swamp	**0.408 (±0.036)**	**0.376 (±0.023)**	**0.301 (±0.012)**	**0.255 (±0.012)**	**0.230 (±0.011)**	**0.227 (±0.011)**		**< 0.001**
SWACN_GUI	River	**0.142 (±0.004)**	**0.137 (±0.004)**	**0.133 (±0.004)**	**0.127 (±0.004)**	**0.116 (±0.004)**	**0.112 (±0.004)**		**< 0.001**
	Swamp	**0.386 (±0.054)**	**0.379 (±0.050)**	**0.366 (±0.042)**	**0.329 (±0.025)**	**0.269 (±0.020)**	**0.260 (±0.022)**		**< 0.001**
SWATH_THS	River	**0.133 (±0.005)**	**0.134 (±0.005)**	**0.135 (±0.004)**	**0.138 (±0.004)**	**0.126 (±0.004)**	**0.125 (±0.004)**		0.185
	Swamp	**0.374 (±0.060)**	**0.372 (±0.058)**	**0.370 (±0.057)**	**0.361 (±0.053)**	**0.327 (±0.037)**	**0.310 (±0.033)**		0.213
Significant	**River**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		
	Swamp	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		

Table S4.9: Number of Markers across WGS for River and Swamp

LD Pruning Target	Species	r^2 Threshold 0.01		0.05		0.1		0.2		0.5		0.8		1	
		Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly	Total	Poly
ALL	River		869 (±77)		2085 (±172)		4308 (±457)		9599.5 (±430.2)		27229 (±989.4)		37582.5 (±1277.3)		36420 (±3618)
	Swamp		639 (±77)		1841 (±219)		4187 (±674)		8179 (±267.0)		14094 (±1509.2)		16357 (±2734.5)		30342 (±4885)
RIVER	River		1533 (±50)		3406 (±134)		7700 (±505)		19460 (±583.6)		32692.5 (±1098.8)		37376.5 (±1263.5)		
	Swamp		521 (±88)		1171 (±186)		3234 (±484)		8448.0 (±1489.0)		14035 (±2452.8)		15859 (±2804.5)		
RIVIT_MED	River		1250 (±91)		1711 (±125)		2284 (±174)		3402 (±88.5)		15068 (±473.4)		26955.5 (±871.1)		
	Swamp		395 (±68)		528 (±89)		716 (±116)		1447 (±267.2)		6376 (±1132.3)		11382 (±2023.7)		
RIVPK_NIL	River		1529 (±84)		1733 (±97)		2349 (±141)		4510.5 (±78.3)		23191 (±743.9)		35762 (±1215.3)		
	Swamp		486 (±85)		539 (±93)		741 (±127)		1933 (±361.0)		9883 (±1743.3)		15158 (±2678.6)		
SWAMP	River		658 (±57)		1534 (±137)		3441 (±310)		13405.5 (±439.0)		18648.5 (±609.5)		20736 (±676.0)		
	Swamp		1469 (±87)		3281 (±216)		7090 (±732)		11216 (±567.3)		13893 (±899.8)		14628 (±1226.3)		
SWACN_GUI	River		573 (±52)		842 (±78)		1174 (±108)		3251 (±105.5)		9557 (±318.4)		12038.5 (±397.9)		
	Swamp		1268 (±96)		1870 (±148)		2618 (±237)		3139 (±136.3)		8795 (583.6)		11035 (±763.0)		
SWATH_THS	River		642 (±62)		683 (±66)		756 (±75)		1904.5 (±59.6)		5156.5 (±157.1)		9506 (±304.5)		
	Swamp		1459 (±128)		1515 (±135)		1656 (±151)		1878 (±59.8)		4832 (±237.1)		8700 (±488.3)		

Table S4.10: F across WGS for River and Swamp. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	River	-0.096 (±0.131)	-0.111 (±0.117)	-0.050 (±0.062)	-0.048 (±0.060)	-0.046 (±0.060)	-0.053 (±0.060)	-0.098 (±0.102)	1.000
	Swamp	-0.059 (±0.122)	-0.030 (±0.102)	-0.021 (±0.074)	-0.023 (±0.081)	0.003 (±0.160)	0.025 (±0.220)	-0.056 (±0.100)	0.998
RIVER	River	-0.099 (±0.100)	-0.100 (±0.101)	**-0.051 (±0.059)**	**-0.055 (±0.057)**	**-0.054 (±0.059)**	**-0.053 (±0.060)**		0.996
	Swamp	-0.005 (±0.166)	-0.025 (±0.142)	**0.026 (±0.227)**	**0.018 (±0.227)**	**0.025 (±0.225)**	**0.025 (±0.228)**		0.977
RIVIT_MED	River	**-0.111 (±0.101)**	-0.101 (±0.101)	**-0.055 (±0.061)**	**-0.054 (±0.059)**	**-0.057 (±0.059)**	**-0.053 (±0.060)**		0.989
	Swamp	**-0.036 (±0.151)**	-0.030 (±0.154)	**0.020 (±0.225)**	**0.017 (±0.227)**	**0.030 (±0.227)**	**0.022 (±0.226)**		0.987
RIVPK_NIL	River	**-0.111 (±0.111)**	**-0.089 (±0.109)**	**-0.054 (±0.062)**	**-0.053 (±0.062)**	**-0.054 (±0.060)**	**-0.053 (±0.060)**		0.994
	Swamp	**0.016 (±0.161)**	**0.010 (±0.152)**	**0.034 (±0.237)**	**0.026 (±0.239)**	**0.027 (±0.226)**	**0.025 (±0.228)**		0.998
SWAMP	River	-0.113 (±0.127)	-0.100 (±0.118)	-0.046 (±0.059)	-0.051 (±0.060)	-0.055 (±0.060)	-0.053 (±0.060)		0.967
	Swamp	-0.024 (±0.101)	-0.030 (±0.100)	-0.020 (±0.090)	-0.019 (±0.103)	0.007 (±0.138)	0.001 (±0.164)		0.872
SWACN_GUI	River	-0.091 (±0.131)	-0.087 (±0.130)	-0.053 (±0.066)	-0.045 (±0.061)	-0.048 (±0.060)	-0.051 (±0.059)		0.993
	Swamp	-0.029 (±0.098)	-0.021 (±0.098)	-0.034 (±0.076)	-0.027 (±0.074)	-0.023 (±0.079)	-0.024 (±0.059)		0.876
SWATH_THS	River	-0.087 (±0.130)	-0.087 (±0.126)	-0.054 (±0.073)	-0.050 (±0.061)	-0.054 (±0.062)	-0.050 (±0.060)		0.997
	Swamp	0.000 (±0.104)	-0.015 (±0.104)	-0.032 (±0.080)	-0.022 (±0.077)	-0.022 (±0.077)	-0.027 (±0.078)		0.991
Significant	**River**	0.970	0.999	0.960	0.993	0.989	0.999		
	Swamp	0.737	0.870	0.715	0.475	0.320	0.227		

Table S4.11: T.test results for between SNP array data vs WGS for Ho, He, and F

LD Prune Target	R^2 Threshold	H_O		H_E		F	
		River	Swamp	River	Swamp	River	Swamp
ALL	1.00	**0.000583**	**0.006610**	**0.000000**	**0.000000**	0.097320	**0.010450**
	0.80	**0.000583**	**0.008658**	**0.000583**	**0.041126**	0.097319	**0.008658**
	0.50	**0.000583**	0.588745	**0.000583**	0.240260	0.097319	**0.015152**
	0.20	**0.000583**	**0.002165**	**0.000583**	**0.002165**	0.208625	0.093074
	0.10	**0.000583**	**0.002165**	**0.000583**	**0.002165**	0.164918	0.484848
	0.05	**0.000583**	**0.002165**	**0.000583**	**0.002165**	0.164918	0.699134
	0.01	**0.000583**	**0.002165**	**0.000583**	**0.002165**	0.164918	0.179654
RIVER	0.80	**0.000583**	0.179654	**0.000583**	0.064935	0.097319	**0.041126**
	0.50	**0.000583**	0.179654	**0.000583**	0.064935	0.097319	**0.025974**
	0.20	**0.000583**	0.179654	**0.000583**	0.132035	0.072844	**0.015152**
	0.10	**0.000583**	0.588745	**0.000583**	0.179654	0.053030	**0.015152**
	0.05	0.164918	0.240260	**0.002331**	0.132035	0.053030	**0.041126**
	0.01	0.901515	0.179654	0.128205	0.132035	0.053030	0.132035
RIVIT_MED	0.80	**0.001166**	**0.041126**	**0.000583**	**0.015152**	0.072844	**0.025974**
	0.50	**0.001166**	**0.041126**	**0.000583**	**0.041126**	**0.026224**	**0.025974**
	0.20	**0.006993**	0.064935	**0.004079**	**0.041126**	**0.017483**	0.064935
	0.10	0.259324	**0.025974**	**0.037879**	**0.008658**	**0.017483**	0.393939
	0.05	0.455711	**0.025974**	**0.037879**	**0.008658**	**0.026224**	0.132035
	0.01	0.317599	**0.015152**	**0.037879**	**0.008658**	**0.037879**	0.064935
RIVPK_NIL	0.80	**0.017483**	**0.015152**	**0.002331**	**0.008658**	0.097319	0.093074
	0.50	**0.037879**	**0.015152**	**0.006993**	**0.008658**	0.053030	0.093074
	0.20	0.534965	**0.025974**	0.097319	**0.008658**	0.164918	0.309524
	0.10	0.455711	**0.025974**	0.164918	**0.008658**	0.128205	0.393939
	0.05	0.534965	**0.041126**	0.128205	**0.064935**	0.053030	0.240260
	0.01	0.620047	0.064935	0.164918	**0.041126**	0.097319	0.393939
SWAMP	0.80	**0.000583**	**0.025974**	**0.000583**	0.179654	0.208625	**0.025974**
	0.50	**0.000583**	0.132035	**0.000583**	0.818182	0.208625	**0.041126**
	0.20	**0.000583**	0.937229	**0.000583**	0.937229	0.164918	0.132035
	0.10	**0.000583**	0.132035	**0.000583**	0.484848	0.097319	0.393939

	0.05	**0.000583**	**0.004329**	**0.000583**	**0.002165**	0.128205	0.309524
	0.01	**0.000583**	0.937229	**0.000583**	0.484848	0.097319	0.699134
SWACN_GUI	0.80	**0.000583**	**0.041126**	**0.000583**	**0.025974**	0.164918	0.240260
	0.50	**0.000583**	**0.041126**	**0.000583**	**0.025974**	0.208625	0.179654
	0.20	**0.000583**	0.240260	**0.000583**	0.240260	0.128205	0.393939
	0.10	**0.000583**	0.484848	**0.000583**	0.393939	0.317599	0.699134
	0.05	**0.000583**	0.588745	**0.000583**	0.393939	0.208625	0.937229
	0.01	**0.000583**	0.484848	**0.000583**	0.393939	0.208625	0.937229
SWATH_THS	0.80	**0.000583**	0.818182	**0.000583**	0.393939	0.208625	0.588745
	0.50	**0.000583**	0.818182	**0.000583**	0.484848	0.164918	0.699134
	0.20	**0.000583**	0.309524	**0.000583**	0.393939	0.164918	0.818182
	0.10	**0.000583**	0.309524	**0.000583**	0.393939	0.128205	0.393939
	0.05	**0.000583**	0.393939	**0.000583**	0.393939	0.208625	0.484848
	0.01	**0.000583**	0.309524	**0.000583**	0.309524	0.455711	0.484848

Table S4.12: FST values within river comparisons. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	**Species**	**r^2 Threshold** 0.01	0.05	0.1	0.2	0.5	0.8	1	**Significant**
ALL	SNP	0.047 (± 0.034)	0.049 (± 0.035)	0.048 (± 0.036)	0.046 (± 0.034)	0.049 (± 0.037)	0.050 (± 0.037)	0.050 (± 0.037)	1.000
	WGS	0.054 (± 0.036)	0.050 (± 0.035)	0.046 (± 0.035)	0.048 (± 0.036)	0.053 (± 0.038)	0.054 (± 0.038)	0.054 (± 0.038)	0.988
RIVER	SNP	0.051 (± 0.036)	0.042 (± 0.030)	0.041 (± 0.029)	0.046 (± 0.034)	0.049 (± 0.037)	0.049 (± 0.037)		0.939
	WGS	0.057 (± 0.041)	0.048 (± 0.032)	0.042 (± 0.027)	0.045 (± 0.030)	0.052 (± 0.036)	0.054 (± 0.038)		0.683
RIVIT_MED	SNP	0.035 (± 0.022)	0.037 (± 0.023)	0.037 (± 0.024)	0.041 (± 0.028)	0.048 (± 0.035)	0.047 (± 0.034)		0.759
	WGS	0.048 (± 0.031)	0.047 (± 0.030)	0.047 (± 0.030)	0.046 (± 0.029)	0.045 (± 0.029)	0.044 (± 0.028)		0.993
RIVPK_NIL	SNP	0.051 (± 0.040)	0.048 (± 0.039)	0.049 (± 0.039)	0.050 (± 0.038)	0.050 (± 0.038)	0.050 (± 0.037)		0.999
	WGS	0.055 (± 0.042)	0.055 (± 0.043)	0.057 (± 0.044)	0.056 (± 0.043)	0.054 (± 0.042)	0.053 (± 0.043)		0.997
SWAMP	SNP	0.049 (± 0.038)	0.048 (± 0.036)	0.048 (± 0.037)	0.049 (± 0.037)	0.049 (± 0.037)	0.049 (± 0.037)		0.999
	WGS	0.053 (± 0.035)	0.052 (± 0.035)	0.052 (± 0.036)	0.050 (± 0.035)	0.051 (± 0.035)	0.052 (± 0.035)		0.999
SWACN_GUI	SNP	0.051 (± 0.036)	0.050 (± 0.036)	0.049 (± 0.036)	0.048 (± 0.036)	0.050 (± 0.037)	0.049 (± 0.038)		0.999
	WGS	0.053 (± 0.034)	0.052 (± 0.033)	0.051 (± 0.033)	0.052 (± 0.033)	0.052 (± 0.034)	0.052 (± 0.034)		0.999
SWATH_THS	SNP	0.051 (± 0.038)	0.051 (± 0.038)	0.050 (± 0.036)	0.049 (± 0.037)	0.049 (± 0.038)	0.049 (± 0.037)		0.999
	WGS	0.054 (± 0.034)	0.054 (± 0.034)	0.057 (± 0.035)	0.055 (± 0.034)	0.053 (± 0.034)	0.054 (± 0.035)		0.998
Significant	**SNP**	0.841	0.921	0.945	0.995	0.999	0.999		
	WGS	0.994	0.997	0.849	0.932	0.997	0.990		

Table S4.13: FST values within swamp comparisons. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	SNP	0.067 (± 0.035)	0.068 (± 0.035)	0.068 (± 0.037)	0.066 (± 0.034)	0.053 (± 0.024)	0.048 (± 0.021)	0.047 (± 0.020)	0.157
	WGS	0.059 (± 0.029)	0.056 (± 0.028)	0.050 (± 0.025)	0.050 (± 0.025)	0.057 (± 0.027)	0.058 (± 0.027)	0.058 (± 0.027)	0.935
RIVER	SNP	0.045 (± 0.020)	0.047 (± 0.020)	0.047 (± 0.021)	0.047 (± 0.020)	0.047 (± 0.021)	0.047 (± 0.021)		0.989
	WGS	0.050 (± 0.024)	0.052 (± 0.026)	0.055 (± 0.025)	0.055 (± 0.025)	0.057 (± 0.026)	0.057 (± 0.026)		0.957
RIVIT_MED	SNP	0.049 (± 0.021)	0.047 (± 0.020)	0.047 (± 0.020)	0.047 (± 0.021)	0.047 (± 0.021)	0.047 (± 0.021)		0.999
	WGS	0.058 (± 0.025)	0.056 (± 0.025)	0.055 (± 0.026)	0.055 (± 0.026)	0.055 (± 0.026)	0.055 (± 0.026)		0.996
RIVPK_NIL	SNP	0.045 (± 0.018)	0.044 (± 0.018)	0.045 (± 0.020)	0.047 (± 0.020)	0.047 (± 0.021)	0.047 (± 0.020)		0.990
	WGS	0.060 (± 0.026)	0.058 (± 0.026)	0.056 (± 0.025)	0.057 (± 0.027)	0.056 (± 0.026)	0.056 (± 0.026)		0.996
SWAMP	SNP	0.065 (± 0.034)	0.053 (± 0.026)	0.052 (± 0.026)	0.055 (± 0.027)	0.053 (± 0.025)	0.052 (± 0.024)		0.704
	WGS	0.070 (± 0.032)	0.063 (± 0.030)	0.055 (± 0.026)	0.053 (± 0.025)	0.058 (± 0.027)	0.059 (± 0.028)		0.511
SWACN_GUI	SNP	0.069 (± 0.039)	0.070 (± 0.039)	0.069 (± 0.038)	0.064 (± 0.037)	0.061 (± 0.034)	0.061 (± 0.034)		0.793
	WGS	0.063 (± 0.032)	0.066 (± 0.034)	0.066 (± 0.034)	0.066 (± 0.034)	0.059 (± 0.032)	0.057 (± 0.032)		0.895
SWATH_THS	SNP	0.071 (± 0.035)	0.073 (± 0.035)	0.072 (± 0.036)	0.068 (± 0.035)	0.066 (± 0.033)	0.065 (± 0.033)		0.824
	WGS	0.066 (± 0.030)	0.066 (± 0.029)	0.065 (± 0.029)	0.067 (± 0.031)	0.065 (± 0.031)	0.062 (± 0.032)		0.997
Significant	**SNP**	**0.031**	**0.024**	0.076	0.169	0.379	0.389		
	WGS	0.621	0.732	0.610	0.620	0.985	0.998		

Table S4.14: FST values between species comparisons. Bold values indicate significant results between species, r^2 thresholds and LD datasets.

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1	Significant
ALL	SNP	**0.242 (± 0.024)**	**0.140 (± 0.025)**	**0.124 (± 0.025)**	0.142 (± 0.023)	**0.278 (± 0.044)**	**0.352 (± 0.050)**	**0.360 (± 0.051)**	**< 0.001**
	WGS	**0.529 (± 0.024)**	**0.281 (± 0.023)**	**0.165 (± 0.024)**	0.133 (± 0.020)	**0.208 (± 0.021)**	**0.275 (± 0.022)**	**0.304 (± 0.023)**	**< 0.001**
RIVER	SNP	**0.372 (± 0.053)**	**0.374 (± 0.052)**	**0.368 (± 0.052)**	**0.366 (± 0.051)**	**0.360 (± 0.051)**	**0.359 (± 0.051)**		0.264
	WGS	**0.352 (± 0.045)**	**0.346 (± 0.043)**	**0.330 (± 0.036)**	**0.316 (± 0.035)**	**0.311 (± 0.035)**	**0.309 (± 0.036)**		**< 0.001**
RIVIT_MED	SNP	0.350 (± 0.051)	**0.357 (± 0.051)**	**0.357 (± 0.051)**	**0.357 (± 0.052)**	**0.355 (± 0.052)**	**0.356 (± 0.051)**		0.917
	WGS	0.331 (± 0.045)	**0.331 (± 0.046)**	**0.331 (± 0.045)**	**0.333 (± 0.041)**	**0.318 (± 0.042)**	**0.311 (± 0.045)**		0.128
RIVPK_NIL	SNP	**0.377 (± 0.053)**	**0.379 (± 0.052)**	**0.376 (± 0.052)**	**0.374 (± 0.052)**	**0.361 (± 0.051)**	**0.359 (± 0.051)**		0.074
	WGS	**0.339 (± 0.046)**	**0.341 (± 0.046)**	**0.341 (± 0.047)**	**0.332 (± 0.046)**	**0.316 (± 0.045)**	**0.303 (± 0.045)**		**< 0.001**
SWAMP	SNP	**0.132 (± 0.028)**	**0.199 (± 0.024)**	**0.227 (± 0.026)**	**0.243 (± 0.027)**	0.277 (± 0.031)	**0.292 (± 0.032)**		**< 0.001**
	WGS	**0.302 (± 0.058)**	**0.293 (± 0.056)**	**0.284 (± 0.052)**	**0.271 (± 0.050)**	0.272 (± 0.049)	**0.272 (± 0.050)**		**0.037**
SWACN_GUI	SNP	**0.137 (± 0.041)**	**0.138 (± 0.040)**	**0.147 (± 0.038)**	**0.167 (± 0.033)**	**0.202 (± 0.028)**	**0.206 (± 0.028)**		**< 0.001**
	WGS	**0.292 (± 0.057)**	**0.292 (± 0.056)**	**0.290 (± 0.057)**	**0.286 (± 0.057)**	**0.268 (± 0.057)**	**0.264 (± 0.058)**		**0.086**
SWATH_THS	SNP	**0.146 (± 0.044)**	**0.154 (± 0.044)**	**0.153 (± 0.043)**	**0.157 (± 0.041)**	**0.190 (± 0.035)**	**0.201 (± 0.035)**		**< 0.001**
	WGS	**0.308 (± 0.057)**	**0.305 (± 0.057)**	**0.307 (± 0.056)**	**0.306 (± 0.057)**	**0.291 (± 0.058)**	**0.286 (± 0.058)**		0.341
Significant	**SNP**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		
	WGS	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**		

Table S4.15: Average Number of ROHs per individual found in river buffalo for SNP array and WGS data

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
ALL	SNP		0.013	0.697	2.197	5.368	9.816	11.750
	WGS			1.000	3.907	15.209	22.395	19.973
RIVER	SNP		0.092	1.711	3.329	7.158	9.592	
	WGS		0.047	0.791	2.884	8.000	9.674	
RIVIT_MED	SNP				0.013	2.816	5.066	
	WGS				0.233	1.930	4.070	
RIVPK_NIL	SNP				0.263	4.289	8.803	
	WGS				0.070	1.791	5.628	
SWAMP	SNP		0.145	1.908	2.500	3.487	4.039	
	WGS		0.163	2.349	7.047	15.721	18.535	
SWACN_GUI	SNP				0.066	2.197	2.645	
	WGS				1.930	8.535	14.581	
SWATH_THS	SNP					0.579	2.250	
	WGS							

Table S4.16: Average Number of ROHs per individual found in swamp buffalo for SNP array and WGS data

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
ALL	SNP		0.028	0.676	1.704	10.915	32.211	41.042
	WGS			1.216	2.514	11.054	16.730	26.674
RIVER	SNP		0.535	5.592	12.761	28.127	35.282	
	WGS		0.162	2.514	6.270	15.351	18.892	
RIVIT_MED	SNP				0.183	11.042	21.761	
	WGS				0.973	5.676	11.811	
RIVPK_NIL	SNP				1.972	16.803	33.085	
	WGS				0.297	5.324	13.432	
SWAMP	SNP		0.239	1.493	2.169	3.887	4.915	
	WGS		0.054	1.054	2.270	4.973	5.297	
SWACN_GUI	SNP				0.014	1.775	2.211	
	WGS				0.676	2.324	3.486	
SWATH_THS	SNP					0.620	1.859	
	WGS							

Table S4.17: Length of ROHs found in River Buffalo

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
ALL	SNP		6.377163	7.770428	9.99534	8.976598	6.358075	5.849415
				3.163246	6.224376	7.216448	5.683332	5.288308
	WGS			6.879351	7.948545	3.273327	2.613099	2.41778
				3.476924	3.364939	1.418265	1.211308	1.127796
RIVER	SNP		6.606411	11.46987	11.29229	7.517095	6.428623	
			1.50716	6.948855	8.160584	6.302996	5.700813	
	WGS		11.10059	11.13674	6.700971	3.870795	3.522569	
			4.994857	4.780819	3.113	1.556985	1.460487	
RIVIT_MED	SNP				8.070136	11.19799	9.050028	
						6.825088	7.301163	
	WGS				8.090354	8.536977	5.704417	
					2.084583	3.214333	1.898155	
RIVPK_NIL	SNP				6.962482	10.0809	6.748153	
					2.243432	7.865282	5.995352	
	WGS				9.04633	10.95363	5.483895	
					0.414428	4.295299	2.142996	
SWAMP	SNP		6.61105	10.89944	12.33063	11.24909	10.47303	
			1.856229	7.398633	7.457768	7.368319	7.329266	
	WGS		7.029019	10.56151	7.664243	4.417576	4.078838	
			1.095541	5.271674	3.735468	2.004296	1.819658	
SWACN_GUI	SNP				6.226767	9.990285	10.42957	
					1.049246	5.904383	6.216093	
	WGS				8.046731	7.5989	5.379694	
					2.959045	3.254798	2.337067	
SWATH_THS	SNP					7.459488	9.369351	
						2.896947	5.460894	
	WGS					7.643293	11.6173	
						2.199452	6.059968	

Table S4.18: Length of ROHs found in Swamp Buffalo

LD Pruning Target	Species	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
ALL	SNP		8.859883	8.10881	10.64231	6.865855	4.697449	4.409093
			0.925171	3.202039	7.012179	5.639626	3.518486	3.238162
	WGS			7.153756	7.50774	3.149178	2.533506	2.355389
				2.986621	3.0692	1.43917	1.175823	1.120654
RIVER	SNP		7.43469	10.27233	8.298671	5.298731	4.685899	
			1.26273	6.126037	5.56217	3.962302	3.490649	
	WGS		7.423859	10.31546	6.825857	3.812199	3.437852	
			1.721703	5.049682	3.169043	1.900606	1.730057	
RIVIT_MED	SNP				8.65044	9.141854	6.233638	
					2.274464	5.088793	4.354573	
	WGS				7.907126	8.297564	5.445422	
					2.152215	3.48878	2.72638	
RIVPK_NIL	SNP				7.705773	7.262316	4.917264	
					2.208629	4.854226	3.74569	
	WGS				10.08309	10.40986	5.30314	
					2.641715	4.807705	2.779106	
SWAMP	SNP		7.529808	11.96813	12.7516	10.15265	9.208415	
			2.217859	7.718384	8.431463	7.976811	7.043697	
	WGS		5.804966	11.86213	7.695908	4.277271	3.976708	
			0.395222	6.581165	2.721006	1.605221	1.532613	
SWACN_GUI	SNP				7.685149	10.51374	11.23017	
						5.94526	6.583303	
	WGS				8.060783	7.788592	5.414443	
					3.076988	3.271926	2.18514	
SWATH_THS	SNP					7.637219	10.45389	
						3.015344	6.786465	
	WGS					8.625152	10.945	
						4.391659	4.585062	

Table S4.19: P Values for t.tests for Length of ROHs

LD Dataset	R2 Threshold	Between Species		Between Data	
		SNP	WGS	River	Swamp
ALL	1.00	**< 0.001**	**0.040**	**< 0.001**	**< 0.001**
	0.80	**< 0.001**	0.069	**< 0.001**	**< 0.001**
	0.50	**< 0.001**	**0.043**	**< 0.001**	**< 0.001**
	0.20	0.850	0.194	**0.003**	**0.007**
	0.10	0.593	0.452	**0.029**	0.121
RIVER	0.80	**< 0.001**	**0.038**	**< 0.001**	**< 0.001**
	0.50	**< 0.001**	0.064	**< 0.001**	**< 0.001**
	0.20	**< 0.001**	0.973	**< 0.001**	**< 0.001**
	0.10	0.118	0.303	0.669	0.561
RIVIT_MED	0.80	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
	0.50	**< 0.001**	0.386	**0.001**	**0.019**
	0.20		0.915		0.408
RIVPK_NIL	0.80	**< 0.001**	**0.006**	0.853	**< 0.001**
	0.50	**< 0.001**	0.126	**< 0.001**	**< 0.001**
	0.20	0.198	0.876	0.091	**0.006**
SWAMP	0.80	**0.002**	0.937	**< 0.001**	**< 0.001**
	0.50	**0.005**	0.982	**< 0.001**	**< 0.001**
	0.20	0.850	0.271	**< 0.001**	**< 0.001**
	0.10	0.226	0.271	0.575	0.673
SWACN_GUI	0.80	0.167	0.603	**< 0.001**	**< 0.001**
	0.50	0.427	0.672	**< 0.001**	**0.001**
	0.20		0.997	0.149	
SWATH_THS	0.80	0.189	0.651	**< 0.001**	0.091
	0.50	0.973	0.851	0.332	0.407

Table S4.20: Number of 1-2Mb ROHs in River Buffalo per individual

Species	LD Pruning Target	Data	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
River	ALL	SNP	0.000	0.000	0.000	0.013	0.013	0.039	0.039
		WGS	0.000	0.000	0.000	0.000	1.279	8.163	12.605
	RIVER	SNP	0.000	0.000	0.013	0.000	0.039	0.053	
		WGS	0.000	0.000	0.000	0.047	0.023	0.047	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.026	0.026	
		WGS	0.000	0.000	0.000	0.000	0.023	0.000	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	0.026	0.000	
		WGS	0.000	0.000	0.000	0.000	0.000	0.000	
	SWAMP	SNP	0.000	0.000	0.000	0.013	0.000	0.013	
		WGS	0.000	0.000	0.000	0.023	0.023	0.047	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.000	0.039	
		WGS	0.000	0.000	0.000	0.000	0.023	0.000	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.000	0.013	
		WGS	0.000	0.000	0.000	0.000	0.000	0.047	
Swamp	ALL	SNP	0.000	0.000	0.000	0.000	0.099	0.155	0.268
		WGS	0.000	0.000	0.000	0.000	0.838	6.865	10.459
	RIVER	SNP	0.000	0.000	0.028	0.056	0.141	0.254	
		WGS	0.000	0.000	0.000	0.000	0.135	0.351	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.085	0.099	
		WGS	0.000	0.000	0.000	0.000	0.000	0.000	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	0.042	0.169	
		WGS	0.000	0.000	0.000	0.000	0.000	0.054	
	SWAMP	SNP	0.000	0.000	0.000	0.014	0.000	0.014	
		WGS	0.000	0.000	0.000	0.000	0.000	0.000	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.000	0.042	
		WGS	0.000	0.000	0.000	0.000	0.027	0.000	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.000	0.000	
		WGS	0.000	0.000	0.000	0.000	0.000	0.000	

Table S4.21: Number of 2-4Mb ROHs per individual

Species	LD Pruning Target	Data	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
River	ALL	SNP	0.000	0.000	0.026	0.237	0.513	3.855	5.395
		WGS	0.000	0.000	0.186	0.070	10.814	11.907	11.860
	RIVER	SNP	0.000	0.000	0.145	0.092	1.566	3.553	
		WGS	0.000	0.000	0.000	0.163	5.279	7.233	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.184	0.566	
		WGS	0.000	0.000	0.000	0.000	0.023	0.581	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	0.197	2.921	
		WGS	0.000	0.000	0.000	0.000	0.023	1.209	
	SWAMP	SNP	0.000	0.000	0.132	0.158	0.118	0.171	
		WGS	0.000	0.000	0.140	0.000	8.977	12.163	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.158	0.171	
		WGS	0.000	0.000	0.000	0.023	0.116	3.907	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.013	0.316	
		WGS	0.000	0.000	0.000	0.000	0.023	0.256	
Swamp	ALL	SNP	0.000	0.000	0.000	0.268	1.859	17.648	25.282
		WGS	0.000	0.000	0.135	0.000	8.216	8.243	8.000
	RIVER	SNP	0.000	0.000	0.451	0.521	10.549	19.169	
		WGS	0.000	0.000	0.162	0.351	10.784	14.757	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.634	4.183	
		WGS	0.000	0.000	0.000	0.000	0.108	2.946	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	1.296	15.817	
		WGS	0.000	0.000	0.000	0.000	0.162	3.892	
	SWAMP	SNP	0.000	0.000	0.085	0.113	0.211	0.211	
		WGS	0.000	0.000	0.054	0.027	2.676	3.324	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.169	0.155	
		WGS	0.000	0.000	0.000	0.027	0.027	0.865	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.000	0.254	
		WGS	0.000	0.000	0.000	0.000	0.000	0.108	

Table S4.22: Number of 4-8Mb ROHs in River Buffalo per individual

Species	LD Pruning Target	Data	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
River	ALL	SNP	0.000	0.013	0.421	0.789	2.776	3.908	4.263
		WGS	0.000	0.000	0.581	2.442	2.884	2.186	2.093
	RIVER	SNP	0.000	0.066	0.487	1.224	3.526	3.987	
		WGS	0.000	0.023	0.233	2.116	2.442	2.233	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.763	2.421	
		WGS	0.000	0.000	0.000	0.116	0.953	3.047	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.171	2.092	3.829	
		WGS	0.000	0.000	0.000	0.000	0.372	3.791	
	SWAMP	SNP	0.000	0.132	0.750	0.474	1.303	1.789	
		WGS	0.000	0.140	0.698	4.977	5.837	5.558	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.066	0.803	0.855	
		WGS	0.000	0.000	0.000	1.140	5.674	9.140	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.368	0.842	
		WGS	0.000	0.000	0.000	0.000	1.163	0.581	
Swamp	ALL	SNP	0.000	0.000	0.423	0.465	6.577	12.056	12.986
		WGS	0.000	0.000	0.649	1.784	1.784	1.486	1.378
	RIVER	SNP	0.000	0.338	1.718	7.563	14.718	13.380	
		WGS	0.000	0.108	0.865	4.568	3.973	3.297	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.085	4.324	14.000	
		WGS	0.000	0.000	0.000	0.486	3.189	7.541	
	RIVPK_NIL	SNP	0.000	0.000	0.000	1.197	11.592	14.577	
		WGS	0.000	0.000	0.000	0.054	1.108	8.216	
	SWAMP	SNP	0.000	0.155	0.493	0.437	1.704	2.606	
		WGS	0.000	0.054	0.297	1.432	2.081	1.811	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.014	0.507	0.493	
		WGS	0.000	0.000	0.000	0.432	1.514	2.243	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.408	0.521	
		WGS	0.000	0.000	0.000	0.000	0.270	0.243	

Table S4.23: Number of 8-16Mb ROHs in River Buffalo per individual

Species	LD Pruning Target	Data	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
River	ALL	SNP	0.000	0.000	0.237	0.829	1.487	1.526	1.566
		WGS	0.000	0.000	0.233	1.302	0.233	0.140	0.116
	RIVER	SNP	0.000	0.026	0.684	1.434	1.513	1.513	
		WGS	0.000	0.023	0.488	0.535	0.256	0.163	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.013	1.329	1.447	
		WGS	0.000	0.000	0.000	0.116	0.860	0.442	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.092	1.395	1.526	
		WGS	0.000	0.000	0.000	0.070	1.233	0.628	
	SWAMP	SNP	0.000	0.013	0.658	1.224	1.421	1.421	
		WGS	0.000	0.023	1.233	1.837	0.860	0.767	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.947	1.171	
		WGS	0.000	0.000	0.000	0.767	2.512	1.442	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.184	0.789	
		WGS	0.000	0.000	0.000	0.000	0.930	2.116	
Swamp	ALL	SNP	0.000	0.028	0.239	0.662	1.859	1.803	1.958
		WGS	0.000	0.000	0.432	0.649	0.216	0.135	0.135
	RIVER	SNP	0.000	0.197	2.718	4.056	2.211	1.958	
		WGS	0.000	0.054	1.216	1.270	0.432	0.459	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.099	5.408	2.901	
		WGS	0.000	0.000	0.000	0.486	2.270	1.189	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.775	3.225	2.000	
		WGS	0.000	0.000	0.000	0.243	3.622	1.054	
	SWAMP	SNP	0.000	0.085	0.606	1.127	1.479	1.577	
		WGS	0.000	0.000	0.486	0.757	0.216	0.162	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.789	1.113	
		WGS	0.000	0.000	0.000	0.216	0.649	0.378	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.197	0.789	
		WGS	0.000	0.000	0.000	0.000	0.162	1.162	

Table S4.24: Number of >16Mb ROHs in River Buffalo per individual

Species	LD Pruning Target	Data	r^2 Threshold 0.01	0.05	0.1	0.2	0.5	0.8	1
River	ALL	SNP	0.000	0.000	0.013	0.329	0.579	0.487	0.487
		WGS	0.000	0.000	0.000	0.093	0.000	0.000	0.000
	RIVER	SNP	0.000	0.000	0.382	0.579	0.513	0.487	
		WGS	0.000	0.000	0.070	0.023	0.000	0.000	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.513	0.605	
		WGS	0.000	0.000	0.000	0.000	0.070	0.000	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	0.605	0.500	
		WGS	0.000	0.000	0.000	0.000	0.163	0.000	
	SWAMP	SNP	0.000	0.000	0.368	0.632	0.645	0.645	
		WGS	0.000	0.000	0.279	0.209	0.023	0.000	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.289	0.408	
		WGS	0.000	0.000	0.000	0.000	0.209	0.093	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.013	0.289	
		WGS	0.000	0.000	0.000	0.000	0.000	0.605	
Swamp	ALL	SNP	0.000	0.000	0.014	0.310	0.521	0.549	0.549
		WGS	0.000	0.000	0.000	0.081	0.000	0.000	0.000
	RIVER	SNP	0.000	0.000	0.676	0.563	0.507	0.521	
		WGS	0.000	0.000	0.270	0.081	0.027	0.027	
	RIVIT_MED	SNP	0.000	0.000	0.000	0.000	0.592	0.577	
		WGS	0.000	0.000	0.000	0.000	0.108	0.135	
	RIVPK_NIL	SNP	0.000	0.000	0.000	0.000	0.648	0.521	
		WGS	0.000	0.000	0.000	0.000	0.432	0.216	
	SWAMP	SNP	0.000	0.000	0.310	0.479	0.493	0.507	
		WGS	0.000	0.000	0.216	0.054	0.000	0.000	
	SWACN_GUI	SNP	0.000	0.000	0.000	0.000	0.310	0.408	
		WGS	0.000	0.000	0.000	0.000	0.108	0.000	
	SWATH_THS	SNP	0.000	0.000	0.000	0.000	0.014	0.296	
		WGS	0.000	0.000	0.000	0.000	0.054	0.162	

Figure S4.1: Posterior Distributions for River to swamp migration

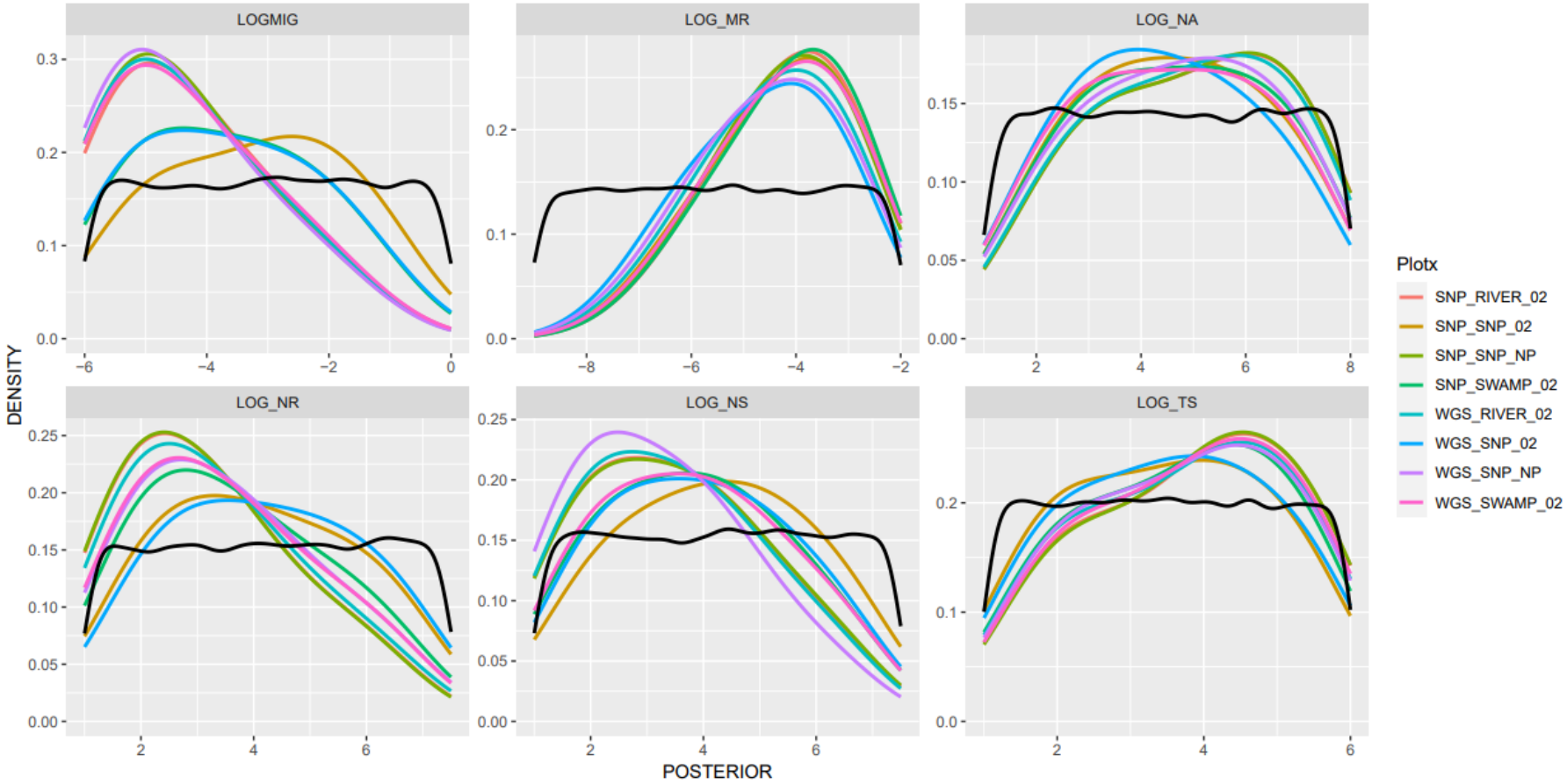
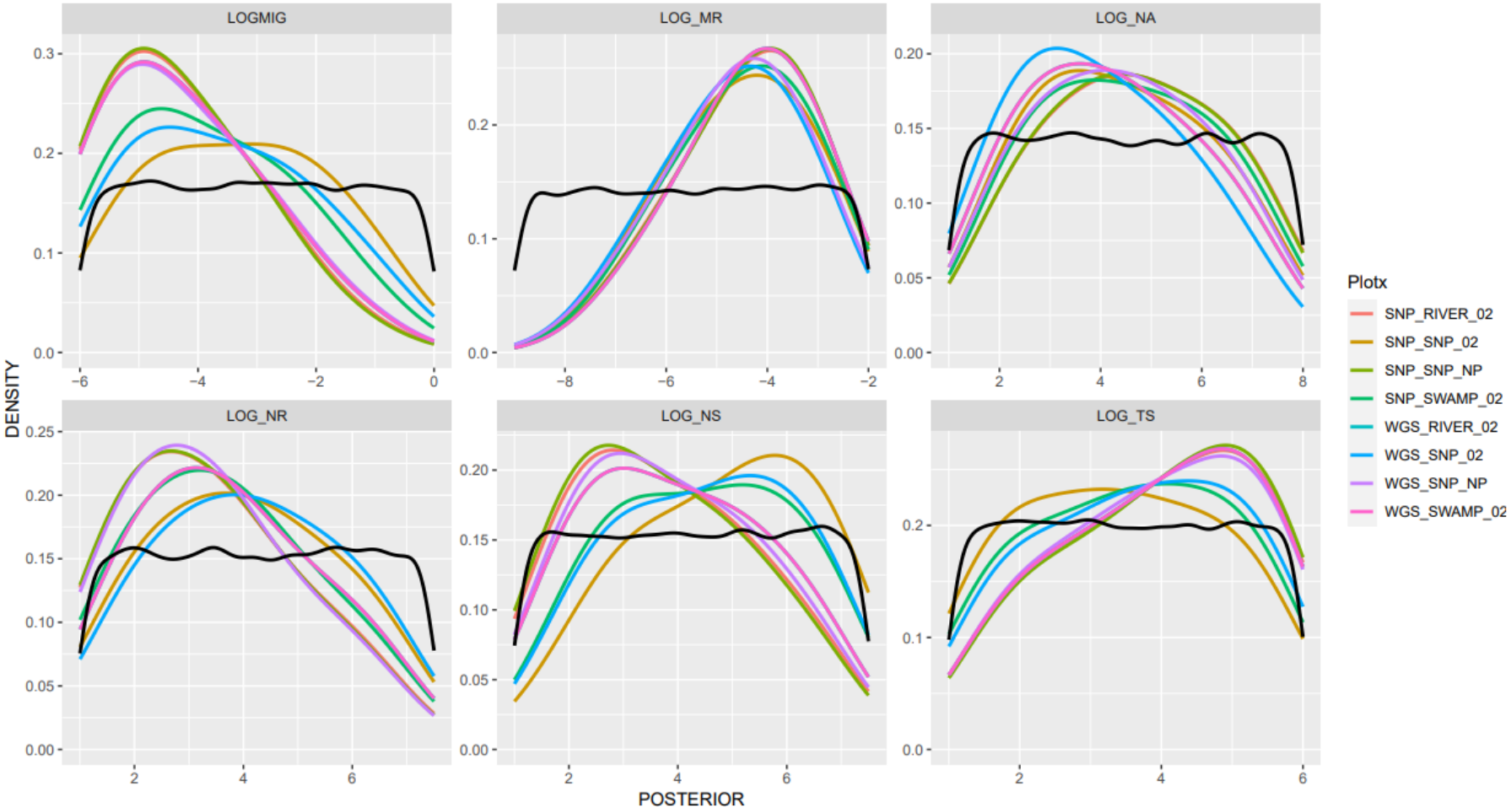


Figure S4.2: Posterior Distribution for Swamp to River migration



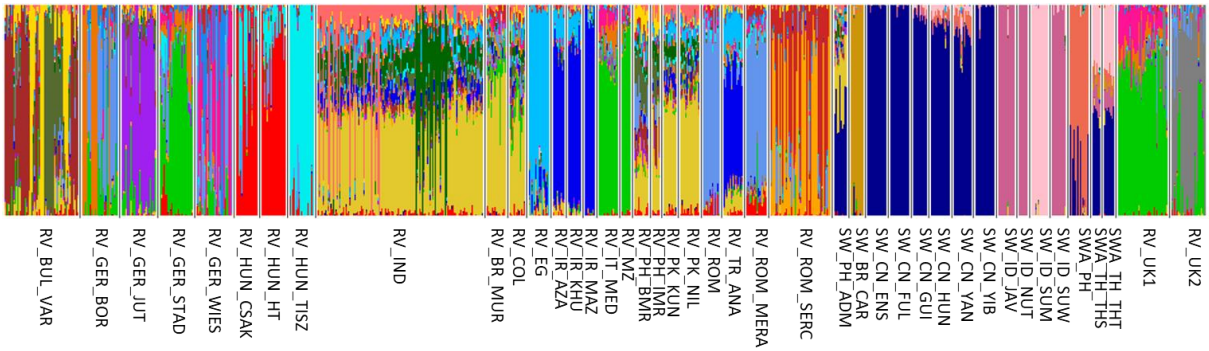
Supplementary Table S5.1: ABC parameter defined and priors used.

Parameter	Description	Scale	Minimum	Maximum
Ne_India	Effective population size of India	Log_{10}	2	7
Ne_Iran	Effective population size of Iran	Log_{10}	2	7
Ne_Italy	Effective population size of Italy	Log_{10}	2	7
Ne_Swamp	Effective population size of Swamp	Log_{10}	1	5
Ne_River1	Effective population size of first ancestor river population	Log_{10}	2	8
Ne_River2	Effective population size of second ancestral river population	Log_{10}	2	8
Ne_Ancestor	Effective population size of pre river and swamp ancestor	Log_{10}	1	7
T_River1	Time of split for most recent river populations	Log_{10}	0	3.3
T_River2	Time of split for most first river populations	Log_{10}	0	3.3
T_River_Swamp	Time of split for river and swamp buffaloes	Log_{10}	2.22	7
MutRate	Mutation Rate	Log_{10}	-8	-2
Mig_India_Iran	Migration between Indian and Iranian buffaloes	Log_{10}	-3	0
Mig_River_Swamp	Migration between river and swamp buffaloes	Log_{10}	-6	0
Mig_Swamp_River	Migration between swamp and river buffaloes	Log_{10}	-6	0

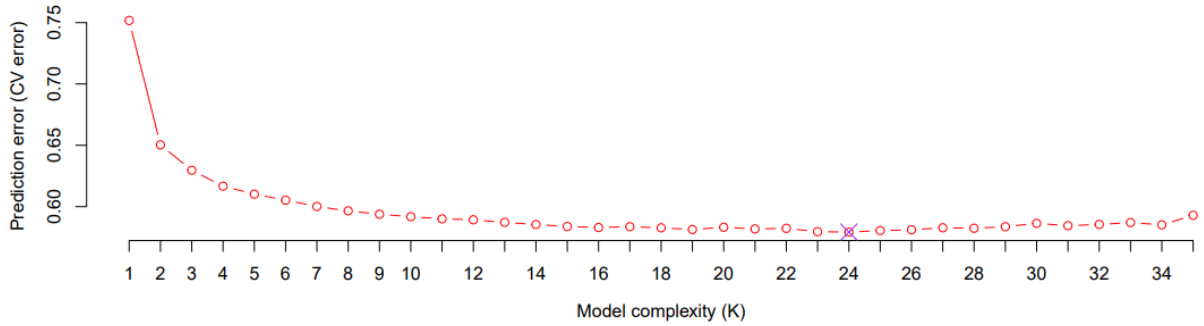
Supplementary Table S5.2: Summary statistics calculated for ABC posterior estimation. Greyed out rows were not used. Population numbers correspond to: India (1), Iran (2), Italy (3), and China (4).

Statistic	Description	Observed Data
K_1	Mean number of alleles over loci in population 1	1.991
K_2	Mean number of alleles over loci in population 2	1.958
K_3	Mean number of alleles over loci in population 3	1.954
K_4	Mean number of alleles over loci in population 4	1.879
Ksd_1	Standard deviation over loci in population 1 of the mean number of alleles	0.097
Ksd_2	Standard deviation over loci in population 2 of the mean number of alleles	0.202
Ksd_3	Standard deviation over loci in population 3 of the mean number of alleles	0.210
Ksd_4	Standard deviation over loci in population 4 of the mean number of alleles	0.326
mean_K	Mean number of alleles over loci of all populations	1.945
sd_K	Standard deviation of the mean number of alleles over all populations	0.047
tot_K	Mean total number of alleles	2.000
H_1	Mean heterozygosity over loci in population 1	0.413
H_2	Mean heterozygosity over loci in population 2	0.386
H_3	Mean heterozygosity over loci in population 3	0.367
H_4	Mean heterozygosity over loci in population 4	0.297
Hsd_1	Standard deviation over loci in population 1 of heterozygosity	0.118
Hsd_2	Standard deviation over loci in population 2 of heterozygosity	0.147
Hsd_3	Standard deviation over loci in population 3 of heterozygosity	0.150
Hsd_4	Standard deviation over loci in population 4 of heterozygosity	0.177
mean_H	Mean heterozygosity over loci of all populations	0.366
sd_H	Standard deviation of the heterozygosity over loci in all populations	0.049
tot_H	Mean total heterozygosity	0.441
FST_2_1	Pairwise F_{ST} between populations 1 and 2	0.050
FST_3_1	Pairwise F_{ST} between populations 1 and 3	0.114
FST_3_2	Pairwise F_{ST} between populations 2 and 3	0.128
FST_4_1	Pairwise F_{ST} between populations 1 and 4	0.288
FST_4_2	Pairwise F_{ST} between populations 2 and 4	0.318
FST_4_3	Pairwise F_{ST} between populations 3 and 4	0.331
PI_2_1	Mean number of differences between populations 1 and 2	759.710
PI_3_1	Mean number of differences between populations 1 and 3	792.624
PI_3_2	Mean number of differences between populations 2 and 3	778.250
PI_4_1	Mean number of differences between populations 1 and 4	900.869
PI_4_2	Mean number of differences between populations 2 and 4	897.710
PI_4_3	Mean number of differences between populations 3 and 4	896.220

Supplementary Figure S5.3: Best fitting value of K in ADMIXTURE (K = 24)



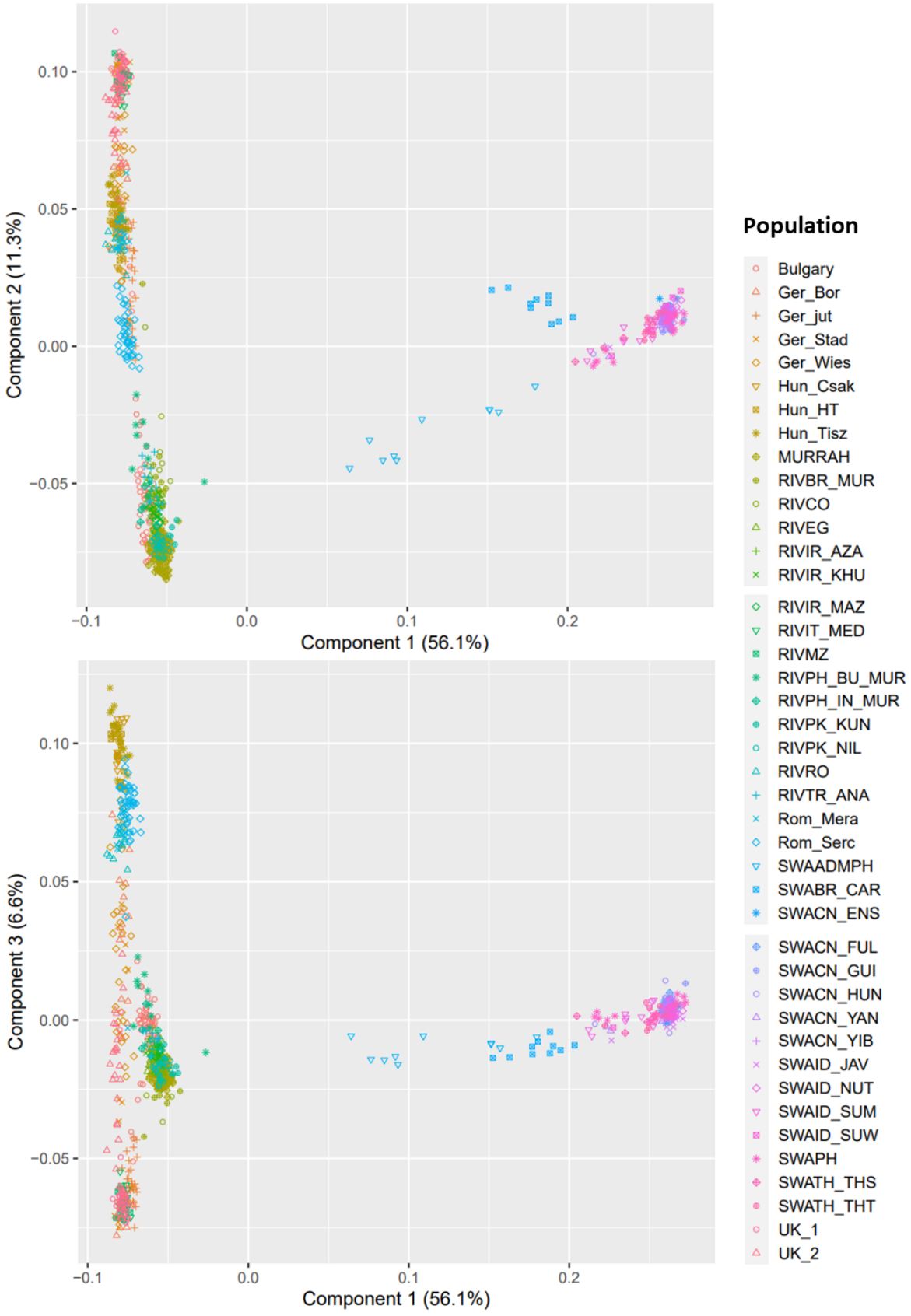
Supplementary Figure S5.4: CV values across values of K tested in ADMIXTURE.



Supplementary Table S5.5: Treemix outputs showing variance explained by each number of migrations included and their log likelihoods. m=2 surpassed the 99.8% threshold.

m	Variance Explained	M sign	Percentage	0 Log Likelihood	M Log Likelihood
0	0.99737	0	0	915.392	915.392
1	0.99774	1	100	906.413	1344.47
2	**0.99853**	**2**	**100**	**915.392**	**3804.48**
3	0.99887	3	100	915.392	4335.2
4	0.99901	4	100	906.413	4542.64
5	0.99922	5	100	892.017	4635.35
6	0.99909	6	100	915.392	4700.76
7	0.99920	7	100	915.392	4926.06
8	0.99946	8	100	915.392	5019.7
9	0.99945	9	100	906.413	5052.11
10	0.99960	10	100	915.392	5351.08
11	0.99959	11	100	915.392	5314.49
12	0.99968	12	100	892.017	5562.1
13	0.99972	13	100	915.392	5643.63
14	0.99969	14	100	892.017	5538.75
15	0.99976	15	100	915.392	5703.14

Supplementary Figure S5.6: MDS analysis showing components 1, 2, and 3



Supplementary Table S5.7: AMOVA result

Source of variation	Degrees of freedom	Sum of squares	Variance component	Percentage variation
Among groups	1	505711.960	887.32416	23.15939
Among populations within groups	40	505403.179	264.05577	6.89192
Among individuals within populations	775	2054294.077	-29.03842	-0.75791
Within individuals	815	2213116.500	2709.03760	70.70659
Total		5278525.742	3831.37910	

Supplementary Table S5.8: Genes found under balancing and divergent selection between river and swamp buffaloes. * Indicates nearest gene within 100,000bp as opposed to SNP featuring within genes

Inferred Selection	SNP	Chromosome	Position	H_O	F_{ST}	P-value	Gene
Divergent	AX-85059449	1	137805119	0.532	0.674	0.043	NLGN1
	AX-85074379	2	117197387	0.496	0.668	0.021	NCKAP5
	AX-85130617	5	86875705	0.718	0.856	0.028	DLG2
	AX-85098108	5	123612840	0.315	0.534	0.049	FGF3*
	AX-85068609	7	13135978	0.461	0.644	0.028	-
	AX-85053103	7	29497048	0.521	0.667	0.038	-
	AX-85092885	7	59228951	0.516	0.654	0.039	TMEM156
	AX-85053669	7	59261804	0.511	0.676	0.021	TMEM156
	AX-85141759	7	76125508	0.480	0.629	0.042	KCNIP4
	AX-85123264	7	87537289	0.348	0.552	0.043	CAMK2D
	AX-85100415	7	113766499	0.394	0.569	0.049	MARCH1
	AX-85122178	8	89393319	0.607	0.731	0.035	-
	AX-85064549	9	36156687	0.465	0.679	0.011	-
	AX-85078335	10	47108337	0.594	0.726	0.038	-
	AX-85123889	14	7117671	0.592	0.724	0.035	-
	AX-85115030	14	34586497	0.416	0.597	0.033	BMP2*
	AX-85092242	14	34644384	0.417	0.598	0.033	-
	AX-85080355	19	1027972	0.419	0.593	0.040	LOC112580673
	AX-85095904	19	4930295	0.420	0.638	0.015	-
	AX-85173042	19	19062193	0.418	0.594	0.041	PDE4D
	AX-85109455	19	19090790	0.424	0.640	0.017	PDE4D
	AX-85156949	20	40534232	0.562	0.703	0.034	PCSK6
	AX-85109516	21	3721217	0.565	0.701	0.032	RBMS3
	AX-85080769	24	31495868	0.630	0.745	0.050	SNX29
	AX-85043173	24	38742317	0.391	0.576	0.039	UBALD1*
Balancing	AX-85077866	1	55546207	0.229	0.023	0.047	LOC112583954
	AX-85108159	1	59994964	0.378	0.014	0.038	NCAM2
	AX-85068810	1	85408765	0.328	0.030	0.049	-
	AX-85114914	1	121168884	0.493	0.016	0.046	LOC112585195
	AX-85090080	1	134494382	0.321	0.026	0.043	LOC112585597
	AX-85159363	1	150365153	0.498	0.012	0.040	OTOL1*
	AX-85052477	2	25475840	0.267	0.010	0.035	NEU1*
	AX-85087584	2	37264933	0.476	0.000	0.034	TREM1

AX-85112910	2	41710900	0.339	0.029	0.047	CPNE5
AX-85105035	2	74209524	0.375	0.015	0.039	ATF2
AX-85120212	3	11177522	0.182	0.007	0.046	LOC102410316
AX-85058807	3	48001841	0.229	0.017	0.045	TMEM132E*
AX-85174824	3	74461115	0.280	0.006	0.035	ACO1
AX-85066277	3	74679688	0.484	0.011	0.040	-
AX-85088622	3	126549607	0.234	0.003	0.035	GABBR2
AX-85122185	4	13975944	0.335	0.009	0.034	CRACR2A
AX-85072461	4	71601048	0.242	0.018	0.044	WIF1*
AX-85083446	4	98709789	0.456	0.010	0.040	-
AX-85077110	4	155750012	0.404	0.028	0.048	-
AX-85067417	5	6622618	0.401	0.015	0.038	-
AX-85086486	5	27591985	0.452	0.017	0.046	KAZN
AX-85095818	5	50309636	0.499	0.014	0.042	SMYD3
AX-85064355	5	100456933	0.368	0.014	0.036	NAV2*
AX-85067692	5	119820501	0.197	0.013	0.049	MAJIN
AX-85169249	6	27619169	0.343	0.028	0.048	VANGL1*
AX-85119775	6	56879244	0.277	0.021	0.043	CLCA4
AX-85068645	7	23640037	0.476	0.003	0.033	-
AX-85075338	7	45668788	0.344	0.018	0.038	AASDH
AX-85153345	7	62586729	0.402	0.015	0.039	-
AX-85069405	7	66942094	0.444	0.010	0.039	-
AX-85063644	7	78250793	0.418	0.017	0.043	-
AX-85090615	8	6647576	0.237	0.025	0.050	-
AX-85081730	8	30931972	0.408	0.016	0.040	LOC112586464
AX-85109656	8	111682591	0.226	0.008	0.038	CNTNAP2
AX-85112337	9	21765795	0.311	0.024	0.046	-
AX-85082311	9	23945538	0.475	-0.005	0.030	-
AX-85124176	9	44887825	0.392	0.017	0.042	-
AX-85092905	9	55481311	0.356	0.034	0.047	-
AX-85092488	9	75375969	0.501	0.016	0.042	-
AX-85056106	11	4811606	0.370	0.000	0.028	-
AX-85051746	11	43792867	0.302	0.013	0.038	SPPL2A
AX-85133095	12	131865	0.337	0.008	0.034	ZC3H8
AX-85103904	12	3024058	0.433	0.017	0.040	TMEM131
AX-85057934	12	40144595	0.500	0.012	0.037	-
AX-85106510	12	77734826	0.227	0.019	0.048	-
AX-85175464	12	85135298	0.391	0.024	0.045	-
AX-85047589	13	70312558	0.275	0.021	0.043	RNASEH2B
AX-85047955	14	16762379	0.423	0.010	0.039	VSTM2L*
AX-85102486	15	43900948	0.481	0.004	0.037	CRISPLD1*
AX-85104025	16	55600588	0.345	0.008	0.033	TREH
AX-85101944	17	2084327	0.225	0.011	0.042	LOC102408479
AX-85053946	17	4390666	0.504	0.006	0.033	NF2
AX-85111891	17	41978525	0.501	0.015	0.045	-
AX-85176273	18	64601486	0.501	-0.001	0.031	LOC102399551
AX-85087562	19	10020060	0.494	0.012	0.042	OCLN
AX-85046689	19	14526456	0.335	0.012	0.037	SHISL2B*
AX-85123100	19	14719215	0.439	0.010	0.037	RGS7BP*
AX-85068339	19	68553602	0.446	0.005	0.034	-
AX-85070718	20	50558766	0.425	0.024	0.046	NTRK3
AX-85073019	21	20176221	0.293	0.009	0.036	-
AX-85077610	22	38659850	0.301	0.027	0.048	ASXL3
AX-85112930	22	39062144	0.502	0.022	0.041	NOL4
AX-85042968	23	16533814	0.324	0.017	0.038	PDLIM1*