Phylogenetic Analysis and Population Structure of the New Zealand Huntaway and Heading Dog

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science

in

Animal Science

At Massey University Al Rae Centre, Hamilton, New Zealand.

Millicent Frances Henry

2025

Abstract

The New Zealand Huntaway and Heading Dog are vital to the success of New Zealand's livestock farming industry. This study presents one of the first large-scale genomic analyses of these two unique working breeds. Utilising phylogenetic and population structure methodology, the primary objectives of this research were to investigate the genetic relationships of the New Zealand Huntaway and Heading Dog with international dog breeds, formalise the major historical breed contributions that have shaped their development, and assess their population structure and diversity. This study utilised a dataset of 1,735 individuals, representing 129 dog breeds and three wild canid species, and included 211 Huntaways and 246 Heading Dog genomic datasets generated as part of this project.

Genomic analyses revealed that both breeds can be traced back to ancestral UK rural working dog populations, occupying positions within the UK rural subclade on the domestic canine phylogeny. Baseline inbreeding scores suggest that neither the New Zealand Huntaway or Heading Dog populations express significantly high inbreeding levels, and the scores obtained are similar to those seen in other formally recognised dog breeds. A list of genes under putative positive selection in both populations is presented and could be used for future genetic selection in these dog populations.

This study highlights the Huntaway and Heading Dog as genetically distinct breeds adapted to the unique demands of New Zealand's pastoral farming systems. These results not only contribute to the current understanding of breed development in geographically isolated environments such as New Zealand but also highlights the opportunity of integrating genomics into working dog management and breeding practices. The genetic resources generated through this study will serve as a valuable reference for future research in the "Right Dog for the Job" project, aimed at preserving the health, diversity and functionality of the New Zealand Huntaway and Heading Dog.

Acknowledgements

A big thank you to my supervisors, Dr. Thomas Lopdell, Professor Matt Littlejohn, and Dr. Nick Sneddon. Thomas, I really appreciated your support, our weekly catch-ups, coding help, and feedback on my drafts. Matt, thank you for your guidance and the opportunity to contribute to the “Right Dog for the Job” project. Nick, thank you for your feedback and support throughout.

I would like to thank Massey University for supporting my research through the Master’s Research Scholarship, and Livestock Improvement Corporation (LIC) for the Pat Shannon Scholarship and use of the R&D office space. I’m grateful to the LIC R&D team for welcoming me and for your support and feedback. Thanks to Pam Stephen Photography for producing Figure 1. A special shoutout to my master’s buddy Florence Smith, thank you for the laughs, support, and riding out the stress with me. Short queens rule.

A big thank you to my Grandparents, Alan and Alison Henry, who have supported me throughout both my undergrad and postgraduate studies. Thank you for always being so interested in my endeavours and I have really appreciated your support throughout my university career. A special mention to my Nana, Frankie Smith, for teaching me to live life with a fearlessly authentic attitude, and the importance of staying true to myself, no matter what.

A massive thank you to my Partner Damien Martinus for being by my side and supporting me through the last 18 months.

Lastly, I would like to dedicate this thesis to my parents, Mark and Lucy Henry. Your support has carried me through every challenge, and you have both always been there for me no matter what. You have inspired me to pursue my passions with confidence, and none of this would have been possible without your ongoing love, encouragement and belief in me.

Animal ethics statement

All experiments were performed in accordance with the rules and guidelines outlined in the New Zealand Animal Welfare Act 1999. Samples were gathered in accordance with protocols approved by the Massey University Animal Ethics Committee (approval MUAEC 23/27). No animals were sacrificed for this study.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CV	Cross validation
DM	Degenerative Myelopathy
DNA	Deoxy-ribonucleic acid
E_{HOM}	Expected number of homozygous markers
eROHi	Extreme runs of homozygosity islands
F_{HET}	Inbreeding based on observed versus expected heterozygosity
F_{HOM}	Genome-wide homozygous excess due to inbreeding
F_{ROH}	Inbreeding coefficient based on runs of homozygosity
Gb	Gigabase
GDV	Gastric dilation-volvulus
GTC	Genotype call
GWAS	Genome wide association study
IBD	Identity-by-descent
IBS	Identity-by-state
IDAT	Illumina raw intensity
IQR	Inter-quartile range
Kb	Kilobase
Kg	Kilogram
L_{auto}	Total length of autosomal genome
LD	Linkage Disequilibrium
L_{ROH}	Sums of ROH lengths

MAF	Minor allele frequency
Mb	Mega base
Mbp	Mega base pair
MDS	Multidimensional scaling
MPS IIIA	Mucopolysaccharidosis IIIA
N	Total number of markers
n	Number
NCL	Neuronal Ceroid Lipofuscinosis
NJ	Neighbour Joining
NZ	New Zealand
O_{HOM}	Observed number of homozygous markers
PC	Principal Component
PCA	Principal component analysis
QC	Quality control
ROH	Runs of homozygosity
ROHi	Runs of homozygosity islands
SNP	Single nucleotide polymorphism
TNS	Trapped Neutrophil Syndrome
UK	United Kingdom
VCF	Variant Call Format
WGS	Whole genome sequence

Chapter 1: Introduction

Many of today's recognised dog breeds can be traced back to the Victorian era when selective breeding aimed at producing animals with specific aesthetic and behavioural traits desired by humans. In contrast, historical working breeds were developed with a focus on functional traits required for specific tasks, with far less priority on aesthetic traits (Ostrander et al., 2017; Parker et al., 2017). In a New Zealand context, an exact date cannot be placed on the introduction of farm working dogs to New Zealand, although their history is closely linked to the development of the country's agricultural sector in the early 1800's. As sheep and beef farming expanded across New Zealand's varied and often challenging terrain, there was a need for capable and versatile working dogs. In response to this, two unique working breeds, the New Zealand Huntaway and New Zealand Heading Dog, were developed specifically to support these farming operations. These breeds have been selectively bred over generations to perform distinct, yet complementary roles. The Huntaway is known for its loud bark and ability to drive livestock across long distances, often commanded from a distance as well. By contrast, the Heading Dog is known for its speed and aptitude for controlling livestock with quiet precision up close.

Today, these dogs remain integral to the success of New Zealand's agricultural industry. Without their contributions, many sheep and beef operations would be neither physically nor economically feasible. Despite their significant and ongoing contributions to agricultural practices in New Zealand, there has been limited scientific research addressing the origins, development and genetic composition of these two breeds. This contrasts with the progress made in characterising modern dog breeds globally through the application of phylogenetic and population structure analyses, which have revealed that factors such as geographic isolation, human-mediated migration and intensive artificial selection have played pivotal roles in shaping breed lineages and traits (Parker et al., 2017). Like many formally recognised breeds, New Zealand Huntaways and Heading Dogs are also prone to a range of health issues, and their distinct working environments, physical demands, diets and housing conditions all contribute to their overall health. Genetic predispositions may also contribute to certain diseases and performance-related limitations within these populations (Isaksen et al., 2020; Jerram, 2013).

In an effort to address these knowledge gaps, Massey University launched the "Right Dog for the Job" project, in partnership with Ministry for Primary Industries (Ministry for Primary Industries, 2025), Focus Genetics (Focus Genetics, 2025), Pāmu (Pāmu, 2025), Ancestry (Ancestry, 2024), Vetlife (Vetlife, 2024), Helical Company Limited (The Helical Company Limited, 2025), New Zealand Sheep Dog Trial Association (New Zealand Sheep Dog Trial Association, 2023) and Livestock Improvement Corporation (Livestock Improvement Corporation, 2025). The central aim of the "Right Dog for the Job" project was to explore the genetics of New Zealand's unique working breeds, with a main focus on improving their health and performance (Massey University, 2025). The primary objective of this study was to examine the genetic relationships between the New Zealand Huntaway and Heading Dog, and a range of formally recognised dog breeds from around the world. Towards this goal, my project aimed to identify and formalise the historical breed contributions that have influenced the development of the Huntaway and Heading Dog. In addition, baseline average inbreeding levels within the populations of these two breeds were calculated, providing an understanding of their current genetic diversity and population structure. Characterising the genetic makeup and breed composition of the Huntaway and Heading Dog provides valuable insights into shared health and performance related traits in these breeds and others with similar ancestry, ultimately supporting more informed breeding and management decisions.

Chapter 2: Literature Review

2.1 Domestication of Dogs

The domestic dog (*Canis lupus familiaris*) fills various roles in everyday human life, from serving as a companion animal, to performing specific working tasks such as herding, droving, guarding, retrieving and guidance. Archaeological evidence indicates that domesticated dogs first appeared 15,000 to 33,000 years ago, pre-dating the establishment of agriculture that occurred around 10,000 years ago (Wayne & Vonholdt, 2012). While dogs are known to have descended from wolves (Freedman & Wayne, 2017), the exact process of canid domestication is quite uncertain. One of the leading hypotheses suggests that the early association between humans and wolves was facilitated by human hunting practices (Mech & Janssens, 2022). Leftover carcasses from hunts likely attracted wolves, and these wolves became increasingly orientated to a human-niche. Wolves eventually established a mutualistic relationship with humans and coordinated activities such as hunting and defence (Wayne & Vonholdt, 2012). Early dogs were subject to a second wave of selection during the development of agriculture, with this selection mostly focused on behavioural traits such as docility, guarding and droving. A third (and the most dramatic) phase of selection in domestic dogs was the Victorian breed explosion which occurred around 200 years ago. Modern dog breeds observed today can largely be traced back to that era, and are the result of intense selective breeding for aesthetic traits such as body size, skeletal mutations (foreshortened limbs and pathologically short faces) and the coat colour and texture traits that are desired by humans (Freedman & Wayne, 2017; Wayne & Vonholdt, 2012). In contrast, historical working breeds were developed with a focus on functional traits required for specific tasks, such as herding or guarding, which are essential to their roles (Ostrander et al., 2017; Parker et al., 2017).

2.2 History and Profiles of New Zealand Working Farm Dogs

2.2.1 History of New Zealand Working Dogs

While an exact date cannot be placed on the arrival of working farm dogs in New Zealand, the history of New Zealand farm dogs coincides with the country's agricultural development. By the mid-1800s, Merino sheep had been introduced to New Zealand for their wool (Hughes, 2013). Alongside the extensive number of sheep imports throughout the 19th century, what were referred to as 'working collies' or 'shepherd working dogs' were also brought to New Zealand (The Border Collie Museum, 2014). Because official breed standards did not exist at this time, these original sheepdogs are thought to have had considerable variation in both appearance and working behaviour (The Border Collie Museum, 2013). Wool breeds of sheep, including the merino, tend to be flighty and are most effectively pushed by quiet herding dogs. However, the introduction of refrigerated shipping in 1882 dramatically altered farming practices, shifting farmers' focus from wool production to meat exports. This shift saw the replacement of flighty wool breeds, with more docile meat breeds, such as Leicester and Romney. These breeds were harder to manage using traditional quiet herding dogs, necessitating the selective breeding of dogs capable of pushing livestock with noise (Hughes, 2013).

A notable figure in the history of New Zealand farm dog breeding is James Lilico, who was active in sheepdog breeding and handling throughout his life. Lilico made a significant impact on the development of the Heading Dog breed by introducing renowned Border Collie bloodlines to New Zealand, including that of Old Hemp (Rennie, 1984). Bred in Northumberland in 1893, Old Hemp was an early Border Collie that became well known for his exceptional herding abilities, particularly in his distinctive herding style that relied on body positioning and an intense stare. The Border Collie is a British herding breed that originated from Scotland and Northern England, but did not achieve formal breed recognition until 1977 since these dogs were originally bred to work, rather than to conform to show standards (Ács et al., 2019). Old Hemp's working reputation made him a highly sought-after stud dog, and it is believed he sired around 200 offspring (Austin, 2018). After Lilico arrived in New Zealand from Northumbria in 1895, he recognised the potential of Border Collies to meet the unique demands of New Zealand's livestock farming. His contributions led to a substantial boost of Border Collie genetics into the Heading Dog population, refining their abilities for silent herding and livestock control (Rennie, 1984). Lilico's contribution to New Zealand Sheepdog breeding supports the anecdotal evidence that the foundation of New Zealand's two primary farm dog breeds lies in the Scottish Working/Border Collies and Bearded Collies brought to New Zealand by European settlers in the 19th century (Dalton, 2009; Hughes, 2013). The expansion of European settlement in New Zealand also resulted in the importation of other working dog breeds, including Smithfield Collies, Scottish Hunting Dogs, Foxhounds, Labradors and Rottweilers. It is thought that some of these breeds may have contributed to the development of the Huntaway, particularly in enhancing its barking ability and robust build (Frank Film, 2019; Hughes, 2013; Jerram, 2013).

As of 2023, New Zealand has approximately 22,005 sheep and beef farms, accounting for 47% of the total farms in the country and a sheep population of about 25.3 million (Beef + Lamb New Zealand, 2023). New Zealand's livestock farms, particularly sheep and beef farms, are often located in areas with challenging terrain. This makes farm work physically demanding, requiring farm dogs in New Zealand to be purposefully bred to suit these strenuous conditions, with selection for both physical and behavioural traits (Jerram, 2013). As of 2023, the number of registered Heading Dogs in New Zealand is 20,000 (Figure NZ, 2023a) and the number of registered New Zealand Huntaways (the second most popular dog breed in the country) is around 30,000 (Figure NZ, 2023b). These numbers likely underrepresent the total population size of these dogs, as many farm dogs are not registered.

Figure 1: Photographs of the New Zealand Huntaway (A) and New Zealand Heading Dog (B). Photography by Pam Stephen Photography.

2.2.3 New Zealand Huntaway Profile

The New Zealand Huntaway, which makes up approximately 70% of the country's farm dog population (Hughes, 2013), is a barking dog bred to push livestock using noise. Formally recognised as a breed in 2013 (Dogs New Zealand, 2013), it was first mentioned in the Otago Daily Times in 1884 (Frank Film, 2019) and was developed to better handle docile sheep breeds. When working, these dogs bark continuously to effectively direct livestock in a preferred direction. Some farmers also employ Huntaway's for yard work. Huntaways 'back' livestock, which involves running over the backs of livestock to push them forward, and 'head', or move to the front of a flock/herd to turn or hold them (Isaksen et al., 2020; Jerram, 2013). These robust dogs typically have short black and tan coats, large heads and deep chests, with an average weight of 28kg. Like the Heading Dog, the New Zealand Huntaway lacks a strict breed standard, as it has been bred for performance rather than appearance (Dogs New Zealand, 2013).

2.2.2 New Zealand Heading Dog Profile

The New Zealand Heading Dog is a dog that works silently to gather livestock. These dogs rely on a “staring eye” and positioning to control sheep. Furthermore, these dogs are often trained to ‘heel’ where they bite below an animal’s hock, and ‘cover’ to collect livestock together (Jerram, 2013). New Zealand Heading Dogs are agile and fast, with a typically black, white and tan smooth coat, often resembling short-haired Border Collies (Isaksen et al., 2020). Their phenotype is variable as they are primarily selected for performance over physical appearance and have an average weight of 19kg. The New Zealand Heading Dog does not have an official breed standard and is not recognised as an official breed (Wag!, 2024).

2.2.4 New Zealand Farm Dog Health

The genetic variability within domestic animals is largely influenced by breeding decisions and practices (Leroy et al., 2006). Modern dog breeds are often characterised by a limited number of founders selected primarily for aesthetic traits such as morphology, size and coat colour. Historical population bottlenecks, the reliance on popular sires and restricted breeding practices have further influenced the development of modern dog breeds. Furthermore, the emphasis on selecting for aesthetic traits has led to a loss of genetic diversity within breeds, inevitably increasing inbreeding levels (Bannasch et al., 2021; Menor-Campos, 2024). Inbreeding can have negative consequences, as it leads to long homozygous regions in the genome. This can increase the likelihood that harmful genetic variants will affect health and overall fitness, due to deleterious variants often showing a recessive mode of inheritance and therefore only affect homozygous individuals (Mata & Mata, 2023). Inbreeding in the New Zealand Huntaway and Heading Dog has not been previously studied. Unlike most registered modern breeds, the Huntaway and Heading Dog have been bred for their functionality rather than aesthetic traits (Jerram, 2013). Because of this, these populations may retain higher genetic variation due to less constrained breeding practices. However, the establishment of sheepdog trials in New Zealand has led to certain high-performing dogs becoming sought after as stud dogs. Successful dogs are often bred extensively, with the use of frozen semen further amplifying their genetic influence. This effect is often referred to as the popular sire effect. Additionally, a population bottleneck occurred when the first herding dogs were introduced to New Zealand, which may have increased the frequency of some deleterious variants in the population due to genetic drift. Such bottlenecks, combined with the popular sire effect, can have lasting impacts on the genetic health of these populations.

Specific dog breeds have strong predispositions for certain inherited diseases. For example, Border Collies have 25 known inherited disorders, with common examples including Trapped Neutrophil Syndrome (TNS), characterised by static neutropenia and hyperplasia of myeloid cells in the bone marrow, and Neuronal Ceroid Lipofuscinosis (NCL), a neurodegenerative lysosomal storage disease characterised by an accumulation of intracellular lipopigments in neurons and other cells (Soh et al., 2021). Neuronal Ceroid Lipofuscinosis has been documented in over 20 canine breeds, including Australian Shepherds and Australian Cattle Dogs (Katz et al., 2017). The Online Mendelian Inheritance in Animals website currently catalogues 552 likely causal variants for disease and non-disease single-gene traits (Nicholas, 2025). Like many dog breeds, New Zealand Huntaways and Heading Dogs are prone to various health issues. Their unique working environment, high exercise demands, diets and housing conditions influence their overall health. Underlying genetic predispositions may also contribute to the prevalence of certain disease in these dogs (Isaksen et al., 2020; Jerram, 2013). The first large-scale investigation into the frequencies of diseases affecting working farm dogs in New Zealand was conducted by Cave et al. (2009). This study surveyed 2,214 farm dogs nationwide, consisting of 51% Huntaways, 39% Heading Dogs, and 10% other breeds. Of these cases, 46% of the veterinary cases were classed as non-traumatic health events, while 38% were classed as trauma-related issues.

Among the 1,024 non-traumatic cases, gastrointestinal issues were the most common, representing 20% of cases. Diagnoses included constipation, dental disease, gastric dilation-volvulus (GDV), vomiting, and foreign body ingestion. Gastric dilation-volvulus is a condition where the stomach twists and fills with gas which can be lethal and is typically seen in large dog breeds (Broome & Walsh, 2003). Skin diseases accounted for 15% of non-traumatic cases which included abscesses, pruritic conditions, and parasitic infections. Reproductive issues made up 19% of non-traumatic events, mis-mating being the most common issue. Huntaways were more prone to pyometra and vaginal prolapse. The high incidence of reproductive conditions is likely because the majority of New Zealand's working farm dog population are kept sexually intact for breeding purposes. The prevalence of mis-mating suggests the need for better management strategies, as this issue is avoidable. Musculoskeletal conditions made up 20% of non-traumatic cases with diagnoses including degenerative arthritis and hip dysplasia. Respiratory issues made up 3% of cases, and poisonings made up 4% of cases. The over-representation of certain non-traumatic events in Huntaways such as GDV, pyometra and vaginal prolapse may suggest there are genetic differences between Huntaways and Heading Dogs. Identifying anatomical, behavioural and genetic risk factors associated with these issues could help determine risk factors that cause these events in Huntaways. A total of 848 trauma related cases were reported in this study, 52% of which were in Heading Dogs and 43% in Huntaways. The most common injuries occurred in the foot, stifle and tarsus. 20% of traumatic cases were inflicted by livestock interaction, 19% caused by vehicle interaction and 12% caused by bites from other dogs (Cave et al., 2009).

The main objective for this phylogenetic study is to formalise the historic breed contributions and genetic diversity of the Huntaway and Heading Dog using SNP array genomic datasets. Another key objective of the wider "Right Dog for the Job" project is to identify genetic variants potentially linked to disease in the Huntaway and Heading Dog populations. As part of the broader "Right Dog for the Job" project, a short communication by Smith et al. (2025) has been released, reporting five high impact disease associated variants identified from 249 whole genome sequences (WGS) of Huntaways (n = 130), Heading Dogs (n = 104) and Huntaway/Heading Dog crossbreeds (n = 15). One of the variants reported in this study is a loss of function mutation in the *CUBN* (cubilin) gene. This gene encodes for the cubilin protein, which is essential for vitamin B12 absorption (Kristiansen et al., 1999). This mutation was discovered in Border Collies in 2013 with a minor allele frequency (MAF) of 1.5% and results in malabsorption of vitamin B12 in homozygotes (Owczarek-Lipska et al., 2013). Six carriers (heterozygotes) for this variant were identified out of the 249 dogs studied and all were Heading Dogs (Smith et al., 2025). The MAF for this variant in Heading Dogs was 2.9%, greater than that reported for Border Collies. If left untreated, vitamin B12 malabsorption can be fatal (Owczarek-Lipska et al., 2013). The frequency of carriers identified in this study highlights the need for further investigation in the wider New Zealand working dog population (Smith et al., 2025).

Another loss of function variant was identified in the *CLN8* gene (ceroid-lipofuscinosis, neuronal 8) that has been linked to a neurodegenerative disorder neuronal ceroid lipofuscinosis (NCL). *CLN8* encodes a membrane protein involved in trafficking materials within the endoplasmic reticulum (Lonka et al., 2000). Neuronal Ceroid Lipofuscinosis is characterised by motor dysfunction, seizures and blindness, and typically results in early euthanasia due to the severity and lack of treatment options (Guo et al., 2014). This variant was found in 21 dogs (13 Heading Dogs, seven Huntaways and one crossbreed), all of which were carriers. The MAF's for Heading Dogs and Huntaways were 6.3% and 2.7%, respectively (Smith et al., 2025), which are notably greater than those reported in other breeds such as the Australian Shepherd (0.67%) (Guo et al., 2014) and the German Shorthaired Pointer (0%) (Guo et al., 2019). While the exact number of affected dogs in New Zealand is unknown, the high frequency of carriers indicates that affected individuals do exist in the population but are likely euthanised at a young age. The variant should be included in genetic testing to prevent further spread of the disease (Smith et al., 2025).

A 1bp insertion variant in the *SGSH* (heparan sulfate sulfamidase) gene was also detected. This gene encodes sulfamidase, a lysosomal enzyme necessary for the degradation of heparan sulfate (a polysaccharide found in the extracellular matrix on cell surfaces) (Meyer et al., 2008). The variant results in enzyme deficiency in homozygotes which causes Mucopolysaccharidosis IIIA (MPS IIIA), a severe, untreatable neurodegenerative disorder. This condition has been previously reported in Huntaways in 2002 (Yogalingam et al., 2002). Five Huntaways in the study by Smith et al. (2025) were identified as carriers. However, four of these dogs were known carriers from a research colony previously used to study the disease and were therefore excluded from allele frequency calculation. Excluding these individuals, the MAF in Huntaways was calculated to be 0.4%, but earlier studies have reported a higher frequency of 3.8% (Smith et al., 2025).

A missense variant in the *SOD1* (superoxide dismutase 1) gene, responsible for encoding an antioxidant enzyme was also discovered. This variant is linked to Degenerative Myelopathy (DM), a degenerative disease of the spinal cord (Santos et al., 2020). In this study, 52 carriers were identified (46 Huntaways, four Heading Dogs and two crosses) and five homozygotes (all Huntaways). The MAF in Huntaways was 21.7%, but this is comparable to the MAF reported in German Shepherds (22%) (Santos et al., 2020). In contrast, the Heading Dog MAF was significantly lower at 1.9%. This mutation exhibits variable penetrance, meaning not all homozygous individuals develop the disease (Zeng et al., 2014). Given the high MAF in Huntaways, further investigation is warranted to assess the disease risk and prevalence in the wider population (Smith et al., 2025).

Lastly, a splice site variant was discovered in the *VWF* (von Willebrand factor) gene. This gene encodes a glycoprotein essential for blood clotting (Hassan et al., 2012). This variant has been proposed to cause type 1 von Willebrand disease, a condition characterised by prolonged bleeding and has been reported previously in other breeds (Crespi et al., 2018; Donner et al., 2016). This mutation was found in nine carriers (all Huntaways) and one homozygote (Huntaway crossbreed). The MAF in Huntaways was calculated at 2.2% (Smith et al., 2025), but it is suggested that this mutation exhibits incomplete penetrance (Crespi et al., 2018). Given the nature of physical work and the risk of injury in farm dogs, testing for this variant may help reduce the risk of adverse bleeding events in affected individuals (Smith et al., 2025).

The limited documentation on the development of the Huntaway and Heading Dog, combined with a lack of research into their health, highlights the need for further genetic investigation. Both breeds are affected by certain health conditions, some of which likely have a genetic basis. Characterising their genetic makeup and breed composition can reveal shared health and performance related traits within these breeds and among other closely related breeds. In addition, assessing inbreeding levels will provide valuable insights into their genetic diversity and population structure. This knowledge will support more informed breeding decisions and improved management strategies for shepherds and farmers.

2.3 Review of Phylogenetic and Population Structure Research on Dog Breeds

2.3.1 Modern Dog Breed Development

Geographic isolation, migration and artificial selection have all shaped the development of modern dog breeds (Parker et al., 2017). Previously published studies have incorporated a range of phylogenetic methods to examine the relationships and history of dog breeds. An important study of the development of dog breeds was conducted by Parker et al. (2017a), who utilised genetic distance, migration patterns and genome-wide haplotype sharing analyses to investigate these developmental histories. Using a global SNP dataset of 161 breeds genotyped on the Illumina Canine HD BeadChip, the study identified 23 multi-breed clades (Parker et al., 2017). Each breed clade contains breeds that share behaviours, physical traits and geographic origin. All dog breeds share small haplotypes within genomic regions of approximately 10-15kb, with selection for and/or against specific traits influencing haplotype retention between breeds over time. Because of this, breed-specific patterns only appear in larger shared haplotypes (Sutter et al., 2004). Identity-by-descent (IBD) haplotype sharing in this study revealed that most dog breeds either do not share large haplotypes beyond their breed clade or share large haplotypes with only one other breed from a different clade. Interestingly, some clades contain dogs that share functional or behavioural traits with other clades, but the limited haplotype sharing between breed clades suggests that many traits associated with specific functions, such as herding or guarding, may have risen independently in different geographic regions. Furthermore, it is likely that the breed clades identified in the study were present prior to the Victorian-era breed explosion 200 years ago. Given the oral history of New Zealand Huntaway's and Heading Dogs, the UK rural clade identified in this study is of particular relevance. This clade consists of breeds primarily used for herding purposes, and some notable breeds include the Border Collie, Shetland Sheepdog, Corgi, Old English Sheepdog, Bearded Collie, Australian Shepherd, Australian Cattle Dog, and Kelpie.

2.3.2 Patagonian Sheepdog

Barrios et al. (2022) investigated the genetic relationships between the Patagonian Sheepdog and modern herding breeds from the United Kingdom. Similar to the New Zealand Huntaway and Heading Dog, the Patagonian Sheepdog has been geographically isolated for over 130 years. During this time, local sheep farmers selectively bred these dogs to meet the environmental and practical demands of sheep farming in Patagonia, resulting in the modern Patagonian Sheepdog. This population of working dogs remains unrecognised as a breed by any kennel clubs (Barrios et al., 2019). In their study, 159 Patagonian Sheepdogs were genotyped using the Illumina CanineHD Whole-Genome Genotyping BeadChip, and the resulting SNP data was incorporated into the previously published dataset of 161 dog breeds from the study of Parker et al. (2017). A total of 150,069 SNP markers were analysed. Genetic analyses revealed that all 159 Patagonian Sheepdogs belonged to the 'UK rural clade' as previously defined (Parker et al., 2017), reflecting the history of European settlement in Patagonia and the introduction of sheep farming (Barrios et al., 2019). Cladograms were constructed to visualise phylogenetic relationships among the breeds in the dataset. Cladograms differ from phylogenetic trees since they focus on common ancestry without indicating the exact amount of evolutionary change between lineages. (Wiley, 1979). Genetic distances were calculated in PLINK v1.9 (PLINK, 2023), and identity-by-state (IBS) distance matrices and a neighbour-joining phylogeny were calculated using Phylip v3.698 (University of Washington, 2021) and visualised in FigTree v1.4.4 (FigTree, 2018). This revealed a strong genetic association between the Patagonian Sheepdog and breeds from the UK rural clade with a resampling confidence value of 92%. Further evidence came from haplotype sharing analysis, which examines sets of genetic variants that tend to be inherited together, providing insight into the genetic relationships of common ancestry (Liu et al., 2008). Using identity-by-descent (IBD) methods in Beagle v4.1 software (Ayres et al., 2012) and visualising these relationships using R Studio (R Studio Team, 2020), this analysis showed that the Patagonian Sheepdog population exhibited the highest levels of haplotype sharing with breeds in the UK rural clade, as well as with the German Shepherd, a member of the New world clade. To explore genetic diversity within the Patagonian Sheepdog population, admixture analysis was conducted using ADMIXTURE v1.3 (Alexander et al., 2015) on a subset of the data (Hoggart et al., 2004). This method revealed that the genomic structure of the Patagonian Sheepdog was most similar to that of the Border Collie, Bearded Collie and Kelpie, suggesting these breeds share a close common ancestor which is most likely to be extinct herding dogs from the UK (Barrios et al., 2019). Principal component analysis (PCA), which reduces complex datasets to principal components while retaining most of the variance in the data (Greenacre et al., 2022), further supported these findings. This was conducted using the R package flashpcaR (Abraham et al., 2017). Although genetic similarities among the Patagonian Sheepdog, Border Collie, Bearded Collie and Kelpie were evident, the PCA plot showed that these breeds clustered separately, indicating clear genetic differentiation.

F_{HOM} inbreeding coefficients, which are generated based on the difference between observed and expected numbers of homozygous genotypes (Nishio et al., 2023) were also calculated for the Patagonian Sheepdog population and 10 other breeds using the PLINK v1.9 --het function. The average F_{HOM} for the Patagonian Sheepdog was 0.03, the lowest observed among all breeds in the analysis. In comparison, the next lowest F_{HOM} was found in the Border Collie (0.15), while the Collie exhibited the highest F_{HOM} at 0.45. These findings suggest that the Patagonian Sheepdog is less inbred than other modern herding breeds, indicating that the bottlenecks affecting modern herding breeds (due to artificial selection) may be more severe compared to the selection history of the Patagonian Sheepdog. However, it is important to note that the sample size for other breeds (n=10) is small in comparison to the number of Patagonian Sheepdogs used (n=159), and further studies with larger sample sizes would help strengthen these conclusions.

2.3.3 Sardinian Fonni's Dog

Dreger et al. (2016) investigated the genetic history of Fonni's Dog, a breed native to Sardinia that has been shaped primarily by geographic isolation and environmental factors, rather than by artificial selection. Fonni's Dog is thought to be genetically related to sighthound and mastiff breeds developed in Italy, the Middle East and North Africa, such as the Komondor, Anatolian Shepherd, Sloughi, Saluki and Azawakh, Cane Corso, Cane Paratore, Mastino Abruzzese, Spinone Italiano, Volpino Italiano, Levriero Meridionale, and Cirneco dell'Etna. Using SNP data from 23 breeds genotyped with the Illumina CanineHD

SNP chip (~150,000 SNP markers post filtering), the author conducted STRUCTURE analysis, a method similar to ADMIXTURE using FastSTRUCTURE v.1.0 (Raj et al., 2014) and constructed neighbour-joining cladograms with Phylip v3.696 (University of Washington, 2021) to investigate the genetic relationships between Fonni's Dog and other Mediterranean and Italian breeds. Consistent with their hypothesis and the human demographic history of Sardinia, this study found that Fonni's Dog originated from a combination of breeds, particularly those resembling sighthounds and mastiffs. Fonni's Dog shares a direct common ancestor with the breeds Mastino Abruzzese, Cane Paratore, and Volpino Italiano as indicated through neighbour-joining cladograms (Dreger et al., 2016).

F_{HOM} was calculated using the --het function in PLINK v1.07. Inbreeding coefficients calculated for the Fonni's dog had a mean of 0.13, while other breeds analysed in this study had inbreeding coefficient estimates of 0.39 (Pharaoh Hound) and 0.03 (Mastino Abruzzese). Therefore, the inbreeding values estimated for the Fonni's dog are within the range observed for other mediterranean dog breeds.

2.3.4 Korean Sapsaree

Similarly, Gajaweera et al. (2019) examined the genetic diversity of the Sapsaree, a historically old breed believed to have originated during Korea's Silla dynasty (57BCE – 935CE) and later restored from near extinction. SNP data genotyped on the Illumina CanineSNP20 BeadChip (~22,000 SNP markers analysed) from the Sapsaree and morphologically similar breeds, including the Afghan Hound, Tibetan Terrier, Lhasa Apso and Great Pyrenees, were analysed using multidimensional scaling (MDS), admixture analysis and cladogram construction. Like PCA, MDS is a multivariate statistical method used to visualise similarities or differences in genetic datasets (Tzeng et al., 2008). The MDS constructed using PLINK v1.9 (PLINK, 2023) and graphed in R Studio (R Studio Team, 2020) revealed that the Sapsaree clustered distinctly from the other breeds, therefore there is enough genetic differentiation for it to be considered a distinct breed. Admixture analysis was evaluated through ADMIXTURE v.1.23 (Alexander et al., 2015), and conducted with an optimal K value of 10. This showed that the Sapsaree displayed high levels of genetic admixture compared to the other breeds, suggesting that remnants of ancestral genetic contribution have persisted in the population. The admixture visualised is also attributed to the non-restrictive breeding practices employed during the breed's restoration process to reestablish this iconic dog breed in South Korea. The cladogram created with SNPhylo (Lee et al., 2014) and illustrated using FigTree v1.4.2 (FigTree, 2018) shows that the Sapsaree shared common ancestry with other Asian developed breeds such as the Tibetan Terrier and Lhasa Apso, as well as the Great Pyrenees and Old English Sheepdog (Gajaweera et al., 2019).

2.3.5 Australian Labradoodle

Another study undertaken by Ali et al. (2020) investigated the development of the Australian Labradoodle, a popular designer breed that is a cross of the Standard Poodle and Labrador Retriever. All dogs in this study were genotyped with the Illumina CanineHD SNP chip (150,106 SNP markers post filtering). Initial insights into the relationship of the Australian Labradoodle, Labrador and three Poodle breeds were explored through PCA calculated using FlashPCA (Abraham G & Inouye M, 2014). From this, they were able to determine that the Australian Labradoodle retains more ancestry from the poodle in comparison to the labrador, as the Australian Labradoodle clustered much closer to the Poodle breeds in comparison to the Labrador. Admixture analysis was conducted using ADMIXTURE v1.23 (Alexander et al., 2015) with K values of 4-6 and further suggested that the three Poodle breeds are the main contributors to the Australian Labradoodle followed by the Labrador, aligning with the results from the PCA. IBD haplotype sharing analysis was then conducted incorporating the dataset from the study by Parker et al. (2017) described above. This was performed with BEAGLE v4.1 (Ayres et al., 2012). As expected, the three Poodle breeds and the Labrador had the most significant haplotype sharing with the Australian Labradoodle.

Inbreeding was also assessed in the Australian Labradoodle and related breeds using the 'het' function in VCFtools v0.14, which calculates F_{HET}. Unlike F_{HOM} calculated in PLINK which measures inbreeding based on homozygosity, F_{HET} detects inbreeding by identifying reduced heterozygosity. As expected, both the Labradoodle and the Australian Labradoodle had the lowest mean inbreeding F_{HET} scores (-0.06 and 0.05, respectively) among the breeds analysed, while the highest F_{HET} score was observed in the American Cocker Spaniel (0.37).

The higher genetic diversity of the Labradoodle and Australian Labradoodle can be explained by their mixed breed origins, which result in greater heterozygosity compared to purebred populations (Ali et al., 2020).

2.3.6 Runs of Homozygosity Analysis in Dogs

Runs of homozygosity (ROH) analysis is a powerful method of estimating genome-wide inbreeding levels and detecting selection signatures. Runs of homozygosity are uninterrupted homozygous segments within the genome (Gorssen et al., 2021), and the length and quantity provide valuable insights into individual and population history. Long, continuous ROH segments (>10Mb) indicate recent inbreeding (within five generations), while shorter ROH segments result from recombination and reflect distant inbreeding. Distant inbreeding typically refers to events occurring 12.5 to 50 generations ago, with ROH segments of 4Mb associated with inbreeding events 12.5 generations ago and segments around 1Mb indicative of events 50 generations ago (Perfilyeva et al., 2023). Runs of homozygosity analysis also provides information about genetic relatedness between individuals. Runs of homozygosity lengths can be used to calculate F_{ROH}, a method of calculating inbreeding based on ROH (Nishio et al., 2023), which is a more accurate inbreeding coefficient estimate compared to pedigree data (Halvoník et al., 2024). PLINK is the most commonly used software for ROH analyses in livestock populations, using the ‘homozyg’ function with a scanning window approach (Meyermans et al., 2020).

Mastrangelo et al. (2018) examined genetic variability and inbreeding in the Braque Français type Pyrénées dog using SNP genotype data. Runs of homozygosity detection was calculated in PLINK v1.9, and molecular inbreeding (F_{ROH}) was estimated per animal by the ratio between the total sum of ROH lengths and the total autosomal genome size. The highest mean F_{ROH} values across all breeds were observed in the Doberman Pinscher (0.371) and English Bulldog (0.357), while breeds with the lowest average F_{ROH} included the Jack Russel Terrier (0.080) and Shar-Pei (0.106). Border Collie F_{ROH} in this study was 0.15, while the Australian Shepherd F_{ROH} was 0.21. Similarly, Sweetalana et al (2024), analysed F_{ROH} by breed groups using the same method. The herding group (15 breeds) included five UK rural breeds (Australian Cattle Dog, Border Collie, Bearded Collie, Pembroke Welsh Corgi and

Shetland Sheepdog) and had a mean F_{ROH} of 0.29. Compared to other breed groups in the study, the Herding group's mean F_{ROH} was moderate, with the highest observed F_{ROH} in the Terrier group (0.41) and the lowest observed in the Chinese Indigenous dogs (0.05) (Sweetalana et al., 2024).

Runs of homozygosity can also be used to investigate highly inbred genomic regions within a population, which are referred to as ROH islands (ROHi). Runs of homozygosity islands provide insights into signatures of putative positive selection in the genome. Perfilyeva et al. (2023) used ROH analysis to estimate genome wide inbreeding levels and identify selection signatures in a Kazakh dog population. Among 30 Kazakh dogs analysed, a total of 1,699 ROH were identified, with 1,143 between 1-2Mb in length, 306 between 2-4Mb, 136 between 4-8Mb, 67 between 8-16Mb and 47 exceeding 16Mb. Five extreme ROH islands (eROHi) were detected on chromosomes 18, 22 and 25, containing 39 genes. The mean ROH based inbreeding coefficient per dog was 0.057, suggesting a low level of inbreeding. This is comparable to previously reported inbreeding estimates of the Jack Russel Terrier (0.061) (Mortlock et al., 2016a) but considerably lower than in breeds such as the German Shepherd (0.119) and the Bulldog (0.151) (Mortlock et al., 2016a; Mortlock et al., 2016b). Similarly, Shihabi et al. (2022) investigated eROHi in livestock guarding dogs to understand the mechanisms of environmental adaptation and artificial selection. Ten eROHi were identified, with two notable genes highlighted. *LRIG3* (Leucine Rich Repeats and Immunoglobulin like Domains 3) and *TRIP11* (thyroid hormome receptor interactor 11) (Shihabi et al., 2022). *LRIG3* is involved in fat metabolism and bone morphogenetic protein signalling (Herdenberg et al., 2025). A study by Zapata et al (2020) identified a genomic region in chromosome 10 near the *LRIG3* gene that predicts aggression directed at unfamiliar dogs (Zapata et al., 2022), though livestock guarding dogs are not thought to often display aggressive behaviour (McConnell et al., 2022). *TRIP11* (thyroid hormone receptor interactor 11) is involved in Golgi apparatus function and encodes a protein that interacts with thyroid hormone receptor beta, though a link to livestock guarding dog behaviour has not been made (Shihabi et al., 2022).

The methodologies employed in previous studies, including cladogram construction, haplotype sharing analysis, admixture modelling and PCA, can be adapted and applied in this study to investigate the genetic structure and breed composition of the Huntaway and Heading Dog. Additionally, ROH analysis will facilitate the estimation of baseline inbreeding levels and allow for the detection of genomic regions under putative positive selection. Identifying these regions can reveal candidate genes associated with morphological, behavioural, or disease-related traits that have been influenced by historical and ongoing selective breeding pressures in these working dog populations.

2.4 Review of SNP Genotyping Technologies and Phylogenetic Software Tools

2.4.1 SNP Genotyping

The availability of genomic data provides the opportunity to explore the origins and breed characteristics of the New Zealand Huntaway and Heading Dog. This study utilises microarray technology from two leading brands, Thermofisher and Illumina, with these respective panels being the Affymetrix Axiom Canine HD Array and the Illumina CanineHD Whole-Genome Genotyping BeadChip. Single nucleotide polymorphism arrays are a cost-effective means of generating genotype calls for a specific number of SNP markers (Nicolazzi et al., 2016). The Affymetrix arrays employ multiple 25-mer probes synthesised in situ, each positioned at pre-defined locations on the array. In contrast, Illumina arrays use 50-mer probes attached to microbeads, which are randomly distributed on the array and decoded using specific tagging sequences. (Barnes et al., 2005). Furthermore, the Affymetrix canine chip contains 712,577 markers (Thermofisher, 2024) while the Illumina canine chip has 172,115 markers (Illumina, 2025a). Single nucleotide polymorphism array data lacks standard file formats and allele-coding conventions across different platforms, which poses a challenge when it comes to integrating data from two or more platforms (Nicolazzi et al., 2016). Limited research has explored the compatibility of SNP datasets generated by the canine SNP array from these two companies, but a study by Howard et al. (2021) examined the concordance of the Illumina Infinium 20K and Affymetrix Axiom 480K SNP arrays in the Apple variety *Malus x domestica* and reported an 81.7% agreement across a subset of 10,295 SNPs in duplicate samples. This study highlighted that compatibility is significantly influenced by alternate probe binding locations and secondary polymorphisms, and while common SNPs between the two datasets can be integrated, this process requires significant data filtering and conversion (Howard et al., 2021).

Since Illumina and Affymetrix platforms use different protocols for allele calling and identification, advanced bioinformatics methods are required to standardise and integrate SNP data across these platforms. SNPConvert is an open-source software developed in Python 2.7 (Nicolazzi et al., 2016) that facilitates cross-platform standardisation. This program works by converting Illumina "FinalReport" and matrix raw file formats into PLINK format, while also modifying allele coding formats and updating genomic coordinates in PLINK files. PLINK is a widely used open-source software in genetic research and is the only common format for SNP data output across Illumina and Affymetrix.

An alternative method for cross-platform standardisation is using the BCFtools 1.20 plugin 'idat2gtc'. This plugin converts Illumina raw intensity IDAT files into genotype call (GTC) files (GitHub, 2024). IDAT files are binary outputs generated during sampling scanning. These files contain raw fluorescence intensity data from the red and green channels for each probe on the chip. These intensities represent hybridisation signals for the two possible alleles (A and B) required for determining genotypes. GTC files are binary files that contain final genotype calls for each SNP, along with sample specific quality control data. The 'idat2gtc' plugin also requires Illumina's CanineHD Manifest and cluster files (Illumina, 2025b) to perform the conversion. The manifest file provides specific genomic information such as the SNP name, chromosome and its position on the reference genome and also provides details about the probe sequences, target alleles and assay design. This allows the alignment of raw intensity data with genomic regions. The cluster file acts as a reference, defining the predefined genotype clusters (AA, AB, BB) for each SNP. These clusters are based on the distribution of the red and green intensities derived from a large reference sample set. Once in GTC file format, the files can then be converted into variant call format (VCF), a widely used text file format for storing genetic data, using the BCFtools 1.20 plugin 'GTC2VCF' (GitHub, 2024). From VCF format, the data can be converted into PLINK binary format using the PLINK software and used for downstream filtering, merging and analysis.

2.4.2 Genomic Data Processing and Analysis Software Tools

2.4.2.1 PLINK

PLINK (Purcell et al., 2007) is an open-source whole genome association study toolset that is designed for the analysis and manipulation of large-scale genotype data. It is commonly used in population genetics, genome wide association studies (GWAS) and analysis of other genotype properties and parameters. Some of the main functions of PLINK include data filtering, basic association testing, linkage disequilibrium analysis, population stratification correction and data format conversion. PLINK supports PLINK .ped/.map, PLINK binary .bed/.bim/.fam and VCF file formats, which are compatible with many downstream genetic analysis tools. In canine genetics research, PLINK has been widely used to investigate breed structures, genetic diversity and ancestry. Studies have applied it for genotype data filtering, principal component analysis (PCA), and the calculation of genetic distances, inbreeding coefficients and allele frequencies (Ali et al., 2020; Barrios et al., 2022; Dreger et al., 2016; Gajaweera et al., 2019). Furthermore, it is one of the most widely used programs for analysing ROH both in human and animal populations (Barrios et al., 2022; Dreger et al., 2016; Gorssen et al., 2021; Meyermans et al., 2020; Perfilyeva et al., 2023; Shihabi et al., 2022). Given its relevance to similar studies, PLINK can be applied in this project to ensure consistency across datasets derived from different genotyping platforms, facilitating comparisons of genetic variation across breeds, and enabling downstream analyses such as haplotype sharing and phylogenetic analyses.

2.4.2.2 Genotype Harmoniser

Genotype Harmoniser (Deelen et al., 2014) is an open-source software tool designed to standardise genetic datasets by resolving strand orientation issues and ensuring format compatibility. It supports PLINK, binary PLINK and VCF file formats and works by aligning SNPs to a specified reference without requiring prior strand orientation knowledge. To address ambiguous A/T and C/G SNPs, Genotype Harmoniser leverages linkage disequilibrium (LD) patterns, selecting nearby non-ambiguous SNPs in LD with both the study and reference data. By comparing estimated haplotype frequencies, it identifies strand mismatches based on correlation patterns and flips strands when negative correlations outweigh positive ones. Additionally, it can optionally align SNPs based on minor allele frequencies. If alignment is not possible due to insufficient surrounding SNPs, the ambiguous SNP is excluded. Genotype Harmoniser is particularly useful for studies integrating datasets from different genotyping platforms, where strand inconsistencies and format differences complicate data merging. By aligning ambiguous SNPs to a common reference and resolving discrepancies, it enhances dataset consistency, reduces potential errors and improves the accuracy of downstream genetic analyses.

2.4.2.3 R Studio

R Studio (R Studio Team, 2020) is a powerful integrated development environment for R, widely used in statistical computing, data visualisation and genomic research. R Studio has an extensive library ecosystem, particularly packages such as ggplot2, dplyr and tidyverse, which enable users to efficiently manipulate and visualise data, making it an ideal tool for preprocessing and visualising outputs from genetic analyses such as haplotype sharing analysis, admixture and PCA.

2.4.2.4 Beagle

Beagle is a widely used software tool capable of handling large-scale datasets that can be used for phasing genotyping datasets (Ayres et al., 2012; Browning & Browning, 2016). Of particular relevance to this study is Beagle’s identity-by-descent (IBD) detection algorithm, which compares genotype data to identify regions of the genome that are identical between individuals (Browning & Browning, 2010). Beagle has been widely used in canine phylogenetic studies (Ali et al., 2020; Barrios et al., 2022; Parker et al., 2017) to investigate the genetic histories of various dog breeds and to assess their relatedness. By identifying genomic segments inherited from a common ancestor, Beagle provides valuable insights into patterns of historical migration and population structure, shedding light on genetic bottlenecks, breed divergence, and the genetic underpinnings of breed-specific traits.

2.4.2.5 Admixture

ADMIXTURE (Alexander et al., 2015) is a widely used software tool for analysing population structure and genetic admixture in large-scale genomic datasets. Using a maximum likelihood approach, it estimates the proportion of ancestry from different populations or ancestral sources for each individual in the dataset and supports PLINK file formats. Admixture software is commonly used in human genetics, animal breeding and conservation genetics and helps to uncover population stratification, identify sub-populations and assess the extent of admixture between groups. In other canine phylogenetic studies (Barrios et al., 2022; Gajaweera et al., 2019), Admixture was used to illustrate the genetic diversity and variability within breeds, as well as identifying shared genomic signatures between closely related breeds. Its ability to detect subtle patterns of ancestry and admixture makes it a valuable tool for understanding the genetic relationships that drive breed formation and diversification.

2.4.2.6 Phylip

Phylip (Phylogeny Inference Package) software (University of Washington, 2021) is a free package of programs for inferring phylogenies of living species and organisms. It is now one of the most widely used packages for computing accurate phylogenetic trees and carrying out certain related tasks. It contains over 35 programs, with the most notable programs being NEIGHBOR and CONSENSE programs. The NEIGHBOR program within Phylip is used to construct phylogenetic trees based on distance matrix methods. It employs a neighbour-Joining (NJ) algorithm, which is one of the most widely used techniques for constructing phylogenies. The NJ method builds a tree by iteratively joining pairs of taxa that are most similar to each other, effectively minimising the total branch length at each step. The result is a tree structure that reflects the genetic distances between the input sequences. The program can also incorporate bootstrap analysis, allowing users to assess the robustness of the resulting tree by resampling the data and observing how often the same clades are recovered across multiple replicates. The CONSENSE program is used for evaluating the reliability of phylogenetic trees generated by various methods, including NEIGHBOR. After generating a set of candidate trees, CONSENSE employs a consensus tree approach to summarise the information from multiple trees into a single representative tree. It calculates the majority-rule consensus tree, which shows the most frequently occurring branching patterns among the trees and also calculates bootstrap values for the nodes of the tree, which provide a measure of confidence in the correctness of the tree. Phylip software was utilised in other canine phylogenetic studies (Barrios et al., 2022; Dreger et al., 2016; Parker et al., 2017) to construct neighbour-joining cladograms of dog breeds from around the world. Whilst neighbour-joining cladograms do not indicate the exact number of evolutionary changes between branches, this method can be used to understand the relationship between the breed of interest and all other breeds within their datasets. This approach also helped to identify breed clades within the canine family tree (Parker et al., 2017).

2.4.2.7 Tree Viewer

TreeViewer is a software tool designed for drawing and analysing phylogenetic trees (Bianchini & Sánchez-Baracaldo, 2024). TreeViewer is user-friendly and contains many built-in customisation options that allow for adjustments in branch lengths, cartooning and collapsing branches, colours and labels to enhance interpretability. It supports both rooted and unrooted trees and can also display metadata such as bootstrap support values and clade or taxa information. Phylogenetic trees drawn in TreeViewer can be readily exported into formats suitable for publication (Bianchini & Sánchez-Baracaldo, 2024).

2.4.3 Runs of Homozygosity Islands Analysis

Gorssen et al. (2021) compiled a comprehensive repository of ROH islands for 442 populations from eight animal species, including dogs. This repository contains data from 50 dog breeds, offering a valuable resource for examining breed-specific ROH patterns. To enable standardised ROH analysis, they developed a publicly available R script for breed-specific quality control and ROH detection (Gorssen et al., 2021). The script utilises PLINK binary files (.bim, .bed, .fam) to perform quality control, identify ROH regions, and generate Manhattan plots that visualise ROH incidence per SNP.

The R script begins with quality control using PLINK v.1.9. Only autosomal SNPs are retained with --autosome, and individual call rates are set to 0.90 with --mind 0.10. Any duplicate SNPs or individuals within the dataset are excluded, and the minimum SNP call rate is set to 0.95 with --geno 0.05. For the ROH analysis (performed using the --homozyg function), heterozygous SNPs are excluded using --homozyg-window-het and --homozyg-het, and a single missing SNP per window was allowed with --homozyg-window-missing. The minimal number of SNPs required per window and in the final ROH segment are calculated using the L-parameter and specified with --homozyg-window-snp and --homozyg-snp, respectively. The minimum window size is set to 1000kb with --homozyg-kb with a density of 150kb per SNP (--homozyg-density). The maximum gap between SNPs in a window is set to 1000 kb with --homozyg-gap, while the window threshold is determined by two outer SNPs (Gorssen et al., 2021).

ROH incidence is calculated as the percentage of animals with a SNP within an ROH segment for a given population, which is then visualised using Manhattan plots in R Studio with the qqman package (Turner, 2018). ROH islands are defined as regions where SNPs have a p-value for ROH incidence that exceeds the population specific threshold. These thresholds are derived from standard normal z-scores based on the distribution of ROH incidences, and the top 0.1% of SNPs are considered to form ROH islands. Location tables for ROH islands in each breed are available in the repository created by Gorssen et al. (2021), and locations are represented as 1Mb bins. These can be used to explore overlapping of ROH island regions across different breeds or populations (Gorssen et al., 2021).

2.5 Summary and Research Objectives

The sheep and beef farms of New Zealand would not be physically or economically viable without the contributions of New Zealand's two unique working farm dogs: the Huntaway and the Heading Dog. These two breeds have been selectively bred to carry out distinct, yet complementary roles in the mustering and management of livestock across New Zealand's farming landscapes. The Huntaway, known for its strong voice and ability to drive livestock over long distances, and the Heading Dog, valued for its precision and control, are both essential to the effective functioning of pastoral farming operations throughout the country. Despite their significant and ongoing contributions to agricultural practices in New Zealand, there has been limited scientific research focused on these working dogs. Furthermore, little has been formally documented regarding the origins, genetic makeup and establishment of the Huntaway and Heading Dog as distinct breeds. Studies on other modern dog breeds have made use of phylogenetic and population structure analyses to explore their development, revealing isolation, human-directed migration, and artificial selection as key factors in shaping breed characteristics and lineage.

The primary objective of this study is to examine the genetic relationships between the New Zealand Huntaway and Heading Dog and a range of formally recognised dog breeds from around the world. Through the use of genetic analyses, this study aims to identify and formalise the historical breed contributions that have influenced the development of the Huntaway and Heading Dog. In addition, average baseline inbreeding levels within the populations of these two breeds will be estimated, providing a clearer understanding of their current genetic diversity and population structure. This research supports one of the central goals of the “Right Dog for the Job” project, which is to identify genetic variants that may impact the health and performance of New Zealand’s working dogs. Understanding the genetic makeup and historical breed composition of the Huntaway and Heading Dog provides additional context for analysing genetic variation related to performance and disease risk. These insights may help explain shared health or performance related traits observed in these breeds and others with overlapping genetic backgrounds, ultimately contributing to more informed breeding and management decisions.

Chapter 3: Methods

3.1 Methods Overview

To investigate the breed structure and origins of New Zealand Huntaways and Heading Dogs, we genotyped 211 Huntaways and 246 Heading Dogs using the Affymetrix Axiom CanineHD array (Thermofisher, 2024). Dog samples were obtained from various regions across New Zealand (Table 2), and genotypes were integrated with publicly available genotypes that were generated on the Illumina CanineHD Whole-Genome Sequencing BeadChip (Illumina, 2025a) from two previous studies (Barrios et al., 2022; Parker et al., 2017). The merged datasets were then used to determine the origin and diversity of the New Zealand Huntaway and Heading Dogs, as outlined below.

3.2 Genotyping

Cheek swab samples from 211 Huntaways and 246 Heading Dogs were taken on farm by Massey University staff. Swabs were then couriered back to the Massey Al Rae Centre in Hamilton, New Zealand, and were then couriered to the Ancestry laboratory in Massachusetts, America (Ancestry, 2024). Genotypes were generated by Ancestry under contract, following company standard commercial procedures using the Affymetrix Axiom CanineHD array (Thermofisher, 2024).

3.3 Data processing

3.3.1 Processing of Previously Published Data

The data uplifted from Parker et al. (2017) and Barrios et al. (2022) was genotyped on the Illumina CanineHD Whole-Genome Sequencing BeadChip. The New Zealand Dog data was genotyped using Affymetrix Axiom CanineHD Array Chips. Due to genotyping platform differences, various formatting and filtering methods were required to make the two datasets compatible with one another.

Data not directly generated through our study (i.e. dogs genotyped from breeds other than Huntaways and Heading Dogs) were sourced from the Gene Expression Omnibus (GEO) database under accession numbers GSE70454, GSE83160, GSE90441, GSE96736 and GSE193384 (Barrios et al., 2022; Parker et al., 2017). These data were available as raw intensity (IDAT) files. Firstly, the IDAT files were converted into genotype call (GTC) format files using the BCFtools v1.20 plugin idat2gtc (GitHub, 2024) with Illumina's provided manifest (BPM) and cluster (EGT) files for the Illumina CanineHD BeadChip. The resulting GTC files were then converted into variant call format (VCF) using BCFtools v1.20 plugin GTC2VCF (GitHub, 2024).

The VCF files were then converted into PLINK binary format (.bed, .bim and .fam) with PLINK v1.90b6.16 (Purcell et al., 2007). The PLINK files generated from the five GEO datasets were merged using PLINK v1.90b6.16, resulting in a unified dataset containing 1,256 previously published CanineHD BeadChip genotypes (172,743 SNPs) representing 127 dog breeds and three wild canid species. Further filtering was applied on the unified dataset to remove problematic variants. The PLINK function --autosome removed 5,560 variants that were located on sex chromosomes or on unidentified chromosomes. The PLINK functions --geno 0.10 and --mind 0.10 were applied, removing a further 5,259 variants with more than 10% missing call rates. The function --mind 0.10 did not remove any samples from the dataset, because no individuals had missing genotypes over 10%. After filtering, the unified dataset contains 161,924 variants from 1,256 canids. This dataset is referred to hereafter as the 'Illumina dataset'.

3.3.2 Processing of New Zealand Data

Following genotyping at Ancestry, the New Zealand Huntaway and Heading Dog results were received as .ped and .map files aligned to the canFam3 reference genome. The .ped and .map files were converted into PLINK binary files using PLINK v1.90b6.16 (Purcell et al., 2007). The resulting dataset included 299 Affymetrix Axiom CanineHD Array genotypes (712,577 SNPs). The same filtering steps described above were also applied to this dataset, with the function --autosome removing 11,216 variants, --geno 0.1 removing 8,601 variants with more than 10% missing genotypes and --mind 0.10 removing one individual that was missing more than 10% of its genotype. After filtering, the dataset contained 692,760 variants from 298 dogs. This dataset will be referred to as the Affymetrix dataset.

3.3.3 Harmonising the Datasets

Since the two datasets were genotyped on different platforms, further processing was required to ensure that the two datasets could be merged successfully. First, variants in the Illumina dataset that had the same position as those in the Affymetrix dataset were renamed to match the variant IDs in the Affymetrix dataset. A total of 160,465 variants in the Illumina dataset were renamed. To ensure consistency and compatibility in downstream genetic analyses, the remaining 1,459 variants in the Illumina dataset that were not renamed since these variants were not present in the Affymetrix dataset were removed from the Illumina dataset. Likewise, the 532,099 variants that were present in the Affymetrix dataset, but were not in the Illumina dataset were also excluded. The two resulting datasets retained the 160,465 variants that were present in both datasets.

For some of the variants present in the Affymetrix dataset, only one allele was observed (the alternate allele was not present in any dogs genotyped). When this occurs, the .map and .ped file format can only record the one allele that was observed, and when these are converted into the binary format, this is represented by including the one allele alongside a zero as a placeholder for the missing allele (e.g. 0/A), rather than including both the reference and alternative alleles. In the Illumina dataset, by contrast, the same variants that only had only one allele recorded had both the reference and alternative alleles listed (e.g. C/A). The placeholder zeros in the allele coding of the Affymetrix dataset created issues in the merging process since the allele coding did not match that in the Illumina dataset, therefore where the major allele in the Illumina dataset matched that of the same variant in the Affymetrix dataset, the zero placeholder was updated with the minor allele from the Illumina dataset (e.g., a variant in the Affymetrix dataset was 0/A, and in the Illumina dataset it was C/A so the Affymetrix allele code was updated to C/A). A total of 13,214 variants in the Affymetrix dataset had the allele coding corrected.

To further ensure the compatibility of the two datasets for merging, the Affymetrix dataset was run through Genotype Harmoniser v1.4.25 (Deelen et al., 2014), using the Illumina dataset as the reference panel. Genotype Harmoniser aligns and standardises genotype datasets by ensuring compatibility in SNP names, positions, alleles and reference strands. It also carefully assesses ambiguous SNPs (e.g. A/T or C/G) by checking alignment with the reference panel and, when necessary, checking nearby SNPs in linkage disequilibrium and using minor allele frequency to resolve strand orientation or exclude problematic SNPs. This process removed 8,797 variants from the Affymetrix dataset. The Affymetrix dataset now contained 151,668 variants from 298 dogs. These 8,797 variants were also removed from the Illumina dataset to ensure that the merged dataset would not have missing calls that were stratified by panel manufacturer.

3.3.4 Merging the Datasets

The Affymetrix and Illumina datasets were merged using PLINK v1.90b6.16 (Purcell et al., 2007) using the --bmerge function and had a genotyping rate of 0.995. The resulting dataset comprised of 151,688 variants from 1,554 dogs, representing 129 breeds and three wild canid species (132 breeds/species in total). To prevent overrepresentation by the most sampled breeds, the Huntaway (n = 108), Heading Dog (n = 141) and Patagonian Sheepdog (n = 146) populations were randomly down sampled to 10 individuals each. This reduced the dataset down to 1,140 canids, and this subset was used for neighbour-joining cladogram construction. An additional subset was created for ADMIXTURE and Principal Component Analysis (PCA), which included the UK rural clade breed individuals and 10 individuals for each of the Huntaway, Heading Dog and Patagonian Sheepdog populations and were randomly selected. This subset contains 104 individuals and is referred to as the UK rural and NZ subset.

After an additional batch of genotyping had been performed by Ancestry, a further 181 Affymetrix genotypes were subsequently merged with the dataset containing 1,554 dogs using the same methods as described above. The merge had a genotyping rate of 0.992, and the resulting dataset contained 151,688 variants from 1,735 dogs, representing 129 breeds and three wild canid species. The total number of New Zealand Huntaways and Heading

Dogs in this dataset was 211 and 246, respectively. This dataset was used for haplotype sharing analysis and runs of homozygosity (ROH) analysis, and to estimate the inbreeding coefficients F_{ROH} and F_{HOM}. The construction of the cladogram, admixture and principal component analyses were not rerun with the additional samples since these analyses used subsamples of the original dataset. The data conversion, filtering and merging pipeline for the Illumina and Affymetrix SNP datasets is outlined in Figure 2.

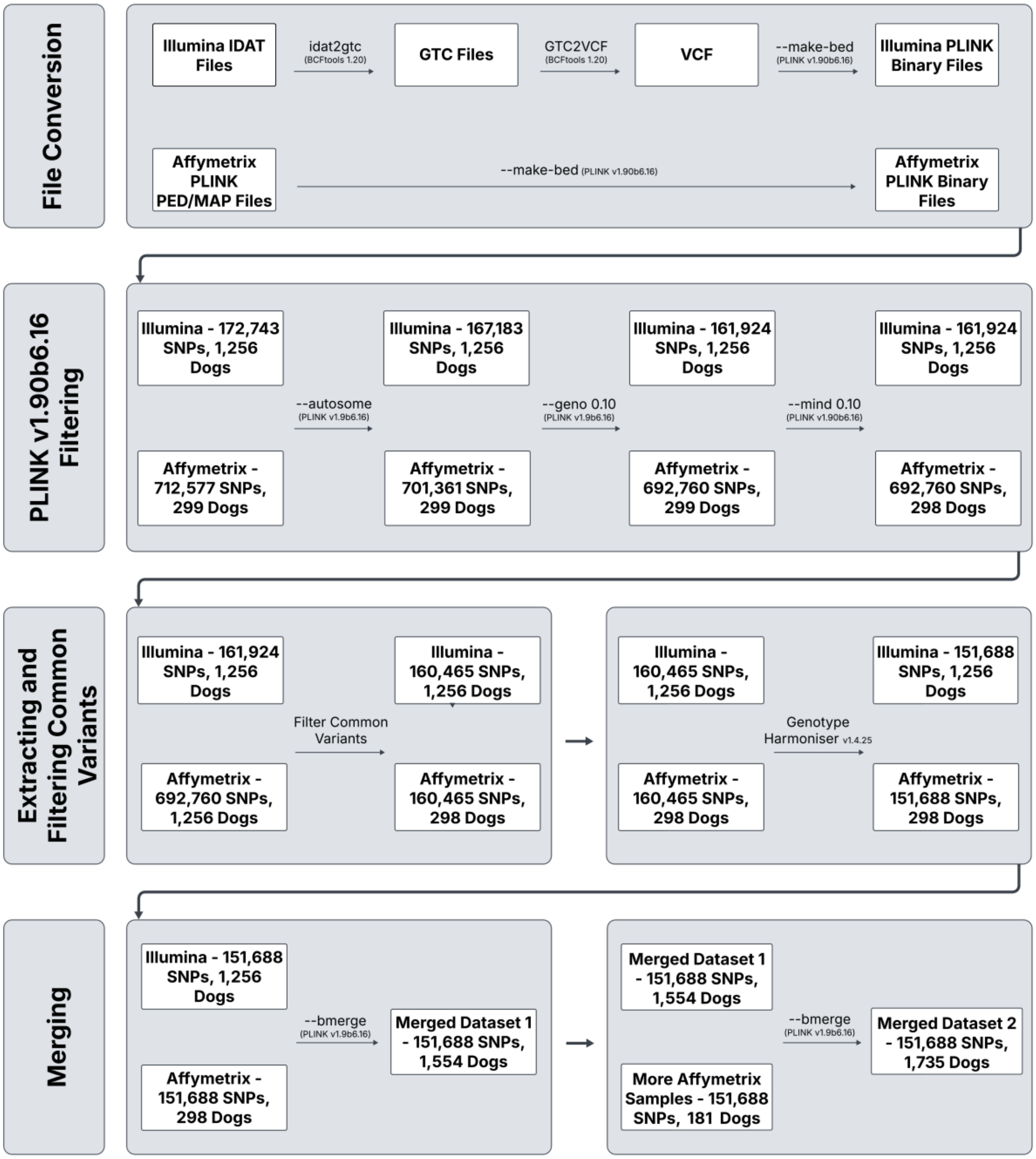

Figure 2: Data conversion and manipulation pipeline. The pipeline is broken into four stages: file conversion, PLINK filtering, extracting and filtering common variants, and merging.

3.3.5 Breed Information

A table that includes the animal genotype IDs, breeds, breed abbreviations and clades was created in R Studio (R Studio Team, 2020). Breed information for New Zealand samples was collected from dog owners in the "Right Dog for the Job" dog survey. The breed abbreviations and clade information for externally sourced dogs were obtained from Parker et al. (2017a) and Barrios et al. (2022). For the purposes of this study, the New Zealand Huntaways and Heading Dogs clade were listed as 'NZ' for the clade and were given the breed abbreviations HUNT and HEAD. Huntaway and Heading Dog crossbreeds were assigned to the NZ breed clade, and the breed abbreviation was set to NZ/CROSS (New Zealand Crossbreed). The NZ/CROSS dogs in the dataset were not used in any genetic analyses in this study. Table 1 presents the breeds, their abbreviations, associated clades and the number of dogs included in this study.

Table 1: Full list of dog breeds with their associated breed abbreviations, breed clade and the number of individual dogs for each breed.

Breed	Breed Abbreviation	Breed Clade	Number of Dogs
Basenji	BSJI	African	10
Akita	AKIT	Asian Spitz	10
Alaskan Malamute	AMAL	Asian Spitz	20
Chinese Shar-pei	SHAR	Asian Spitz	10
Chow Chow	CHOW	Asian Spitz	10
Samoyed	SAMO	Asian Spitz	10
Siberian Husky	HUSK	Asian Spitz	20
Tibetan Mastiff	TIBM	Asian Spitz	11
Xigou	XIGO	Asian Spitz	5
Pekingese	PEKE	Asian Toy	10
Shih Tzu	SHIH	Asian Toy	10
Tibetan Spaniel	TIBS	Asian Toy	5
Tibetan Terrier	TIBT	Asian Toy	5
Belgian Sheepdog	BELS	Continental Herder	10

Belgian Tervuren	TURV	Continental Herder	10
Briard	BRIA	Continental Herder	10
Black Russian Terrier	BRTR	Drover	4
Doberman Pinscher	DOBP	Drover	10
Giant Schnauzer	GSNZ	Drover	10
Rottweiler	ROTT	Drover	11
Bernese Mountain Dog	BMD	European Mastiff	7
Great Swiss Mountain Dog	GSMD	European Mastiff	6
Saint Bernard	STBD	European Mastiff	17
American Staffordshire Terrier	AMST	European Mastiff	2
Boston Terrier	BOST	European Mastiff	10
Boxer	BOX	European Mastiff	10
Bull Terrier	BULT	European Mastiff	10
Bulldog	BULD	European Mastiff	9
Bulmastiff	BULM	European Mastiff	10
Cane Corso	CANE	European Mastiff	9
Dogue de Bordeaux	DDBX	European Mastiff	6
English Mastiff	MAST	European Mastiff	12
French Bulldog	FBUL	European Mastiff	10
Great Dane	DANE	European Mastiff	10
Miniature Bull Terrier	MBLT	European Mastiff	10
Neapolitan Mastiff	NEAP	European Mastiff	12
Staffordshire Bull Terrier	STAF	European Mastiff	10
Puli	PULI	Hungarian	2
Pumi	PUMI	Hungarian	5
Afghan Hound	AFGH	Mediterranean	10
Anatolian Shepherd	ANAT	Mediterranean	6
Azawakh	AZWK	Mediterranean	5
Cane Paratore	CPAT	Mediterranean	2
Cirneco dell'Etna	CIRN	Mediterranean	5
Fonni's Dog	FONN	Mediterranean	6
Great Pyrenees	GPYR	Mediterranean	13
Ibizan Hound	IBIZ	Mediterranean	10
Kuvasz	KUVZ	Mediterranean	10
Leonberger	LEON	Mediterranean	10
Levriero Meridionale	LVMD	Mediterranean	2
Mastino Abruzzese	MAAB	Mediterranean	2
Saluki	SALU	Mediterranean	29
Sloughi	SLOU	Mediterranean	5
German Shepherd Dog	GSD	New World	10
Keeshond	KEES	Nordic Spitz	10

Norwegian Elkhound	NELK	Nordic Spitz	1
Swedish Valhund	SVAL	Nordic Spitz	6
New Zealand Heading Dog	HEAD	NZ	246
New Zealand Huntaway	HUNT	NZ	211
Heading Dog Crossbreed	NZ/CROSS	NZ	11
Heading Dog x Huntaway	NZ/CROSS	NZ	12
Huntaway Crossbreed	NZ/CROSS	NZ	13
Brittany	BRIT	Pointer Setter	10
Dalmatian	DALM	Pointer Setter	4
Gordon Setter	GORD	Pointer Setter	2
Irish Setter	ISET	Pointer Setter	3
Large Munsterlander	LMUN	Pointer Setter	3
Vizsla	VIZS	Pointer Setter	7
Wirehaired Pointing Griffon	WHPG	Pointer Setter	6
Miniature Poodle	MPOO	Poodle	10
Portuguese Water Dog	PTWD	Poodle	11
Standard Poodle	SPOO	Poodle	10
Toy Poodle	TPOO	Poodle	10
Curly Coated Retriever	CCRT	Retriever	4
Flat-coated Retriever	FCR	Retriever	10
Golden Retriever	GOLD	Retriever	11
Irish Water Spaniel	IWSP	Retriever	10
Labrador Retriever	LAB	Retriever	10
Newfoundland	NEWF	Retriever	10
Nova Scotia Duck Tolling Retriever	NSDT	Retriever	3
Basset Hound	BASS	Scent Hound	10
Beagle	BEAG	Scent Hound	10
Blood Hound	BLDH	Scent Hound	10
Dachshund	DACH	Scent Hound	10
Foxhound	FOXH	Scent Hound	10
Petit Basset Greffon Vendeen	PBGV	Scent Hound	10
Miniature Schnauzer	MSNZ	Schnauzer	10
Standard Schnauzer	SSNZ	Schnauzer	17
Borzoi	BORZ	Sight Hound	10
Greyhound	GREY	Sight Hound	10
Irish Wolfhound	IWOF	Sight Hound	10
Italian Greyhound	ITGY	Sight Hound	17
Scottish Deerhound	DEER	Sight Hound	10
Whippet	WHIP	Sight Hound	10
American Cocker Spaniel	ACKR	Spaniel	10
Cavalier King Charles Spaniel	CKCS	Spaniel	10

English Cocker Spaniel	ECKR	Spaniel	9
English Springer Spaniel	ESSP	Spaniel	10
Field Spaniel	FIEL	Spaniel	4
Airedale Terrier	AIRT	Terrier	3
Australian Terrier	AUST	Terrier	10
Bedlington Terrier	BEDT	Terrier	4
Border Terrier	BORT	Terrier	4
Cairn Terrier	CAIR	Terrier	10
Irish Terrier	IRIT	Terrier	7
Kerry Blue Terrier	KERY	Terrier	4
Norwich Terrier	NOWT	Terrier	10
Parsons Russell Terrier	PARS	Terrier	2
Scottish Terrier	SCOT	Terrier	10
Soft Coated Wheaten Terrier	SCWT	Terrier	4
West Highland White Terrier	WHWT	Terrier	10
Yorkshire Terrier	YORK	Terrier	10
American Hairless Terrier	AHRT	Toy Dog	10
Chihuahua	CHIH	Toy Dog	10
Havanese	HAVA	Toy Dog	10
Maltese	MALT	Toy Dog	2
Miniature Pinscher	MPIN	Toy Dog	10
Papillon	PAPI	Toy Dog	10
Pomeranian	POM	Toy Dog	10
Rat Terrier	RATT	Toy Dog	2
Volpino Italiano	VPIN	Toy Dog	4
Pug Dog	PUG	Toy Spitz	10
Australian Cattle Dog	AUCD	UK	2
Australian Shepherd	AUSS	UK Rural	10
Border Collie	BORD	UK Rural	12
Cardigan Welsh Corgi	CARD	UK Rural	10
Collie	COLL	UK Rural	10
Kelpie	KELP	UK Rural	5
Old English Sheepdog	OES	UK Rural	10
Patagonian Sheepdog	PGOD	UK Rural	146
Pembroke Welsh Corgi	PEMB	UK Rural	10
Shetland Sheepdog	SSHP	UK Rural	10
Dingo	DING	Wild	1
Golden Jackal	GDJK	Wild	2
Grey Wolf	WOLF	Wild	7
Total No. Individuals			**1,735**

Table 2: Regional distribution of New Zealand Huntaway and Heading Dog samples.

Region	**Huntaway**	**Heading Dog**
Hawkes Bay	31	26
Manawatū-Whanganui	30	13
Waikato	32	74
Marlborough	7	3
Canterbury	8	13
Otago	21	40
Southland	60	52
Other	22	25
Total	**211**	**246**

3.4 Genetic Analyses

3.4.1 Cladogram

First, the IIDs in the .fam PLINK file were updated to reflect the breed abbreviation for downstream visualisation, followed by a number as a unique identifier. To build a consensus tree, 100 subsets of the genetic data was randomly generated. The .bim PLINK file that contained the genetic variant IDs was resampled without replacement to produce 100 subsets of the original dataset. Subsets were produced because of PLINK v1.9's limitation in handling duplicate individual and variant IDs, any duplicate variants generated during resampling were automatically removed when generating distance matrices. Genetic distance matrices of each subset were generated using PLINK v1.9 (Purcell et al., 2007), encompassing 129 dog breeds and three wild canid species.

As mentioned above, random down-sampling for the number of Huntaways, Heading Dogs and Patagonian Sheepdogs to 10 individuals per breed was conducted to prevent the large sample sizes of these three breeds from dominating the cladogram, and avoid bias since the majority of other breeds from the dataset have only 10 individuals per breed. The total number of dogs used to generate the cladogram was 1140 individuals. The PLINK functions --distance, --l-ibs, --square and --flat-missing were used to calculate the genetic distances.

The subset distance matrices were then analysed using PHYLIP v3.6 (University of Washington, 2021). The NEIGHBOR program was used to construct neighbour-joining cladograms for each subset (n = 100), and the results from this analysis were further processed using the majority-rules consensus method in the CONSENSE program, which combined the subset results into one consensus tree. To visualise the consensus neighbour-joining cladogram, TreeViewer software was used (Bianchini & Sánchez-Baracaldo, 2024). However, resampling confidence values in the PHYLIP output required formatting adjustments for compatibility with TreeViewer. All resampling confidence values in the .outtree file were enclosed in square brackets, and all circular brackets in the file were replaced with square brackets. For visual clarity, dog breeds were represented as cartooned nodes, with colours indicating their respective breed clades. Breed abbreviations were used as identifiers, and resampling confidence values are represented by the colours of points on the internal nodes.

An unrooted cladogram for the UK rural clade subset was created using the same pipeline as above. This was created using the UK rural and NZ subset that contains 104 dogs.

3.4.2 Haplotype Sharing Analysis

Phasing was required for the identity-by-descent (IBD) haplotype sharing analysis. To do this, the final PLINK dataset was converted into VCF format, then utilised for phasing in Beagle v4.1. The window size was set to 1900 variants following the approach by Ali et al., (2020), the overlap was set to 50 variants and the ‘ibd’ setting was set to ‘true’ for phasing. The resulting IBD file was imported into R Studio for further manipulation.

The breed table was merged with the IBD dataset so that each dog in the dataset was assigned to its recorded breed. All comparisons between dogs of the same breed were removed from the IBD dataset. All known crossbreeds sampled in the “Right Dog for the Job” project were excluded from the IBD dataset to reduce any bias from dogs of other breeds sampled in New Zealand. This reduced the total number of individuals in the dataset to 1,699 dogs. Haplotype lengths were calculated using the ‘end position’ and ‘start position’ columns from the IBD dataset. The haplotype lengths for each unique pair of dogs were then summed to give the total haplotype sharing for that pair of dogs. The total haplotype sharing with each unique dog breed was then averaged for each individual Huntaway and Heading Dog. For example, the haplotype sharing for Huntaway 1 was measured against 10 individual Border Collies. The total haplotype sharing of those 10 Border Collies was averaged for Huntaway 1, and so on for all breeds and all individual Huntaways and Heading Dogs. The averages for each Huntaway and Heading Dog were then plotted onto boxplots (one each for Huntaways and Heading Dogs) in R studio v4.2.1 (R Studio Team, 2020). Each box represents a separate breed or wild canid species that the Huntaway and Heading Dogs were compared against. A 95% threshold was calculated and added to the graph to visualise breeds with significant haplotype sharing. Breeds where the median of the box sits above this threshold are considered statistically significant. The boxes are coloured by the breed clade, while the breed abbreviations are labelled on the x-axes of the graphs. Graphs were constructed with and without showing outliers.

3.4.3 Admixture Analysis

The extent of admixture between the New Zealand Huntaway, Heading Dog and the closely related UK herding breeds in the dataset was evaluated through the model-based clustering algorithm implemented in the ADMIXTURE software v1.3 (Alexander et al., 2009). The UK rural and NZ subset (n = 104 dogs) was used for this analysis. Variants were pruned for linkage disequilibrium using PLINK’s --indep-pairwise function; variants with R^2 values greater than 0.2 with any other variant within a 50-SNP sliding window (advanced by 5 SNPS at a time) were excluded. The resulting pruned dataset contains 36,747 SNPs from 104 dogs.

To reduce prediction error, ADMIXTURE cross-validation (CV) was performed to identify the optimal K value as described in the ADMIXTURE documentation (Alexander et al., 2015). The K value represents the number of ancestral populations assigned during each clustering run. Cross-validation determined the optimal K value to be 4 (CV = 0.57432), although this value was only marginally lower than the CV error calculated for K = 3 (0.57491) and K = 5 (0.57513). Admixture results were visualised in R studio v4.2.1 (R Studio Team, 2020), with plots generated for K = 3, K = 4 and K = 5.

3.4.4 Principal Component Analysis

Principal component analysis (PCA) was performed on the same the LD-pruned UK rural and NZ dataset described above, using the --pca function in PLINK allowing for the visualisation of the genetic differences between the breeds of interest and the most closely related breeds from the UK rural clade. This produced a total of 20 principal components (PCs). Plots of PC1 vs PC2, PC1 vs PC3 and PC2 vs PC3 were produced with R studio version 4.2.1 (R Studio Team, 2020). Points representing individual dogs are coloured according to the breed of dog based on owner reporting or the designations reported by Barrios et al. (2022) and Parker et al. (2017).

3.4.5 Inbreeding Coefficient Estimates

Inbreeding coefficients were estimated for 211 Huntaways and 246 Heading Dogs using two genomic methods (Yoshida et al., 2020). Inbreeding coefficients based on runs of homozygosity (F_{ROH}) were estimated for each animal based on all ROHs as follows:

$$F_{ROH} = \frac{L_{ROH}}{L_{auto}}$$

Where L_{ROH} is the sum of ROH lengths and L_{auto} is the total length of the CanFam3 autosomal genome (2.204675Gb) (NCBI, 2011). L_{ROH} was calculated during ROH analysis using the R script from Gorssen et al. (2021).

Inbreeding coefficients based on the excess of homozygous regions in the genome (F_{HOM}) was calculated in PLINK v1.9 with the function --het as follows:

$$F_{HOM} = \frac{O_{HOM} - E_{HOM}}{N - E_{HOM}}$$

where O_{HOM} is the observed number of homozygous markers in each individual, E_{HOM} is the expected number of homozygous markers under Hardy-Weinberg equilibrium calculated from the allele frequencies estimated based on the sample and N is the total number of markers (Yoshida et al., 2020).

3.4.6 Runs of Homozygosity Island Analysis

Runs of homozygosity island (ROHi) analysis was performed using an R script developed in the study by Gorssen et al. (2021). This script utilises PLINK software (Purcell et al., 2007) for quality control and ROHi analysis, while R Studio is used for the visualisation of results. A subset of New Zealand Huntaways (n=211) and Heading Dogs (n=246) was generated for ROH analysis. The default quality control and ROHi analysis filters were utilised as described in Gorssen et al. (2021). ROH incidence (the percentage of animals with a SNP within an ROH segment for a given population) were plotted as Manhattan plots (Gorssen et al., 2021). Genes within the highest incidence ROH segments per population were collated using the Genome Data Viewer (National Center for Biotechnology Information, 2025).

Chapter 4: Results

4.1 Cladogram

Phylogenetic analysis of 1,140 dogs representing 129 breeds was performed to determine the genetic relationship between the Huntaway and Heading Dog and other formally recognised dog breeds. This analysis constructed neighbour-joining cladograms using PHYLIP v3.6 (see methods section 3.3.1). Figure 3 displays the neighbour-joining cladogram visualised in TreeViewer v2.2.0. Dogs are grouped and coloured by breed clade, with breed cartooning used for visual clarity (refer to Table 1 for breed names and associated breed abbreviations). Breeds in the UK rural clade are highlighted in light blue and includes 12 breeds. The New Zealand Huntaway (HUNT) and Heading Dog (HEAD) appear as distinct branches in dark and light pink, respectively, within the UK rural clade. Based on the oral history of the Huntaway and Heading Dog mentioned above, it was expected that both breeds would cluster within the UK rural clade.

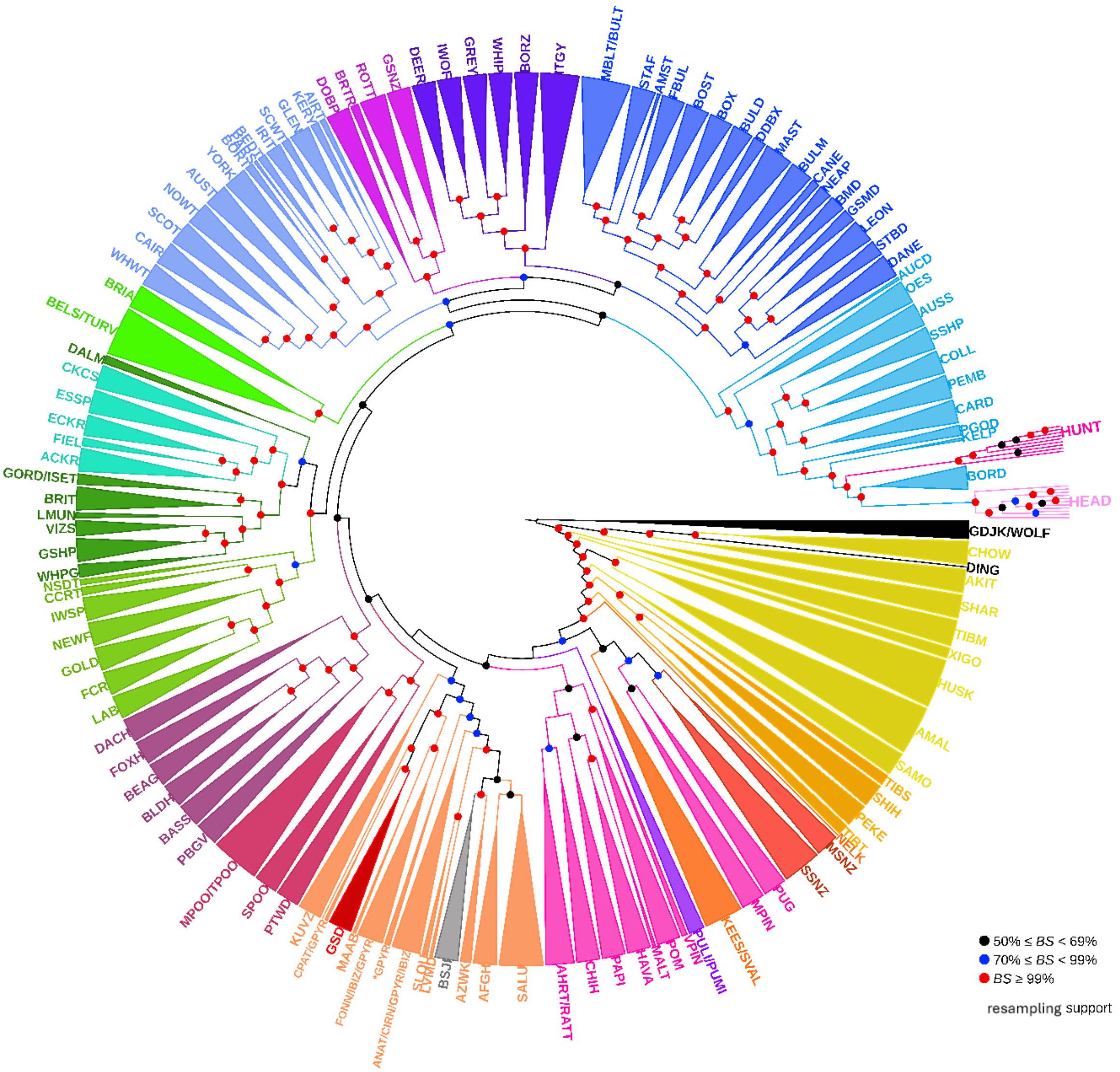

Figure 3: Neighbour-joining cladogram constructed using genetic distance matrices of 1,140 individuals that represent 129 domestic dog breeds and three wild canid species, based on 100 subsets of the original dataset. Each colour represents a breed clade, and individual dogs of the same breed are cartooned for visualisation (refer to Table 1 for breed names and associated breed abbreviations). Breeds in the UK rural clade are in light blue, and the New Zealand Huntaway and Heading Dog are represented by the branches in dark and light pink, respectively, within the UK rural clade. Points on the internal nodes of the cladogram represent resampling confidence values. The UK rural clade is supported ≥99% resampling confidence, and the New Zealand Huntaway and Heading Dogs cluster to the UK rural clade with ≥99% resampling confidence.

In Figure 3, both the Huntaway and Heading Dog cluster with the UK rural clade ≥99% of the time, and the UK rural clade itself shows ≥99% resampling support. Both NZ breeds also form their own breed-specific nodes with ≥99% resampling support. Within the UK rural clade, the Huntaway and Heading Dog form a monophyletic subclade with the Patagonian Sheepdog (PGOD), Australian Kelpie (KELP) and Border Collie (BORD). Additionally, the Heading Dog diverges directly from the Border Collie.

Figure 4 presents an unrooted neighbour-joining cladogram of the UK rural clade generated through the same pipeline as above. This subset represents 104 dogs from the UK rural clade, including the random down-sampled Huntaways, Heading Dogs and Patagonian Sheepdog (see methods section 3.3.1).

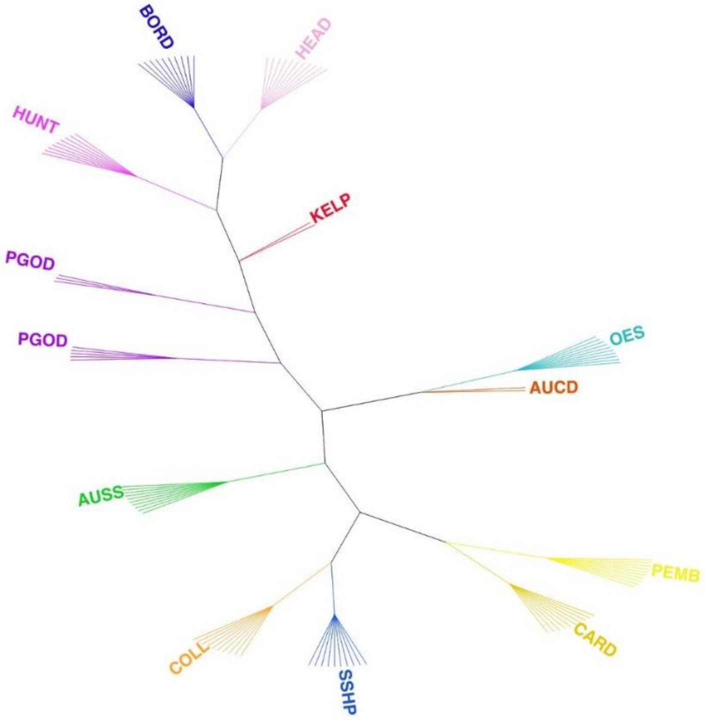

Figure 4: Unrooted neighbour joining tree of the UK rural and NZ subset data frame indicating the relationships between the 12 breeds in the UK rural- and NZ subset data (104 dogs).

Figure 4 further highlights the phylogenetic structure within the UK rural clade, reinforcing the clustering observed in Figure 3. The New Zealand Huntaway and Heading Dog again form distinct nodes, and both breeds remain monophyletic with the Patagonian Sheepdog, Australian Kelpie and Border Collie. The Heading Dog and Border Collie again diverge from one another at a shared node. This cladogram also highlights other close relationships between other breeds in the UK rural clade. For example, it can be seen that the Collie (COLL) and Shetland Sheepdog (SSHP) diverge from each other from a shared node. This is also observed between the Pembroke Welsh Corgi (PEMB) and Cardigan Welsh Corgi (CARD). Unlike other breeds, the Patagonian Sheepdog (PGOD) branches at two locations in Figure 4. In the study by Barrios et al. (2022), it was highlighted that there are geographically associated genetic differences in the Patagonian Sheepdog population. The randomly generated subset of Patagonian Sheepdog individuals used for the cladogram construction likely reflects these genetic differences within the Patagonian Sheepdog population.

4.2 Haplotype Sharing Analysis

Haplotypes of 1,699 dogs were phased using Beagle v4.1 and calculated and visualised using R Studio (see methods section 3.3.2). Each box represents a separate breed or wild canid species that the Huntaway and Heading Dogs were compared against. Breeds where the median of the box sits above the 95% significance threshold are considered statistically significant. The boxes are coloured by breed clade and the breed abbreviations are labelled on the x-axes of the graphs (refer to Table 1 for expanded breed names and associated breed abbreviations). Figures 5 and 6 show identity-by-descent (IBD) haplotype sharing between 127 previously published dog breeds and three wild canids, against the Huntaway and Heading Dog, respectively. To better visualise the overall haplotype sharing patterns at the population level, extreme outliers (defined as individuals with haplotype sharing scores exceeding 1.5 times the interquartile range (IQR) above the third quartile) were excluded from the plots since these individuals are assumed to represent recent crossbreeding events. Figures 7 and 8 show zoomed in views of the haplotype sharing against the UK rural subset for the Huntaway and Heading Dog, respectively. Consistent with the cladogram, it is expected that both the Huntaway and Heading Dog will share larger haplotype lengths with the breeds from the UK rural clade. Considering the oral history of the development of the Huntaway, the Huntaway may also share significant haplotypes from breeds outside of the UK rural clade.

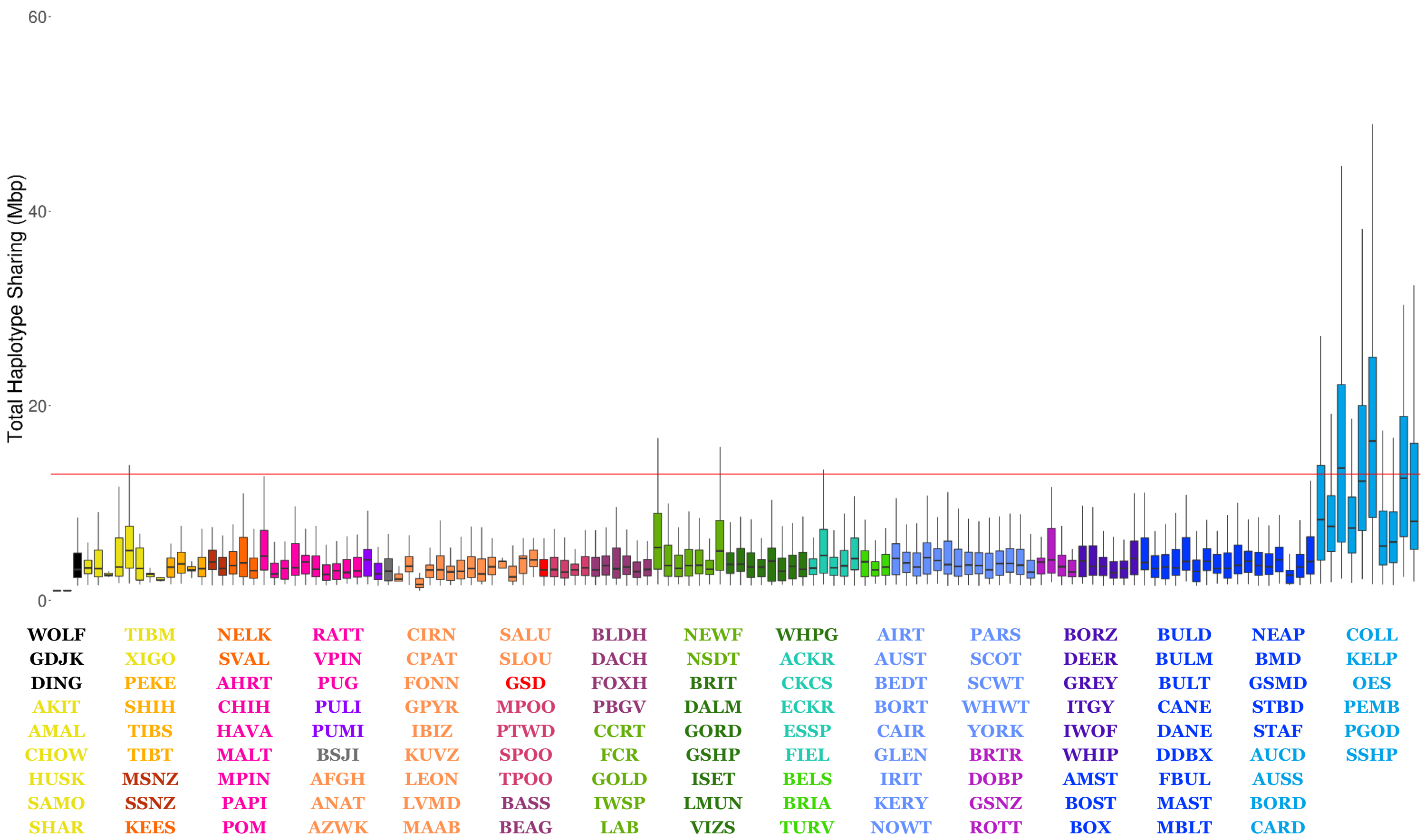

Figure 5: Identity-by-descent (IBD) Haplotype sharing analysis of 128 previously published breeds and three wild canids against the New Zealand Huntaway. Each boxplot represents haplotypes shared between the New Zealand Huntaway and each of the other breeds and species. The y-axis represents the total amount of haplotype sharing in mega base pairs (Mbp). Dog breed abbreviations are on the x-axis and coloured by previously defined breed clades. Breeds within the UK rural clade are represented in light blue. The red horizontal line indicates the 95% significance threshold, and breeds where the total haplotype sharing median values is above this threshold are considered to share significant haplotypes with the New Zealand Huntaway.

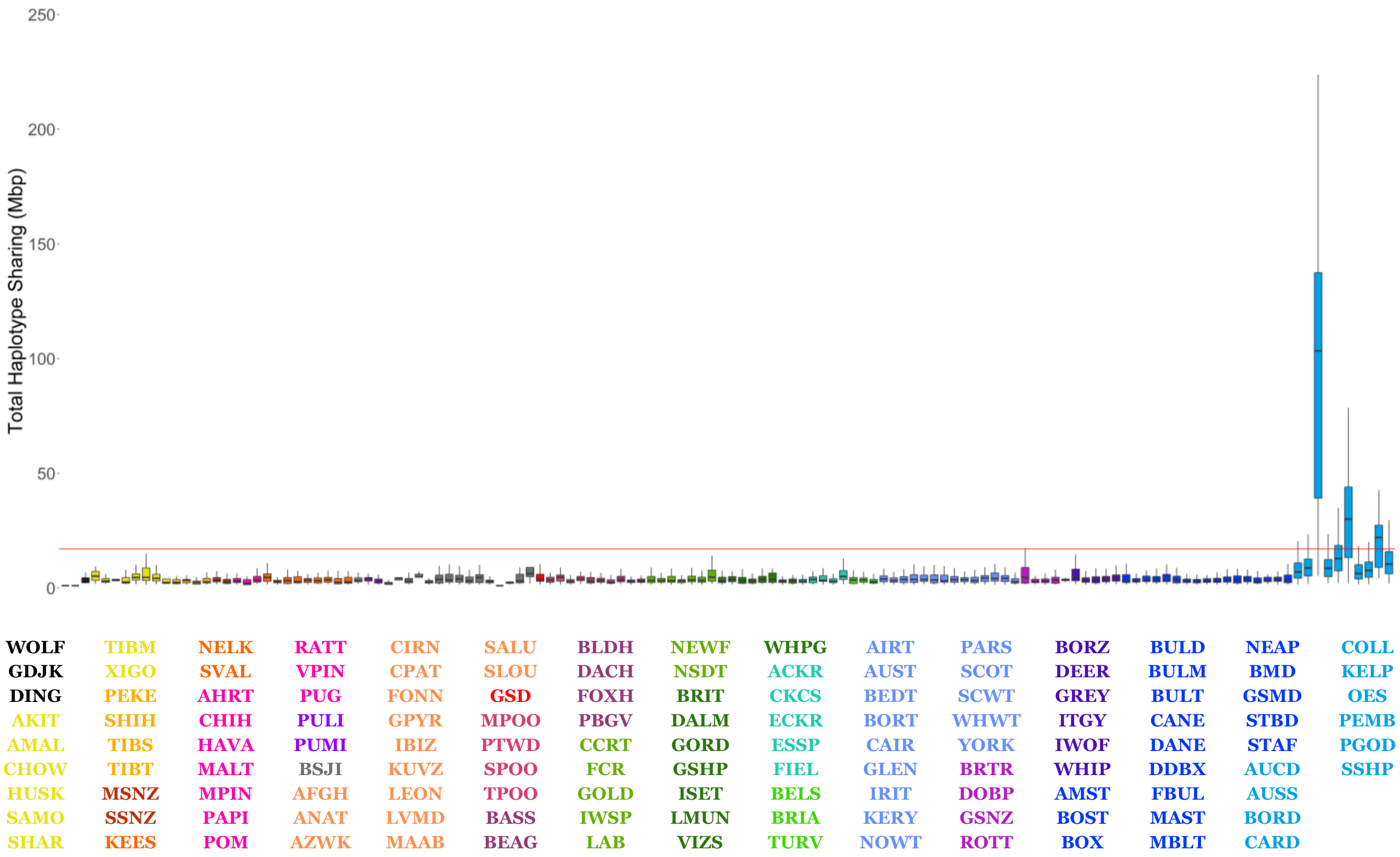

Figure 6: Identity-by-descent (IBD) Haplotype sharing analysis of 128 previously published breeds and three wild canids against the New Zealand Heading Dog. Each boxplot represents haplotypes shared between the New Zealand Heading Dog and each of the other breeds and species. The y-axis represents the total amount of haplotype sharing in mega base pairs (Mbp). Dog breed abbreviations are on the x-axis and coloured by previously defined breed clades. Breeds within the UK rural clade are represented in light blue. The red horizontal line indicates the 95% significance threshold, and breeds where the total haplotype sharing median values is above this threshold are considered to share significant haplotypes with the New Zealand Heading Dog.

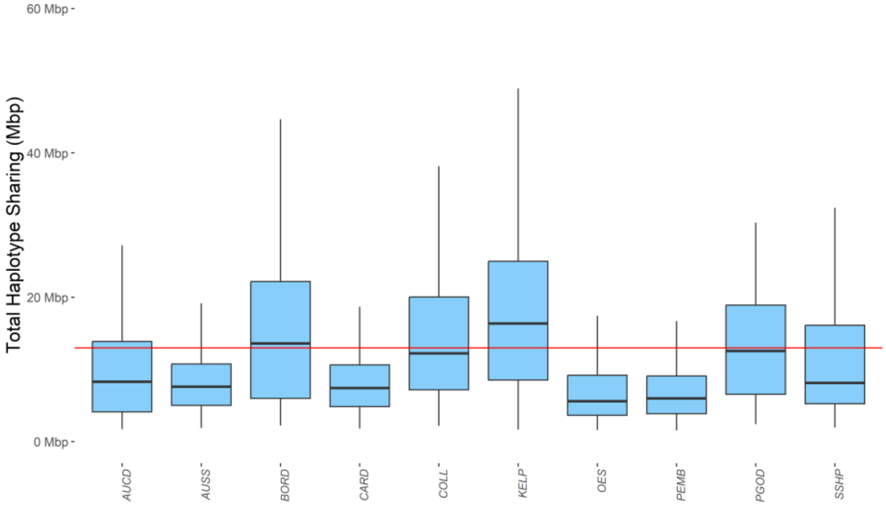

Figure 7: Identity-by-descent (IBD) Haplotype sharing zoomed in on the UK rural subset, against the New Zealand Huntaway. The y-axis represents the total haplotype sharing in mega base pairs (Mbp) and breed abbreviations are on the x-axis. The red horizontal line represents the 95% significance threshold across all breeds, and breeds where the total haplotype sharing median values is above this threshold are considered to share significant haplotypes with the New Zealand Huntaway.

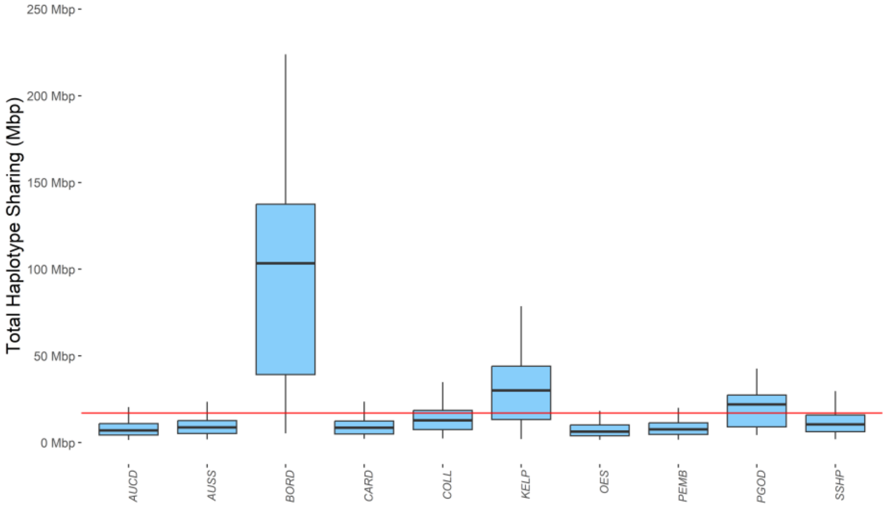

Figure 8: Identity-by-descent (IBD) Haplotype sharing zoomed in on UK rural subset, against the New Zealand Heading Dog. The y-axis represents the total haplotype sharing in mega base pairs (Mbp) and breed abbreviations are on the x-axis. The red horizontal line represents the 95% significance threshold across all breeds, and breeds where the total haplotype sharing median values is above this threshold are considered to share significant haplotypes with the New Zealand Heading Dog.

The Huntaway exhibits strong haplotype sharing with UK rural clade breeds, reflecting shared ancestry. The Border Collie (mean = 16.23 Mbp; median = 13.60 Mbp) and Australian Kelpie (mean = 18.60 Mbp; median = 16.38 Mbp) exceed the significance threshold of 12.989 Mbp. While the Collie (mean = 15.07 Mbp; median = 12.24 Mbp) and Patagonian Sheepdog (mean = 12.84; median = 12.57 Mbp) fall just below the threshold, the haplotype sharing shown here is higher in comparison to other UK rural breeds that have not been mentioned. No breeds outside the UK rural clade show significant haplotype sharing with the Huntaway.

The 95% threshold was calculated at 17.022Mbp for haplotype sharing against the Heading Dog. A similar trend is observed, with the UK rural clade breeds sharing longer average haplotypes with the Heading Dog compared to other breed clades. The Border Collie (mean = 93.38 Mbp; median = 103.44 Mbp), Australian Kelpie (mean = 30.55 Mbp; median = 30.05 Mbp) and the Patagonian Sheepdog (mean = 19.12 Mbp; median = 22.02 Mbp) all show significant haplotype sharing with the Heading Dog. While the Collie does not reach the threshold (mean = 14.14 Mbp; median = 12.90 Mbp), the breed still displays notable sharing. In Figure 6, as with the Huntaway, no breeds outside of the UK rural clade exhibit significant haplotype sharing with the Heading Dog.

4.3 Admixture Analysis

The extent of admixture between the Huntaway, Heading Dog and the closely related UK herding breeds was evaluated using ADMIXTURE software using the LD-pruned subset of 104 dogs (see methods section 3.3.3). Figure 9 presents ADMIXTURE results, highlighting genetic diversity and ancestral contributions within the UK rural clade and the two New Zealand breeds. The lowest cross-validation (CV) error was observed at K = 4 (0.57432), though values for K = 3 (0.57491) and K = 5 (0.57513) were similarly low. Therefore K = 4 was the optimal number of ancestral populations assigned but given the minimal difference between the K = 3, 4 and 5, results for K values 3 to 5 are presented. In this analysis, each colour represents a distinct inferred ancestral component: ancestry 1 (green), ancestry 2 (orange), ancestry 3 (red), ancestry 4 (blue) and ancestry 5 (purple). Unless otherwise stated, all statistics discussed refer to K = 4.

Figure 9: Admixture analysis of the UK NZ subset (104 dogs). Each column represents an individual dog, and each block represents a breed. The number of ancestral populations assigned (K) is labelled on the y-axis. The ADMIXTURE cross validation optimal K value was K = 4, but K values for 3 and 5 are also presented. The breed abbreviations are on the x-axis.

Across all K values, ancestry 1 dominates several breeds, with mean percentages for each breed as follows: Heading Dogs (99.8%), Border Collies (96.1%), Huntaways (80.8%), Australian Kelpies (69.4%), Patagonian Sheepdogs (64.6%), Australian Cattle Dogs (58.1%) and Australian Shepherds (53.5%). Lower levels of ancestry 1 are also seen in the Cardigan Welsh Corgi (20.8%). The Heading Dog and Border Collie show exclusively ancestry 1 profiles, reflecting a strong historical link. Huntaways also show high ancestry 1 levels, but have higher percentages of other ancestral components, including 7% ancestry 2, 9.9% ancestry 3 and 1.5% ancestry 4. Notably, the proportion patterns seen in the Huntaways is quite similar to those of the Patagonian Sheepdogs, Australian Cattle Dogs and Kelpies. This pattern is consistent across all K values analysed. Other close breed relationships are also observed in Figure 7. For example, the percentages of ancestry 3 for K = 4 suggest that the Pembroke Welsh Corgi (PEMB) and Cardigan Welsh Corgi (CARD), which are 66% and 99% respectively, share a close genetic relationship. This is also seen for ancestry 4 between the Collie (COLL) and Shetland Sheepdog (SSHP), where the percentages are 99% and 97%, respectively.

4.4 Principal Component Analysis

Principal component analysis (PCA) was performed using PLINK v1.9 on the LD-pruned subset of 104 dogs to further explore the genetic relationships between the UK rural clade breeds and the New Zealand Huntaway and Heading Dogs (see methods section 3.3.4). The first three principal components, PC1, PC2 and PC3, explain 13.6%, 9.8% and 9.2% of the total genetic variance, respectively. Figure 10 shows PC1 and PC2, Figure 11 shows PC1 and PC3 and Figure 12 shows PC2 and PC3. Breeds are colour coded and labelled on the x-axes.

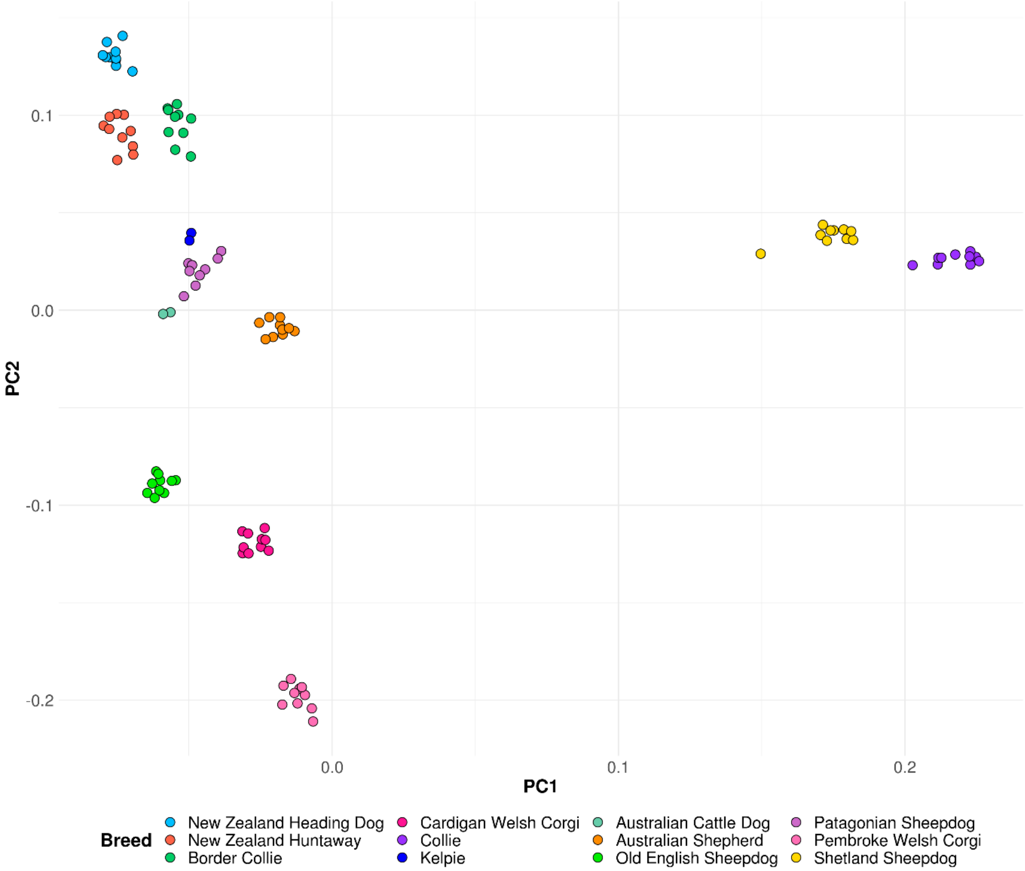

Figure 10: Principal Component Analysis (PCA) of principal component 1 (PC1) on the x-axis and principal component 2 (PC2) on the y-axis, showing the distribution of the 12 breeds in the UK rural clade. PC1 represents 13.6% of the total genetic variation, and PC2 accounts for 9.8%. Dogs are coloured by breed. The New Zealand Huntaway is in dark orange, and the New Zealand Heading Dog is in light blue.

Figure 10 reveals distinct breed-specific clusters for all breeds analysed. Notably, the Huntaway and Heading Dog form clusters that are positioned close together and are primarily separated by PC2. Both New Zealand breeds also cluster closely to the Border Collie. The Patagonian Sheepdog, Australian Cattle Dog and Kelpie cluster closely together, distinct from the New Zealand breeds. The Pembroke Welsh Corgi and Cardigan Corgi, and Collie and Shetland Sheepdog cluster further away from other breeds in the UK rural clade, but it can be seen again that these breed pairs have close genetic relationships, but there is enough genetic variation that these closely related breeds form breed specific clusters.

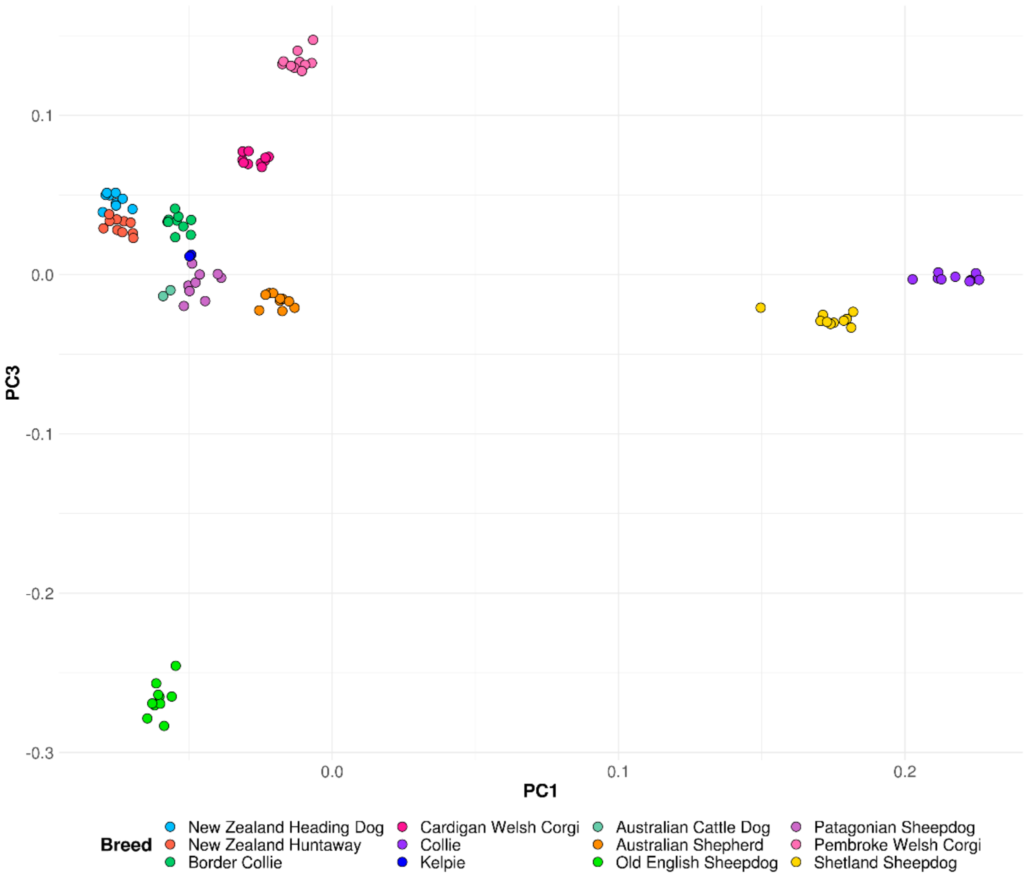

Figure 11: Principal Component Analysis (PCA) of principal component 1 (PC1) on the x-axis and principal component 3 (PC3) on the y-axis, showing the distribution of the 12 breeds in the UK rural clade. PC1 represents 13.6% of the total genetic variation, and PC3 accounts for 9.2%. Dogs are coloured by breed. The New Zealand Huntaway is in dark orange, and the New Zealand Heading Dog is in light blue.

Figure 11 continues to show breed-specific clusters, with the Border Collie, Huntaway and Heading Dog now clustering more closely with the Patagonian Sheepdog and the Australian Breeds. The Huntaway and Heading Dog remain distinct from one another, with PC1 separating them from the Border Collie, and PC3 further distinguishing the Huntaway from the Heading Dog. This figure reinforces the closer genetic relationship between the Patagonian Sheepdog and the Australian breeds, as seen in Figure 10.

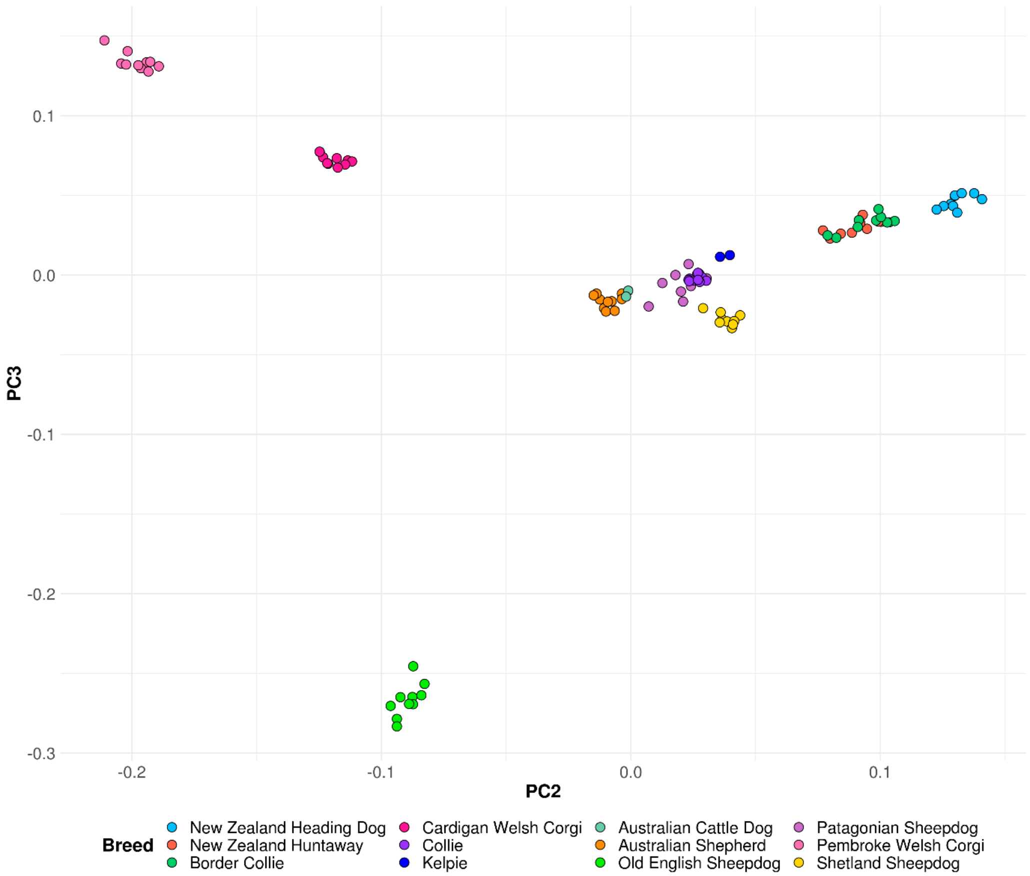

Figure 12: Principal Component Analysis (PCA) of principal component 2 (PC2) on the x-axis and principal component 3 (PC3) on the y-axis, showing the distribution of the 12 breeds in the UK rural

clade. PC2 represents 9.8% of the total genetic variation, and PC3 accounts for 9.2%. Dogs are coloured by breed. The New Zealand Huntaway is in dark orange, and the New Zealand Heading Dog is in light blue.

In Figure 12, a larger group is formed by the Patagonian Sheepdog, the Australian breeds, the Collie and the Shetland Sheepdog. However, the Border Collie, Huntaway and Heading Dog remain distinct from this larger group. Notably, the Border Collie and Huntaway cluster together in PC2 vs. PC3, while the Heading Dog forms its own distinct cluster. This pattern likely reflects the closer genetic association between the Huntaway and breeds like the Patagonian Sheepdog and Australian Kelpie, in contrast to the Heading Dog.

Overall, Figures 10, 11 and 12 demonstrate that the New Zealand Huntaway and Heading Dog share a close genetic relationship, but form distinct breed-specific populations, both from one another and from other herding breeds originating from the United Kingdom. The clustering patterns also indicate that the Border Collie is closely associated with both New Zealand breeds. The Huntaway also clusters closer to the Patagonian Sheepdog and Australian Kelpie in comparison to the Heading Dog.

4.5 Inbreeding Analysis

Inbreeding coefficients were calculated for 211 Huntaways and 246 Heading Dogs using F_{ROH} and F_{HOM} (see methods section 3.3.5). Table 2, and Figures 13, 14 and 15 present descriptive plots of the runs of homozygosity (ROH) identified through the calculation of $F_{ROH.}$ A total of 15,584 and 22,644 runs of homozygosity (ROH) were identified in New Zealand Huntaways and Heading Dogs, respectively. Figure 16 presents the F_{ROH} (A) and F_{HOM} (B) scores calculated for each Huntaway and Heading Dog. As mentioned above, dog breeds selected for functionality over aesthetic traits tend to have lower inbreeding scores (Barrios et al., 2022; Dreger et al., 2016), therefore we expect both breeds to have lower than average inbreeding scores in comparison to other formally recognised dog breeds.

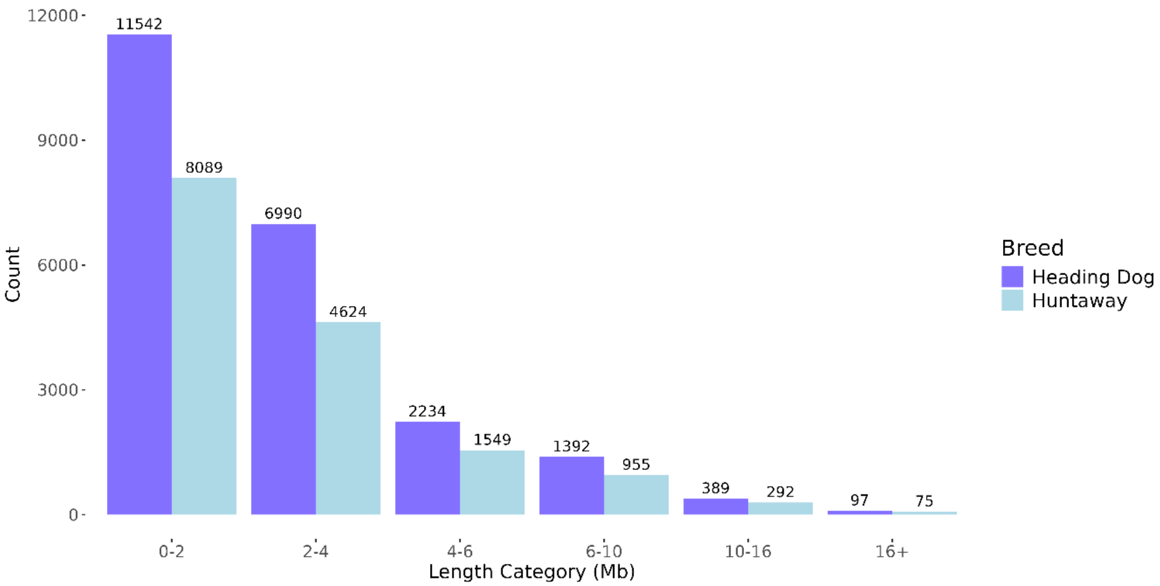

Figure 13: Observed counts of runs of homozygosity (ROH) for each class of length (Mb).

Table 3: Percentage of the total number of runs of homozygosity (ROH) identified per class of length.

	0-2Mb	2-4Mb	4-6Mb	6-10Mb	10-16Mb	16+Mb	Total No. ROH
NZ Huntaway	51.90%	26.67%	9.94%	6.13%	1.87%	0.48%	15,584
NZ Heading Dog	50.97%	30.87%	9.86%	6.15%	1.72%	0.43%	22,644

In both breeds, the majority of ROH were short in length, with ROH between 0-2Mb comprising more than 50% of the total number of ROH detected. Despite differences in the total number of ROH identified between breeds, the proportional distribution across length classes was very similar. Long ROH exceeding 16Mb were observed in both populations, with 75 identified in Huntaways and 97 in Heading Dogs. These longer ROH represented only 0.48% and 0.43% of the total number of ROH in Huntaways and Heading Dogs, respectively. The longest ROH observed in the Huntaway was 53.536Mb, while in the Heading Dogs the longest ROH was 56.727Mb.

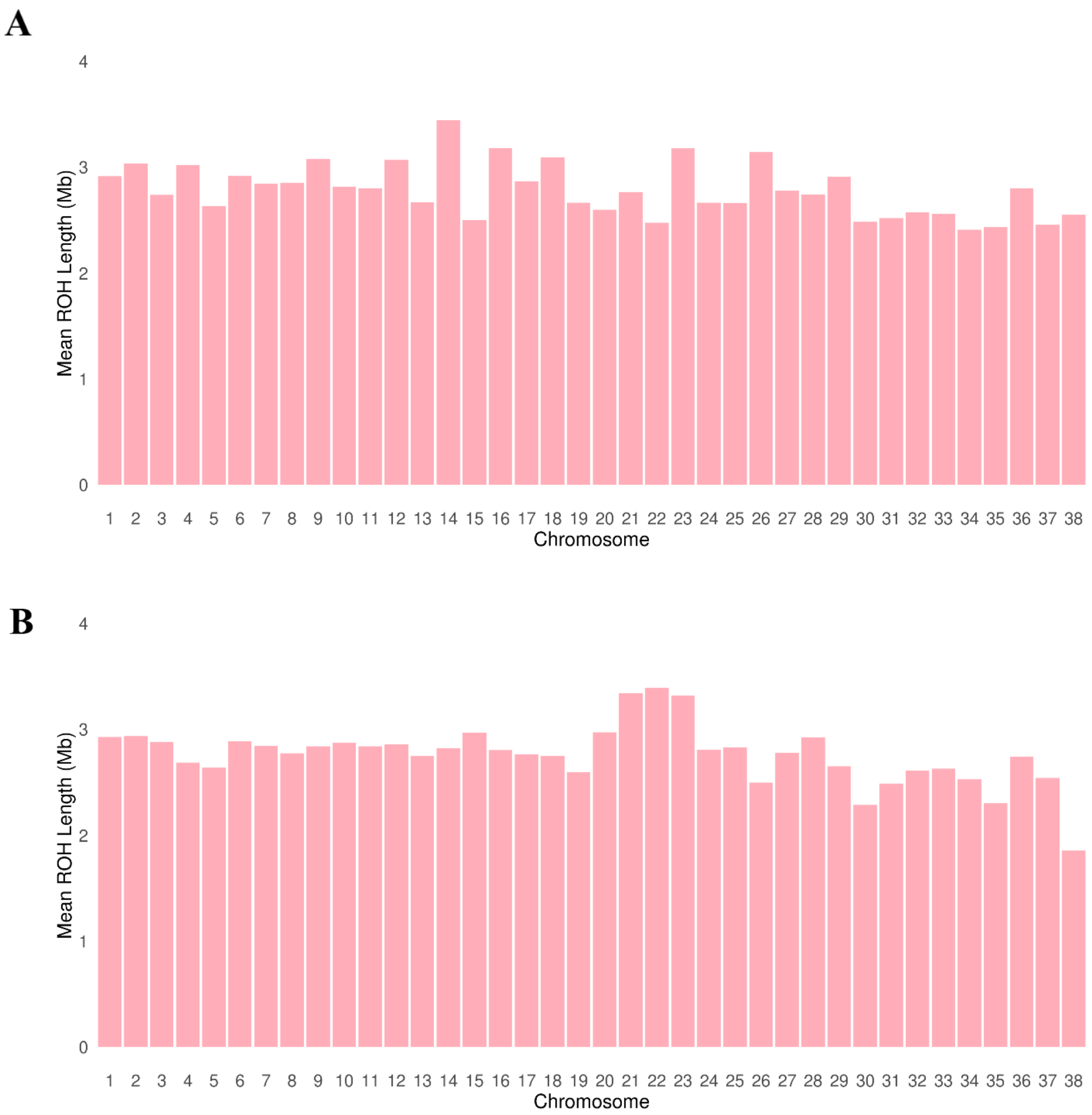

Figure 14: Mean runs of homozygosity (ROH) Length observed in each autosomal chromosome Huntaway (A), Heading Dog (B). The y-axis represents the mean ROH length in megabases (Mb) and the x-axis represents the chromosome number.

ROH were identified on all autosomes in both breeds. In New Zealand Huntaways, the mean ROH length exceeded 3Mb on chromosomes 2, 4, 9, 12, 14, 16, 18, 23 and 26, representing 21.5% of all autosomes. In contrast, only chromosomes 21, 22 and 23 in Heading Dogs had a mean ROH length greater than 3Mb, representing 7.89% of autosomes. The shortest mean ROH length in Huntaways was observed on chromosome 34 (2.41Mb), while in Heading Dogs, the shortest was on chromosome 38 (1.85Mb). The longest mean ROH length in

Huntaways occurred on chromosome 14 (3.44Mb), whereas in Heading Dogs, it was observed on chromosome 22 (3.39Mb).

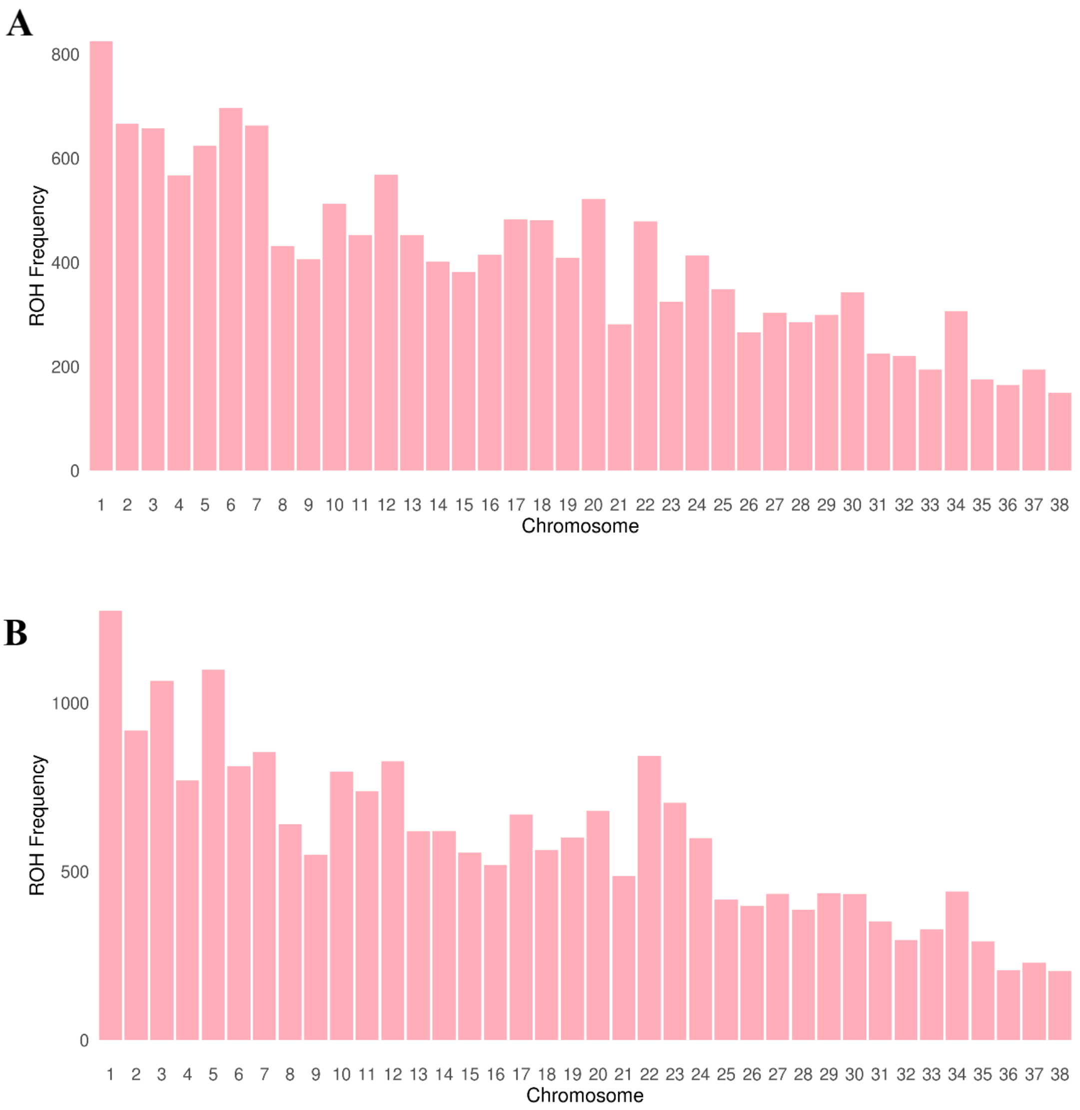

Figure 15: Frequencies of runs of homozygosity (ROH) per autosomal chromosome in the New Zealand Huntaway (A) and Heading Dog (B). The y-axis represents the number of ROH (frequency) in each chromosome, and the x-axis represents the chromosome number.

The frequencies of ROH per chromosome are presented in figure 15. In both breeds, the highest ROH frequencies were observed on chromosome 1, with 825 ROH identified in New Zealand Huntaways and 1,273 ROH identified in Heading Dogs. The distribution of ROH frequency per chromosome is influenced by the length of each chromosome and as expected, the distribution of ROH frequency is associated with the length of each chromosome with longer chromosomes generally exhibiting higher ROH counts than shorter ones.

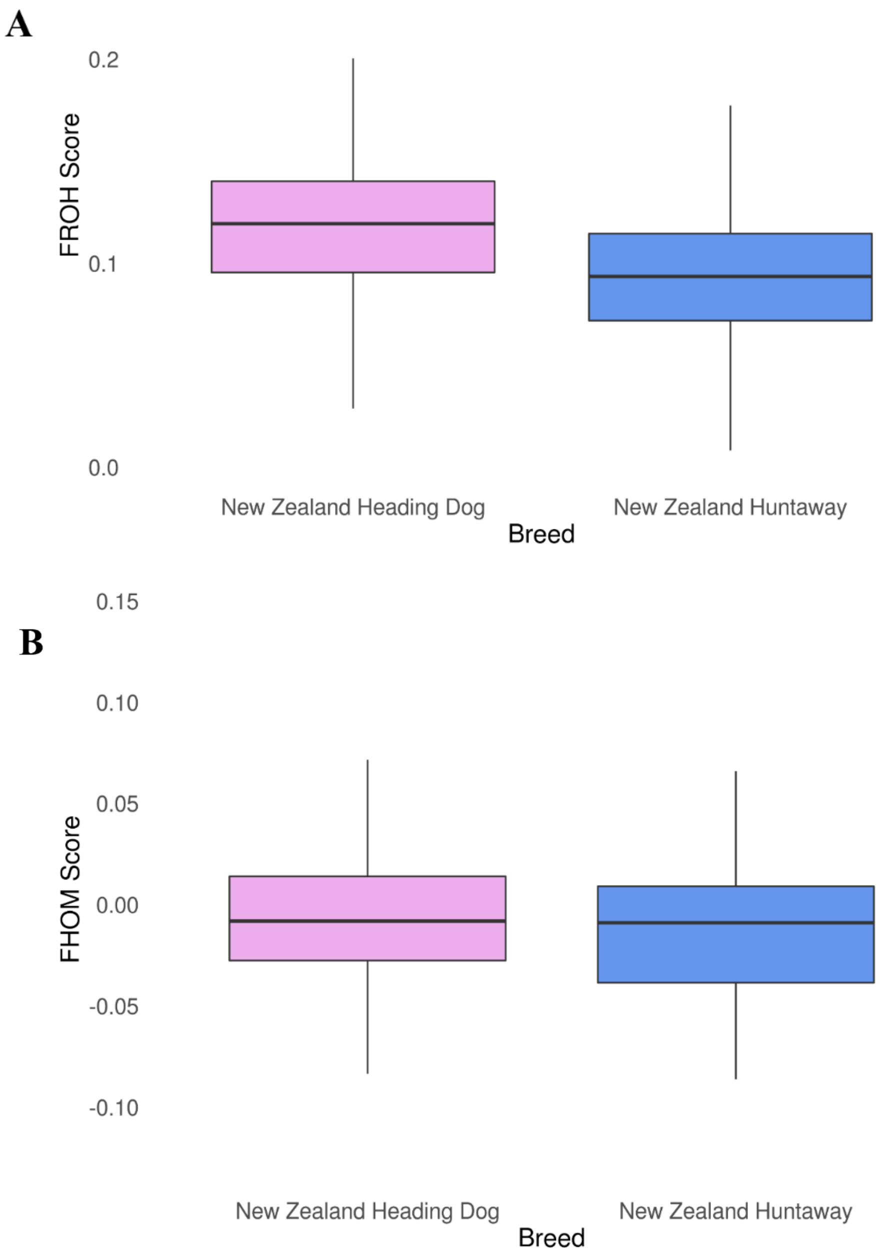

Figure 16: Boxplots of F_{ROH} (A) and F_{HOM} (B) for the New Zealand Huntaway and Heading Dog. The y-axes represent the FROH score (A) and FHOM score (B), and the x-axes represent the breed.

The inbreeding levels in New Zealand Huntaways and Heading Dogs were estimated using F_{ROH} (Figure 16A) and F_{HOM} (Figure 16B) calculations for each Huntaway and Heading Dog and compared the results between breeds. New Zealand Huntaways had an average F_{ROH} of 0.094, whereas Heading Dogs had an average F_{ROH} of 0.119. The minimum and maximum scores observed for Huntaways were 0.002 and 0.268, respectively, while for Heading Dogs, the minimum and maximum scores were 0.018 and 0.274, respectively.

For F_{HOM}, Huntaways had an average score of 0.001, while Heading Dogs had an average score of 0.002. The minimum and maximum F_{HOM} scores observed for Huntaways were -0.122 and 0.224, respectively, while for Heading Dogs, these scores ranged from -0.154 to 0.239.

4.6 Runs of Homozygosity Island Analysis

Runs of homozygosity island (ROHi) analysis was conducted using an R script developed in the study by Gorssen et al. (2021) (see methods section 3.3.6). Figures 17 and 18 show the ROH incidence plot for the Huntaway and Heading Dog, respectively. The Manhattan plots provide an overview of population-specific baseline ROH levels and ROH islands (ROHi). Table 3 presents the regions of the genome where ROHi were identified for the Huntaway and the Heading Dog. Table 4 presents the genes within the ROHi identified in the two breeds.

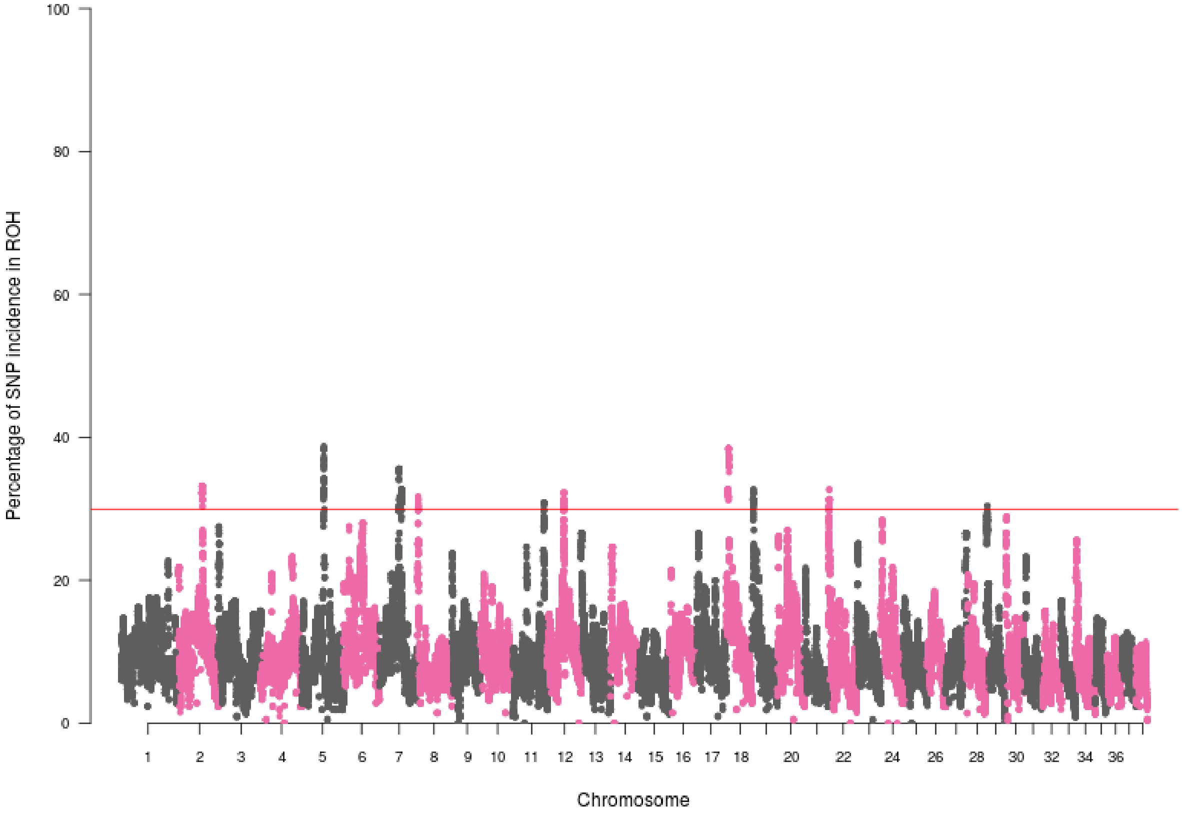

Figure 17: Incidence plots of single nucleotide polymorphisms (SNPs) in runs of homozygosity (ROH) for the New Zealand Huntaway. The y-axis represents the percentage of animals with a SNP within an ROH and the x-axis represents the chromosome number. The red horizontal line indicates the threshold where SNPs above this line are defined to sit in runs of homozygosity islands (ROHi).

In Figure 17, the population-specific minimal threshold based on ROH incidence distribution for detection of ROHi was 30%, meaning an ROH has to be present in at least 30% of the population to be included in an ROHi. In New Zealand Huntaways, ROH incidence levels generally sit between 0-20%. ROHi are observed on chromosome 2, 5, 7, 8,11, 12, 18, 19, 22 and 29.

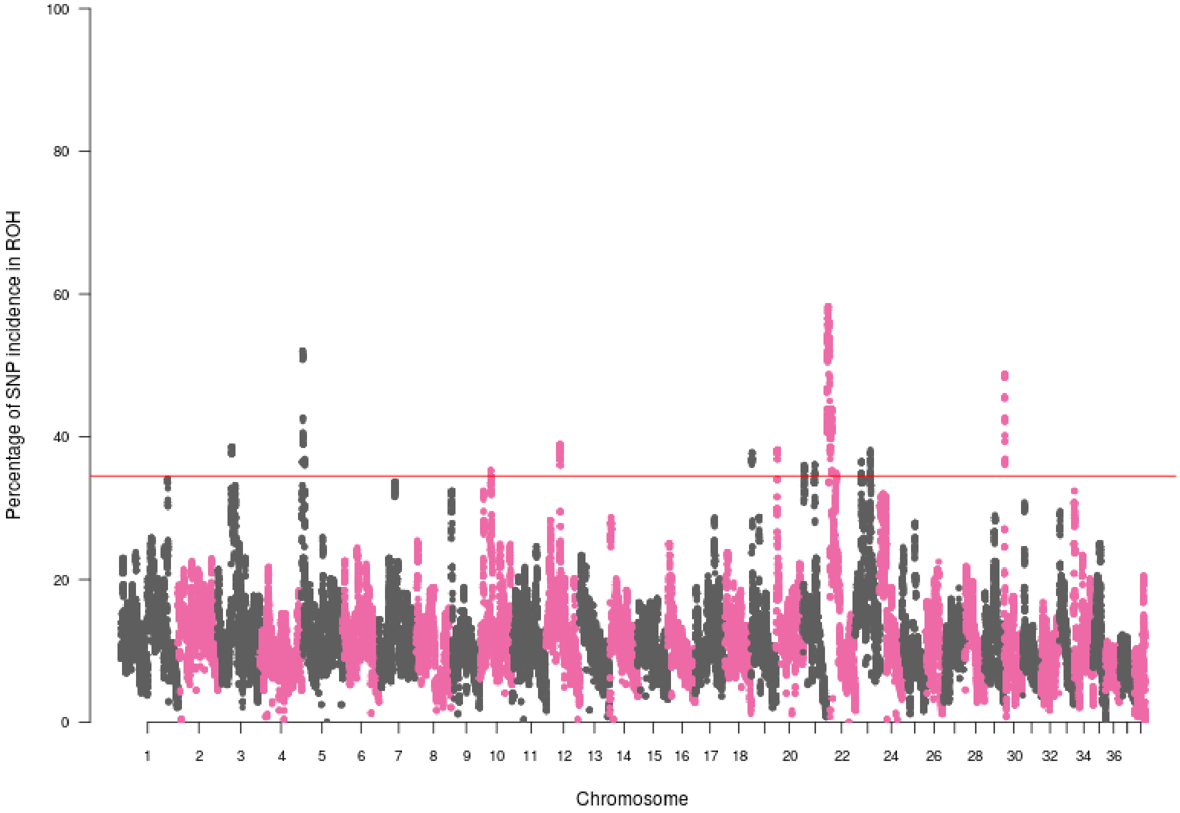

Figure 18: Incidence plots of single nucleotide polymorphisms (SNPs) in runs of homozygosity (ROH) for the New Zealand Heading Dog. The y-axis represents the percentage of animals with a SNP within an ROH and the x-axis represents the chromosome number. The red horizontal line indicates the threshold where SNPs above this line are defined to sit in runs of homozygosity islands (ROHi).

In Figure 18, the minimal threshold for detection of ROHi was 38.3%. ROH incidence levels generally sit between 0-25%, and ROHi are observed on chromosomes 3, 5, 10, 12, 19, 20, 21, 22, 23 and 30.

Table 4: Runs of homozygosity island (ROHi) regions for the New Zealand Huntaway and Heading Dog on the CanFam3 reference genome. The genomic region of each ROHi is calculated in one mega base (Mb) bins. Only chromosomes that had ROHi are listed.

Chromosome	NZ Huntaway	NZ Heading Dog
2	50-52	
3		30-31
5	46-48	2-4 ; 7-8
7	41-43 ; 46-48	
8	1-2	
10		23-25
11	67-68	
12	35-36	27-30
18	3-6	
19	1-3	0-3
20		2-3
21		1-3 ; 24-25
22	2-3	0-12 ; 19-21
23		11-14 ; 31-33
29	2-3	
30		1-3

Table 4 indicates the location of ROHi for New Zealand Huntaways and Heading Dogs. A total of 11 ROHi were identified in Huntaways, and 13 ROHi were identified in Heading Dogs. For both breeds, ROHi tend to sit within regions of chromosomes that are 1-2Mb in length. The largest ROHi for Huntaways is on chromosome 18 (3Mb), and for Heading Dogs, there is a very large ROHi on chromosome 22 (12Mb).

Table 5: Genes identified within runs of homozygosity islands (ROHi) in the New Zealand Huntaway and Heading Dog on the CanFam3 reference genome. Genes listed in bold have been previously studied in domestic dogs.

Breed	**Chromosome**	**Position (Mb)**	**Genes in ROHi**
Huntaway	2	50-52	HTR1A, RNF180, RGS7BP, FAM159B, SREK1IP1, CWC27, ADAMTS6, CENPK, PPWD1, TRIM23, TRAPPC13, SGTB, NLN, ERBIN, SREK1
Huntaway	5	46-48	ROR1, PGM1, EFCAB7, ITGB3BP, ALG6, FOXD3, ATGFC, DOCK7, ANGPTL3, USP1, KANK4, L1TD1, PATJ
Huntaway	7	41-43	ARHGEF11, PEAR1, NTRK1, INSRR, SH2D2A, PRCC, HDGF, MRPL24, RRNAD1, ISG20L2, CRABP2, NES, BCAN, HAPLN2, GPATCH44, NAXE, TTC24, IQGAP3, MEF2D, MIR9-1, RHBG, TSACC, CCT3, GLMP, TMEM79, SMG5, PAQR6, BGLAP, SLC25A44, SEMA4A, LMNA, MEX3A, RAB25, LAMTOR2, UBQLN4, SSR2, ARHGEF2, KIAA0907, RIT1, SYT11, GON4L, MSTO1, DAP3, ASH1L, RUSC1, FDPS, PKLR, HCN3, CLK2, SCAMP3, FAM189B, GBA, MTX1, THBS3, MIR92B, MUC1, TRIM46, KRTCAP2, DPM3, SLC50A1, EFNA1, EFNA3, ADAM15, DCST1, DCST2, ZBTB7B, LENEP, FLAD1, CKS1B, SHC1, PYGO2, PBXIP1, PMVK, KCNN3, ADAR, CHRNB2, UBE2Q1, SHE, IL6R, ATP8B2, AQP10, HAX1, UBAP2L
Huntaway	7	46-48	SETBP1, SYT4, RIT2
Huntaway	8	1-2	CATSPERB, TC2N, FBLN5, TRIP11, ATXN3, CPSF2, **SLC24A4**, RIN3, LGMIN, GOLGA5, CHGA, ITPK1
Huntaway	11	67-68	KIAA1958, INIP, SNX30, SLC46A2, ZFP37, SLC31A2, FKBP15, SLC31A1, CDC26, PRPF4, RNF183, WDR31, BSPRY, HDHD3, ALAD, POLE3, RGS3
Huntaway	12	35-36	KCNQ5, DPPA5, OOEP, KHDC3L, DDX43, MB21D1, MTO1, EEF1A1, SLC17A5, CD109
Huntaway	18	3-6	VSTM2A, SEC61G, EGFR
Huntaway	19	1-3	INPP4B, IL15, ZNF330, UCP1, ELMOD2, MGAT4D, CLGN, SCOC, MAML3
Huntaway	22	2-3	**KPNA3**, EBPL, **RCBTB1**, PHF11, **SETDB2**, **CAB39L**, **CDADC1**, **MLNR**, **FNDC3A**, **CYSLTR2**
Huntaway	29	2-3	SNTG1
Heading Dog	3	30-31	IQGAP2, SV2C, POC5, POLK, COL4A3BP, HMGCR
Heading Dog	5	2-4	OPCML, NTM

Heading Dog	5	7-8	**KIRREL3**
Heading Dog	10	23-25	NFAM1, TCF20, MIR8829, CYP2D15, NDUFA6, SMDT1, FAM109B, NAGA, WBP2NL, SEPT3, CENPM, TNFRSF13C, SHISA8, MIR33A, MEI1, SNU13, XRCC6, DESI1, PMM1, CSDC2, POLR3H, ACO2, PHF5A, TOB2, TEF, ZC3H7B, RANGAP1, CHADL, L3MBTL2, EP300, RBX1, XPNPEP3, DNAJB7, ST13, SLC25A17, MCHR1, MKL1, SGSM3, ADSL, TNRC6B
Heading Dog	12	27-30	PTP4A1, PHF3, EYS
Heading Dog	19	0-3	GAB1, USP38, INPP4B, IL15, ZNF330, UCP1, ELMOD2, MGAT4D, CLGN, SCOC, MAML3
Heading Dog	20	2-3	KBTBD12, SEC61A1, RUVBL1, EEFSEC, DNAJB8, GATA2, RPN1, RAB7A, H1FX, HMCES, COPG1, CNBP, ISY1, RAB43
Heading Dog	21	1-3	CNTN5
Heading Dog	21	24-25	P4HA3, PPME1, C2CD3, UCP3, UCP2, DNAJB13, PAAF1, CHCHD8, MRPL48, RAB6A, PLEKHB1, FAM168A, RELT, ARHGEF17, P2RY6, P2RY2
Heading Dog	22	0-12	NEK5, ALG11, **ATP7B**, CCDC70, WDFY2, INTS6, SERPINE3, FAM124A, RNASEH2B, MIR15A, MIR16-1, KCNRG, TRIM13, SPRYD7, **KPNA3**, EBPL, **RCBTB1**, PHF11, **SETDB2**, **CAB39L**, **CDADC1**, **MLNR**, **FNDC3A**, **CYSLTR2**, RCBTB2, RB1, LPAR6, ITM2B, MED4, NUDT15, SUCLA2, HTR2A, ESD, LRCH1, RYBCNL, LRRC63, LCP1, CPB2, ZC3H13, SIAH3, SPERT, ERICH6B, COG3, SLC25A30, TPT1, GTF2F2, KCTD4, GPALPP1, NUFIP1, MIR8814, TSC22D1, SERP2, SMIM2, LACC1, CCDC122, ENOX1, DNAJC15, EPSTI1, FAM216B, TNFSF11, AKAP11, DGKH, VWA8, RGCC, NAA16, MTRF1, KBTBD7, WBP4, ELF1, SUGT1, CNMD, MIR759, OLFM4
Heading Dog	22	19-21	*no characterised genes*
Heading Dog	23	11-14	ULK4, TRAK1, CCK, LYZL4, VIPR1, SEC22C, SS18L2, NKTR, ZBTB47, KLHL40, HHATL, CCDC13, HIGD1A, ACKR2, ZNF662, FAM198A, POMHNT2, MIR8806, DYNC1LI1, CMTM6, CMTM7, CMTM8, GPD1L, OSBPL10, STT3B, GADL1, TGFBR2
Heading Dog	23	31-33	RYK, AMOTL2, ANAPC13, CEP63, KY, **EPHB1**, PPP2R3A, MSL2, PCCB, STAG1
Heading Dog	30	1-3	CHRM5, AVEN, **RYR3**, FMN1, GREM1, SCG5, ARHGAP11A, GJD2, ACTC1, AQR, ZNF770

Table 4 displays the genes associated with the ROHi in each breed. Genes that have not yet been assigned a gene symbol (LOC genes) are not listed. Genes in bold (n = 13) have been previously studied in dogs and are suggested or known to be linked to traits or diseases in dogs and will be discussed further. A total of 391 genes are reported in this table, of which 176 genes were reported in ROHi identified in Huntaways and 215 genes were reported in ROHi in Heading Dogs.

Chapter 5: Discussion

5.1 Dataset Standardisation

SNP microarrays are a cost-effective method for generating genotype calls for a generally large, fixed number of SNP markers. With the increasing number of microarray platforms available, cross-platform integration of SNP data enables the comparison, replication and combination of results obtained by independent research groups. To investigate the genetic relationships of the New Zealand Huntaway and New Zealand Heading Dog with other breeds, genome-wide SNP data from 211 Huntaways and 246 Heading Dogs were combined with previously published genotypes from 1,256 individuals representing 127 recognised dog breeds and three wild canids. These datasets were generated using two leading high-density genotyping platforms: the Affymetrix Axiom CanineHD Array (Thermofisher, 2024), and the Illumina CanineHD Whole-Genome Genotyping BeadChip (Illumina, 2025a). As these platforms differ in SNP marker content, marker naming conventions and file output formats, standardising the datasets from the two platforms was essential to minimise platform-specific biases that could have influenced the results in downstream analyses. Cross-platform compatibility of canine SNP arrays has received limited investigation.

Because PLINK software was required for downstream genetic analyses, the initial analysis focus was to get the publicly sourced data into PLINK format. One conversion pipeline that was identified in the literature is to convert the raw Illumina IDAT files (available in the public repositories) into VCF files using the BCFtools plugins ‘idat2gtc’ and ‘GTC2VCF’ (GitHub, 2024), then subsequently converting these into PLINK binary format using PLINK (Purcell et al., 2007). The resulting PLINK files then match the file format used for the Affymetrix SNP data, facilitating the merging of the two datasets. Once in PLINK format and prior to merging, both datasets underwent a series of quality control procedures applied using PLINK (Purcell et al., 2007). Variants located on sex chromosomes or unplaced variants (e.g. those on unassembled scaffolds) were excluded to avoid confounding effects from sex-linked inheritance patterns or ambiguous genomic positioning. Additionally, both SNPs and individuals with greater than 10% missing genotype calls were removed. This filtering step aligns with standard practices in large-scale genotype QC pipelines, where thresholds ranging from 5-20% are commonly used to exclude unreliable data and minimise genotyping error rates (Marees et al., 2018; PLINK, 2023; Saravanan et al., 2022).

The phylogenetic methods utilised in this study rely on the assumption that the genetic data being analysed represents a consistent and comparable set of loci across all samples. Therefore, it was essential to retain only those SNP markers that were shared between the two genotyping platforms. The inclusion of platform-specific markers would have introduced bias, potentially leading to inferred genetic variation that were due to differences in genotyping platform markers, rather than true biological variation. A total of 160,465 SNPs were identified as common to both datasets and were retained for further analysis. This number represents 93.23% of the total SNP markers on the Illumina CanineHD Whole-Genome Genotyping BeadChip and 22.52% of the total SNP markers on the larger Affymetrix Axiom CanineHD Array. Additional standardisation was performed using Genotype Harmoniser (Deelen et al., 2014) to minimise further discrepancies in SNP identifiers, genomic positions, reference/alternate allele orientation and strand alignment between the common SNPs on the two platforms. Problematic ambiguous SNPs were also identified and subsequently removed to avoid the introduction of further errors due to strand misidentification. This reduced the datasets to a total of 151,688 high-quality, cross-platform standardised SNP markers. The retention of this number of SNP markers in this study is comparable to the marker densities used in similar phylogenetic studies. For instance, Barrios et al. (2022) used a dataset containing 150,069 SNP markers, Ali et al. (2020) used 150,069 SNP markers and Dreger et al. (2016) used approximately 150,000 SNP markers. This finding provides confidence that the dataset used for this study retains enough marker density to tag the majority of haplotypes present in the population, and allowing effective investigation of genetic diversity, breed to breed relationships, population structure, inbreeding and ROH for the New Zealand Huntaway and Heading Dog breeds.

One of the issues highlighted by this study is the current lack of standardised tools for cross-platform integration of SNP array data from different providers, such as Affymetrix and Illumina. Given the widespread use of SNP array technologies, and the increasing range of available genotyping platforms, the development of cross-platform harmonisation methods would enable researchers to integrate datasets from multiple sources, allowing for the generation of larger, more robust datasets. The data processing and harmonisation workflow used for this study provides a framework for integrating SNP data generated from the Affymetrix Axiom Canine HD Array and the Illumina CanineHD Whole-Genome Genotyping BeadChip. While the current workflow has not been tested on other SNP arrays, it could be tested and adapted for cross-platform harmonisation between other Affymetrix and Illumina microarray panels, provided that a sufficient proportion of SNPs are shared between platforms.

5.2 Phylogenetic Relationship Among Dog Breeds

Herding dogs have been utilised for centuries all around the world to fill various roles (Ostrander et al., 2017). The precise date for the arrival of shepherd working dogs in New Zealand is unknown, but it likely coincided with the introduction of sheep from Britain during the mid-1800s, suggesting that the origins of the New Zealand Huntaway and Heading Dog lie in historic working dog populations that existed in the UK prior to the Victorian-era dog breed explosion (Freedman & Wayne, 2017; Ostrander et al., 2017).

This study confirms that both the New Zealand Huntaway and the Heading Dog can be traced back to historic working breeds from the UK and form a part of the UK rural subclade in the domestic canine family tree. The breeds within this clade were all developed in Europe or descended from UK herding dogs, and all bear traits characteristic of dogs bred to herd livestock. This is supported by the neighbour-joining cladogram generated using genetic distance matrices from 100 subsets, which assigned both the New Zealand Huntaway and Heading Dog to the UK rural breed clade with ≥99% resampling support. Both the New Zealand Huntaway and Heading Dog form a well-supported monophyletic clade with the Patagonian Sheepdog (PGOD), Australian Kelpie (KELP) and the Border Collie (BORD), and this is evident in the phylogeny generated (Figure 3). All three of these formally recognised breeds are traditionally used for herding livestock, performing working roles similar to those of the New Zealand Huntaway and Heading Dog (Ács et al., 2019; Barrios et al., 2019; Chew et al., 2019). This finding is further supported by identity-by-descent haplotype sharing analysis. For the Huntaway (Figures 5 and 7), statistically significant haplotype sharing was observed with both the Border Collie and the Kelpie. High levels (though not statistically significant) of sharing were also seen with the Patagonian Sheepdog and the Collie. The Heading Dog (Figures 6 and 8) showed significant haplotype sharing with the Border Collie, Kelpie and the Patagonian Sheepdog. While haplotype sharing with the Collie did not reach statistical significance, it fell just below the significance threshold. In general, for both breeds haplotype sharing was substantially higher with breeds from the UK rural clade than with those from outside of it.

5.3 The Divergence of Historic UK Herding Breeds

The divergence of UK herding breeds such as the Patagonian Sheepdog, Australian Kelpie, Border Collie, New Zealand Huntaway and Heading Dog has occurred in a manner consistent with historical European migration patterns, and the isolation of these dog populations has resulted in distinct phenotypes that are better suited to their environment. Barrios et al. (2022) proposed that the Patagonian Sheepdog represents the closest living descendant of the original UK herding dog populations, as they descended from working dogs bought by European immigrants between 1877 and 1910. Additionally, the Australian Kelpie is thought to have originated from Scottish Smooth Collies and working farm Collies, with the breed subsequently refined from a small number of individuals introduced to Australia in the late 19th century (Chew et al., 2019). The timing of migration in these dog breeds is very similar to that of the Huntaway and Heading Dogs, which are thought to have been brought to New Zealand in the mid-1800s. This suggests that the breeds that form the observed monophyletic clade all diverged around the same time between the mid-1800s and early 1900s. As described in Figure 9, the admixture analysis supported this, showing a close resemblance in the overall ancestral component pattern between the Huntaway and the Patagonian Sheepdog, Australian Shepherd, Australian Cattle dog and the Australian Kelpie across all K values tested.

Furthermore, Ancestry 1 (Figure 9) was shared across multiple breeds including the Heading Dog, Huntaway, and Border Collie at percentages above 80%, in Kelpies and Patagonian Sheepdogs at percentages above 60-70%, in Australian Shepherds and Australian Cattle dogs between 50-60% and in Cardigan Welsh Corgis at 20.8%. This indicates the presence of a shared ancestral genetic background. A similar trend was observed in the study by Barrios et al. (2022), where a common genomic signature was reported across the Patagonian Sheepdog, Kelpie and Border Collie populations at high proportions, and in the Bearded Collies and Australian Shepherds at lower proportions and infers a common ancestor between these breeds. The common genetic signature detected through ADMIXTURE likely stems from a historical population of herding/droving dogs that predated modern breed differentiation. Following this differentiation, each of these breeds has been selected for differing roles, and further differentiation occurred between these breeds during the Victorian breed explosion when breeders began selecting heavily for morphological characteristics over behavioural traits. Despite this, many of these formally recognised UK rural herding breeds are still used as working dogs, and the common genetic signature could be a factor in the retention of herding behaviours in these breeds.

5.4 Genetic Relationship Between New Zealand Heading Dogs and Border Collies

Both phylogenies generated (Figures 3 and 4) show the New Zealand Heading Dog branching directly from the Border Collie, indicating a particularly close genetic relationship between these breeds. Furthermore, the haplotype sharing between the Heading Dog and Border Collie (Figures 6 and 8) was much larger in comparison to all other breeds (median = 103.44Mbp), and bar plots generated from admixture analysis (Figure 9) also reveal a strong genetic connection between these breeds as they have high percentages of ancestral component 1 across all K values. Similar genetic relationships can be seen in the unrooted neighbour joining tree of the UK rural clade (Figure 4) and the admixture analysis between other closely related dog breeds, such as the Pembroke Corgi (PEMB) and Cardigan Corgi (CARD), and the Collie (COLL) and Shetland Sheepdog (SSHP). The close phylogenetic proximity of these breeds reflects not only recent shared ancestry but also their continued phenotypic and behavioural similarity.

By way of example, the Pembroke and Cardigan Welsh Corgis are recognised as distinct modern breeds, but these breeds can both be traced back to now-extinct Welsh herding dogs (American Kennel Club, 2019). Historically, both breeds were used as drovers, a role which involves moving livestock over large distances rather than close proximity herding. The divergence between the two breeds was largely shaped by geographical separation, with Pembroke Corgis originating in Pembrokeshire (southern Wales) while the Cardigan Corgis arose in Cardiganshire (north Wales) (Loades, 2020). Phenotypically, both breeds have similar low-set, elongated bodies adapted for cattle work, but Pembroke Corgis have naturally short tails, while Cardigan Corgis have full bushy tails and are slightly larger and more robust (American Kennel Club, 2019).

Similarly, the Collie and Shetland Sheepdog share a recent common ancestor in Scottish herding dogs (Loades, 2020). The most notable difference between the breeds is their size, with the Shetland Sheepdog being smaller than the Collie. Other than size, these breeds show strong morphological resemblance, with both having long rough coats, similar colour patterns and similar working behaviour (American Kennel Club, 2022). The Collie also has two coat variations, rough and smooth coat. Rough coat Collies were originally bred for managing flocks in the open fields and pastures in Scotland, Ireland and England. In contrast, smooth coat Collies, are preferred in rougher terrains and used more often for droving sheep and cattle over large distances. Smooth coat collies also have a double coat that offers insulation and regulates body temperatures (American Kennel Club, 2025). The Shetland Sheepdog was developed later, and was used to work Shetland sheep, a small breed of sheep hailing from the Shetland Islands. Smaller dogs were favoured because the sheep were smaller, and also because of the harsh environmental conditions of the Shetland Islands (American Kennel Club, 2022).

The direct branching of the Border Collie and the New Zealand Heading Dog reflects a relationship comparable to that observed in the closely related breed pairs discussed above. As discussed earlier, anecdotal evidence suggests that Border Collies were routinely bred with early herding dogs in New Zealand to produce working dogs that retained the behavioural and some physical characteristics of the Border Collie but were better suited to New Zealand's farming environments (Rennie, 1984). Furthermore, it is still common today for Heading Dogs to be crossed back to Border Collies. New Zealand Heading Dogs are generally similar in size to Border Collies, but differ notably in coat type, typically exhibiting a shorter dense coat compared to the longer, rough coat characteristic of Border Collies. This difference in coat characteristic is comparable to the distinction between the rough and smooth collies, where smooth-coated Collies were historically preferred for work in harsher or more rugged environments due to the practicality of a coat that is easier to maintain, and provides thermal regulation. A similar selection pressure appears to have influenced the development of the Heading Dog, favouring shorter coats better suited to New Zealand's working environment. Despite the minor phenotypic differences, both Border Collies and Heading Dogs continue to fill similar functional roles. Behaviourally, both breeds demonstrate the characteristic 'staring eye' behaviour, highlighting the retention of key Border Collie traits in the Heading Dog population despite the selective adaptation to New Zealand's farming environment.

5.5 Huntaways and Other Dog Breeds

Anecdotal evidence suggests that other breeds, such as Mastiff, Retriever or Pointer type breeds, may have contributed historically to the breed development of the Huntaway (Frank Film, 2019; Hughes, 2013). However, based on the average haplotype sharing patterns within the Huntaway cohort analysed (Figure 5), there is no evidence to support substantial genetic contributions from breeds outside of the UK rural clade. Some breed boxplots, in particular the Samoyed, Curly Coat Retriever and Nova Scotia Duck Tolling Retriever, do have upper whiskers that extend above the significance line and slightly greater haplotype sharing medians in comparison to other breeds in the graph. This could be indicative of genomic signatures of historical introgression that happened many generations ago from these breeds, but the signatures of these genomic introgressions now appear at such low frequencies or short lengths that they are no longer easily detectible. Haplotype sharing analysis is biased toward detecting more recent crossbreeding events, therefore short IBD haplotypes that represent historical gene flow often fall below the detection threshold, making them very difficult to identify (Browning & Browning, 2012).

It is common for functional haplotypes that have been inherited from other breeds to be retained through selection, while other haplotypes that have not been selected for are gradually diluted over generations. In cases where a crossbreed is backcrossed (where a crossbred offspring is crossed back to the breed of interest) for specific traits, the genetic contribution of the crossbreed can eventually be reduced to very small regions of the genome

(Hospital, 2005). Such breeding techniques are often employed by animal and plant breeders to introduce desired traits from one breed into the genetic background of another. Examples of such breeding programs include Livestock Improvement Corporations 'slick cattle' breeding program, where Senepol cattle were crossed with Holstein Friesians to introduce a single base pair variant in the *PRLR* (prolactin receptor) gene. This genetic variant results in a short, sleek hair coat. These crosses were then back crossed to Holstein Friesian cattle until offspring are essentially Holstein Friesians with a short, sleek coat and have better heat tolerance (Donkersloot et al., 2021; Littlejohn et al., 2014). A similar approach has also been used in the Wiltshire-Romney program at Massey University, where Wiltshire sheep are crossed with Romneys and then backcrossed to retain the Wiltshire self-shedding trait with an otherwise Romney genetic background (Duffin, 2022). The proportion of the genome that is retained from the undesirable breeds in these back crossed offspring becomes very difficult to detect through methods such as haplotype sharing analysis. Some individual Huntaways did display evidence of significant haplotype sharing with breeds outside the UK rural clade, but this is likely reflective of more recent crossing decisions by individual farmers. While the overall population trend did not suggest widespread genetic influence from breeds outside of the UK rural clade, it is still likely that breeds outside of the UK rural clade have contributed to the Huntaway, but contributions from these breeds are now very difficult to identify using the methodologies applied in this study.

5.6 Population Structure of the Huntaway and Heading Dog

Both the Huntaway and Heading Dog populations form breed specific clusters, which indicates that enough genetic variation exists between these populations for them to be considered separate breeds. Despite the findings from haplotype sharing (Figures 6 and 8) and ADMIXTURE (Figure 9), where a very close relationship between the Heading Dog and Border Collie is observed, the PCA (Figures 10, 11 and 12) reveals sufficient genetic differentiation to distinguish the two as genetically distinct breeds. This differentiation suggests that although the Heading Dog has been developed using a substantial amount of Border Collie genetics, enough genetic divergence has occurred through selection for the two breed populations to form separate breed-specific clusters. The Huntaway, in comparison to the Heading Dog, clustered closer to other members of the UK rural clade, such as the Kelpie, Patagonian Sheepdog and Collie.

The phylogeny and PCA results reveal that the Huntaway and Heading Dog also share a close genetic affinity. This is unsurprising, given that both breeds were developed alongside each other and from a common ancestral population. But what is particularly noteworthy is the striking contrast in their morphology and working styles despite this genetic similarity. The Huntaway is larger and more robust, relying on a deep, loud bark to drive livestock, whereas the Heading Dog is smaller and more agile, and uses body positioning and a staring eye to control livestock. These differences highlight how artificial selection for specific traits, such as vocal herding versus silent stalking, can rapidly lead to substantial divergence in phenotype in genetically similar breeds. This observation also underscores the influence of selection in shaping breed characteristics over relatively short timescales.

The results reported here also raise questions about breed classification. Currently, the New Zealand Huntaway is recognised as a formal breed by the New Zealand Kennel Club, reflecting its national importance and established role in New Zealand's farming systems. In contrast, the Heading Dog is not formally recognised as a distinct breed. Genetic analyses conducted in this study indicate that the Heading Dog and Border Collie are closely related breeds, but despite this there is enough genetic variation that the breed branches separately in the cladogram and forms a breed-specific cluster in the PCA. These findings, alongside the importance of their role in New Zealand's farming systems warrant that the Heading Dog should be considered a genetically and functionally distinct breed.

Although the Huntaway is formally recognised in New Zealand, it is not currently recognised by major international kennel clubs and lacks an official international breed standard. This is largely because New Zealand Huntaways have been developed and maintained as a strictly functional breed, with selection based on their ability to perform specific working tasks rather than conforming to a standardised set of physical characteristics. International breed recognition would require the establishment of a formal breed standard and participation in conformation events, such as dog shows. While dog shows do include working dog classes, there is scepticism about the inclusion of the Huntaway (Fallon, 2023) that stems from examples of other working breeds, such as the Border Collie, where the development of show lines has led to divergence from traditional working lines (Border Collie Society of America, 2014; Jeong et al., 2025). In these cases, show dogs are strongly selected for aesthetic traits, sometimes at the expense of their original behavioural and functional characteristics. The working traits of the Huntaway, which include their barking, stamina and ability to push livestock, are not traits prioritised in the show ring. As such, there is concern that any selection for conformation of Huntaways, biased towards their aesthetic traits, could dilute the breeds iconic working behaviours. Instead, preserving the Huntaways functional integrity remains a priority for those who rely on these dogs in practical working environments.

5.7 Runs of Homozygosity and Baseline Inbreeding Estimation

Inbreeding in dog populations is heavily influenced by breeding decisions and management practices (Leroy et al., 2006). To date, the level of inbreeding in the New Zealand Huntaway and Heading Dog populations has not been reported. In this study, baseline average inbreeding levels for both breeds were calculated using two methods based on the proportion of the genome in runs of homozygosity (F_{ROH}) and the excess of homozygosity (F_{HOM}) (Figure 16). The results suggest that the New Zealand Huntaway population has lower inbreeding levels than the Heading Dog population. Furthermore, the inbreeding levels in the New Zealand Huntaway population align with other breeds with low inbreeding scores such as the Jack Russel Terrier and Shar-Pei, whereas the Heading Dog population inbreeding score aligns with breeds such as the Border Collie and German Shepherd.

Runs of homozygosity were identified on all autosomes in both breeds, and the ROH statistics calculated for both breeds align with other studies (Boccardo et al., 2020; Perfilyeva et al., 2023). The majority (>50%) of ROH identified in both breeds were between 0-2Mb, which reflects distant inbreeding that occurred 12.5 to 50 generations ago, while long ROH greater than 16Mb which indicate recent inbreeding within five generations made up 0.48% and 0.43% of the total ROH in Huntaways and Heading Dogs, respectively (Table 2). Some ROH identified were very long (the longest ROH in Huntaways and Heading Dogs were 56.727Mb and 53.536Mb, respectively).

F_{ROH} is a measure of inbreeding based on ROH regions of the genome, whereas F_{HOM} is a measure based on the difference between observed and expected numbers of homozygous genotypes (Nishio et al., 2023). The mean F_{ROH} values calculated were 0.094 for the Huntaway and 0.119 for the Heading Dog. The greater F_{ROH} score observed for the Heading Dog may be partly explained by the close genetic relationship to the Border Collie, which are known to have higher inbreeding levels as indicated in a study by Mastrangelo et al (2018) that observed a F_{ROH} score of 0.15 in the Border Collie across 123,666 SNPs. Though, it is important to note that the F_{ROH} scores calculated in other studies were generated with a different number of SNPs to this study and therefore warrants careful interpretation. The Heading Dog population in New Zealand is also smaller in comparison to Huntaways, with roughly 20,000 Heading Dogs being registered to the New Zealand Kennel club in 2023 (Figure NZ, 2023a). The New Zealand Huntaway is the second most popular dog breed in the country, with around 30,000 Huntaways registered to the New Zealand Kennel Club in 2023 (Figure NZ, 2023b). The statistics obtained from Figure NZ (2023a) and Figure NZ (2023b) likely underrepresent the true number of Huntaways and Heading Dogs in New Zealand, as many farm dogs are not registered. Despite the higher average F_{ROH} value observed in the Heading Dog population, the value is within the inbreeding range typical of other herding breeds (Sweetalana et al., 2024).

The F_{HOM} scores calculated for the Huntaway exhibited a mean of 0.0001 and Heading Dogs a mean of 0.002, which is around 20 times greater than the Huntaway score, but both scores suggest low levels of inbreeding in both populations. These scores are lower than those reported by Barrios et al. (2022), where the Patagonian Sheepdog population had an average F_{HOM} of 0.03, and Dreger et al. (2016) for the Fonni's dog that has an average F_{HOM} of 0.13. Again, the F_{HOM} scores generated in other studies have been calculated with different numbers of SNPs, and therefore the comparison to the scores obtained in this study should be interpreted carefully. It is also important to note that inbreeding levels vary on an individual scale. Both populations had some individuals showing elevated inbreeding levels, with maximum observed F_{ROH} values of 0.268 for the Huntaway and 0.274 for the Heading Dog. Similarly, the maximum F_{HOM} values were 0.224 in Huntaways and 0.239 in Heading Dogs. It is also notable that despite the historical population bottleneck that occurred when working shepherd dogs were first introduced to New Zealand, the inbreeding level within the current populations is within a similar range or lower in comparison to other modern-day breeds. These results suggest that while both working farm dogs and other international breeds have strong recent histories of selection, the intense morphological selection that defines the latter group may have been a more powerful driver of inbreeding.

5.8 Detection of Selection Signatures

An analysis examining 'runs of homozygosity islands' (ROHi) (Gorssen et al., 2021) was performed and identified a total of 11 ROHi present in New Zealand Huntaways and 13 ROHi in New Zealand Heading Dogs (Table 3). Within these ROHi, a total of 176 characterised genes were identified in Huntaways, and 215 characterised genes identified in Heading Dogs (Table 4). Table 4 provides a framework to potentially unearth genes that might be linked to structural or behavioural traits in the New Zealand Huntaway and Heading Dog populations. Thirteen genes found in ROHi in Huntaways and/or Heading Dogs have previously been studied in relation to disease, structural or behavioural traits in other dog breeds and will be discussed further. Importantly, while some of these genes are linked to disease, the putative positive selection of genes linked to disease does not imply that disease-associated variants are being selected for or against in these populations. It is more likely that the putative positive selection for these genes is reflective of selection for an unknown structural or behavioural trait in New Zealand working dogs.

5.9 New Zealand Huntaway Genes Within Identified ROHi

A gene identified within an ROHi on chromosome 8 in the Huntaway population and has been previously linked to a study in dogs is *SLC24A4*.

5.9.1 SLC24A4

A study by Pedersen et al. (2017) identified a large run of homozygosity containing the *SLC24A4* gene. This gene encodes sodium/potassium/calcium exchanger 4 protein which is involved in potassium dependent sodium/calcium exchange and transport and is important for enamel formation. Specific to Samoyed dogs, an autosomal recessive mutation in the *SLC24A4* causes enamel hypoplasia (Pedersen et al., 2017). The Samoyed hails from the Asian Spitz clade and was originally bred in Northern Russia and Siberia as reindeer herding/driving dogs and sled dogs. Pedersen et al. also note that a UK couple, Sir Ernest and Clara Kilburn-Scot were among the first to import Samoyeds to England and were known to have spent time in Australia between 1907 and 1910, during which they became involved in the handling of sled dogs returning from Antarctic expeditions. Some of these dogs were sent to sled dog breeding kennels in the South Island, although many were quarantined under restrictive conditions and exhibited in zoos. Further anecdotal evidence supports the early presence of Samoyeds in New Zealand. An article in the New Zealand Kennel Gazette (2006) cites Carsten Borchgrevink's book First on the Antarctic Continent, which describes the offloading of sled dogs on Native Island, near Stewart Island. The article also suggests that a breeding pair of Samoyeds was gifted between 1901 and 1902 to former prime minister of New Zealand Sir Jospeh Ward, who became the first known breeder of Samoyeds in New Zealand (Auckram, 2006).

No significant haplotype sharing was found between the Huntaway and the Samoyed, but as mentioned earlier the upper whisker of the boxplot does touch the significance threshold, and it has a slightly higher median haplotype sharing in comparison to other breeds. Considering the presence of Samoyeds in the early 1900s and the shared genomic region under selection between these breeds, this could be evidence that Samoyeds may have contributed to the genetic ancestry of the modern Huntaway, and that the ancestry has been diluted over time and is no longer detected through methods such as IBD haplotype sharing analysis. Alternatively, the selection of this gene in both breeds could suggest that they are being selected for a shared trait. As mentioned above, Samoyeds were historically used to herd and drive reindeer in the Arctic. Rather than controlling herds through close range commands, they were mostly used to guide and accompany reindeer over large grazing areas to keep herds together (Hill's, 2020). Huntaways are employed to push livestock from a distance and rely less on being in close proximity to livestock. This could potentially suggest that the Samoyed and Huntaway are under selection for similar livestock herding traits, however the exact functional relevance of this gene is unclear in the Huntaway population, and this gene has not been previously linked to any behavioural traits.

5.10 Genes in ROHi for both the New Zealand Huntaway and Heading Dog

Both the Huntaway and Heading Dog exhibit ROHi on chromosome 22. Selective sweeps and patterns of directional selection on chromosome 22 have been reported in several dog breeds (Akkad et al., 2015; Perfilyeva et al., 2023; Vaysse et al., 2011). Vaysse et al. (2011) identified a selective sweep on chromosome 22 that was shared by eight breeds analysed (Beagle, Border Terrier, English Bulldog, Gordon Setter, Irish Wolfhound, Newfoundland, Rottweiler and Weimaraner). Considering the prevalence of this sweep across multiple breeds, it is suggested to be a conserved region of the genome in dogs (Vaysse et al., 2011). This region, in particular, contained the genes *FNDC3A*, which encodes a protein involved in various cellular processes including fertilisation, spermatid and Sertoli cell development (Obholz et al., 2006) and *CYSTLR2*, which encodes the cysteinyl leukotriene receptor 2, a protein that acts as a receptor for cysteinyl leukotrienes which play a role in inflammation (Ghatak et al., 2023). Akkad et al. (2015) later identified an overlapping region in a comparison of pointing and herding dog genomes, but this region also included the gene *SETDB2* which encodes a histone lysine methyltransferase, which modifies histones (Roqueta-Rivera et al., 2016). The methylation of histones results in the chromatin becoming more compact, and this can influence gene expression (Torrano et al., 2019). Akkad et al. (2015) speculated that the region identified by Vaysse et al. may be linked to hunting traits in dogs, and the larger selection region that includes *SETDB2* may represent a signature specifically linked to pointing behaviour. This was supported by SNPs within *CYSTLR2* and *SETDB2* that were fixed in six out of seven pointing breeds studied and were not fixed in herding breeds (Akkad et al., 2015). *SETDB2* is also associated with left-right asymmetry. Left-right asymmetry is a critical process during early embryonic development that alters the bilateral symmetry of the embryo, playing a key role in establishing a body plan (Hamada et al., 2002). Interestingly, motor laterality, which is associated to left-right asymmetry, is linked to success in guide dog training (Tomkins et al., 2012) and therefore variations in SETDB2 may influence learning ability (Akkad et al., 2015).

Given the presence of regions of selection in chromosome 22 in the Huntaway and Heading Dog that include the same genes, it seems likely that this region does play a role in hunting related traits. While specific variants may differentiate pointing from other hunting behaviours, the consistent selection of this region across a variety of breeds (Vaysse et al., 2011) suggests a relevance to hunting behaviours. Furthermore, many herding behaviours such as following and staring at sheep are modified hunting traits where the lethal step, catching and killing prey, has been bred out. Cognitive traits such as motor laterality could also be important for herding dogs and could explain why the *SETDB2* gene is under selection in both New Zealand working breeds.

Perfilyeva et al. (2023) also observed patterns of putative positive selection on chromosome 22 in the Tazy breed, which is considered a sighthound. Eight genes identified in the selection sweep (*KPNA3*, *RCBTB1*, *SETDB2*, *CAB39L*, *CDADC1*, *MLNR*, *FNDC3A* and *CYSTLR2*), clustered within a single protein-protein interaction (PPI) network, a representation of the relationships between proteins, that showed high PPI enrichment. These eight genes are present in the ROHi on chromosome 22 for both the New Zealand Huntaway and Heading Dog. While the hunting behaviours of pointing breeds and sighthounds differ, this further suggests that the genomic region on chromosome 22 under selection in multiple dog breeds likely contributes to broader hunting traits, and more specific variation within this region may differentiate hunting behaviours. A gene identified by Perfilyeva et al. but was not reported in previous studies is *CAB39L*. This gene is involved in the regulation of AMP-activated protein kinase pathway and helps maintain energy homeostasis during physical activity (Burton et al., 2023), indicating that *CAB39L* may contribute to running endurance and speed in the Tazy.

Owner surveys collected in the “Right Dog for the Job” project confirm that full time working Huntaways and Heading Dogs are typically used in the field for more than three hours a day. Endurance and speed are traits essential in livestock herding dogs, as these directly relate to their ability to perform sustained, physically demanding work over long periods of time. The presence of *CAB39L* in the ROHi on chromosome 22 for both the Huntaway and the Heading Dog may reflect a common selection pressure for traits that prolong physical exertion. The detection of this gene in the Huntaway and Heading Dog populations may provide further evidence that *CAB39L* contributes to running endurance and speed, but further research is needed to evaluate its functional significance.

5.11 New Zealand Heading Dog Genes Within Identified ROHi

The genes identified within the ROHi's identified specifically in the Heading Dog population and have been previously linked to studies in dogs are *KIRREL3*, *ATP7B*, *RYR3* and *EPHB1*.

5.11.1 KIRREL3

The Heading Dog cohort in this study have a ROHi on chromosome 5, and the only gene in this ROHi is *KIRREL3*. A study by Olsson et al. (2015) that focused on identifying genomic regions associated with low immunoglobulin A levels, found that *KIRREL3* is one gene significantly associated to immunoglobulin A levels. *KIRREL3* encodes a transmembrane protein that plays a role in both in synapse formation and as a regulator of haematopoiesis. It is thought that variations in this gene can lead to an alteration in early B-cell development (Ueno et al., 2003). *KIRREL3* knock-out mice display a loss of male-male aggression implicating changes in behaviour are also connected to this gene, and it is speculated that the selective pressure of this gene in dogs for behavioural traits may have also enriched it for immune defects (Olsson et al., 2015). The putative positive selection for this gene in Heading Dogs could therefore reflect the selection for behavioural traits such as assertiveness or controlled aggression towards livestock. However, the precise role of this gene and whether it is linked to any herding or hunting behaviours in dogs remains unclear. Further research is needed to investigate whether there is any specific canine behaviours influenced by the *KIRREL3* gene.

5.11.2 ATP7B

A gene present in the ROHi on chromosome 22 that was only under selection in the Heading Dog population studied is *ATP7B*. This gene encodes a copper transporting ATP-ase protein that plays a critical role in maintaining copper homeostasis. Genetic variants in the *ATP7B* gene have previously been associated with the development of copper toxicosis in Labrador Retrievers and Doberman Pinschers (Pindar & Ramirez, 2019; Wu et al., 2019). Copper toxicosis (also known as Wilson disease) is caused by defects in the *ATP7B* protein which prevents the transport of cytosolic copper into the trans-Golgi network, limiting the amount of copper available for incorporation into the protein ceruloplasmin for excretion in the bile. Furthermore, the accumulation of excess copper in the system eventually leads to liver failure (Pindar & Ramirez, 2019). While genetic variants in this gene have been linked to copper toxicosis in other dog breeds, the putative positive selection of this gene in the Heading Dog population does not suggest that Heading Dogs are affected by this disease. Furthermore, the *ATP7B* mutation in dogs that is thought to be responsible for Wilson disease (ATP7B:c.4358G>A) is common in a large number of dog breeds that are not reported to have copper toxicosis. Therefore, while this mutation is a risk factor it may not be the sole cause of the disease (Haywood et al., 2023). There are currently no other studies that link the *ATP7B* gene to any specific traits in dogs. Therefore, the potential role of this gene in influencing desirable working dog traits that are under putative positive selection in the Heading Dog population is still uncertain but warrants further investigation in future studies.

5.11.3 EPHB1

The *EPHB1* gene was identified in an ROHi on chromosome 23 in the New Zealand Heading Dogs studied. This gene encodes ephrin type-B receptor 1, which plays key roles in axon growth, spatial memory and neuropathic pain perception. It also regulates the preoptic regulatory factor-2, a molecule involved in various neural pathways (Cibert-Goton et al., 2013; Huang et al., 2018). A study by Jeong et al. (2025) investigated the genetic underlying selection for herding traits in Border Collies by comparing working lines, bred for functional performance, with conformation (show) lines, which are bred for the adherence to breed standards. In this study it was hypothesised that the different selection pressures between working and conformation lines of Border Collies could reveal genetic signatures that may be specific to herding traits (Jeong et al., 2025).

A total of 24 genes in the Border Collie population were identified under statistically significant selection, with the strongest signatures of selection being associated with the ephrin type-B receptor (*EPHB1*) gene. To assess the selection associated with the *EPHB1* gene, they constructed haplotypes within the *EPHB1* region in eight dog breed populations, including both herding and non-herding breeds. They identified six distinct haplotypes spanning ~57kb in *EPHB1*, and haplotype 1 was nearly fixed in the working Border Collie population (84%). In contrast, haplotype 1 was only present in 22% of conformation Border Collies, and nearly absent in all non-herding breeds tested (Jeong et al., 2025).

These findings suggest that the genomic region that includes the *EPHB1* has undergone selection in working Border Collies and could be associated with herding behaviours. Given that the EPHB1 gene is involved in social cognition and communication, it is likely that associated traits are important for dogs that are required both to learn and respond to a wide range of commands and to use their body positioning effectively to control livestock (Jeong et al., 2025). The close behavioural traits and shared ancestry between the New Zealand Heading Dog and the Border Collie and the presence of a selective sweep including the *EPHB1* gene in the Heading Dog population could further suggest that *EPHB1* is in fact linked to herding behaviour. Investigating the frequency of haplotype 1 from the study by Jeong et al. in the New Zealand Heading dog population could provide further insight into whether this specific haplotype represents a selection signature for herding, rather than being specific to working Border Collies. This would strengthen the hypothesis that variations in the *EPHB1* gene play a role in the behavioural phenotype expressed among herding breeds.

5.11.4 RYR3

A gene located on an ROHi on chromosome 30 that was only identified in the Heading Dog is the *RYR3* gene. This gene encodes a ryanodine receptor involved in skeletal muscle contraction and aids the release of calcium from the sarcoplasmic reticulum. This is a crucial step in the initiation of muscle fibre contraction (Lynch et al., 2013). A study by Yu et al. (2022) identified *RYR3* as a candidate gene under putative positive selection in German Longhaired Pointers and other hunting breed dogs. This study suggests that the putative positive selection of *RYR3* may be associated with the development of muscular phenotypes and physical endurance in hunting dog populations (Yu et al., 2022). Given the physical endurance and stamina required by Heading Dogs to manage livestock across steep and rugged terrain, genes involved in muscle function are likely important to their working roles. The presence of the *RYR3* gene in an ROHi in the Heading Dog population may therefore reflect selection for traits that support muscular performance, such as fast recovery from exertion and responsiveness during bursts of movement which would be advantageous in livestock herding dogs. This finding provides further evidence that the *RYR3* gene may contribute to muscle related functions in hunting and herding dogs.

5.2 Study Limitations

The publicly available datasets used in this study included genotype data from 1,256 individuals representing 127 breeds and three wild canid species. This dataset captures a wide range of dog breeds and therefore a large spectrum of genotypic variation between breeds. Samples were taken from many different geographic locations, including North America, Europe, Africa and Asia (Parker et al., 2017). While the dataset clearly captures global diversity across breeds, limitations arise in relation to sampling design. Although it is evident that breeds were sampled from different countries, the specific criteria used to select individual dogs within each breed is unknown. For the majority of breeds, around 10 individuals were genotyped, but this did vary with some breeds having 20 samples (e.g. Alaskan Malamute and Siberian Husky) while others have only one or two samples (Dingo, Kelpie and Australian Cattle dog). A key limitation is that the dataset does not provide details about whether the sampled individuals within each breed were related, nor whether they originated from the same geographic area. This is important to note, as sampling dogs from a narrow geographic or familial background may lead to an overestimation of within-breed genetic similarity. This could also overestimate the differences between breeds, if the only real difference between breeds was the familial lines of individuals. In contrast, if individuals were sampled from diverse geographic regions within a breed, the resulting genetic profile would better reflect the breed's overall diversity. Without this context, it is difficult to assess how representative the genotyped individuals are of their breed as a whole, which may influence the interpretation of patterns of genetic diversity and population structure.

The full dataset used in the study by Parker et al. (2017) comprised of 1,346 individuals representing 161 dog breeds. However, in the publicly available files, several breeds included in the original study are missing. The reasons for the exclusion of some breeds were not stated but may relate to permission constraints. One notable breed that was excluded from the publicly available files is the Bearded Collie, which is particularly relevant given its potential historical and genetic significance to this study, as this breed belongs to the UK rural clade and has been shown to share significant haplotypes with the Patagonian Sheepdog (Barrios et al., 2022). In addition, anecdotal evidence suggests that the Bearded Collie was one of the founding breeds of both the Huntaway and Heading Dog (Dalton, 2009; Hughes, 2013). The breed continues to be used as a working dog on New Zealand farms, supporting its ongoing genetic connection to these populations. Including the Bearded Collie in the haplotype sharing analyses may have provided more insights into the genetic foundations of New Zealand's working breeds. It might have also helped clarify the origins of breed-specific traits, such as the Huntaways distinctive deep bark.

For the New Zealand Huntaway and Heading Dog data, breed identification for each individual was provided by the owner. However, some of the dogs genotyped were not genetically consistent with being 100% Huntaway or 100% Heading Dog, which led to some outliers in the data visualisation. While this may have influenced some of the observed trends in breed contributions to the Huntaway and Heading Dog populations, extreme outliers, which are very clearly recent crossbreeds, were removed from haplotype sharing analyses. Additionally, since this study focused on population level patterns rather than individual variation, the influence of any remaining undetected crossbreeds was reduced.

This study was conducted using high density SNP chip data, which was appropriate for the research aims. However, whole-genome sequencing (WGS) data could offer deeper insight, particularly into potential contributions from breeds outside the UK rural clade, into the genetic makeup of the Huntaway. While SNP data is useful for detecting broad patterns of genetic structure and ancestry, its lower marker density may limit the resolution of analyses such as inbreeding estimation.

The neighbour joining cladogram (Figure 3) was generated using 100 subsets of the original dataset. Subsets of the original dataset were generated since PLINK v1.9 is unable to handle duplicate variant IDs and automatically removes them when constructing distance matrices. For future analyses, cladograms should ideally be built using bootstrapped (resampled with replacement) datasets, as this approach provides a more robust framework for cladogram construction.

The inbreeding scores calculated here provide an initial indication of inbreeding levels in both the Huntaway and Heading Dog populations. These scores were derived from 151,688 SNP markers and are based on data from 246 Heading Dogs and 211 Huntaways. It Is difficult to compare the F_{ROH} and F_{HOM} scores generated in this study to those calculated in other studies, as other scores have been calculated for different numbers of SNPs. It is also important to note that SNP chip based inbreeding estimates may underestimate true inbreeding levels, particularly because they are less effective at capturing the short runs of homozygosity that are indicative of more ancient inbreeding, due to lacking sufficient marker density. Moreover, comprehensive assessments of inbreeding typically rely on larger sample sizes and higher resolution genomic data. Therefore, the inbreeding scores presented in this study cannot be extrapolated on a population level and are best interpreted as preliminary estimates.

Several genes identified within ROHi regions in both the Huntaway and Heading Dog populations have previously been associated with specific behaviours or diseases in other dog breeds. While their presence within ROHi suggests that these genes are under putative positive selection in these populations, it is important to note that the traits linked to these genes are associated with specific variants in these genes, rather than the genes' presence alone. This study did not investigate whether particular functional variants known to cause disease or influence behaviour in other breeds are also present in the Huntaway or Heading Dog populations. As such, although these genes have been implicated in certain traits in other breeds, their effects may not be expressed in the same way in the New Zealand working dog populations. Further work is required to determine whether the variants influencing these traits exist and are relevant in Huntaways and Heading Dogs.

5.3 Future Research

This study confirms that the New Zealand Huntaway and Heading Dog can be traced back to ancestral UK working dogs and form a part of the UK rural subclade of the domestic dog phylogeny. There was no evidence of significant contributions from breeds outside of the UK rural clade in the Huntaways. Future studies using whole-genome sequencing (WGS) data, which has much higher SNP density over the entire genome, could re-run the IBD haplotype sharing and potentially identify more genetic contributions from breeds outside of the UK rural clade and provide further insight into this. It could also be worth running haplotype sharing per chromosome, as this may capture smaller ROH that are being drowned out by larger ones in other areas of the genome. Additionally, increasing the sample size in haplotype sharing analyses would enhance the statistical power of the results and strengthen conclusions drawn from the study.

The inbreeding scores presented in this study represent preliminary estimates. These scores were calculated with SNP genotyping data (151,688 SNP markers) and from only a moderate number of individuals (246 Heading Dogs and 211 Huntaways). While the scores calculated offer initial insights into the levels of inbreeding within New Zealand's working dogs, further analysis using larger sample sizes is needed, and the calculations performed here should be redone as more SNP genotyping data becomes available. The "Right Dog for the Job" project aims to genotype approximately 2,400 individual dogs, which will significantly increase the sample size, and therefore the statistical power of the inbreeding scores obtained. Furthermore, if more genomic datasets from other dog breeds can be included as part of this analysis, it will make the results for the Huntaway and Heading Dog more comparable to other breeds.

Since the calculation of F_{ROH} and F_{HOM} in the Huntaway and Heading Dog does not require the externally available data used in this study, the inbreeding should also be recalculated using the full set of Affymetrix CanineHD Array markers, which includes approximately 712,000 SNPs. In comparison to the Illumina array, this high-density array has a larger coverage of the canine genome and will potentially identify more shared haplotypes that were missed in this study. Future studies should also compare scores against those calculated with WGS data since it gives a much higher resolution view of the genome and will potentially pick up runs of homozygosity that were missed in the SNP based estimates.

The ROHi analysis conducted in this study identified genomic regions of the canine genome under putative positive selection in the New Zealand Huntaway and Heading Dog populations. These regions under putative positive selection harbour genes that may be involved in traits that are critical to the roles performed by Huntaways and Heading Dogs, including physical endurance, cognitive ability and livestock herding behaviours, as highlighted above. A full functional analysis of all of the genes listed was outside the scope of the project, but the list provides a foundation for future research. In particular, this dataset could be used to identify genes that have not been previously studied in dogs but could be linked to behavioural traits in New Zealand working dogs, and also for variant discovery to identify genetic variations associated with key behavioural or disease related traits, such as the deep bark observed in the Huntaway. Some of the genes identified in this study have been previously associated with functional traits in other breeds, but it is important to note that the presence of a gene under selection does not confirm the selection of the same functional traits identified in other studies discussed in the New Zealand working dog populations. Further research is needed to confirm whether the specific variants reported in other breeds linked to such traits are present in these dogs, and whether they have the same functional effects.

Chapter 6: Conclusion

The New Zealand Huntaway and Heading Dog are vital to the ongoing success of New Zealand's sheep and beef farming industries. Without their contributions, many New Zealand sheep and beef operations would be neither physically nor economically feasible. These two breeds have been selectively bred to carry out distinct, yet complementary roles in the mustering and management of livestock. The Huntaway, known for its strong voice and ability to drive livestock over long distances, and the Heading Dog, valued for its precision and control, are both essential to the effective functioning of pastoral farming throughout the country. Until now, there has been limited scientific research addressing the origins, development and genetic composition of these two breeds. This study represents the first large-scale genomic investigation into the origins and population structure of New Zealand's two unique working farm dogs breeds. Through the integration of high-density SNP genotype data from the two New Zealand breeds with publicly available datasets representing 127 formally recognised dog breeds and three wild canid species, this study has formalised the genetic heritage of the Huntaway and Heading Dog within the broader domestic canine phylogeny.

The analyses presented demonstrate that both breeds can be traced back to ancestral UK rural working dog populations, and both breeds occupy a position within the UK rural subclade on the domestic canine phylogeny. This finding is consistent with historical patterns of European settlement and dog importation during the agricultural expansion of New Zealand in the 19th century. The geographic isolation and targeted breeding practices within New Zealand have since shaped the Huntaway and Heading Dog into genetically differentiated populations, well suited to the specific demands of New Zealand's farming environment. While anecdotal evidence suggests that the New Zealand Huntaway was developed using breeds outside of the UK rural clade, this study cannot confirm what other breeds contributed to its genetic makeup outside of the UK rural clade. Analyses also highlight that the New Zealand Heading Dog is closely related to the Border Collie, but the population has undergone specific selection pressures in New Zealand and has led to the development of a genetically distinct and functionally specialised working dog. The New Zealand Heading Dog should be a formal breed in its own right.

The calculation of F_{ROH} and F_{HOM} revealed that neither the Huntaway or Heading Dog population express significantly high inbreeding levels, and the scores obtained are similar to that of other recognised dog breeds, but inbreeding scores calculated in this study should only be interpreted as preliminary and warrant further investigation. Runs of homozygosity islands identified in both populations provides initial insights into genomic regions that are under putative positive selection in the Huntaway and Heading Dog. While some of the genes identified in ROHi in this study have been previously linked to disease or behavioural traits in other dog breeds, further research is needed to determine the exact functional effects of these genes in the Huntaway and Heading Dog populations. The full list of genes within the ROHi identified could be used for future studies looking to identify genes linked to behavioural traits in New Zealand working dogs, and also for variant discovery to identify genetic variation associated with key behavioural or disease traits.

Overall, this study provides genomic evidence demonstrating the distinctiveness of the Huntaway and Heading Dog and their adaptation to the unique demands of New Zealand's pastoral farming systems. These results not only contribute to the current understanding of breed development in geographically isolated environments but also highlight the future opportunity of integrating genomics into working dog management and breeding practices. The genetic resources generated through this study will serve as a valuable reference for future research in the "Right Dog for the Job" project, aimed at preserving the health, diversity and functionality of the New Zealand Huntaway and Heading Dog.

References

Abraham, G., Qiu, Y., & Inouye, M. (2017). FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics (Oxford, England)*, *33*(17), 2776–2778. https://doi.org/10.1093/bioinformatics/btx299

Abraham G, & Inouye M. (2014). Fast Principal Component Analysis of Large-Scale Genome-Wide Data. *PLoS ONE*, *9*(4), e93766. https://doi.org/10.1371/journal.pone.0093766

Ács, V., Bokor, Á., & Nagy, I. (2019). Population structure analysis of the border collie dog breed in hungary. *Animals*, *9*(5), 250. https://doi.org/10.3390/ani9050250

Akkad, D. A., Gerding, W. M., Gasser, R. B., & Epplen, J. T. (2015). Homozygosity mapping and sequencing identify two genes that might contribute to pointing behavior in hunting dogs. *Canine Genetics and Epidemiology*, *2*, 5. https://doi.org/10.1186/s40575-015-0018-5

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

Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). *Admixture 1.3 Software Manual*. https://dalexander.github.io/admixture/admixture-manual.pdf

Ali, M. B., Evans, J. M., Parker, H. G., Kim, J., Pearce-Kelling, S., Whitaker, T., Plassais, J., Khan, Q. M., & Ostrander, E. A. (2020). Genetic analysis of the modern Australian labradoodle dog breed reveals an excess of the poodle genome. *PLoS Genetics*, *16*(9), e1008956. https://doi.org/https://doi.org/10.1371/journal.pgen.1008956

American Kennel Club. (2019, October 15). *Meet Two Similar Yet Different Breeds: The Cardigan Welsh Corgi and the Pembroke Welsh Corgi*. https://www.akc.org/expert-advice/lifestyle/cardigan-welsh-corgi-pembroke-welsh-corgi/#:~:text=Physically%2C%20the%20Cardigan%20and%20Pembroke,besides%20regular%20bathing%20and%20brushing.

American Kennel Club. (2022, January 11). *Collie vs. Shetland Sheepdog: How to Tell the Difference*. https://www.akc.org/expert-advice/dog-breeds/collie-versus-shetland-sheepdog-tell-difference/

American Kennel Club. (2025). *Collie*. https://www.akc.org/dog-breeds/collie/

Ancestry. (2024). *Know Your Pet DNA*. https://petdna.ancestry.com/about

Auckram, V. (2006). *New Zealand Kennel Gazette*. https://www.thesamoyedclub.org.nz/pdfs/samoyed-supplement-new-zealand-kennel-club-gazette-may-2006-volume-46-no-4.pdf

Austin, F. (2018, October 19). *Border Collie History: From Old Hemp to New Beginnings*. https://colliepoint.com/border-collie-history/comment-page-1/#comments

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., & Suchard, M. A. (2012). BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Systematic Biology*, *61*(1), 170–173. https://doi.org/10.1093/sysbio/syr100

Bannasch, D., Famula, T., Donner, J., Anderson, H., Honkanen, L., Batcher, K., Safra, N., Thomasy, S., & Rebhun, R. (2021). The effect of inbreeding, body size and morphology on health in dog breeds. *Canine Medicine and Genetics*, *8*(1), 12. https://doi.org/10.1186/s40575-021-00111-4

Barnes, M., Freudenberg, J., Thompson, S., Aronow, B., & Pavlidis, P. (2005). Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic Acids Research*, *33*(18), 5914–5923. https://doi.org/10.1093/nar/gki890

Barrios, N., Fuenzalida, A., Gómez, M., Heuser, C., Muñoz, R., Ostrander, E. A., Parker, H. G., & González-Lagos, C. (2019). The Patagonian sheepdog: Historical perspective on a herding dog in Chile. *Diversity*, *11*(12), 245. https://doi.org/10.3390/D11120245

Barrios, N., González-Lagos, C., Dreger, D. L., Parker, H. G., Nourdin-Galindo, G., Hogan, A. N., Gómez, M. A., & Ostrander, E. A. (2022). Patagonian sheepdog: Genomic analyses trace the footprints of extinct UK herding dogs to South America. *PLoS Genetics*, *18*(4), e1010160. https://doi.org/10.1371/journal.pgen.1010160

Beef + Lamb New Zealand. (2023). *Farm Facts 47th Edition*. https://beeflambnz.com/knowledge-hub/PDF/farm-facts-2023.pdf

Bianchini, G., & Sánchez-Baracaldo, P. (2024). TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees. *Ecology and Evolution*, *14*(2), e10873. https://doi.org/10.1002/ece3.10873

Boccardo, A., Marelli, S. P., Pravettoni, D., Bagnato, A., Busca, G. A., & Strillacci, M. G. (2020). The German shorthair pointer dog breed (Canis lupus familiaris): Genomic inbreeding and variability. *Animals*, *10*(3), 498. https://doi.org/10.3390/ani10030498

Border Collie Society of America. (2014). *The Border Collie*. https://www.doc-developpement-durable.org/file/Elevages/Chiens/Chiens%20de%20conduite/presentationBorderCollie.pdf

Broome, C. J., & Walsh, V. P. (2003). Gastric dilatation-volvulus in dogs. *New Zealand Veterinary Journal*, *51*(6), 275–283. https://doi.org/10.1080/00480169.2003.36381

Browning, S. R., & Browning, B. L. (2010). High-Resolution Detection of Identity by Descent in Unrelated Individuals. *American Journal of Human Genetics*, *86*(4), 526–539. https://doi.org/10.1016/j.ajhg.2010.02.021

Browning, S. R., & Browning, B. L. (2012). Identity by descent between distant relatives: Detection and applications. *Annual Review of Genetics*, *46*, 617–633. https://doi.org/10.1146/annurev-genet-110711-155534

Burton, M. A., Antoun, E., Garratt, E. S., Westbury, L., Baczynska, A., Dennison, E. M., Harvey, N. C., Cooper, C., Patel, H. P., Godfrey, K. M., & Lillycrop, K. A. (2023). Adiposity is associated with widespread transcriptional changes and downregulation of longevity pathways in aged skeletal muscle. *Journal of Cachexia, Sarcopenia and Muscle*, *14*(4), 1762–1774. https://doi.org/10.1002/jcsm.13255

Cave, N. J., Bridges, J. P., Cogger, N., & Farman, R. S. (2009). A survey of diseases of working farm dogs in new zealand. *New Zealand Veterinary Journal*, *57*(6), 305–312. https://doi.org/10.1080/00480169.2009.60926

Chew, T., Willet, C. E., Haase, B., & Wade, C. M. (2019). Genomic characterization of external morphology traits in kelpies does not support common ancestry with the Australian dingo. *Genes*, *10*(5), 337. https://doi.org/10.3390/genes10050337

Cibert-Goton, V., Yuan, G., Battaglia, A., Fredriksson, S., Henkemeyer, M., Sears, T., & Gavazzi, I. (2013). Involvement of EphB1 Receptors Signalling in Models of Inflammatory and Neuropathic Pain. *PLoS ONE*, *8*(1), e53673. https://doi.org/10.1371/journal.pone.0053673

Crespi, J. A., Barrientos, L. S., & Giovambattista, G. (2018). von Willebrand disease type 1 in Doberman Pinscher dogs: genotyping and prevalence of the mutation in the Buenos Aires region, Argentina. *Journal of Veterinary Diagnostic Investigation*, *30*(2), 310–314. https://doi.org/10.1177/1040638717750429

Dalton, C. (2009). Farm Dogs. In *Te Ara*. https://teara.govt.nz/en/farm-dogs

Deelen, P., Bonder, M. J., Van Der Velde, K. J., Westra, H. J., Winder, E., Hendriksen, D., Franke, L., & Swertz, M. A. (2014). Genotype harmonizer: Automatic strand alignment and format conversion for genotype data integration. *BMC Research Notes*, *7*(1), 901. https://doi.org/10.1186/1756-0500-7-901

Dogs New Zealand. (2013). *Breed Standard - NZ Huntaway*. https://www.dogsnz.org.nz/breeds/info/nz-huntaway/535/breed-standard

Donkersloot, E., Worth, G., Yeates, A., McNaughton, L., Spelman, R., & Davis, S. (2021). *The Benefit of a Slick Hair Coat for Heat Tolerance in New Zealand Dairy Cattle*.

Donner, J., Kaukonen, M., Anderson, H., Möller, F., Kyöstilä, K., Sankari, S., Hytönen, M., Giger, U., & Lohi, H. (2016). Genetic panel screening of nearly 100 mutations reveals new insights into the breed distribution of risk variants for canine hereditary disorders. *PLoS ONE*, *11*(8), e0161005. https://doi.org/10.1371/journal.pone.0161005

Dreger, D. L., Davis, B. W., Cocco, R., Sechi, S., Di Cerbo, A., Parker, H. G., Polli, M., Marelli, S. P., Crepaldi, P., & Ostrander, E. A. (2016). Commonalities in development of pure breeds and population isolates revealed in the genome of the Sardinian Fonni's dog. *Genetics*, *204*(2), 737–755. https://doi.org/10.1534/genetics.116.192427

Duffin, Paul. (2022, March 24). *Grading up to a Wiltshire flock of shedding sheep*. https://www.massey.ac.nz/about/news/grading-up-to-a-wiltshire-flock-of-shedding-sheep/

Fallon, V. (2023, August 13). "They're absolute athletes": All hail the huntaway, NZ's only indigenous dog. *Waikato Times*. https://www.waikatotimes.co.nz/nz-news/350053589/theyre-absolute-athletes-all-hail-huntaway-nzs-only-indigenous-dog

FigTree. (2018, November 25). *FigTree*. http://tree.bio.ed.ac.uk/software/figtree/

Figure NZ. (2023a). *Registered purebred Heading dogs in New Zealand*. https://figure.nz/chart/7sHd7bTcp2D1LFhI-Xyh2NxcKIA8dVmOw

Figure NZ. (2023b). *Registered purebred Huntaways in New Zealand*. https://figure.nz/chart/7sHd7bTcp2D1LFhI-3cqjUr3z8T3p4FZx

Focus Genetics. (2025). *Focus Genetics*. https://www.focusgenetics.com/

Frank Film. (2019, November 8). *The Huntaway* [Video recording]. https://frankfilm.co.nz/frank-changing-south/changing-south-2019/huntaway/

Freedman, A. H., & Wayne, R. K. (2017). Deciphering the Origin of Dogs: From Fossils to Genomes. *Annual Review of Animal Biosciences*, *5*, 281–307. https://doi.org/10.1146/annurev-animal-022114-110937

Gajaweera, C., Kang, J. M., Lee, D. H., Lee, S. H., Kim, Y. K., Wijayananda, H. I., Kim, J. J., Ha, J. H., Choi, B. H., & Lee, S. H. (2019). Genetic diversity and population structure of the Sapsaree, a native Korean dog breed. *BMC Genetics*, *20*, 66. https://doi.org/10.1186/s12863-019-0757-5

Ghatak, S., Satapathy, S. R., & Sjölander, A. (2023). DNA Methylation and Gene Expression of the Cysteinyl Leukotriene Receptors as a Prognostic and Metastatic Factor for Colorectal Cancer Patients. *International Journal of Molecular Sciences*, *24*(4), 3409. https://doi.org/10.3390/ijms24043409

GitHub. (2024). *gtc2vcf*. https://github.com/freeseek/gtc2vcf

Gorssen, W., Meyermans, R., Janssens, S., & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, *53*(1), 2. https://doi.org/10.1186/s12711-020-00599-7

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*, 100. https://doi.org/10.1038/s43586-022-00184-w

Guo, J., Johnson, G. S., Brown, H. A., Provencher, M. L., da Costa, R. C., Mhlanga-Mutangadura, T., Taylor, J. F., Schnabel, R. D., O'Brien, D. P., & Katz, M. L. (2014). A CLN8 nonsense mutation in the whole genome sequence of a mixed breed dog with neuronal ceroid lipofuscinosis and Australian shepherd ancestry. *Molecular Genetics and Metabolism*, *21*(4), 302–309. https://doi.org/10.1016/j.ymgme.2014.05.014

Guo, J., Johnson, G. S., Cook, J., Harris, O. K., Mhlanga-Mutangadura, T., Schnabel, R. D., Jensen, C. A., & Katz, M. L. (2019). Neuronal ceroid lipofuscinosis in a German Shorthaired Pointer associated with a previously reported CLN8 nonsense variant. *Molecular Genetics and Metabolism Reports*, *21*, 100521. https://doi.org/10.1016/j.ymgmr.2019.100521

Halvoník, A., Moravčíková, N., Chalupková, M., & Kasarda, R. (2024). Commonly used genomic estimators of individual inbreeding in livestock. *Czech Journal of Animal Science*, *69*(7), 269–279. https://doi.org/10.17221/91/2024-CJAS

Hamada, H., Meno, C., Watanabe, D., & Saijoh, Y. (2002). Establishment of vertebrate left-right asymmetry. *Nature Reviews Genetics*, *3*(2), 103–113. https://doi.org/10.1038/nrg732

Hassan, M. I., Saxena, A., & Ahmad, F. (2012). Structure and function of von Willebrand factor: The protein that is deficient and/or abnormal in inherited von Willebrand disease. *Blood Coagulation and Fibrinolysis*, *23*(1), 11–22. https://doi.org/10.1097/MBC.0b013e32834cb35d

Haywood, S., Swinburne, J., Schofield, E., Constantino-Casas, F., & Watson, P. (2023). Copper toxicosis in Bedlington terriers is associated with multiple independent genetic variants. *Veterinary Record*, *193*(4), e2832. https://doi.org/10.1002/vetr.2832

Herdenberg, C., Henriksson, R., Hedman, H., & Rondahl, V. (2025). Lrig3-deficient mice exhibit strain-specific alterations in liver fat accumulation, intestinal morphology, and middle ear inflammation. *Gene*, *960*, 149539. https://doi.org/10.1016/j.gene.2025.149539

Hill's. (2020). *Samoyed Dog Breed Information and Personality Traits*. https://www.hillspet.co.nz/dog-care/dog-breeds/samoyed#:~:text=The%20Samoyed%20is%20truly%20an,in%20much%20of%20the%20world.

Hoggart, C. J., Shriver, M. D., Kittles, R. A., Clayton, D. G., & Mckeigue, P. M. (2004). Design and Analysis of Admixture Mapping Studies. *American Journal of Human Genetics*, *74*, 965–978. https://doi.org/10.1086/420855

Hospital, F. (2005). Selection in backcross programmes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1459), 1503–1511. https://doi.org/10.1098/rstb.2005.1670

Howard, N. P., Troggio, M., Durel, C. E., Muranty, H., Denancé, C., Bianco, L., Tillman, J., & van de Weg, E. (2021). Integration of Infinium and Axiom SNP array data in the outcrossing species Malus × domestica and causes for seemingly incompatible calls. *BMC Genomics*, *22*, 246. https://doi.org/10.1186/s12864-021-07565-7

Huang, G. H., Guo, L., Zhu, L., Liu, X. D., Sun, Z. L., Li, H. J., Xu, N. J., & Feng, D. F. (2018). Neuronal GAP-Porf-2 transduces EphB1 signaling to brake axon growth. *Cellular and Molecular Life Sciences*, *75*(22), 4207–4222. https://doi.org/10.1007/s00018-018-2858-0

Hughes, P. (2013). *History and Development of the Farm Working Dog in New Zealand*. https://www.sciquest.org.nz/search/results-2/downloadfulltext/166049

Illumina. (2025a). *Canine HD (12 Sample)*. https://emea.illumina.com/library-prep-array-kit-selector/kits-and-arrays/canine-hd_12-sample.html

Illumina. (2025b). *CanineHD Product Files*. https://support.illumina.com/downloads/caninehd_product_files.html

Isaksen, K. E., Linney, L., Williamson, H., Cave, N. J., Beausoleil, N. J., Norman, E. J., & Cogger, N. (2020). TeamMate: A longitudinal study of New Zealand working farm dogs. I. Methods, population characteristics and health on enrolment. *BMC Veterinary Research*, *16*, 59. https://doi.org/10.1186/s12917-020-2273-2

Jeong, H., Ostrander, E. A., & Kim, J. (2025). Genomic evidence for behavioral adaptation of herding dogs. *Science Advances*, *11*(18), 4591. https://doi.org/10.1126/sciadv.adp4591

Jerram, A. (2013). *Demographics and management of dogs used to herd livestock in New Zealand* [Massey University]. https://mro.massey.ac.nz/server/api/core/bitstreams/3e2e5bc3-5c53-4539-8cff-afd6a1cab616/content

Katz, M. L., Rustad, E., Robinson, G. O., Whiting, R. E. H., Student, J. T., Coates, J. R., & Narfstrom, K. (2017). Canine neuronal ceroid lipofuscinoses: Promising models for preclinical testing of therapeutic interventions. *Neurobiology of Disease*, *108*, 277–287. https://doi.org/10.1016/j.nbd.2017.08.017

Kristiansen, M., Kozyraki, R., Jacobsen, C., Nexø, E., Verroust, P. J., & Moestrup, S. K. (1999). Molecular Dissection of the Intrinsic Factor-Vitamin B 12 Receptor, Cubilin, Discloses Regions Important for Membrane Association and Ligand Binding. *Journal of Biological Chemistry*, *274*(29), 20540–20544. https://doi.org/10.1074/jbc.274.29.20540

Lee, T. H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics*, *15*, 162. https://doi.org/10.1186/1471-2164-15-162

Leroy, G., Rognon, X., Varlet, A., Joffrin, C., & Verrier, E. (2006). Genetic variability in French dog breeds assessed by pedigree data. *Journal of Animal Breeding and Genetics*, *123*(1), 1–9. https://doi.org/10.1111/j.1439-0388.2006.00565.x

Littlejohn, M. D., Henty, K. M., Tiplady, K., Johnson, T., Harland, C., Lopdell, T., Sherlock, R. G., Li, W., Lukefahr, S. D., Shanks, B. C., Garrick, D. J., Snell, R. G., Spelman, R. J., & Davis, S. R. (2014). Functionally reciprocal mutations of the prolactin signalling pathway define hairy and slick cattle. *Nature Communications*, *5*, 5861. https://doi.org/10.1038/ncomms6861

Liu, N., Zhang, K., & Zhao, H. (2008). Haplotype-Association Analysis. *Advances in Genetics*, *60*, 335–405. https://doi.org/10.1016/S0065-2660(07)00414-2

Livestock Improvement Corporation. (2025). *Livestock Improvement Corporation*. https://www.lic.co.nz/

Loades, M. (2020). *Dogs: Working Origins and Traditional Tasks*. https://books.google.co.nz/books?hl=en&lr=&id=dy8NEAAAQBAJ&oi=fnd&pg=PT8&dq=corgi+origins&ots=_aBDKK9_dZ&sig=PGoS9y2RChYsv2Ivhkpkhs6hBvg&redir_esc=y#v=onepage&q=corgi%20origins&f=false

Lonka, L., Kyttälä, A., Ranta, S., Jalanko, A., & Lehesjoki, A.-E. (2000). The neuronal ceroid lipofuscinosis CLN8 membrane protein is a resident of the endoplasmic reticulum. *Human Molecular Genetics*, *9*(11), 1691–1697. https://doi.org/10.1093/hmg/9.11.1691

Lynch, A. I., Irvin, M. R., Boerwinkle, E., Davis, B. R., Vaughan, L. K., Ford, C. E., Aissani, B., Eckfeldt, J. H., Arnett, D. K., & Shrestha, S. (2013). RYR3 gene polymorphisms and cardiovascular disease outcomes in the context of antihypertensive treatment. *Pharmacogenomics Journal*, *13*(4), 330–334. https://doi.org/10.1038/tpj.2012.22

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., & Derks, E. M. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *International Journal of Methods in Psychiatric Research*, *27*(2), e1608. https://doi.org/10.1002/mpr.1608

Massey University. (2025, February 17). *Genetic research aims to improve health and performance of New Zealand working dogs*. https://www.massey.ac.nz/about/news/genetic-research-aims-to-improve-health-and-performance-of-new-zealand-working-dogs/

Mata, F., & Mata, A. (2023). Investigating the relationship between inbreeding and life expectancy in dogs: mongrels live longer than pure breeds. *PeerJ Life & Environment*, *11*, e15718. https://doi.org/10.7717/peerj.15718

McConnell, I., Marker, L., & Rooney, N. (2022). Preliminary investigation into personality and effectiveness of livestock guarding dogs in Namibia. *Journal of Veterinary Behavior*, *48*, 11–19. https://doi.org/10.1016/j.jveb.2021.10.006

Mech, L. D., & Janssens, L. A. A. (2022). An assessment of current wolf Canis lupus domestication hypotheses based on wolf ecology and behaviour. *Mammal Review*, *52*(2), 304–314. https://doi.org/10.1111/mam.12273

Menor-Campos, D. J. (2024). Ethical Concerns about Fashionable Dog Breeding. *Animals*, *14*(5), 756. https://doi.org/10.3390/ani14050756

Meyer, A., Kossow, K., Gal, A., Steglich, C., Mühlhausen, C., Ullrich, K., Braulke, T., & Muschol, N. (2008). The mutation p.Ser298Pro in the sulphamidase gene (SGSH) is associated with a slowly progressive clinical phenotype in mucopolysaccharidosis type IIIA (Sanfilippo A syndrome). *Human Mutation*, *29*(5), 770. https://doi.org/10.1002/humu.20738

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (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), 94. https://doi.org/10.1186/s12864-020-6463-x

Ministry for Primary Industries. (2025). *Ministry for Primary Industries*. https://www.mpi.govt.nz/

Mortlock, S. A., Booth, R., Mazrier, H., Khatkar, M. S., & Williamson, P. (2016a). Visualization of genome diversity in german shepherd dogs. *Bioinformatics and Biology Insights*, *9*, 37–42. https://doi.org/10.4137/BBI.S30524

Mortlock, S. A., Khatkar, M. S., & Williamson, P. (2016b). Comparative analysis of genome diversity in bullmastiff dogs. *PLoS ONE*, *11*(1), e0147941. https://doi.org/10.1371/journal.pone.0147941

National Center for Biotechnology Information. (2025). *Genome Data Viewer*. https://www.ncbi.nlm.nih.gov/gdv/

NCBI. (2011). *Genome Assembly CanFam3.1*. https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002285.3/

New Zealand Sheep Dog Trial Association. (2023). *NZ Sheep Dog Trial Association*. https://www.nzsheepdogs.co.nz/

Nicholas, Frank. T. Imke. S. I. Hub. (2025). *Online Mendelian Inheritance in Animals (OMIA)*. https://omia.org

Nicolazzi, E., Marras, G., & Stella, A. (2016). SNPConvert: SNP Array Standardization and Integration in Livestock Species. *Microarrays*, *5*(2), 17. https://doi.org/10.3390/microarrays5020017

Nishio, M., Inoue, K., Ogawa, S., Ichinoseki, K., Arakawa, A., Fukuzawa, Y., Okamura, T., Kobayashi, E., Taniguchi, M., Oe, M., & Ishii, K. (2023). Comparing pedigree and genomic inbreeding coefficients, and inbreeding depression of reproductive traits in Japanese Black cattle. *BMC Genomics*, *24*, 376. https://doi.org/10.1186/s12864-023-09480-5

Obholz, K. L., Akopyan, A., Waymire, K. G., & MacGregor, G. R. (2006). FNDC3A is required for adhesion between spermatids and Sertoli cells. *Developmental Biology*, *298*(2), 498–513. https://doi.org/10.1016/j.ydbio.2006.06.054

Olsson, M., Tengvall, K., Frankowiack, M., Kierczak, M., Bergvall, K., Axelsson, E., Tintle, L., Marti, E., Roosje, P., Leeb, T., Hedhammar, Å., Hammarström, L., & Lindblad-Toh, K. (2015). Genome-wide analyses suggest mechanisms involving early B-cell development in canine IgA deficiency. *PLoS ONE*, *10*(9), e0138405. https://doi.org/10.1371/journal.pone.0133844

Ostrander, E. A., Wayne, R. K., Freedman, A. H., & Davis, B. W. (2017). Demographic history, selection and functional diversity of the canine genome. *Nature Reviews Genetics*, *18*(12), 705–720. https://doi.org/10.1038/nrg.2017.67

Owczarek-Lipska, M., Jagannathan, V., Drögemüller, C., Lutz, S., Glanemann, B., Leeb, T., & Kook, P. H. (2013). A Frameshift Mutation in the Cubilin Gene (CUBN) in Border Collies with Imerslund-Gräsbeck Syndrome (Selective Cobalamin Malabsorption). *PLoS ONE*, *8*(4), e61144. https://doi.org/10.1371/journal.pone.0061144

Pāmu. (2025). *Pāmu Farms of New Zealand*. https://www.pamunewzealand.com/

Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., & Ostrander, E. A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, *19*(4), 697–708. https://doi.org/10.1016/j.celrep.2017.03.079

Pedersen, N. C., Shope, B., & Liu, H. (2017). An autosomal recessive mutation in SCL24A4 causing enamel hypoplasia in Samoyed and its relationship to breed-wide genetic diversity. *Canine Genetics and Epidemiology*, *4*, 11. https://doi.org/10.1186/s40575-017-0049-1

Perfilyeva, A., Bespalova, K., Bespalov, S., Begmanova, M., Kuzovleva, Y., Vishnyakova, O., Nazarenko, I., Abylkassymova, G., Perfilyeva, Y., Plakhov, K., Djansugurova, L., & Bekmanov, B. (2023). Homozygosity mapping in the Kazakh national dog breed Tazy. *Scientific Reports*, *13*, 10735. https://doi.org/10.1038/s41598-023-37990-5

Pindar, S., & Ramirez, C. (2019). Predicting copper toxicosis: relationship between the ATP7A and ATP7B gene mutations and hepatic copper quantification in dogs. *Human Genetics*, *138*(5), 541–546. https://doi.org/10.1007/s00439-019-02010-y

PLINK. (2023). *PLINK 1.9 Manual*. https://www.cog-genomics.org/plink/1.9/

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Studio Team. (2020). *The Comprehensive R Archive Network*. https://cran.r-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Rennie, N. (1984). *Working Dogs Breeding, Feeding, Training and Care*.

Roqueta-Rivera, M., Esquejo, R. M., Phelan, P. E., Sandor, K., Daniel, B., Foufelle, F., Ding, J., Li, X., Khorasanizadeh, S., & Osborne, T. F. (2016). SETDB2 Links Glucocorticoid to Lipid Metabolism through Insig2a Regulation. *Cell Metabolism*, *24*(3), 474–484. https://doi.org/10.1016/j.cmet.2016.07.025

Santos, C. R. O., de Simoni Gouveia, J. J., Gouveia, G. V., Bezerra, F. C. M., Nogueira, J. F., & Baraúna, D. (2020). Molecular screening for the mutation associated with canine degenerative myelopathy (SOD1:c.118G > A) in German Shepherd dogs in Brazil. *PLoS ONE*, *15*(11), e0242347. https://doi.org/10.1371/journal.pone.0242347

Saravanan, K. A., Panigrahi, M., Kumar, H., & Bhushan, B. (2022). Advanced software programs for the analysis of genetic diversity in livestock genomics: a mini review. *Biological Rhythm Research*, *53*(3), 358–368. https://doi.org/10.1080/09291016.2019.1642650

Shihabi, M., Glavaš, V., Cubric-Curik, V., Zorc, M., Dovč, P., & Curik, I. (2022). *Genome-wide signals of positive selection identified in Livestock Guardian Dogs*. 3101–3104. https://doi.org/10.3920/978-90-8686-940-4_753

Smith, F., Lopdell, T., Stephen, M., Henry, M., Dittmer, K., Hunt, H., Sneddon, N., Williams, L., Rolfe, J., Garrick, D., & Littlejohn, M. (2025). Survey of Mendelian-effect functional variants in New Zealand Huntaway and Heading dog breeds. *Animal Genetics*, *56*(5), e70042. https://doi.org/doi:10.1111/age.70042

Soh, P. X. Y., Hsu, W. T., Khatkar, M. S., & Williamson, P. (2021). Evaluation of genetic diversity and management of disease in Border Collie dogs. *Scientific Reports*, *11*, 6243. https://doi.org/10.1038/s41598-021-85262-x

Sutter, N. B., Eberle, M. A., Parker, H. G., Pullar, B. J., Kirkness, E. F., Kruglyak, L., & Ostrander, E. A. (2004). Extensive and breed-specific linkage disequilibrium in Canis familiaris. *Genome Research*, *14*(12), 2388–2396. https://doi.org/10.1101/gr.3147604

Sweetalana, Nataneli, S., Huang, S., Mooney, J. A., & Szpiech, Z. A. (2024). Genotypic and phenotypic consequences of domestication in dogs. *BioRxiv*. https://doi.org/10.1101/2024.05.01.592072

The Border Collie Museum. (2013). *The Collie Breeds of Australia and New Zealand*. https://www.bordercolliemuseum.org/BCCousins/AusAndNZ.html

The Border Collie Museum. (2014). *Introduction to Border Collie Cousins*. https://www.bordercolliemuseum.org/BCCousins/BC_Cousins_2.html

The Helical Company Limited. (2025). *The Helical Company Limited*. https://www.helicalco.com/

Thermofisher. (2024). *Axiom Canine HD array*. https://www.thermofisher.com/order/catalog/product/550869

Tomkins, L. M., Thomson, P. C., & McGreevy, P. D. (2012). Associations between motor, sensory and structural lateralisation and guide dog success. *Veterinary Journal*, *192*(3), 359–367. https://doi.org/10.1016/j.tvjl.2011.09.010

Torrano, J., Al Emran, A., Hammerlindl, H., & Schaider, H. (2019). Emerging roles of H3K9me3, SETDB1 and SETDB2 in therapy-induced cellular reprogramming. *Clinical Epigenetics*, *11*(1), 43. https://doi.org/10.1186/s13148-019-0644-y

Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, *3*(25), 731. https://doi.org/10.21105/joss.00731

Tzeng, J., Lu, H., & Li, W. H. (2008). Multidimensional scaling for large genomic data sets. *BMC Bioinformatics*, *9*, 179. https://doi.org/10.1186/1471-2105-9-179

Ueno, H., Sakita-Ishikawa, M., Morikawa, Y., Nakano, T., Kitamura, T., & Saito, M. (2003). A stromal cell-derived membrane protein that supports hematopoietic stem cells. *Nature Immunology*, *4*(5), 457–463. https://doi.org/10.1038/ni916

University of Washington. (2021, January 1). *PHYLIP*. https://phylipweb.github.io/phylip/index.html

Vaysse, A., Ratnakumar, A., Derrien, T., Axelsson, E., Pielberg, G. R., Sigurdsson, S., Fall, T., Seppälä, E. H., Hansen, M. S. T., Lawley, C. T., Karlsson, E. K., Bannasch, D., Vilà, C., Lohi, H., Galibert, F., Fredholm, M., Häggström, J., Hedhammar, Å., André, C., ... Webster, M. T. (2011). Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. *PLoS Genetics*, *7*(10), e1002316. https://doi.org/10.1371/journal.pgen.1002316

Vetlife. (2024). *Vetlife*. https://vetlife.co.nz/

Wag! (2024). *New Zealand Heading Dog* . https://wagwalking.com/breed/new-zealand-heading-dog

Wayne, R. K., & Vonholdt, B. M. (2012). Evolutionary genomics of dog domestication. *Mammalian Genome*, *23*(1–2), 3–18. https://doi.org/10.1007/s00335-011-9386-7

Wiley, E. O. (1979). Cladograms and Phylogenetic Trees. *Systematic Biology*, *28*(1), 88–92. https://doi.org/10.1093/sysbio/28.1.88

Wu, X., Mandigers, P. J. J., Watson, A. L., van den Ingh, T. S. G. A. M., Leegwater, P. A. J., & Fieten, H. (2019). Association of the canine ATP7A and ATP7B with hepatic copper accumulation in Dobermann dogs. *Journal of Veterinary Internal Medicine*, *33*(4), 1646–1652. https://doi.org/10.1111/jvim.15536

Yogalingam, G., Pollard, T., Gliddon, B., Jolly, R. D., & Hopwood, J. J. (2002). Identification of a mutation causing mucopolysaccharidosis type IIIA in New Zealand Huntaway dogs. *Genomics*, *79*(2), 150–153. https://doi.org/10.1006/geno.2002.6699

Yoshida, G. M., Cáceres, P., Marín-Nahuelpi, R., Koop, B. F., & Yáñez, J. M. (2020). Estimates of autozygosity through runs of homozygosity in farmed coho salmon. *Genes*, *11*(5), 490. https://doi.org/10.3390/genes11050490

Yu, Y., Liu, L., Windig, J., Bosse, M., Groenen, M. A. M., & Crooijmans, R. P. M. A. (2022). Unique genetic signature and selection footprints in Dutch population of German Longhaired Pointer dogs. *Animal Genetics*, *53*(6), 829–840. https://doi.org/10.1111/age.13253

Zapata, I., Lilly, M. L., Herron, M. E., Serpell, J. A., & Alvarez, C. E. (2022). Genetic testing of dogs predicts problem behaviors in clinical and nonclinical samples. *BMC Genomics*, *23*(1), 102. https://doi.org/10.1101/2020.08.13.249805

Zeng, R., Coates, J. R., Johnson, G. C., Hansen, L., Awano, T., Kolicheski, A., Ivansson, E., Perloski, M., Lindblad-Toh, K., O'Brien, D. P., Guo, J., Katz, M. L., & Johnson, G. S. (2014). Breed distribution of SOD1 alleles previously associated with canine degenerative myelopathy. *Journal of Veterinary Internal Medicine*, *28*(2), 515–521. https://doi.org/10.1111/jvim.12317

Appendix

Table A.1. Number of individual dogs per breed clade (excluding Huntaways and Heading Dogs). The breed clade colours represent the colours of the breed clades in the cladogram and haplotype sharing analysis.

Breed Clade	**No. Dogs per Clade**
African	10
Asian Spitz	96
Asian Toy	30
Continental Herder	30
Drover	35
European Mastiff	160
Hungarian	7
Mediterranean	115
New World	10
Nordic Spitz	17
Pointer Setter	35
Poodle	41
Retriever	58
Scent Hound	60
Schnauzer	27
Sight Hound	67
Spaniel	43
Terrier	88
Toy Dog	68
UK Rural	223
Wild	10

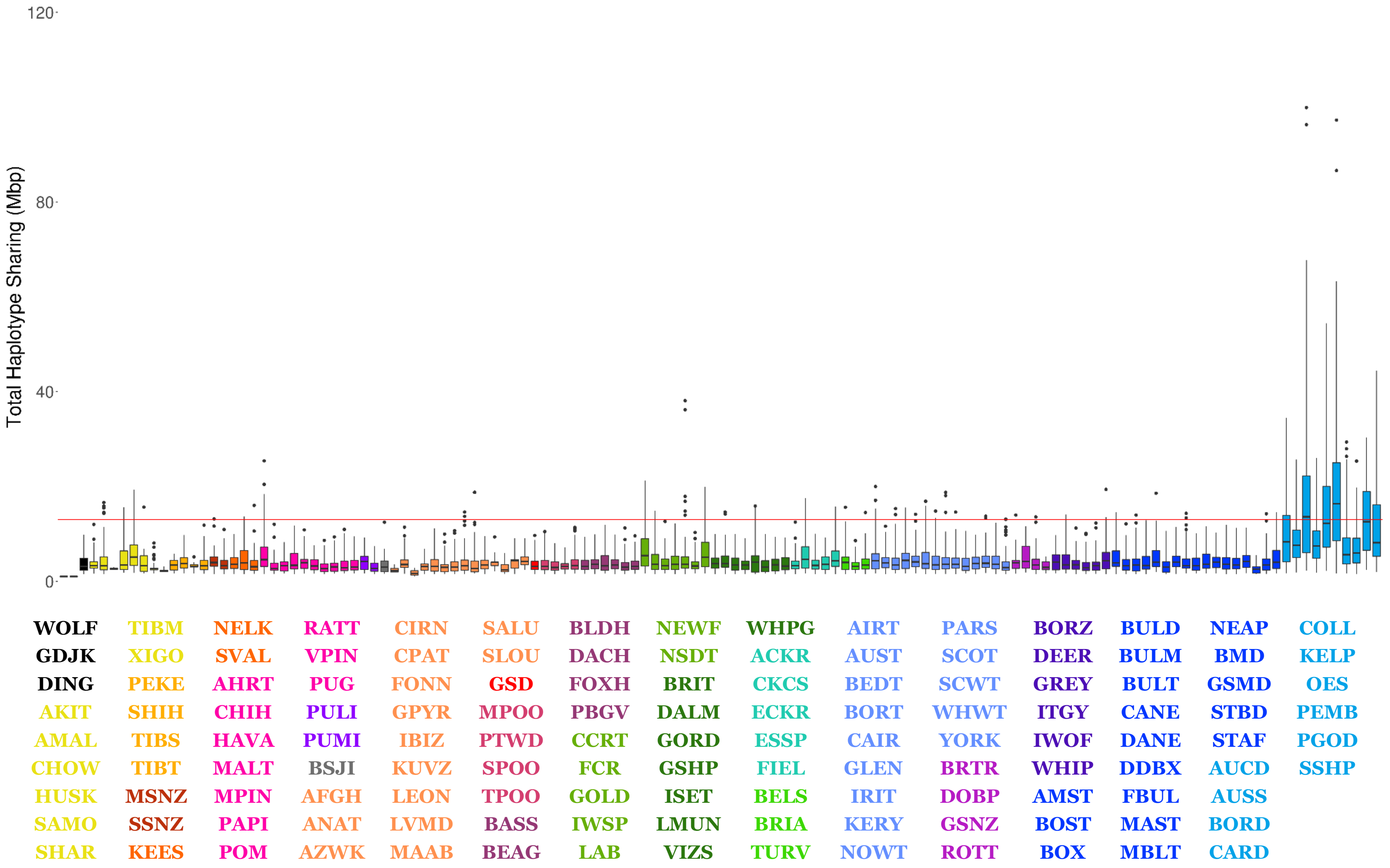

Figure A.1. Identity-by-descent (IBD) Haplotype sharing analysis of 128 previously published breeds and three wild canids against the New Zealand Huntaway with outliers.

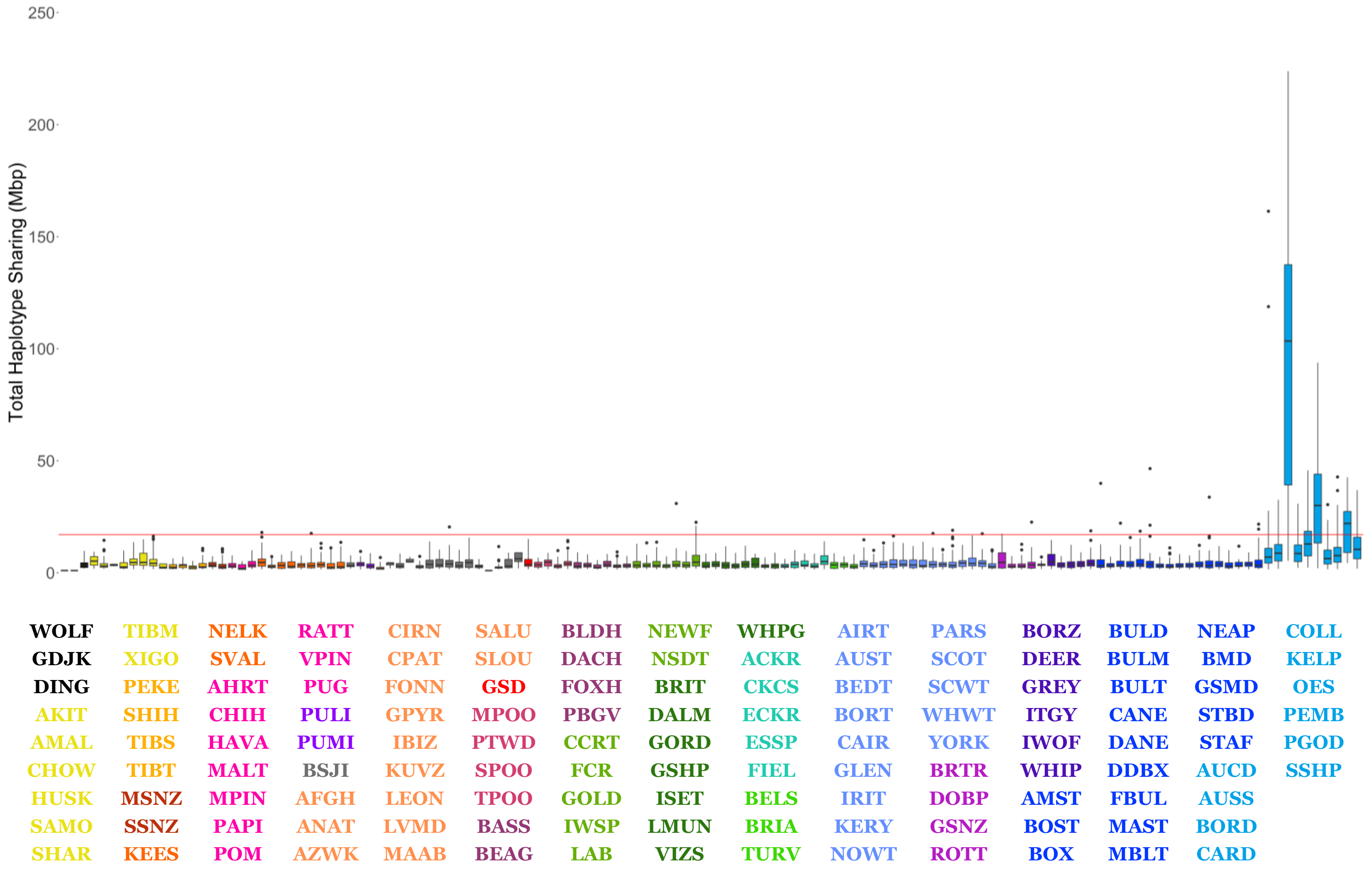

Figure A.2. Identity-by-descent (IBD) haplotype sharing analysis of 128 previously published breeds and three wild canids against the New Zealand Heading Dog with outliers.

Table A.1. Identity-by-descent (IBD) haplotype sharing analysis summary stats for the New Zealand Huntaway against the UK rural breeds.

Breed	Mean	Median	Standard Deviation	Minimum	Maximum
AUCD	9752051.4	8300194.3	6574824.2	1720545.5	34444773.5
AUSS	8261100.9	7606993.5	4227867.5	1873601.333	25661033.2
BORD	16226746.7	13602400.7	13903107.4	2224053	99911169.4
CARD	8123536.0	7433626.8	4317936.6	1815466.5	26037001.1
COLL	15064613.6	12237918.1	10672377.4	2192179	54429048.1
KELP	18601547.4	16381464.0	14421606.1	1675136	97268361.5
OES	7082642.9	5599925.2	4857812.2	1614418	29368839.2
PGOD	12837803.1	12568197.6	6793974.6	2420563.095	30350002.41
SSHP	11273305.2	8128733.0	7993980.3	1952285	44389317.8
PEMB	6895499.7	5996522.3	3881487.5	1562376	25352437.4

Table A.2. Identity-by-descent (IBD) haplotype sharing analysis summary stats for the New Zealand Heading Dog against the UK rural breeds.

Breed	Mean	Median	Standard Deviation	Minimum	Maximum
AUCD	9038058.6	7029974.3	13355872.5	1525035	161358619
AUSS	9386534.7	8737808.3	5116525.3	1795027.667	32538237.2
BORD	93384423.6	103435387.7	53100542.7	5335843.889	223849640
CARD	9531609.1	8621745.2	5421625.2	2186036.4	30836022
COLL	14142620.5	12868902.2	8748465.8	2315318.75	45702190.6
KELP	30553574.6	30053572.0	18363945.9	2029117.5	93836017
OES	7454371.5	6334552.6	4453583.6	1593890.4	30446365.2
PGOD	19123374.8	22023480.9	9606142.5	4330965.57	42623065.2
SSHP	11755559.7	10488339.2	6894634.1	1890554.75	36825367.7
PEMB	8859349.0	7674173.9	5825595.1	1681862	42774850.1

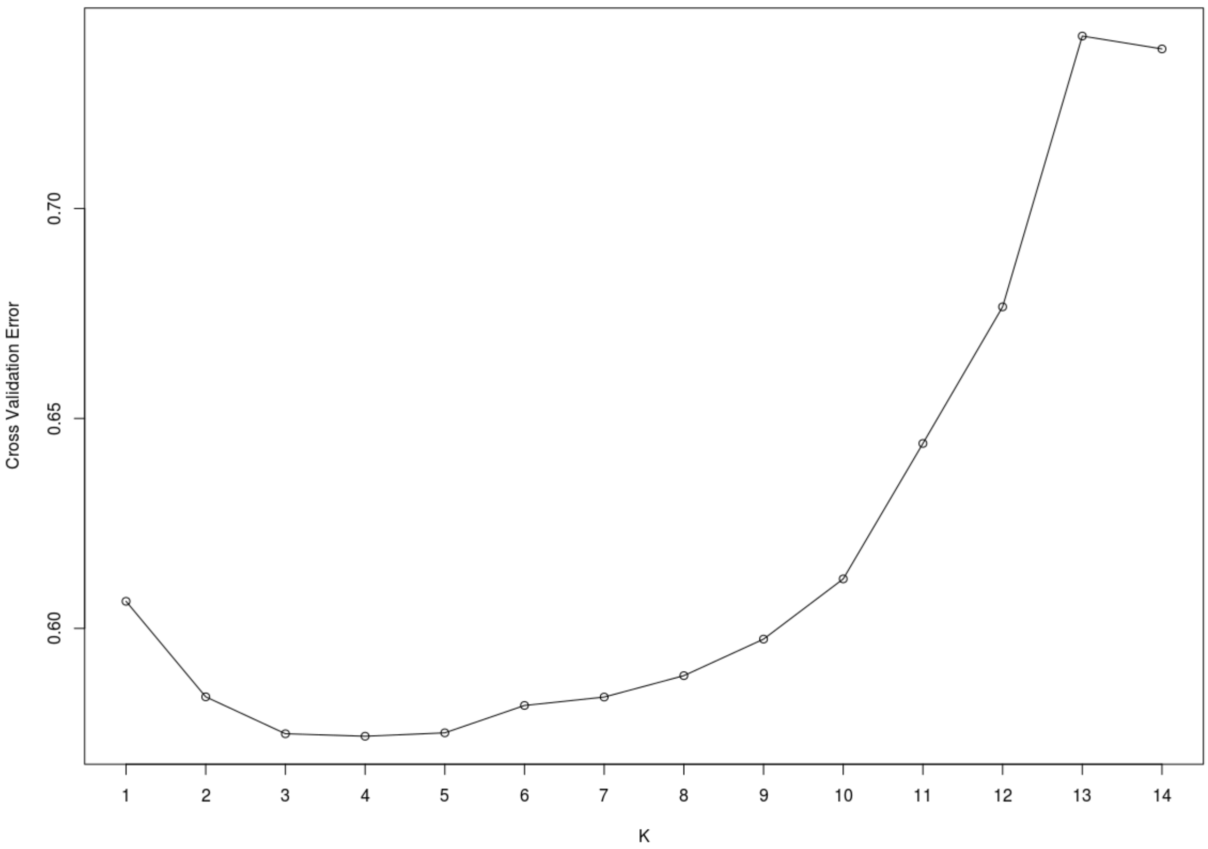

Figure A.3. Cross-validation (CV) plot for ADMIXTURE analysis. The y-axis represents the CV error value, and the x-axis represents the K values.

Table A.3. ADMIXTURE summary statistics for K = 4.

Breed	Ancestral Component	Mean	SD	Min	Max
AUCD	Ancestry 1 (green)	0.581	0.0038	0.579	0.584
AUCD	Ancestry 2 (orange)	0.180	0.0092	0.174	0.187
AUCD	Ancestry 3 (red)	0.206	0.0007	0.206	0.207
AUCD	Ancestry 4 (blue)	0.032	0.0047	0.028	0.035
AUSS	Ancestry 1 (green)	0.535	0.0117	0.518	0.556
AUSS	Ancestry 2 (orange)	0.149	0.0075	0.138	0.162
AUSS	Ancestry 3 (red)	0.186	0.0122	0.173	0.206
AUSS	Ancestry 4 (blue)	0.129	0.0135	0.109	0.147
BORD	Ancestry 1 (green)	0.951	0.0467	0.863	0.994
BORD	Ancestry 2 (orange)	0.009	0.0161	0.000	0.041
BORD	Ancestry 3 (red)	0.011	0.0187	0.000	0.050
BORD	Ancestry 4 (blue)	0.028	0.0179	0.006	0.054
CARD	Ancestry 1 (green)	0.208	0.0168	0.185	0.238

CARD	Ancestry 2 (orange)	0.088	0.0074	0.080	0.099
CARD	Ancestry 3 (red)	0.657	0.0201	0.627	0.687
CARD	Ancestry 4 (blue)	0.047	0.0142	0.030	0.064
COLL	Ancestry 1 (green)	0.000	0.0000	0.000	0.000
COLL	Ancestry 2 (orange)	0.000	0.0000	0.000	0.000
COLL	Ancestry 3 (red)	0.000	0.0000	0.000	0.000
COLL	Ancestry 4 (blue)	1.000	0.0000	1.000	1.000
HEAD	Ancestry 1 (green)	0.998	0.0066	0.979	1.000
HEAD	Ancestry 2 (orange)	0.001	0.0033	0.000	0.011
HEAD	Ancestry 3 (red)	0.000	0.0003	0.000	0.001
HEAD	Ancestry 4 (blue)	0.001	0.0030	0.000	0.009
HUNT	Ancestry 1 (green)	0.808	0.0226	0.776	0.850
HUNT	Ancestry 2 (orange)	0.077	0.0128	0.059	0.095
HUNT	Ancestry 3 (red)	0.100	0.0108	0.082	0.118
HUNT	Ancestry 4 (blue)	0.015	0.0107	0.000	0.032
KELP	Ancestry 1 (green)	0.694	0.0100	0.687	0.701
KELP	Ancestry 2 (orange)	0.106	0.0064	0.102	0.111
KELP	Ancestry 3 (red)	0.146	0.0078	0.140	0.151
KELP	Ancestry 4 (blue)	0.055	0.0042	0.052	0.058
OES	Ancestry 1 (green)	0.000	0.0000	0.000	0.000
OES	Ancestry 2 (orange)	1.000	0.0000	1.000	1.000
OES	Ancestry 3 (red)	0.000	0.0000	0.000	0.000
OES	Ancestry 4 (blue)	0.000	0.0000	0.000	0.000
PEMB	Ancestry 1 (green)	0.000	0.0000	0.000	0.000
PEMB	Ancestry 2 (orange)	0.000	0.0000	0.000	0.000
PEMB	Ancestry 3 (red)	1.000	0.0000	1.000	1.000
PEMB	Ancestry 4 (blue)	0.000	0.0000	0.000	0.000
PGOD	Ancestry 1 (green)	0.646	0.1089	0.566	0.950
PGOD	Ancestry 2 (orange)	0.140	0.0543	0.000	0.207
PGOD	Ancestry 3 (red)	0.148	0.0528	0.002	0.180
PGOD	Ancestry 4 (blue)	0.066	0.0165	0.047	0.094
SSHP	Ancestry 1 (green)	0.009	0.0280	0.000	0.089
SSHP	Ancestry 2 (orange)	0.009	0.0184	0.000	0.048
SSHP	Ancestry 3 (red)	0.004	0.0120	0.000	0.038
SSHP	Ancestry 4 (blue)	0.978	0.0553	0.825	1.000