MASTERS IN EVOLUTIONARY BIOLOGY

School of Biology and Environmental Science
University College Dublin

BIOL40050
Research Project Thesis

A Population Genomic Analysis of the Endangered Galway Sheep Breed

Author: Gillian McHugo

Student Number: 15204236

Supervisor: Professor David MacHugh

Date of Submission: 22/08/2016

Declaration of Authorship

I hereby certify that the submitted work is my own work, was completed while registered as a candidate for the degree stated on the Title Page, and I have not obtained a degree elsewhere on the basis of the research presented in this submitted work

Name: Gillian McHugo

Date: 22/08/2016

Procedures for Submission and Late Submission
Ensure that you have checked the School's procedures for the submission of assessments.
Note: There are penalties for the late submission of assessments. For further information please see the University's *Policy on Late Submission of Coursework*, (http://www.ucd.ie/registry/academicsecretariat/late_sub.pdf)

Plagiarism is a form of academic dishonesty. In any assignment, plagiarism means that you have presented information or ideas belonging to someone else falsely as being your own original thoughts on a subject. The penalties associated with plagiarism reflect the seriousness of the University's commitment to academic integrity.
Ensure that you have read the University's Briefing for Students on Academic Integrity and Plagiarism and the University's Plagiarism Statement, Plagiarism Policy and Procedures, (http://www.ucd.ie/registry/academicsecretariat/info_stud.htm)

RESEARCH THESIS (40 % OF FINAL GRADE)

The text, including figures and tables should not exceed 5000 words.

The thesis should not repeat the literature review. The introduction should be concise and focus on the research questions of your project.

Two printed copies of the thesis must be signed in to the School Office, Room 213 Science Centre (West), and a **third electronic copy** submitted via the safeAssign link in Blackboard.

SUMMARY

The endangered Galway sheep is the only native Irish sheep breed, and an important genetic resource. Population genomic analyses of the Galway breed in comparison with other sheep populations of European origin revealed their recent evolutionary history. This confirmed the use of the Border Leicester breed in the formation of the Galway breed and more recent gene flow with the Suffolk breed. The Galway breed also had unique signatures of artificial selection when compared to other breeds, which highlighted candidate genes that may be involved in meat and wool production. When compared to both commercial and endangered breeds, the Galway breed was intermediate in effective population size (N_e), genomic inbreeding and runs of homozygosity (ROH). This indicates that, although the Galway breed is declining, it is still relatively genetically diverse and careful conservation and management plans may aid its recovery.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 DOMESTICATED SPECIES AND EVOLUTIONARY BIOLOGY

Domesticated animal species are those that are bred in captivity and modified from their wild ancestors to make them more useful to humans (Diamond 2002). This has been done through thousands of years of artificial selection by humans that has impacted many genes (genotypes) and altered a wide range of physical, physiological and behavioural traits (phenotypes) (Diamond 2002; Andersson and Georges 2004). Since the signatures of natural and artificial selection are essentially the same, domesticated species provide fascinating models for practical evolutionary biology (de Simoni Gouveia *et al.* 2014).

Sheep (*Ovis aries*) in particular are excellent study organisms for evolutionary biology (Haenlein 2007). Sheep were domesticated approximately 11 000 years ago and have since been selected for a variety of uses including meat, milk and wool production (Meadows 2014; Leroy *et al.* 2015). Due to this artificial selection and their long history of domestication there are over 1400 diverse sheep breeds recorded, which were even commented on by Darwin (Darwin 1859; Kijas *et al.* 2009). However, the global sheep industry focuses on a subset of breeds with enhanced productivity and high-quality outputs, which has resulted in many localised breeds becoming endangered or extinct (Taberlet *et al.* 2008).

1.2 STUDYING SHEEP

Due to their multiple uses sheep are very important economically and have therefore been studied to increase their productivity (Byrne *et al.* 2010; Renwick 2013). This research has been aided by the genomic resources that have recently become available, such as a reference sheep genome (Archibald *et al.* 2010; Jiang *et al.* 2014). Sheep can also be genotyped for

more than 54 000 SNPs (single nucleotide polymorphisms-a type of genetic marker) using the Illumina Ovine SNP50 Beadchip, a commercial SNP array platform generated from sheep genome re-sequencing experiments using Illumina GA, Roche 454 and Sanger DNA sequencing (Kijas *et al.* 2012). This SNP genotyping platform consists of a microarray of thousands of oligonucleotide probes that can differentially bind to specific genomic DNA segments containing individual SNPs, allowing massively parallel SNP genotyping of sheep genomic DNA samples (Gunderson *et al.* 2006). SNP data from these chips can be used to study many sheep breeds, including the Galway breed.

1.4 Establishment of the Galway Sheep Breed

Galway sheep, pictured in **Figure 1**, are the only sheep breed native to Ireland (Curran 2010). They are medium-sized, hornless sheep with white, medium-length fleece and a characteristic bob of wool on their heads (Curran 2010). Galway sheep are primarily raised for meat, but also have high-quality wool (Curran 2010; DAD-IS 2016). The breed is thought to have originated in the 1840s from the now presumed extinct Roscommon breed, due to their morphological similarity (Curran 2010; DAD-IS 2016). Other breeds thought to have been involved in the formation of the Galway sheep include the Cotswold, Leicester, Southdown, Ryeland, Lincoln, Merino and Border Leicester breeds since there is evidence of these being introduced to Ireland in the 1800s (Curran 2010). The number of different breeds potentially involved in the formation of the Galway breed raises questions about the genetic distinctiveness of the Galway breed and which breeds they are more closely related to.

Figure 1. Galway ewe with lambs (Louise Byrne, 2003, modified from www.fao.org/dadis).

The *Galway Sheep Breeders' Society* was formed in 1923 to maintain the flock book, hold shows and improve the breed by distributing high quality Galway rams (Curran 2010). By the 1950s the Galway breed had replaced the Roscommon as the focus of lowland sheep farming in western Ireland (Curran 2010). This was because, compared to the Roscommon breed, Galway sheep were faster to mature, better at producing early lambs and less susceptible to disease (Curran 2010). A grant scheme for the purchase of Galway rams by the Galway County Committee of Agriculture in 1949 also aided their rise in popularity (Curran 2010). The Galway breed reached a peak of ~950 registered pedigree ewes in 1960 but declined after this time, following the trend of the total sheep population in Ireland and illustrating the effect the state of the sheep industry can have on individual breeds (Martin 1975a).

1.5 IMPROVEMENT OF THE GALWAY SHEEP BREED

A breeding programme was started in 1963 to improve the meat quality and yield of the Galway breed (Curran 2010). A recording programme was also introduced in 1969 and premium ram scheme in the 1970s, which both aimed to improve the litter size of the Galway breed genetically (Martin 1975a; Hanrahan 1977). This was because the Galway sheep had smaller litter sizes than similar British breeds: about 1.4 compared to 1.7 for comparable British ewes (Hanrahan 1976). However, studies found that the low heritability of litter size in Galway sheep meant that improvement through within-breed selection would be extremely slow and crossing with other, more prolific breeds would provide a better solution (Hanrahan 1976; More O'Ferrall 1976; Hanrahan 1977). This led to 'Fingalway' (Finnish Landrace and Galway) crosses that were an improvement on the reproductive performance of the Galway breed (Hanrahan and Quirke 1975; Quirke 1978b; Quirke 1978a). The existence of known crosses between Galway sheep and other breeds raises questions regarding the level of admixture or gene flow from other breeds into the Galway breed.

1.6 DECLINE OF THE GALWAY SHEEP BREED

Even at a peak in 1960, the population size of pedigree Galway sheep was smaller than other pedigree breeds (Martin 1975b). By the mid-1970s the Galway was still the principal Irish lowland sheep breed and accounted for approximately 35% of the national sheep flock, but was localised within a 180km^2 area in the west of Ireland (Martin 1975b; Martin 1975a). The breed further declined so that by 1985 Galway ewes made up only 12% of the total flock of breeding ewes (Curran 2010). The increase in popularity of the more productive Suffolk and Texel breeds is linked with this decline (Howard 2008). By 1994, the Galway breed had reached the 'critical' status for sheep breeds as defined by the UK Rare Breeds' Survival Trust, by having only 300 registered pedigree breeding ewes (Curran 2010).

However, it is important to note that non-pedigree Galway sheep not involved in the Society were not counted (Curran 2010).

Starting in 1991, the EU Rural Environmental Protection Scheme (REPS) paid premiums for *in situ* conservation of Galway sheep (Curran 2010). The Galway Sheep Conservation Project was then started with the aim of computerising the flock book, identifying further flocks and formulating mating strategies to conserve genetic variation in the population (Curran 2010). Despite this, the breed was officially classed as endangered by the Department of Agriculture in 1998 due to their low population size (Curran 2010). Since being classed as endangered, the numbers of pedigree Galway sheep have increased; however, the breed population size is currently decreasing (Food and Agriculture Organization 2015; DAD-IS 2016). This is likely due to the closure of the REP Scheme to new entrants since 2009, but the replacement schemes may help the Galway breed (Murphy *et al.* 2014). However, the Galway remains the only endangered sheep breed in Ireland (McHugh *et al.* 2014). Additionally, a small population of Galway sheep are farmed in the UK, where they are also classified as endangered (DAD-IS 2016). **Figure 2** shows the changes in the numbers of ewes of all breeds, the Galway breed and pedigree Galway breed in Ireland since the 1950s. The clear decline of the census population size of the Galway sheep breed raises questions regarding the viability of the Galway breed, as well as the level of genetic diversity remaining.

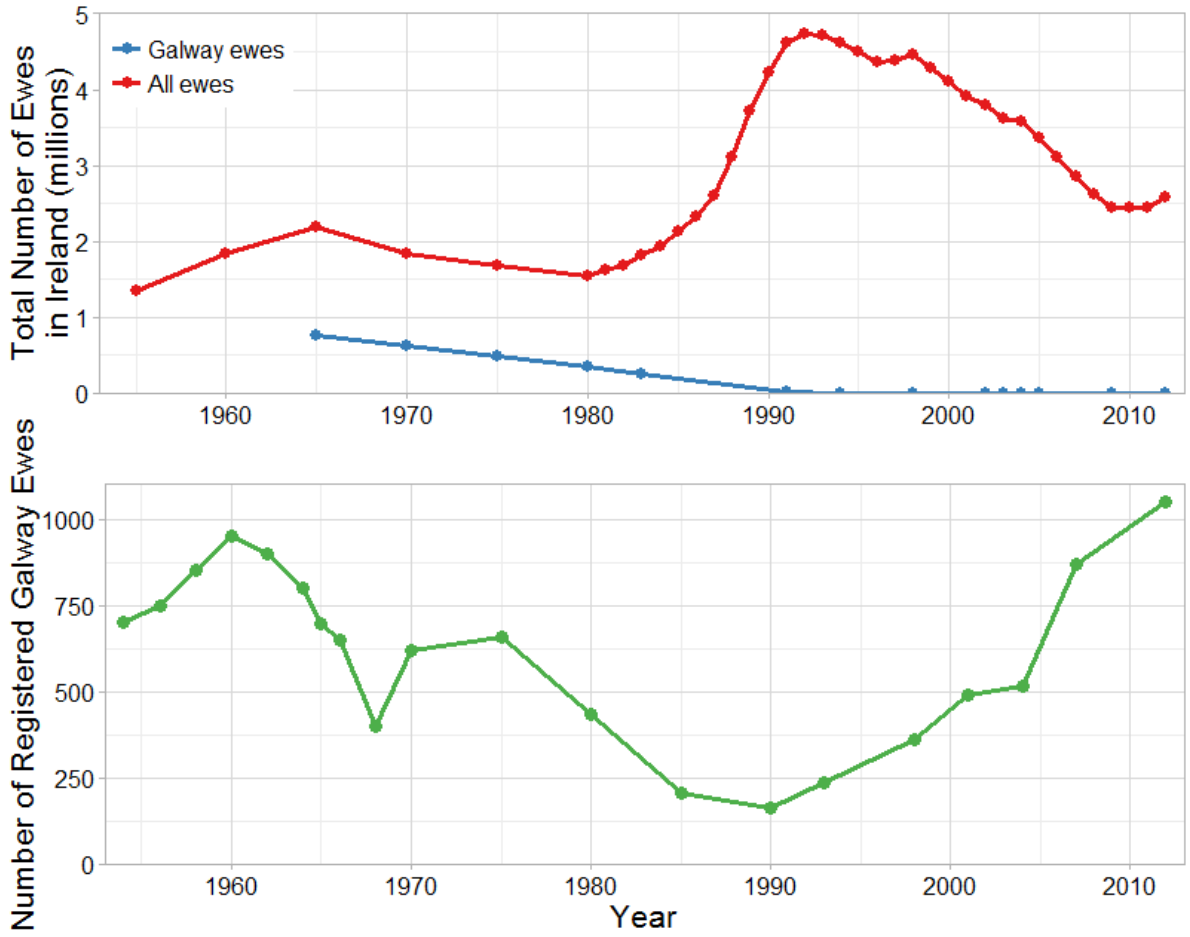

Figure 2. Graphs comparing the total number of ewes (in millions) in Ireland from 1955 to 2012 (red), total number of Galway ewes (in millions) from 1965 to 2012 (blue) and number of registered Galway ewes from 1954 to 2012 (green). Information from the StatBank database of the Central Statistics Office, analysis of Galway sheep flock books, Galway Sheep Breeders Association website and FAO database and produced using the ggplot2 package and multiplot function in R (Martin 1975a; Hanrahan 2007; Wickham 2009; Teetor 2011; R Core Team 2015; CSO 2016; DAD-IS 2016).

1.7 Conservation of the Galway Sheep Breed

The purpose of biological conservation is to protect the genetic diversity of genotypes within species, as well as the diversity of species and the ecosystems they inhabit (Allendorf *et al.* 2013). It is important to conserve the diversity of domesticated, as well as wild, species because increasing demands for productivity has led to a decline in diversity and the extinction of many traditionally farmed breeds (Taberlet *et al.* 2008). These breeds are important because they act as reservoirs of domestic species genetic diversity (Taberlet *et al.*

2008; Bowles 2015). In particular, local sheep breeds on the periphery of Northern Europe are important sources of genetic diversity that should be conserved (Tapio *et al.* 2005). The conservation of the Galway breed is important for these scientific and agricultural reasons, as well as cultural reasons (Curran 2010). These include their history as part of the Irish landscape, use in traditional craft products and rural tourism (Gandini and Villa 2003; Curran 2010).

Since the Galway sheep is a local breed with a low population size, the main threat it faces is replacement by more productive, commercial breeds, which would further reduce the population size and increase inbreeding (Taberlet *et al.* 2008; Curran 2010). Inbreeding is mating between related individuals and increases homozygosity which can lead to inbreeding depression (Allendorf *et al.* 2013). This is a reduction in the fitness of the offspring of related individuals compared to unrelated individuals, primarily due to the emergence of recessive genetic conditions and a reduction in overdominance (heterozygote advantage) at many genomic loci (Allendorf *et al.* 2013). Other problems for local breeds include increased genetic drift (change in allele frequencies due to random sub-sampling of alleles from generation to generation), poor animal husbandry and management which can increase inbreeding, crossbreeding which reduces adaptation and geographic confinement which increases the risk of extinction (Taberlet *et al.* 2008; Allendorf *et al.* 2013).

Conservation genetics and genomic technologies have been proposed as tools that could be used as part of the conservation programme for Galway sheep (McHugh *et al.* 2014). Conservation genetics and genomics (the use of genetic and genomic information in conservation) have been used to aid the conservation of endangered wild species and the management of endangered domestic populations (Der Sarkissian *et al.* 2015; Meszaros *et al.* 2015). Additionally, genomic and phenotypic information could be used to increase the

productivity of local breeds through selection, which could reduce the likelihood of their replacement by more productive breeds (Biscarini *et al.* 2015). The recommendations of *Teagasc* and *SheepIreland* include the collection of genetic, phenotypic and pedigree information which would facilitate research on genetic associations, parentage, genomic selection and genetic evaluations that could improve breeding programmes (McHugh *et al.* 2014).

1.8 TYPES OF ANALYSIS

The fixation index (F_{ST}) is normally considered a measure of the genetic diversity of subpopulations (i.e., breeds) as a proportion of the total population (Weir and Cockerham 1984). Pairwise F_{ST} values, calculated using allele frequencies, can be used to examine the relatedness of pairs of sheep breeds since $F_{ST} = 1$ indicates the breeds do not share any genetic diversity and are completely separate, while $F_{ST} = 0$ indicates the breeds are interbreeding freely and are completely mixed (Weir and Cockerham 1984; Kijas *et al.* 2009). Comparing pairwise F_{ST} values will allow the relative distinctness of the Galway breed to be assessed and may indicate related breeds.

The uniqueness of some sheep breeds can also be demonstrated using principal component analysis (PCA) (Lawson Handley *et al.* 2007). PCA involves plotting data points for individual samples or discrete populations along axes of variation that explain the main patterns of genetic variability (Reich *et al.* 2008). This analysis can also reveal important features of the histories of particular populations (Patterson *et al.* 2006).

The relationships between domestic animal populations or breeds can be examined using recently developed phylogenetic graph methods, which can infer the patterns of population splits and mixtures in multiple populations (Pickrell and Pritchard 2012). Initially, a maximum likelihood tree that yields the highest probability of evolving the observed data is

built with the root fixed by a known outgroup (Felsenstein 1981). Populations that fit the model poorly are then identified using residuals and migration events or gene flow between these populations are modelled to increase the likelihood of the graph (Pickrell and Pritchard 2012).

Gene flow between breeds can be examined in greater detail using admixture or population structure analysis (Raj *et al.* 2014). Individual sheep can be assigned to populations that have different allele frequencies using a model that allows for individuals to be admixed or contain alleles from more than one of the populations (Pritchard *et al.* 2000). The number of assumed populations (breeds) or model complexity (indicated by the parameter *K*) can be tested and a range of probable values can be found (Raj *et al.* 2014).

The genetic effective population size (N_e) is the size of the ideal population that would display the same level of genetic drift as the actual population and can be important in assessing the viability of endangered populations (Allendorf *et al.* 2010). Trends in the historic effective population size can be estimated using the relationship between N_e, r^2 (a measure of linkage disequilibrium, the non-random association of alleles) and c (the recombination rate) (Barbato *et al.* 2015).

Inbreeding can be defined as mating between individuals that are more related to each other than the average relatedness of the population (Curik *et al.* 2014). *F* coefficients of inbreeding are traditionally estimated using pedigree information and measure the expected homozygosity of an individual, with $F = 1$ being completely homozygous, but can also be calculated using allele frequencies and the probabilities of SNPs being homozygous (Purcell *et al.* 2007). These different approaches have been shown to have a reasonable correlation in sheep (Li *et al.* 2011).

Pedigree-based inbreeding coefficients have also been found to correlate with runs of homozygosity (ROH) (Purfield *et al.* 2012). ROH are long stretches of homozygosity that most likely arise due to inbreeding and can be identified through surveys of genome-wide SNP data in populations (Purfield *et al.* 2012; Curik *et al.* 2014). Long ROH are more likely due to more recent inbreeding while short ROH, which can be formed by the breakdown of long ROH over time, are more likely to indicate ancestral inbreeding (Curik *et al.* 2014). Examining the different number and lengths of ROH can show differences in breed origins and management (Purfield *et al.* 2012). Additionally, some regions of the genome containing more ROH (ROH hotspots) may contain genes under selection (Pemberton *et al.* 2012; Kim *et al.* 2013).

Artificial selection is the selection of heritable phenotypes by humans to elicit a desired phenotypic change in succeeding generations and leaves genetic signatures, such as increases in the frequencies of specific alleles or genomic segments within, or proximal to genes and/or regulatory elements controlling the trait under selection (Allendorf *et al.* 2013; Andersson 2013). There are many methods for detecting signatures of selection in domestic animal populations including: SNP-specific fixation index (F_{ST}); directional change in selected allele frequency (ΔSAF), which estimates the differences in allele frequencies between populations with unknown ancestral alleles; and the cross-population extended haplotype homozygosity test (XP-EHH), which compares the probabilities that different populations or breeds have homozygosity that extends beyond a given haplotype (a discrete genomic segment that tends to be transmitted as a single unit as a consequence of population genetic processes specific to a particular population) (Sabeti *et al.* 2007; Kijas *et al.* 2012; Randhawa *et al.* 2014). These tests can be combined into one composite selection signal (CSS), which increases the power to detect selection (Randhawa *et al.* 2014). This approach can be used to examine SNP data

from the Galway breed and other sheep populations to detect specific signals of selection across the genome (Kijas *et al.* 2012; Randhawa *et al.* 2014).

2. AIMS AND OBJECTIVES

The aim of the research work described in this thesis is to use population genomic approaches to examine the recent evolutionary history of the Galway sheep breed. This includes assessing the genetic uniqueness of the Galway sheep as a distinct breed with its own particular history of crossbreeding and artificial selection. The effective population size and level of inbreeding can also be estimated to provide information relevant for genetic conservation of the Galway breed. These analyses can be undertaken using genome-wide SNP data available from the International Sheep Genomics Consortium HapMap Project (www.sheephapmap.org/hapmap.php) for the Galway breed and other sheep breeds from Ireland, Britain and Northern Europe.

The specific research questions to be addressed are as follows:

1. Is the Galway breed distinct from other sheep breeds? Is it distinct enough to be considered a unique breed?
2. What is the phylogenetic history of the Galway breed? What breeds are closely related to the Galway breed? Do genomic results concur with existing information concerning breeds involved in the formation of the Galway breed?
3. Is there any evidence of admixture or migration into the Galway breed from other breeds? Does genomic admixture concur with existing information about crossbreeding of the Galway sheep with other breeds to improve quality and production?
4. What is the genetic effective population size of the Galway breed? Is the effective population size comparable to other local breeds known to be of conservation concern or is the population similar to larger commercial breeds?

5. What is the level of inbreeding in the Galway breed? How does the level of inbreeding in Galway sheep compare to other breeds of known conservation concern versus commercial breeds?
6. How much of the genome of the Galway sheep population is covered by runs of homozygosity (ROH) on average? What is the proportion of long to short ROH in the Galway sheep population? How do ROH in Galway sheep compare to other breeds either endangered or of least concern? Are these results consistent with results from the inbreeding analysis?
7. Is there any evidence of selection in the Galway breed when examined using genome-wide-SNP data from a number of different sheep breeds? What are the candidate genes under selection and what are their functions? Does selection of these genes make sense in light of domestication and subsequent natural and artificial selection?

3. MATERIALS AND METHODS

3.1 DATA SOURCE

The genome-wide SNP data set obtained from the International Sheep Genomics Consortium HapMap Project consisted of 2819 sheep from 74 breeds DNA-sampled using buccal swabs, whole blood, hair or other biological tissues (Kijas *et al.* 2012). Genomic DNA samples were genotyped for 49 034 evenly-spaced SNPs using the Illumina Ovine SNP50 Beadchip, including Galway breed DNA samples collected by the UCD Animal Genomics Group (Kijas *et al.* 2012). All breeds were sampled from multiple flocks to obtain representative samples of within-breed diversity and the average SNP genotype call rate was 99.8% (Kijas *et al.* 2012).

3.2 SHEEP BREEDS FOR ANALYSIS

To focus on the Galway breed, a subset of breeds, shown in **Table 1**, was selected for analysis. This included breeds known to be related or crossed with the Galway breed, such as the Merino and Border Leicester breeds, which have been proposed to be involved in the formation of the Galway breed (Curran 2010). The Border Leicester breed was also involved in the formation of the Texel breed (Howard 2008). Galway sheep have been known to be crossbred with Finnish Landrace, Scottish Blackface and Suffolk sheep (Quirke 1978b; Howard 2008). The Dorset Horn, New Zealand Romney, Soay and Wiltshire breeds have no known connection to the Galway breed. These breeds were selected for comparison due to their origins in the United Kingdom, shown in **Figure 3**, and history of study (DAD-IS 2016).

Table 1. Breed name, sample size and origin of the breeds selected for analysis.

Breed Name	Sample Size	Breed Origin
Australian Merino	50	Spain/ Unknown
Border Leicester	48	England
Dorset Horn	21	England
Finnish Landrace	99	Finland
Galway	49	Ireland
Irish Suffolk	55	England
New Zealand Romney	24	England
Scottish Blackface	56	Scotland
Scottish Texel	80	Netherlands
Soay	110	Scotland
Wiltshire	23	England
Total	615	

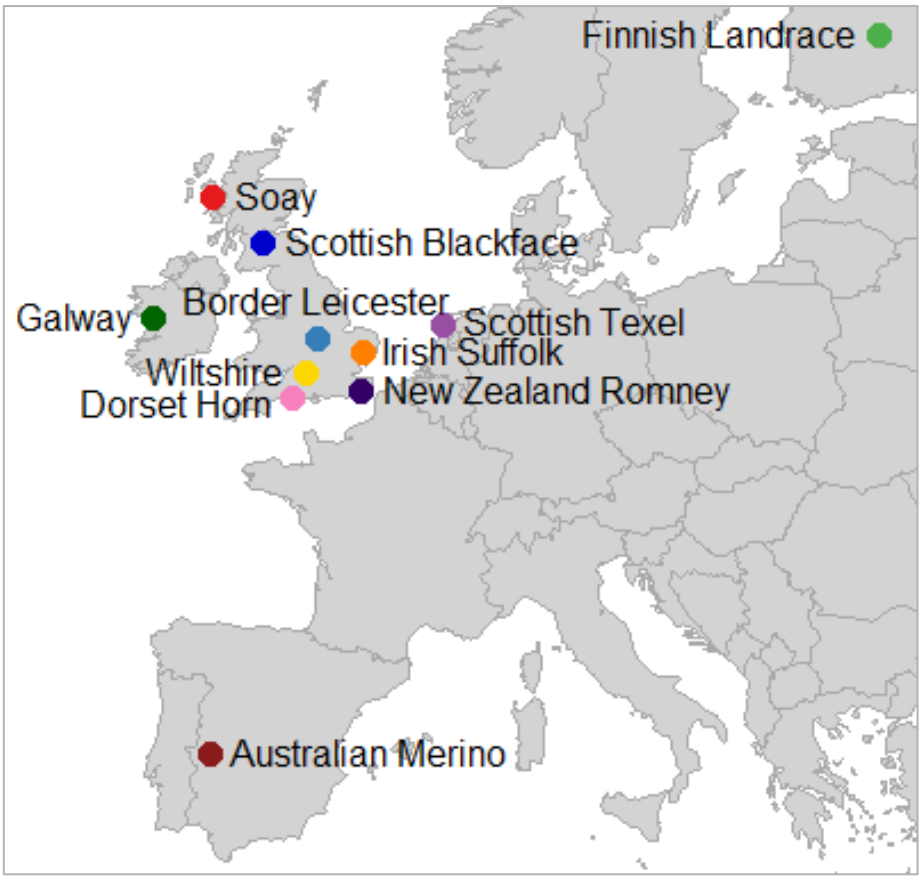

Figure 3. Map showing the origin of the breeds, adapted from Kijas *et al*. (2012) and produced in R using the maps, mapdata and ggmaps packages (Becker *et al.* 2013a; Becker *et al.* 2013b; Kahle and Wickham 2013; R Core Team 2015).

3.3 DATA FILTERING

The SNP data set was obtained as PLINK MAP and PED files and transformed into binary files using the `--make-bed` command, allowing for faster computation, and using the `--sheep` command to account for 26 autosomal sheep chromosomes (Purcell *et al.* 2007). A text file obtained by editing the resulting FAM file to contain only the individual sheep from the selected breeds shown in **Table 1** was used with the `--keep` command to remove the other sheep samples from the files. The resulting data set was filtered to remove 110 SNPs unassigned to any chromosome, 281 SNPs with no position and 1231 SNPs assigned to the X and Y chromosomes to avoid problems caused by unassigned or sex-linked SNPs in some of the analyses (Patterson *et al.* 2006; Purfield *et al.* 2012). This was done using the `--exclude` command with a text file containing a list of the SNPs to be excluded obtained by editing the BIM file. This left 47 412 SNPs in 615 individual sheep with a total genotyping rate of 99.7%.

3.4 F_{ST} ANALYSIS

Pairwise F_{ST} was calculated for each pair of breeds using the `--fst --within` commands in PLINK1.9, PLINK files of the selected sheep breeds and text files containing the individuals from the pair of breeds to be analysed obtained by editing the FAM file (Weir and Cockerham 1984; Chang *et al.* 2015). The weighted values were chosen to account for different sample sizes.

3.5 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) was carried out using the EIGENSOFT package (Patterson *et al.* 2006). The PLINK files were saved using the appropriate file extensions (.bim = .pedsnp, .fam = .pedind), the end of line symbol was converted for a Linux

environment and the -9 value for missing data in the PEDIND file was changed to 1. The files were converted using the `convertf -p par.file` command and a parameter file containing the filenames, output format, confirmation that the files contained family (breed) names and number of autosomes (26). The analysis was performed using the command `smartpca -p par.data` and a parameter file containing the filenames. The number of outlier removal iterations was set to 0 to turn outlier removal off since outliers could flag sheep that were the result of crossbreeding. The analysis was repeated with the 110 Soay sheep removed following the filtering procedure described above due to their high level of differentiation from the other breeds. PCA plots were generated using R and the ggplot2 and scatterplot3d packages (Ligges and Mächler 2002; Wickham 2009; R Core Team 2015).

3.6 PHYLOGENETIC ANALYSIS

Phylogenetic analysis was performed using the TreeMix software package with PLINK files filtered as described above to include the Comisana breed as an outgroup due to their geographic and genetic isolation from the selected breeds (Pickrell and Pritchard 2012; Ciani *et al.* 2014). An allele frequency cluster file was produced using the PLINK `--freq --within data.clust` commands, where data.clust is a text file generated by editing the FAM file to contain the breed name, individual ID and cluster (breed) of each sheep. The resulting .frq.strat file was gzipped and converted to TreeMix format using the `plink2treemix` python script (Pickrell and Pritchard 2012). TreeMix was run with the Comisana breed as the root and four migration edges. The analysis was repeated using a larger subset of the breeds of European origin (**Table S1**), the Indian Garole breed as an outgroup and two migration edges (Beynon *et al.* 2015). Graphs were produced using R and the R script provided with TreeMix (Pickrell and Pritchard 2012; R Core Team 2015).

3.7 ADMIXTURE ANALYSIS

Admixture analysis was performed using fastSTRUCTURE with the PLINK files of the selected sheep breeds (Raj *et al.* 2014). The analysis was carried out with the model complexity or number of populations (*K*) set from two to eleven. The `chooseK` function was used to test the outputs to find a range of values of *K* that best accounted for the structure in the data (Raj *et al.* 2014). The results were graphed using the DISTRUCT program (Rosenberg 2004).

3.8 EFFECTIVE POPULATION SIZE

Analysis of trends in historical effective population size was performed using the SNeP program (Barbato *et al.* 2015). Separate PLINK files for each breed were generated following the filtering procedure described above with the `--recode` command since SNeP does not use binary files (Purcell *et al.* 2007; Barbato *et al.* 2015). Each file was analysed in SNeP using the `-samplesize 1` command to account for sample size since the gametic phase was unknown (Barbato *et al.* 2015). The results were graphed in R using the ggplot2 package and the multiplot function (Wickham 2009; Teetor 2011; R Core Team 2015). Since the distribution of results was not normal, a Kruskal-Wallis test was carried out followed by pairwise Wilcox tests using Bonferroni corrections (R Core Team 2015).

3.9 INBREEDING ANALYSIS

Analysis of genomic inbreeding was carried out using the PLINK `--het` command with the PLINK file of genome-wide SNP data from the selected breeds since similar inbreeding results have been found using pruned and unpruned data for a similar-sized data set (Purcell *et al.* 2007; Binns *et al.* 2012). The results were graphed in R using the ggplot2 package (Wickham 2009; R Core Team 2015). Since the design was unbalanced due to unequal

sample sizes, a Kruskal-Wallis test was carried out followed by pairwise Wilcox tests using Bonferroni corrections (R Core Team 2015).

3.10 Runs of Homozygosity Analysis

Analysis of ROH was performed using the PLINK `--homozyg` command following the method of a previous study using a similar-sized data set (Purcell *et al.* 2007; Purfield *et al.* 2012). A window of 50 SNPs was moved in one SNP intervals along the genome to estimate homozygosity, a maximum of two SNPs with missing genotypes and one heterozygous SNP was allowed in each window and the minimum SNP density allowed was set to one SNP every 120 kb (Purfield *et al.* 2012). There was no minimum number of SNPs in a ROH but the maximum gap between SNPs in a ROH was set to 1000 kb to account for the genotype density (Purfield *et al.* 2012). The inbreeding coefficient based on ROH (F_{ROH}) was calculated by dividing the total length of ROH for each individual by the length of the genome covered in SNPs (2449.65 Mb for the sheep autosomes) (Purfield *et al.* 2012; Jiang *et al.* 2014). The results were graphed in R using the ggplot2 package (Wickham 2009; R Core Team 2015). Since the design was unbalanced due to unequal sample sizes, a Kruskal-Wallis test was carried out followed by pairwise Wilcox tests using Bonferroni corrections (R Core Team 2015).

3.11 Selection Analysis

The data was analysed for signatures of selection through collaboration with Dr Imtiaz Randhawa, using the CSS method he has recently developed (Randhawa *et al.* 2014). The 49 Galway sheep were analysed against 50 sheep selected from the other 10 breeds, five animals from each breed. The results were averaged over 1 Mb sliding windows to produce 'smooth' Manhattan plots (Randhawa *et al.* 2014). This procedure was performed three times

to ensure the results were not due to differences in the selected sheep. Ensembl was used to obtain a list of sheep genes and an R script was used to search for genes 1 Mb upstream and downstream of clusters of SNPs within the top 0.1% CSS (Randhawa *et al.* 2014). The Ensembl IDs of these candidate genes were converted to HGNC (HUGO Gene Nomenclature Committee) symbols using Biomart (Smedley *et al.* 2015), duplicates were removed to leave 119 genes that were analysed using Ingenuity Pathway Analysis (IPA-www.ingenuity.com) to identify enriched biological pathways for this particular set of 119 genes.

4. RESULTS

4.1 F_{ST} RESULTS

The pairwise weighted F_{ST} values for each pair of breeds are shown in **Table S1** and ranged from 0.080 (Australian Merino and Scottish Blackface) to 0.326 (Soay and Wiltshire). The Galway breed exhibited lower pairwise F_{ST} values of 0.110, 0.118 and 0.119 with the New Zealand Romney, Australian Merino and Scottish Texel breeds, respectively. The Galway breed also exhibited higher pairwise F_{ST} values of 0.250 and 0.207 with the Soay and Wiltshire breeds, respectively.

4.2 PRINCIPAL COMPONENT ANALYSIS RESULTS

The results of the PCA and percentage of variation explained by each component are shown in **Figure 4**. The x-axis, which represents the first principal component (PC1) and explains 29.36% of the variation in the data, separates the Soay breed from the other populations. The second principal component (PC2) on the y-axis accounts for 14.98% of the variation and separates the other breeds into groups. The Galway breed is grouped with the Scottish Texel breed and close to the Border Leicester breed but does not clearly separate as a distinct breed. Additionally, Irish Suffolk individual ISF25 groups with the Galway and Scottish Texel breeds rather than the rest of the Irish Suffolk breed. Other Irish Suffolk individuals do not group with any particular breed.

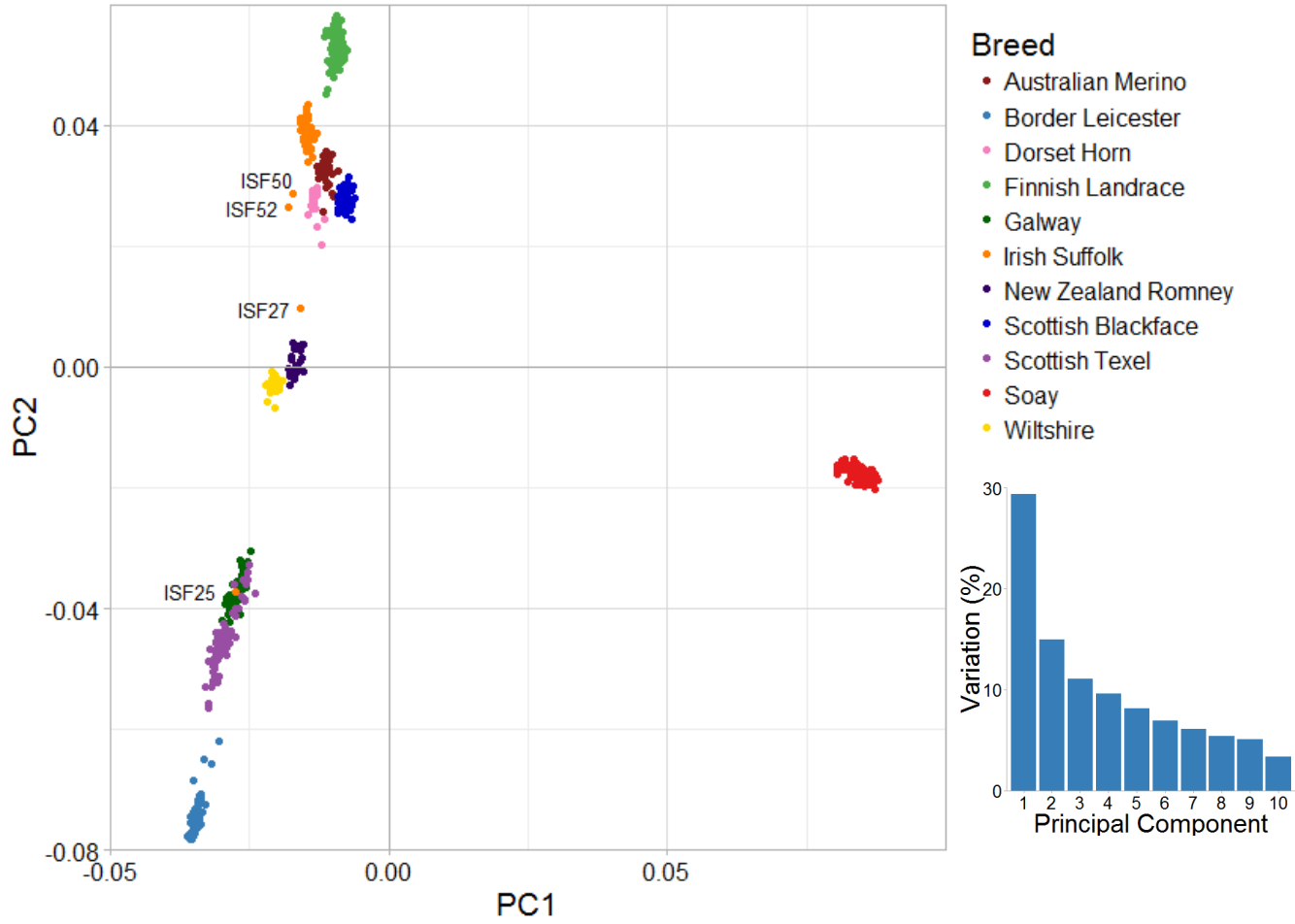

Figure 4. Scatter plot showing the results of the PCA with the first principal component (PC1) on the x-axis and the second principal component (PC2) on the y-axis. Each breed is designated a different colour and individuals that do not group with the rest of their breed are labelled. The bar chart shows the percentage of variation explained by each principal component.

Figure 5 shows results from the same analysis but includes the third principal component (PC3), which explains 11.05% of the variation. This clearly separates the Galway breed from the others as a distinct breed. It also shows that the Irish Suffolk individual ISF25 groups with a subset of Scottish Texel sheep and not the Galway breed.

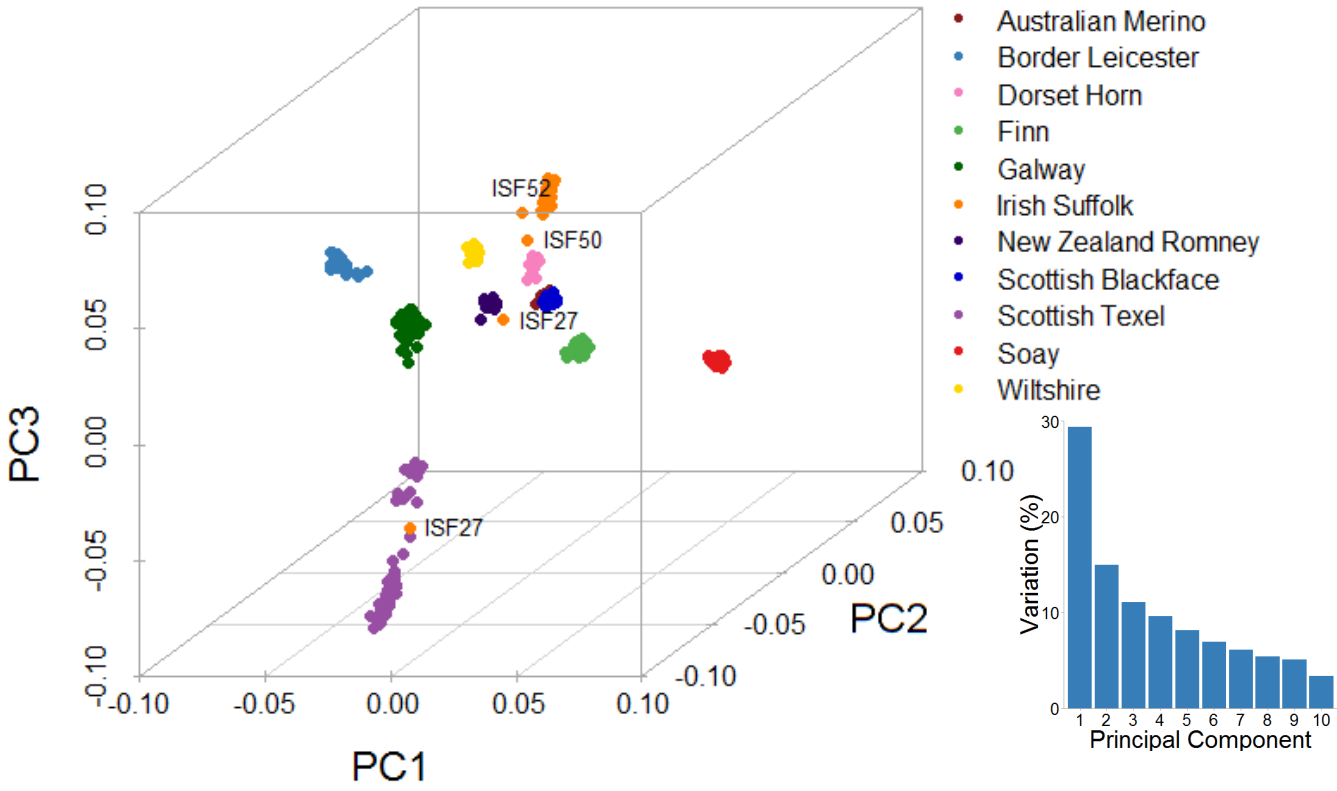

Figure 5. Scatter plot showing the results of the PCA with the first principal component (PC1) on the x-axis, the second principal component (PC2) on the y-axis and the third principal component (PC3) on the z-axis. Each breed is designated a different colour and individuals that do not group with the rest of their breed are labelled. The bar chart shows the percentage of variation explained by each principal component.

Due to the high level of differentiation of the Soay sheep breed another PCA was performed without the samples from this breed; the results of this analysis and percentage of variation explained by each PC are shown in **Figure 6**. Without the Soay breed all of the remaining breeds, except the Australian Merino and Scottish Blackface breeds, clearly separate using only PC1 and PC2, as do a subset of the Scottish Texel breed.

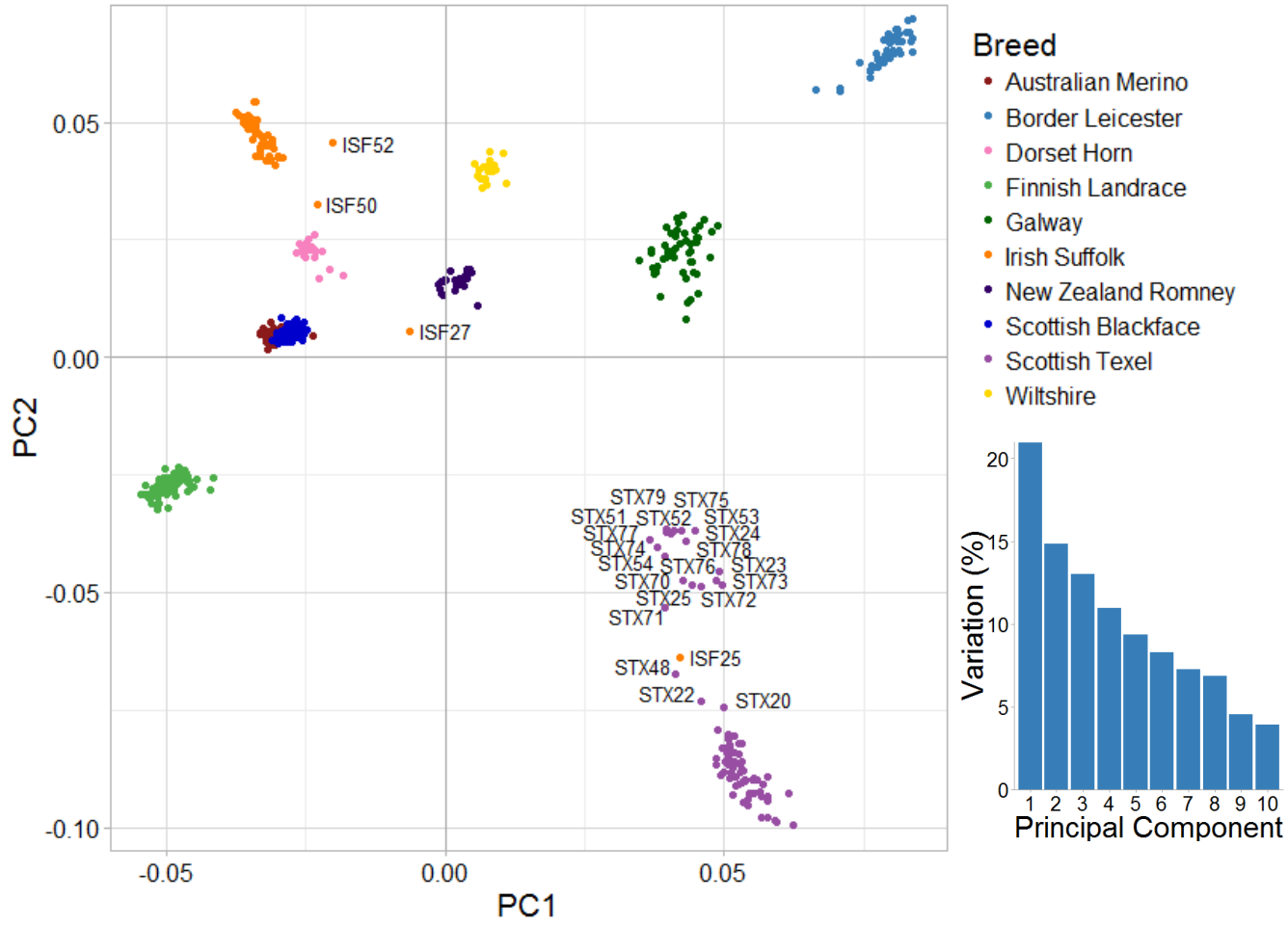

Figure 6. Scatter plot showing the results of the PCA without the Soay sheep breed with the first principal component (PC1) on the x-axis and the second principal component (PC2) on the y-axis. Each breed is designated a different colour and individuals that do not group with the rest of their breed are labelled. The bar chart shows the percentage of variation explained by each principal component.

4.3 PHYLOGENETIC RESULTS

In the TreeMix phylogenetic graph, shown in **Figure 7**, the Galway breed is placed within a group containing the breeds of English origin including the New Zealand Romney, Wiltshire, Border Leicester, Dorset Horn and Irish Suffolk breeds. The Scottish Texel sheep, which originate in the Netherlands, are also included in this group. The arrows (graph edges) indicate migration or gene flow between sheep breeds and the colour indicates the weight of this migration. The arrow with the lowest migration weight is between the Galway and Irish Suffolk breeds.

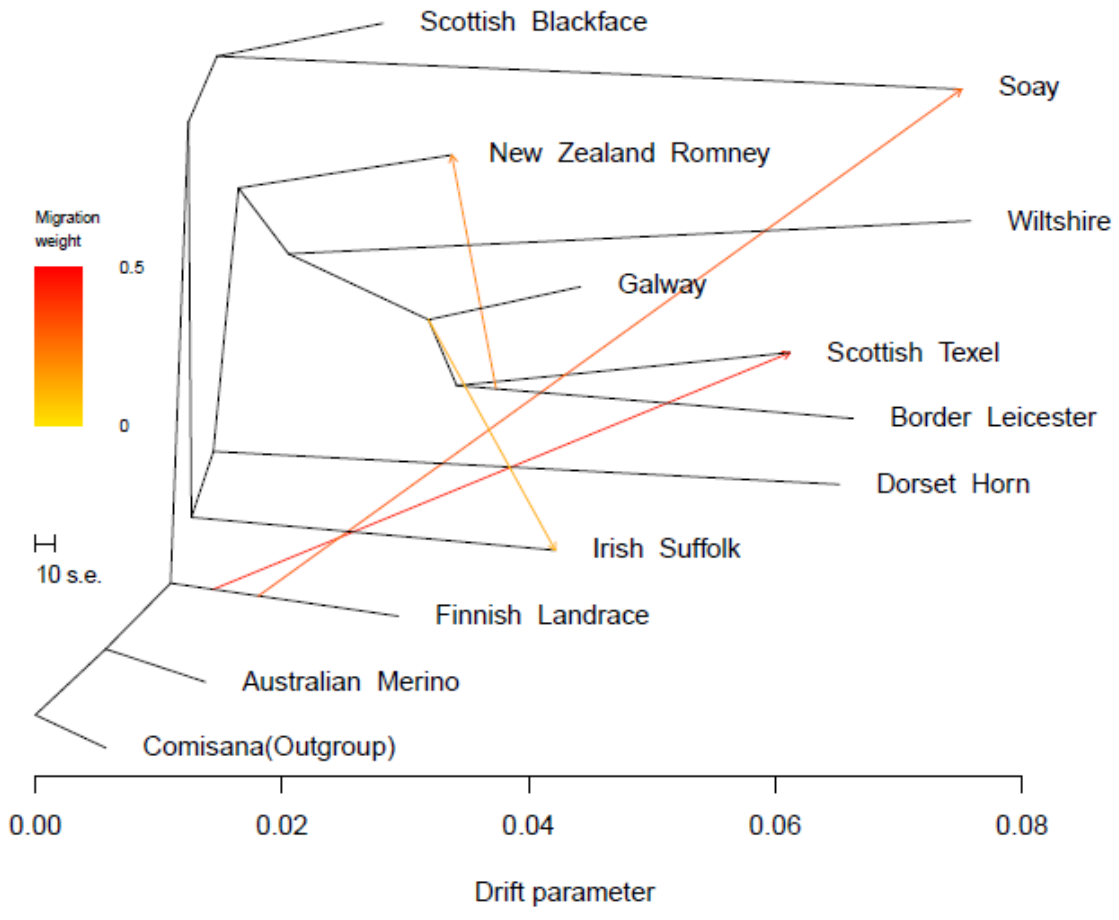

Figure 7. Phylogenetic graph showing the relationships between the sheep breeds. The arrows indicate migration or gene flow between the breeds, the colour of the arrow indicates the weight of this migration. The residuals are shown in **Figure S1.**

The TreeMix phylogenetic graph generated with additional European breeds for a broader picture of the phylogenetic history of the Galway breed in the context of other European breeds, shown in **Figure 8**, supports the results of **Figure 7**. The Galway sheep breed is again included in a group containing the sheep breeds of English origin and placed in a sub-group containing the Border Leicester and Scottish Texel breeds, with the addition of the German Texel breed.

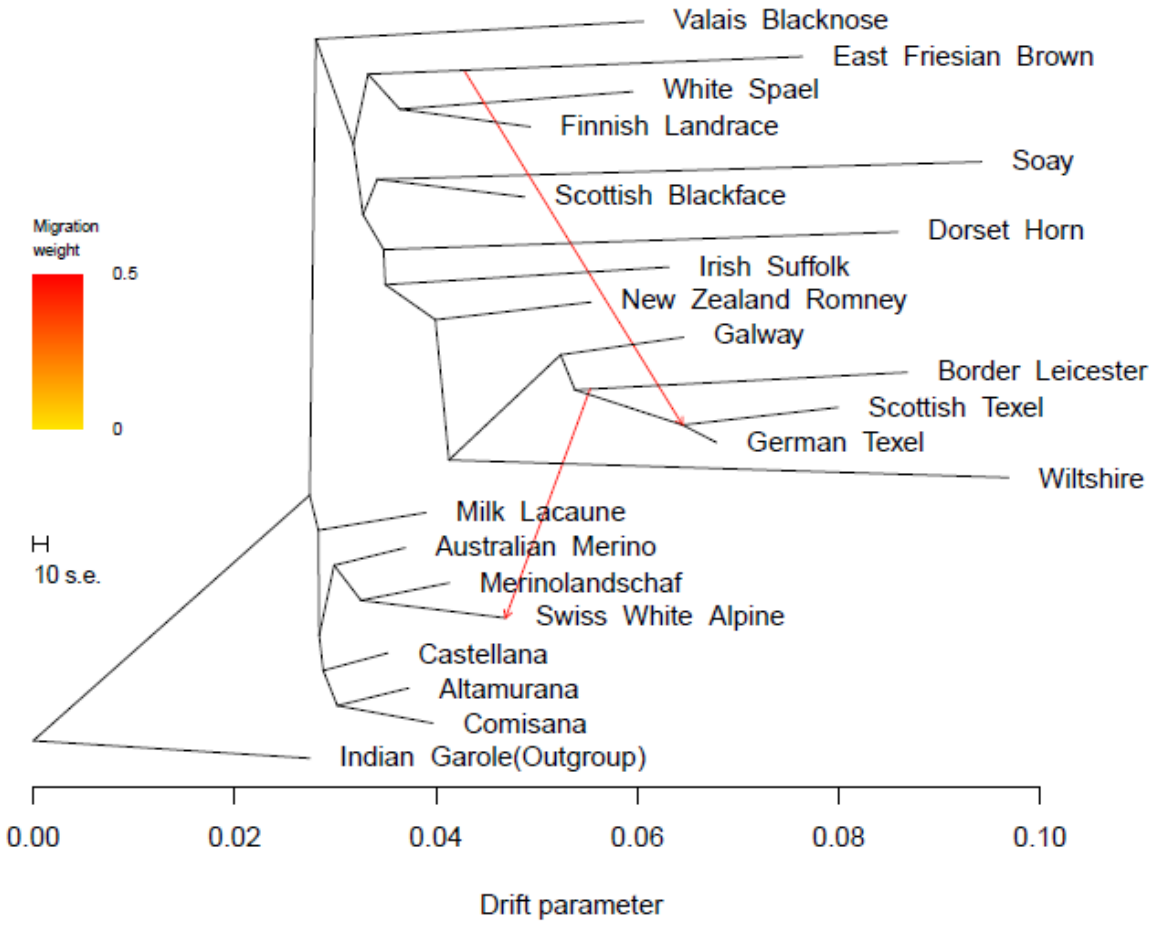

Figure 8. Phylogenetic graph showing the relationships between European sheep breeds. The arrows indicate gene flow or admixture between the breeds, the colour of the arrow indicates the weight of this migration. The residuals are shown in **Figure S2.**

4.4 ADMIXTURE RESULTS

The fastSTRUCTURE admixture results in **Figure 9** are shown for model complexity or number of assumed populations (*K*) from two to eleven because the suitable range of *K* values ranged from *K* = 2 to explain the structure in the data to *K* = 11 to maximise the marginal likelihood.

The colours indicate the placement of the sheep into the assumed populations. From *K* = 4 and onwards it is clear that the Galway breed is made up of different proportions of admixture than the other breeds. By *K* = 9 the Galway breed is shown to be distinct from the others and even makes up a proportion of the variation in the New Zealand Romney breed. At *K* = 11 each breed has its own distinct colour. However, some sheep show evidence of admixture which can be seen in the lines representing individuals that are different colours or proportions of colours than the rest of their breeds. Some individual Galway sheep are 10% or more admixed with other breeds. The Galway breed shows evidence of admixture with the Border Leicester, Scottish Texel and Irish Suffolk breeds. There is evidence of admixture confined to one Galway individual each for the Finnish Landrace, Wiltshire, Australian Merino and Scottish Blackface breeds. There is also evidence of admixture with Galway sheep within the blocks of the Border Leicester, Scottish Texel, Irish Suffolk, New Zealand Romney, Australian Merino and Scottish Blackface breeds.

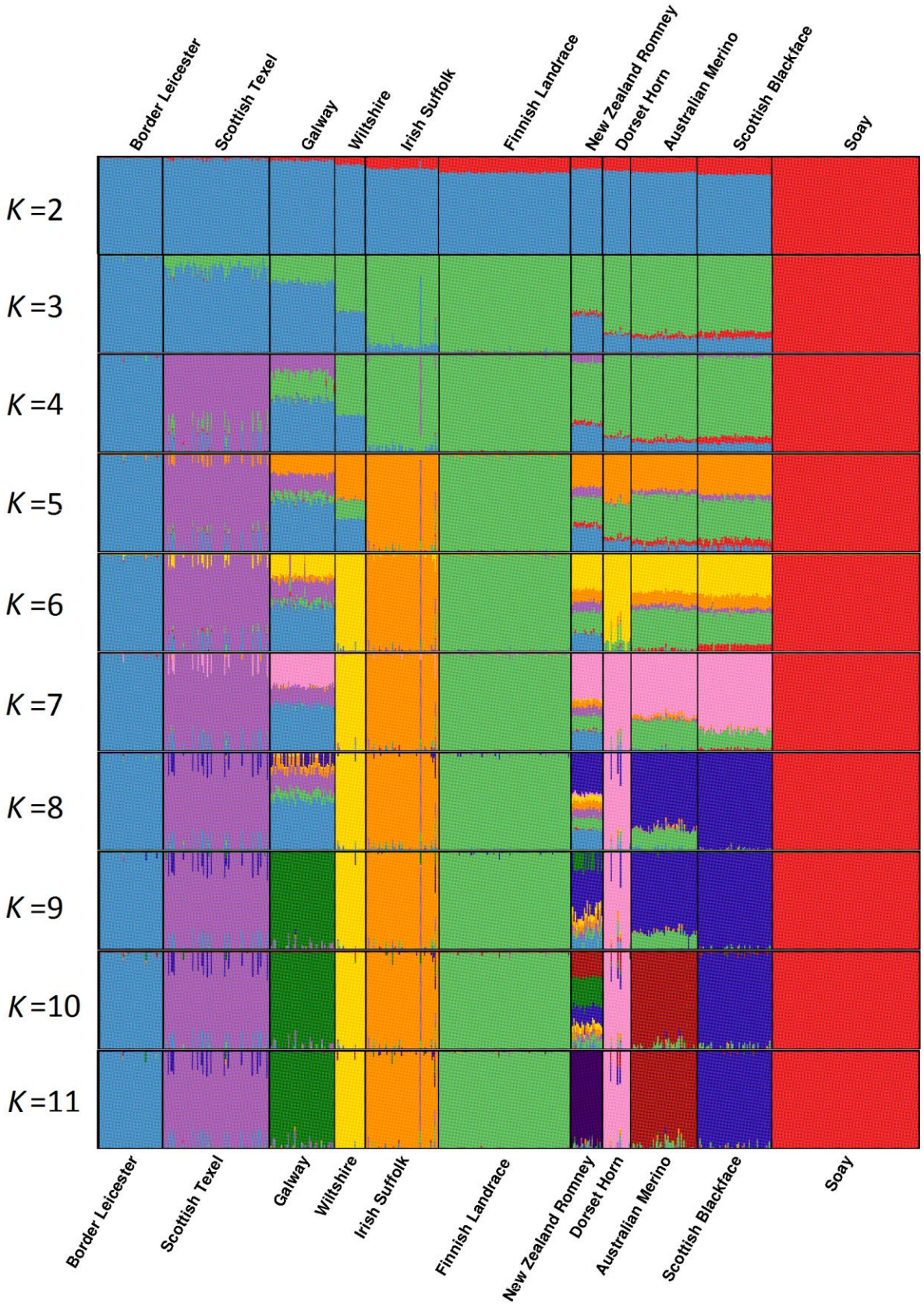

Figure 9. Graph of the results from the fastSTRUCTURE admixture analysis for *K* = 2 to 11.

4.5 Effective Population Size Results

All of the historical trends in effective population size (N_e) for the different sheep breeds shown in **Figure 10** decline towards the present; however, there was a significant difference between the breeds (H=87.408, df=10, P value < 0.001). The main difference between the breeds is the maximum N_e estimated in the past and the shape if the curve declining towards the present. Some breeds, such as the Australian Merino, Finnish Landrace and Scottish Blackface, have high N_e values in the past (~4000) followed by a very steep decline towards the present. In contrast, others, such as the Soay and Wiltshire breeds, exhibit much smaller historical N_e values (~1500) and an almost linear decline towards the present. There were significant differences between these types of breeds (**Table S2**). The Galway breed falls between these two extremes with a medium-sized historical N_e value (2427) and a more gradual decline towards the present and only had a significant difference with the Australian Merino breed (**Table S2**).

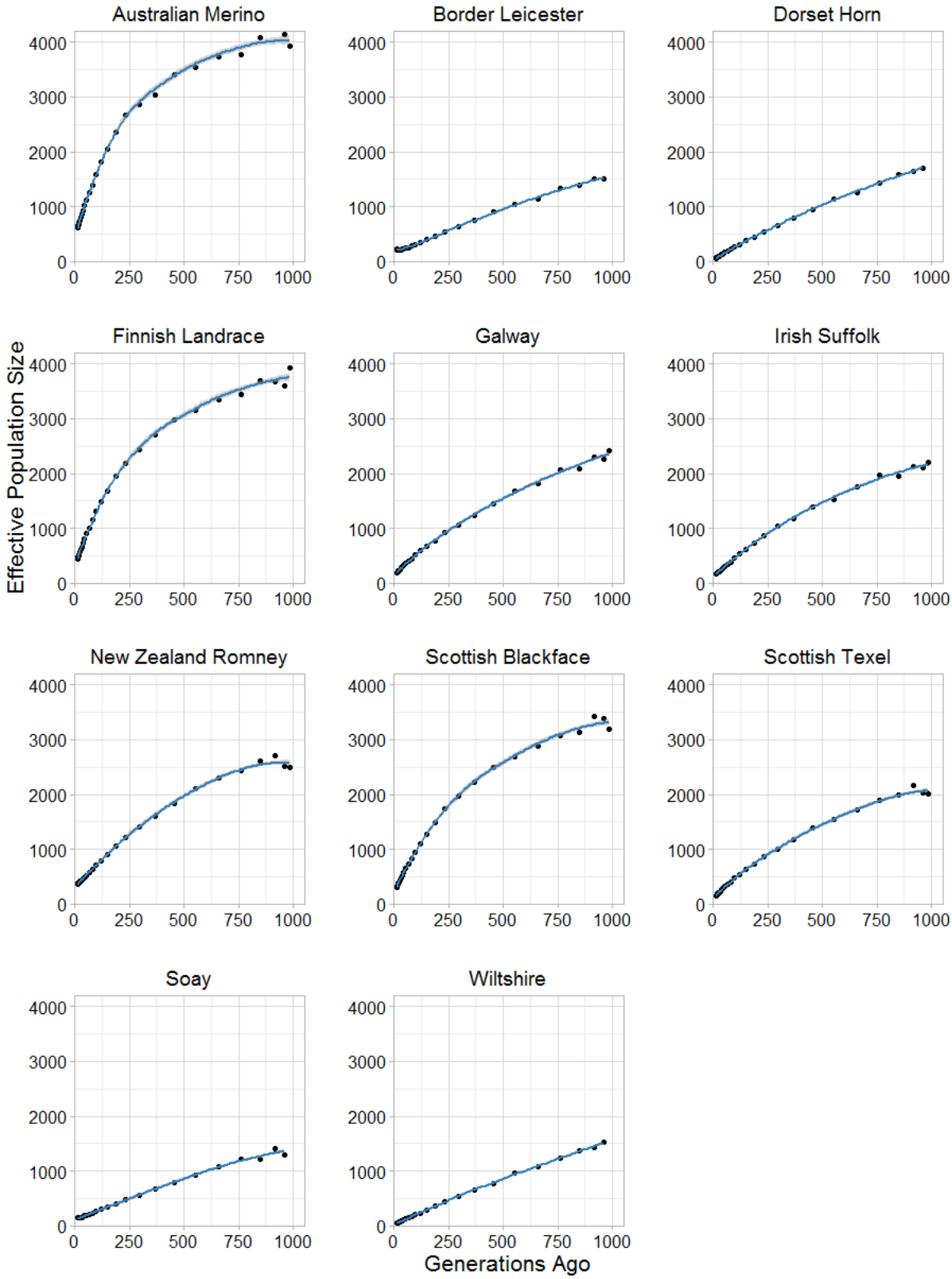

Figure 10. Graphs showing the trend in historical effective population size (N_e) for each breed up to 1000 generations ago.

4.6 INBREEDING RESULTS

The genomic inbreeding coefficient estimates (*F*) for the individual sheep, shown in **Figure 11**, range from less than 0.000 to 0.389. There was a significant difference between the breeds (H=477.57, df=10, P value < 0.001). The Australian Merino breed has the lowest mean genomic inbreeding coefficient estimate of $F = 0.056$ while the Soay breed has the highest ($F = 0.306$). The Galway breed falls between these two values with a mean genomic inbreeding coefficient estimate of $F = 0.129$. There were significant differences between almost all the breeds; however, the Galway breed had no significant difference with the New Zealand Romney and Scottish Texel breeds (**Table S3**). Some individual outliers are more inbred than the rest of their breed, such as Galway individuals GAL26, GAL18 and GAL15.

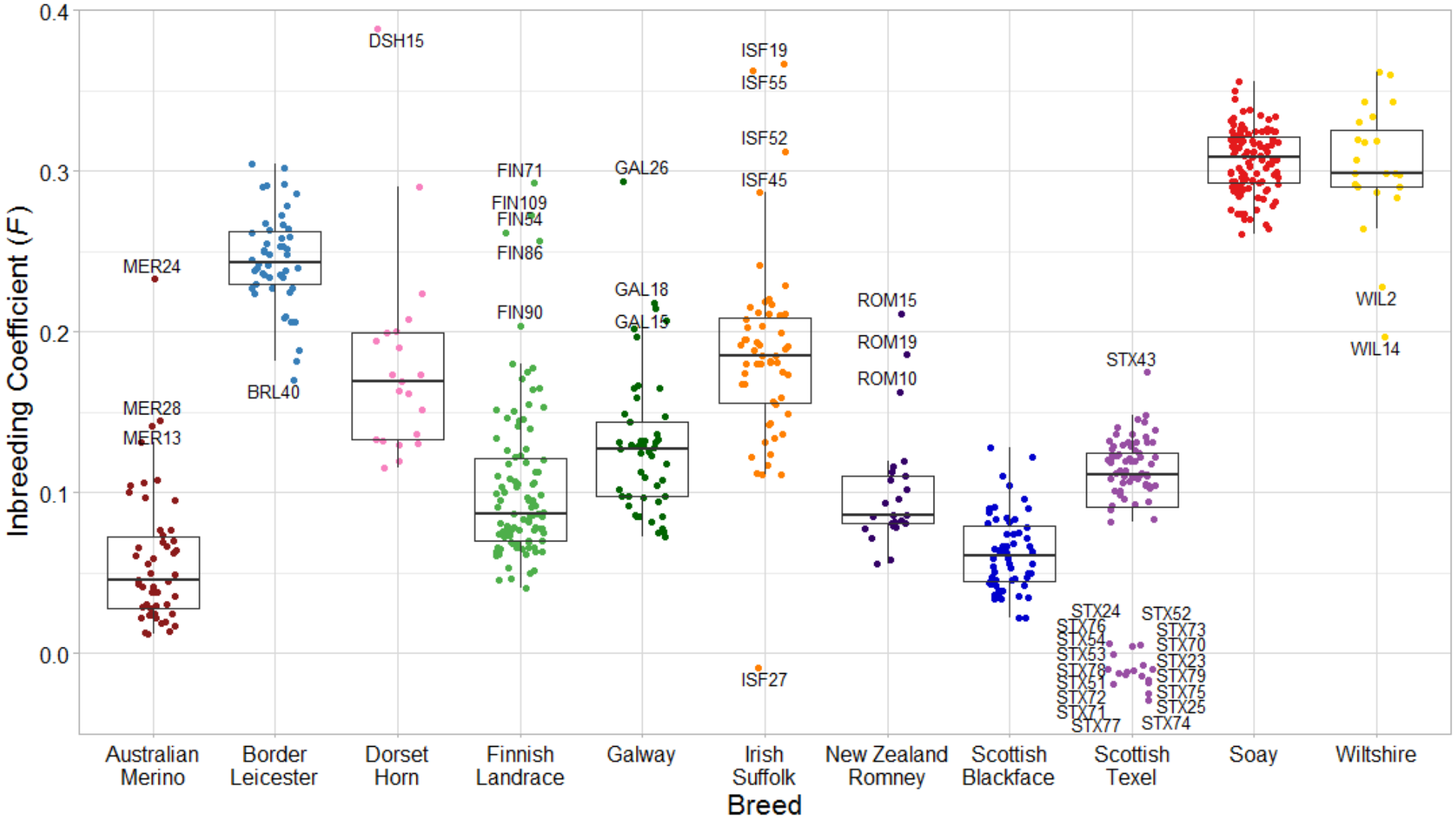

Figure 11. Jitter plots overlaid with box plots showing the genomic inbreeding coefficient estimate (*F*) for each individual sheep, arranged by breed. Individual outliers from each breed are labelled.

4.6 Runs of Homozygosity Results

The inbreeding coefficient calculated using ROH (F_{ROH}), shown in **Figure 12**, ranges from 0.000 to 0.339. F_{ROH} represents the proportion of each individual's genome covered by ROH. There was a significant difference between the breeds (H=343.71, df=10, P value < 0.001). The Scottish Blackface breed has the lowest mean F_{ROH} = 0.032 while the Wiltshire breed has the highest mean F_{ROH} = 0.251. The Galway breed falls between these two values with F_{ROH} = 0.061. There were significant differences between almost all of the breeds; however, the Galway breed had no significant difference with the Finnish Landrace, New Zealand Romney and Scottish Texel breeds (**Table S4**). Some individual outliers are more inbred than the rest of their breed, including Galway individuals GAL26, GAL15, GAL18, GAL16 and GAL36.

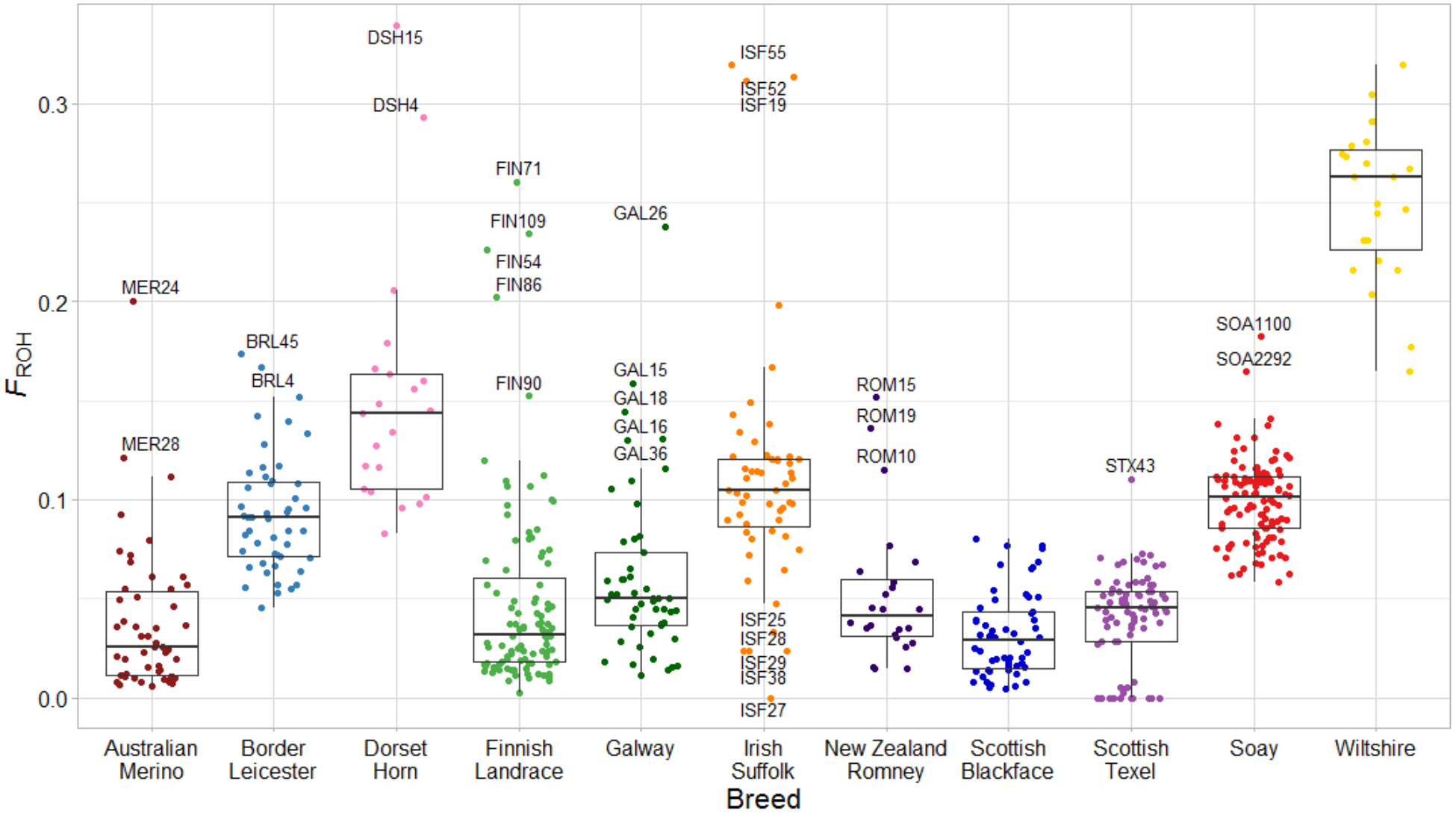

Figure 12. Jitter plot overlaid with box plots showing F_{ROH} for each individual sheep, arranged by breed. Individual outliers from each breed are labelled.

The mean sum of ROH in each breed for different length categories is shown in **Figure 13**. The proportion of the genome that is covered in ROH from the different length categories varies between the breeds; however, none of the breeds have large mean total length of ROH in the 1-5 Mb category. The Wiltshire breed has large mean total lengths in all the categories, apart from the first. Some breeds, such as the Australian Merino and Scottish Blackface, have smaller mean total lengths of ROH in all the categories. The Galway sheep fall between the extremes of the Wiltshire and Australian Merino breeds but is more similar to the Australian Merino. The Galway mean sum of ROH decreases with each category starting from the second, but increases again in the last 30 Mb+ category.

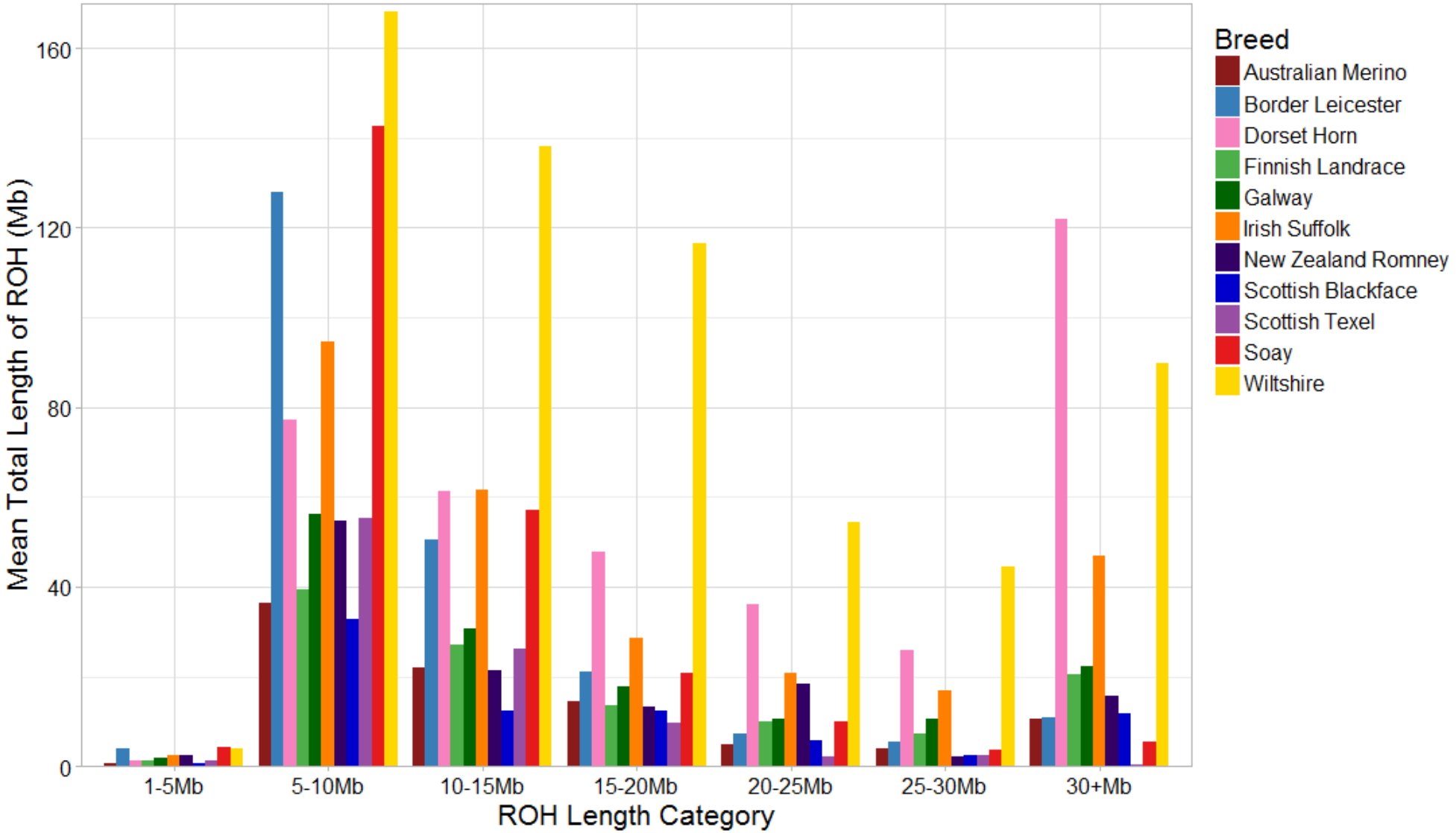

Figure 13. Bar graph showing the mean total length of runs of homozygosity for each breed in each length category in Mb.

4.6 SELECTION RESULTS

There are five significant peaks in the CSS results generated through collaboration with Dr Randhawa and shown in **Figure 14**; two on Chromosome 1 (OAR1), one on Chromosome 3 (OAR3) and two on chromosome 8 (OAR8) which are next to each other but merge into one peak in the graph.

The precise locations of the clusters of SNPs within the top 0.1% CSS score in these peaks, the number of SNPs and maximum CSS *P*-value in the clusters, number of genes in 1 Mb flanking the clusters upstream and downstream and IDs of Galway sheep with the flanking region included in ROH are provided in **Table 2**. The 197 candidate genes within these regions are listed in **Table S6**. The top five physiological system development and function pathways enriched for the panel of 119 genes with HGNC symbols in are listed in **Table 3**.

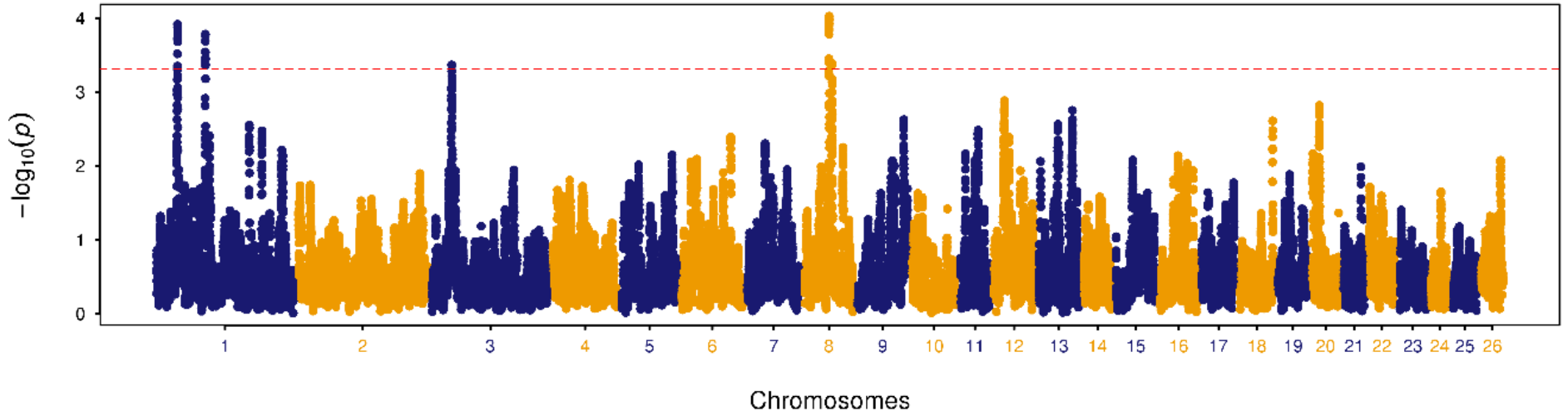

Figure 14. Smooth Manhattan plot showing -$\log_{10}$ *P*-values of Composite Selection Signal (CSS) scores for the 49 Galway sheep versus 50 randomly selected sheep from other breeds. The chromosomes are arranged in order along the x-axis and the red dashed line indicates the genome-wide 0.1% (upper cut-off) significance threshold.

Table 2. Table showing the regions of the clusters of SNPs with the top 0.1% CSS, number of SNPs and maximum CSS *P*-value in the clusters, number of genes in the 1 Mb flanking the clusters upstream and downstream and IDs of Galway sheep with the flanking region included in a ROH.

Cluster Region	Number of SNPs in Cluster	Maximum CSS in Cluster	Number of Genes in Flanking 1Mb	Galway Sheep with ROH in Flanking 1Mb
Chr 1:42.29-43.08Mb	14	3.917	23	GAL2, GAL15, GAL26, GAL49, GAL50
Chr 1:102.45-103.2Mb	14	3.785	104	GAL11, GAL13 GAL45
Chr 3:40.64-40.68Mb	2	3.367	13	GAL4, GAL12, GAL 19, GAL27, GAL35, GAL44
Chr 8:49.15-49.85Mb	16	4.027	30	GAL18, GAL30, GAL35, GAL36, GAL38, GAL45, GAL50
Chr 8:56.82-56.88Mb	2	3.390	26	GAL15, GAL18, GAL26, GAL30, GAL45, GAL46

Table 3. Table showing the top five physiological system development and function pathways enriched for the 119 candidate genes with HGNC symbols, number of genes and range of *P*-values for each pathway.

Pathway	Number of Genes	Range of *P*-values
Tissue Development	28	0.037-0.000
Haematological System Development and Function	26	0.037-0.000
Hair and Skin Development and Function	7	0.016-0.000
Immune Cell Trafficking	13	0.037-0.001
Connective Tissue Development and Function	15	0.037-0.001

5. DISCUSSION

5.1 F_{ST}

The Galway breed did not exhibit high pairwise F_{ST} values with every breed but also did not display low pairwise F_{ST} values, indicating their genetic distinctness from most of the other breeds. The lowest Galway pairwise F_{ST} values were with the Romney, Merino and Texel breeds. The Merino and Texel breeds have known connections with the Galway breed, while the Romney breed is another common commercial breed which may have been crossed with Galway sheep in the past (Curran 2010; DAD-IS 2016). Additionally, the Border Leicester breed did not have a particularly low pairwise F_{ST} value with the Galway breed despite its documented involvement in the formation of the Galway breed (Curran 2010). This could indicate that the Galway breed is now genetically distinct from this ancestral breed.

5.2 PRINCIPAL COMPONENTS ANALYSIS

The PCA shows that the Galway breed does separate as a distinct breed but only if three principal components are plotted (PC1-3) or if the Soay breed are removed from the analysis (see **Figures 4-6**). This is likely due to the genetic distinctiveness of the Soay breed, which is likely due to their geographic isolation and is demonstrated by the large proportion of the variation accounted for by PC1 in **Figure 4** (Patterson *et al.* 2006; Barrett 2012). However, differences between groups with large sample sizes can be 'up-weighted' in PCA (Patterson *et al.* 2006) and the Soay breed have the largest sample size in this data set (n = 110; see **Table 1**). The grouping of Galway, Border Leicester and Texel sheep is likely due to their shared ancestry and supports a previous study based on microsatellite genetic markers (Howard 2008). Some individual sheep samples do not cluster with the rest of their respective

breeds, which is likely due to crossbreeding (Patterson *et al.* 2006), or possible mislabelling of certain samples.

5.3 PHYLOGENETIC ANALYSIS

Phylogenetic graph analyses using TreeMix placed the Galway breed with the English sheep breeds (see **Figures 7** and **8**), which agrees with known information on the origins of the Galway breed (Curran 2010). The inclusion of the Texel breed in this branch is in agreement with previous results and the clustering of these breeds in the PCA (Kijas *et al.* 2012; Fariello *et al.* 2013); this is likely due to their shared ancestry with the Border Leicester breed (Howard 2008). The analysis also shows evidence of gene flow between the Irish Suffolk and Galway sheep, which is supported by documented historical evidence of crossbreeding between these populations (Curran 2010).

5.3 ADMIXTURE

The fastSTRUCTURE admixture analysis (see **Figure 9**) confirms the genetic distinctiveness of the Galway sheep population as a breed. This is because a model complexity or number of populations of $K = 11$ is required to maximise the marginal likelihood of the results. Since the number of breeds used in the analysis was 11 this shows that the breeds can be considered discrete populations, supporting the interpretation of sheep breeds as separate genetic units (Taberlet *et al.* 2008)

The pattern of admixture among Galway and Border Leicester, Scottish Texel and Irish Suffolk breeds is in concordance with their known breed histories and origins as well as both the PCA and TreeMix results, which is expected due to the known concordance of PCA and STRUCTURE analysis (Patterson *et al.* 2006; Howard 2008; Curran 2010; Raj *et al.* 2014). The lower levels of admixture observed with the Australian Merino, Finnish Landrace and

Scottish Blackface breeds is also supported by documented historical information on the Galway sheep breed (Quirke 1978b; Howard 2008; Curran 2010). The observed signature of Galway admixture within the New Zealand Romney breed was unexpected but supports the F_{ST} value for these breeds (see **Figure S2**). The observed signature of Wiltshire admixture within the Galway breed was also unexpected; however, this admixture is restricted to a small number of individual animals.

5.4 Effective Population Size

As expected, all the historical trends in genetic effective population size (N_e) show a decline towards the present (see **Figure 10**); however, the most recent and oldest estimates of N_e should be considered with caution (Barbato *et al.* 2015). The Galway breed exhibited a curve showing a trend in historical N_e between that of the commercial Merino and endangered Wiltshire breeds (DAD-IS 2016). This indicates that, while the Galway sheep does not have the demographic history of a large commercial sheep breed, it is still more genetically diverse than many landrace or localised sheep breeds. This is most likely a consequence of the widespread use of the Galway breed for lowland sheep production in Ireland up until the 1980s (see **Figure 2**).

5.5 Inbreeding

It is expected that genetic effective population size would affect the level of inbreeding experienced by a breed. As observed for the analysis of historical N_e trends, the Galway breed is between the extremes of the commercial Merino and isolated or endangered Soay and Wiltshire breeds (DAD-IS 2016). The mean genomic inbreeding coefficient estimate of $F = 0.129$ for the Galway breed is also different to the estimates using pedigree information for Galway lambs born in 1969 ($F = 0.019$) and 2012 ($F = 0.023$) (Martin 1975b;

McHugh *et al.* 2014). However, the level of inbreeding in Galway sheep was predicted to increase from the 1969 level if the population continued to decline, which it has (Martin 1975b). In addition, the 2012 survey recommended further collection of pedigree information to increase the accuracy (McHugh *et al.* 2014). It is also important to note that estimates of genomic inbreeding based on observed and expected homozygosity at genome-wide SNP genetic markers are likely to be higher and more realistic compared to estimates from pedigree data with a defined starting point of zero inbreeding (Kardos *et al.* 2015; Zhang *et al.* 2015; Wang 2016)

The individual animals that are outliers from their respective breeds can indicate more recent inbreeding or outbreeding. Sheep that are outliers based on these analyses are important to identify for conservation and management plans so that individual animals that are crossbred with other breeds or more inbred than the rest of the breed, such as Galway breed sample GAL26, are not prioritised for breeding programmes.

5.6 RUNS OF HOMOZYGOSITY

The results of the ROH analysis are in concordance with that of the genomic inbreeding and genetic effective population size analyses (see **Figure 12**). This is because individual animals that are part of breeds with larger genetic effective population sizes, such as the Australian Merino, are less likely to become inbred and therefore less likely to have large ROH segments in their genomes (Curik *et al.* 2014). The converse of this is true for breeds with lower N_e values and large ROH tracts in their genomes, such as the endangered Wiltshire breed (Curik *et al.* 2014). The Galway breed is between these extremes, indicating their intermediate genetic effective population size and level of inbreeding. However, the increase in mean total ROH length in the 30+ Mb category in the Galway breed could indicate more recent inbreeding due to their reduced population size (see **Figure 13**). This is

because ROH of more than 20 Mb can indicate recent inbreeding, while ROH of less than 20 Mb can indicate historical inbreeding (Purfield *et al.* 2012). None of the breeds have a high mean sum of ROH of less than 5 Mb because the density of the SNP data is too low to accurately detect ROH of this size (Purfield *et al.* 2012). The F_{ROH} analysis also confirms many of the outliers from the inbreeding and other analyses, including individual Galway animals, indicating that these should not be used in breeding programmes for conservation purposes.

5.6 SELECTION

The presence of selection peaks when genome-wide SNP data from the Galway animals is examined with data from the other breeds demonstrated that the population has experienced different artificial and natural selection processes during their unique history as a breed (see **Figure 14**). Five peaks of selection were detected, each of which was also located in ROH for at least three Galway animal samples (see **Table 2**). Of the 119 candidate genes under selection in the Galway breed with HGNC symbols, 28 are involved in tissue development and 15 are involved in connective tissue development and function (see **Table 3**). This is likely because the Galway breed is primarily used for meat production (DAD-IS 2016). Other sheep breeds farmed for meat, such as the Texel and Romney, possess specific mutations in genes involved in muscle development (Cockett *et al.* 2005; Clop *et al.* 2006; Wang *et al.* 2016). Seven genes are involved in hair and skin development and function, which may be explained by the high quality wool the Galway breed is also known for (Curran 2010). Selection of animals that are more resistant to disease, and therefore with better immune system function, is important in domestic animal populations, including many sheep breeds (González *et al.* 2012; Bishop 2015). The Galway breed, in particular, is considered to be less susceptible to infectious disease than the Roscommon breed, which may explain the

13 genes involved in immune cell trafficking (Curran 2010). The reason for putative selection of 26 genes involved in haematological system development and function is unknown.

5.7 CONCLUSIONS

The results of analyses presented in this thesis are mutually supportive and consistent. The F_{ST} and PCA results showed that the Galway breed is distinct from the other European sheep breeds, but more closely related to the Border Leicester and Texel breeds, which have documented shared ancestry with the Galway breed. The phylogenetic graph showed that the Galway breed is closely related to these breeds and that crossbreeding has taken place between the Galway and Irish Suffolk breeds, which was previously documented. The admixture analysis highlighted the genetic distinctiveness of the Galway breed and indicated that admixture has taken place between the Galway and Border Leicester, Scottish Texel and Irish Suffolk breeds.

The analysis of genetic effective population size trends demonstrated that the Galway breed has a smaller effective population size than commonly used commercial sheep breeds, such as the Australian Merino. However, it is larger than the endangered Wiltshire and isolated Soay breeds. The inbreeding analysis showed that the mean genomic inbreeding coefficient of the Galway breed was intermediate between that of the commercial and endangered breeds. This analysis also identified individual animals that were more inbred than the rest of their breed and should therefore, like those that are crossbred, not be used in conservation breeding programmes. The F_{ROH} analysis confirmed that many of these animals were inbred. The ROH analysis showed that the Galway breed falls between the commercial and endangered breeds, but that there is evidence of more recent inbreeding in the Galway breed that could increase if conservation programmes are not put in place. Genome-wide analysis of signature of selection analysis detected regions that may be under selection in the

Galway sheep when compared to the other breeds, highlighting their unique evolutionary history. These regions were located in ROH for individual Galway animals and contain genes that may be involved with meat and wool production and immune system function, which are known to be important in the Galway breed. These results support the usefulness of genome-wide marker data for conservation genomics in livestock populations, particularly for endangered local populations such as the Galway sheep breed.

6. ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor David MacHugh, for all his guidance and feedback. I would also like to thank Dr Imtiaz Randhawa for collaboration on the analysis of genome-wide selection signatures, Dr James Kijas and the International Sheep Genome Consortium for providing the SNP data, Carolina Correia for her assistance with some of the computer programs and Professor Kay Nolan for her feedback on the literature review. Finally, I would like to thank Dr Jon Yearsley, Professor Kay Nolan, John Leech, Lorenzo de Jonge, Charlotte Gilbert, Han Haige and, in particular, Sam Browett for their helpful discussions.

7. REFERENCES

Allendorf, F. W., P. A. Hohenlohe and G. Luikart, 2010 Genomics and the future of conservation genetics. Nat Rev Genet 11**:** 697-709.

Allendorf, F. W., G. Luikart and S. N. Aitken, 2013 *Conservation and the Genetics of Populations*. John Wiley & Sons, Oxford, UK.

Andersson, L., 2013 Molecular consequences of animal breeding. Curr Opin Genet Dev 23**:** 295-301.

Andersson, L., and M. Georges, 2004 Domestic-animal genomics: deciphering the genetics of complex traits. Nature Reviews Genetics 5**:** 202-212.

Archibald, A. L., N. E. Cockett, B. P. Dalrymple, T. Faraut, J. W. Kijas *et al.*, 2010 The sheep genome reference sequence: a work in progress. Anim Genet 41**:** 449-453.

Barbato, M., P. Orozco-terWengel, M. Tapio and M. W. Bruford, 2015 SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet 6**:** 109.

Barrett, R. D., 2012 Bad coat, ripped genes: cryptic selection on coat colour varies with ontogeny in Soay sheep. Mol Ecol 21**:** 2833-2835.

Becker, R. A., A. R. Wilks and R. Brownrigg, 2013a mapdata: Extra map databases. R package version**:** 2.2-2.

Becker, R. A., A. R. Wilks, R. Brownrigg and T. P. Minka, 2013b Maps: draw geographical maps. R package version 2.

Beynon, S. E., G. T. Slavov, M. Farre, B. Sunduimijid, K. Waddams *et al.*, 2015 Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genet 16**:** 65.

Binns, M. M., D. A. Boehler, E. Bailey, T. L. Lear, J. M. Cardwell *et al.*, 2012 Inbreeding in the Thoroughbred horse. Anim Genet 43**:** 340-342.

Biscarini, F., E. L. Nicolazzi, A. Stella, P. J. Boettcher and G. Gandini, 2015 Challenges and opportunities in genetic improvement of local livestock breeds. Front Genet 6**:** 33.

Bishop, S. C., 2015 Genetic resistance to infections in sheep. Vet Microbiol 181**:** 2-7.

Bowles, D., 2015 Recent advances in understanding the genetic resources of sheep breeds locally-adapted to the UK uplands: opportunities they offer for sustainable productivity. Front Genet 6**:** 24.

Byrne, T. J., P. R. Amer, P. F. Fennessy, A. R. Cromie, T. W. J. Keady *et al.*, 2010 Breeding objectives for sheep in Ireland: a bio-economic approach. Livestock Science 132**:** 135-144.

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell *et al.*, 2015 Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4**:** 7.

Ciani, E., P. Crepaldi, L. Nicoloso, E. Lasagna, F. M. Sarti *et al.*, 2014 Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. Anim Genet 45**:** 256-266.

Clop, A., F. Marcq, H. Takeda, D. Pirottin, X. Tordoir *et al.*, 2006 A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 38**:** 813-818.

Cockett, N. E., M. A. Smit, C. A. Bidwell, K. Segers, T. L. Hadfield *et al.*, 2005 The callipyge mutation and other genes that affect muscle hypertrophy in sheep. Genet Sel Evol 37 Suppl 1**:** S65-81.

CSO, 2016 StatBank database, pp. Central Statisitics Office.

Curik, I., M. Ferencakovic and J. Solkner, 2014 Inbreeding and runs of homozygosity: A possible solution to an old problem. Livestock Science 166**:** 26-34.

Curran, P. L., 2010 *The native lowland sheep of Galway and Roscommon: a history*. Patrick Leonard Curran, Tara, Co. Meath.

DAD-IS, 2016 Domestic Animal Diversity Information System (DADIS), pp. Food and Agriculture Organization of the United Nations.

Darwin, C., 1859 *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. J. Murray.

de Simoni Gouveia, J. J., M. V. da Silva, S. R. Paiva and S. M. de Oliveira, 2014 Identification of selection signatures in livestock species. Genet Mol Biol 37**:** 330-342.

Der Sarkissian, C., L. Ermini, M. Schubert, M. A. Yang, P. Librado *et al.*, 2015 Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. Curr Biol 25**:** 2577-2583.

Diamond, J., 2002 Evolution, consequences and future of plant and animal domestication. Nature 418**:** 700-707.

Fariello, M. I., S. Boitard, H. Naya, M. SanCristobal and B. Servin, 2013 Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. Genetics 193**:** 929-941.

Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17**:** 368-376.

Food and Agriculture Organization, 2015 *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*. edited by B.D. Scherf & D. Pilling. FAO Commission on Genetic Resources for Food and Agriculture Assessments., Rome, Italy.

Gandini, G. C., and E. Villa, 2003 Analysis of the cultural value of local livestock breeds: a methodology. J Anim Breed Genet 120**:** 1-11.

González, L., M. Jeffrey, M. P. Dagleish, W. Goldmann, S. Sisó *et al.*, 2012 Susceptibility to scrapie and disease phenotype in sheep: cross-PRNP genotype experimental transmissions with natural sources. Vet Res 43**:** 55.

Gunderson, K. L., F. J. Steemers, H. Ren, P. Ng, L. Zhou *et al.*, 2006 Whole-genome genotyping. Methods Enzymol 410**:** 359-376.

Haenlein, G., 2007 About the evolution of goat and sheep milk production. Small Ruminant Research 68**:** 3-6.

Hanrahan, J. P., 1976 Response to Selection for Litter Size in Galway Sheep. Irish Journal of Agricultural Research 15**:** 291-300.

Hanrahan, J. P., 1977 Sources of variation and repeatability of litter size in pedigree Galway sheep flocks. Irish Journal of Agricultural Research 16**:** 285-291.

Hanrahan, J. P., 2007 The Galway Breed - Origins and the Future, pp. Galway Sheep Breeders Association.

Hanrahan, J. P., and J. F. Quirke, 1975 Repeatability of the duration of oestrus and breed differences in the relationship between duration of oestrus and ovulation rate of sheep. J Reprod Fertil 45**:** 29-36.

Howard, D. J., 2008 PhD thesis: A Comparative Study of Molecular Genetic Variation in Three Irish Sheep Breeds; the Galway, Texel and Suffolk, pp. University College Dublin.

Jiang, Y., M. Xie, W. Chen, R. Talbot, J. F. Maddox *et al.*, 2014 The sheep genome illuminates biology of the rumen and lipid metabolism. Science 344**:** 1168-1173.

Kahle, D., and H. Wickham, 2013 ggmap: Spatial Visualization with ggplot2. R Journal 5.

Kardos, M., G. Luikart and F. W. Allendorf, 2015 Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. Heredity (Edinb) 115**:** 63-72.

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

Kijas, J. W., D. Townley, B. P. Dalrymple, M. P. Heaton, J. F. Maddox *et al.*, 2009 A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS ONE 4**:** e4668.

Kim, E. S., J. B. Cole, H. Huson, G. R. Wiggans, C. P. Van Tassell *et al.*, 2013 Effect of artificial selection on runs of homozygosity in U.S. Holstein cattle. PLoS ONE 8**:** e80813.

Lawson Handley, L. J., K. Byrne, F. Santucci, S. Townsend, M. Taylor *et al.*, 2007 Genetic structure of European sheep breeds. Heredity 99**:** 620-631.

Leroy, G., C. Danchin-Burge, I. Palhière, M. SanCristobal, Y. Nédélec *et al.*, 2015 How do introgression events shape the partitioning of diversity among breeds: a case study in sheep. Genetics Selection Evolution 47**:** 48.

Li, M. H., I. Stranden, T. Tiirikka, M. L. Sevon-Aimonen and J. Kantanen, 2011 A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. PLoS ONE 6**:** e26256.

Ligges, U., and M. Mächler, 2002 Scatterplot3d-an r package for visualizing multivariate data.

Martin, I., 1975a A genetic analysis of the Galway sheep breed: 1. some aspects of population dynamics of the pedigree and non-pedigree Galway sheep breed. Irish Journal of Agricultural Research**:** 245-253.

Martin, I., 1975b A genetic analysis of the Galway sheep breed: 3. level of inbreeding in the pedigree Galway sheep breed. Irish Journal of Agricultural Research**:** 269-274.

McHugh, N., D. Berry, S. McParland, E. Wall and T. Pabiou, 2014 Irish sheep breeding: current status and future plans, pp., edited by Teagasc and Sheep Ireland.

Meadows, J. R. S., 2014 Sheep: Domestication, pp. 6597-6600 in *Encyclopedia of Global Archaeology*. Springer.

Meszaros, G., S. A. Boison, A. M. Perez O'Brien, M. Ferencakovic, I. Curik *et al.*, 2015 Genomic analysis for managing small and endangered populations: a case study in Tyrol Grey cattle. Front Genet 6**:** 173.

More O'Ferrall, G. J., 1976 Phenotypic and genetic parameters of productivity in Galway ewes. Animal Science 23**:** 295-304.

Murphy, G., C. O'Donoghue, S. Hynes and E. Murphy, 2014 Modelling the participation decision in agri-environmental schemes, pp. in *2014 International Congress, August 26-29, 2014, Ljubljana, Slovenia*. European Association of Agricultural Economists.

Patterson, N., A. L. Price and D. Reich, 2006 Population structure and eigenanalysis. PLoS Genet 2**:** e190.

Pemberton, T. J., D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg *et al.*, 2012 Genomic patterns of homozygosity in worldwide human populations. Am J Hum Genet 91**:** 275-292.

Pickrell, J. K., and J. K. Pritchard, 2012 Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8**:** e1002967.

Pritchard, J. K., M. Stephens and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics 155**:** 945-959.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81**:** 559-575.

Purfield, D. C., D. P. Berry, S. McParland and D. G. Bradley, 2012 Runs of homozygosity and population history in cattle. BMC Genet 13**:** 70.

Quirke, J. F., 1978a Onset of puberty and oestrous activity in Galway, Finnish Landrace and Finn-Cross ewe lambs during their first breeding season. Irish Journal of Agricultural Research 17**:** 15-23.

Quirke, J. F., 1978b Reproductive performance of Galway, Finnish Landrace and Finn-Cross ewe lambs. Irish Journal of Agricultural Research 17**:** 25-32.

R Core Team, 2015 R: A language and environment for statistical computing, pp. R Foundation for Statistical Computing,, Vienna, Austria.

Raj, A., M. Stephens and J. K. Pritchard, 2014 fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197**:** 573-589.

Randhawa, I. A., M. S. Khatkar, P. C. Thomson and H. W. Raadsma, 2014 Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BMC Genet 15**:** 34.

Reich, D., A. L. Price and N. Patterson, 2008 Principal component analysis of genetic data. Nat Genet 40**:** 491-492.

Renwick, A., 2013 The Importance of the Cattle and Sheep Sectors to the Irish Economy, pp. The Irish Farmers' Association, University College Dublin, Ireland.

Rosenberg, N. A., 2004 DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4**:** 137-138.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter *et al.*, 2007 Genome-wide detection and characterization of positive selection in human populations. Nature 449**:** 913-918.

Smedley, D., S. Haider, S. Durinck, L. Pandini, P. Provero *et al.*, 2015 The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res 43**:** W589-598.

Taberlet, P., A. Valentini, H. Rezaei, S. Naderi, F. Pompanon *et al.*, 2008 Are cattle, sheep, and goats endangered species? Mol Ecol 17**:** 275-284.

Tapio, M., I. Tapio, Z. Grislis, L. E. Holm, S. Jeppsson *et al.*, 2005 Native breeds demonstrate high contributions to the molecular variation in northern European sheep. Mol Ecol 14**:** 3951-3963.

Teetor, P., 2011 *R Cookbook*. O'Reilly Media.

Wang, J., 2016 Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? Theor Popul Biol 107**:** 4-13.

Wang, J., H. Zhou, J. Hu, S. Li, Y. Luo *et al.*, 2016 Two single nucleotide polymorphisms in the promoter of the ovine myostatin gene (MSTN) and their effect on growth and carcass muscle traits in New Zealand Romney sheep. J Anim Breed Genet 133**:** 219-226.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-statistics for the analysis of population structure. Evolution**:** 1358-1370.

Wickham, H., 2009 *ggplot2 : elegant graphics for data analysis*. Springer, New York.

Zhang, Q., M. P. Calus, B. Guldbrandtsen, M. S. Lund and G. Sahana, 2015 Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. BMC Genet 16**:** 88.

8. APPENDIX

Table S1. Table showing breed name, origin and sample size of the breeds used for the European phylogenetic tree in **Figure 8**.

Breed Name	Sample Size	Breed Origin
Altamurana	24	Italy
Australian Merino	50	Spain/Unknown
Border Leicester	48	England
Castellana	23	Spain
Comisana	24	Sicily
Dorset Horn	21	England
East-Friesian Brown	39	Germany
Finnish Landrace	99	Finland
Galway	49	Ireland
German Texel	46	Netherlands
Indian Garole	26	India
Irish Suffolk	55	England
Merinolandschaf	24	Germany
Milk Lacaune	103	France
New Zealand Romney	24	England
Scottish Blackface	56	Scotland
Scottish Texel	80	Netherlands
Soay	110	Scotland
White Spael	32	Norway
Swiss White Alpine	24	Switzerland
Valais Blacknose	24	Switzerland
Wiltshire	23	England
Total	1004	

Table S2. Table showing pairwise weighted F_{ST} results calculated using PLINK 1.9 (Chang *et al.* 2015).

	Australian Merino	Border Leicester	Dorset Horn	Finnish Landrace	Galway	Irish Suffolk	New Zealand Romney	Scottish Blackface	Scottish Texel	Soay
Border Leicester	0.174									
Dorset Horn	0.160	0.264								
Finnish Landrace	0.084	0.182	0.177							
Galway	0.118	0.132	0.202	0.129						
Irish Suffolk	0.115	0.206	0.206	0.127	0.148					
New Zealand Romney	0.096	0.167	0.183	0.110	0.110	0.133				
Scottish Blackface	0.080	0.179	0.169	0.091	0.121	0.118	0.100			
Scottish Texel	0.128	0.169	0.214	0.131	0.119	0.161	0.132	0.128		
Soay	0.208	0.302	0.302	0.203	0.250	0.247	0.236	0.204	0.251	
Wiltshire	0.188	0.261	0.286	0.199	0.207	0.229	0.200	0.193	0.222	0.326

Table S3. Table showing pairwise Bonferroni-corrected *P*-values for Wilcox tests on genetic effective population size results (N_e) calculated using R (R Core Team 2015). Significant values are in bold font.

	Australian Merino	Border Leicester	Dorset Horn	Finnish Landrace	Galway	Irish Suffolk	New Zealand Romney	Scottish Blackface	Scottish Texel	Soay
Border Leicester	**0.000**									
Dorset Horn	**0.000**	1.000								
Finnish Landrace	1.000	**0.001**	**0.000**							
Galway	**0.006**	1.000	0.916	0.067						
Irish Suffolk	**0.003**	1.000	1.000	**0.042**	1.000					
New Zealand Romney	0.092	**0.036**	**0.042**	1.000	1.000	1.000				
Scottish Blackface	1.000	**0.011**	**0.006**	1.000	1.000	0.478	1.000			
Scottish Texel	**0.002**	1.000	1.000	**0.036**	1.000	1.000	1.000	0.531		
Soay	**0.000**	1.000	1.000	**0.000**	0.516	1.000	**0.015**	**0.003**	1.000	
Wiltshire	**0.000**	1.000	1.000	**0.000**	0.230	0.477	**0.007**	**0.003**	0.490	1.000

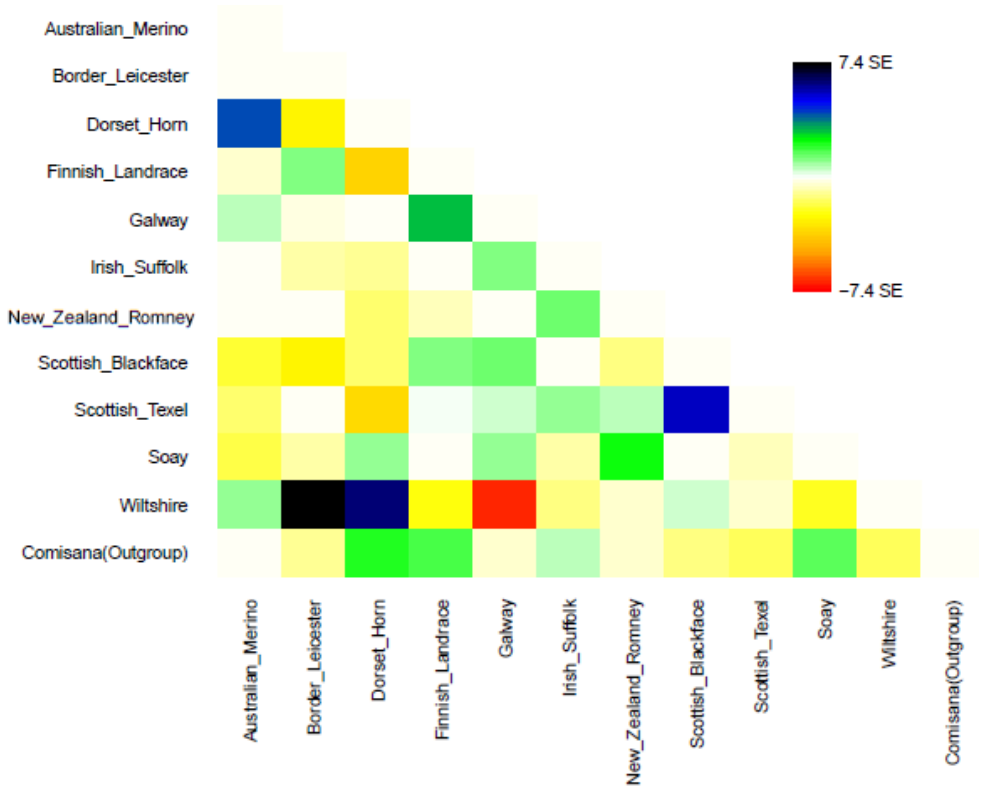

Figure S1. Graph showing the residuals for the phylogenetic tree in **Figure 7**.

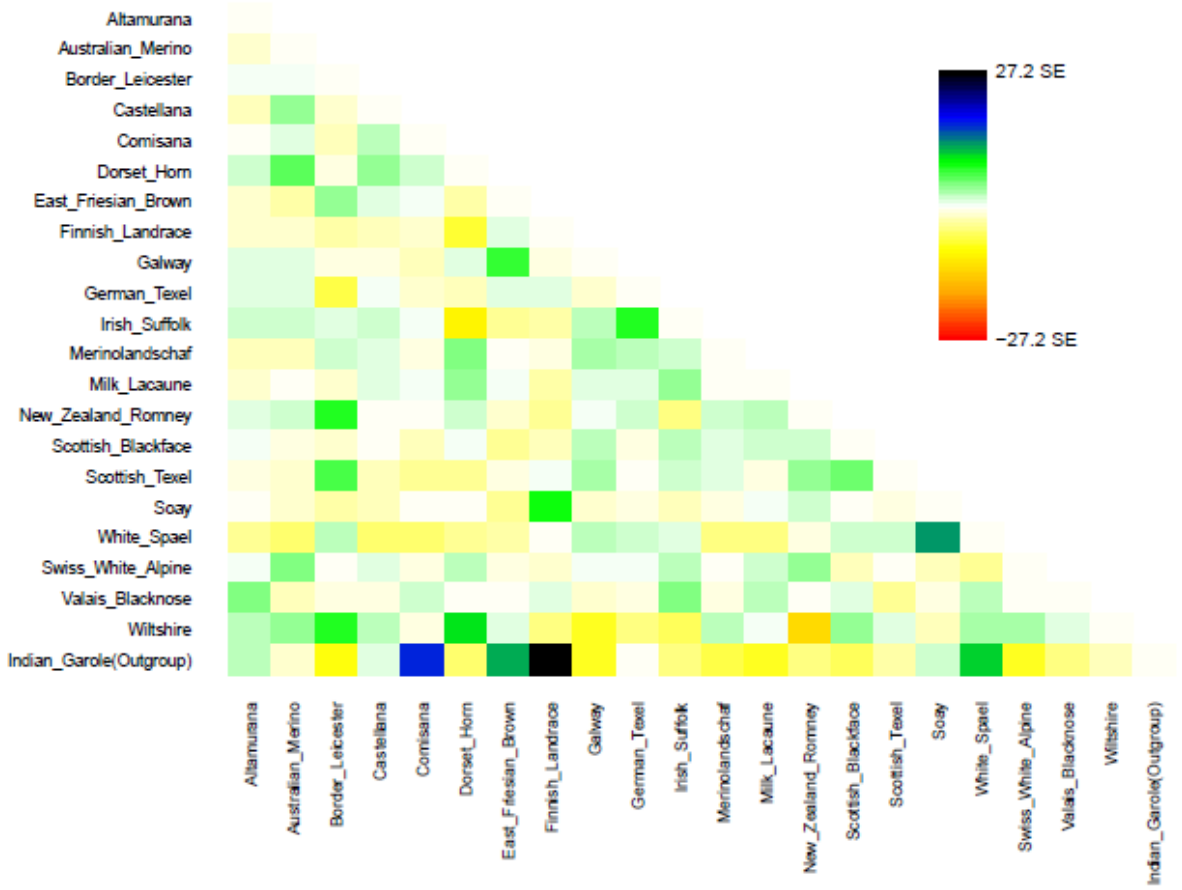

Figure S2. Graph showing the residuals for the European phylogenetic tree in **Figure 8**.

Table S4. Table showing pairwise Bonferroni-corrected *P*-values for Wilcox tests on genomic inbreeding coefficient (*F*) results calculated using R (R Core Team 2015). Significant values are in bold font.

	Australian Merino	Border Leicester	Dorset Horn	Finnish Landrace	Galway	Irish Suffolk	New Zealand Romney	Scottish Blackface	Scottish Texel	Soay
Border Leicester	**0.000**									
Dorset Horn	**0.000**	**0.000**								
Finnish Landrace	**0.000**	**0.000**	**0.000**							
Galway	**0.000**	**0.000**	**0.003**	**0.002**						
Irish Suffolk	**0.000**	**0.000**	1.000	**0.000**	**0.000**					
New Zealand Romney	**0.000**	**0.000**	**0.000**	1.000	0.058	**0.000**				
Scottish Blackface	1.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**			
Scottish Texel	**0.003**	**0.000**	**0.000**	1.000	0.298	**0.000**	1.000	**0.000**		
Soay	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	
Wiltshire	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	1.000

Table S5. Table showing pairwise Bonferroni-corrected p-values for Wilcox tests on F_{ROH} inbreeding coefficient results calculated using R (R Core Team 2015). Significant values are in bold font.

	Australian Merino	Border Leicester	Dorset Horn	Finnish Landrace	Galway	Irish Suffolk	New Zealand Romney	Scottish Blackface	Scottish Texel	Soay
Border Leicester	**0.000**									
Dorset Horn	**0.000**	**0.000**								
Finnish Landrace	1.000	**0.000**	**0.000**							
Galway	**0.018**	**0.000**	**0.000**	0.155						
Irish Suffolk	**0.000**	1.000	0.087	**0.000**	**0.000**					
New Zealand Romney	1.000	**0.000**	**0.000**	1.000	1.000	**0.000**				
Scottish Blackface	1.000	**0.000**	**0.000**	1.000	**0.001**	**0.000**	0.389			
Scottish Texel	1.000	**0.000**	**0.000**	1.000	1.000	**0.000**	1.000	1.000		
Soay	**0.000**	1.000	**0.000**	**0.000**	**0.000**	1.000	**0.000**	**0.000**	**0.000**	
Wiltshire	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Table S6. Table showing the selection peak cluster region, Ensembl ID, HGNC symbol and description of the candidate genes under selection in the Galway breed.

Cluster Region	Ensembl ID	Gene Symbol	Description
Chr 1:42.29-43.08	ENSOARG00000010091	-	Uncharacterized protein
Chr 1:42.29-43.08	ENSOARG00000010122	*DIRAS3*	DIRAS family GTPase 3
Chr 1:42.29-43.08	ENSOARG00000010521	*PDE4B*	phosphodiesterase 4B
Chr 1:42.29-43.08	ENSOARG00000010617	*SGIP1*	SH3 domain GRB2 like endophilin interacting protein 1
Chr 1:42.29-43.08	ENSOARG00000010702	*TCTEX1D1*	Tctex1 domain containing 1
Chr 1:42.29-43.08	ENSOARG00000010717	*INSL5*	insulin like 5
Chr 1:42.29-43.08	ENSOARG00000010792	*WDR78*	WD repeat domain 78
Chr 1:42.29-43.08	ENSOARG00000010903	*MIER1*	MIER1 transcriptional regulator
Chr 1:42.29-43.08	ENSOARG00000010986	*SLC35D1*	solute carrier family 35 member D1
Chr 1:42.29-43.08	ENSOARG00000011004	*C1orf141*	chromosome 1 open reading frame 141
Chr 1:42.29-43.08	ENSOARG00000011083	*IL23R*	interleukin 23 receptor
Chr 1:42.29-43.08	ENSOARG00000011173	*IL12RB2*	interleukin 12 receptor subunit beta 2
Chr 1:42.29-43.08	ENSOARG00000011282	*SERBP1*	SERPINE1 mRNA binding protein 1
Chr 1:42.29-43.08	ENSOARG00000011356	*GADD45A*	growth arrest and DNA damage inducible alpha
Chr 1:42.29-43.08	ENSOARG00000011369	*GNG12*	G protein subunit gamma 12
Chr 1:42.29-43.08	ENSOARG00000011383	*WLS*	wntless Wnt ligand secretion mediator
Chr 1:42.29-43.08	ENSOARG00000011470	*RPE65*	retinal pigment epithelium-specific protein 65kDa
Chr 1:42.29-43.08	ENSOARG00000011569	*DEPDC1*	DEP domain containing 1
Chr 1:42.29-43.08	ENSOARG00000021718	-	Small nucleolar RNA SNORA70
Chr 1:42.29-43.08	ENSOARG00000023397	-	U6 spliceosomal RNA
Chr 1:42.29-43.08	ENSOARG00000024412	-	-
Chr 1:42.29-43.08	ENSOARG00000025521	-	-
Chr 1:42.29-43.08	ENSOARG00000025522	-	-
Chr 1:102.45-103.2	ENSOARG00000000259	*C1orf68*	chromosome 1 open reading frame 68

Chr 1:102.45-103.2	ENSOARG00000000272	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000287	*KPRP*	keratinocyte proline rich protein
Chr 1:102.45-103.2	ENSOARG00000000297	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000314	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000330	-	-
Chr 1:102.45-103.2	ENSOARG00000000342	*SPRR3*	small proline rich protein 3
Chr 1:102.45-103.2	ENSOARG00000000351	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000368	-	-
Chr 1:102.45-103.2	ENSOARG00000000377	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000389	*PGLYRP3*	peptidoglycan recognition protein 3
Chr 1:102.45-103.2	ENSOARG00000000406	*PGLYRP4*	peptidoglycan recognition protein 4
Chr 1:102.45-103.2	ENSOARG00000000417	*S100A9*	S100 calcium binding protein A9
Chr 1:102.45-103.2	ENSOARG00000000432	*S100A12*	S100 calcium binding protein A12
Chr 1:102.45-103.2	ENSOARG00000000450	*S100A8*	S100 calcium binding protein A8
Chr 1:102.45-103.2	ENSOARG00000000463	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000477	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000000593	-	Protein S100
Chr 1:102.45-103.2	ENSOARG00000000610	*S100A6*	S100 calcium binding protein A6
Chr 1:102.45-103.2	ENSOARG00000000623	*S100A5*	S100 calcium binding protein A5
Chr 1:102.45-103.2	ENSOARG00000000641	*S100A4*	S100 calcium binding protein A4
Chr 1:102.45-103.2	ENSOARG00000000657	*S100A3*	S100 calcium binding protein A3
Chr 1:102.45-103.2	ENSOARG00000000674	*S100A2*	S100 calcium binding protein A2
Chr 1:102.45-103.2	ENSOARG00000000691	*S100A16*	S100 calcium binding protein A16
Chr 1:102.45-103.2	ENSOARG00000000707	*S100A1*	S100 calcium binding protein A1
Chr 1:102.45-103.2	ENSOARG00000000735	*CHTOP*	chromatin target of PRMT1
Chr 1:102.45-103.2	ENSOARG00000000758	*SNAPIN*	SNAP associated protein
Chr 1:102.45-103.2	ENSOARG00000000787	*ILF2*	interleukin enhancer binding factor 2

Chr 1:102.45-103.2	ENSOARG00000000858	*NPR1*	natriuretic peptide receptor 1
Chr 1:102.45-103.2	ENSOARG00000001095	*INTS3*	integrator complex subunit 3
Chr 1:102.45-103.2	ENSOARG00000001265	*SLC27A3*	solute carrier family 27 member 3
Chr 1:102.45-103.2	ENSOARG00000001449	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000001486	*GATAD2B*	GATA zinc finger domain containing 2B
Chr 1:102.45-103.2	ENSOARG00000001584	*DENND4B*	DENN domain containing 4B
Chr 1:102.45-103.2	ENSOARG00000001610	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000001630	*SLC39A1*	solute carrier family 39 member 1
Chr 1:102.45-103.2	ENSOARG00000001743	*CREB3L4*	cAMP responsive element binding protein 3 like 4
Chr 1:102.45-103.2	ENSOARG00000001762	*JTB*	jumping translocation breakpoint
Chr 1:102.45-103.2	ENSOARG00000001777	*RAB13*	RAB13, member RAS oncogene family
Chr 1:102.45-103.2	ENSOARG00000001813	*NUP210L*	nucleoporin 210 like
Chr 1:102.45-103.2	ENSOARG00000001845	*TPM3*	tropomyosin 3
Chr 1:102.45-103.2	ENSOARG00000001925	*C1orf189*	chromosome 1 open reading frame 189
Chr 1:102.45-103.2	ENSOARG00000002028	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000002170	*UBAP2L*	ubiquitin associated protein 2 like
Chr 1:102.45-103.2	ENSOARG00000002271	*HAX1*	HCLS1 associated protein X-1
Chr 1:102.45-103.2	ENSOARG00000002349	*AQP10*	aquaporin 10
Chr 1:102.45-103.2	ENSOARG00000002448	*ATP8B2*	ATPase phospholipid transporting 8B2
Chr 1:102.45-103.2	ENSOARG00000002562	*IL6R*	interleukin 6 receptor
Chr 1:102.45-103.2	ENSOARG00000002674	*SHE*	Src homology 2 domain containing E
Chr 1:102.45-103.2	ENSOARG00000002772	*TDRD10*	tudor domain containing 10
Chr 1:102.45-103.2	ENSOARG00000002800	*UBE2Q1*	ubiquitin conjugating enzyme E2 Q1
Chr 1:102.45-103.2	ENSOARG00000002921	*CHRNB2*	cholinergic receptor nicotinic beta 2 subunit
Chr 1:102.45-103.2	ENSOARG00000002952	*ADAR*	adenosine deaminase, RNA specific
Chr 1:102.45-103.2	ENSOARG00000002973	*KCNN3*	potassium calcium-activated channel subfamily N member 3

Chr 1:102.45-103.2	ENSOARG00000003056	*PMVK*	phosphomevalonate kinase
Chr 1:102.45-103.2	ENSOARG00000003126	*PBXIP1*	PBX homeobox interacting protein 1
Chr 1:102.45-103.2	ENSOARG00000003199	*SHC1*	SHC adaptor protein 1
Chr 1:102.45-103.2	ENSOARG00000003292	*CKS1B*	CDC28 protein kinase regulatory subunit 1B
Chr 1:102.45-103.2	ENSOARG00000003310	*FLAD1*	flavin adenine dinucleotide synthetase 1
Chr 1:102.45-103.2	ENSOARG00000003333	*ZBTB7B*	zinc finger and BTB domain containing 7B
Chr 1:102.45-103.2	ENSOARG00000003355	*DCST2*	DC-STAMP domain containing 2
Chr 1:102.45-103.2	ENSOARG00000003448	*DCST1*	DC-STAMP domain containing 1
Chr 1:102.45-103.2	ENSOARG00000003543	*ADAM15*	ADAM metallopeptidase domain 15
Chr 1:102.45-103.2	ENSOARG00000003565	*EFNA4*	ephrin A4
Chr 1:102.45-103.2	ENSOARG00000003579	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000003647	*EFNA1*	ephrin A1
Chr 1:102.45-103.2	ENSOARG00000003734	*SLC50A1*	solute carrier family 50 member 1
Chr 1:102.45-103.2	ENSOARG00000003747	*KRTCAP2*	keratinocyte associated protein 2
Chr 1:102.45-103.2	ENSOARG00000003769	*TRIM46*	tripartite motif containing 46
Chr 1:102.45-103.2	ENSOARG00000003788	*MUC1*	mucin 1, cell surface associated
Chr 1:102.45-103.2	ENSOARG00000003894	*THBS3*	thrombospondin 3
Chr 1:102.45-103.2	ENSOARG00000003912	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000003937	*GBA*	glucosylceramidase beta
Chr 1:102.45-103.2	ENSOARG00000003954	*FAM189B*	family with sequence similarity 189 member B
Chr 1:102.45-103.2	ENSOARG00000004032	*SCAMP3*	secretory carrier membrane protein 3
Chr 1:102.45-103.2	ENSOARG00000004102	*CLK2*	CDC like kinase 2
Chr 1:102.45-103.2	ENSOARG00000004200	*HCN3*	hyperpolarization activated cyclic nucleotide gated potassium channel 3
Chr 1:102.45-103.2	ENSOARG00000004229	*PKLR*	pyruvate kinase, liver and RBC
Chr 1:102.45-103.2	ENSOARG00000004341	-	-
Chr 1:102.45-103.2	ENSOARG00000004448	*ASH1L*	ASH1 like histone lysine methyltransferase

Chr 1:102.45-103.2	ENSOARG00000006463	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000006502	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000006529	-	-
Chr 1:102.45-103.2	ENSOARG00000006552	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000006578	*PRR9*	proline rich 9
Chr 1:102.45-103.2	ENSOARG00000006605	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000006634	*LENEP*	lens epithelial protein
Chr 1:102.45-103.2	ENSOARG00000006665	*DPM3*	dolichyl-phosphate mannosyltransferase subunit 3
Chr 1:102.45-103.2	ENSOARG00000022891	-	U6 spliceosomal RNA
Chr 1:102.45-103.2	ENSOARG00000023246	-	Small nucleolar RNA SNORA51
Chr 1:102.45-103.2	ENSOARG00000023446	-	Small nucleolar RNA SNORA58
Chr 1:102.45-103.2	ENSOARG00000024783	-	U6 spliceosomal RNA
Chr 1:102.45-103.2	ENSOARG00000024820	-	-
Chr 1:102.45-103.2	ENSOARG00000025019	-	Small nucleolar RNA SNORA58
Chr 1:102.45-103.2	ENSOARG00000025190	-	small proline-rich protein type II
Chr 1:102.45-103.2	ENSOARG00000025191	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000025193	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000025195	-	Uncharacterized protein
Chr 1:102.45-103.2	ENSOARG00000025551	-	-
Chr 1:102.45-103.2	ENSOARG00000025552	-	-
Chr 1:102.45-103.2	ENSOARG00000025553	-	-
Chr 1:102.45-103.2	ENSOARG00000025554	-	-
Chr 1:102.45-103.2	ENSOARG00000025555	-	-
Chr 1:102.45-103.2	ENSOARG00000025556	-	-
Chr 3:40.64-40.68	ENSOARG00000019897	*APLF*	aprataxin and PNKP like factor
Chr 3:40.64-40.68	ENSOARG00000019899	*FBXO48*	F-box protein 48
Chr 3:40.64-40.68	ENSOARG00000019902	*PLEK*	pleckstrin
Chr 3:40.64-40.68	ENSOARG00000019907	*CNRIP1*	cannabinoid receptor interacting protein 1

Chr 3:40.64-40.68	ENSOARG00000019910	-	Uncharacterized protein
Chr 3:40.64-40.68	ENSOARG00000019914	*PNO1*	partner of NOB1 homolog
Chr 3:40.64-40.68	ENSOARG00000019929	-	Uncharacterized protein
Chr 3:40.64-40.68	ENSOARG00000019931	*C1D*	C1D nuclear receptor corepressor
Chr 3:40.64-40.68	ENSOARG00000019948	*ETAA1*	Ewing tumor associated antigen 1
Chr 3:40.64-40.68	ENSOARG00000019962	*MEIS1*	Meis homeobox 1
Chr 3:40.64-40.68	ENSOARG00000024135	-	-
Chr 3:40.64-40.68	ENSOARG00000025922	-	-
Chr 3:40.64-40.68	ENSOARG00000025970	-	-
Chr 8:49.15-49.85	ENSOARG00000012863	*SRSF12*	serine and arginine rich splicing factor 12
Chr 8:49.15-49.85	ENSOARG00000012871	*PNRC1*	proline rich nuclear receptor coactivator 1
Chr 8:49.15-49.85	ENSOARG00000012911	*RNGTT*	RNA guanylyltransferase and 5'-phosphatase
Chr 8:49.15-49.85	ENSOARG00000012955	*SPACA1*	sperm acrosome associated 1
Chr 8:49.15-49.85	ENSOARG00000012963	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000012967	*AKIRIN2*	akirin 2
Chr 8:49.15-49.85	ENSOARG00000012992	*ORC3*	origin recognition complex subunit 3
Chr 8:49.15-49.85	ENSOARG00000013003	*RARS2*	arginyl-tRNA synthetase 2, mitochondrial
Chr 8:49.15-49.85	ENSOARG00000013049	*SLC35A1*	solute carrier family 35 member A1
Chr 8:49.15-49.85	ENSOARG00000013071	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000013104	*C6orf163*	chromosome 6 open reading frame 163
Chr 8:49.15-49.85	ENSOARG00000013115	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000013132	*ZNF292*	zinc finger protein 292
Chr 8:49.15-49.85	ENSOARG00000013153	*CGA*	glycoprotein hormones, alpha polypeptide
Chr 8:49.15-49.85	ENSOARG00000013156	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000019995	*CNR1*	cannabinoid receptor 1 (brain)
Chr 8:49.15-49.85	ENSOARG00000020002	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000020007	*HTR1E*	5-hydroxytryptamine receptor 1E

Chr 8:49.15-49.85	ENSOARG00000020008	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000020012	-	Uncharacterized protein
Chr 8:49.15-49.85	ENSOARG00000021301	-	Small nucleolar RNA SNORA70
Chr 8:49.15-49.85	ENSOARG00000021683	-	U6 spliceosomal RNA
Chr 8:49.15-49.85	ENSOARG00000023346	-	5S ribosomal RNA
Chr 8:49.15-49.85	ENSOARG00000024253	-	-
Chr 8:49.15-49.85	ENSOARG00000024586	-	Small nucleolar RNA SNORD22
Chr 8:49.15-49.85	ENSOARG00000024732	-	U6 spliceosomal RNA
Chr 8:49.15-49.85	ENSOARG00000027025	-	-
Chr 8:49.15-49.85	ENSOARG00000027026	-	-
Chr 8:49.15-49.85	ENSOARG00000027027	-	-
Chr 8:49.15-49.85	ENSOARG00000027028	-	-