The Genetics of *Mycobacterium bovis* infection in Irish Cattle

Ian William Richardson

A thesis submitted to the Dublin University, Trinity College for the degree of Doctor of Philosophy

Department of Genetics

Dublin University

Trinity College

June 2016

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed by Author_____

Ian William Richardson

June 2016

In memoriam

This thesis is dedicated to the memory of Dr. W T Irwin, my grandfather and a Trinity graduate, who passed away shortly after the completion of this thesis.

Thank you Papa incepting a love and curiosity for science in me.

Acknowledgements

Completing a PhD is a harrowing but rewarding process; though it appears on the surface to be the work of one person, it is ultimately the support and guidance of everyone around the student that mold the thesis. As such there are more than a few people to thank for their selflessly given help in producing the work presented here. Firstly and most importantly I'd like to thank my supervisors Prof. Dan Bradley and Dr Donagh Berry for their immense patience and guidance; having come from a biochemistry background, to be thrust into quantitative and molecular genetics would have been a much more tumultuous experience without their guiding hand. Moreover they provided me with the skills needed to form a sound basis for critical thinking and communication of my work. On top of all of this they supported my athletic career throughout my studies. For this I am more than grateful.

Thanks must be given to Science foundation Ireland for funding the work in this thesis as well as the Irish Cattle Breeding Federations, the Department of Agriculture, Food and the Marine, the Centre for Veterinary Epidemiology and Risk Analysis (in particular Simon More and Isabella Higgins) and Teagasc for providing the data used in this thesis.

To all my colleagues (Eppie, Rui, Lara, Kevin, Valeria, Marta, Victoria) thank you for the years of friendship, support and advice, in particular Russell and Matthew who were always available and on hand to help when I felt at a dead end, despite having a mountain of their own work to deal with. To my decade long friends, Dogan, Phil and Suzy, thank you for always being there even when I had not been in contact for weeks at a time.

Finally to my parents, words cannot describe my gratitude; you've been there for me in my darkest times and never lost hope in me. You've constantly believed in my abilities

both academically and athletically. For all of this and supporting me through my decade in University I am eternally grateful.

Contents

Figures

Tables

Summary

Infection of livestock with *Mycobacterium bovis*, is the causing agent of bovine tuberculosis and is of major economic and health concern globally, with the cost of livestock loss estimated at €2 billion globally. In developed nations the majority of the cost involved with bovine tuberculosis is due to controlling the disease, while in developing nations, infection is endemic leading to reduced productivity and loss of livestock. Zoonotic infection of bovine tuberculosis also constitutes a major human health risk in developing nations. In Ireland approximately 15,000-20,000 cattle are infected with bovine tuberculosis each year, with control of bovine tuberculosis in Ireland incurring a cost of €63 million to the Irish government. While an extensive eradication program has been in place in Ireland since the 1960s, eradication of the disease from the island has remained elusive. This is in part due to the presence of a natural wildlife reservoir for the disease in the form of Eurasian badgers (*Melese meles*). Ireland has had a successful nation-wide breeding strategy, carried out by the Irish Cattle breeding federation, in place for cattle since the early 2000s. This breeding strategy has been based around developing accurate genetic evaluations for all breeding sires for economically important traits (i.e. milk yield, fertility etc.), using both progeny records and genotype information.

This thesis delves into the underlying genetics of *Mycobacterium bovis* infection in Irish dairy and beef cattle, utilizing a wealth of phenotypic, genetic and pedigree information supplied by national databases operated by Teagasc, the Department of Agriculture, Food and the Marine, the Irish Cattle Breeding Federation and the Centre for Veterinary Epidemiology and Risk Analysis. Variance components and heritability of *Mycobacterium bovis* susceptibility in Irish Dairy and beef cattle was estimated using a

total of 105,914 cow, 56,904 heifer and 21,872 steer single intra-dermal comparative tuberculin test records. A heritability of 0.11 was estimated for all animals analyzed (with a variation of 0.08 to 0.19 for each group separately). The genetic correlation between dairy and beef animals (and all groups) were all positive. This work indicated a much higher heritability for *Mycobacterium bovis* infection than other documented disease traits (~0.03). Moreover the positive genetic correlation between all animal groups and the correlation between increased prevalence of the disease in sire daughters and estimated breeding value, indicates that implementing a national breeding strategy for increased *Mycobacterium bovis* infection could be possible.

As *Mycobacterium bovis* infection has been documented to be an elusive polygenic traits, several of GWAS methods were implemented in this thesis to elucidate the underlying genes involved in host susceptibility. A total of 841 dairy sire genotypes with estimated breeding values for *Mycobacterium bovis* susceptibility were available. One quantitative trait locus on chromosome 23 was common to all three genome-wide analyses for association to *Mycobacterium bovis* susceptibility. Imputation to whole genome sequence of this QTL region revealed the gene *FKBP5* which encodes a protein from the immunophillin protein family, a family of proteins often targeted by immunosuppressant drugs.

Finally a search for signatures of selection and ancestry mapping was performed on two anciently admixed cattle groups, East African zebu cattle breeds and cattle originating from the Near Eastern Anatolian peninsula. These breeds provide an interesting intersection of the two domestic lines of cattle, *Bos indicus* and *Bos taurus*. Moreover mapping signatures of selection and ancestral haplotypes can often have a higher power of

analysis that traditional GWAS methods. A number of key identifiers of selection were identified in both breed groups, which East African Zebu breeds having suggestive signatures of selection for innate immunity and signatures of selection in Anatolian breeds indicating selection towards beef and dairy production traits.

The work presented in this thesis has helped the development of a national breeding strategy towards improved resistance to *Mycobacterium bovis* infection in Irish cattle herds which is expected to be implemented within the next two years. Furthermore it has identified a novel QTL region associated with *Mycobacterium bovis* infection.

Chapter One - Literature Review

1.1 Cattle production in Ireland

The agri-food sector in Ireland accounts for approximately 7.7% of gross value added (GVA) to the Irish economy, 8.4% of total employment (~163,000 jobs) and 11.4% of total merchandise exports in 2014 [1]. The land area of Ireland is 6.9 million hectares; of which 4.3 million is used for agriculture [2]. Approximately 80% of agricultural area is assigned to grass (3.4 million hectares), 11% to rough grazing and ~9% to crop production.

Table 1.1 Total number (000's) of livestock in Ireland from 2008-2014 source CSO

Farm animal	2008	2009	2010	2011	2012	2013	2014
Total cattle	6,303.9	6,231.7	5,917.7	5,925.3	6,253.2	6,309.1	6,243.1
Dairy cows	1,024.1	1,022.4	1,006.9	1,035.6	1,060.3	1,082.5	1,127.7
Total sheep	3,422.9	3,182.6	3,122.0	3,321.3	3,430.3	3,324.1	3,324.9
Breeding sheep	2,603.0	2,440.0	2,419.1	2,524.2	2,576.3	2,546.3	2,503.6
Total pigs	1,510.7	1,501.9	1,500.4	1,552.9	1,493.0	1,468.5	1,505.0
Breeding pigs	152.1	159.8	150.9	147.7	145.1	145.2	146.4

Cattle production in Ireland is the primary form of agriculture in Ireland in terms of livestock numbers Table 1 and percentage share of agri-food exports (Table 2) with beef and dairy production accounting for 51% or €5.2 billion of agri-food exports in 2014, while pig and sheep meat accounted for 5% and 2% respectively. Cattle production is the primary enterprise on Irish farms, with >100,000 or 90% of farms having a cattle enterprise while also being the main enterprise on almost 60,000 farms.

Table 1.2 contribution of each agri-food section to total agri-food exports for 2013/14 source CSO

	2013 (€m)	2014 (€m)	% Share of agri-food exports
Dairy products	2,969	3,055	29%
Beef	2,248	2,270	22%
Prepared foods	1,669	1,805	17%
Beverages	1,197	1,205	12%
Pig meat	552	570	5%
Seafood	496	540	5%
Edible Cereals	222	230	2%
Poultry	259	310	3%
Sheep meat	216	218	2%
Live Animals	245	245	2%
Total	10,072	10,448	100%

Cattle production in Ireland is split into two sectors; the dairy production system and the beef production system. The dairy industry in Ireland is the biggest single contributor to agri-food exports accounting for 29% of all agri-food exports or €3bn, with an annual production of ~6 billion liters from ~1 million milking cows and remains the main enterprise for 28,000 farm households with it being the most profitable farm enterprise [3]. The climate and large percentage of available agricultural land in Ireland gives an economic advantage to Irish dairy farmers over their European competitors as animals are able to be fed primarily on grassland for the majority of the year, while being fed silage during winter months, which was produced and stored during the peak growing periods.

Since 2015 and the abolition of milk quotas (which represented a cap on the amount of milk that a farmer could sell every year without paying a levy), Irish dairy production is set to increase significantly, to feed the increasing demand for dairy products

outside of the EU: in particular south-east Asia (China currently being the largest market for Irish argi-food outside of the EU) and Africa. Teagasc have predicted the creation of 15,000 jobs due to the abolition of milk quotas. The food harvest 2020 report has also predicted a 50% increase in dairy production over the next 5 years. New Zealand, which has not had a cap on its milk production, increased its milk production from 7.6 to 19 billion liters during the three decades in which milk quotas have been in place.

Beef production in Ireland is split into two sources; beef from suckler herds and beef from dairy herds. The beef industry in Ireland is much more than self-sufficient with eight times the national requirements being produced, resulting in >80% of beef produced in Ireland being exported. The combined value of Irish meat and livestock exports for 2014 was estimated at €3.6 billion with the primary export market for Irish beef being the United Kingdom with exports to the UK valued at €1.04 billion compared to €980 million for the rest of the European Union. Suckler cows are kept specifically for the production of beef, rather than milk. Their calves stay with the herd and are allowed to be fed by the dam. Suckler cows keep their offspring with them until the calves are ready to be sold on for either fattening or for slaughter. Ireland has an approximate beef suckler cow herd of 1.1 million spread across ~80,000 farms. The sires are generally beef breeds (Hereford, Angus, Limousin, Charolais or Simmental) crossed with dairy breeds, sourced from dairy herds. These cross-bred dams are then breed with pure-bred beef sires and will produce one calf per year, typically in the spring. The remainder of the ~1.9 million beef animals slaughtered in 2014 came from animals born from dairy herds, which were sold on for fattening after weaning or as a yearling [4]. The majority of animals used for beef are

steers (37% as of 2014) which are easier to manage than bulls and have heavier slaughter weights than cows or heifer. There has been an increase in the use of bulls as beef

Table 1.3 Output value (€m) and numbers (000's) of beef cattle and calves exported 2013/14 source CSO

	2013		2014	
	Value	Volume	Value	Volume
Live exports	117.41	213	117.83	240.93
Total slaughterings	1,929.16	1,589	1,954.75	1,745.91
Total disposals	2.058.18	1,802	2,088.71	1,986.84
Imports	10.98	13	10.35	11.09
Changes in stock	104.61	56	-67.74	-66.0
Total	2,151.81	1,845	2,010.63	1,909.75

animals in recent years with these animals being fed by winter finishing cows during the winter months [5]. Due to the heavy reliance of exports by the beef industry it can be greatly influenced by fluctuations in the market for beef products, particularly to fluctuations to market demand in the UK which makes up nearly 50% of the total export market for Irish beef. The food harvest 2020 report have predicted that a growth of 20% in the output of the beef sector will be achievable in the next 5 years.

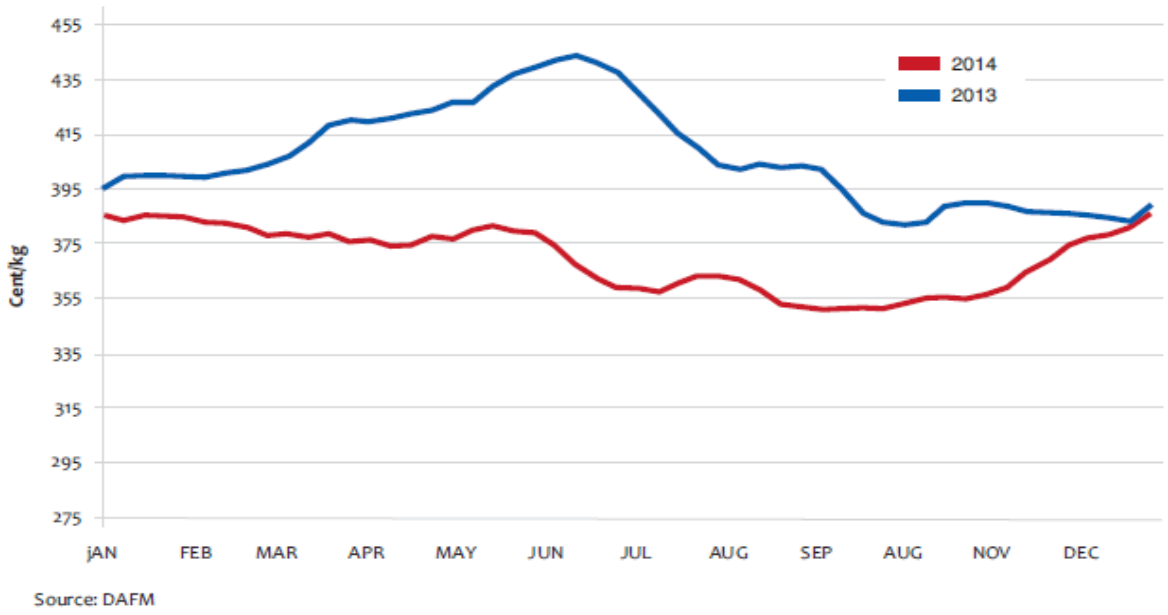


Figure 1.1 Steer prices for 2013/14

1.2 Irish cattle breeding federation

The Irish cattle breeding federation is a non-profit organization set up in 1998 whose aims are to provide the Irish dairy and beef industries with high quality breeding information. Their mission statement is "achieving the greatest possible genetic improvement in the national cattle herd". In order to bring about improvements in genetic gain the ICBF requires information on ancestry and quantitative data on traits of importance to industry partners and farmers for a large number of animals across generations, a system to identify genetically superior animals, a breeding scheme that ensures farmers use genetically superior animals and accurate data from farms and industry partners on their breeding and farming decisions. All of this information is ultimately collected and stored in the national cattle breeding database which contains information on

nearly all animals registered nationally, which includes all available performance data and information on the herds in which these animals belong.

The national cattle breeding database is a combination of two main databases; the CMMS database (Centralised Movement & Monitoring System), operated by the Department of Agriculture, Food and the Marine (DAFM), which monitors all movements of animals in and out of both herds and the country. And the ICBF database, operated on behalf of Irish dairy and beef farmers, which stores all phenotypic information on virtually all animals registered and genotypic information on all breeding sires and an increasing number of breeding dams also. Both databases are synchronised meaning both Irish farmers and the cattle breeding industry must only provide information once, with relevant data being allowed to be shared between all relevant stakeholders. The ICBF routinely performs genetic evaluations for both dairy and beef animals, using data extracted from the national cattle breeding database for ~60 traits in total. These evaluations are then fed back into the database and shared with relevant stakeholders and farmers through the ICBF website and various cattle breeding reports.

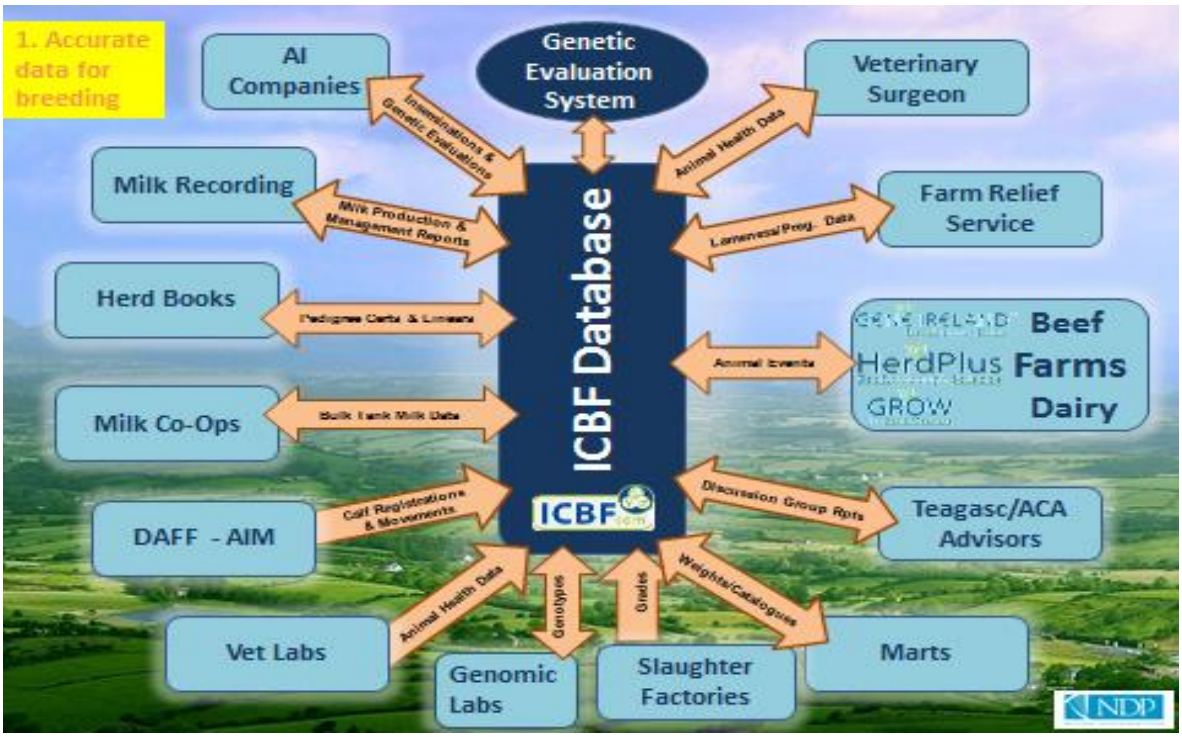


Figure 1.2 Flow chart of the sources and output of the ICBF national cattle database

1.3 Breeding strategies in Ireland

In order to best communicate the genetic merit of one sire over another to farmers the ICBF development the economic breeding index (EBI) in 2001, which replaced the RBI (Relative Breeding Index), a breeding scheme which focused only on milk production traits (Milk yield, protein yield, fat yield and protein percentage). The EBI is a single profit index, communicating to farmers the average profit (in euro) per lactation of daughters relative to the base population performance of cows born in 1995, who calved and milked in 2000, also referred to as the "base cow".

Table 1.4 Base population performance - Cows born 1995, calved and milk recorded in 2000

	Milk kg	Fat kg	Prot kg	Fat %	Prot %	CI days	Surv %
Base Cow Performance	5192	196	171	3.79	3.30	404	80

The EBI comprises of seven sub-indexes related to profitable milk production. The sub-indexes of the EBI are 1) Milk production, 2) fertility, 3) calving performance 4) beef carcass, 5) cow maintenance 6) cow management and 7) health [6]. Each of these sub-indexes are assigned a different weighting towards the overall EBI or genetic merit of each animal. The weighting of each trait in the EBI is given an economic weight, which were derived from a bio-economic model and quantify the change in profit per unit change in the target trait, holding all other traits constant. A summary of the sub-indexes, including traits and relative weightings within the EBI are given in Table 1.5.

Table 1.5 Economic values and % emphasis of various traits in the EBI formula. Source, ICBF

2014 Economic values and % emphasis for traits in the EBI				
Sub-Index	Trait	Economic Weight (€)	Trait emphasis %	Overall Emphasis %
Production	Milk	-0.09	10.6	33
	Fat	1.04	3.4	
	Protein	6.64	18.9	
Fertility	Calving Interval	12.43	24.0	35
	Survival	12.01	10.9	
Calving	Direct Calving Difficulty	-3.52	2.8	9
	Maternal Calving Difficulty	-1.73	1.3	
	Gestation Length	-7.49	4.1	
	Calf Mortality	-2.58	1.0	
Beef	Cull Cow Weight	0.15	0.7	9
	Carcass Weight	1.38	5.1	
	Carcass Confirmation	10.32	1.7	
	Carcass Fat	-11.71	1.1	
Maintenance	Cull Cow Weight	-1.65	7.2	7
Management	Milking Time	-0.25	2.1	4
	Milking Temperament	33.69	1.9	
Health	Lameness	-54.26	0.6	3
	Somatic Cell Count	-43.49	1.8	
	Mastitis	-77.10	0.8	

Before the introduction of the EBI the relative breeding index (RBI) was the breeding scheme of choice for Ireland. The RBI focus was on dairy traits only, comprising of four traits, 1) milk yield, 2) milk fat yield, 3) milk protein yield and 4) milk protein %, however a negative genetic correlation between milk yield and fertility [7, 8] meant that continued breeding of cattle using only milk related traits resulted in a detrimental decline of fertility in dairy cattle. Since the introduction of the EBI in 2001 this negative correlation has been

alleviated despite the continued selection for milk traits in dairy cattle, with fertility in dairy cattle again on the rise with milk yields also continuing to increase albeit at a slower rate (figure 1.3) [9]. The EBI has continued to evolve since its inception, adding more sub-indexes and re-evaluating the economic benefits that can be attained through genetic selection. The evolution of the EBI is shown in figure 1.4.

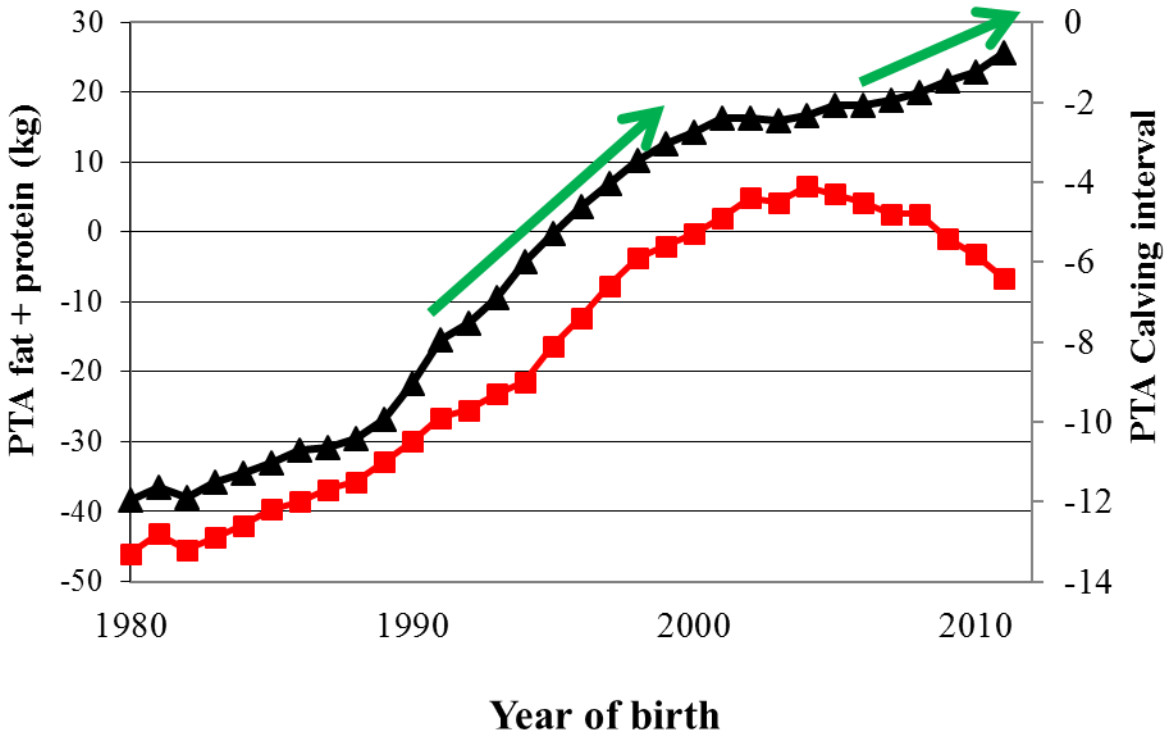


Figure 1.3 Genetic trends in milk yield (black line) and fertility (red line) in Irish dairy cattle since 1980. Source [9]

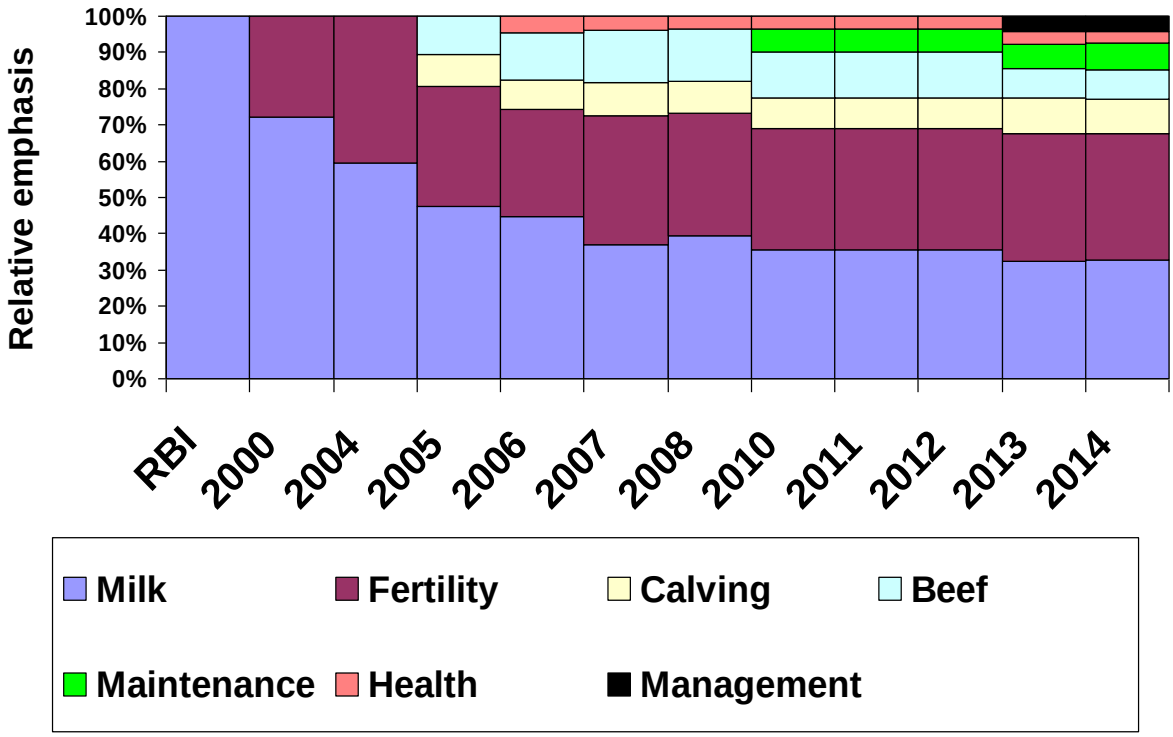


Figure 1.4 Evolution of the Economic Breeding Index

As the EBI is weighted more towards dairy farming and production an alternative beef breeding index, the Suckler Beef Value (SBV) was introduced in 2007. The SBV comprised of 5 sub-indexes; three of the sub-indexes (Calving traits, weanling export and beef slaughter) describe traits of interest for beef value whereas the other two (maternal milk & fertility, maternal calf quality) describe the traits of interest for producing high quality replacement heifers. The SBV however, was replaced in 2012 with the €uro-Star indexes, which comprised of a maternal and terminal beef indexes. The €uro-Star index can be broken down into three economic indexes; 1) a maternal index, for farmers wishing to identify animals for breeding replacements, 2) a terminal index, for farmers wishing to produce animals for breeding cattle for sale/slaughter and 3) a dairy beef index, for dairy

farmers wishing to identify bulls for breeding cattle for sale/slaughter. A summary of the relative weightings in the €uro-Star index is shown in table 1.6.

Table 1.6 Relative weighting % and traits in the Maternal and Terminal indexes

	Maternal	Terminal
Calving	21	30
Docility	4	3
Beef	26	49
Feed Intake	23	18
Daughter Milk	9	
Daughter Fertility	17	

The G€N€ Ireland breeding program set up by the ICBF in 2007 aimed at producing genetically elite Sires by contract mating the bulls selected from the top 1% of bulls in the EBI with cows identified as elite, high EBI Irish cows of unique pedigree, with proven performance for fertility and milk solids production. This program tests 100 bulls each year from the top 1% of cows on the EBI, each with 100 daughters. The program focuses on three key areas 1) selection of high EBI young bulls, 2) progeny testing the young bulls of a targeted set of herds and 3) returning elite bulls to AI. The primary benefit of G€N€ Ireland to the farmer is the availability of the top genetic bulls at a discounted rate. The use of genomic selection means that the G€N€ Ireland program can increase genetic gain by 48% or by €20-30 per cow per year, at a low cost. The G€N€ Ireland program initially focused on improving the genetics of dairy animals, but since 2010 beef animals have also been included.

1.4 Quantitative genetics

1.4.1 Basic model

As seen in the EBI the prediction of breeding values is an integral part of programs aimed at improving genetics in a cattle population. In order to produce high quality breeding values the availability of records on the traits of interest is of vital importance, as shown in the national cattle breeding database. Prior to the introduction of genomic breeding values the prediction of breeding values was based on records from individual animals and relatives. The phenotypic record of an animal can be express by the basic model for quantitative traits as;

Phenotype = environmental effects + genetic effects + residual effects

The genetic effects are a combination of the additive, dominance and epistatic genetic values of the genotype of an animal. The additive genetic effects, also referred to as the breeding value is average additive genetic effects of the genes an individual receive from both of its parents [10]. Since the additive genetic effects as based on the transmittable genes from parents to offspring, it is the only component that can be selected for, thus in most models the dominance and epistatic genetic value are presumed to be of little significance, so are included in the environmental effects of the basic linear model, expressed as.

$$y = \mu + g_a + e \qquad \textit{equation 1.1}$$

Where y is the phenotypic record of an animal, μ population mean for that phenotypic value, g_a is the additive genetic effects (or breeding value) and e is the environmental effects (which include the dominance and epistatic genetic effects). The breeding value a_i

of animal i can also be expressed as

$$a_i = g_{ai} = \frac{1}{2}a_s + \frac{1}{2}a_d + m_i \qquad \textit{equation 1.2}$$

Where a_s and a_d are the breeding values of the sire and dam and m_i is the Mendelian sampling, or the deviation of the breeding value of animal i from the average breeding value of its parents. This helps to express that each parent passes on a sample ½ of its genes and that there is variation in the genetic variation between animals with the same parents [10].

The ratio of the additive genetic variance to the phenotypic variance in a population can define the heritability of a trait, namely the narrow sense heritability which expresses the extent to which a trait is determined by the genes inherited from their parents. The broad sense heritability is the ratio of the genetic variance (including additive, dominance and epistatic effects) to the variance in the phenotype [11]. The narrow sense heritability is more important to breeding programs as it disentangles the environmental effects from the genetic effects in determining the phenotype. Thus when selective pressure is applied to a population such as in artificial selection of livestock, the response of the trait is related directly to the narrow-sense heritability. The narrow sense heritability can be expressed as follows.

$$h^2 = \frac{Var(A)}{Var(P)} = \frac{cov(a,y)}{var(y)} = \frac{\sigma_a^2}{\sigma_y^2} \qquad \textit{equation 1.3}$$

Where σ_a^2 is the additive genetic variance and σ_y^2 is the phenotypic variance. The accuracy of evaluations of variance components, reported as the reliability or repeatability (r^2).

When only one phenotypic record is the only information available, the accuracy can be expressed in terms of the heritability as.

$$r_{a,y} = \frac{cov(a,y)}{(\sigma_a \sigma_y)} = \frac{\sigma_a^2}{(\sigma_a \sigma_y)} = h \qquad \textit{equation 1.4}$$

Thus the reliability (r^2) equals h^2. The expected response to selection (R) can also be expressed in terms of the heritability, on the basis of a single record per individual [12].

$$R = ir_{a,y}^2 \sigma_y = ih^2 \sigma_y \qquad \textit{equation 1.5}$$

Where i, is the intensity of selection, or the superiority of selected animals over the population average in terms of the phenotypic standard deviation. For breeding schemes it is important to be able to estimate the annual rate of genetic gain in a population, thus the above equation can be translated into the following.

$$\Delta G = \frac{i.r.\sigma_g}{L} \qquad \textit{equation 1.6}$$

Where ΔG is the annual rate of genetic gain, r the accuracy of selection, σ_g the genetic standard deviation and L the generation interval. This is often referred to as the breeder's equation [12].

1.4.2 Best Linear Unbiased Prediction (BLUP)

One of the most popular models for estimating breeding values is the best linear unbiased prediction (BLUP) due to its ability to estimate fixed effects and breeding values simultaneously [12]. BLUP maximizes the correlation between true (a) and predicted breeding values ($\hat{a}$) and minimizing the prediction error variance, while estimating unbiased realized values for random variables such as animal breeding values and functions of fixed effects [13]. With the steady increase in computing power the utilisation of BLUP methods for computing breeding values and variance components has increased from the use of simpler BLUP models such as the sire model to more complex models such as the animal, multivariate, maternal and random regression models [10]. A wealth of computer programs using BLUP methodology are widely available such as PEST, BREEDPLAN and ASREML [14, 15]. BLUP can be expressed in the following equation for a mixed linear model.

$$y = Xb + Za + e \qquad \textit{equation 1.7}$$

Where y is an n x 1 vector of phenotypic observations, b is a p x 1 vector of fixed effects, a is a q x 1 vector of random animals effects, e a n x 1 vector of random residual effects. X is a matrix of fixed effects with an order of n x p and Z a matrix of random animal effects of order n x q. Both X and Z are termed incidence matrices. In matrix form the BLUP model can be represented as such.

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \frac{\sigma_e^2}{\sigma_a^2}A^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'Y \\ Y'Z \end{bmatrix} \qquad \textit{equation 1.8}$$

Where A is a numerator relationship matrix of the animal, based on their pedigree data, σ_e^2 is the random residual effect for each observation, $\hat{b}$ are the effects of sex on the trait and $\hat{u}$ are the estimated breeding values. The BLUP mixed model equation (MME) makes a number of assumptions however. These assumptions are; the distributions of y, u and e are assumed to be under multivariate normal, meaning that traits are assumed to be determined by many additive genes of small effect. The inclusion of the relationship matrix in the model accounts for genetic variance resulting from genetic drift, inbreeding or gametic disequilibrium [16]. The variances and covariances of the base population are assumed to be known, in reality these values are often not known but can be estimated by restricted maximum likelihood (REML). Several models exist for computing breeding values and variance components using BLUP, the most popular of these include the linear animal mode (LAM) threshold animal model (TAM), linear sire model (LSM) and threshold sire model (TSM).

Linear animal models are the most appropriate models for estimating variance components and use all relationships, however, are inappropriate when the target trait is a binary phenotype (i.e. in health/disease traits); in these cases the assumption that mean and variance are independent is violated [17]. Threshold animal models, are more appropriate for binary trait analysis and assume an underlying distribution for the trait and apply a threshold within this distribution [18]. The linear sire model has a greatly reduced computation load to the LAM however it will only evaluate the genetic merit of the sire and will not account for the genetic merit of the dam; meaning the LSM assumes all mates are of similar genetic merit. The threshold sire model have the same benefits as the TSM

with faster computation times [19], however the TSM ignores the relationship between daughter-dam pairs, which may lead to upwardly biased estimates [20].

A product of estimating breeding values using the BLUP methodology is that the breeding values tend to regress towards the mean, implying that animals estimated at having the best and worst breeding values are under and overestimated respectively. Furthermore using EBVs estimated from BLUP will be a product of the information available on all related animals (known as parent average), while this is useful for developing breeding programs, it can be troublesome when EBVs are used as response variables for genome wide analysis or genomic prediction. A method was proposed by Garrick et al. [21] to remove the parent average from estimated breeding values, effectively de-regressing the EBVs. The equation of de-regressing EBVs is as follows.

$$\tilde{y} = R(R^{-1} + A^{-1})\hat{a} \qquad \textit{equation 1.9}$$

where $\tilde{y}$ is a vector of deregressed EBVs, R is a diagonal matrix containing one divided by the animal's reliability from his progeny less one, A is the numerator relationship matrix among animals and $\hat{a}$ is a vector of the EBVs.

1.5 Genome wide association analysis

The availability of dense DNA marker information, particularly in the form of single nucleotide polymorphisms, has led to the rapid spread of genome-wide association studies (GWAS) as the *de facto* approach to identifying candidate regions for association with disease traits in both human medicine and livestock genetics. GWAS has also been a powerful tool for studying genetic variation and population structure for human health, agriculture, evolutionary history and ancestry, among others. GWAS exploits the population level associations between markers, known as linkage disequilibrium (LD). These associations arise due to small segments of the genome in a population being descended from the same common ancestor. These segments which have been inherited from a common ancestor without undergoing recombination, will have identical alleles or haplotypes across members of the population. If a causative mutation (also referred to as a quantitative trait locus or QTL) is present within these segments, then this causative mutation will be identical in members which inherited that segment [22]. While GWAS has been successful in identifying variants that are definitively associated with traits, in some cases discovering causal variants of disease traits, in many cases GWAS has shown that a large number of traits are polygenic, with the genetic effect on a trait being scattered across the entire genome. Despite the underlying biology of many traits remaining a mystery, the availability of genomic information still allows for the prediction of phenotypes based on genomic information. Furthermore advancements in methodology such as haplotype based methods and methods which fit all markers simultaneously could aid in removing the elusiveness of polygenic traits. The simplest and most commonly used approach for GWAS is a marker by marker analysis also known as single marker regression.

1.5.1 Single marker regression

The model for single marker regression, which is the association between a marker and a trait, given a population with random mating and no population structure is as follows.

$$y = Wb + Xg + e \qquad \textit{equation 1.10}$$

Where y is a vector of phenotype observations, *W* is a matrix assigning records to the phenotype fixed effects, *b* is a vector of fixed effects, *X* is a matrix allocating records to marker effects, *g* is the effect of the marker and *e* is a vector of random deviates. This is an additive model where the effect of the marker is treated as a fixed effect, and where it is assumed that a marker will only affect a trait if it is in LD with an unobserved QTL. The null hypothesis in the model is where the marker has no effect on the trait while the alternative hypothesis is that the marker does have an effect on the trait.

The power to detect a QTL using a single marker regression model is affected by the following factors. The strength of LD (r^2) between the marker and QTL [23], the proportion of total phenotypic variance explain by the QTL, the number of phenotypic records available, the minor allele frequency of the marker (the power of analysis is particular sensitive when this is low i.e. <0.1) and the significance level [22]. Luo [24] described a method for predicting the power of analysis when using a single marker regression approach. Hayes [22] described an example in dairy cattle; in dairy cattle the strength of LD (r^2) is ~0.2 at 100kb. To detect a QTL in dairy cattle with phenotypic variance of 0.05 and 1,000 records a minimum of 15,000 markers would be needed to achieve a power of analysis of 0.8 and ensure that there is a marker 100kb from every

QTL. However this assumes that all markers are evenly spaced and that all have a minor allele frequency greater than 0.2, which is not the case in practice.

In GWAS tens of thousands to potentially millions of markers are used and with the problem of multiple testing a question of which significance level is appropriate arises. As a result of the number of tests being performed in GWAS the significance of each individual test do not correspond to the actual significance in the whole experiment. With the number of markers being tested even if the expected number of false positive is 5% this could result in many thousand results being false positives. The significance level can be adjusted using a Bonferroni correction, however this does not take into account that markers are not independent from one another due to LD and thus can be very conservative. An alternative to seeking avoidance of false positives is to set a significance level with an acceptable number of false positive to positive results. The false discovery rate (FDR) is the expected proportion of significant tests that are in fact false positives [25-27]. FDR can be calculated for a GWAS as

$$mP_{max}/n \qquad \textit{equation 1.11}$$

Where P_{max} is a chosen P value significance threshold, n is the number of QTL which exceeds the significance threshold, and m is the number of markers tested.

1.5.2 Haplotype methods for GWAS

A haplotype is a set of polymorphisms that tend to be inherited together and can refer to a combination of alleles of SNPs that are found on the same chromosome. Instead of using single markers haplotypes can be used for GWAS. Using haplotypes in GWAS has the potential to increase the power as marker haplotypes may have greater LD with a QTL than single markers. If we consider two identical haplotypes, drawn from a random pair of individuals from a population, there can be two ways in which these haplotypes are identical. Either they are identical by descent (IBD) meaning they were inherited from the same common ancestor or they are identical by state (IBS) meaning they are the same due to chance recombination. As the length of these haplotypes increase from being single marker haplotypes to being many markers the chance of these haplotypes being IBD rather than IBS increases. Several methods exist for detecting IBD haplotypes; The fastIBD [28, 29] method is based on estimating frequencies of shared haplotypes using a probabilistic model, GERMLINE [30] uses an algorithm to find shared haplotypes exceeding a length threshold and more recently Refined IBD [31] combined these two methods.

Single marker regression, haplotype marker regression and IBD mapping approaches have been compared for their power and precision in GWAS [32, 33] using simulated data. These papers concluded that single marker regression had comparable power and precision to IBD methods but greater power and precision than haplotype regression methods. Real world results from Hayes et al. [34] and Calus et al. [35] contradict these results. Hayes et al. [34] found that using marker haplotypes gain greater accuracy that single markers for predicting QTLs and Calus et al. [35] found that an IBD approach to genomic selection gave greater accuracies for breeding values than haplotype

regression or single marker regression. In general, given the densities of markers available, IBD approaches are beginning to appear redundant. However for traits where the QTLs have low minor allele frequencies, using haplotype or IBD approaches may have benefits. For example a rare sequence variant associated with human type 1 diabetes was reported which was only detectable by IBD [36].

1.5.3 Bayesian approaches to GWAS

While single marker regression is the most common approach used for GWAS and has been used extensively to detect many associations, these significant associations only explain a small fraction of the genetic variance of a quantitative trait [37-39]. However there are two disadvantages to single marker regression approaches: firstly, the problem of multiple testing, meaning a very stringent significance level must be used; and secondly the region containing the true QTL can be hard to define as a large number of SNPs can be in LD with a QTL. This is a particular problem in cattle (and other livestock) as LD in some special cases can extend for mega-bases (Mb). A solution to both of these problems is to fit all SNPs simultaneously as random effects. An approach such as this has the added advantage of explaining a higher proportion of the genetic variance than single marker regression [40-42].

Bayesian multiple–regression models can simultaneously fit many more markers than the number of observations available. As it is often the case that many more markers are available than observations a regression method based on ordinary least squares cannot be used to estimate the effects of all of the markers. Thus Bayesian multiple regression

models use the Bayes theorem to combine prior beliefs of marker effects with the observed phenotypes and can then estimate all marker effects [43]. This method was originally intended for genomic prediction but has been used for GWAS also.

Bayesian methods to GWAS have an alternative approach to account for the problem of multiple testing, which is to control for the proportion of false positives among all positive results. The advantage of this approach is that the power of detecting QTLs is not inversely related to the number of markers tested, as is the case in single marker regression [25, 26, 44].

There are several methods to Bayesian analysis to GWAS and genomic prediction. The primary methods are BayesA, BayesB proposed by Meuwissen et al. [43], BayesC and BayesCπ [45]. Each of these methods can begin with a mixed linear model in the form

$$y = X\beta + Z\alpha + e \qquad \textit{equation 1.12}$$

Where y is a vector of phenotypic values, X is a matrix relating the vector β of fixed effects to y, Z is a matrix of genotype covariates (coded as 0, 1 or 2) for SNP markers, α is a vector of random partial regression coefficients of the SNPs and e is a vector of residual effects. In Bayesian regression models a prior distribution must be specified for β, α and e; each of the methods only differ in their prior used for α.

In BayesA, the prior assumption is that marker effects have identical effects on the trait and display an independent univariate t-distribution each with a null mean scale parameter S and v degrees of freedom. BayesA effectively assumes non-null effects for all loci, but in reality many loci are not segregating and a few have genetic variance. In BayesB the prior assumption is that marker effects have identical and independent mixture

distributions, where each has a point mass at zero with probability π and a univariate *t*-distribution with probability 1-π having a null mean. In simpler terms π is a prior belief of the proportion of SNPs thought to be truly associated with the trait. Thus BayesA can be thought of as a special case of BayesB with π=0. BayesC assumes a common variance for all SNPs in the model [46] while in BayesCπ, π is treated as unknown with a uniform prior. BayesCπ can be useful for getting a best estimate for the value of π to use in BayesB, while BayesC can be used to estimate the starting variances to use in BayesB.

Another advantage of Bayesian approaches to GWAS is the removal of the P-value as a measure of significance, which can often be ambiguous and does not translate well from experiment to experiment. The Bayes factor is an alternative to classical frequentist hypothesis testing. Bayes factors (BF) can be calculated as the odds ratio between posterior and prior probabilities for each individual marker [47].

$$BF = (f_i/(1-f_i))/(\pi_1/\pi) \qquad \textit{equation 1.13}$$

Where f is the model frequency of each SNP calculated, π is the posterior probability that a SNP is not associated with the trait and π_1 is the posterior probability that a SNP is associated with the trait (represented as 1- π). Bayes factors can be used to determine the level of association a marker has with a trait. Bayes factors above 3.2 are taken as 'substantially' associated, above 10 are 'strongly' associated and above 100 are 'decisively' associated with a trait [48].

1.5.4 Genomic Selection

Genomic selection increases the rate of genetic gain in a population by reducing the time needed to get reliable progeny tested records by using dense DNA marker information to predict an individual's genetic merit [43, 49]. Since in genomic selection, evaluations are based on SNP genotypes, breeding values can be calculated as soon as DNA information is available [50]. As per the breeders formula (equation 1.6) genomic selection reduces the generation interval (L) with genomic estimated breeding values (GEBV) obtaining the same reliability as traditional EBVs in about half the time (~3 compared to ~6 years). Several methods have been described for genomic prediction including the Bayesian methods described in the previous section. BLUP methodology has also been adapted to genomic selection, where, in short, the numerator relationship matrix A is replaced with a genomic relationship matrix (GRM). One of these BLUP based methods, gBLUP [50, 51] can use animals with or without phenotypic information. In this case the genomic information of the population of animals with phenotypic information (referred to as the training or reference population) can be used to predict the genetic merit of the animals without phenotypic information.

In Ireland, genomic prediction has been in place since 2009 with initial genomic selection being performed on a population of 945 Holstein-Friesian AI (artificial insemination) bulls to identify the optimal genotype for Irish production systems [52]. The reliability of the genomic evaluations have risen from 32% since its inception to 54% with significant increases in reliability each year. The model used for prediction GEBVs in Ireland is as follows

$$GEBV = G(R+G)^{-1}\tilde{y} \qquad \textit{equation 1.14}$$

where G is the genomic relationship matrix; *R* is a diagonal matrix containing one divided by the animal's reliability from his daughters less one and $\tilde{y}$ is the deregressed EBV (as per equation 1.9) for the trait under investigation [52].

1.6 Bovine Tuberculosis

Bovine tuberculosis (bTB) is a chronic respiratory disease caused by a bacterial infection from *Mycobacterium bovis* (*M. bovis*), a member of the *Mycobacterium tuberculosis* complex (MTBC), comprising of four sub-species (*Mycobacterium tuberculosis, M. bovis, M. africanum* and *M. microti*). Infection of livestock with bTB has an estimated global cost of €2 billion annually [53], primarily due to the loss of animals and animal productivity in developing countries, where bTB is widespread. In contrast the economic loss in developed countries due to bTB infection is in the control of the disease. Estimated costs to the Irish and UK governments in 2010 and 2011 due to bTB control were estimated at €63 million and €179 million [54], respectively. Bovine tuberculosis is also ranked as the fourth most important livestock disease globally [55], due to its importance in developing countries.

BTB is a zoonotic disease and is responsible for some cases of human TB, however, the majority of TB cases result from infection by *Mycobacterium* tuberculosis[56]. *M. bovis* has a much broader host range than M. tuberculosis as it has been documented in many wild animals and most domestic animals. *M. bovis* is capable of infecting humans through inhalation, ingestion and less commonly through contact with mucous membranes and broken skin[57].

Over the last two decades, *M. bovis* human infection has accounted for only a small amount, 0.5-7.2%, of TB cases in industrialised countries, however, it is a significant

problem in developing countries with *M. bovis* infection being major threat to public health[57].

in Ireland there were 470 cases of TB provisionally notified in 2008, among this there were 209 culture positive cases and only ten were *M. bovis* (4.8%)[58]. This is a lower percentage than the developing countries as milk pasteurisation is widespread thus limiting the risk of TB especially among the urban population.

Infection with *M. bovis* in cattle leads to the development of granulomatous lesions or tubercles in diseased tissues [20]. All of the MTBC members are genetically very similar with *M. bovis* being >99% sequence homology with the causative agent for human tuberculosis *Mycobacterium tuberculosis* [59]. The evolution of the bacterium was originally thought to have occurred from zoonosis with *M. tuberculosis* evolving from *M. bovis* around the time of cattle domestication (10,000-15,000 years ago). However since the sequencing of the *M. bovis* genome [59] and deletion analysis of MTBC strains [60] it appears that *M. bovis* forms a separate lineage that branched from a common ancestor of the MTBC when members were already present in humans. These small genetic difference between MTBC strains and *M. bovis* has resulted in biological differences to other MTBC strains including pathogenesis, antigen variation and range of hosts [61].

The most widely recognized source of bTB infection in cattle is from the infection of aerosols containing bacilli excreted from the lungs of infected animals [62, 63]. Infection is also possible through interaction with cattle and their environment, such as at pasture or feeding/drinking from troughs that have been contaminated by waste from infected animals. Over 60% of naturally infected cattle have been shown to be caused by aerosol infection[64]. In contrast to *M. tuberculosis* which almost exclusively only infects

humans, *M. bovis* has a much broader host range and can infect most domestic animal species and many wild mammal species. However, only a small number of species can act as a "reservoir" for *M. bovis*, as this requires *M. bovis* to be self-sustaining in these populations over generations [64]. In the UK and Ireland the Eurasian badger (*Meles meles*) is a reservoir for *M. bovis* continued infection of cattle. Wildlife reservoirs in other countries include, the brushtail possum (*Trichosurus vulpecula*) and the ferret (*Mustela furo*) in New Zealand, the Cape buffalo (*Syncerus caffer*) in South Africa, antelope (*Kobus leche kafuensis*) and other Bovidae in Africa, the North American bison (*Bison bison*) in Canada and the white-tailed deer (*Odocoileus virginianus*) in Michigan, USA [65].

Due to the small number of infected cattle found in *M. bovis* infected herds, it is thought that transmission of *M. bovis* amongst cattle is low [66]. As such most of the focus for permanently eradicating bTB from countries has been on control or culling of wildlife reservoirs. In 2006 an eight yearlong Randomized Badger Culling Trial (RBCT) concluded whose aim was to investigate the effectiveness of badger culling on reducing bTB infection in cattle [67, 68]. The RBCT performed both proactive (i.e. repeated annual culling across all land) and reactive culling (i.e. culling on a single occasion locally on or near farmland where a bTB outbreak had occurred). Reactive localized culling was halted in 2003 before the end of the trial as early results showed an increase of 18.9% in bTB infection in cattle herds when compared to control areas. Reactive culling was thus deemed highly unlikely to have anything but negative effects on bTB control, as the reactive culling disrupted localized badger populations, driving them into neighboring areas and forcing more contact between badgers and other species. Proactive culling, while initially seeing an increase of 24.5% in bTB herd incidence in areas outside of the culling area (with a decrease of 23.2%

within the culling area), after a few years an overall decrease in bTB incidence of 34.1% within culling areas and 5.6% in areas neighboring culling areas. Despite this the number of breakdowns prevented from badger culling was not deemed to be sufficient to offset the financial cost of culling [68-70].

1.7 Bovine tuberculosis eradication programs in Ireland

During the 1950's bTB was a major cause of loss of livestock and perhaps human illness in Ireland. The early success of a bTB eradication program in the UK that commenced in 1950 meant the prospect of a loss in exports to the UK due to the continued endemic bTB infection in Irish cattle. In 1954 a pilot eradication program was started with a compulsory and national scale scheme underway by 1962. This program includes as a primary program measure an annual single intradermal cervical comparative tuberculin test (SICTT) of all herds, as well as animal identification and movement system (AIMS) which monitors all movements of animals in and out of herds as well as recording birth. In more recent years this program has expanded to control the wildlife reservoirs of bTB infection also. From the inception of the eradication program animal incidence fell rapidly in the early years with national incidence falling from 17% in cows only to 0.44% in all animal categories by 1965. However since then progress has been stagnant with annual incidence only falling below 0.5% in recent years [54, 71]. bTB in Irish badgers was first reported in 1974 [72], however the significance of this finding was not realized for 15 years until further evidence emerged that showed the disease to be endemic in this species suggesting the badgers involvement in continued infection of bTB in cattle [71].

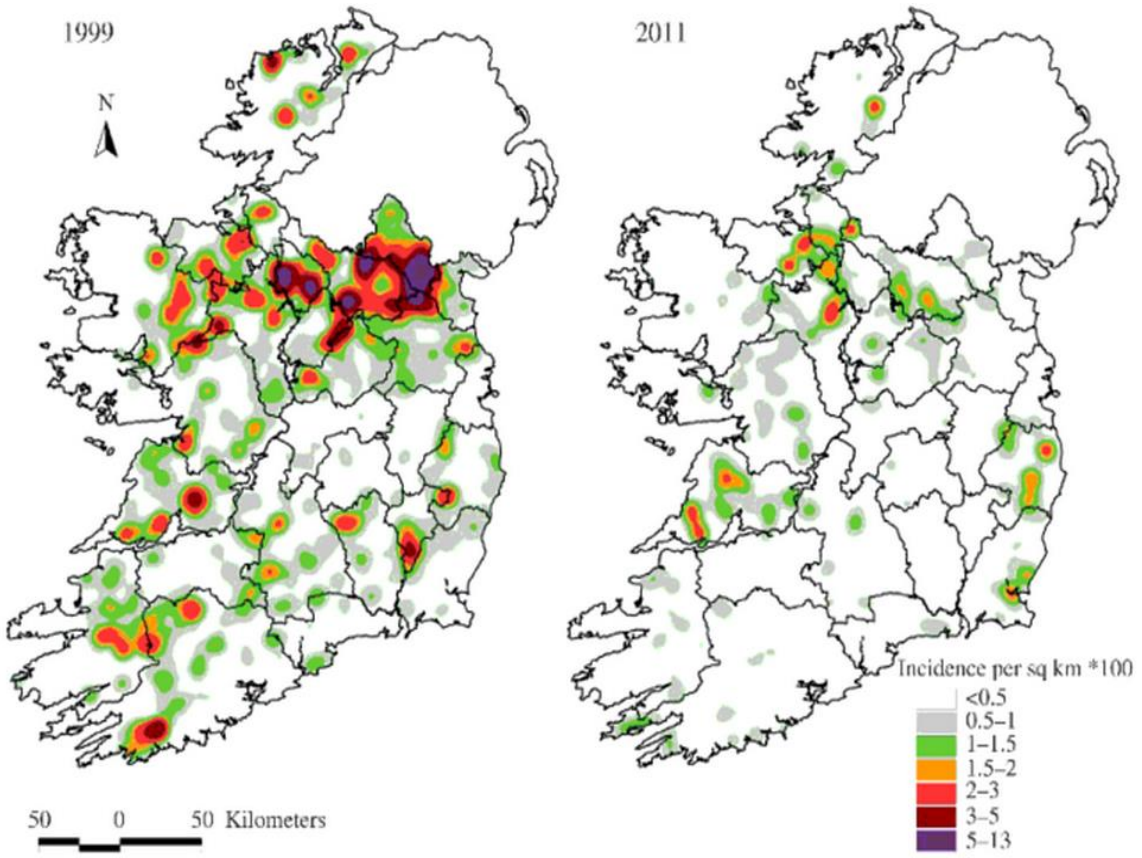


Figure 1.5 Density of TB incidence per km^2 for 1999 and 2011 source:[71]

The 1980s saw the formation of the Eradication of Animal Disease (ERAD), now since renamed Centre for Veterinary Epidemiology and Risk Analysis (CVERA) who were tasked with identifying the factors which prevent eradication of bTB in Ireland and identify means of improving the rate of eradication. CVERA set up research programs aimed at improving the diagnostic tools available for use in the eradication program and studying the role played by badgers in the maintenance of bTB in cattle [71].

Several badger culling studies have been conducted in Ireland which have shown both a beneficial impact on the risk of future breakdowns and no evidence of a negative impact on the risk of bTB infection in herds in surrounding areas [73-75]. Due to

this evidence in Ireland, areas which are implicated in a serious outbreak of bTB, badgers are captured as part of the national bTB eradication program. The DAFM and CVERA have also been collaborating in the development of a vaccine to control *M. bovis* infection in badgers. To date the use of oral vaccination of badgers has generated high levels of protective immunity against challenge with bTB [76]. Field trials are ongoing to investigate the effectiveness of a vaccine, however it will be a number of years before benefits may accrue [77].

With the combination of wildlife reservoir and primary host control of bTB; the eradication program in Ireland has seen a fall in the incidence of the disease from 7.5% in 2000 to 4.2% in 2011, with the number of serious outbreaks significantly reduced. Furthermore the number of reactors in 2011 was 18,500 compared to an average of 33,000 in the 1990s, resulting in the first time that reactor number have fallen below 20,000 since the bTB eradication program officially began [54, 71, 78, 79].

The current bTB eradication program has the potential capability of reducing and ultimately eradicating bTB from cattle herds in Ireland. However its success is ultimately hinged on the ability to reduce further the rate of infection in badgers. With questions persisting on the efficacy and cost-effectiveness of vaccinating badger populations, further options have been explored. The variance component and heritability of bTB in cattle has been extensively studied [80-82] with a view of developing EBVs that could be used for a breeding strategy to increase resistance in the Irish cattle population to bTB infection [83]. Considering the current version of the eradication program combined with a successful vaccination program for the wildlife reservoir and the possibility of a breeding scheme

targeted at improving resistance to infection of bTB, there is good reason to be optimistic about the eventual eradication of bTB from the Irish herd.

Additional policy changes to the bTB eradication program in Ireland were introduced in recent years. These included the confinement of inconclusive reactors to the herd of origin with a life-long movement restriction, except direct to slaughter and increased restrictions on herds contiguous to high-risk bTB breakdowns [84]. During the last 15 years the primary focus of research into bTB has been in improved understanding of risk of bTB infection, which has led to improvements in policy resulting in both increased surveillance and control of the disease. Current research is now focused on two key areas: how to limit cattle-to-cattle transmission and sustainable methods to limit badger-to-cattle transmission, in particular the development of a bTB vaccine for badgers. It is likely that improvements in these areas will greatly reduce bTB infection in cattle but not eradicate it completely, but discussions on measures to be taken after this are already commencing [84].

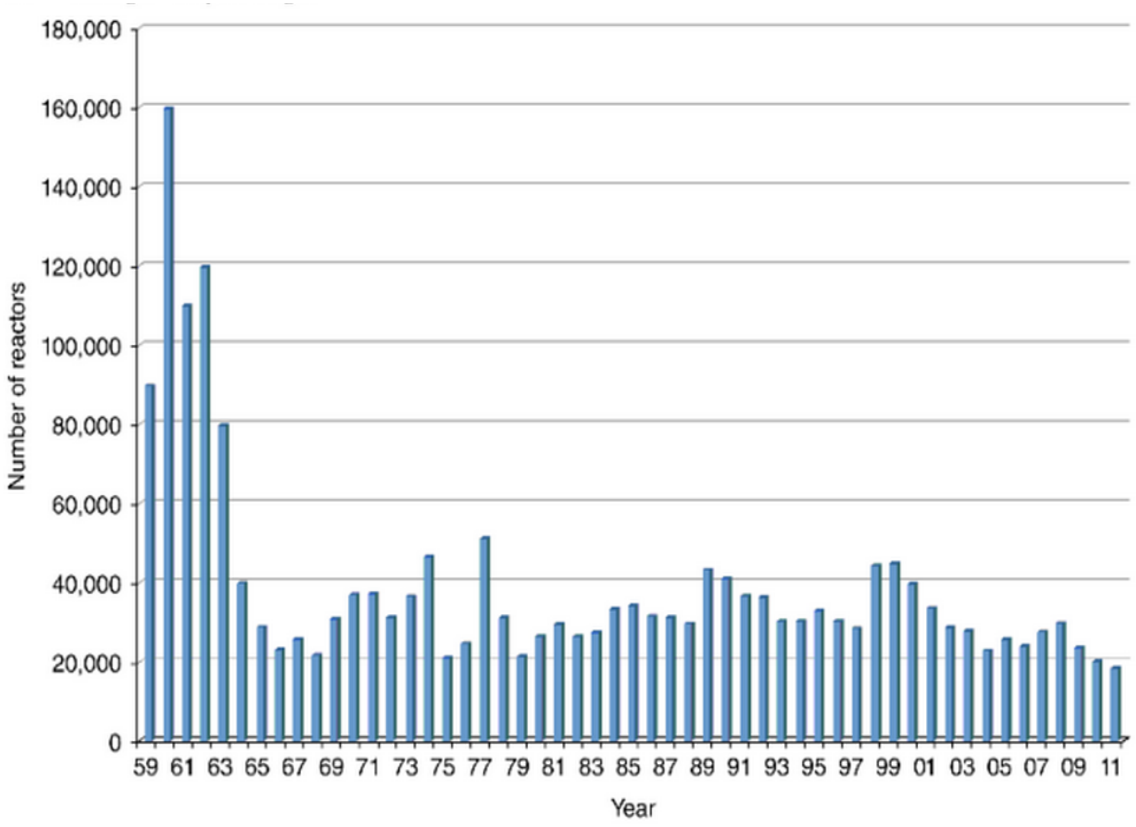


Figure 1.6 Reactor animals identified per year, 1959-2011 source:[71]

The following chapter has been published in

Genetics Selection Evolution, volume 46, issue 77, under the title;

"Variance components for susceptibility to Mycobacterium bovis infection in dairy and beef cattle"

Authors were

Ian W Richardson, Daniel G Bradley, Isabella M Higgins, Simon J More, Jennifer McClure, and Donagh P Berry

Contributions by Authors

IWR. Made substantial contribution to conception, design and undertook the analysis, interpretation and editing of data. I.W.R. also gave substantial contribution to the drafting of the manuscript

DGB. Gave final approval of the version to be published

IMH. Undertook the collection of data and gave substantial contribution to the interpretation of data

SJM. Was substantially involved in the drafting and revising of the manuscript for important intellectual content.

JM. Was involved in the revising of the manuscript for important intellectual content

DPB. Made substantial contribution to the concept, design and interpretation of data. D.P.B. also made substantial contributions to the drafting and review of the manuscript.

Chapter two - Variance components for susceptibility to *Mycobacterium bovis* infection in dairy and beef cattle

2.1 Introduction

Bovine tuberculosis (bTB) is an infectious chronic respiratory disease caused by *Mycobacterium bovis*. Infection of livestock with bTB has an estimated global cost of €2 billion annually [53], due mostly to the lack of bTB control in developing countries, where infection is endemic, which results in reduced productivity of livestock. The primary cost of bTB infection in developed countries is the control of bTB, with estimated spends by the Irish and UK governments in 2010 and 2011 of €63 million and €179 million [54], respectively. Bovine tuberculosis is ranked as the fourth most important livestock disease globally [55], primarily due to its importance in developing countries. Eradication of bTB from cattle herds has been successful in some countries such as Australia and throughout Scandinavia. Epidemiological circumstances in developed countries where bTB is present vary. The UK and Ireland have yet to gain an official bTB free status [85], although both countries introduced eradication programs during the late 1950s. In the UK, bTB is an expanding epidemic, while in Ireland, it forms a stable endemic with substantial year-on-year reductions, particularly recently [54]. A compulsory bTB eradication program for cattle was introduced in Ireland in the late 1950s. The initial response was rapid but by the mid-1960s, progress had reached a plateau, with the number of animal infections per year in excess of 30 000 [54]. In recent years, the number of animal infections has decreased to less than 20 000 per year; however, progress towards a bTB free status in Ireland remains slow [78, 79]. When the eradication program began in the late 1950s, confirmed animal infection prevalence was 17%; in 2003, this prevalence was 0.4% [86]. Re-emergences of

bTB have been reported in France and Spain [87-89], with concerns that bTB could expand into neighboring countries [90].

The existence of wildlife reservoirs for bTB is a major contributing factor to the continued infection of bTB in livestock. In Ireland and the UK, cattle share their environment with wild badger (*Meles meles*) [86], which act as a reservoir for the disease.

Genetic selection for bTB resistance requires information on the extent of genetic variation for bTB susceptibility within the Irish cattle herd. Heritability for susceptibility to *M. bovis* infection has previously been documented to be 0.18 in both Irish [17] and UK dairy herds [81]. However, these studies were confined to dairy cattle and, to date, no study has been undertaken on beef cattle. Moreover, the potential of genotype by environment (GxE) effects for susceptibility to *M. bovis* infection between herds of different disease prevalence levels has not been quantified.

The objective of the present study was to estimate variance components for bTB susceptibility in both dairy and beef cattle and to quantify whether GxE for bTB susceptibility exists across environments that differ in bTB prevalence. In this study, tuberculin skin test results from both dairy and beef cattle were used as a measure of bTB susceptibility. The variance components estimated in this study are needed to derive multi-breed estimated breeding values (EBV) for bTB resistance in national genetic evaluations. As well as being useful for incorporation into breeding programs, these EBV could also be useful phenotypes for subsequent genomic analysis.

2.2 Methods

2.2.1 Testing procedures

Ireland has operated a compulsory bovine tuberculosis (bTB) eradication program since the late 1950s. The current national program consists of a mandatory registration system for herds, an individual bovine identification system, a computerized movement monitoring system (CMMS), an animal health computer system (AHCS), as well as a comprehensive program of disease surveillance and control. Two forms of surveillance are used in Ireland. Field surveillance for bTB is performed annually in each herd through the routine use of the single intra-dermal comparative tuberculin test (SICTT) [91, 92]. Testing is carried out by veterinarians approved by the Irish Department of Agriculture, Food and the Marine (DAFM). In addition, abattoir surveillance is conducted on all animals at slaughter, relying on the gross examination of each carcass by an approved veterinarian for pathology suggestive of bTB.

Field surveillance using the SICTT involves intradermal injection of bovine tuberculin (a purified protein derived from *M. bovis*) into the neck of the animal. This inoculate will cause an immune reaction to the tuberculin in animals that are sensitized to antigens in bovine tuberculin and results in an inflammatory response and localized swelling at the injection site that reaches its greatest intensity 48 to 72 h post-injection. To distinguish animals infected with different strains of *Mycobacterium* from those infected with *M. bovis*, animals are also injected at a different site with avian tuberculin (from *M. avium*). The outcome of the SICTT is determined by the relative difference in the thickness of skin-folds in reaction to the bovine and avian tuberculins. Using the so-called standard interpretation (as would be used in herds with bTB free status that can, therefore, be traded

without restrictions), if the relative bovine-avian difference in skin-fold test is greater than 4 mm, the animal is considered a ‘standard reactor’, if between 1 and 4 mm, a ‘standard inconclusive reactor’, and otherwise, a ‘non-reactor’ [92]. Inconclusive reactors are re-tested 60 days after the initial test. Post-mortem examinations and/or laboratory culture of tissue samples can be used to confirm the presence of bTB in reactors.

A herd loses its official bTB free status if at least one reactor is detected, and then movement of animals from the herd is restricted. Subsequently, infected animals are removed for slaughter and further SICTT testing is conducted at approximately two-month intervals, until two consecutive clear herd tests (that is, herd tests with no positive or inconclusive reactors disclosed) are achieved, at which point herd bTB free status is regained. In restricted herds, a lower threshold (so-called, severe interpretation) of the SICTT is used, such that all animals with a bovine-avian difference of at least 2 mm are considered a ‘reactor under severe interpretation’ and animals with a bovine reaction less than 2 mm greater than an avian tuberculin reaction are considered an ‘inconclusive reactor under severe interpretation’. Additional skin tests are always undertaken on herds that are contiguous to, or otherwise linked to, infected herds, and in herds that are located in 'at-risk' areas. Herds are also re-tested for bTB six months after they have re-gained bTB free status.

2.2.2 Herd-episodes of bTB infection

In the present study, a 'herd-episode of bTB infection' refers to the full period of herd restriction triggered by disclosure of bTB infection within a herd [93]. The herd-episode starts with the disclosure or discovery of at least one infected animal (from either field or abattoir surveillance) and ends immediately following two consecutive clear herd tests at approximately two-month intervals. In the present study, episodes were only included in the analysis if at least two standard reactors were detected from field surveillance, of which at least one had to be home-born. Herd-episodes with 10 reactors or more were only included if at least 20% of the reactors, based on SICTT results, presented bTB lesions postmortem from abattoir testing.

2.2.3 Data collection and editing

Tuberculin skin test results from 22 381 herd-episodes in 16 717 dairy, beef and mixed herds were collected between 2001 and 2010, inclusive. To investigate any difference in genetic variance between herd types, herds were classified as either dairy, beef or mixed, determined by the average breed composition of cows in each herd. Herds were defined as dairy herds if the average breed composition of that herd was >= 90% dairy breed, while herds were defined as beef herds if the average breed composition of that herd was >= 30% beef breed. All other herds were defined as mixed herds. The total number of breeds represented across all herd-episodes was 35, with the mean number of breeds per herd-episode being 2.3; 5571 herd-episodes had more than one breed present during a herd-episode, across both dairy and beef herds.

Only positive and inconclusive results were recorded in the database. During these herd-episodes, there were a total of 183,955 'standard reactors' and 'reactors under severe interpretation', 20,884 'standard inconclusive reactors', and 1896 'inconclusive reactors under severe interpretation'. Confirmed bTB lesions were identified in 50,614 reactors from postmortem examinations and/or tissue culture samples. The national animal identification and movement (AIM) system, which monitors animal movements in and out of herds, was used to identify animals present in herds at the time of testing. In total, 1 357 791 animals were present during herd-episodes but did not test positive or were inconclusive for bTB. Animals were only included if they had moved into a herd more than 6 weeks prior to the start of a herd-episode. In this analysis, only 'standard reactors' and 'reactors under severe interpretation' were considered to be infected with bTB, thus the inconclusive reactors were removed from the dataset. Animals with no recorded sire were also removed from the dataset.

Animal age at the start of each herd-episode was determined. Female animals that were more than 30 months old or had calved at least once by the start of the herd-episode were classified as cows, and otherwise as heifers. Male animals that were less than 36 months at the start of the herd-episode were classified as steers (i.e., castrated bulls), and otherwise as bulls. Due to the paucity of data (n=1,822), bulls were subsequently discarded from the dataset. The final dataset consisted of 21,872 steers, 105,914 cows and 56,904 heifers; of these, 22,573 were reactors and 162,117 were nonreactors.

2.2.4 Heterosis and recombination effects

Considerable cross-breeding exists in Irish herds. For example 58% of herd-episodes in beef herds contained dairy ancestry. Many cows in beef herds are first or second crosses from dairy herds. Following VanRaden and Sanders [94], heterosis was calculated as $1 - \sum_{i=1}^{n} sire_i \cdot dam_i$, and recombination loss as $1 - \sum_{i=1}^{n} (sire_i^2 + dam_i^2)/2$, here $sire_i$ and dam_i are the proportions of breed i in the sire and dam, respectively.

2.2.5 Data analysis

Variance components for bTB susceptibility, as a binary trait (i.e. reactor or not reactor), were estimated using both an animal linear mixed model and a threshold animal model in ASreml [15]. Variance components were estimated separately for cows, heifers, and steers, as well as within beef, dairy and all herds, the latter also including mixed herds. Fixed effects considered for inclusion in all models were herd-episode and heterosis and recombination loss coefficients of the animal (as covariates). Parity (1, 2, 3, 4, ≥ 5) and stage of lactation (0-60, 61-120,121-180, 181-240, 241-300, > 300 days) were included as fixed effects in the analysis of cows. Age (in days) was included as a covariate in the analysis of steers and heifers. Animal sex was included in models for the combined analysis of cows, heifers and steers (with animal class removed in order to avoid confounding). Animal was included in all models as a random genetic effect and founder animals were allocated to genetic groups by breed; preliminary analysis revealed no permanent environmental effects and therefore was ignored in the analysis.

Genetic correlations for bTB susceptibility between animal type (cows, heifers and steers) were estimated using bivariate sire models to evaluate whether bTB was genetically the same trait in the different populations. To test whether the genetic correlations between animal types differed from 1, using the linear model, a likelihood ratio test was performed between each model and an analysis where the genetic correlation in each model was constrained to 1. Additionally, using the linear model, a likelihood ratio test was performed between each model and a model where the genetic variances in each animal type were constrained to be equal.

Sire estimated breeding values (EBV) for bTB susceptibility were generated from the univariate animal model, using only cow records. The mean EBV per breed was obtained for (1) all sires with daughter records for bTB in the data and (2) all sires with at least 10 daughter records for bTB in the data. Mean EBV per year of birth was calculated to quantify genetic trends. Only sires born after 1992 were included to calculate the mean EBV per breed and genetic trends because animals that were born before 1992 were too distant from the actual phenotypic data. The prevalence of bTB infection in the daughters of sires was calculated as the ratio of non-infected to infected daughters present during a herd-episode. Only sires with more than 50 daughters in more than 10 herds were included in this analysis.

To determine whether bTB in the different environments was genetically the same trait, herd-episodes were split into four categories based on mean bTB infection prevalence (calculated as the percentage of animals infected during a herd episode) to define different environments: very low (< 0.25 prevalence), low (0.25 to 0.50 prevalence), high (0.50 to 0.75 prevalence) and very high (> 0.75 prevalence). (Co)variance components were

estimated within each environment using both linear and threshold animal models. For comparative purposes, variance components estimated using the linear animal model were also transformed to the liability scale using the standard Robertson transformation [95]. Genetic correlations between bTB environments were estimated using a series of bivariate sire linear models to determine whether bTB in the different environments was genetically the same trait. For the linear models, a likelihood ratio test was performed to investigate the existence of GxE by comparing the unconstrained bivariate models to a bivariate analysis where the genetic correlation between environments was set to 1. A separate likelihood ratio test was performed for the linear models to investigate the existence of re-scaling effects between environments by comparing each bivariate model to a bivariate analysis where the variances were constrained to be equal.

2.3 Results

The mean prevalence of bTB infection for the entire dataset (cows, heifers and steers) was 0.11 (Table 2.1). Prevalences of bTB in dairy and beef herds as well as in heifers, steers and cows are summarized in Table 2.1. The greatest prevalence of bTB infection was observed for cows, with a slightly greater prevalence for beef than for dairy cows. Prevalence of bTB infection was lower for heifers than for cows or steers and was lowest for heifers from dairy herds.

2.3.1 Variance components

The estimate of heritability of susceptibility to *M. bovis* infection for the entire dataset, using a linear model, was 0.11 (Table 2.1). The estimate of heritability was highest for heifers from beef herds (0.19) and lowest (0.08) for heifers from dairy herds (Table 2.1). Estimates of heritability of susceptibility to *M. bovis* infection were similar for dairy and beef herds. Younger animals (both heifers and steers) had higher estimates of heritability than cows. The genetic standard deviation (SD) for susceptibility to *M. bovis* infection, when estimated across all animals, was 0.094 and was highest for heifers from beef herds and lowest for heifers from dairy herds (Table 2.1). The estimate of genetic variance for bTB susceptibility was greater in beef herds than in dairy herds (Table 2.1) and was also greater ($P < 0.001$) for cows than for younger animals (i.e., heifers and steers) but did not differ significantly between heifers and steers. Increased breed heterozygosity was associated with lower susceptibility to bTB (Table 2.1), for example, 100% heterozygosity (i.e. as in a F1 crossbred) in dairy herds resulted in a 3.2% reduction in bTB

prevalence. Estimates of recombination loss effects were different from 0 in all sub-populations, except for cows, heifers in beef herds and steers in dairy herds.

Estimates of the genetic correlation for bTB susceptibility was 0.55 between cows and heifers (standard error (se) = 0.048), 0.10 between cows and steers (se = 0.104) and 0.64 between steers and heifers (se = 0.082). The likelihood ratio test between nested models revealed that all genetic correlation estimates were significantly different from 1 ($P < 0.001$).

Table 2.1 Results and summary statistics for univariate analyses for susceptibility to bovine tuberculosis (M. bovis) infection.

Model	N	Mean	σ_g	Heterosis (se)	Recombination (se)	h^2(se)
All herds	184,302	0.11	0.094	-0.002 (0.004)	-0.011 (0.011)*	0.11 (0.006)
Beef herds	32,191	0.14	0.114	-0.026 (0.007)*	-0.092 (0.027)*	0.13 (0.016)
Dairy herds	126,646	0.09	0.098	-0.032 (0.005)*	-0.031 (0.012)*	0.12 (0.007)
All cows	105,526	0.20	0.110	-0.026(0.005)*	-0.002 (0.014)	0.14 (0.009)
Cows in beef herds	10,202	0.22	0.108	0.004(0.016)	-0.003 (0.062)	0.10 (0.027)
Cows in dairy herds	87,918	0.19	0.109	-0.025(0.006)*	-0.009(0.015)	0.15 (0.010)
All heifers	56,904	0.09	0.098	0.003 (0.006)	-0.018 (0.018)*	0.15 (0.014)
Heifers in beef herds	13,468	0.10	0.124	-0.002 (0.011)	0.018 (0.037)	0.19 (0.030)
Heifers in dairy herds	33,987	0.07	0.068	-0.012 (0.009)*	-0.019 (0.019)*	0.08 (0.013)
All steers	21,872	0.10	0.102	0.000 (0.009)	-0.042 (0.032)*	0.15 (0.030)
Steers in beef herds	8,610	0.11	0.119	-0.009 (0.012)	-0.044 (0.044)*	0.18 (0.040)
Steers in dairy herds	4,741	0.09	0.095	-0.000 (0.018)	-0.032 (0.046)	0.17 (0.047)

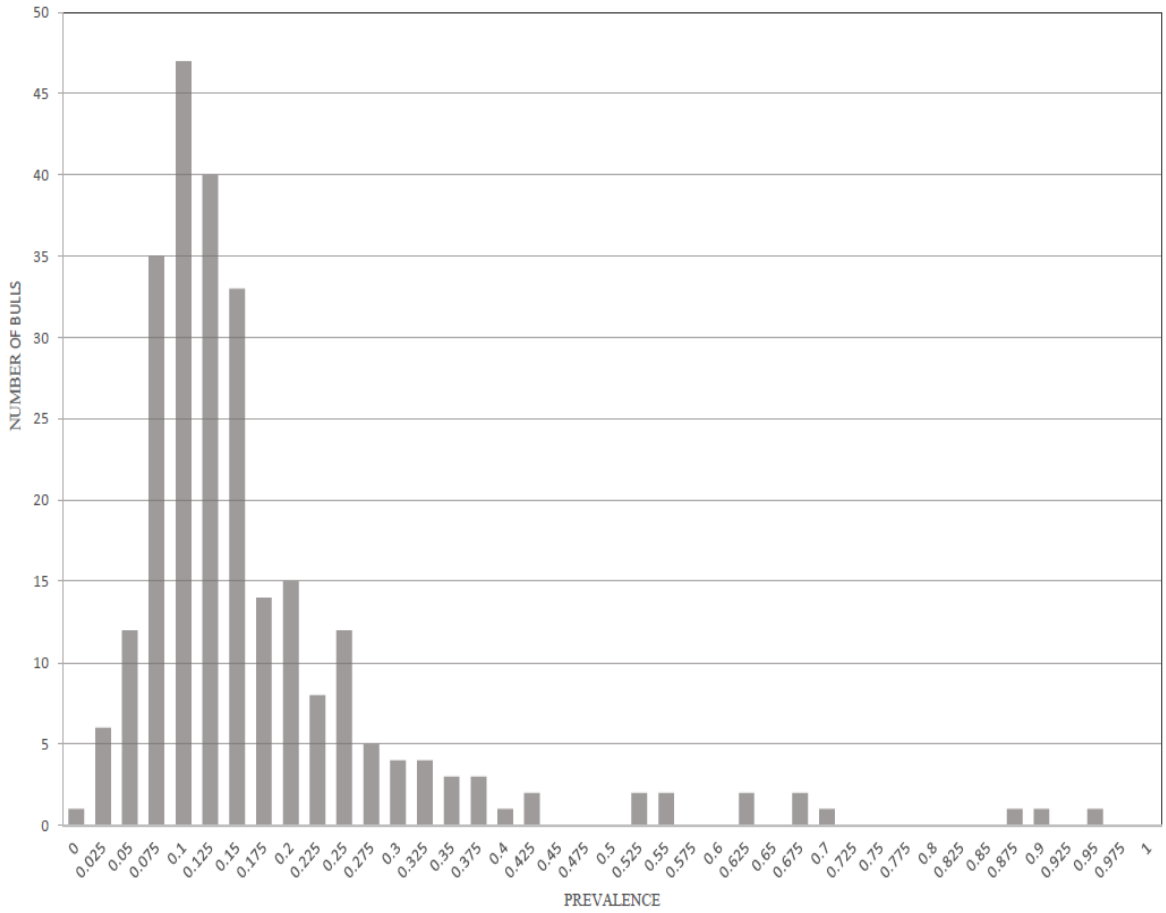
Number of records (N), mean prevalence of infection (Mean), estimates of the genetic standard deviation (σ_g), heterosis, recombination loss effects, and heritability (h^2)

**, denotes significance (P<0.05) from zero.*

Figure 2.1 shows the mean bTB prevalence for daughters of sires with more than 50 female progeny in more than 10 herds in the population under study. The mean bTB prevalence for daughters per sire ranged from 0.00 to 0.95. Figure 2.2 shows the mean bTB prevalence plotted against the EBV for bTB susceptibility for sires with more than 50 female progeny in more than10 herds; the correlation was 0.27, the correlation using all available sire information was 0.13. To investigate whether the correlation shown in figure 2.2 was not comprised of two clusters the correlation for sires with more than 50 daughters in more than 10 herds, with a daughter prevalence of <0.5 and a second group where the

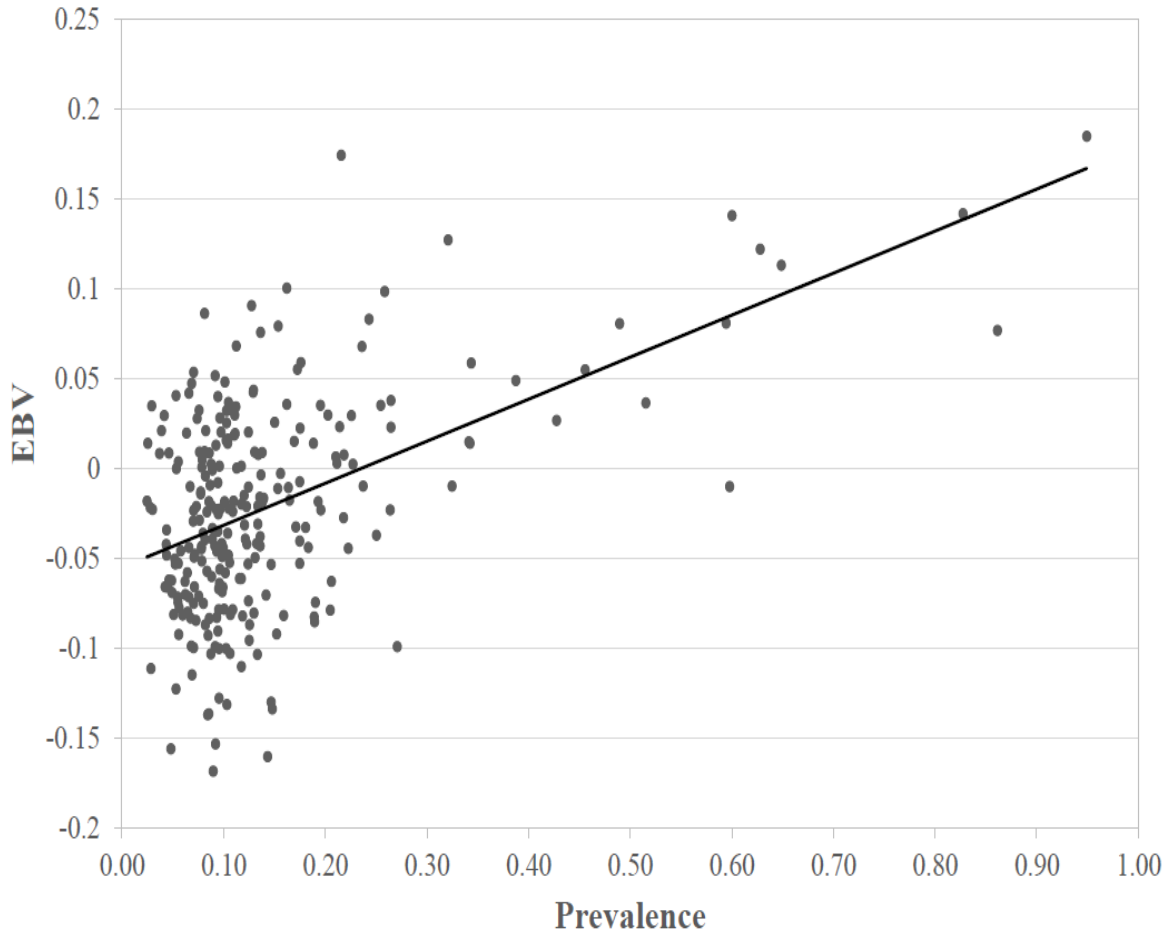
daughter prevalence was >0.5 was estimated. The correlation for the population with a daughter prevalence <0.5 was 0.06 ($P<1x10^{-7}$) and the correlation for the population with a daughter prevalence >0.5 was 0.23 (P=0.07). The distribution of the EBVs for the entire population is shown in figure 2.3. A summary of data used to plot figures 2.1-2.4 is shown in table 2.2.

Figure 2.1 Distribution of mean bTB prevalence in daughters of sires with bTB data.*



*Only sires with more than 50 daughters from more than 10 herds are represented. A wide variation in mean bTB prevalence in daughters of sires was observed

*Figure 2.2 Relationship of sire estimated breeding value (EBV) ** and mean prevalence of bTB infection (Prevalence)*.*



*mean prevalence calculated using daughters of sires.
**Only sires with more than 50 daughters from more than 10 herds are represented.
An increase in EBV correlated to an increase in herd prevalence.

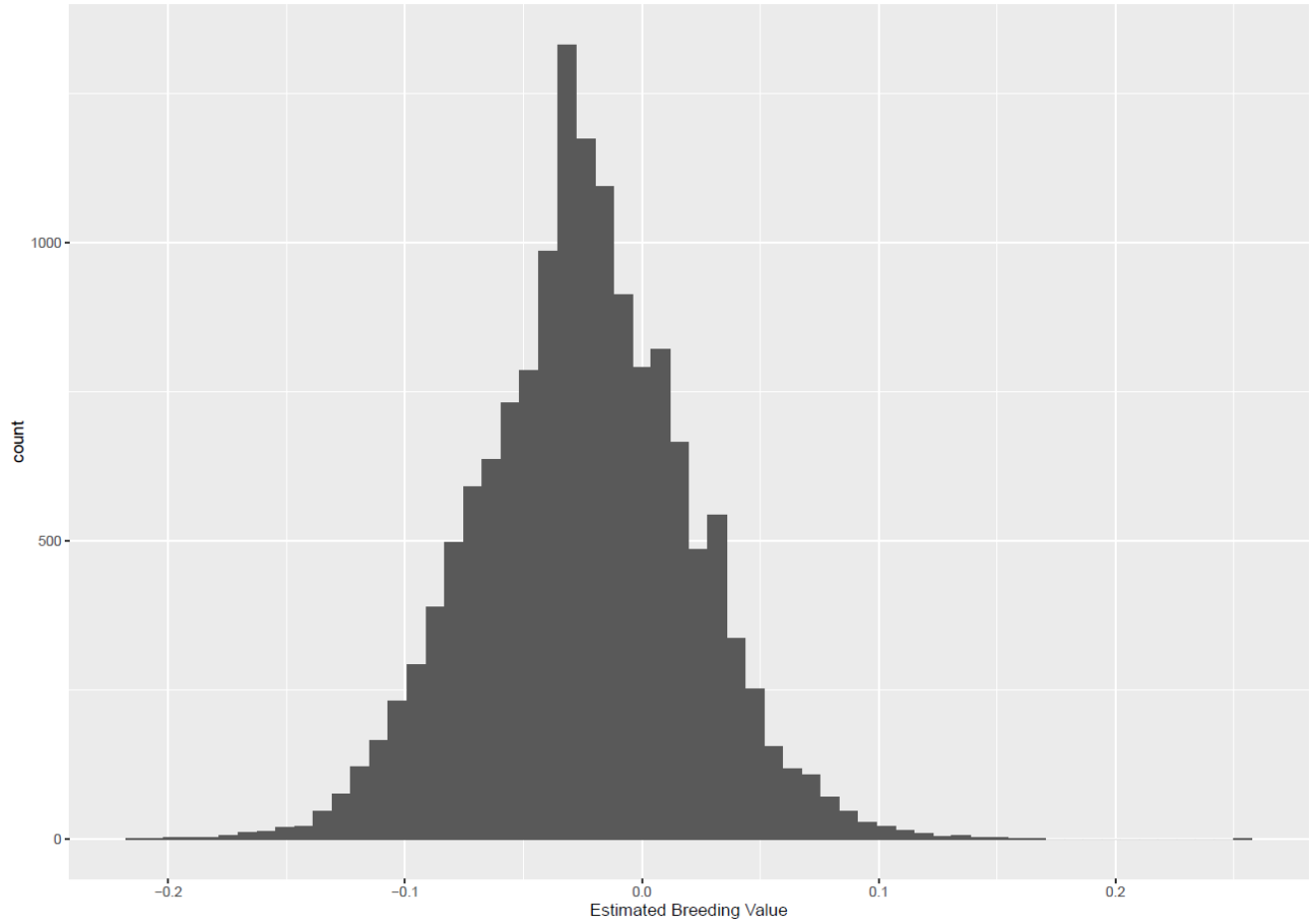


Figure 2.3 Distribution of estimated breeding values in the population.

Table 2.2 Summary of data used to produce figures 2.1-2.4

	Year of Birth	EBV (using all data)	Reliability (all data)
Minimum	1953	-0.3608	-0.20187
1st Quartile	1996	-0.03479	0.09806
Median	2000	-0.01246	0.2284
mean	1998	-0.01076	0.23796
3rd Quartile	2003	0.01252	0.33628
Max	2009	0.3085	0.98973
	# Offspring	# Daughters (cows)	# Daughters (Heifers)
Minimum	1	0	0
1st Quartile	1	0	0
Median	2	1	0
mean	12.59	7.205	3.887
3rd Quartile	7	3	2
Max	8178	6104	1552
	# offsprint bTB negative	# Cows bTB positive	# Cows bTB negative
Minimum	0	0	0
1st Quartile	1	0	0
Median	2	0	1
mean	11.04	1.087	6.117
3rd Quartile	6	1	2
Max	7337	667	5437
	# steers bTB positive	# steers bTB negative	Prevalence (all offspring)
Minimum	0	0	0
1st Quartile	0	0	0
Median	0	0	0
mean	0.1335	1.361	0.1772
3rd Quartile	0	0	0.1875
Max	52	470	1
	EBV (using daugther data)	Reliability (Daughter data)	# sons (steers)
Minimum	-0.2128	-0.24521	0
1st Quartile	-0.05368	0.03237	0
Median	-0.02526	0.14968	0
mean	-0.025	0.17404	1.495
3rd Quartile	0.0043	0.29007	1
Max	0.2549	0.98188	522
	# offspring bTB positive	# heifers bTB positive	# heifers bTB negative
Minimum	0	0	0
1st Quartile	0	0	0
Median	0	0	0
mean	1.543	0.322	3.565
3rd Quartile	1	0	2
Max	841	122	1430
	Prevalence (All daughters)	# Herds Sire has offspring	
Minimum	0	1	
1st Quartile	0	1	
Median	0	1	
mean	0.231	4.8	
3rd Quartile	0.333	0	
Max	1	1744	

2.3.2 Breed effects

From the linear model using the full dataset, Holstein-Friesian sires had the lowest mean EBV compared to all other breeds (for sires born from 1992 onwards; Table 2.3), which indicates a lower average susceptibility to bTB. Simmental and Charolais sires had the greatest mean EBV (i.e. most susceptible to bTB infection) compared to all other breeds. The average EBV (weighted against the number of sires in each breed against the total number of sires across all breeds), across beef breeds was 0.025 when considering sires with more than 10 daughters, which was greater than the weighted average of -0.012 obtained across dairy breeds when considering only sires with more than 10 daughters. The distribution of EBV per breed is shown in figure 2.4.

*Table 2.3 Per breed mean estimated breeding values for susceptibility to bovine tuberculosis infection**

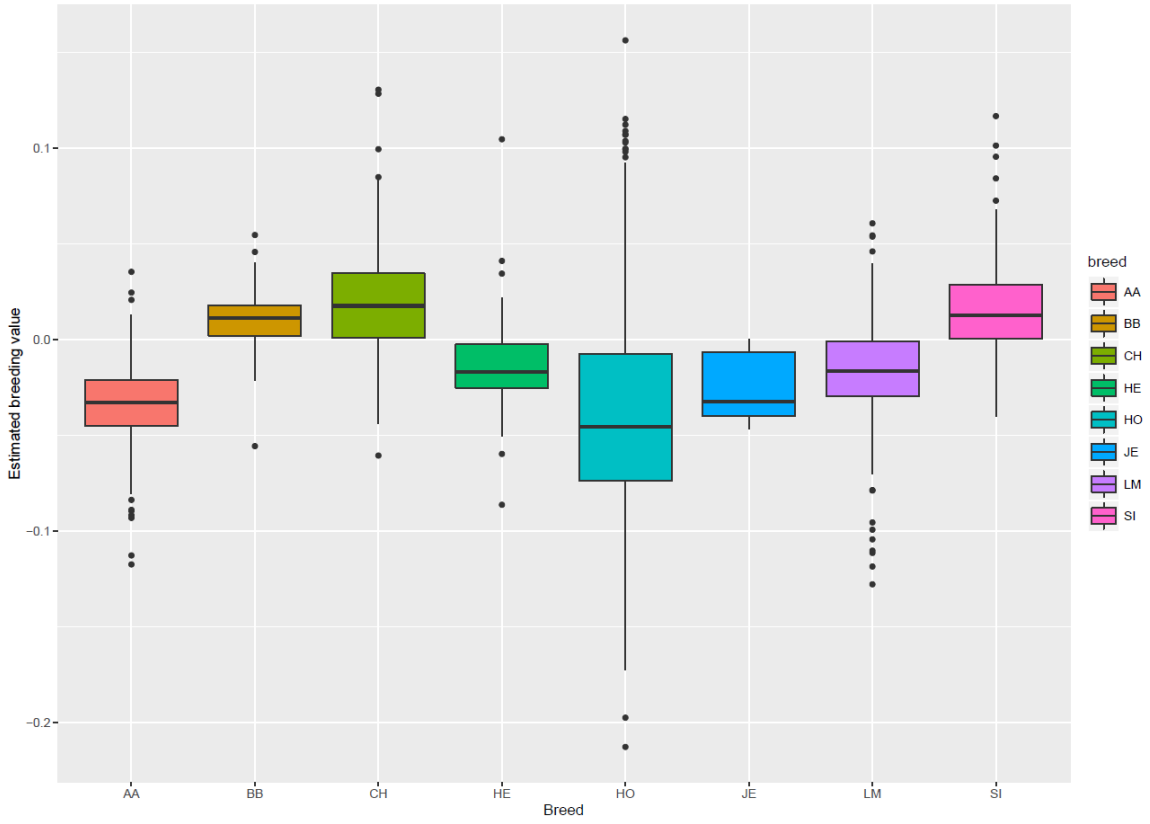
	All sires		**Sires with ≥ 10 progeny**	
Breed	**N**	**EBV (se)**	**N**	**EBV(se)**
Aberdeen Angus	379	0.005 (0.037)^{D}	27	0.005 (0.050)^{D}
Belgian Blue	102	0.042 (0.029)^{B}	11	0.047 (0.054)^{BAC}
Charolais	513	0.049 (0.034)^{A}	48	0.050 (0.044)^{BA}
Hereford	257	0.017 (0.028)^{C}	11	0.017 (0.038)^{BDC}
Holstein-Friesian	4977	-0.008 (0.045)^{E}	959	-0.013 (0.0051)^{E}
Jersey	84	0.015 (0.028)^{C}	21	0.021 (0.036)^{BAC}
Limousin	536	0.016 (0.035)^{C}	50	0.010 (0.059)^{DC}
Simmental	245	0.052 (0.033)^{A}	16	0.053 (0.043)^{A}

**for all sires and sires with more than 10 progeny.*

Number of sires (N) and mean estimated breeding value (EBV; standard error (se) in parentheses [1].

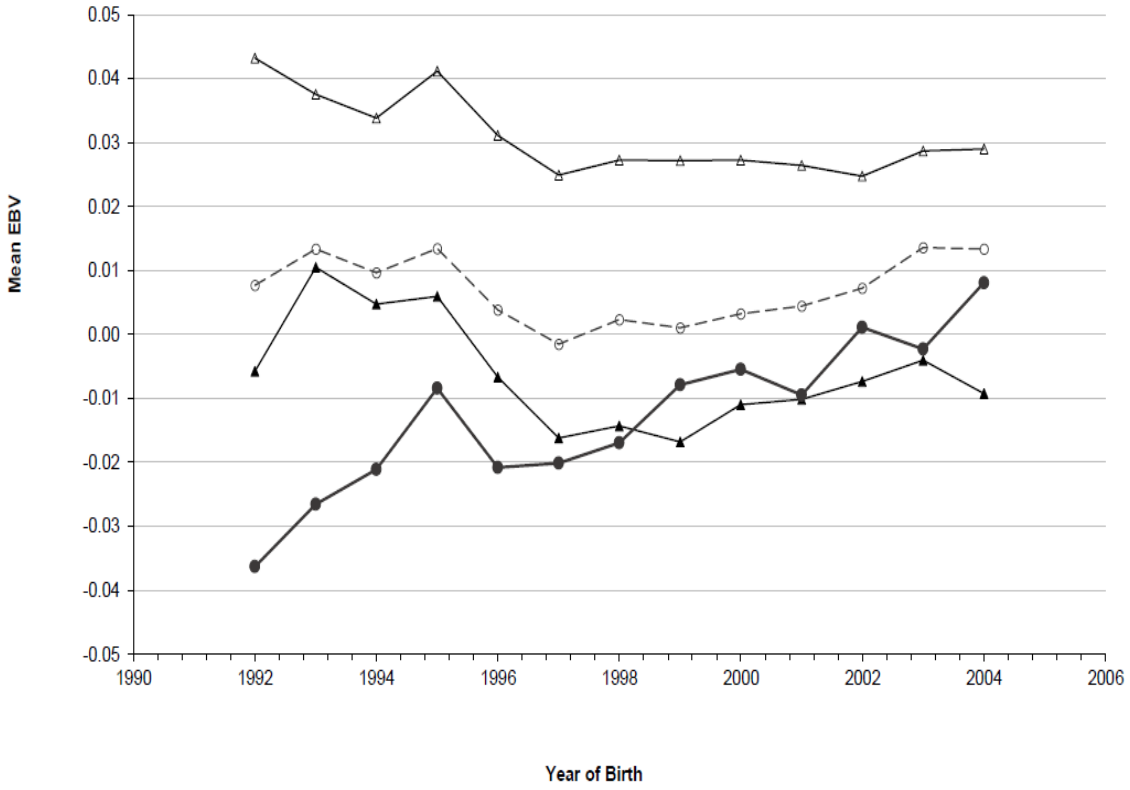
[1]EBV with the same superscripted letter are not significantly different from each other.

Figure 2.4 Distribution of estimated breeding value per breed.



Genetic trends of the mean EBV of sires born between 1992 and 2004 are in Figure 2.5. Since 1992, the mean EBV was greater in beef sires than in dairy sires. The genetic susceptibility to bTB infection in beef and dairy animals has increased since 1999 but only the increase in dairy breeds was significant ($P < 0.05$).

Figure 2.5 Mean estimated breeding values (EBV) for bTB susceptibility against year of birth.



All sires (clear dot), sires with more than 10 daughters (solid dot), dairy breeds (solid triangle) and beef breeds (clear triangle).

2.3.3 Genotype by environment effects

Summary statistics for bTB susceptibility across different bTB prevalence categories using a linear and threshold animal model are in Table 2.3. Using a linear animal model, the estimate of heritability of susceptibility to *M. bovis* infection was greatest for animals within herd-episodes of very high prevalence of bTB infection (> 0.75 prevalence) and lowest for those that had experienced herd-episodes of very low prevalence (< 0.25 prevalence). Using a threshold animal model, the estimate of heritability was lowest for animals that had experienced herd-episodes of very low prevalence of bTB infection (< 0.25 prevalence) and greatest for those that had suffered

herd-episodes of very high prevalence (> 0.75). Genetic SD for bTB susceptibility, estimated using the linear animal model, was lowest for the extreme bTB prevalence categories, with the genetic SD of very low (< 0.25 prevalence) and very high (> 0.75 prevalence) prevalence categories approximately equal to half that of the low (0.25 to 0.5 prevalence) and high (0.5 to 0.75 prevalence) prevalence categories (Table 2.3). The genetic SD for bTB susceptibility, estimated using the threshold animal model, increased greatly when the prevalence of bTB infection was higher than 0.75. After transforming heritability estimates from the linear model to the liability scale using the Robertson transformation, estimates increased as the prevalence of bTB infection increased from very low (< 0.25) to very high (> 0.75) prevalence, with a pronounced increase in heritability estimates in herds with a very high (> 0.75) prevalence.

Table 2.4 Results and summary statistics for univariate analyses for susceptibility to bovine tuberculosis infection across disease prevalence's.

	V. Low	**Low**	**High**	**V. High**
N	87,609	16,812	3,082	3,359
Mean	0.07	0.34	0.62	0.93
σ_g (LAM)	0.07^{A}	0.16^{B}	0.12^{A}	0.09^{C}
Heritability (LAM)	0.07 (0.008)	0.16 (0.024)	0.08 (0.048)	0.25 (0.072)
σ_g (TAM)	0.71	0.89	0.83	1.39
Heritability (TAM)	0.13 (0.012)	0.19 (0.016)	0.17 (0.037)	0.37 (0.081)
Transformed	0.25	0.27	0.31	0.88

Number of records (N), mean prevalence (Mean), estimates of the genetic standard deviation (σ_g[1]) and heritability from the linear animal model (LAM) and the threshold animal model (TAM) and based on the standard Robertson transformation (Transformed), for environments with very low (<0.25), low (0.25-0.5), high (0.5-0.75) and very high (>0.75) prevalence of bTB infection
[1]Genetic variance estimates with the same superscripted letter were not different (P<0.05) from each other.

Estimates of genetic correlations between susceptibility to bTB infection in different environments were all positive (Table 2.4) and ranged from 0.06 (environments with low and very high bTB prevalence) to 0.86 (environments with low and very low bTB prevalence). All genetic correlation estimates were less than 1 ($P < 0.05$) and estimates of the genetic variance in different environments also differed from each other ($P < 0.05$) with the exception of the very low (<0.25) and high (0.5-0.75) bTB prevalence environments (Table 2.3).

Table 2.5 Estimates of genetic correlations[1] for susceptibility to bTB infection between environments of differing disease prevalence.

	V. Low	**Low**	**High**
Low	0.86 (0.043)		
High	0.65 (0.114)	0.84 (0.133)	
V. High	0.08 (0.107)	0.06 (0.147)	0.31(0.253)

Very low (< 0.25), low (0.25 to 0.5), high (0.5 to 0.75) and very high (> 0.75) bTB prevalence.
[1]All genetic correlations differed (P < 0.05) from 1.

2.4 Discussion

The contribution of breeding programs to phenotypic gains in animal performance has been well documented [96, 97]. Hence, a breeding program aimed at improving resistance to bTB can form an integral component of a national eradication program if heritable variation in bTB exists and can be exploited. The economic importance of bTB in Ireland is well established [86] and the centrally stored phenotypic data on bTB status at the individual level could be used to estimate breeding values. Therefore, the objective of our study was to quantify the genetic variation of the susceptibility to bTB in cattle. A novel component of this study was the estimation of genetic variation for bTB susceptibility in beef cattle.

2.4.1 Variance components

In this study on dairy and beef cattle, a heritability of 0.11 for susceptibility to bTB was estimated using a linear model, which is between documented estimates of most production diseases, i.e. < 0.05 in [<0.05; 97], and animal performance traits, i.e. ~ 0.30; [Circa 0.30; 98, 99, 100]. The heritability estimates of susceptibility to bTB reported here, especially those for dairy cows, are consistent with previous linear model estimates for both Irish dairy [17] and UK dairy [81] cows. Moreover, the reported heritability estimates of bTB, including those from the present study, are similar to those documented for paratuberculosis (i.e., *Mycobacterium avium*) [101-103], which shares some epidemiological similarities with bTB [104]. Furthermore, our linear model heritability estimates of susceptibility to bTB did not vary considerably across the populations

investigated (beef, dairy, cows, heifers, steers); differences in estimates were due to a combination of differences in residual and genetic variances between populations.

Despite the similarities in heritability estimates between populations the genetic variance estimates for susceptibility to bTB reported here were two to three times greater than those documented by Bermingham et al. [17] in a sub-population of the animals used here, i.e. that was limited to dairy cows and heifers. The greater genetic variance reported here could be due to the greater diversity of breeds included. Our estimates of genetic variance for susceptibility to bTB were also three to six times greater than those estimated for Johne's disease in a group of 4603 US Holstein-Friesian cows [102], which was also treated as a binary trait and estimated using a linear model.

2.4.2 Breed effects

Previously reported data on between-breed differences in susceptibility to bTB corroborate our results. For example, Ameni et al. [105] reported a greater susceptibility and severity in the pathology of bTB in Holsteins compared to East African Zebu and Zebu-Holstein crossbreds. Other studies also documented significant breed differences for many health traits, such as somatic cell score [106], lameness [107, 108], and bovine respiratory disease [109].

The observed recent genetic trends in susceptibility to bTB (Figure 2.3) suggest that, although the breeding goals for Irish dairy cattle are to select for longevity, fertility and milk production [110], genetic merit for resistance to bTB appears to be deteriorating in dairy cattle. Although the estimate of the genetic trend for susceptibility to bTB in beef

cattle was not different from zero, genetic resistance to bTB has not improved since the 1990s (Figure 2.3). It is therefore important to consider including resistance to bTB in the Irish national breeding goals, especially for dairy cows. Another recommendation is that other countries also evaluate genetic trends for susceptibility to bTB in their cattle populations and, where necessary, take remedial action. The contribution of breeding programs for improving genetic merit of breeds for animal health traits has been clearly documented for Scandinavian dairy cow populations [111, 112]. In countries like Ireland, where routine testing of animals for bTB status is undertaken, it should be possible to generate accurate genetic evaluations for bTB to inform better selection decisions. Some fears have been expressed that breeding for disease resistance will lead to (or otherwise be associated with) reduced responsiveness to the SICTT. This concern was considered in detail by Bishop and Woolliams [113] and they concluded there was no evidence that such a scenario would arise.

2.4.3 Genotype by environment effects

The genetic variance of susceptibility to bTB in environments with different bTB prevalence levels has not previously been analyzed. The pattern of genetic variances of susceptibility to bTB in environments with different bTB prevalence levels, estimated using an animal linear mixed model, was similar to the trend in genetic variances estimated for somatic cell score in environments with different herd average somatic cell scores [114], i.e. it was lowest for the most extreme environments. However, the genetic variance pattern that we observed was in direct contrast to the pattern of genetic variance estimates reported for Johne's disease in Dutch Holstein-Friesian cows [115], which increased with

the herd's prevalence for Johne's disease. However, the pattern of genetic variances of susceptibility to bTB estimated with a threshold animal model in our study were similar to those estimated for Johne's disease [115] and increased with herd prevalence (Table 2.3). Estimates of variance components in the different prevalence groups were similar to van Hulzen et al. [115], using a linear model, with the exception of the very high prevalence environment. Furthermore, estimates of the genetic variance for the low and high prevalence groups were greater than the estimate for the very low prevalence groups, both with the linear and threshold animal models. An inverse relationship existed between estimates of genetic variance and heritability using a linear model, across the different prevalence groups, while one would expect a positive relationship [113]. This could be due to the smaller number of animals in the more extreme environmental categories or selection/culling; to ensure that there was no impact of the estimation algorithm on estimates of the variance components, we switched the 0s and 1s in the high prevalence environment using the linear model, but this did not impact the estimated variance components. Although heritability estimates differed between the threshold animal model and the linear model estimates transformed to the liability scale, the heritability of susceptibility to bTB in the very high prevalence herds (> 0.75) was always greatest, regardless of the model used. However, due to paucity of data available on animals within herd-episodes of very high bTB prevalence (> 0.75) the heritability estimate transformed to the liability scale will be overestimated.

The estimates of the genetic correlation of susceptibility to bTB between cows and heifers from our study (0.55) was similar to the estimate of Bermingham et al. [17] (0.53). While differences in genetic variance for disease susceptibility between females and males

have been documented, e.g., for tick infection in Australian beef cattle [116], there is little information on genetic correlations for disease susceptibility between sexes. However, estimates of genetic correlations between performance traits in different sexes have been reported. For example, Pabiou et al. [100] reported genetic correlations that ranged from 0.54 to 0.81 for wholesale carcass cut traits in steers and heifers, while Stalhammar and Philipsson [117] documented genetic correlations for post-weaning and weanling gain in male and female Swedish beef cattle that ranged from 0.4 to 1.0. Moreover, genetic correlations between the same performance traits for animals of different ages have also been reported, for example, genetic correlations of animal weight at weaning, post-weaning weight and weight of cows were found to range from 0.16 to 0.79 [118]. These results corroborate estimates of genetic correlations for susceptibility to bTB being less than 1 reported here. These less than 1 genetic correlations between animal types ($P < 0.05$), indicates that the bTB phenotype is possibly under different genetic control in the cows, heifers, and steers.

Estimates of genetic correlations of susceptibility to bTB between environments with different bTB prevalence levels were weakest for the environment with the greatest pathogen load (i.e., > 0.75 bTB prevalence). This suggests that the genetic background of susceptibility to bTB in environments under a very high pathogen load is different to that in environments under a lower pathogen load. Calus et al. [119] observed a similar phenomenon when they investigated the genetics of somatic cell score in Dutch Holstein cows across environments (i.e., herds) based on different mean bulk tank somatic cell scores. Furthermore, these authors found the weakest genetic correlations for animals in environments with the highest somatic cell counts, when compared to all other

environments. The weak genetic correlations for bTB observed within the environment with the greatest prevalences of bTB could be due to the extremely high pathogen load, which could constitute a very different context in terms of genes or biological systems (e.g., innate, adaptive) that contribute to whether or not the animal becomes infected with bTB. Another explanation could be the high likelihood that all animals in such an environment are exposed to the pathogen, thus the susceptibility to bTB in this environment, as defined here may better reflect the true phenotype for bTB susceptibility. The genetic variance of susceptibility to bTB was also low in the environment with the greatest bTB prevalence, but its heritability was greatest. The high pathogen and the high heritability (0.25) for bTB in the environment with the greatest bTB prevalence (>0.75) is somewhat analogous to an experimental inoculation of bTB and is closer to the heritability of 0.48 reported for deer experimentally infected with *M. bovis* [0.48; 120]. It must be noted however, that this heritability estimate of susceptibility to bTB in deer is considered to be high due to the naivety of the deer to TB. Despite this, it could be suggested that the heritability of susceptibility to bTB may actually be greater than the value of 0.11 reported here across all data once the assumption that all animals are equally exposed to the pathogen is fulfilled.

2.4.4 Methodological strengths and weaknesses

A number of factors that contribute to the under-estimation of the heritability of disease susceptibility in livestock have been documented [113] using simulated data. Bishop and Woolliams [113] identified three factors which contribute to the underestimation of the heritability of disease susceptibility in livestock:

(1) Incomplete exposure of animals to the pathogen, resulting in some animals not having the opportunity to express their genotype for the disease. The degree of underestimation is linear with the decreased probability of complete exposure. The strict inclusion criteria for herd-episodes imposed in the present study aimed at maximizing the probability of complete exposure. However, complete exposure may not have been reached in some cases. With the observed prevalence of bTB of 0.11 in the entire dataset (Table 2.1), according to Bishop and Wolliams [115] equations, an exposure probability of 50% would have been needed in our analysis to estimate a sufficiently accurate heritability of susceptibility to bTB.

(2) Incomplete sensitivity and specificity of the diagnostic tests used to classify animals as healthy or diseased. Bermingham et al. [121] showed that the effects of imperfect SICTT sensitivity resulted in an underestimation of the heritability of susceptibility to bTB in Irish and UK dairy cows. Furthermore, Bishop and Woolliams [113] concluded that imperfect sensitivity to SICTT has a large effect on the estimate of heritability of susceptibility to bTB, especially when the disease prevalence is as low as 0.10, similar to that observed here. It is also well-documented that the standard SICTT can detect only between 40 and 80% of infected animals [122, 123]. In our study, to account for the sensitivity of SICTT, only the herd-episodes for which 20% of 'reactors' presented bTB lesions were included, which implies that bTB infection was reasonably established in these herd-episodes.

(3) Misclassification of animals due to the dynamic expression of the disease over the course of an infection. With bTB infection, many animals will not present the infection for some time [92] and may therefore not be identified when testing is done. Furthermore,

during the period of bTB infection, animals within a herd will be at different stages of infection. Animals at the earlier stages of infection will be difficult to detect (since sensitivity to testing is low), while animals at later stages will be easier to detect. This misclassification will cause underestimation of the heritability, analogous to the effect of imperfect test sensitivity. To reduce the probability of misclassification because of dynamic expression, a number of standard testing procedures for bTB are carried out by the Irish Department of Agriculture, Food and Marine, including annual testing of all animals in herds and continuous surveillance of animals at slaughter. Re-testing of herds after the identification of a 'standard reactor' or an 'inconclusive reactor', at two-month intervals also aims at reducing the likelihood of misclassifying animals due to this effect.

2.5 Conclusions

For a trait to be included in a breeding goal it must have some relevance (e.g., economical or social), be measurable or correlated with a measureable phenotype, and exhibit genetic variation. Bovine tuberculosis is of economic importance [55, 86], which justifies its consideration for inclusion in the breeding goal. The annual testing of whole herds for bTB and the storage of these data in a national database, as well as access to pedigree information, animal movements and other systematic environmental effects pertaining to the test (e.g., herd, season, age of animal), imply that routine genetic evaluations are possible for bTB if genetic variation in susceptibility to bTB exists. Results from this study clearly show that exploitable genetic variation in susceptibility to bTB exists. Moreover, the current genetic trend for sires with more than 50 progeny with bTB data in the national population, suggests that susceptibility to bTB infection in the progeny

of these animals may be deteriorating. All these points constitute a strong argument to consider increased resistance to bTB as a trait for selection in Irish dairy and beef breeding and part of the national strategy for bTB eradication in cattle. Moreover, its usefulness in breeding strategies for other cattle populations that are not free of bTB should be investigated. While fears have been expressed that selection for bTB resistance will reduce sensitivity to SICTT testing, it is thought that such a scenario will not arise. Although some estimates of the genetic correlation of susceptibility to bTB between sub-populations were less than 0.8, which means that it cannot be considered as the same trait across these sub-populations, performing separate genetic evaluations for each animal type or for different environments in a multi-trait genetic evaluation may be impractical. Moreover, since all genetic correlation estimates between the different sub-populations were positive, a genetic evaluation using all animal types and environments combined is expected to improve genetic merit across all sub-populations.

The following chapter was published in the journal

Genetic Selection Evolution, under the title

"A genome wide association study for genetic susceptibility to *Mycobacterium bovis* infection in dairy cattle identifies a susceptibility QTL on chromosome 23."

Authors were

Ian W Richardson, Donagh P Berry, Heather Wiencko, Isabella M Higgins, Simon J More, Jennifer McClure, David J Lynnand Daniel G Bradley

Authors' contribution

IWR made substantial contribution to the conception and design of the study, to the drafting of the manuscript, and undertook the GWAS analysis, interpretation and editing of data.
DPB made substantial contribution to the concept, design and interpretation of data, and to the drafting and review of the manuscript.
HW undertook the network analysis and contributed to the drafting and review of the manuscript.
IMH collected the data and contributed substantially to the interpretation of data. SJM was largely involved in the drafting and revising of the manuscript for important intellectual content.
JM and DJL were involved in revising the manuscript for important intellectual content.
DGB made substantial contribution to the concept, design and interpretation of data, and to the drafting and review of the manuscript. All authors read and approved the final manuscript.

Chapter three - Genome wide association study for genetic susceptibility to *Mycobacterium bovis* infection in dairy cattle.

3.1 Introduction

Both within [17, 80, 81] and between [80, 105] breed variation in susceptibility to bTB in dairy and beef cattle has been previously reported. Documented heritability estimates for susceptibility to bTB in cattle range from 0.08 to 0.19 [17, 80, 81] as reported in chapter two. Heritability estimates for Johne's disease (i.e., *Mycobacterium avium* ssp paratuberculosis), which shares some epidemiological similarities with bTB [104], range from 0.07 to 0.15 [101-103].

Although several genome-wide association studies have been undertaken for Johne's disease in cattle [124-126], only three genome wide association studies have emerged for bTB [127-129]. A medium density genome wide association study (GWAS) was performed by Finlay et al. [127] in a population based on the progeny of 307 Holstein-Friesian bulls. This study proposed one genomic region on BTA 22 as being putatively associated with bTB susceptibility[127]. A high density genome wide association study performed by Bermingham et al. [128] used a population of 1151 Holstein-Friesian cows (592 cases and 559 age-matched controls). These authors identified two genes putatively associated with bTB: *PTPRT* on BTA13 and *MYO3B* on BTA2. The SNPs that were identified as significantly associated with bTB susceptibility by Finlay et al. [127] were however not validated by Bermingham et al. [128] and *vice versa*. Kassahun et al. [130] used an admixture mapping approach with a low density marker set in African-European hybrid field populations to find some evidence of association relative susceptibility and the toll-like receptor region on chromosome 6.

Several genome wide association studies for susceptibility to human tuberculosis (TB, caused by infection with *M. tuberculosis*) have also been performed in human populations [131], with several genes being identified and validated in multiple studies, including the genes *NRAMP1, IFNG, NOS2A, MBL, VDR* and four toll-like receptors [131]. A whole genome scan performed in mice experimentally infected with *M. tuberculosis* [132] identified loci in the region of the *sst1* gene that is involved in the immune response of macrophages to TB [131]. Several single gene investigations to bTB related traits have been performed, with *NRAMP1* (*SLC11A*) being identified as a susceptibility locus in both cattle and humans [133-135].

The objective of the present study was to detect regions of the bovine genome associated with bTB susceptibility in Irish dairy cattle and elucidate any likely candidate genes and pathways contributing to these putative quantitative trait loci (QTLs). High density genotypes of artificial insemination dairy bulls (n=841) were associated with estimated breeding values (EBV) for bTB susceptibility calculated from epidemiological information on 105,914 daughters. This information was used in two distinct approaches to GWAS and a haplotype association approach, to search for QTLs associated with bTB susceptibility in dairy cattle.

3.2 Materials and Methods

3.2.1 Phenotypic data

Individual animal EBVs for bTB susceptibility were calculated using an animal linear mixed model in ASREML [15]. The data, statistical model and variance components used to estimate the EBVs were outlined in detail in chapter two. In brief, the dependent trait was the single intra-dermal comparative tuberculin test (SICTT) result [91, 92] (i.e., binary trait) from individual animals present during one of 4,048 herd bTB episodes (periods of herd restriction triggered by detection of bTB infection in one or more animals) in 3,240 Irish dairy herds. A number of data editing criteria, described in detail in chapter two, were implemented to maximize the likelihood of equal bTB exposure among all animals that were enrolled from each bTB episode. After edits, 105,914 cow single intra-dermal comparative tuberculin skin test records remained.

A univariate animal linear mixed model with the binary cow SICTT result as the dependent variable was used to estimate EBVs for bTB susceptibility. Fixed effects included in the statistical model were herd-episode, heterosis and recombination loss coefficients of the animal, parity and stage of lactation [80]. Individual animal was included in the model as a random effect and founder animals were assigned to different breed groups. Relationships among animals were considered through the use of a numerator relationship matrix. Only EBVs of sires with a reliability of at least 0.30 were retained for use in the GWAS.

Estimated breeding values for sires from the animal model were deregressed as:

$$\tilde{y} = R(R^{-1} + A^{-1})\hat{a}$$

where $\tilde{y}$ is a vector of deregressed EBVs, R is a diagonal matrix containing one divided by the animal's reliability from his progeny less one, A is the numerator relationship matrix among animals and $\hat{a}$ is a vector of the EBVs.

3.2.2 Genotypes

Illumina high density SNP genotypes (n=777,962 SNPs) were available on 770 Holstein-Friesian bulls. All animals had a genotype call rate >95%. Where genotypes were available on both sire and son(s), the integrity of the animal genotype identity was confirmed based on the extent of Mendelian inconsistencies between sire-son pairs. A total of 3,554 autosomal SNPs had >2% Mendelian inconsistencies and were discarded. The genotype of both the sire and son were set to missing for the remaining autosomal SNPs where sporadic Mendelian inconsistencies existed.

A total of 1,574 SNPs with an Illumina GenTrain score <0.55 were also discarded as were 40,934 non-autosomal SNPs; duplicate SNPs on the Illumina high density manifest were also discarded. Of the remaining SNPs, 17,273 with a call rate of <95% were discarded along with a further 83,833 SNPs with a minor allele frequency <0.02; finally any SNP that deviated ($P<0.1x10^{-8}$) from Hardy-Weinberg equilibrium was excluded.

Following all edits, 597,144 autosomal SNPs remained for inclusion in the analysis. Missing genotypes were imputed using Beagle [29, 136].

Illumina Bovine50 beadchip genotypes (i.e. 54,001 SNPs) were available on 5,313 Holstein-Friesian dairy bulls; 639 of these animals were part of the 770 animals with high density genotypes. Animal and SNP editing criteria imposed were as previously described. Following all edits, 45,239 SNPs remained. Genotypes were imputed to high density for each chromosome separately using BEAGLE [29, 136, 137] (BTau 4.6.1 reference build) and the available 639 high density genotypes as reference. A total of 579 Holstein-Friesian bulls with high density genotypes and an additional 262 Holstein-Friesian bulls with imputed genotypes had EBVs for bTB susceptibility with reliabilities greater than 0.30. These 841 bull SNP genotypes (597,144 SNPs) were used for the genome wide association scan for bTB susceptibility.

3.2.3 Association analyses

Two alternative strategies were used to test for genome wide association with bTB susceptibility: a single SNP regression approach and a Bayesian approach. Prior to inclusion in the association analyses, the deregressed EBVs were weighted using the formula outlined by Garrick et al., [21].

$$\omega_i = \frac{1-h^2}{\left[c + (1-r_i^2)/r_i^2\right]h^2}$$

Where ω_i is the weighting factor of the i^{th} EBV, h^2 is the heritability estimate for bTB susceptibility (here assumed to be 0.14; chapter two), r_i^2 is the reliability of the i^{th} EBV and c is the proportion of genetic variance not accounted for by the SNPs. Weighting factors were calculated with a c value of 0.10 for the Bayesian approach and 0.90 for the single SNP regression approach.

3.2.4 Single SNP regression method for genome wide analysis

SNP effects for bTB susceptibility were estimated using mixed models implemented in WOMBAT [138, 139]. Weighted deregressed EBVs were the dependent variable and individual animal was included as a random effect with relationships among animals accounted for *via* the pedigree relationship matrix. Allele dosage for each SNP was included separately in the model as a fixed effect. Population structure was accounted for by a numerator relationship matrix (which describes the genetic relationships between individuals of a population), which Wombat calculates from inclusion of the populations pedigree information. Significance levels for each SNP were calculated from the resulting t-statistic assuming a two tailed t-distribution. SNPs with a P-value of $<1x10^{-5}$ were considered to be genome wide significant based on a false discovery rate [25] of 1%. The percentage of genetic variance that a SNP accounted for was calculated using $2pqa^2$, where p was the major allele frequency, q the minor allele frequency and a the allele substitution effect.

3.2.5 Bayesian approach

Estimation of SNP effects for bTB susceptibility using a Bayesian model was implemented in the GenSel software package [46]. Weighted deregressed EBV was the dependent variable. A BayesC model averaging approach as described by Kizilkaya et al. [140] was first implemented to estimate the genetic and residual variances for bTB. BayesC assumes a common variance for all SNPs in the model [46]. Starting values for the genetic and residual variance for BayesC were from a linear mixed model implemented in ASREML [15].

A BayesB algorithm, as described by Hayes and Goddard [43], was subsequently invoked to estimate SNP effects. The genetic and residual variances used for bTB were those estimated with the BayesC algorithm. Population structure was accounted for by a numerator relationship matrix as with the single SNP regression approach. The posterior probability that a SNP was not associated with susceptibility to bTB infection (π) was assumed to be 0.999. A burn in of 10000 and chain length of 41000 were applied in both BayesC and BayesB algorithms. Bayes factors (BF) for each SNP were calculated from the BayesB analysis as an alternative to classical frequentist hypothesis testing. These were calculated as [141]:

$$BF_i = (f_i/(1-f_i))/(\pi_1/\pi)$$

where f_i is the proportion of times SNP i was included in the model, π is the posterior probability that SNP i is not associated with bTB susceptibility, and π_1 is the posterior probability that SNP i is associated with bTB susceptibility (represented as 1- π). Bayes factors above 3.2 are considered 'substantially' associated, those above 10 are 'strongly'

associated, and Bayes factors above 100 are 'decisively' associated with a trait [48]. The percentage of genetic variance accounted for by each SNP was calculated using GenSel.

3.2.6 Defining quantitative trait locus regions

Quantitative trait locus regions of interest were defined using the local linkage disequilibrium (LD) structure surrounding each SNP that had a genome wide association significance level of $P < 1x10^{-5}$ or a Bayes factor over 10. Pairwise LD as measured by r^2 between all SNPs within 5 Mb upstream and downstream of significant SNPs was calculated using PLINK [142]. Within this 10 Mb region, a LD level of $r^2 \geq 0.5$ was applied and the SNPs furthest upstream and downstream of a significant SNP with this level of LD were noted and used to form the beginning and end of a QTL. QTLs which overlapped one another were combined into a single QTL. These QTLs were then investigated for the presence of genes within these regions.

3.2.7 Network analysis

Genes identified within QTLs can be used to construct a biological network from available biological network databases, such as InnateDB [143]. This network can then be investigated to identify over-represented pathways/ontologies in the network or sub-networks. Network analysis was performed separately on the gene lists, derived from the single SNP regression (SSR) or the Bayesian analyses. Genes identified within QTLs were sent as a query to the InnateDB website [143] to return lists of interacting networks that were over-represented in any pathways possibly associated with bTB susceptibility. The

analysis was performed using the hypergeometric algorithm and the Benjamini Hochberg correction for multiple hypotheses testing.

3.2.8 Identity-by-descent analysis

A subset of prolific artificial insemination (AI) sires (with ≥50 daughter from ≥10 herds) identified in chapter two as having mean daughter prevalence of bTB infection >60% (n=11) were used as a case group to investigate the presence of identical by descent (IBD) haplotypes that may be associated with bTB susceptibility. The control group used were a subset of the most prolific AI sires identified in chapter one that had a mean daughter prevalence of bTB infection <30% (n=154). Genotypes of the 11 cases and 154 controls were phased using BEAGLE [29, 136] and pair-wise IBD segments were called using fastIBD, an algorithm implemented within BEAGLE. The proportion of IBD shared within the case and within the control groups was calculated by dividing the total IBD shared per segment by the number of total possible pairs in the case or control groups (i.e., $n*(n-1)/2n$), where n is the number of animals within each group. The case group IBD proportions per segment were normalized against the control group IBD proportions per locus by

$$\frac{(X-Y)^2}{2X}$$

where X is the case-case IBD proportions and Y is the control-control IBD proportion, per segment.

3.2.9 Targeted imputation of sequence data in candidate regions

A 3 Mb window, flanking QTL regions that were identified as significant (P $<1x10^{-5}$, Bayes factor > 10 and normalized proportion of IBD shared > 0.5) across all of our three analyses were imputed from high density (HD; 597,144 SNPs) up to full sequence. Imputation was performed using BEAGLE [144], using all 1,147 individual sequences from the 1000 bull's genome project data [145] as the reference population. The 1000 bull's data covered 24 pure breeds and 11 crossbred animals. The majority of animals within the 1000 bull's genome data were Holsteins (n=389) and the average genome coverage for the entire dataset was 11.0. A single SNP regression approach to GWAS was then performed on this imputed region, using the same model described previously in this study.

3.3 Results

3.3.1 Single SNP regression approach

The association between each SNP and bTB susceptibility from the single SNP regression approach is illustrated in Figure 3.1. A total of 74 SNPs had genome wide significance level at P $<1x10^{-6}$. A Quantile-Quantile plot of the observed against expected SNP P-values (uncorrected for multiple testing) is in Figure 3.1. The heritability accounted for by all SNPs was estimated at 0.10. A total of 28 QTL regions were identified from these SNPs. The length of these QTLs ranged from 0.6 to 183kb. The most significant (P $<1x10^{-8}$) SNP, BovineHD0100019801, was located on BTA1 and accounted for 0.089% of the genetic variance for bTB susceptibility. BovineHD0100019801 was part of a 12kb QTL and is located within intron eleven of the *KALRN* gene. The major A allele was associated with an increase in resistance to bTB infection, with animals homozygote for this allele having a mean EBV of -0.099 (n=378, SD=0.1390), compared to animals with a homozygote for the minor allele (MAF=0.33) having a mean EBV of -0.0120 (n=105, SD=0.2010).

A total of four genes (Table 3.1) were identified overlapping with, or within the detected QTL regions from single SNP regression analysis. Of these genes, only two (*AP3B1* on BTA10 and *FKBP5* on BTA23) have any known involvement in immune response. The major T allele at the SNP (Hapmap50596-BTA-121389, P-value = $4.09x10^{-06}$) within the region on BTA23 was associated with an increase in resistance to bTB. Animals with a homozygote for the major allele of Hapmap50596-BTA-121389 had increased resistance to bTB with a mean EBV of -0.0909 (n=703, SD=0.1548) compared

to a mean EBV of 0.0778 (n=14, SD=0.1924) associated with a homozygote of the minor allele (MAF=0.1361).

Table 3.1 Genes identified, chromosome, SNP name of the significant SNPs within quantitative trait locus (QTL), P-value, minor allele frequency (MAF), and base pair length of QTL. QTL regions identified using SNP effects from single SNP regression approach to GWAS.

chromosome	Gene name	SNP name	P-value (FDR=1%)	MAF	QTL length (bp)
1	*KALRN*	BovineHD0100019801	$9.44x10^{-8}$	0.33	12165
4	*DPP6*	BovineHD0400033997	$7.42x10^{-6}$	0.45	32568
		BovineHD0400033998	$7.42x10^{-6}$	0.45	
10	*AP3B1*	BovineHD1000002985	$2.23x10^{-6}$	0.24	13726
		BovineHD1000002982	$2.43x10^{-6}$	0.24	
		BovineHD1000030516	$3.10x10^{-6}$	0.25	
23	*FKBP5*	Hapmap50596-BTA-121389	$4.09\ x10^{-6}$	0.13	74693

Network analysis using results from the SSR provided a list of networks which contained 292 genes, with "Antigen processing and presentation" (*PID Biocarta 4144*) as the most significantly overrepresented pathway ($P < 1x10^{-4}$). The 10 most significantly represented pathways from the single SNP approach are listed in Table 3.2.

Table 3.2 *Top ten biological pathways identified as being associated with bTB susceptibility using SNP effects estimated from a single SNP regression approach to GWAS. Pathway name, corrected P-value, pathway source, and genes associated with this pathway as listed.*

Pathway name	Corrected P-value	Source	Genes
Antigen processing and presentation	0.0011	PID BIOCARTA	BTA91437, PSMB8, PSMB9, TAP1
Arf1 pathway	0.0011	PID NCI	AP2A1, AP2M1, CD4, CYTH2, KDELR1
Validated targets of C-MYC transcriptional activation	0.0115	PID NCI	BAX, BTA19231, EIF4A1, NME2, PDCD10, RUVBL2, TP53
Nef Mediated CD4 Down-regulation	0.0145	REACTOME	AP2A1, AP2M1, CD4
ABC transporters	0.0285	KEGG	ABCC3, BTA104466, BTA87418, BTA91437, TAP1
Classical Kir channels	0.0321	REACTOME	BTA62324, KCNJ12
Antigen processing and presentation	0.0719	KEGG	BOLA-DMA, BTA91437, CD4, TAP1
ER-Phagosome pathway	0.0719	REACTOME	BTA91437, PSMB8, PSMB9, PSMD2, TAP1
Nef Mediated CD8 Down-regulation	0.0786	REACTOME	AP2A1, AP2M1
TNF/stress related signaling	0.0838	PID BIOCARTA	ATF1, MAP2K3, TANK

3.3.2 Bayesian approach

Bayes factors for each SNP are in Figure 3.1. The proportion of additive genetic variance accounted for by all SNPs from the Bayesian approach was 15% with a h^2_{SNP} of 0.025. A total of 484 SNPs were identified as being "strongly associated" (i.e., entered the Bayesian model at least 0.01 of the time and therefore had a Bayes factor of > 10). Two SNPs were "decisively associated" (i.e., each SNP entered the Bayesian model in at least 0.091 of the Gibbs chains and therefore had a Bayes factor > 100). A total of 57 QTL regions were identified around SNPs with a Bayes factor ≥20, and ranged in length from 0.8kb to 273kb.

The SNP BovineHD1700021001 with the greatest Bayes factor (i.e., 285) resided on BTA17 and accounted for 0.0018% of the additive genetic variance. BovineHD1700021001 was part of a 3kb QTL and was located within intron three of the RNF185 gene. Animals with a homozygote for the major allele of BovineHD1700021001 indicated a decreased resistance to bTB and had a mean EBV of 0.0978 (n=628, SD=0.14901) compared to -0.0191(n=16, SD=0.1143) for animals with a homozygote for the minor allele (MAF=0.3234).

A total of 40 genes were identified overlapping with, or within the 57 QTL regions (Table 3.3). Six of these genes were predicted to have an involvement in host immune response to infection; *PTN* on BTA4, *AP3B1* on BTA10, *HSF1*on BTA14, *SHARPIN* on BTA14, *TRAF4* on BTA20 and *FKBP5* on BTA23.

Table 3.3 Genes identified, chromosome, SNP name of the SNP with significant Bayes factor within quantitative trait loci, Bayes factor, minor allele frequency (MAF) and QTL length for QTLs identified using SNP effects from a Bayesian approach to GWAS.

Chromosome	Gene name	SNP name	Bayes factor	MAF	QTL length
1	*KALRN*	BovineHD0100019801	57.58	0.33	12165
1	*TMPRSS2*	BovineHD0100041198	23.41	0.12	3575
2	*TMEFF2*	BovineHD0200023336	29.83	0.43	22453
3	*NFIA*	BovineHD0300025132	26.77	0.41	17003
		BovineHD0300024222	22.78	0.39	
3	*TMEM125*	ARS-BFGL-NGS-114804	20.69	0.39	28366
4	*PTN*	BovineHD4100003190	30.57	0.41	6142
5	*DENND5B*	BovineHD0500022394	38.49	0.49	30818
		BovineHD0500022395	20.38	0.49	
6	*PGM2*	BovineHD0600016172	51.47	0.42	62372
		BovineHD0600016181	49.70	0.43	
6	*LIMCH1*	BovineHD0600017122	20.69	0.21	31027
6	*SLC4A4*	BovineHD0600024174	35.37	0.34	6800
8	*PGM5*	BovineHD0800013359	21.01	0.47	16615
10	*AP3B1*	BovineHD1000002985	41.29	0.24	13726
14	*CPSF1, ADCK5, SLC52A2 SCRT1, DGAT1,*	BovineHD1400000249	134.16	0.48	266275
	HSF1, MROH1, MIR1839, MAF1	BovineHD1400000246	37.41	0.48	
	SHARPIN, CYC1, GPAA1, EXOSC4, OPLAH	BovineHD1400000246	37.41	0.48	
17	*OSBP2*	BovineHD1700020903	41.62	0.26	22175
		BovineHD1700020906	28.88	0.26	
		BovineHD1700020907	28.88	0.26	
		BovineHD1700020908	22.36	0.26	
17	*RNF185*	BovineHD1700021001	285.55	0.32	3141
18	*SLC6A2*	ARS-BFGL-NGS-2712	20.07	0.096	12244
18	*KCNK6*	BovineHD1800014284	28.46	0.26	101126
20	*SDF2, SUPT6H, PROCA1, RAB34, RPL23A*	BovineHD2000019220	28.46	0.41	36609
	TLCD1, NEK8, TRAF4	BovineHD2000019231	23.72	0.41	
		BovineHD2000019229	20.69	0.41	
23	*FKBP5*	Hapmap50596-BTA-121389	21.74	0.13	74693
29	*ME3*	BovineHD2900002648	26.14	0.45	1263

Genes within QTLs defined from SNPs with a Bayes factor greater than 10 were considered for pathway analysis, which returned more than 350 statistically ($P < 1x10^{-4}$) overrepresented pathways. Several pathways associated with immunity were identified, including *TNF alpha* (*Netpath 15913*, $P < 10^{-7}$), Adipocytokine signalling pathway (*KEGG 590*, P-value $< 10^{-6}$), and *Fas* signalling pathway (*INOH 16231*, P-value $< 10^{-6}$). The ten most significant pathways from the Bayesian approach are listed in Table 3.4.

Table 3.4 Top ten biological pathways identified as being associated with bTB susceptibility using SNP effects estimated from a Bayesian approach to GWAS. Pathway name, corrected P-value, pathway source, and genes associated with this pathway as listed.

Pathway name	Corrected P-value	Source	Genes
Signalling events mediated by HDAC Class II	$2.66x10^{-14}$	PID NCI	BCL6, BCOR, GNB1, GNG2, HDAC4, HDAC7, HSP90AA1, NCOR2, RAN, YWHAB, YWHAE
Cell cycle	$4.63x10^{-9}$	KEGG	EP300, GADD45A, GADD45B, GADD45G, HDAC2, MDM2, SFN, YWHAB, YWHAE, YWHAG, YWHAQ, YWHAZ
Insulin receptor signalling	$9.85x10^{-9}$	INOH	PPP2R1A, PPP2R1B, PRKAA1, PRKAA2, SFN, YWHAB, YWHAE, YWHAG, YWHAQ, YWHAZ
Glucocorticoid receptor regulatory network	$1.05x10^{-8}$	PID NCI	BT.26220, BT.55996, BT.60919, EGR1, HDAC2, HSP90AA1, MDM2, NCOA2, SFN
Mechanism of gene regulation by peroxisome proliferators via ppara	$1.18x10^{-8}$	PID BIOCARTA	BT.26220, BT.55996, BT.78456, EP300, HSP90AA1, NCOR1, NCOR2, NR0B2
Androgen Receptor	$2.88x10^{-8}$	NETPATH	BT.55996, BT.61976, EP300, HDAC7, HSP90AA1, MDM2, NCOA2, NCOR1, NCOR2, NR0B2, RAN, RANBP9
Leptin	$3.72x10^{-8}$	NETPATH	BT.26220, BT.55996, BT.60919, BT.61976, EP300, IKBKG, ITGB5, LIMK1, PRKAA1, PRKAA2
Notch	$4.10x10^{-8}$	NETPATH	APP, BT.26220, BT.60919, BT.61976, EP300, HDAC2, NCOR1, NCOR2, SPEN
Pathways in cancer	$4.36x10^{-8}$	KEGG	BT.25770, BT.26220, BT.56113, BT.60919, BT.78456, EP300, HDAC2, HSP90AA1, IKBKG, MDM2, PPARD, RAC2, RHOA, RXRB, RXRG
TNF alpha	$7.31x10^{-8}$	NETPATH	BT.26220, BT.60919, BT.61976, BT.78456, EP300, GLUL, HDAC2, HSP90AA1, IKBKG, YWHAB, YWHAE, YWHAG, YWHAQ, YWHAZ

3.3.3 IBD analysis

Eighteen prolific sires in the test group (i.e. sires with ≥ 50 daughters in ≥ 10 herds) were identified in chapter two as belonging to a small group with extremely high mean daughter prevalence of bTB infection (≥60%, compared to an average of 19%). Eleven of these outlying high infection score animals were highly related, splitting into four lineages. One lineage traced six of the high bTB sires to one ancestor and the four lineages had multiple shared sires. We hypothesized that the shared susceptibility among these lineages would manifest in specific genomic regions which are shared by common descent from founders. The normalized proportion of identity by descent (IBD) shared across the genome is plotted in Figure 3.1. The most represented, *i.e.* the top 20 commonly shared IBD blocks, are given in Table 3.5 with the most extreme a 5 Mb block on BTA13. Fifty-four percent of the high bTB sires used in IBD analysis shared this haplotype compared to 12% of sires with a low (<30%) prevalence of bTB infection in their daughters.

Table 3.5 Top ten biological pathways identified as being associated with bTB susceptibility using SNP effects estimated from a Bayesian approach to GWAS. Pathway name, corrected P-value, pathway source, and genes associated with this pathway as listed.

CHR	Start SNP	End SNP	Start position	End position	High prop	Low prop
13	BovineHD1300014366	BovineHD1300014692	49589499	51444238	0.54	0.12
13	BovineHD1300013610	BovineHD1300014364	46502709	49576039	0.52	0.12
7	BovineHD0700014051	BovineHD0700014305	48615285	49344575	0.52	0.15
7	BovineHD0700014306	BovineHD0700015867	49347339	54833963	0.50	0.18
13	BovineHD1300013374	BovineHD1300013492	45829062	46168983	0.47	0.14
7	BovineHD0700019982	BovineHD0700020720	68220084	70459483	0.45	0.10
7	BovineHD0700013664	BovineHD0700014049	47130142	48606465	0.43	0.14
3	BovineHD4100002136	BovineHD0300018020	56353864	60108292	0.41	0.09
9	BovineHD0900024988	BovineHD4100007649	88657158	90474884	0.41	0.10
13	BovineHD1300014695	BovineHD1300016474	51460099	57393129	0.41	0.11
7	BovineHD0700019273	BovineHD0700019979	66064958	68207807	0.41	0.09
7	BovineHD0700015911	BovineHD0700016486	55087353	57559898	0.40	0.17
7	ARS-BFGL-NGS-33898	BovineHD0700033819	65520210	66054838	0.40	0.09
7	BovineHD0700020722	BovineHD0700021987	70463539	74679215	0.38	0.09
3	BovineHD0300018022	BovineHD0300018594	60114338	61974915	0.38	0.09
7	BovineHD4100006127	BovineHD0700019111	63918463	65517703	0.38	0.09
3	BovineHD0300024252	BovineHD0300024614	84878201	86233867	0.38	0.17
13	BovineHD1300011097	BovineHD1300012983	38486971	44519866	0.38	0.13
23	BovineHD2300001843	BovineHD2300002752	7308515	11189811	0.38	0.09

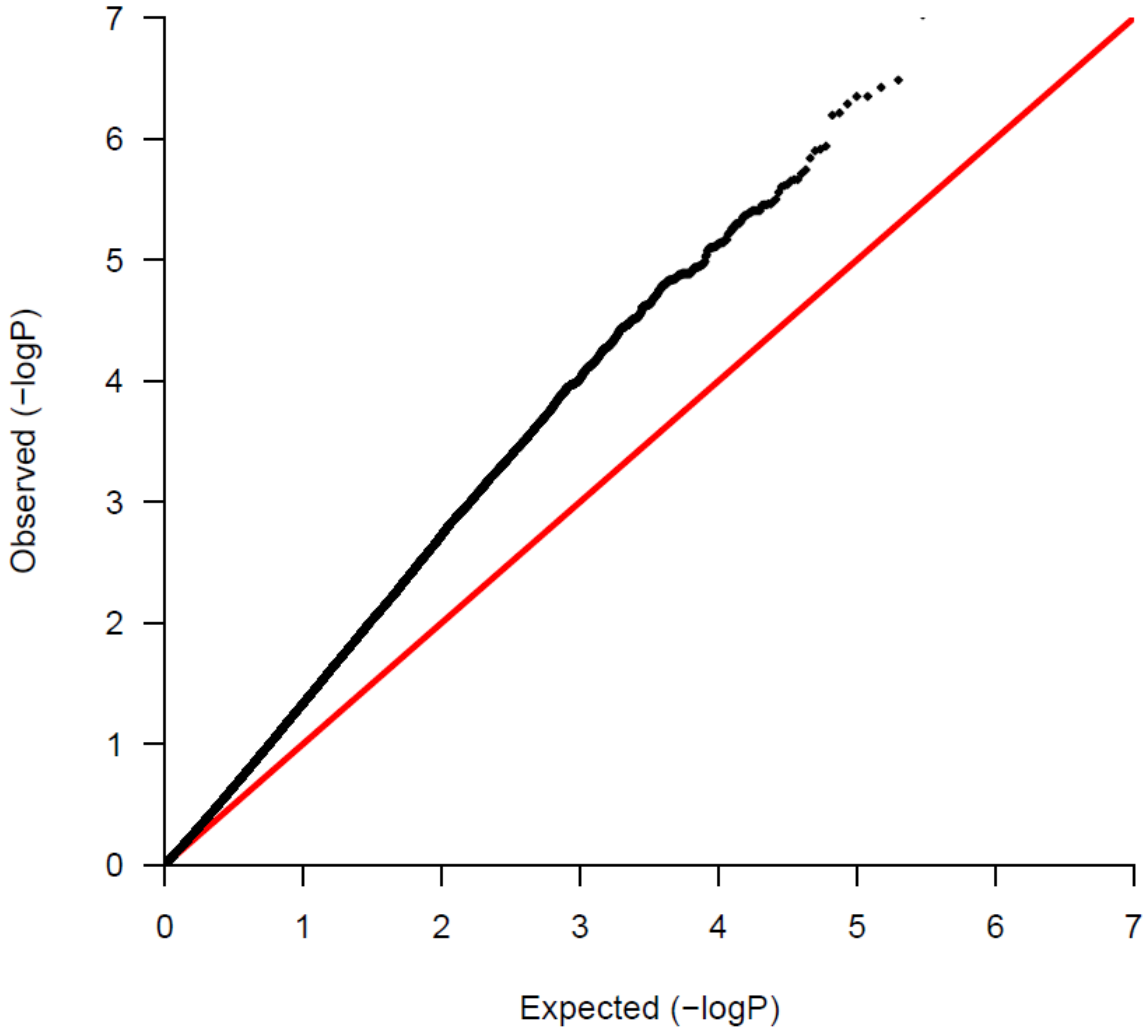


Figure 3.1 QQ-plot for P-values (-Log10) estimated using the single SNP regression approach.

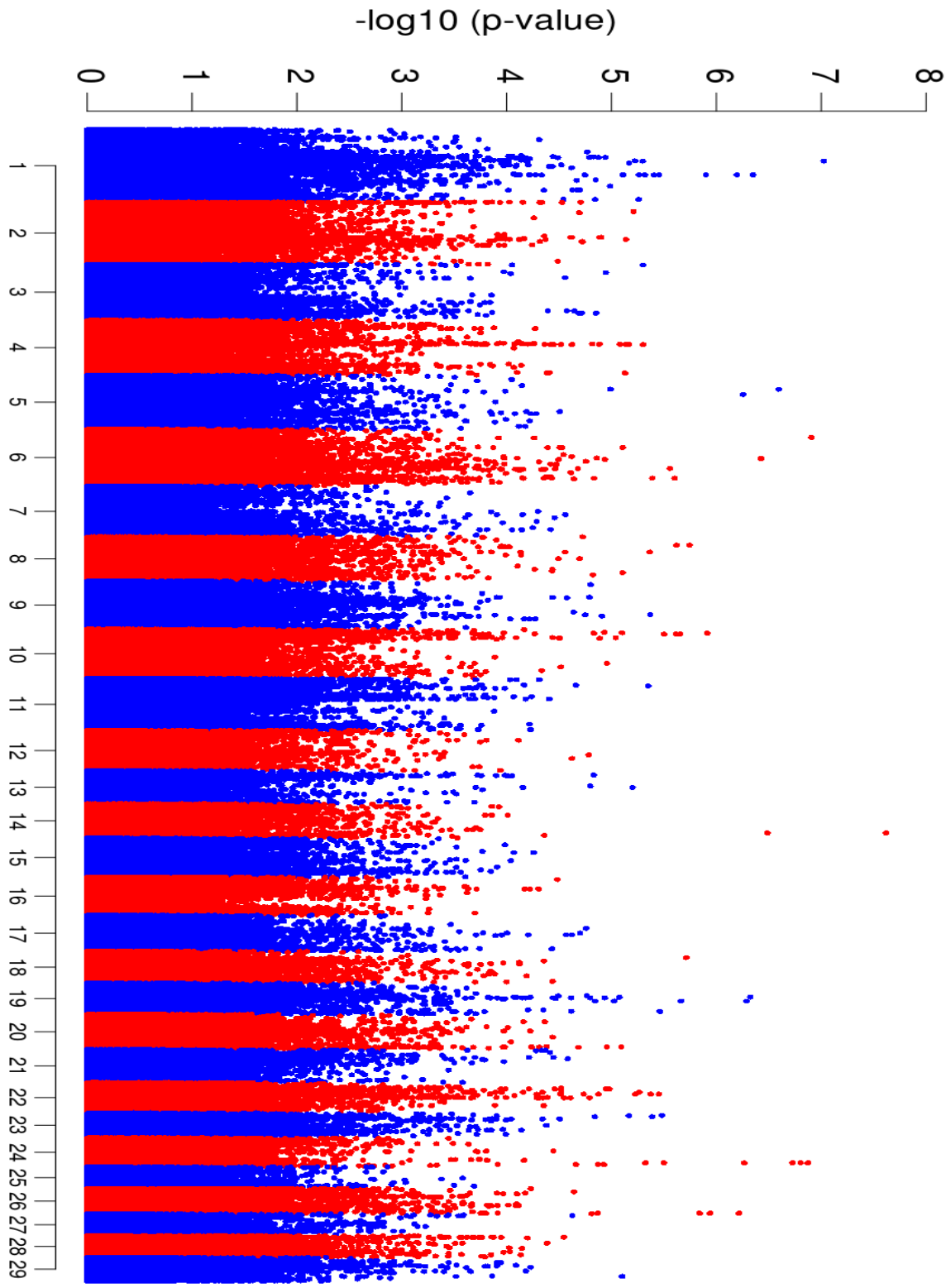


Figure 3.2 Manhattan plot of P-values estimated from a single SNP regression approach to GWAS for bTB susceptibility

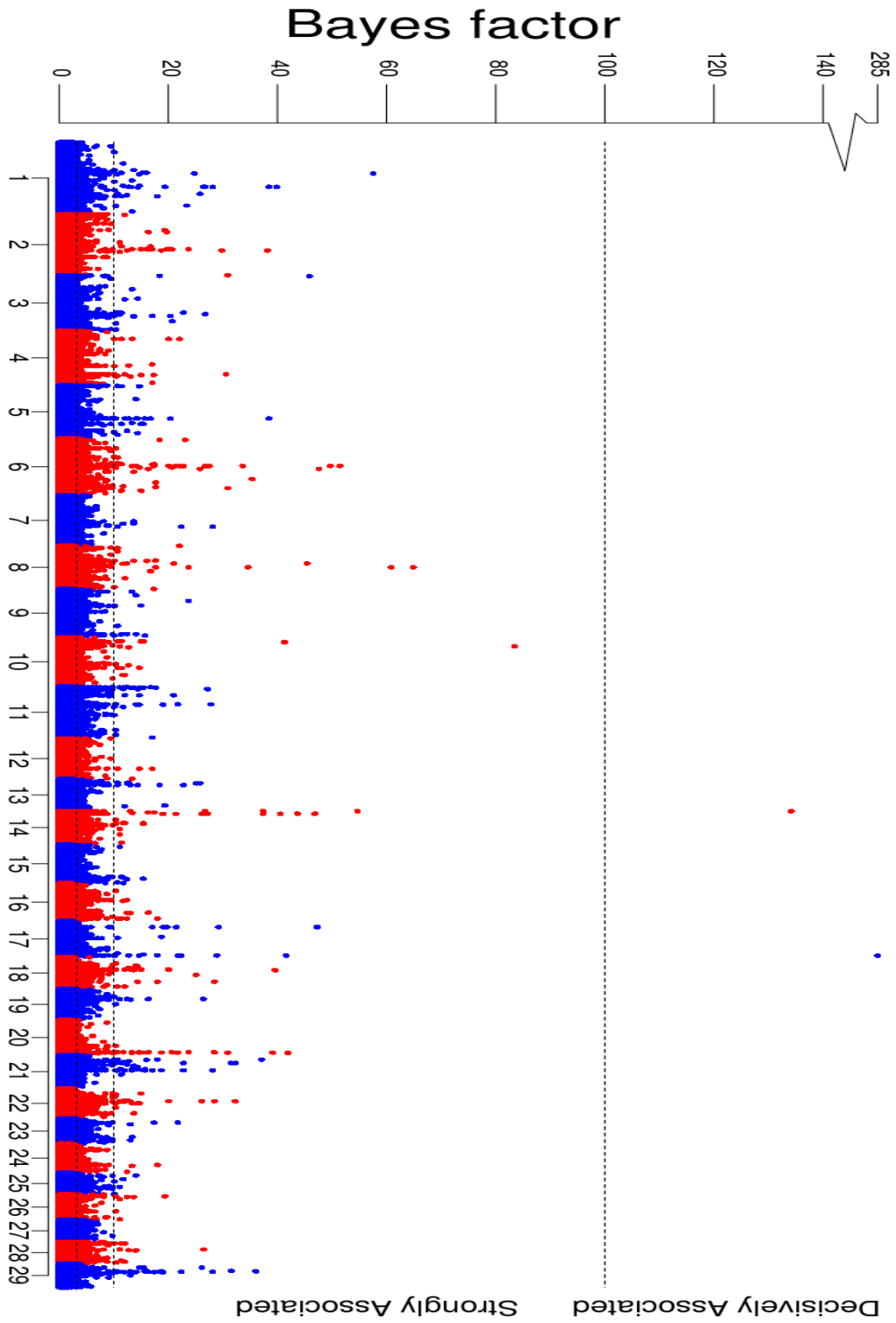


Figure 3.3 Manhattan plot of the association between each SNP and bTB estimated using the Bayesian approach for bTB susceptibility

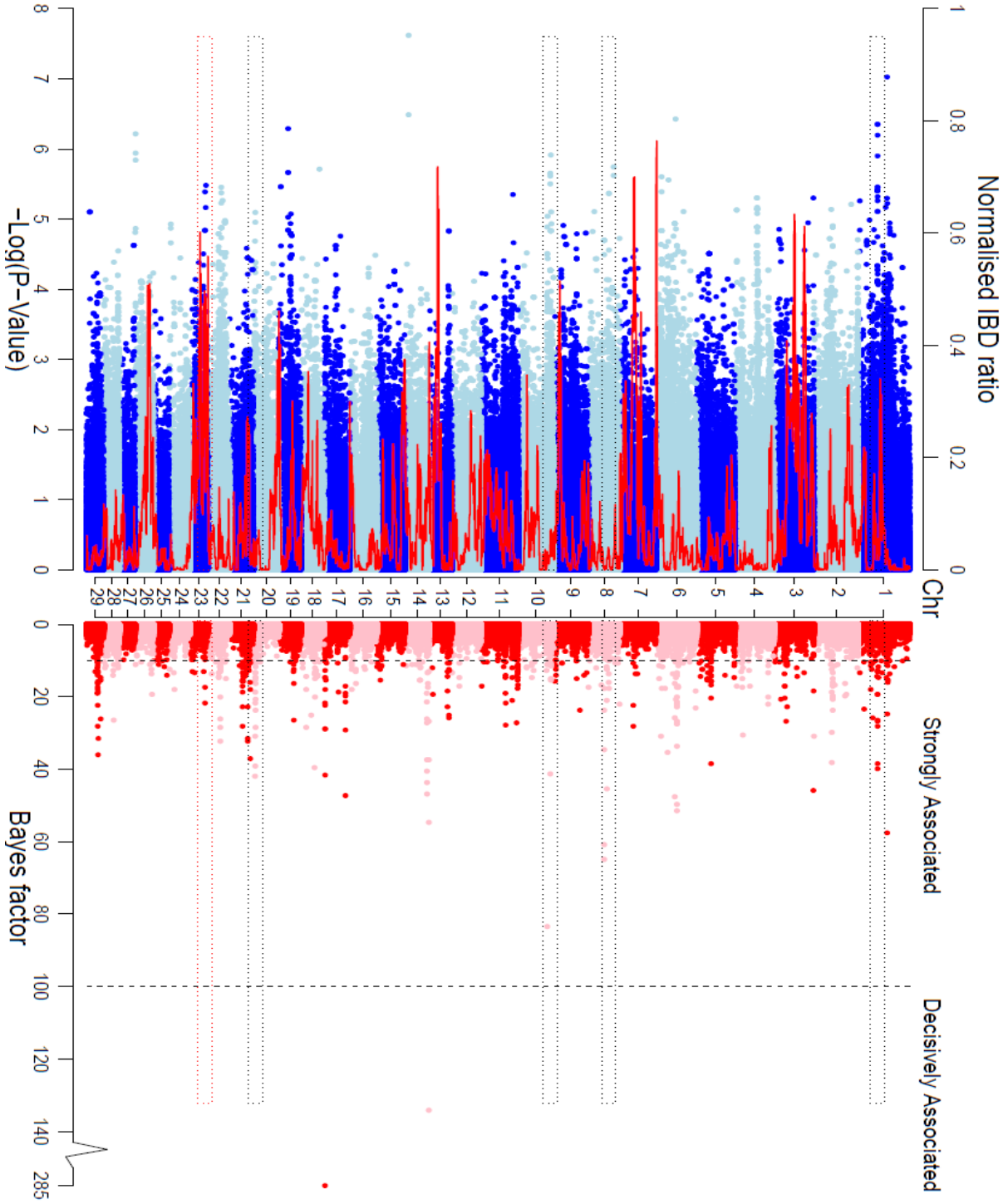


Figure 3.4 Manhattan plot of the association between each SNP and bTB susceptibility estimated using the single SNP regression approach (top Manhattan plot) vs Manhattan plot of the association between each SNP and bTB estimated using the Bayesian approach (bottom Manhattan plot). The proportion of shared IBD across the genome is represented by the red dotted line. QTLs that showed significance in both the Bayesian and SSR approaches are highlighted. The only QTL that was observed to have significance in all three approaches is highlighted by the red rectangle.

3.3.4 Cross-method comparisons

The Bayesian and single SNP regression approaches to GWAS identified 17 SNPs that were significant (i.e. Bayes Factor > 10 and P-Value < $1x10^{-5}$) in both approaches (Table 3.6). Three QTLs were common to both the Bayesian and SSR approaches on BTA1, BTA10 and BTA23. One IBD block (~3 Mb in length) shared in 38% of high bTB sires on BTA23 overlapped QTL regions from the other analyses and contained 63 genes, 5 of which were predicted to have involvement in host immune response (*TAPBP*, *DEF6*, *FKBP5*, *MAPK14* and *MAPK13*).

Table 3.6 Single nucleotide polymorphisms estimated to be significantly associated with bTB susceptibility in both the Bayesian and single SNP regression approached to GWAS.

SNP	CHR	Genetic Position	Bayes Factor	P-value*
BovineHD0100019699	1	69368803	24.77	$6.03x10^{-6}$
BovineHD0100019801	1	69548446	57.58	$9.55x10^{-8}$
BovineHD0100029053**	1	101890992	38.49	$1.26x10^{-6}$
Hapmap53234-rs29020933**	1	101892420	26.45	$7.94x10^{-6}$
BovineHD0100029057**	1	101896647	28.14	$5.50x10^{-6}$
BovineHD0100029059**	1	101899126	39.89	$4.47x10^{-7}$
BovineHD0100029060**	1	101899991	26.77	$6.46x10^{-7}$
BovineHD0100029061**	1	101903117	19.34	$4.47x10^{-7}$
BovineHD0100029062**	1	101910522	12.74	$3.55x10^{-6}$
Hapmap43413-BTA-95698**	1	101918115	19.34	$3.55x10^{-6}$
BovineHD0100029132	1	102058031	13.05	$3.98x10^{-6}$
BovineHD0300000514	3	1943906	45.86	$5.01x10^{-6}$
BovineHD0800010906	8	36652588	17.69	$4.37x10^{-6}$
BovineHD1000002985**	10	9147391	41.29	$2.24x10^{-6}$
BovineHD2000019456	20	67224771	21.74	$8.13x10^{-6}$
Hapmap50596-BTA-121389**	23	9591806	21.74	$4.17x10^{-6}$
BovineHD2600014591	26	50406685	11.11	$1.45x10^{-6}$

*FDR level of 1%. ** Indicates a QTL could be formed using this SNP.

3.3.5 Imputation of candidate region

The candidate region on BTA23 was the only QTL that showed a significant result across each test and was selected for targeted imputation. The strength of association with bTB for each SNP from the imputed sequence flanking a 3 Mb window around the QTL on BTA23 is plotted in Figure 3.3. Twenty two SNPs in the imputed BTA23 region had a significance level of $P < 1x10^{-5}$. All but two of the top ten of the most significant ($P < 1x10^{-5}$) were located within intron 1, 2, 5 and 8 of the *FKBP5* gene (including the two most significant, $P < 4x10^{-6}$; positions 9590819 and 9591806) the other two located within intron eight of *MAPK14*. Furthermore a distinct peak of SNPs with strong P-values were located around the significant SNPs located in *FKBP5* (Figure 3.3).

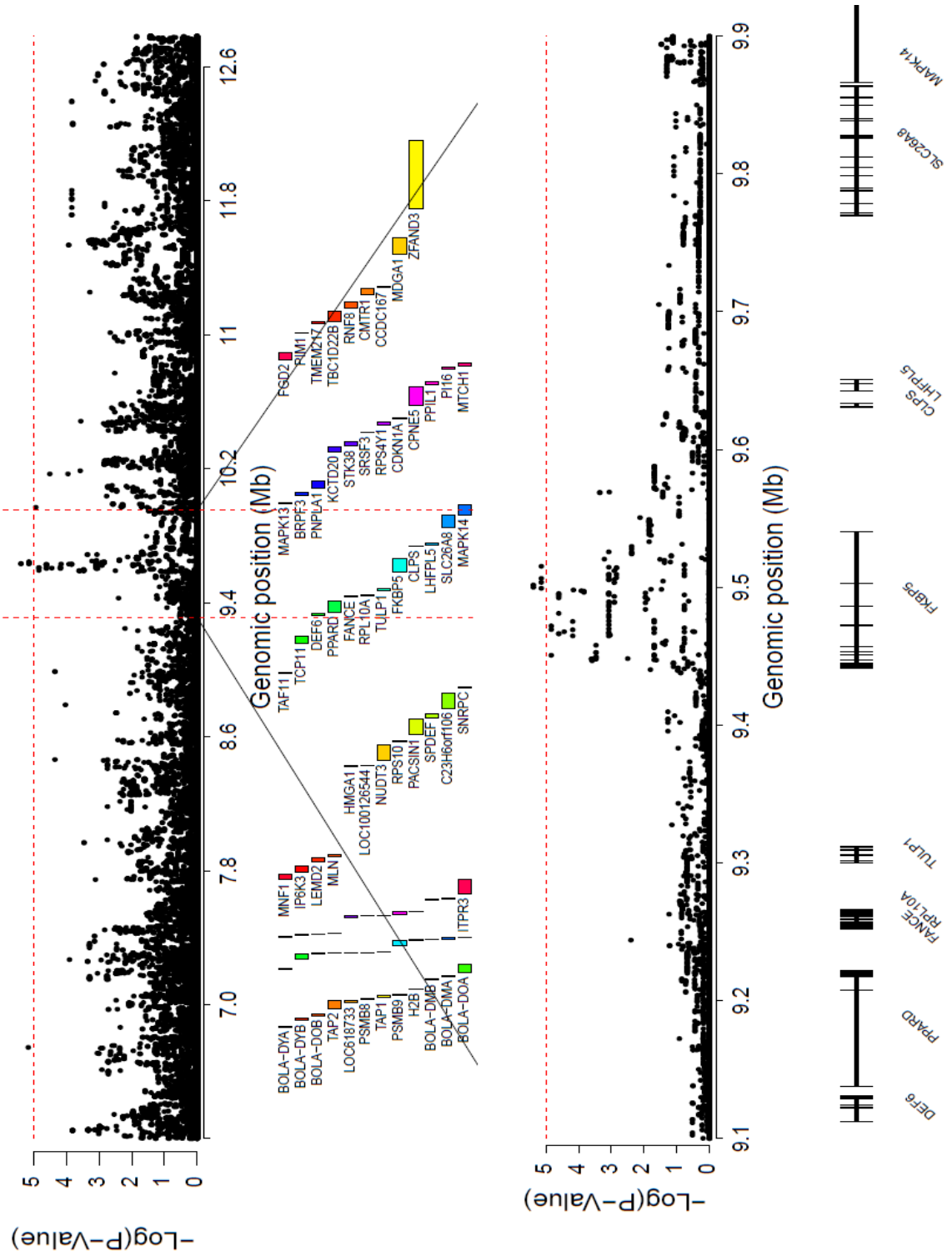


Figure 3.5 Manhattan plot of the association between each SNP and bTB susceptibility estimated within a 5 Mb imputed region on BTA23. Using the single SNP regression approach (top), with gene annotated below the plot. Manhattan plot of the association between each. Gene names and positions were acquired from the UCSC genome browser using the UMD 3.1 build. Red dotted line shows significance threshold using an FDR level of 0.05.

3.4 Discussion

In the developed world, current strategies for reducing bTB infection in cattle primarily focus on efforts to detect infection in herds (through annual herd tests and abattoir surveillance), to effectively control infection in known infected herds, and to limit the spread of infection between herds and regions. Increasingly, these efforts are supplemented by active programmes to manage infection in known wildlife reservoirs. Additional strategies are currently being considered, including increased host resistance through focused breeding practices [80] or via the use of vaccinations either in cattle or wildlife reservoirs [146]. Because both across and within breed differences in susceptibility to bTB exist, breed substitution or within breed selection are viable options to supplement on-going eradication programs for bTB. Moreover, the existence of considerable heritable genetic variation for bTB suggests the existence of genomic variation governing differences in bTB susceptibility among animals. However, only three genome wide studies in cattle [127, 128, 130] have attempted to locate genomic regions contributing to the additive genetic variance in bTB. Potentially, elucidating the genomic variation in bTB could add to the development of pharmaceuticals, or incorporation of genomic information into genetic evaluations could facilitate greater rates of genetic gain.

3.4.1 Methodological comparisons

In a SSR approach, each SNP is included and tested for significance individually in the model. In the Bayesian approach all SNPs are fitted simultaneously within the same model. Of the SNPs estimated to be significant in the Bayesian and SSR approaches, a total of 17 SNPs were common between both approaches (Table 3.6). However, the most

significant SNP in the SSR approach was not estimated as the most significant SNP in the Bayesian approach and *vice versa*. Significant associations estimated using SSR or other frequentist approaches generally only explain a small fraction of genetic variation of quantitative trait [37]; in contrast, Bayesian approaches to GWAS can account for most of the genetic variation by simultaneously fitting all of the genetic markers as random effects.

The Q-Q plot for the expected versus observed P-values from the SSR approach showed a clear inflation of P-values against those expected. This is likely due to plotting P-values in the Q-Q plot that were uncorrected for multiple testing. As other methods for multiple testing such as bonferroni correction were too stringent, resulting in no SNPs reaching significance, we chose false discovery rate as our method for multiple testing correction. It must be noted also that inflation of P-values in polgenic traits (such as in bTB susceptibility) is a common problem [147]. Furthermore while a numerator relationship matrix was included in the model for the single SNP regression, it assumes an equal inheritance of genetic material from both parents to offspring, which is not the case in reality. Modern genomic prediction methods now account for this unequal inheritance from parents with the inclusion of a genomic relationship matrix.

3.4.2 Comparison to previous livestock literature

Results from the current study do not corroborate those of Finlay et al., [127], who identified a region on BTA22 for bTB susceptibility. Our results supersede this prior work and have greater statistical power due to a greater number of experimental units (n=841 vs. n=307), higher density genome coverage and improved epidemiological data. Also, using

field case/control data (592 cases and 559 controls), as opposed to a quantitative trait in the present study, Bermingham et al., [128] suggested two novel resistance loci for bTB infection with chromosome-wide significance, located on BTA2 and BTA13 which did not replicate here. This may be a result of differences in phenotypes employed (Quantitative vs binary phenotype), a difference in population structure or of limited analytical power. Although the numbers of test animals included were similar, the current study is based on the phenotypes of several thousand animals (n=105,526). An admixture mapping analysis for bTB in Ethiopian Zebu/Holstein-Friesian hybrids performed by Kassahun et al., [130] suggested a toll-like receptor gene cluster on BTA6, which was also not replicated in the present study. Genomic regions identified to have association with bTB susceptibility in the present study were also not consistent with those documented to be associated with paratuberculosis in cattle [124-126], which shares some similarities with bTB.

3.4.3 Novel candidate quantitative trait loci

The Bayesian, SSR and IBD approaches identified a region on BTA23 as significantly associated with bTB infection in dairy cattle containing two genes *FKBP5* and *MAPK14*. The *FKBP5* gene encodes a protein from the immunophillin protein family, a family of proteins often targeted by immunosuppressant drugs. The *FKBP5* protein is a *cis-trans* prolyl isomerase that binds to the immunosuppressants *FK506* and rapamycin [148]. *FKBP5* is also involved in the *TNF alpha/NF-kB* signaling pathway, a major pathway involved in host immune response to disease and other harmful stressors and is highly expressed in T-lymphocytes [148]. All of the significant SNPs located within *FKBP5* from analysis using imputed full sequence were within introns of this gene, and intronic associations with disease traits have been previously described [149-151].

Furthermore the distinct peak of SNPs with strong P-values present around the significant SNPs located within *FKBP5* (figure 3.3), increases the likelihood that this gene is truly associated with host susceptibility to bTB.

It must be noted however, that while the QTL region encompassing the *FKBP5* gene was the only region implicated across all three of our analyses, it was not the most significant in any individual analysis. The most significant SNPs ($P<1x10^{-8}$; n=2) estimated from the SSR approach were either located within genes not directly implicated with immune response (BovineHD0100019801, located within the gene *KALRN*), or in QTLs not containing genes (BovineHD1400020824). The most significant SNP from the Bayesian analysis (BovineHD1700021001, Bayes factor = 285) was identified to be within a gene with no known involvement in immune response; *RNF185*. Due to the strong association of this SNP however, further investigation of this region could lead to more insight into the host genetic susceptibility to bTB. Moreover while the QTL containing the second most significant SNP from the Bayesian analysis (BovineHD1400000249, Bayes factor = 134) contained two genes (*HSF1* and *SHARPIN*) with known immune response, this signal also associated with the *DGAT1* gene which is known to play a key role in milk fat production [152].

A QTL on BTA1 that was estimated as significant ($P < 1x10^{-5}$, Bayes Factor > 10) in both the Bayesian and SSR approaches contained 9 of the 17 significant SNPs between these approaches, but was not identified as significant in the IBD approach. While no genes were identified within this QTL, due to the number of common significant SNPs between the Bayesian and SSR approach, this QTL could be a candidate for further investigation.

3.4.4 Network Analysis

Identification of TB susceptibility genes in humans has been documented to be more difficult compared to other infections, even with large sample sizes [153]. This is possibly the result of selection against variants with large or even moderate effect during this long-standing infection. However, the most significantly associated pathway using results from the SSR approach was the antigen processing and presentation pathway. Multiple mutations in this pathway could lead to the under-expression or non-expression of antigens on the surface of infected cells, which highlights cells for degradation by the host immune system. *M. bovis* replicates within the lysosomes of the host macrophages. It can be hypothesised that an impaired antigen processing and presentation pathway could allow *M. bovis* to survive longer in the host [154].

3.5 Conclusion

Results from the present study point to the contribution of multiple genes associated with the susceptibility of cattle to bTB infection. While the genes identified in the present study have not been validated in other cattle studies, several studies have documented their contribution to immune responses in humans. Furthermore the pathway analysis performed in the current study has not been performed previously for bTB infection in cattle and points towards several possible pathways significantly associated with bTB infection in cattle. Results presented in the present study could be used to investigate further the physiological role of the genes and pathways identified, in order to develop pharmaceuticals to affect host resistance to bTB infection.

Chapter four - Mapping of ancestral haplotypes and footprints of selection in Near Eastern cattle breeds and East African Zebu

4.1 Introduction

A major goal in livestock genetics is the identification of quantitative trait loci (QTL) associated with predisposition to complex diseases. Genome wide association scans have been successful in the discovery of alleles associated with disease in both cattle and humans [22, 127, 128, 155-158], however it is becoming more apparent that most complex traits are polygenic, with multiple alleles of small effect contributing to the traits [38]. Due to the small effect of alleles contributing to complex traits, having enough power to identify these alleles requires thousands of cases and controls [159], which is costly. An alternative approach to identifying loci of functional importance is that of inferring adaptive history, which is detection of population genetic signatures of selection. The methods explored in this chapter could have future potential in identifying genomic regions with strong associations with polygenic traits such as bovine tuberculosis. Furthermore understanding what ancestral haplotypes tend to be inherited more readily in environments with intense selection for adaption to disease could give insight into common genomic regions associated with disease resistance, which could have applications in bTB research.

If an allele is under strong selective pressure in a population it can increase in allele frequency rapidly, compared to that of a neutral mutation which can take many generations before it reaches a high frequency through genetic drift. Signatures of recent selection in cattle have been reported previously [160, 161], using methods based on the extended haplotype homozygosity (EHH) statistic [162]. EHH works on the principle that

a novel mutation under positive selection pressure will rapidly increase in frequency accompanied by a surrounding haplotype giving an extended region of reduced homozygosity, referred to as a selective sweep [163]. Qanbari et al. [161] performed a genome wide EHH test on a population of German Holstein-Friesian sires, identifying a number of genes to have known involvement in milk production and fertility. Gautier et al. [160] used an extension of the EHH statistic known as the integrated haplotype score (iHS; [164]), which can be used to detect regions with high levels of haplotype homozygosity over an unexpected long distance within a populations. However information on the ancestral and derived allele of each SNP is needed to calculate iHS accurately as iHS is based on the ratio of the extent of EHH associated with each allele [160]. Gautier and Naves [160] performed a genome wide analysis of New World Creole cattle from Guadalupe (n=140) identifying 23 genomic regions displaying strong signals of selection, containing several genes associated with reproductive function, metabolism and immune response. A whole genome iHS scan has been performed on a population of 162 Kenyan dairy crossbreeds [165]. This study identified a number of candidate regions showing positive signatures of selection for genes involved in immune response and survivability, including *IL-2*, toll like receptors and *NFKB2*. Whole genome iHS and Fst analysis has also been performed on a population of (n=462) East African shorthorn zebu, using Illumina bovineSNP50 chip genotypes. Bahbahani et al. [166] identified a number of signatures of selection associated with innate immune response, such as LOC512672 on chromosome 23 a major histocompatibility complex class I gene, located within the *BoLA* region. Kemper et al. [167] used a targeted approach to identify the usefulness of iHS in identifying genes known to be under strong selection in eight taurus breeds. They were

able to show clear signatures of selection when intense selection has been applied to a single locus, however in regions of the genome associated with complex traits under selection, only weak evidence for signatures of selections was observed[167]. However more recent work by Zhao et al. [168] was able to utilize a combination of iHS and Fst analysis in seven dairy and beef breeds (Angus, Belgian Blue, Charolais, Hereford, Holstein-Friesian, Limousin and Simmental) to identify genes known to be under strong selective pressure (i.e. *DGAT1* in Holstein-Friesian and *MSTN* in Belgian Blue) as well as novel candidate regions which were common to both iHS and Fst analysis. The use of iHS analysis for identifying signatures of selection in cattle breeds has also been documented using whole genome sequences of 43 Fleckvieh animals [169] and was successful in identifying known (i.e. *MC1R* for coat pigmentation) and novel genes under selective pressure.

Ancestry mapping provides a more powerful and attractive approach to GWAS for identifying QTL in admixed populations, where the QTL is expressed more in individuals with a greater frequency of ancestral alleles [170]. There are several advantages to ancestry mapping, firstly if reliable and accurate phenotypes are available it can display a higher power than traditional GWAS [159]; secondly the SNP density required for analysis is greatly reduced to GWAS, due to the assumption that SNPs are inherited as ancestral "blocks", although this advantage decreases with the number of recombination events; lastly ancestry mapping is robust to genetic diversity as a cofounder of association [129, 171]. Ancestry mapping has been successful in identifying associations with a number of diseases in both humans and livestock. A genome-wide survey carried out by Kopp et al. [172] identified evidence that the *MYH9* gene pertains a higher predisposition to focal

segmental glomerulosclerosis in African-American individuals with the African variant of *MYH9*. Kassahun et al. [129] mapped the ancestral haplotypes in a population of Ethiopian Holstein-Zebu hybrids to identify QTL associated with bovine tuberculosis infection. They identified one marginally significant locus on chromosome 6 that contained a cluster of toll-like receptor genes. Moreover Kassahun et al. [129] also mapped ancestral haplotypes to identify associations for red-coat color in Holstein-Zebu hybrids and identified a highly significant peak of SNPs over the *MC1R* gene, which has a known function in coat color pigmentation.

Admixed populations of cattle also provide a alternative model for identifying recent footprints of selection [160]. This arises from the existence of two major domestic genomes of cattle, those of *Bos indicus* and *Bos taurus* that were domesticated from distinct wild populations [173]. These two populations diverged ~280 thousand years ago leading to marked genetic divergence, facilitating the identification of chromosome ancestry. These two cattle lineages have the ability to interbreed resulting in a diversity of cattle populations that admixed at different times throughout history, from several thousand years (Near Eastern breeds and African breeds), to a few centuries (New World cattle breeds; i.e. Creole) to only a few decades (American recent hybrids i.e. Beefmaster, Brangus etc.). These different examples of breed admixture potentially allow for the mapping of footprints of selection in recent and ancient hybrids, as well as estimation of the time depth of admixture.

Taurine and indicine cattle have some major phenotypic differences. Indicine cattle possess a fatty hump at their withers and a large dewlap, they also have increased heat tolerance and an increased ability to digest low quality forage compared to taurine cattle

[174, 175]. Also, with divergent deep histories and ecologies, it would seem likely that they have genomic adaptations to different suites of pathogenic challenge. While indicine cattle comprise the majority of cattle populations worldwide, taurine cattle make up the majority of beef and dairy production breeds due to a number of desirable agricultural production traits which came about through extensive artificial selection in Europe [176]. In hybrids, is is possible to map the extent of admixture on a genomewide basis and selection signals may manifest as regions with outlying values of ancestry from either ancestral type. For example, introgression of zebu genetics may be enhanced around genomic regions with advantageous adaptations in arid environments and retarded around loci harboring immunogenes with alleles adapted to local pathogens.

Dating the time of admixture between indicus and taurus breeds has been attempted in four sub-populations of cattle breeds [176]; namely 1) breeds of known recent admixture (dated to the early 1900's) i.e. Beefmaster and Santa Gertudis, 2) New World cattle (Texas Longhorn, Corriente and Romosinuano), 3) a West African Taurine breed (N'Dama) and 4) an East African Zebu breed (Boran). McTavish et al. [176] developed a scaled median introgressed block size (SBS) statistic to estimate the time since admixture, by comparing the SBS signature of recent hybrids to older admixed populations, however McTavish et al. [176] was only able to gain very rough estimates of the time since admixture. Many methods for estimating absolute time since admixture exist for human genetic data, such as haplotype-based methods which use the variance in shared haplotype length to infer demographic history [177, 178], pairwise allele sharing methods [179], LD-based inference methods [180] and more recently a "chromosome painting" method based on patterns of haplotype similarity [181, 182]. Dating absolute admixture in cattle using

some of these methods is currently not possible due to the lack of an accurate recombination map [176].

A number of studies mentioned previously in this section have been conducted to investigate the admixture structure of East African Zebu [129, 165, 176]. Taurine cattle have been present in Northern Africa since ~7,000 years ago, with indicine cattle thought to have been introduced to Eastern Africa ~3,000 years ago and were introduced to Western Africa by ~1,000 years ago [183, 184]. It is thought that African taurus animals may have been derived from either the same domestication process as European cattle, or from an independent domestication region in Northern Africa [185, 186]. Various estimates of divergence between African and European taurine cattle are younger than those for indicine and taurine cattle; 9,000—15,000 years ago [187, 188] compared to 84,000-1 million years ago [173, 187-189], respectively. The differences in time of introduction of taurine and indicine cattle into Africa has resulted in a historic cline of admixture across Africa; with more indicine ancestry in Eastern Africa and more taurine ancestry in Western Africa [176].

Taurine animals in the Near East represent some of the first domesticated herds with the earliest estimated dating to ~10,000 years ago based on archeological evidence [190]. Several studies have shown clear admixture of Near Eastern cattle and a significant cline with zebu cattle [191, 192]. Archeological estimates place zebu cattle in Jordan ~3,400 years ago [193], suggesting that this cline is a result of trade links across the region [191].

This chapter will focus on ancestry mapping, identifying signatures of selection and estimating time since admixture in two admixed cattle populations; the East African Zebu

breeds and Near Eastern cattle breeds. Identifying signatures of selection and admixture structure in these breeds may give new insight into the mechanisms of selection that these breeds have undergone, as well as giving a possible insight into their history.

4.2 Materials and Methods

4.2.1 Genotyping

Illumina high density SNP genotypes (n=786,799 SNPs) were available for 3,325 cattle from 97 breeds with a wide geographical range of breed origin. Genotypes were provided by a collaboration between Prof Daniel Bradley (Trinity College Dublin), Dr Heather Huson (Cornell University), Dr Tad Sonstegard (United States Department of Agriculture), Cedric Gondro (University of New England), Dr Ehud Lipkin (Hebrew University of Jerusalem), Prof Johan Solkner (University of natural resources and life sciences, Vienna) and Yuri Tani Utsunomiya (Sao Paolo State University). The breakdown of number of genotypes available, including breeds and number of genotypes are shown in table 4.1. Sub-groups of each breed were defined from known breed histories and origins [194]. A total of 125 cattle were removed due to a missing genotype rate >0.10. SNPs with a minor allele frequency <0.01 and a genotype call rate <0.10 were removed, with 6,323 and 17,849 SNP discarded, respectively. After filtering and quality control, 762,627 SNPs and 3,200 genotypes remained.

Table 4.1 Summary of genotypes available (n) per breed.

Sub-group	Breed	n	Sub-group	Breed	n
African Taurus	Kuri	2	Outlier	Yak	2
African Taurus	Lagunaire	5	Taurus	Alentejana	10
African Taurus	Muturu	36	Taurus	Angus	160
African Taurus	N'Dama	87	Taurus	Angus x Romosinuano	12
African Taurus	Somba	10	Taurus	Baladi	20
Anatolian	Anatolian Black	38	Taurus	Blonde dAquitaine	5
Anatolian	South Anatolian Red	59	Taurus	Brown Swiss	77
Anatolian	Turkish Grey	46	Taurus	Braunvieh	48
Anatolian	East Anatolian Red	52	Taurus	Carora	69
Hybrid	Belmont Red	30	Taurus	Charolais	40
Hybrid	Beefmaster	24	Taurus	Curraleiro	17
Hybrid	Brangus	13	Taurus	Guernsey	21
Hybrid	Brahman x Romosinuano	1	Taurus	Galloway	10
Hybrid	Holstein x Bunaji	28	Taurus	Hereford	48
Hybrid	Hanwoo	20	Taurus	Scotch Highland	8
Hybrid	Keteku	25	Taurus	Holstein	496
Hybrid	Nganda	27	Taurus	Holstein Slick	53
Hybrid	Nanyang	21	Taurus	Jersey	142
Hybrid	Santa Gertrudis	35	Taurus	Kerry	9
Hybrid	Qinchuan	23	Taurus	Limousin	51
Asian Indicus	Brahman	79	Taurus	Limonero	83
Asian Indicus	Gyr	57	Taurus	Maremmana	5
Asian Indicus	Harianna	10	Taurus	Mongolian	18
Asian Indicus	Kankrej	2	Taurus	Montbeliard	5
Asian Indicus	Leizhou	15	Taurus	Normande	5
Asian Indicus	Nellore	40	Taurus	Norwegian Red	17
Indicus	Adamawa Gudali	24	Taurus	Pantaneiro	18
Indicus	Ankole	25	Taurus	Piedmontese	24
Inducus	Azawak	2	Taurus	Pinzgaur	118
Indicus	Bunaji	27	Taurus	Red Angus	11
Indicus	Gobra	10	Taurus	Red Holstein	2
Indicus	Maure	10	Taurus	Romosinuano	2
Indicus	Red Bororo	26	Taurus	Romagnola	34
Indicus	Sokoto Gudali	28	Taurus	Red Poll	10
Indicus	Yakanaji	25	Taurus	Senepol	82
ME Indicus	Achai	8	Taurus	Sikia	5
ME Indicus	Bhagnari	4	Taurus	Simmental	10
ME Indicus	Cholistani	13	Taurus	Senegus	6
ME Indicus	Dhanni	10	Taurus	Taleshi	25
ME Indicus	Dajal	10	Taurus	Tyrol Grey	24
ME Indicus	Lohani	13	Taurus	Wagyu	175
ME Indicus	Red Sindhi	34	East Afr. Zebu	Butana	10
ME Indicus	Sistani	19	East Afr. Zebu	Shorthorn East African Zebu	116
ME Indicus	Sahiwal	45	East Afr. Zebu	Karamoja	16
ME Indicus	Tharparkar	13	East Afr. Zebu	Kenana	10
ME Indicus	Wadara	3	East Afr. Zebu	Sheko	18
Outlier	Water Buffalo	27	East Afr. Zebu	Seleze Zebu	15
Outlier	Gaur	2			

4.2.2 Population structure of cattle breeds

In order to gain an understanding of the population structure of East African Zebu breeds and Anatolian breeds related to all breeds in our dataset principle component analysis was performed using the R package SNPRelate [195]. Outlier species were removed. Herefords were removed to remove the effects of ascertainment bias resulting from the influence of the primary genome sequence. Finally the number of animals per breed was restricted to a maximum of twenty animals so that over-represented breeds (such as Holsteins; n=496) would not skew the results. This left 1,332 animals for analysis.

To gain an insight into the ancestry of each breed unsupervised hierarchical clustering of individuals based on SNP genotypes was performed using a maximum likelihood method implemented in the program ADMIXTURE [196]. As above with principle component analysis, outliers were removed and the number of animals per breed restricted to a maximum of 20 for ADMIXTURE analysis.

To investigate the level of haplotype sharing within and between breeds, pairwise identity by descent was performed on all individuals in the dataset. The entire dataset was phased using BEAGLE [29, 136] and pairwise IBD segments were called using fastIBD, an algorithm implemented within BEAGLE. Mean IBD blocks shared between and within breed was calculated as the total number of blocks between pairs of animals divided by the total possible IBD pairings within/between breeds. A distance matrix was formed from these results by classical multidimensional scaling [197] of a matrix with values for between breed mean IBD blocks shared.

Identity by state (IBS) was also performed between/within all breeds which was performed in the R package SNPRelate [195]. IBS results were plotted as a multi-dimensional scaling.

4.2.3 Ancestry mapping

To infer the ancestry of chromosomal segments in admixed cattle populations a haplotype-based Hidden Markov Model (HMM) was utilized by the program HAPMIX [198]. HAPMIX assumes that the population being analyzed has risen from the admixture of two ancestral populations. Phased genotypes of un-admixed ancestral population or populations that are closely related to the true ancestral population are required as reference populations. HAPMIX outputs the likelihood of each SNP comprising of 0,1 or 2 alleles of the ancestral allele. HAPMIX can also be used to estimate the absolute date of admixture [198].

A target population of 86 East African Zebu animals (comprised of Shorthorn East African Zebu, n=20; Butana n= 7; Karamoja, n=16; Kenana, n=10; Sheko, n=18 and Seleze Zebu, n=15) were analyzed in HAPMIX using a population of African taurine (Lagunaire, n=5; Muturu, n=20; N'Dama, n=20; Somba, n=10) and Asian indicine (Cholastani, n=13; Dajal, n=10; Lohani, n=13; Hariana, n=10; Gyr, n=20 and Nelore, n=20) animals as reference populations. A target population of 174 Near Eastern animals from the Anatolian peninsula (Anatolian Black, n=32; East Anatolians red, n=50; South Anatolian red, n=50 and Turkish Grey, n=42) were analyzed in HAPMIX using a population of European taurine (Hereford, n=20, Kerry, n=9, Romagnola, n=20, Red Poll,

n=10 and Galloway, n=9) and the same indicine animals as above for reference populations. Reference populations were chosen based on their ancestry estimated by ADMIXTURE analysis with the number of populations to assign being 3 (i.e. *K*=3); only animals that appeared to belong to pure European taurus, African taurus, or indicine were chosen. Furthermore only indicine breeds originating from Asia were selected as a reference population.

The parameters used for HAPMIX presumed that both ancestral populations contributed to the admixed population equally ($\theta = 0.5$) with both reference populations having the same recombination value of 600, mutation values for the reference and admixed populations of 0.2 and 0.1, respectively and equal miscopying rates in each reference population of 0.05. Reference populations were phased using BEAGLE version 3 [29, 136]. Recombination between SNPs was assumed to be the same, with 1 Mb representing 1 centiMorgan.

To determine the date since admixture each model was run at differing generation times in intervals from 10-6000 generations. The model with the log likelihood estimate closest to zero was the best fit for the data and was used as both the estimate for date of admixture and as the model to be used for mapping ancestry across the genome.

To determine regions with over-represented ancestral segments in the East African Zebu and Anatolians populations, the overall probability of a SNP being inherited from indicus ancestry was calculated as follows

$$P_{ind} = (px1) + (pqx0.5) + (qx0)$$

Where at each SNP p is the probability that an individual is homozygous for indicus, pq is the probability than an individual is heterozygous and q is the probability that an individual

is homozygous for taurus (African or European depending on population) ancestry. These scores were then visualized in the form of a Manhattan plot.

As a validation of our methodology HAPMIX analysis was performed on a population of 12 Brangus animals. A population of 20 Brahman and 20 Angus animals were used as a reference population. Brangus is a breed that was incepted in the United States by the USDA in the early 1900's with an initial breed composition of 5/8 Angus and 3/8 Brahman. Hapmix analysis was also performed on the New World breed the Carora, which are an admixed population which are thought to have been admixed in the late 1400-1500's [199]. Analysis was performed using the same protocol as mentioned above. These two breeds would provide a suitable comparison for the ability of HAPMIX to date absolute time of admixture.

4.2.4 Identifying signatures of selection using the integrated haplotype score

A whole genome scan for signatures of selection was performed on the East African Zebu and Anatolian populations listed above. To perform this analysis an extension of the extended haplotype homozygosity (EHH) statistic, the integrated haplotype score (iHS)[162] was calculated for each SNP in each population. iHS implementation requires both the ancestral and derived allele of each SNP to be known. To estimate the derived and ancestral SNP, the methodology implemented by Gautier & Naves [160] was used, which utilizes results from ADMIXTURE to determine ancestral and derived alleles. The ancestral allele was defined as the allele with the highest ancestral frequency estimated as

$$f_a = w_{TAU} f_{TAU} + w_{IND} f_{IND}$$

Where w_{TAU} and w_{IND} represent the average proportions of taurus and indicus ancestries in the target population, while f_{TAU} and f_{IND} represent the allele frequencies in the taurus and indicus reference populations. Calculation of the iHS statistic and plotting of the results were performed using the R package rehh [200] and genotypes were phased using BEAGLE.

Quantile-Quantile plots were produced from P-values estimated in all iHS analyses to investigate the extent to which the iHS estimated followed the expected (null) distribution. Deviation from the expected distribution could provide evidence towards systematic bias (*i.e.* unrecognized population structure, analytical approach or genotyping artefacts) in our analysis

4.2.5 Meta-analysis of admixture and integrated haplotype score

The results of both the iHS and Hapmix analysis were combined into a single analysis by multiplying the relative rankings of the SNPs in each analysis *i.e.* the genomewide rank of the SNPs significance divided by the number of SNPs in the analysis.

4.2.6 Identifying genes associated with adaption, admixture and disease

Regions formed from SNPs above genome wide significance pertaining to a false discovery rate of 1% in iHS analysis were investigated for the presence of genes. Significant SNPs were held to form a region if they were in proximity (within 100kb) to each other. If a region could not be formed around a SNP with significance greater than the FDR level of 1% then a 50kb window upstream and downstream of this SNP was

investigated. In HAPMIX analysis regions formed by SNPs which were in the top $1x10^{-4}$ ranked fraction for estimated probability of belonging to either ancestral population were investigated for the presence of genes. SNPs ranked in the top 100 for the combined HAPMIX/iHS analysis were investigated.

4.3 Results

4.3.1 Population structure

The principle source of genetic variation from principle component analysis (Figure 4.1) was between *Bos taurus* and *Bos indicus* lineages (principle component 1) with a clear separation of the indicus and taurus breeds. The second principle component (figure 4.1) represented the split between African taurus animals and European taurus and indicus animals. The first principle component represented 5.33% of the variation within the dataset while the second principle component represented 2.55% of the variation. East African Zebu breeds fell into a cluster between that of African taurus and indicus breeds, with the cluster positioned more toward indicus clusters than African taurus clusters. The Near Eastern Anatolian breeds, clustered between Northern European taurus and Indicus breeds, close to clusters of Southern European taurus and modern indicus/taurus hybrids (i.e. Beefmaster, Brangus and Santa Gertudis).

Mean number of IBD blocks between and within breed was converted to a distance matrix and plotted using multi-dimensional scaling - shown in figure 4.2. The first dimension split *Bos indicus* from *Bos taurus*, with the second dimension splitting European taurus and African taurus. Most of the animals did however cluster to the center of the plot, but the split between the populations was still observed. The top 20 within and between breed IBD block sharing for the East African Zebu population and Anatolian population are shown in tables 4.2 and 4.3, respectively. The East African Zebu population had an expectedly high level of IBD blocks shared within and between each other, with the next highest level of IBD sharing seen between African indicus breeds such as Azawak, Ankole and Gobra, furthermore a medium level of IBD (compared to within breed IBD) sharing

was estimated between the Seleze Zebu and Nganda, a known hybrid of East African Zebu and Ankole Sanga. As with the East African Zebu population the Anatolian population had an expectedly high level of IBD sharing within each breed. A medium level of IBD (compared to within breed IBD) sharing was estimated between Anatolian breeds and southern European breeds (Sikias, Alentejana), African taurus (Kuri) and indicus (Sahiwal).

Table 4.2 top 20 mean IBD blocks sharing scores between/within East African zebu animals and other breeds

Breed 1	Breed 2	Mean number of IBD blocks shared
Kenana	Kenana	137.55
Butana	Butana	83.86
Sheko	Sheko	79.33
Seleze Zebu	Seleze Zebu	65.78
Karamoja	Karamoja	24.08
Sahiwal	Seleze Zebu	23.44
East African Zebu	East African Zebu	13.80
East African Zebu	Karamoja	7.28
Azawak	East African Zebu	7.07
Butana	Kenana	6.94
Ankole	East African Zebu	6.07
Karamoja	Seleze Zebu	5.85
Azawak	Karamoja	5.68
Ankole	Karamoja	5.60
Nganda	Seleze Zebu	5.04
East African Zebu	Sheko	4.82
East African Zebu	Seleze Zebu	4.76
Azawak	Seleze Zebu	4.33
Karamoja	Sheko	4.24
Gobra	Kenana	4.20

Table 4.3 top 20 mean IBD blocs sharing scores between Anatolian animals and other breeds

Breed 1	Breed 2	Mean number of IBD blocks shared
Turkish Gray	Turkish Gray	91.64
East Anatolian Red	East Anatolian Red	52.96
South Anatolian Red	South Anatolian Red	48.26
Anatolian Black	Anatolian Black	18.50
Sikia	Turkish Gray	4.60
Kuri	East Anatolian Red	3.69
Anatolian Black	South Anatolian Red	3.44
Baladi	South Anatolian Red	3.30
South Anatolian Red	East Anatolian Red	3.20
South Anatolian Red	Turkish Gray	3.05
Anatolian Black	Kuri	3.01
Anatolian Black	South Anatolian Red	2.46
Sahiwal	Anatolian Black	2.26
Baladi	East Anatolian Red	1.61
Sikia	East Anatolian Red	1.22
Alentejana	East Anatolian Red	1.21
Anatolian Black	Turkish Gray	1.19
Kuri	Turkish Gray	1.06
Montbeliard	East Anatolian Red	0.99
Simmental	East Anatolian Red	0.99

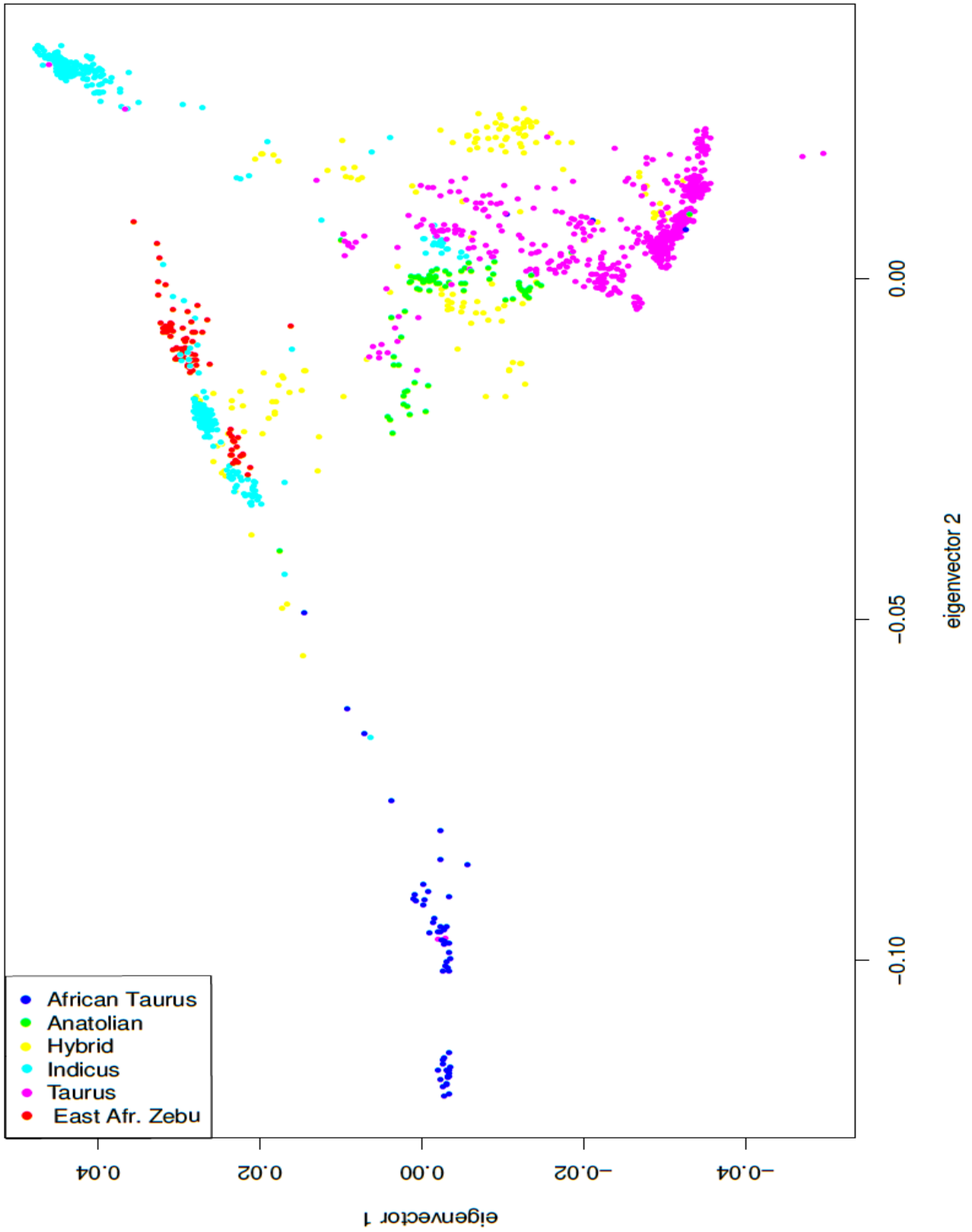


Figure 4.1 Principle component analysis results obtained from a subset of the data-set (n=1,323) representing all of the breeds in the dataset.

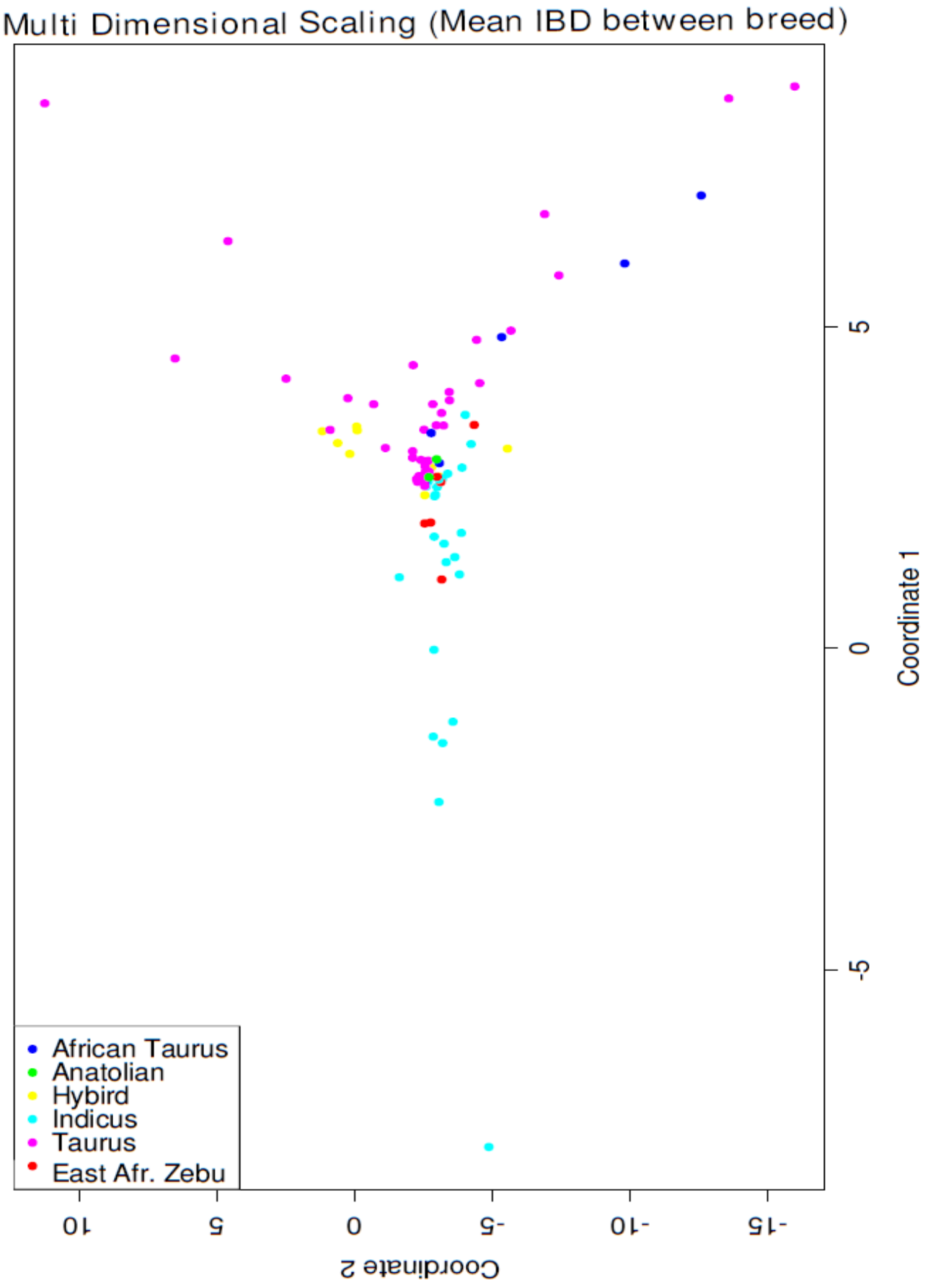


Figure 4.2 Multi-dimensional scaling of the mean number of IBD blocks shared within and between breeds in the data-set.

Multi-dimensional scaling of the IBS estimates are shown in figure 4.3. The IBS estimates mirrored that of the principle component analysis with the first dimension splitting the *Bos indicus* and *Bos taurus* lineages, and the second dimension splitting the European taurus from African taurus breeds. The same clustering of East African Zebu breeds and Near Eastern Anatolian breeds was observed in the IBS estimates as was seen in the principle component analysis, described previously in this section.

Results from ADMIXTURE analysis with K=3 ancestral populations are shown in figure 4.4. ADMIXTURE results for K=3 were in best agreement with the first and second principle components and best represented our understanding of cattle ancestry. Each of the East African Zebu breeds showed a high level of admixture between indicus and African taurus ancestries, with each showing slightly more indicus than African taurus ancestry, with a population average of 33% for African taurus and 66% for indicus ancestry. The Anatolian breeds all showed admixture for the European taurus, African taurus and indicus ancestries, with more taurus ancestry seen in each of the four Anatolian breeds, with a population average of 51% European taurus, 23% African taurus and 25% indicus. Similar admixture patterns to that of the Anatolian breeds were observed in other breeds such as Sikias, Taleshi, Baladi, Curraleiro, Creole and Belmont Red which all have similar breed histories, all of which are hybrids between Indicus/Taurus breeds which were incepted in the last 4-5 centuries.

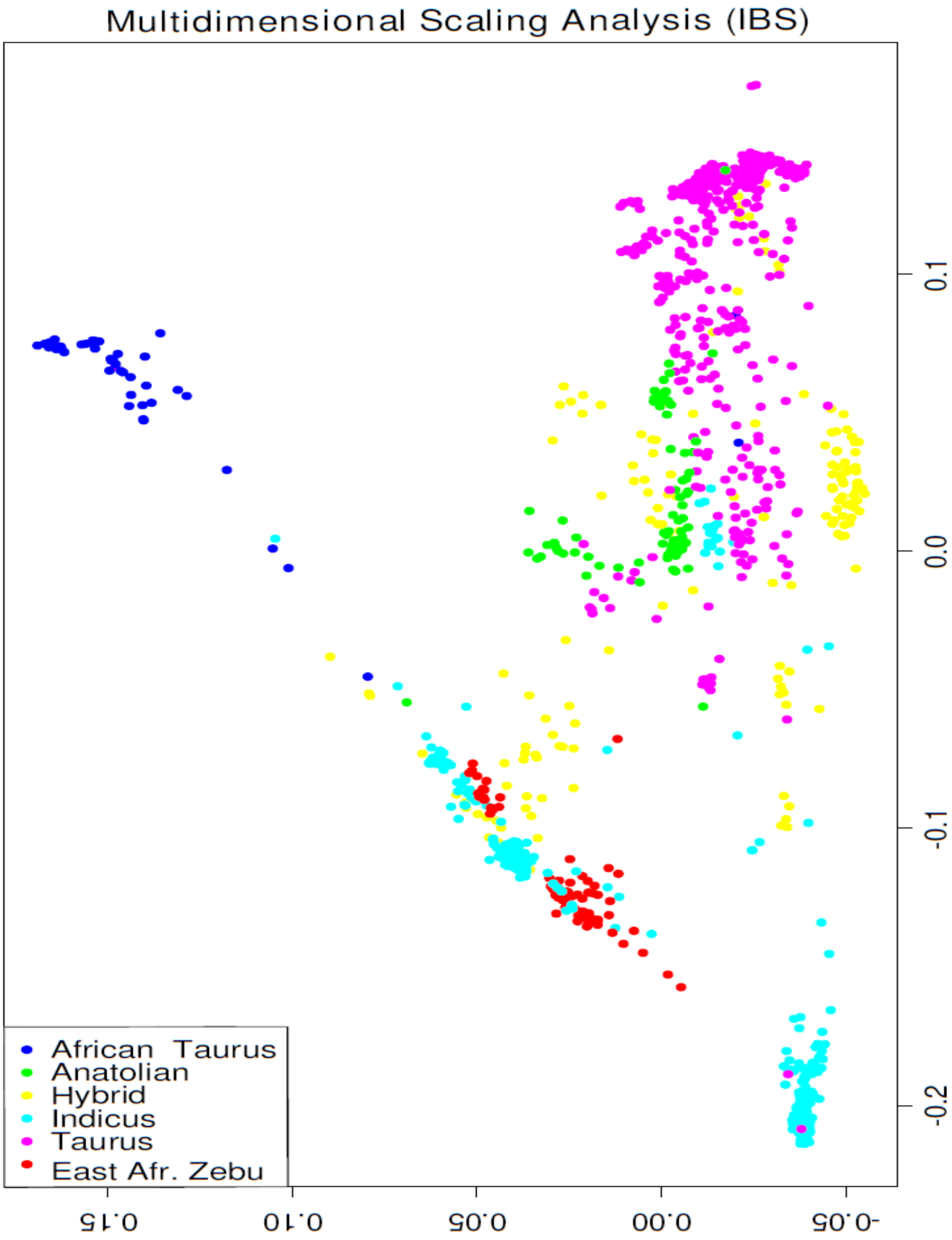


Figure 4.3 Multi-dimensional scaling of the identity by state results between and within individual animals in our data-set

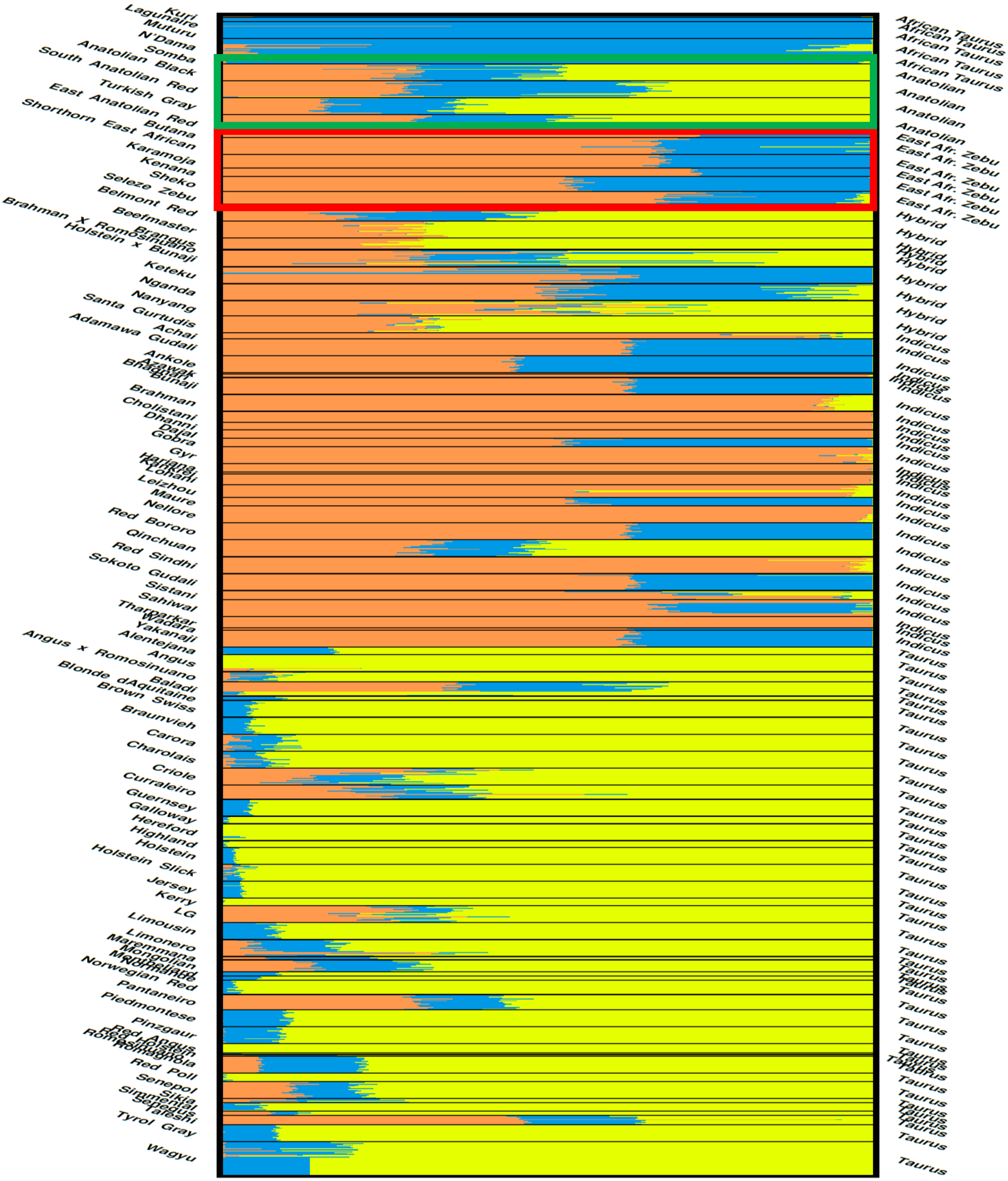


Figure 4.4 Admixture results for K=3 ancestral populations (yellow=European taurus, blue=African taurus, orange= indicus). Red and green squares indicate the target populations; East African zebu and Anatolian breeds respectively.

4.3.2 Ancestry mapping and dating time since admixture

The mapping of ancestral proportions can aid in identifying ancestral segments which have undergone selective pressure in admixed populations. The genomewide probability that a SNP was inherited from indicus or African taurus ancestry in the East African Zebu population is shown in figure 4.5. East African Zebu were estimated at having more indicus ancestry than African taurus with an overall probability of indicus ancestry for all genotypes of 0.617±0.113. Investigation of the top 60 SNPs, which represent the top $1x10^{-4}$ fraction of ranked positions for proportion of indicus ancestry (Table 4.4) revealed two genes, *HERC3* and *TLR3*, on chromosomes 6 and 27, respectively, both of which have involvement in immunity. Investigation of the top 60 SNPs for proportion of African taurus ancestry (Table 4.5) identified three genes; *AOAH*, *TUSC3* and *NUP133*, on chromosomes 4, 27 and 28, respectively.

The genome-wide probability that a SNP was inherited from indicus or European taurus ancestry in Anatolian breeds is shown in figure 4.6. Anatolian breed were estimated at containing slightly more European taurus ancestry than indicus ancestry, with an overall probability of European taurus ancestry for all genotypes of 0.569±0.117. Investigation of the top $1x10^{-4}$ of SNPs for ranked levels of indicus ancestry (Table 4.6) identified three genes within these regions; *LHFPL3*, *KCNIP1* and *CDH10*. Two genes were identified within regions formed by the top $1x10^{-4}$% of SNPs for estimated taurus ancestry *KYNU* on chromosome 2 and *FUT6* on chromosome 7 (Table4.7).

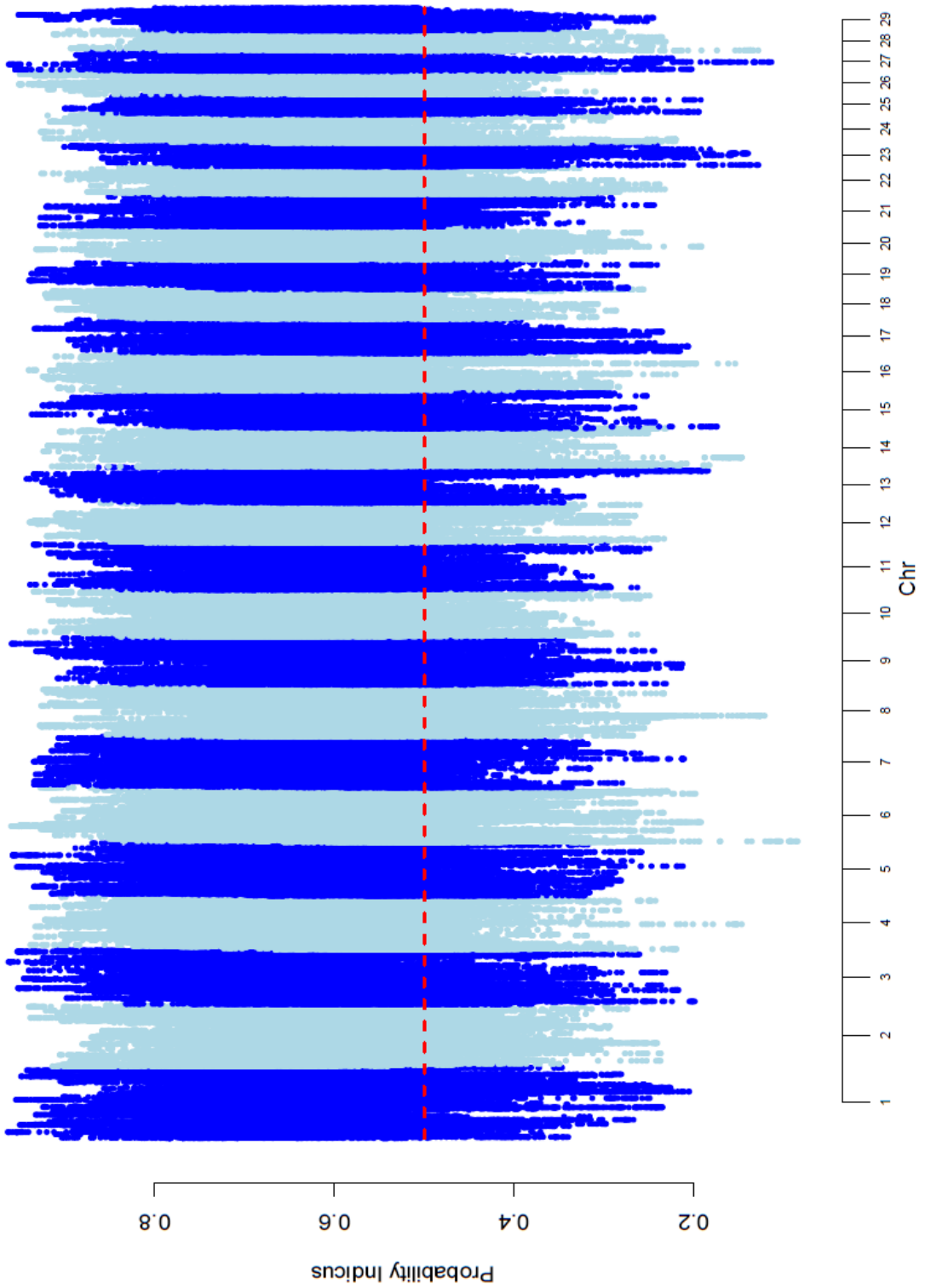


Figure 4.5 Genome wide results for HAPMIX analysis of East African Zebu breeds. Red dashed line indicates regions with equal probability of Taurus or Indicus ancestry.

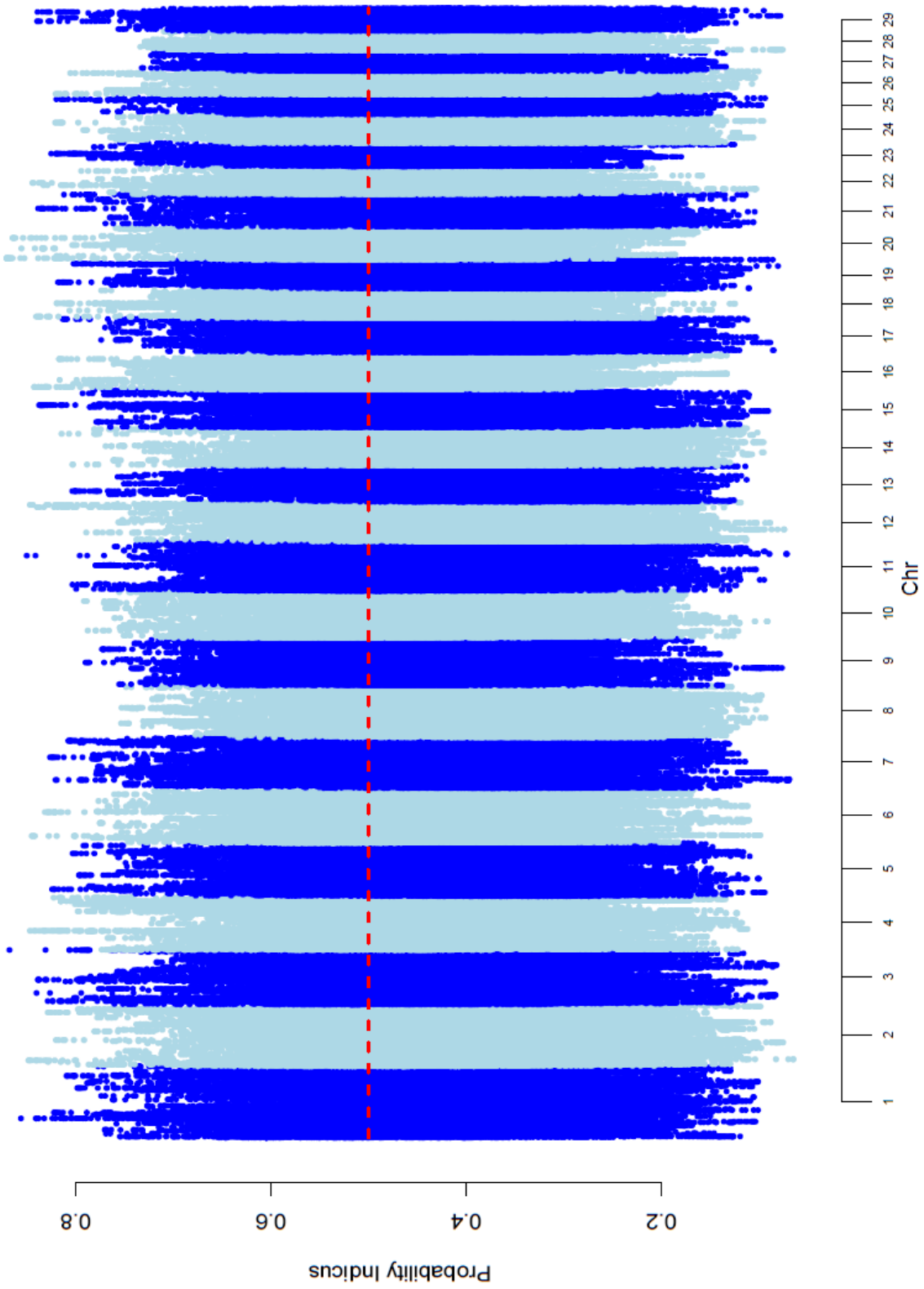


Figure 4.6 Genome wide results for HAPMIX analysis of Anatolian breeds. Red dashed line indicates regions with equal probability of Taurus or Indicus ancestry.

Table 4.4 Top 60 SNPs for probability of indicus ancestry from HAPMIX analysis of East African Zebu breeds.

Chr	SNP	Position	Prob. Afr. taurus	Prob. indicus	Genes
1	ARS-BFGL-BAC-11034	12839893	0.002188	0.917493	
1	BovineHD0100003970	12842398	0.002043	0.921739	none
1	BovineHD0100003972	12846641	0.001739	0.927855	
1	BovineHD0100003973	12848174	0.001797	0.92742	
1	BovineHD0100003974	12850144	0.001841	0.926565	
3	BovineHD0300010890	35105890	0.013594	0.915174	
3	BovineHD0300010891	35106756	0.013667	0.916565	
3	BovineHD0300010892	35110282	0.013667	0.916681	none
3	Hapmap49798-BTA-92389	35113194	0.013623	0.917159	
3	ARS-BFGL-NGS-2680	35133372	0.013725	0.915783	
3	BovineHD0300010896	35135984	0.013725	0.915449	
3	BovineHD0300027824	97045586	0.002217	0.915696	
3	BovineHD0300027826	97048623	0.001841	0.923145	
3	BovineHD0300027827	97049900	0.001623	0.926507	
3	BovineHD0300027828	97052393	0.00171	0.925101	
3	BovineHD0300027829	97054346	0.001768	0.924116	
3	BovineHD0300027830	97055401	0.001826	0.923768	
3	BovineHD0300027832	97059595	0.002145	0.923739	
3	BovineHD0300027833	97060822	0.002188	0.923826	none
3	BovineHD0300027834	97062821	0.00229	0.923507	
3	BovineHD0300027835	97063935	0.002319	0.923362	
3	BovineHD0300027836	97065471	0.002377	0.923159	
3	BovineHD0300027837	97066701	0.002609	0.922884	
3	BovineHD0300027839	97068397	0.00287	0.922681	
3	BTA-35586-no-rs	97069254	0.002942	0.922667	
3	BovineHD0300027840	97069988	0.003043	0.922087	
3	BovineHD0300034252	117570370	0.011145	0.91658	none
3	BovineHD0300034253	117571177	0.011203	0.915507	
5	BovineHD0500026972	94968128	0.006986	0.91613	
5	BovineHD0500026973	94970635	0.007232	0.917594	
5	BovineHD0500026974	94974273	0.007449	0.920797	
5	BovineHD0500026975	94976515	0.007652	0.921522	
5	BovineHD0500026976	94978574	0.007928	0.922362	
5	BovineHD0500026977	94979447	0.008	0.922884	none
5	BovineHD0500026978	94980123	0.008072	0.923072	
5	BovineHD0500026979	94981190	0.008087	0.92342	
5	BovineHD0500026980	94981747	0.00813	0.923609	
5	BovineHD0500026982	94984140	0.008101	0.922725	
5	BovineHD0500026983	94986942	0.008203	0.919145	
5	BovineHD0500026984	94988733	0.008319	0.916783	
6	BovineHD0600034276	37541623	0.001841	0.915058	*HERC3*
6	BovineHD0600010413	37544792	0.001667	0.920652	
9	BovineHD4100007670	93701421	0.002928	0.920174	
9	BovineHD0900026520	93702202	0.002884	0.920855	none
9	BovineHD0900026522	93704329	0.002928	0.918768	
9	BovineHD0900026523	93705281	0.002942	0.918739	
9	BovineHD0900026524	93708507	0.003435	0.916768	
11	BovineHD1100019683	69276596	0.002203	0.91629	
11	BovineHD1100019684	69277444	0.002188	0.918261	none
11	BovineHD1100019685	69278067	0.00213	0.919638	
11	BovineHD1100019686	69278998	0.002145	0.919348	
11	BovineHD1100019687	69279862	0.002217	0.916986	

27	BovineHD2700004406	15239159	0.00313	0.915609	
27	BovineHD2700004407	15240037	0.003029	0.917797	
27	BovineHD2700004408	15241354	0.002913	0.921145	
27	BovineHD2700004409	15244179	0.002681	0.926928	*TLR3*
27	BovineHD2700004410	15245034	0.002768	0.925884	
27	BovineHD2700004411	15245767	0.002841	0.924652	
27	BovineHD2700004412	15247313	0.003058	0.920232	
27	BovineHD2700004413	15248419	0.003217	0.917638	

Table 4.5 Top 60 SNPs for probability of African taurus ancestry from HAPMIX analysis of East African Zebu breeds.

Chr	SNP	position	Prob. Afr. taurus	Prob. indicus	Genes
4	BovineHD0400016660	61012381	0.774333	0.079536	
4	BovineHD0400016661	61013251	0.783971	0.078188	*AOAH*
4	ARS-BFGL-NGS-115237	61015911	0.773449	0.077754	
6	BovineHD0600000744	2985855	0.812304	0.030174	
6	BovineHD0600000745	2986904	0.803652	0.031783	
6	BovineHD0600000746	2988278	0.792319	0.033826	
6	BovineHD0600000748	2991082	0.81829	0.028913	
6	BovineHD0600000749	2997295	0.819855	0.027348	
6	BovineHD0600000750	2998027	0.831609	0.014072	
6	BovineHD0600000751	2998606	0.82571	0.014797	
6	BovineHD0600000752	3001129	0.828261	0.014145	
6	BovineHD0600000753	3001917	0.828986	0.013942	none
6	BovineHD0600000754	3002810	0.829	0.013928	
6	BovineHD0600000755	3004083	0.839957	0.012435	
6	BovineHD0600000756	3005935	0.842319	0.012217	
6	BovineHD0600000757	3007736	0.838449	0.012957	
6	Hapmap30828-BTA-143720	3011115	0.827681	0.014696	
6	BovineHD0600000758	3014517	0.804101	0.016652	
6	BovineHD0600000759	3015449	0.801029	0.017261	
6	BovineHD0600000760	3016178	0.787145	0.020464	
8	BovineHD0800013223	44267279	0.795725	0.073725	
8	BovineHD0800013224	44269229	0.780913	0.074928	
8	BTA-81130-no-rs	44276510	0.816725	0.072087	
8	BovineHD0800013226	44277181	0.818609	0.071957	
8	BovineHD0800013227	44277865	0.818884	0.070884	
8	BovineHD0800013228	44279003	0.816319	0.069391	
8	BovineHD0800013229	44280462	0.813087	0.067087	
8	BovineHD0800013230	44282167	0.810217	0.064478	none
8	BovineHD0800013231	44285286	0.808058	0.051522	
8	BovineHD0800013232	44288048	0.797826	0.051507	
8	BovineHD0800013233	44289192	0.791652	0.050957	
8	BovineHD0800013234	44290619	0.794174	0.051145	
8	BovineHD0800013235	44292707	0.789594	0.050725	
8	BovineHD0800013236	44293365	0.788725	0.050652	
8	BovineHD0800013237	44295539	0.783478	0.050768	
8	BovineHD0800013239	44297173	0.77742	0.051507	
11	BovineHD1100004785	14881149	0.782522	0.08971	none
11	BovineHD1100004786	14882052	0.777507	0.089667	
23	BovineHD2300000942	4117893	0.779232	0.038551	none
23	BovineHD2300000943	4120663	0.781246	0.038145	
23	BovineHD2300000944	4123451	0.776928	0.038435	
27	BovineHD2700005815	20378644	0.780435	0.025435	
27	BovineHD2700005816	20380816	0.775391	0.026986	
27	BovineHD2700005817	20384687	0.78658	0.022565	
27	BovineHD2700005818	20385474	0.788116	0.021942	*TUSC3*
27	BovineHD2700005819	20393556	0.794319	0.022174	
27	BovineHD2700005820	20398371	0.791246	0.024391	
27	BovineHD2700005833	20465016	0.772826	0.085768	
28	BovineHD2800000198	452551	0.775029	0.053406	
28	BovineHD2800000199	453342	0.78029	0.053391	
28	BovineHD2800000200	454388	0.779362	0.053609	*NUP133*
28	BovineHD2800000201	455302	0.778754	0.053754	

28	BTB-00972943	456341	0.775072	0.054565
28	BovineHD2800000206	464645	0.776261	0.058739
28	BovineHD2800000207	465366	0.782261	0.057594
28	BovineHD2800000208	466141	0.788536	0.056536
28	BovineHD2800000209	467472	0.797957	0.055217
28	BovineHD2800000210	468715	0.786942	0.056406
28	BovineHD2800000211	469673	0.771696	0.058551

Table 4.6 Top 60 SNPs for probability of indicus ancestry from HAPMIX analysis of Anatolian breeds.

Chr	SNP	position	Prob. taurus	Prob. indicus	Genes
2	BovineHD0200003687	13152306	0.078265823	0.75978481	
2	BovineHD0200003688	13155017	0.072481013	0.763037975	none
2	BovineHD0200003689	13155951	0.071063291	0.764924051	
2	BovineHD0200003690	13157159	0.069518987	0.765101266	
4	BovineHD0400012545	45954167	0.082316456	0.763227848	
4	BovineHD0400012546	45955629	0.088822785	0.761835443	*LHFPL3*
4	BovineHD0400012547	45956971	0.094113924	0.760139241	
4	BovineHD0400012548	45958623	0.094772152	0.758848101	
7	BovineHD0700020147	68791854	0.119405063	0.760088608	none
7	BovineHD0700020148	68793058	0.112936709	0.759987342	
15	BovineHD1500014819	51535402	0.108987342	0.766379747	
15	BovineHD1500014820	51536580	0.108481013	0.772405063	
15	BovineHD1500014822	51537880	0.108265823	0.775734177	
15	BovineHD1500014823	51538412	0.108202532	0.777151899	
15	BovineHD1500014824	51539316	0.108126582	0.779582278	
15	BovineHD1500014825	51539773	0.108088608	0.780860759	
15	BovineHD1500014826	51540454	0.108088608	0.781848101	
15	BovineHD1500014827	51541107	0.108151899	0.77943038	
15	BovineHD1500014828	51541601	0.108113924	0.777936709	none
15	BovineHD1500014829	51542159	0.108075949	0.776189873	
15	BovineHD1500014830	51542614	0.107936709	0.774848101	
15	BovineHD1500014831	51543394	0.107835443	0.772607595	
15	BovineHD1500014832	51543970	0.107708861	0.770911392	
15	BovineHD1500014833	51545180	0.107594937	0.767708861	
15	BovineHD1500014834	51546473	0.107468354	0.764316456	
15	BovineHD1500014835	51548685	0.107341772	0.75921519	
15	BovineHD1500014851	51581813	0.132949367	0.761443038	
15	BovineHD1500014854	51585025	0.137556962	0.768417722	
18	BovineHD1800001108	3766436	0.087240506	0.764037975	none
18	BovineHD1800001109	3767512	0.087493671	0.767278481	
18	BovineHD1800001110	3769954	0.087050633	0.762012658	
20	BovineHD2000000857	2385439	0.034468354	0.766240506	
20	BovineHD2000000858	2386456	0.034253165	0.770367089	
20	BovineHD2000000859	2386904	0.034240506	0.77078481	
20	BovineHD2000000860	2387665	0.033772152	0.771987342	
20	BovineHD2000000861	2388477	0.033417722	0.772987342	
20	BovineHD2000000862	2389250	0.033481013	0.772291139	
20	ARS-BFGL-NGS-85031	2389754	0.03335443	0.773088608	
20	BovineHD2000000863	2390712	0.033291139	0.773721519	
20	BovineHD2000000864	2392037	0.033746835	0.773708861	*KCNIP1*
20	BovineHD2000000865	2392491	0.033797468	0.773531646	
20	BovineHD2000000866	2392939	0.033962025	0.773392405	
20	BovineHD2000000867	2393779	0.034607595	0.771	
20	BovineHD2000000868	2394909	0.03556962	0.764253165	
20	BovineHD4100014454	2396684	0.034139241	0.765493671	
20	BovineHD2000006926	22970215	0.10064557	0.764632911	
20	BovineHD2000006927	22973057	0.094329114	0.761531646	none
20	BovineHD2000006928	22979292	0.092620253	0.759151899	
20	Hapmap49634-BTA-50049	22985128	0.078075949	0.760772152	
20	BovineHD2000006932	22987991	0.072481013	0.782594937	
20	BovineHD2000013641	49011636	0.053341772	0.776189873	
20	BovineHD2000013642	49012381	0.052379747	0.778417722	*CDH10*
20	BovineHD2000013643	49015441	0.052329114	0.778734177	

20	BovineHD2000013644	49018665	0.051974684	0.780265823	
20	BovineHD2000013645	49021073	0.053367089	0.760405063	
29	BovineHD2900006239	21768418	0.092860759	0.770746835	none
29	BovineHD2900006240	21770469	0.091012658	0.767443038	
29	BovineHD2900006241	21772263	0.089911392	0.761936709	

Table 4.7. Top 60 SNPs for probability of taurus ancestry from HAPMIX analysis of Anatolian breeds.

Chr	SNP	position	Prob. taurus	Prob. indicus	Genes
1	BovineHD0100036254	128288325	0.84964557	0.055683544	
1	BovineHD0100036255	128289131	0.853012658	0.055886076	none
1	BovineHD0100036256	128289890	0.851873418	0.056253165	
1	BovineHD0100036257	128291751	0.84935443	0.057151899	
2	BovineHD0200004401	15471127	0.874303797	0.015075949	
2	BovineHD0200004402	15471807	0.88421519	0.015	none
2	BovineHD0200004403	15472537	0.883974684	0.014962025	
2	BovineHD0200004404	15474919	0.850443038	0.016455696	
2	BovineHD0200015338	53842667	0.848962025	0.025367089	
2	BovineHD0200015339	53843645	0.854	0.02443038	
2	BovineHD0200015340	53844875	0.859848101	0.023227848	
2	BovineHD0200015341	53846259	0.869759494	0.021481013	
2	BovineHD0200015342	53846985	0.861468354	0.020721519	
2	BovineHD0200015343	53848699	0.858367089	0.020151899	*KYNU*
2	BovineHD0200015344	53849656	0.854721519	0.019405063	
2	BovineHD0200015355	53893832	0.857101266	0.023620253	
2	BTB-01390801	53894600	0.857227848	0.023873418	
2	BovineHD0200015356	53895769	0.857417722	0.023810127	
2	BovineHD0200015357	53898497	0.865379747	0.022911392	
7	BovineHD0700005495	19736297	0.860810127	0.005063291	
7	BovineHD0700005496	19737393	0.861101266	0.004987342	
7	BovineHD0700005497	19738587	0.862379747	0.004848101	*FUT6*
7	BovineHD0700005498	19739162	0.864253165	0.004734177	
7	BovineHD0700005499	19740182	0.864708861	0.005139241	
7	BovineHD0700005500	19741020	0.862240506	0.005544304	
7	BovineHD0700006239	22673488	0.849708861	0.013417722	
7	BovineHD0700006240	22675834	0.866772152	0.011898734	none
7	BovineHD0700006241	22676731	0.872658228	0.01135443	
7	ARS-BFGL-NGS-117268	22681472	0.856556962	0.012189873	
9	BovineHD0900011278	40654924	0.853063291	0.026202532	
9	BovineHD0900011279	40656111	0.861860759	0.026202532	
9	BovineHD0900011280	40657173	0.866050633	0.026924051	none
9	BovineHD0900011281	40658860	0.873873418	0.027759494	
9	BovineHD0900011282	40659768	0.864556962	0.028265823	
9	BovineHD0900011283	40661207	0.857873418	0.02864557	
11	BovineHD1100023318	81569662	0.864379747	0.009291139	
11	BovineHD1100023319	81570557	0.864607595	0.00843038	none
11	BovineHD1100023320	81571853	0.862620253	0.007316456	
12	BovineHD1200007538	24911124	0.855898734	0.005379747	none
12	BovineHD1200007539	24913958	0.855683544	0.00535443	
17	BovineHD1700006233	21664275	0.856632911	0.043265823	
17	BovineHD1700006234	21664881	0.864911392	0.042848101	none
17	ARS-BFGL-BAC-35973	21665559	0.860759494	0.042924051	
28	BovineHD2800000717	2044751	0.851683544	0.025379747	
28	BovineHD2800000718	2045860	0.856278481	0.024379747	
28	BovineHD2800000719	2046737	0.857683544	0.023860759	
28	ARS-BFGL-NGS-60834	2047488	0.854189873	0.023886076	
28	BovineHD2800000722	2050339	0.853987342	0.022329114	none
28	BovineHD2800000723	2051287	0.855835443	0.021721519	
28	BovineHD2800000724	2052158	0.857810127	0.021265823	
28	BovineHD2800000725	2053843	0.856987342	0.020873418	
28	BovineHD2800000728	2080478	0.85164557	0.027063291	
28	BovineHD2800000731	2085101	0.870303797	0.02743038	

28	BovineHD2800000733	2087602	0.875164557	0.027974684
28	BovineHD2800000734	2090208	0.877594937	0.031949367
28	BovineHD2800000735	2091193	0.874696203	0.032822785
28	BovineHD2800000736	2091837	0.871253165	0.032759494
28	ARS-BFGL-NGS-57890	2092581	0.863556962	0.033278481

Dating of admixture by HAPMIX estimated the time since admixture for East African Zebus at 2000 generations and the time since admixture for the Anatolian breeds at 3000 generations. The log likelihood curve of the dates of admixture for East African Zebu and Anatolians are shown in figures 4.7 and 4.8, respectively. The date of admixture for each chromosome individually for the East African Zebu and Anatolian breeds are shown in figures 4.9 and 4.10, respectively. The range of date of admixture per chromosome in the East African Zebu breeds was from 1700 to 3000 generations. The range of the date of admixture per chromosome in the Anatolian breeds was from 2000 to 4500 generations.

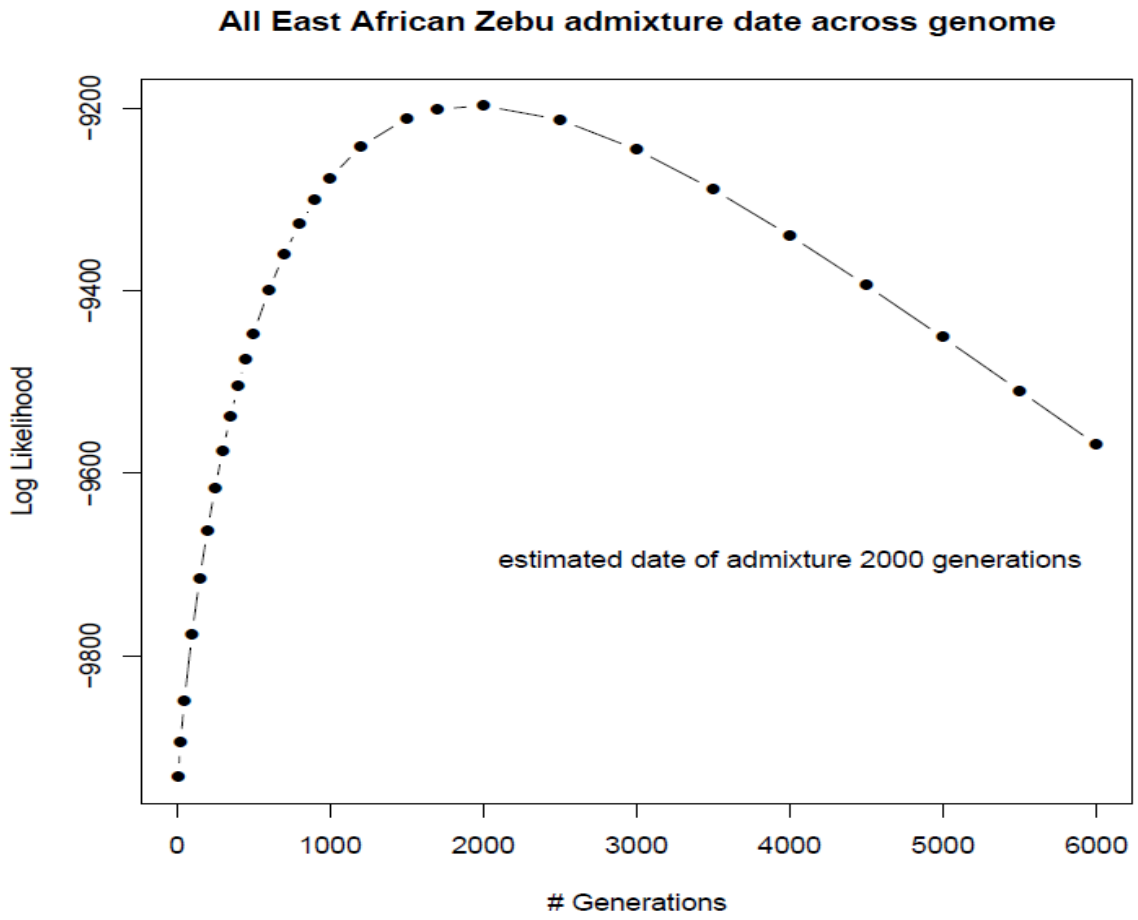


Figure 4.7 Log likelihood estimates for the number of generations since admixture for East African Zebu breeds, estimated in HAPMIX

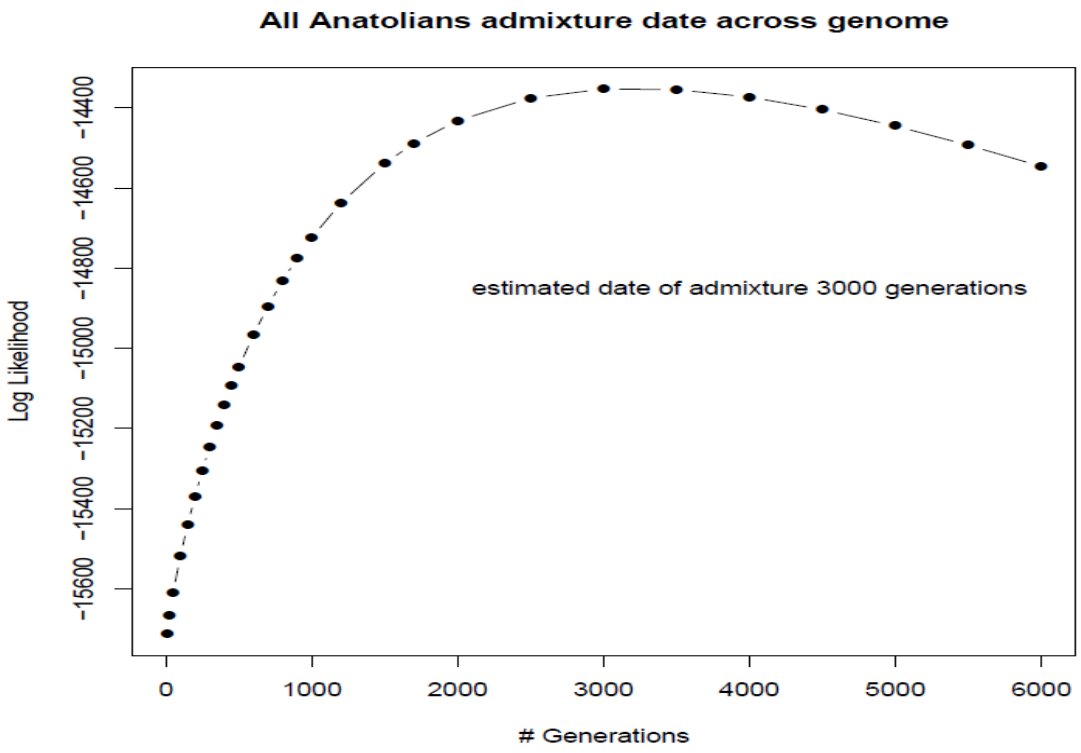


Figure 4.8 Log likelihood estimates for the number of generations since admixture for Anatolian breeds, estimated in HAPMIX

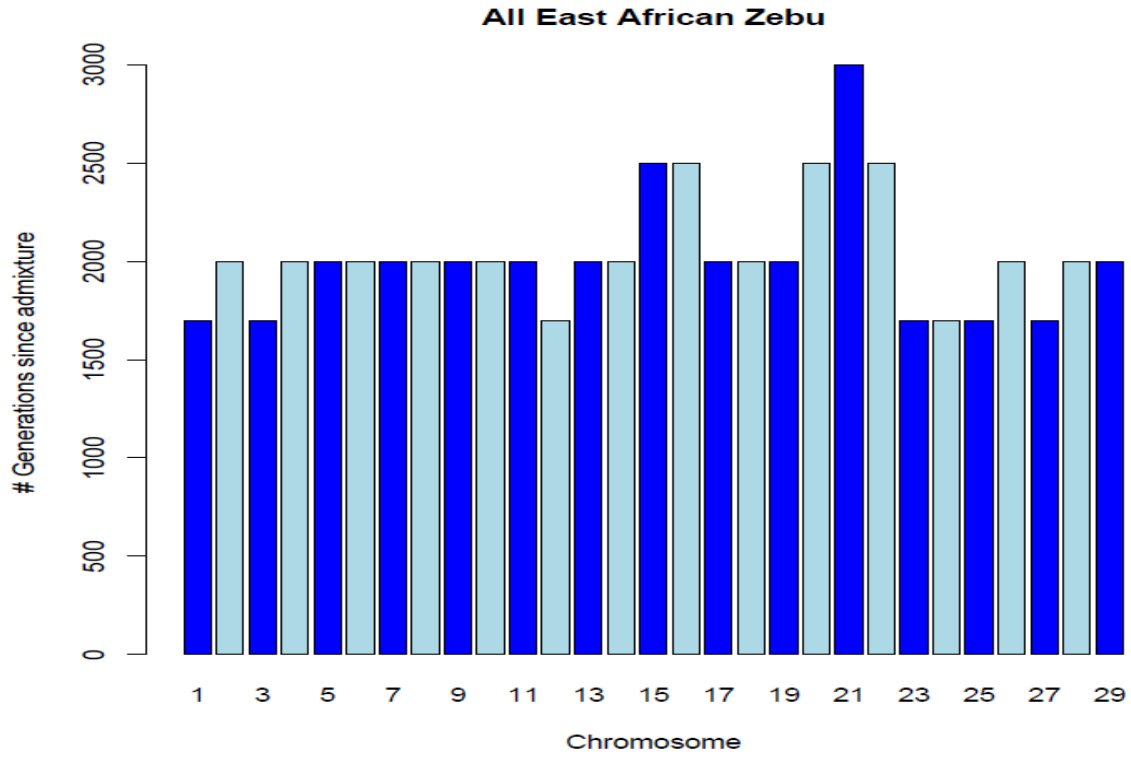


Figure 4.9 Estimates of time since admixture per chromosome for East African Zebu breeds. Estimated in HAPMIX

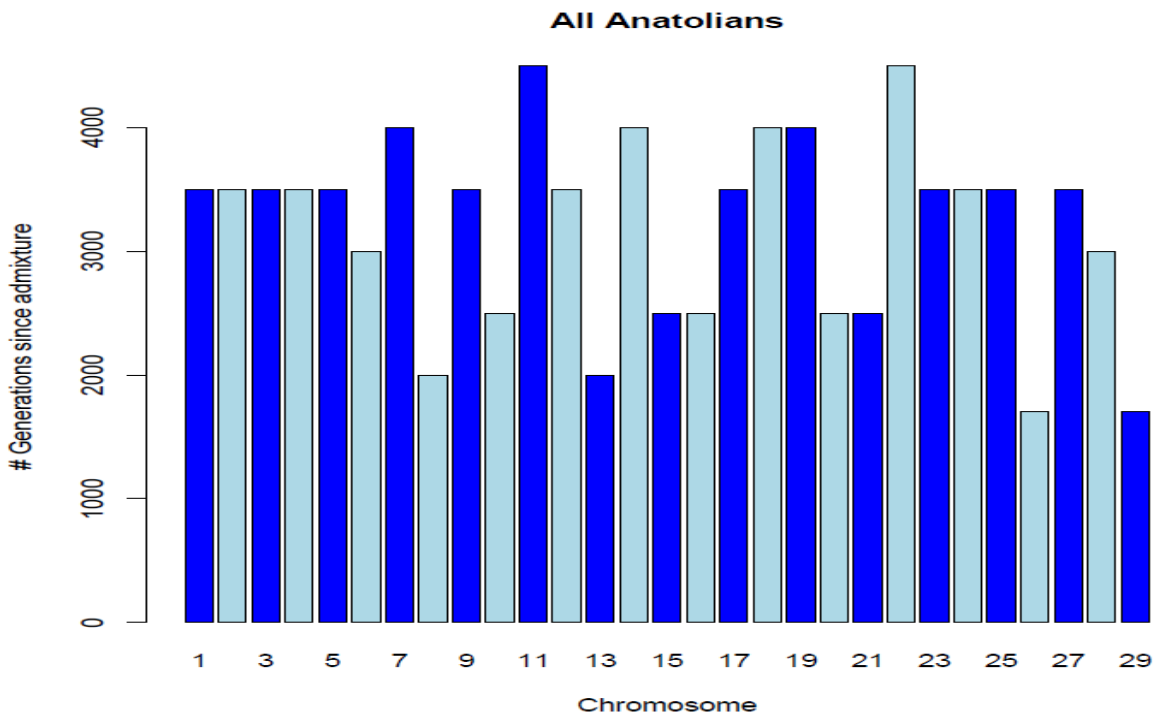


Figure 4.10 Estimates of time since admixture per chromosome for Anatolian breeds. Estimated in HAPMIX

The estimated date of admixture for the Brangus breed by HAPMIX was 1700 generations (figure 4.11) and the estimated date of admixture for the Carora breed was 2000 generations (figure 4.12). The range in date of admixture per chromosome in the Brangus and Carora breeds were 1000-3000 (figure 4.13) and 700-4500 (figure 4.14) generations, respectively.

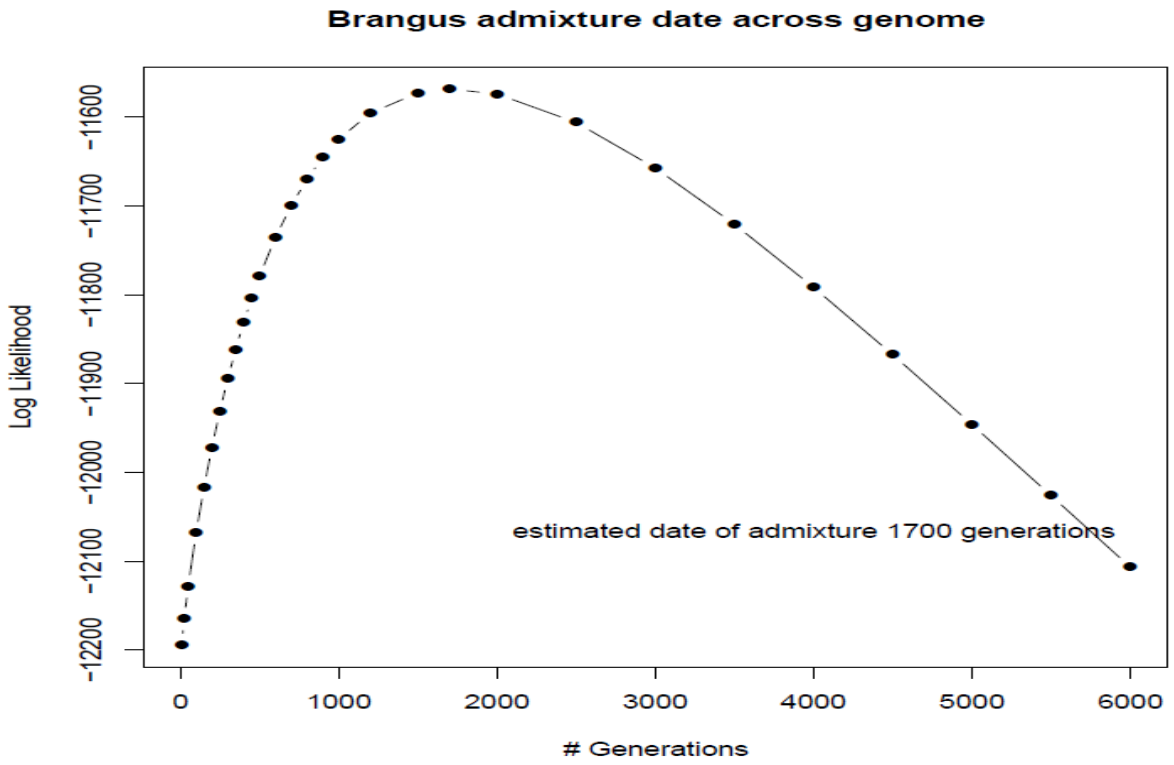


Figure 4.11 Log likelihood estimates for the number of generations since admixture for Brangus animals, estimated in HAPMIX

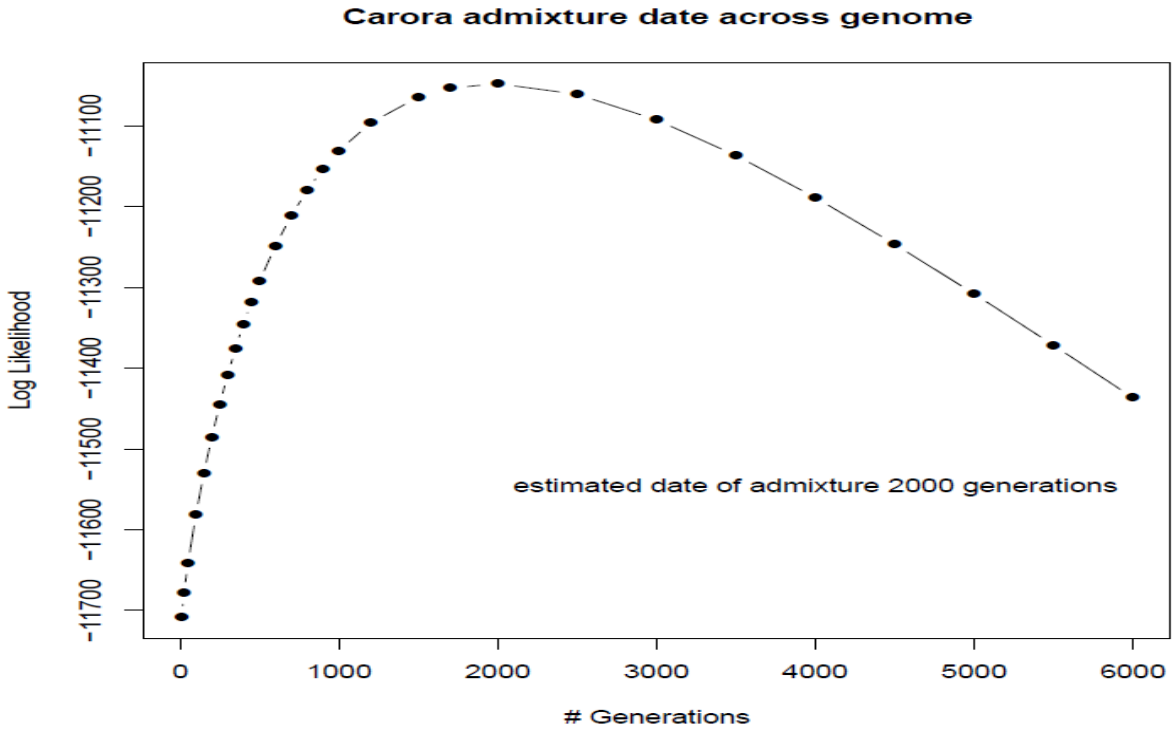


Figure 4.12 Log likelihood estimates for the number of generations since admixture for Carora animals, estimated in HAPMIX

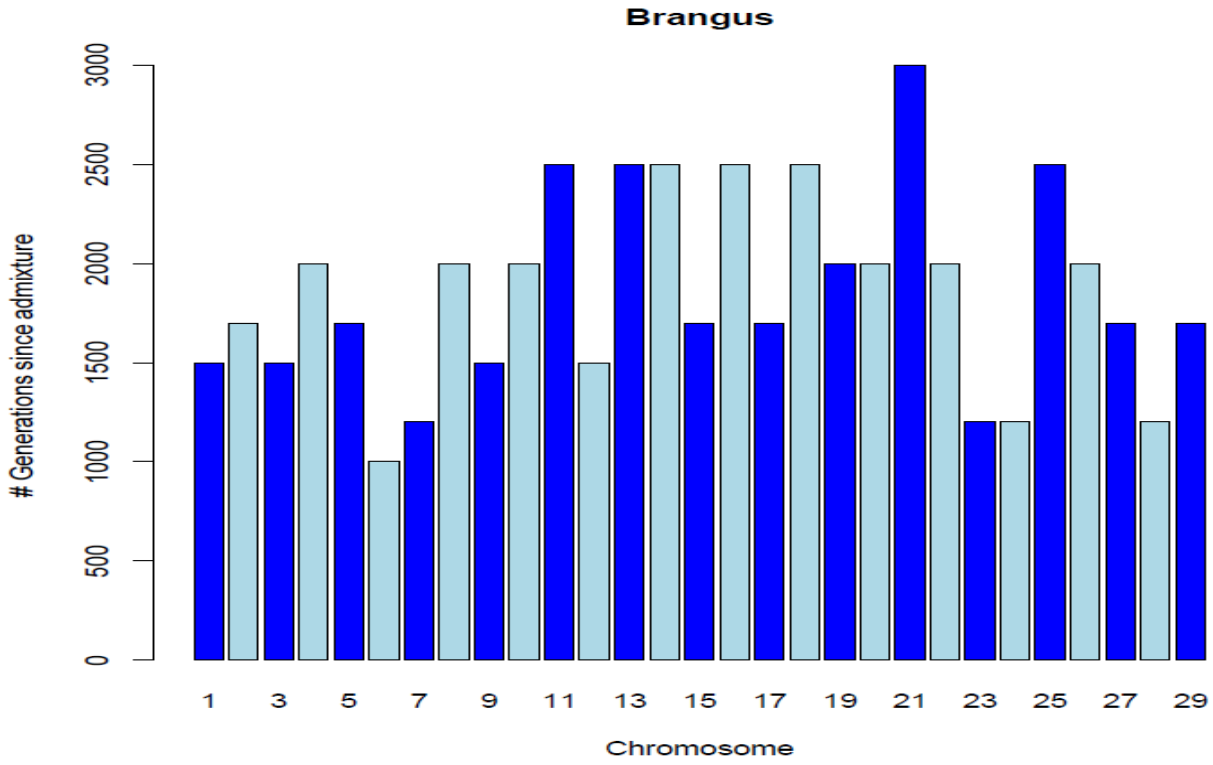


Figure 4.13 Estimates of time since admixture per chromosome for Brangus Animals. Estimated in HAPMIX

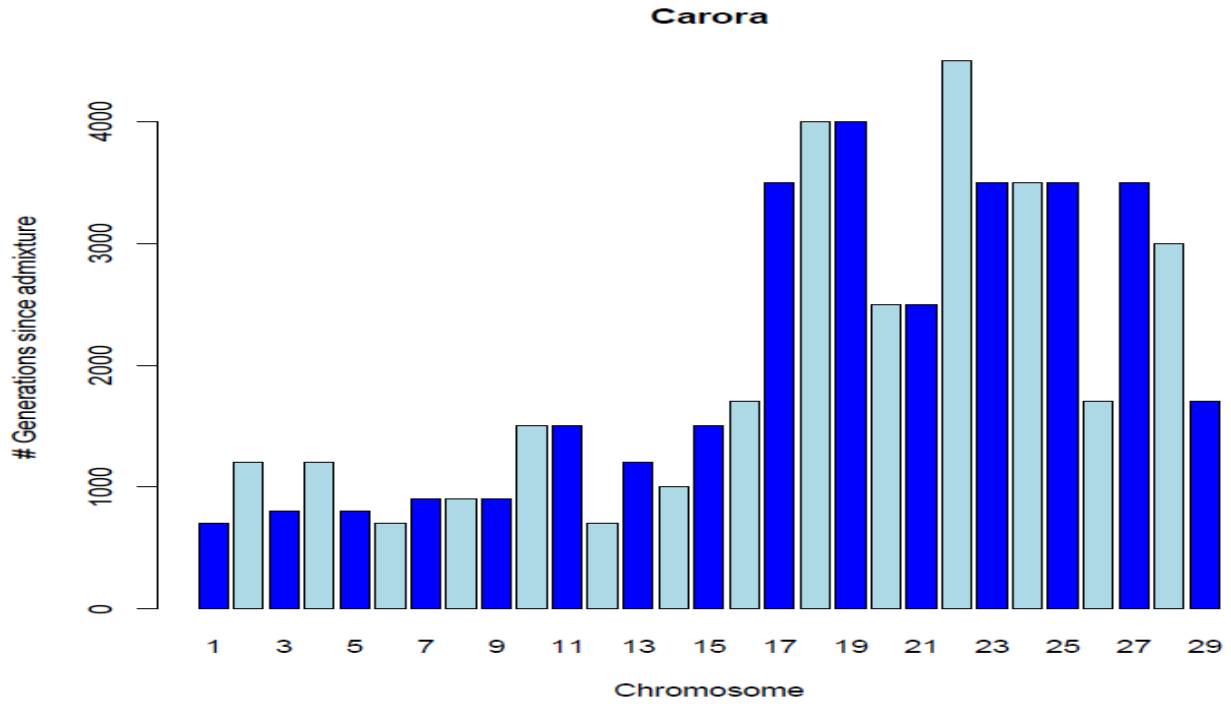


Figure 4.14 Estimates of time since admixture per chromosome for Carora animals. Estimated in HAPMIX

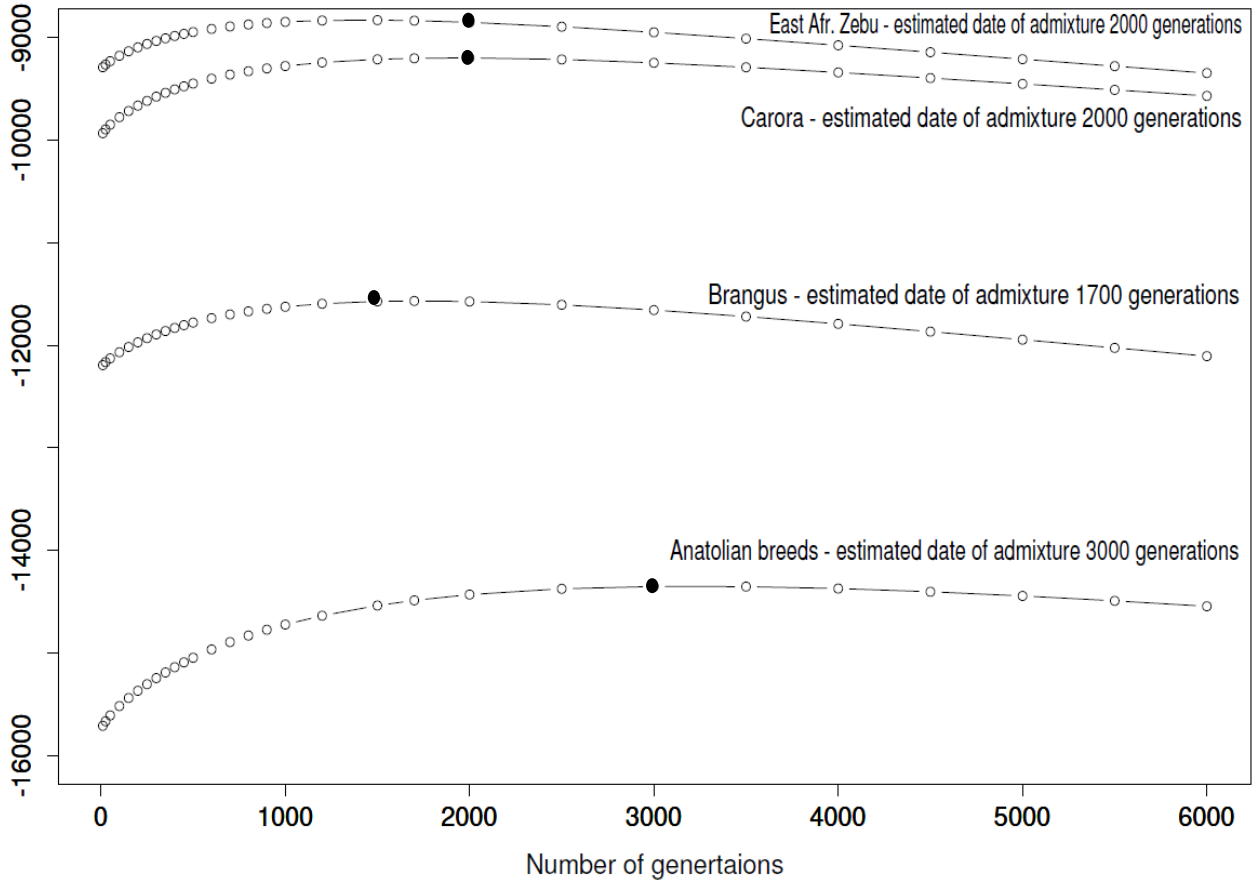


Figure 4.15 A comparison of the log likelihood estimates for the number of generations since admixture for East African Zebu breeds, Anatolian breeds, Brangus and Carora. Credible interval indicated by a solid circle.

4.3.3 Identifying signatures of selection using integrated haplotype score

Whole genome wide iHS estimates (converted to P-values) for the East African Zebu, and Anatolian breeds are shown in figures 4.15 and 4.16 respectively. Analysis of the East African Zebu breeds revealed 77 SNPs with a significance of $P < 1x10^{-7}$ (False discovery rate of 1% applied; Table 4.8). The four most significant SNPs were all located within a few Kb of each other in a peak on chromosome 6. No genes were identified in this region, with the nearest gene (*PRDM5*) being ~200kb from the most significant SNP. Notable genes identified are *NOTCH4*, *PBX2* and *AGER* which are located at the start of the bovine MHC region.

Whole genome estimation of iHS of Anatolian breeds revealed 108 SNPs with significance $P < 1x10^{-6}$ (FDR of 1% applied; Table 4.9). A total of 14 genes were identified from investigation of SNPs above a significance of $P < 1x10^{-6}$ (Table 4.9). The genes identified which were of particular interest were *DCAF16*, *DGAT2*, *PLXNA2* and *HSPB8*.

Quantile-quantile plots for SNP P-values estimated from iHS analysis of East African Zebu and Anatolian breeds both showed a slight inflation of all P-value from a uniform distribution. The most extreme P-values did however fall within the 0.95 pointwise confidence band.

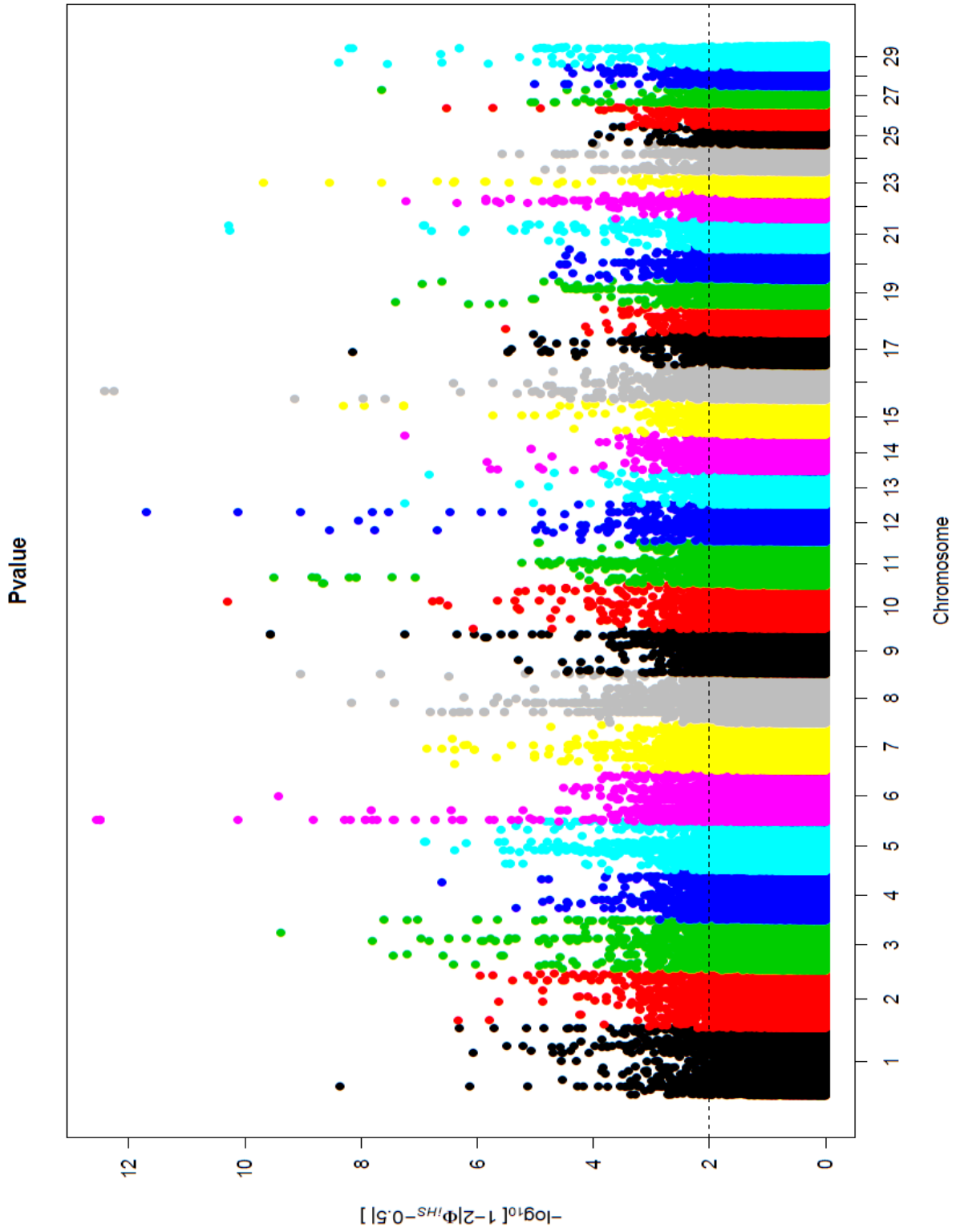


Figure 4.16 Genome wide iHS (converted to P-values) estimated for East African Zebu breeds.

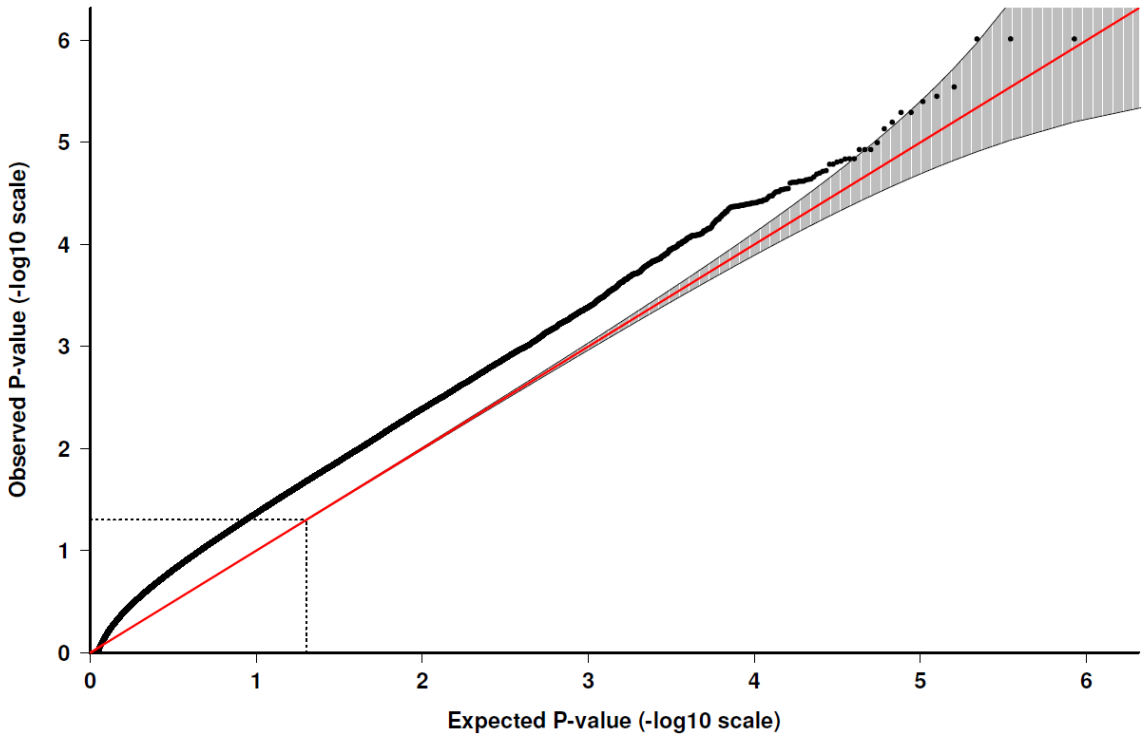


Figure 4.17 Quantile-Quantile plot for the expected vs observed P-value for iHS analysis of East African Zebu breeds. Shaded region indicates a pointwise confidence band of 0.95. All P-values were slightly inflated from what was expected given a uniform distribution (red line).

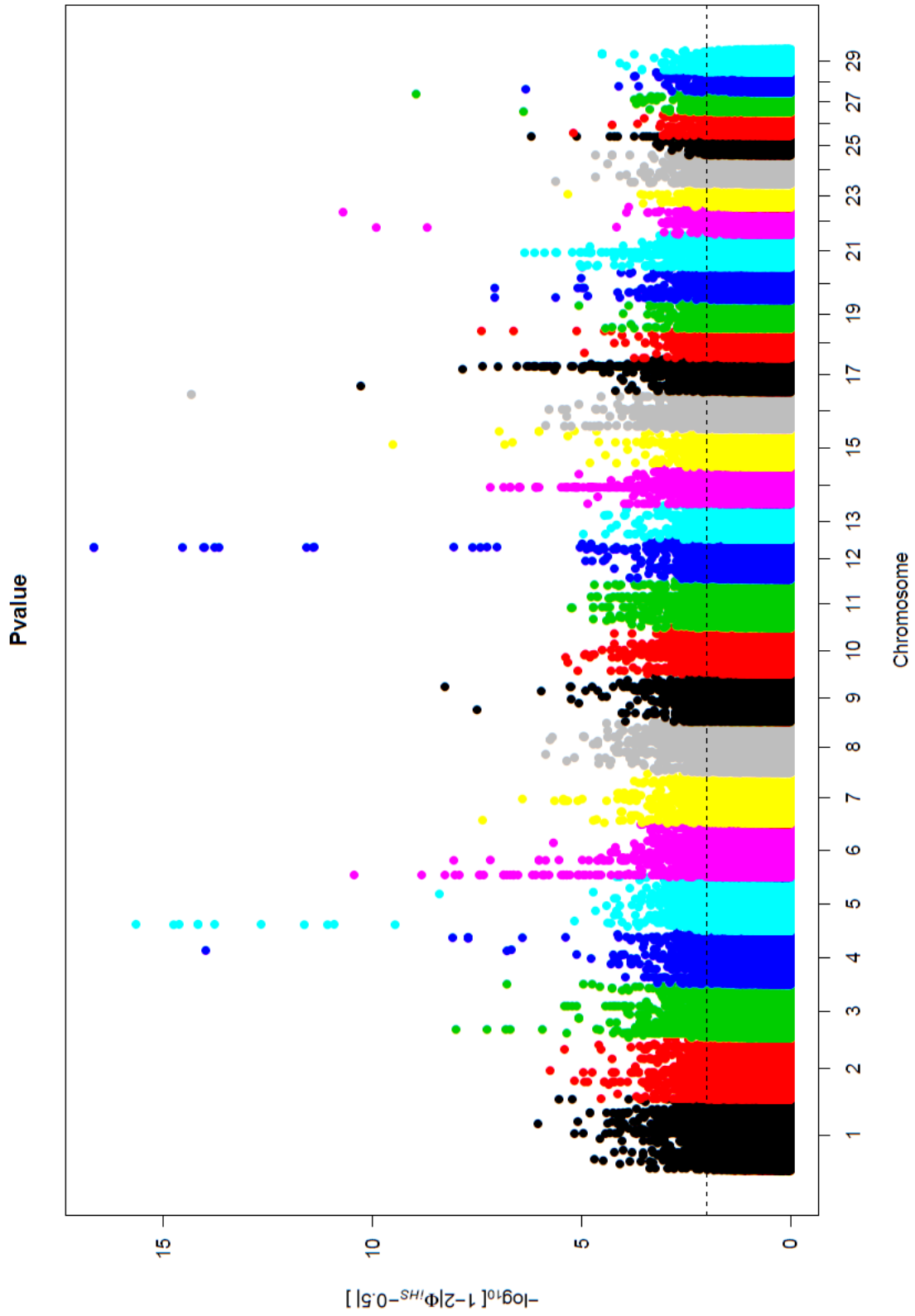


Figure 4.18 Genome wide iHS (converted to P-values) estimated for Anatolian breeds

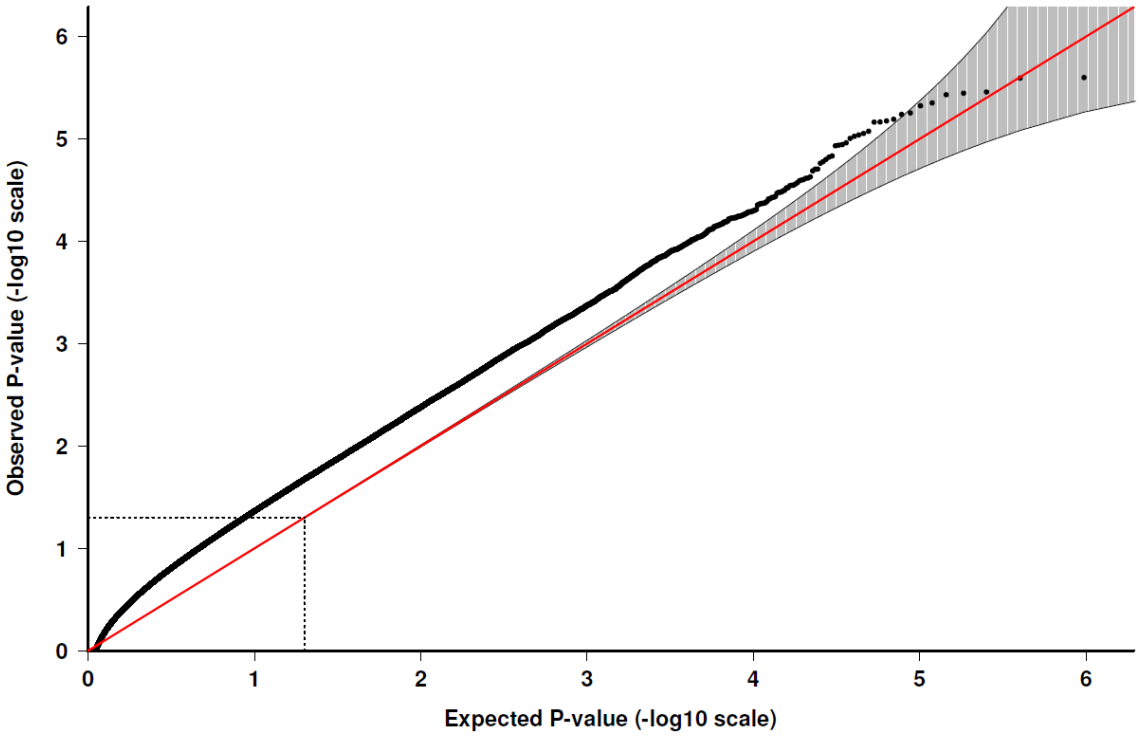


Figure 4.19 Quantile-Quantile plot for the expected vs observed P-value for iHS analysis of Anatolian breeds Shaded region indicates a pointwise confidence band of 0.95. All P-values were slightly inflated (black points) from what was expected given a uniform distribution (red line).

Table 4.8 Genes identified from investigation of SNPs with significant P<1x10-7 (FDR of 1% applied) from iHS analysis of East African Zebu breeds

SNP	CHR	POSITION	iHS	-log$_{10}$(P-value)	Prob. Afr. Taurus	Prob. indicus	Genes
BovineHD0100006443	1	21776519	-5.87	8.35	0.11	0.43	none
BovineHD0300011200	3	35941955	5.41	7.21	0.05	0.69	none
BovineHD0300020290	3	68706675	5.65	7.81	0.25	0.47	none
ARS-BFGL-NGS-21298	3	90669666	6.25	9.38	0.11	0.68	none
BovineHD0300035333	3	120915438	5.41	7.21	0.14	0.53	*SEPT2, STK25,*
BovineHD0300035424	3	121228333	-5.57	7.60	0.06	0.61	*BOK, THAP4,*
BovineHD0300035433	3	121252568	-5.33	7.02	0.29	0.36	*ATG4B, DTYMK, ING5, D2HGDH*
BovineHD0600001303	6	4942854	-5.36	7.07	0.20	0.52	
BovineHD0600001304	6	4944524	-5.51	7.45	0.35	0.26	
BovineHD0600001311	6	4964608	-6.51	10.13	0.35	0.26	
BovineHD0600001320	6	5012389	-5.50	7.42	0.36	0.25	
BovineHD0600001330	6	5055597	-5.80	8.19	0.32	0.31	
BovineHD0600001332	6	5063040	-5.84	8.28	0.25	0.43	
BovineHD0600001333	6	5065951	-7.30	12.55	0.28	0.42	
BovineHD0600001335	6	5071197	-7.29	12.50	0.31	0.39	none
BovineHD0600001336	6	5074619	-7.29	12.50	0.35	0.36	
BovineHD0600001337	6	5076651	-7.29	12.50	0.34	0.36	
BovineHD0600001341	6	5090033	-5.65	7.80	0.34	0.36	
BovineHD0600001344	6	5097695	-6.05	8.83	0.31	0.34	
BovineHD0600034231	6	5108046	-5.70	7.92	0.30	0.34	
BovineHD0600001346	6	5118327	-5.70	7.92	0.29	0.34	
BovineHD0600001350	6	5141190	-5.62	7.72	0.29	0.33	
BovineHD0600007744	6	27991491	5.66	7.83	0.30	0.35	none
BovineHD0600016903	6	61416793	6.26	9.41	0.37	0.23	*APBB2*
BovineHD0800013219	8	44259605	5.50	7.42	0.71	0.08	none
BovineHD0800013229	8	44280462	-5.80	8.17	0.81	0.07	
BovineHD0800032734	8	108924130	6.12	9.04	0.25	0.38	none
BovineHD0800032812	8	109138367	5.60	7.67	0.25	0.34	
BovineHD0900027022	9	95307827	-5.43	7.25	0.13	0.49	
BovineHD4100007681	9	95310420	6.31	9.56	0.16	0.43	none
BovineHD0900027024	9	95310973	6.31	9.56	0.16	0.43	
BovineHD1000019532	10	68179385	6.57	10.29	0.12	0.47	none
BovineHD1100001449	11	3962378	5.98	8.65	0.11	0.48	none
BovineHD1100001450	11	3966538	5.98	8.66	0.16	0.46	
BovineHD1100005271	11	16915845	5.81	8.21	0.20	0.41	
BovineHD1100005274	11	16924102	6.02	8.77	0.24	0.38	
BovineHD1100005277	11	16927003	5.77	8.09	0.24	0.38	
BovineHD1100005278	11	16928946	5.77	8.09	0.27	0.39	none
BovineHD1100005281	11	16934770	5.35	7.06	0.12	0.49	
BovineHD1100005413	11	17596883	6.29	9.50	0.23	0.38	
BovineHD1100005518	11	17901347	5.52	7.47	0.16	0.49	
BovineHD1100005657	11	18497556	6.05	8.84	0.08	0.55	
BovineHD1200007136	12	23716595	5.93	8.53	0.13	0.50	none
BTA-115056-no-rs	12	23751037	5.64	7.76	0.27	0.27	
BovineHD1200013931	12	50727877	5.75	8.05	0.27	0.27	*TBC1D4*
BovineHD1200020972	12	75284908	-5.65	7.80	0.28	0.28	
BovineHD1200020975	12	75287049	7.03	11.70	0.28	0.28	none
BovineHD1200020976	12	75289075	-5.54	7.53	0.25	0.34	

BovineHD1200020983	12	75299277	-6.13	9.05	0.19	0.45	
BovineHD1200020985	12	75300825	-6.50	10.11	0.44	0.38	
BovineHD4100009847	13	1984924	5.43	7.24	0.44	0.38	none
BovineHD1400023127	14	81913109	5.43	7.24	0.44	0.36	none
BovineHD1500020564	15	71316630	5.44	7.26	0.46	0.31	
BovineHD1500020565	15	71317968	5.44	7.26	0.57	0.28	
BovineHD1500020567	15	71319703	5.44	7.26	0.16	0.45	none
BovineHD1500020578	15	71333232	5.71	7.95	0.24	0.40	
BovineHD1500020584	15	71343817	5.85	8.30	0.14	0.54	
BovineHD1600000429	16	1702147	5.56	7.58	0.40	0.26	none
BovineHD1600000438	16	1728962	6.16	9.15	0.30	0.24	
BovineHD1600000669	16	2393030	5.71	7.96	0.10	0.63	none
BovineHD1600006017	16	21537974	7.21	12.25	0.13	0.41	none
BovineHD1600006022	16	21552492	7.26	12.42	0.23	0.43	
BovineHD1700008318	17	29326593	5.79	8.15	0.35	0.41	none
BovineHD4100013893	19	8835302	5.49	7.40	0.33	0.33	none
BovineHD2100012671	21	44285105	6.56	10.26	0.50	0.18	none
BovineHD2100016189	21	56351450	6.57	10.29	0.48	0.19	*FRMD5*
BovineHD2200012496	22	43144534	-5.42	7.23	0.47	0.20	none
BovineHD2300007364	23	26985409	6.36	9.69	0.29	0.43	*NOTCH4, PBX2, AGER*
BovineHD2300007376	23	27007738	5.94	8.55	0.21	0.47	
BovineHD2300007378	23	27010815	5.59	7.64	0.10	0.52	
BovineHD2700010418	27	36385854	5.59	7.65	0.14	0.58	none
BovineHD2900001929	29	6888110	5.55	7.55	0.24	0.45	none
ARS-BFGL-NGS-67211	29	9881743	-5.88	8.38	0.11	0.43	*SYTL2*
ARS-BFGL-NGS-24243	29	44325408	5.81	8.20	0.05	0.69	*SCYL1, LTBP3*
BovineHD4100019100	29	44405678	5.78	8.13	0.25	0.47	

Table 4.9 Genes identified from investigation of SNPs with significant P<1x10-6 (FDR of 1% applied) from iHS analysis of Anatolian breeds

SNP	CHR	POSITION	iHS	-log10 (P-value)	Prob. Taurus	Prob. indicus	Genes
BovineHD0100030444	1	107545399	-4.90	6.02	0.49	0.19	*PPM1L*
BovineHD0300005890	3	18295964	-5.43	7.24	0.49	0.20	
BovineHD0300005892	3	18297924	-5.72	7.98	0.50	0.20	none
BovineHD0300005895	3	18305656	-5.25	6.81	0.49	0.21	
BovineHD0300005904	3	18318836	-5.19	6.69	0.41	0.26	
BovineHD4100002584	3	121330322	-5.24	6.78	0.23	0.44	none
BovineHD0400021843	4	78775928	5.24	6.79	0.42	0.16	
BovineHD0400022147	4	80085378	7.73	13.97	0.46	0.17	*POU6F2*
BovineHD0400022707	4	82291363	5.19	6.68	0.28	0.25	
BovineHD0400030093	4	106282311	5.61	7.70	0.28	0.26	
BovineHD0400030468	4	107123428	5.07	6.39	0.43	0.16	
BovineHD0400030471	4	107135505	5.76	8.07	0.45	0.15	
BovineHD0400030477	4	107145065	5.62	7.71	0.45	0.14	none
BovineHD0400030478	4	107148814	5.62	7.71	0.44	0.15	
BovineHD0400030479	4	107151956	5.62	7.71	0.43	0.15	
BovineHD0400030480	4	107156419	5.62	7.71	0.41	0.15	
BovineHD0500003797	5	12653748	6.27	9.44	0.43	0.30	
BovineHD0500003800	5	12657773	6.83	11.06	0.43	0.29	
BovineHD0500003821	5	12682896	7.01	11.63	0.25	0.48	
Hapmap49501-BTA-15676	5	13007538	7.95	14.75	0.35	0.25	
BovineHD0500004094	5	13554241	7.33	12.65	0.39	0.29	
BovineHD0500004097	5	13562373	7.91	14.61	0.35	0.36	none
BovineHD0500004100	5	13573024	8.28	15.65	0.33	0.40	
BovineHD0500004111	5	13615733	6.77	10.90	0.35	0.25	
Hapmap59589-rs29018532	5	13632142	7.67	13.76	0.42	0.16	
BovineHD0500004123	5	13661629	7.79	14.16	0.71	0.03	
BovineHD0500004124	5	13663137	7.79	14.16	0.70	0.03	
BovineHD0500024150	5	85383176	5.88	8.38	0.54	0.15	none
BovineHD0600001311	6	4964608	-5.23	6.76	0.35	0.27	
BovineHD0600001344	6	5097695	-5.74	8.03	0.33	0.28	
BovineHD0600001346	6	5118327	-5.83	8.25	0.31	0.26	
BovineHD0600001356	6	5174278	-5.12	6.51	0.32	0.23	
BovineHD0600001360	6	5193923	-4.93	6.08	0.35	0.22	
BovineHD0600001361	6	5197676	-5.46	7.32	0.35	0.22	
BovineHD0600001362	6	5201867	-5.26	6.85	0.36	0.21	none
Hapmap35597-SCAFFOLD359101_527	6	5205468	-6.04	8.81	0.37	0.20	
BovineHD0600001364	6	5213570	-5.51	7.44	0.36	0.20	
BovineHD0600001366	6	5221411	-5.70	7.92	0.33	0.22	
BovineHD0600001369	6	5225746	-4.97	6.17	0.36	0.22	
BovineHD0600001371	6	5228902	-6.62	10.44	0.36	0.23	
BovineHD0600001374	6	5234762	-5.17	6.62	0.36	0.22	
BovineHD0600001375	6	5237576	-5.17	6.62	0.35	0.21	
BovineHD0600001442	6	6717220	-5.49	7.41	0.50	0.24	
BovineHD0600001443	6	6724624	-5.26	6.85	0.46	0.24	none
BovineHD0600001444	6	6736226	-5.20	6.69	0.42	0.24	
ARS-BFGL-NGS-112812	6	38627070	5.74	8.04	0.19	0.32	*DCAF16,*
BovineHD4100004597	6	39220461	-5.39	7.16	0.46	0.22	*NCAPG, LCORL*
BovineHD0700001984	7	7825464	-5.48	7.37	0.50	0.28	*AP1M1*
BovineHD0700016186	7	56096544	-5.07	6.40	0.24	0.38	*ARHGAP26*
BovineHD0900007692	9	28388947	5.53	7.50	0.59	0.10	none
BovineHD0900022756	9	81753987	-5.83	8.26	0.40	0.26	*AIG1*

BovineHD1200019890	12	72271847	5.57	7.60	0.48	0.16	
BovineHD1200019906	12	72301304	7.89	14.54	0.48	0.15	
BovineHD1200019907	12	72307576	6.94	11.40	0.70	0.10	
BovineHD1200019929	12	72329951	8.51	16.65	0.66	0.11	
BovineHD1200019930	12	72331910	9.60	16.65	0.68	0.11	
BTA-95991-no-rs	12	72341102	6.99	11.56	0.67	0.11	
BovineHD1200028249	12	72349503	7.64	13.67	0.64	0.11	
BovineHD1200019946	12	72366676	7.67	13.76	0.62	0.11	
BovineHD1200019947	12	72369351	5.50	7.42	0.63	0.10	none
BovineHD1200019949	12	72374093	7.74	14.01	0.63	0.10	
BovineHD1200019950	12	72375072	7.75	14.04	0.62	0.10	
BovineHD1200020968	12	75280650	5.43	7.26	0.62	0.10	
BovineHD1200020969	12	75282655	-5.33	7.00	0.21	0.36	
BovineHD1200020972	12	75284908	-6.93	11.39	0.20	0.37	
BovineHD1200020975	12	75287049	5.75	8.06	0.19	0.38	
BovineHD1400010606	14	36932647	-4.93	6.09	0.46	0.15	
BovineHD1400010609	14	36955735	-4.89	6.00	0.54	0.06	
BovineHD1400010616	14	36989311	-4.92	6.06	0.49	0.13	
BovineHD1400024091	14	36998787	-4.94	6.10	0.49	0.10	
BovineHD1400010621	14	37007897	-5.39	7.16	0.45	0.11	
BovineHD1400010624	14	37020549	-5.27	6.86	0.40	0.16	none
BovineHD1400010644	14	37096559	-5.11	6.50	0.44	0.15	
BovineHD1400010645	14	37099268	-5.11	6.50	0.45	0.15	
BovineHD1400010648	14	37112656	-5.19	6.68	0.52	0.14	
BovineHD1400010649	14	37115932	-5.19	6.68	0.52	0.13	
BovineHD1400010650	14	37117181	-5.10	6.46	0.52	0.13	
BovineHD1500014914	15	51780028	-5.26	6.83	0.26	0.41	none
BovineHD1500014915	15	51780484	-6.29	9.50	0.20	0.56	
BovineHD1500016213	15	55960725	5.18	6.66	0.25	0.51	*DGAT2*
BovineHD1500023560	15	80978010	5.31	6.95	0.45	0.16	
BovineHD1500023564	15	80989716	4.90	6.01	0.46	0.16	none
BovineHD1500023581	15	81039900	4.89	6.01	0.39	0.20	
BovineHD1600022338	16	77196022	7.83	14.31	0.26	0.29	*PLXNA2*
ARS-BFGL-NGS-5282	17	12319235	6.56	10.26	0.46	0.20	none
BovineHD1700014515	17	51683589	5.67	7.85	0.54	0.12	none
BovineHD1700016422	17	58012578	-5.32	7.00	0.50	0.10	
BovineHD1700016485	17	58260294	-5.48	7.36	0.45	0.18	
BovineHD1700016486	17	58261005	-5.00	6.25	0.45	0.19	
BovineHD1700016495	17	58273760	-4.95	6.13	0.41	0.20	*HSPB8*
BovineHD1700016543	17	58388161	-5.13	6.53	0.51	0.20	
BovineHD1700016544	17	58389183	-5.02	6.29	0.50	0.20	
BovineHD1700016549	17	58394868	-4.94	6.11	0.49	0.22	
BovineHD1700016562	17	58419775	-4.93	6.09	0.64	0.19	
BovineHD1800017249	18	59379519	5.17	6.62	0.24	0.27	none
BovineHD1800017353	18	60043680	5.49	7.38	0.24	0.26	
BTA-50642-no-rs	20	3794030	-5.35	7.06	0.39	0.13	none
BovineHD2000007310	20	24282146	5.35	7.06	0.20	0.35	none
BovineHD2100009210	21	31935003	-5.05	6.35	0.51	0.11	none
BovineHD2100009211	21	31941072	-4.94	6.11	0.51	0.11	
BovineHD2200004623	22	15990347	6.43	9.89	0.24	0.29	*ANO10, TOPAZ1*
BovineHD2200018043	22	16322562	5.99	8.69	0.22	0.33	
BovineHD2200014486	22	50745402	6.70	10.69	0.31	0.20	none
BovineHD2500011394	25	40475227	4.98	6.19	0.78	0.05	none
BovineHD2700011840	27	40778889	6.09	8.95	0.27	0.25	none
BovineHD2800001776	28	5987323	5.03	6.31	0.36	0.21	none

4.3.4 Meta-analysis

Product ranks of the SNP estimated from combining the iHS and HAPMIX analysis for the East African zebu and Anatolian breeds are shown in figures 4.21 and 4.22, respectively. Of the top 100 ranked SNPs from combined HAPMIX/iHS estimates from analysis of East African Zebu breeds, 26 were in the top 1% of ranked SNPs in both HAPMIX and iHS analyses. A total of 16 genes were identified from investigation of the top 100 ranked SNPs from combined HAPMIX/iHS analysis (table 4.10) of East African Zebu breeds.

Within the top 100 combined ranked SNPs from HAPMIX and iHS analysis for Anatolian breeds, 30 SNPs were ranked in the top 1% of SNPs from HAPMIX and iHS analysis. A total of 11 genes were identified from investigation of the top 100 combined ranked SNPs (table 4.11). Of these genes two were found to have involvement in economically important traits or immune response, *IGSF11* and *PRKCA*.

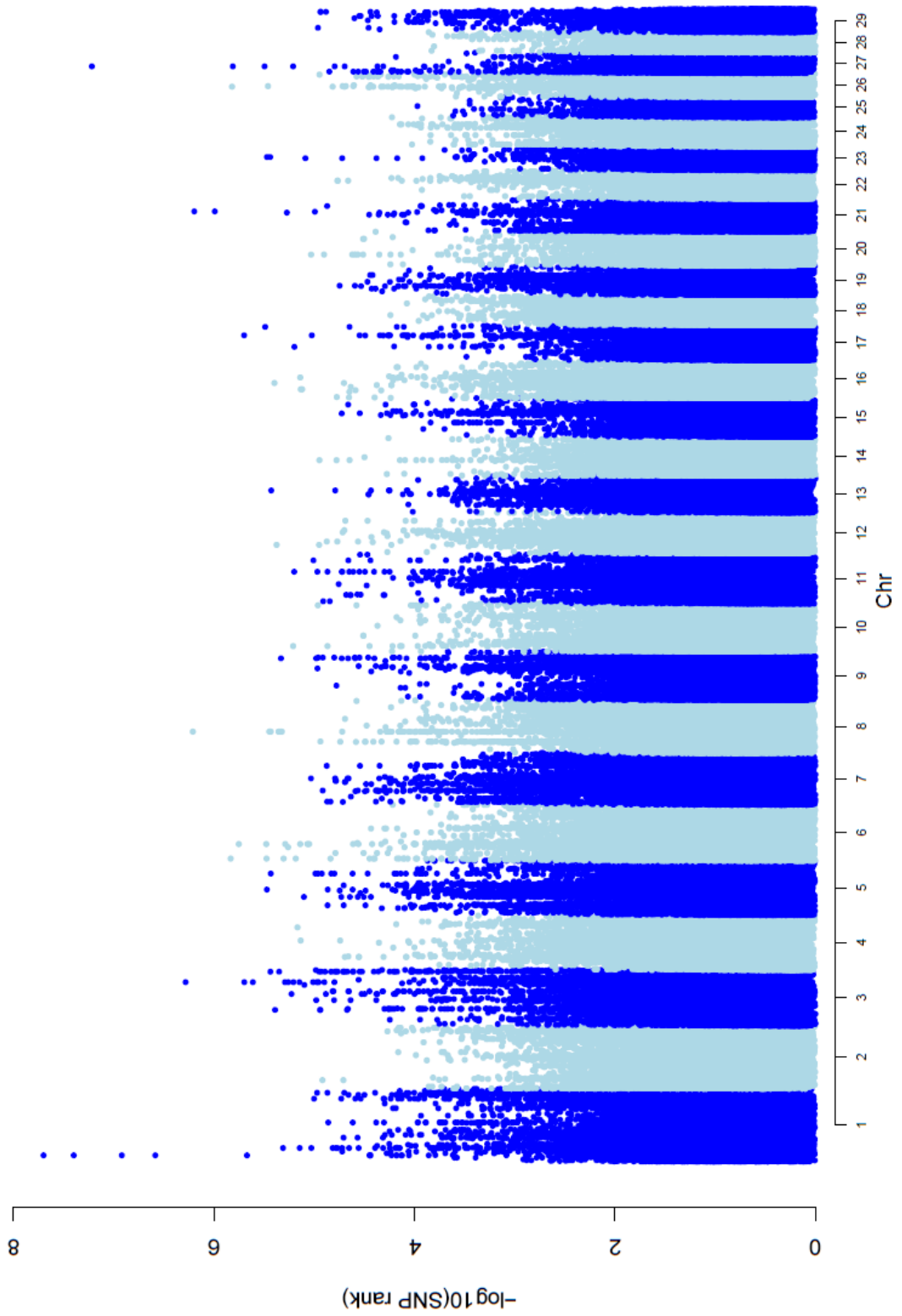


Figure 4.24 Genome wide SNP rankings estimated from combining the rank of SNP from HAPMIX and iHS analysis of East African Zebu breeds

Table 4.10 Genes identified from investigation of the top 100 combined rankings of SNPs from HAPMIX and iHS analysis of East African Zebu breeds

Chr	position	SNP	HAPMIX Rank %	iHS rank %	Genes
1	28927255	BovineHD0100046933*	0.2014%	0.2347%	
1	28957300	BovineHD0100008599	0.0250%	3.6142%	GBE1
1	28962463	BovineHD0100008601	0.0320%	2.1653%	
5	41177386	BovineHD0500011752*	0.2948%	0.2589%	SLC2A13
6	37533544	BovineHD4100004493	0.0086%	5.7567%	
6	37539187	BovineHD0600010412	0.0077%	2.2126%	
6	37541623	BovineHD0600034276	0.0071%	4.5109%	HERC3
6	37547900	BovineHD0600010414	0.0074%	9.4403%	
6	37569913	BovineHD0600010417	0.0411%	2.0912%	
6	37573850	BovineHD0600010418	0.0395%	2.0914%	
8	23982891	BovineHD0800007193*	0.2129%	0.5232%	MLLT3
9	93702202	BovineHD0900026520	0.0036%	27.834%	
9	93704329	BovineHD0900026522	0.0045%	9.8997%	MIR2481
9	95310420	BovineHD4100007681	49.9503%	0.0021%	
12	16882032	BovineHD1200005129*	0.3892%	0.2777%	HTR24
12	16884063	BovineHD1200005131*	0.4323%	0.0942%	
13	47836310	BovineHD1300013973	1.0288%	0.0350%	CDS2
16	34429942	BovineHD1600009855*	0.7544%	0.0514%	SDCCAG8
16	47876872	BovineHD1600013223	0.0429%	1.6353%	ACOT7
17	55290523	BovineHD1700015686	0.0829%	1.1018%	CLIP1
17	55293108	BovineHD1700015689*	0.0988%	0.1956%	
17	71788904	Hapmap47287-BTA-41904*	0.7774%	0.0403%	OSBP2
21	42483941	BovineHD2100012115*	0.7594%	0.0679%	NUBPL
23	26985409	BovineHD2300007364	40.0603%	0.0020%	NOTCH4
26	26076199	BovineHD2600006875*	0.2014%	0.2347%	SORCS3
29	9881743	ARS-BFGL-NGS-67211	0.0250%	3.6142%	SYTL2

*denotes that SNPs where ranked in the top 1% in both iHS and Hapmix analyses.

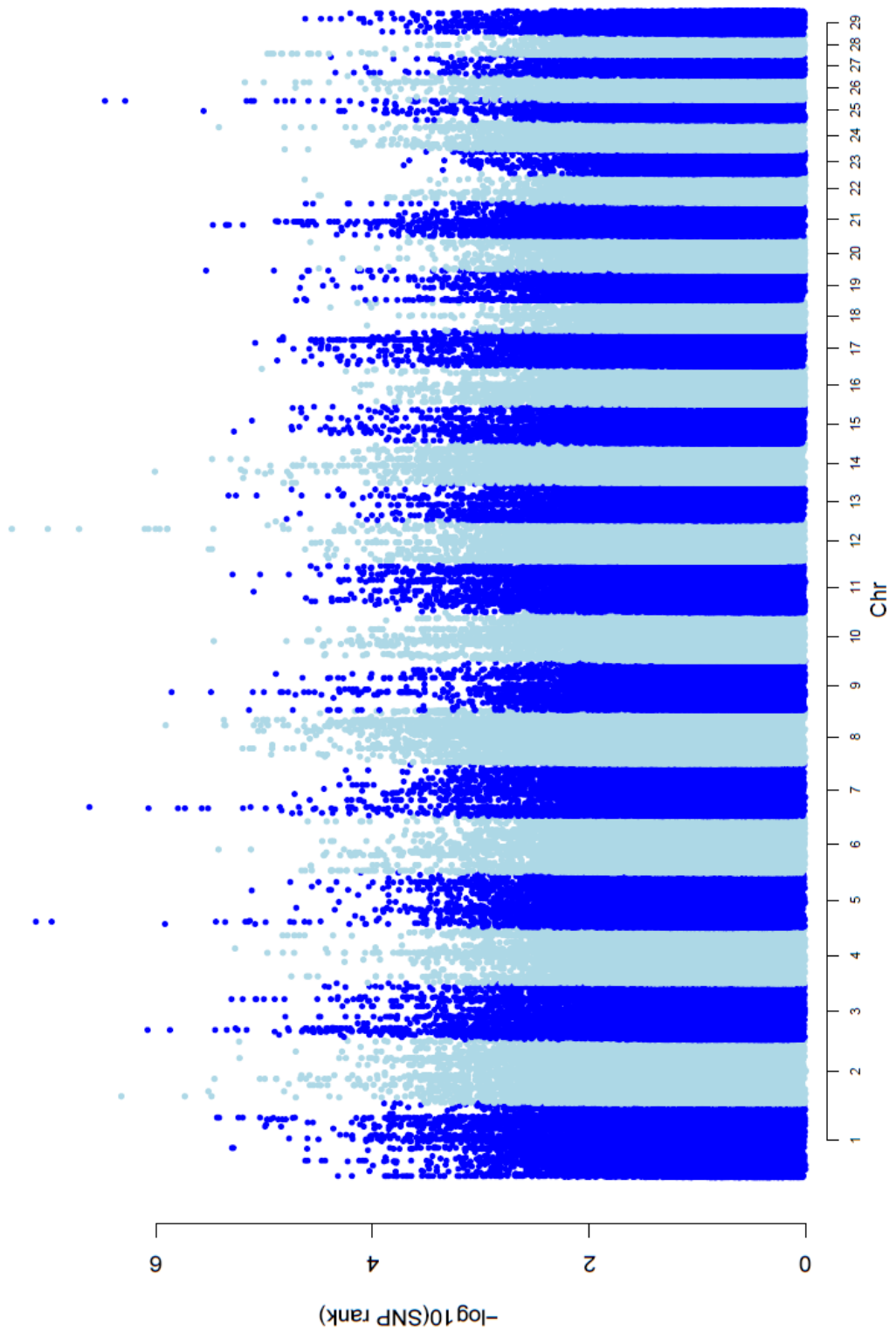


Figure 4.25 Genome wide SNP rankings estimated from combining the rank of SNP from HAPMIX and iHS analysis of Anatolian breeds

Table 4.11 Genes identified from investigation of the top 100 combined rankings of SNPs from HAPMIX and iHS analysis of Anatolian breeds

Chr	position	SNP	HAPMIX rank %	iHS rank %	Genes
1	64479049	BovineHD0100018242*	0.1903%	0.3027%	IGSF11
1	64480510	BovineHD0100018244*	0.1845%	0.3029%	
1	131779198	BovineHD0100037492*	0.9373%	0.0685%	MRAS
1	131780727	BovineHD0100037494	1.0085%	0.0403%	
1	131781670	BovineHD0100037495	1.0521%	0.0405%	
1	131786763	BovineHD0100037500	1.2170%	0.0577%	
2	133617891	BovineHD0200038960	0.0383%	1.6964%	HTR6
3	19810149	BovineHD0300006300	0.0150%	3.3335%	FAM63A, ANXA9
3	19811302	BovineHD0300006302	0.0170%	4.4938%	
3	19816107	BovineHD0300006303*	0.0106%	0.8750%	
3	19816641	BovineHD0300006304	0.0097%	1.5403%	
3	19819542	BovineHD0300006306	0.0111%	3.5308%	
3	19823920	BovineHD0300006309	0.0262%	2.4017%	
3	23656626	BovineHD0300007401*	0.4741%	0.1271%	HMGCS2
5	13007538	Hapmap49501-BTA-15676	79.7965%	0.0008%	SYT1
7	19736297	BovineHD0700005495	0.0017%	20.1661%	FUT6
7	19737393	BovineHD0700005496	0.0015%	19.3642%	
7	19738587	BovineHD0700005497	0.0011%	19.3644%	
7	19739162	BovineHD0700005498	0.0008%	12.5067%	
7	19740182	BovineHD0700005499	0.0006%	29.2120%	
13	54929745	BovineHD1300015594*	0.9600%	0.0983%	SLC17A9, GID8
13	54934270	BovineHD1300015596*	0.8477%	0.0612%	
13	54948937	BovineHD1300015600	1.0426%	0.0583%	
19	63555529	BovineHD1900018398	0.0295%	1.0902%	PRKCA

*denotes that SNPs where ranked in the top 1% in both iHS and Hapmix analyses.

4.4 Discussion

4.4.1 Population structure of Zebu and Anatolian breeds

Population strucuture analysis of the African zebu and Anatolian breeds all corroborated previous work indicating *Bos indicus-Bos taurus* admixture in all Anatolian breeds and African taurus and indicus admixture in East African zebu breeds. These admixture pattern are the most clear from ADMIXTURE analysis with a value of K=3 ancestral populations (Figure 4.4), but is also apparent from principle component analysis (figure 4.1) and multi-dimensional scaling of IBS results (figure 4.3).

While multi-dimesnional scaling of the mean number of IBD blocks (figure 4.2) between breeds was able to distinguish the major divergences between breeds (i.e. between *Bos indicus*, European taurus, and African taurus), it could not distinguish the finer variations between breeds, with most breeds clustering together. Closer investigation into these results reveals that this clustering may be due to an overall low number of IBD block sharing between the majority of breeds. Despite this, analysis of the East African Zebu and Anatolian breeds revealed some interesting relationships. East African Zebu breeds showed a medium level of IBD (compared to within breed IBD) sharing between each other, with the minimum sharing between two East African zebu breeds being an average of 4.24 IBD blocks (Karamoja vs Sheko). Furthermore some East African Zebu breeds also showed a medium level of IBD (compared to within breed IBD) sharing with West African (Azawak, Gobra) breeds.

The Anatolian breeds showed a lower level of IBD sharing between each other than the East African Zebu breeds. However the Anatolian breeds showed a medium level of IBD (compared to within breed IBD) sharing between a number of southern European

taurus breeds (Sikia, Alentejana), African taurus (Kuri) and an indicus breed from the middle east (Sahiwal). The level of IBD sharing between Anatolian breeds and some southern European breeds (Sikia, Alentejana) along with the similar patterns of ADMIXTURE (at *K*=3), as well as their clustering in both PCA and IBD analysis suggests a similar genetic structure in these animals.

Admixture and principle component analysis shown in this chapter corroborates a number of other studies of cattle population structure [160, 185], using a greater number of samples (n=3,325 Vs. n=1,543) and a higher density of markers (n=786,799 Vs. n=43,043).

4.4.2 Admixture outlier identification and dating

The ability to identify regions of retarded or enhanced introgression of zebu genomic origin has been reported previously [129]. The present study is the first use of high density genotypes to map ancestral segments in East African Zebu and Near Eastern Anatolian breeds. Ancestry mapping of East African Zebu breeds revealed three genes involved with economic traits and immune response, *HERC3, AOAH* and *TLR3*. *HERC3* has been identfied from a study on signatures of selection in a population of ~2000 Holstein cattle from China [201], however Pan et al. could not elaborate on the possible function of this gene, however as *HERC3* ubiquitin ligase family, it could have many roles in the regulation of cellular processes. *TLR3,* like other members of toll-like receptor genes, has a major involvement in immunity [202], particularly in recognizing viral components. *AOAH* is a lysosomal enzyme found in neutrophils and macrophages which

has known involvement in inflammation and immunity [203]. While five genes were identified from ancestry mapping of Anatolian breeds (*KYNU, FUT6, CDH10, KCNIP1* and *LHFPL3*) no discernable involvement in important economical or health traits are noted for these genes.

Estimates of aboslute dates since admixture using HAPMIX gave results that reflected the correct order of admixture age, with the oldest being Anatolian breeds and the youngest being Brangus, with Carora and East African Zebu falling in between. However this approach greatly over-estimated the absolute time depths of the admixture events. The Brangus breed was incepted in ~1912, with the formation of a breeding society in 1949 [204, 205]. Assuming a generation interval of 4-6 years for cattle, this should give an estimate of 17-26 generations for the time depth of admixture for Brangus; that estimated by HAPMIX was 1,700 generations. The use of an estimated recombination map in this chapter (where recombination was considered to be uniform with 1Mb representing 1 centimorgan), may have lead to the overestimation of time since admixture. Given that the most accurate and widely used methods for dating absolute time since admixture require highly accurate recombination maps [179, 180, 198], it can be argued that without such a recombination map for cattle, deterimining absolute time since admixture will prove to be very difficult. Other studies have tried to circumvent this problem [176], by developing methods which use the distribution of haplotype block size, however only rough estimates calibrated using breeds with known admixture times can be obtained using these methods.

4.4.3 Signatures of selection estimated from iHS

To the authors knowledge this work represents the first investigation of signatures of selection in populations of East African Zebu breeds and Near Eastern Anatolian breeds using high density genotypes. iHS analysis of East African Zebu breeds identified three genes involved with immunity which may be under positive selection, *NOTCH4*, *PBX2* and *AGER*. These three genes lie at the beginning of the bovine MHC region (*i.e. BoLA*)[206]. Furthermore *NOTCH4* is a member of the *NOTCH* family, *NOTCH1* and *NOTCH3* are known to interact with bovine herpevirus 1 [207, 208]. As East African Zebu breed are known to have a greater resistance to infection from diseases than Eurpoean taurus breeds [105], signatures of selection in this region suggests that the *BoLA* region could be responsible for this difference in breed resistances. Bahbahani et al. [166] performed iHS analysis on a population (n=425) of East African shorthorn zebu using 46,171 SNPs and identified a number of signatures of selection in East African shorthorn zebu in gene regions associated with innate immune response suggesting that immunity genes are hot spots of natural selection in East African shorthorn zebu in response to the high pathogen challenge in their local environment. While the signatures of selection in East African zebu presented in this chapter are not identical to those of Bahbahani et al. [166], the identification of positive selective signatures at the start of the *BoLA* regions is in line with their identification of a signal over the *LOC512672* gene, which is located within the *BoLA* region and is involved in presenting antigen peptides to cytotoxic T-cells.

Several genes to have known involvement with economically important traits were identified by iHS analysis of Near Eastern Anatolian breeds; these were *DCAF16*, *DGAT2*, *PLXNA2* and *HSPB8*. *DCAF16* has been documented to be a significant signal of selection

in Spanish beef breeds (using *Fst* analysis) [209] and in Murnau-Werdenfelser and Franken Gelbvieh breeds (using EHH analysis) [210], suggesting a possible involvement in beef traits. *DGAT2* located on the same chromosomal segment as well as being a member of the same family as *DGAT1*, a major gene involved in milk traits in cattle [152, 211]. Furthermore *DGAT2* is known to have a major involvement in milk and growth traits in cattle, established from studies on different polymorphisms in the *DGAT2* region and their effects on production [212-214]. *PLXNA2* has been suggested to be associated with temperment in Holstein-Charolais crosbreeds [215] while *HSPB8* has been suggested to be associated with beef traits from a genome wide scan for meat quality traits in Nelore cattle [216]. The Anatolian Breeds (Anatolian Black, Turkish Grey, East Anatolian Red and South Anatolian Red) are used as dual purpose breeds throughout Turkey and the Near East, which perhaps explains the signatures of selection presented here.

4.4.4 Meta-analysis of whole genome admixutre and signatures of selection

Product ranks of the significance of SNP estimated from the iHS and HAPMIX analysis for the East African zebu indentified 12 genes that were not identified by either iHS or HAPMIX analysis alone. One of these genes *CDS2* (CDP–diacylglycerol synthase) has been documented to have involvement in milk traits (via phospholipid biosynthesis) and was identified to be under selective pressure in Italian dairy breeds [217]. None of the other genes identified from combined product ranks of the significance of SNP estimated from the iHS and HAPMIX analysis of East African Zebu breeds had any discernable function in desirable cattle traits. Meta-analysis of Anatolian breeds identified ten genes which had not been identified from either iHS or HAPMIX analysis. Two of these genes,

IGSF11, and *PRKCA*. *IGSF11* was identified to have an association with milk protein traits from a runs of homozygosity analysis in US Holstein [218]. While *PRKCA* has been reported to have reduced expression in response to *Mycobacterium bovis* infection [219].

4.5 Conclusion

This chapter represents the first investigation of signatures of selection in East African Zebu and Anatolian breeds using high density genotypes. Both ancestry mapping and iHS analysis of these breed groups identified possible signatures of selection for economically desirable traits. A number of signals of selection were identified from iHS pertaining to economically important traits for dairy and beef production in Anatolians breeds, which has not been reported previously. Furthermore iHS analysis of East African Zebu breeds indicated selective pressure in the *BoLA* region (the bovine major histocompatibility complex region). Mapping of loci with outlying proportions from zebu and taurus ancestral haplotypes identified some genes of importance to animal health and production, however results were not as promising as those from iHS analysis.

4.6 Acknowledgments

The following people contributed to this work by providing the genotypes used for the analysis; Heather Huson, Tad Sonstegard, Laercio Porto, Joao Panetto, Johann Soelkner, Cedric Gondro, Ehud Lipkin, Fernando Garcia, Macros Vinicius and Yuri Tani, who collected and processed the genotypes for use in this chapter.

General Conclusion

Infection of cattle with *Mycobacterium bovis* not only constitutes a major economic burden on the Irish government, but leads to a major financial and in many ways emotional loss for the owners of infected farms. While incredible effort to eradicate the disease from the island has been undertaken for over half a century, the advent of high power computing and high density, low cost genotyping has allowed for a much more accurate and detailed insight into the genetics underlying this disease. The use of strategic national breeding programs, based on genetic evaluations of sires for production traits, has led to incredible increases in output from livestock throughout the world. As production output of livestock increases the next bottleneck for production will be livestock health. The estimation of variance components and heritability presented in this thesis, provides the basis for the development of a future breeding strategy towards improved resistance to *Mycobacterium bovis* infection in Irish cattle. The correlation between increased estimated breeding values and the prevalence of infection in sire daughters and the relatively high heritability (for a disease trait), presented in this thesis indicates that not only could developing a breeding strategy around genetic evaluations be possible but that the increase in genetics gain could be observable within two to three decades. Due to this the Irish Cattle Breeding Federation and Teagasc have begun work in creating a new health sub-index of the national breeding strategy, which will incorporate genetic evaluations for *Mycobacterium bovis* infection susceptibility, bovine viral diarrhea infection and cystic ovaries. Selection using this new sub-index could lead to an overall healthier and hardier Irish herd.

Genetic susceptibility to *Mycobacterium tuberculosis* in humans has been generally considered an elusive polygenic trait, with a greater power of analysis needed to identify

QTL than most disease traits; *Mycobacterium bovis* is no different. To date no study has reached a very high power of analysis to identify QTL for *Mycobacterium bovis* infection in cattle, either due to insufficient number of markers or samples. Meta-analysis has been performed to combine GWAS for *M bovis* infection to increase the power of analysis to great effect. The work in this thesis tried to circumvent the issue of polygenicity of *M. bovis* susceptibility by using several methods for GWAS, one QTL was identified across all three analyses, however it was not the most significant signal in either one alone. An increase in the number of samples to over 1000's genotypes could lead to the identification of more believable signals of association. Furthermore the use of whole genome sequences would increase the power of analysis greatly. As imputation accuracy from high density (770,000 SNPs) to whole genome sequence in production breeds has been documented to be over 90% and with the creation of the 1,000 bull project the use of whole genome sequences for GWAS analysis in cattle is now a reality. Work has already begun to identify QTL associated with a wealth of production and health traits in Irish cattle, including *M. bovis* susceptibility. Furthermore the development of the low density "Irish Dairy Beef" chip means that genomic prediction for all economically important traits (including *M. bovis* susceptibility) could be a possibility, increasing the rate of annual genetic gain in Irish cattle,thus increasing their production and health. Finally recent work by Wu et al., has identified the potential of gene editing techniques in increasing resistance of bTB incattle through the creation of transgenic cattle with the added *SP110* mouse gene which is known the increase resistance to tuberculosis in mice [132, 220].

Mapping ancestral haplotypes and identifying signatures of selection in cattle breeds is a promising alternative to traditional GWAS, as it can yield a greater power of

analysis and provide an understanding of the underlying mechanisms of selection. Signatures of selection have identified a wealth of genes associated with production and innate immunity in cattle. *Bos indicus* cattle breeds have for some time been known to be hardier and more resistant to disease than European *Bos taurus* breeds such as Holstein-Friesians. Identifying the genetics underlying this increased resistance could aid in future work in developing pharmaceuticals which target the mechanisms governed by these genes. While accurate absolute dating of admixture events in cattle will not possible until the availability of an accurate recombination map, a wealth of methods are available to do so and have yielded important insights into human population history. Future work aimed at dating absolute time since admixture could give new insight into the population histories of cattle populations.

The work presented in this thesis represents an important move towards the eradication of bovine tuberculosis from Ireland. A combination of the continued national eradication program, an extensive badger culling program, the development of a *M.bovis* vaccine for badger and a national breeding strategy towards increased resistance to infection of *M. bovis* infection in cattle will hopefully see bovine tuberculosis eradicated from the national herd in the coming decades.

Supplementary Results

Supplementary table 1 All Single nucleotide polymorphisms estimated to be significant ($P < 1x10^{-6}$) from single SNP regression approach

SNP	CHR	Genetic Position	P-Value	MAF
BovineHD1400020824	14	74190131	$2.41x10^{-8}$	0.06
BovineHD0100019801	1	69548446	$9.44x10^{-8}$	0.33
ARS-BFGL-BAC-1936	14	74189329	$3.26x10^{-7}$	0.04
BovineHD0600017374	6	62873573	$3.75x10^{-7}$	0.02
BovineHD0100029059	1	101899126	$4.46x10^{-7}$	0.25
BovineHD0100029061	1	101903117	$4.46x10^{-7}$	0.25
BovineHD1900010742	19	37028145	$5.14x10^{-7}$	0.06
BovineHD2600014594	26	50413589	$6.10x10^{-7}$	0.14
BovineHD0100029060	1	101899991	$6.37x10^{-7}$	0.25
BovineHD2600014592	26	50407301	$1.15x10^{-6}$	0.16
BovineHD1000002509	10	7710447	$1.22x10^{-6}$	0.27
BovineHD0100029053	1	101890992	$1.25x10^{-6}$	0.33
BovineHD2600014591	26	50406685	$1.44x10^{-6}$	0.16
BTB-01874657	8	20165909	$1.80x10^{-6}$	0.08
BovineHD1800003695	18	10365503	$1.94x10^{-6}$	0.06
BovineHD1900010746	19	37038060	$2.17x10^{-6}$	0.06
BovineHD1900010749	19	37043601	$2.17x10^{-6}$	0.06
BovineHD1000002985	10	9147391	$2.23x10^{-6}$	0.24
BovineHD0800006206	8	20172157	$2.39x10^{-6}$	0.08
BovineHD1000002982	10	9133666	$2.43x10^{-6}$	0.24
Hapmap53924-rs29022499	6	104939658	$2.50x10^{-6}$	0.32
BovineHD0600023346	6	84880614	$2.76x10^{-6}$	0.24
BovineHD1000030516	10	9143665	$3.14x10^{-6}$	0.25
BovineHD2300001590	23	6530681	$3.30x10^{-6}$	0.18
BovineHD1900017036	19	59645985	$3.45x10^{-6}$	0.07
BovineHD1900017037	19	59646633	$3.45x10^{-6}$	0.07
BovineHD0100029062	1	101910522	$3.51x10^{-6}$	0.29
Hapmap43413-BTA-95698	1	101918115	$3.51x10^{-6}$	0.29
BovineHD2200007065	22	23830884	$3.52x10^{-6}$	0.07

BovineHD0100029132	1	102058031	$3.92x10^{-6}$	0.25
BovineHD0100029134	1	102063784	$3.92x10^{-6}$	0.25
BovineHD0100029136	1	102066870	$3.92x10^{-6}$	0.25
BovineHD0100029137	1	102073696	$3.92x10^{-6}$	0.25
BovineHD0100029143	1	102095932	$3.92x10^{-6}$	0.25
BovineHD0600029336	6	104931557	$4.08x10^{-6}$	0.32
Hapmap50596-BTA-121389	23	9591806	$4.09x10^{-6}$	0.13
Hapmap24296-BTA-136423	22	23729069	$4.23x10^{-6}$	0.09
BovineHD0800010906	8	36652588	$4.31x10^{-6}$	0.36
BovineHD4100006697	8	36784989	$4.31x10^{-6}$	0.36
BTB-00563665	11	14037552	$4.47x10^{-6}$	0.02
BovineHD0100029058	1	101897521	$4.78x10^{-6}$	0.29
BovineHD0400015737	4	57716679	$4.99x10^{-6}$	0.39
BovineHD0400015738	4	57722157	$4.99x10^{-6}$	0.39
BovineHD0300000514	3	1943906	$5.00x10^{-6}$	0.31
BovineHD0100019829	1	69606726	$5.05x10^{-6}$	0.34
Hapmap50449-BTA-70525	22	20074083	$5.44x10^{-6}$	0.09
BovineHD0100029057	1	101896647	$5.47x10^{-6}$	0.29
BovineHD0100044839	1	153840214	$5.50x10^{-6}$	0.14
BovineHD2200007059	22	23774905	$5.91x10^{-6}$	0.09
BovineHD0100019699	1	69368803	$5.96x10^{-6}$	0.33
BovineHD0200006087	2	21417927	$6.16x10^{-6}$	0.10
BovineHD0100021155	1	73788522	$6.80x10^{-6}$	0.05
BovineHD0400015521	4	56846638	$6.83x10^{-6}$	0.25
BovineHD2300001730	23	6906265	$6.84x10^{-6}$	0.20
BovineHD0400015515	4	56836287	$7.22x10^{-6}$	0.25
BovineHD0400015516	4	56838779	$7.22x10^{-6}$	0.25
BovineHD0400015517	4	56839981	$7.22x10^{-6}$	0.25
BovineHD0400015518	4	56841399	$7.22x10^{-6}$	0.25
BovineHD0200023904	2	83619501	$7.28x10^{-6}$	0.23
BovineHD0400033997	4	116938805	$7.42x10^{-6}$	0.45
BovineHD0400033998	4	116940082	$7.42x10^{-6}$	0.45
Hapmap53234-rs29020933	1	101892420	$7.78x10^{-6}$	0.37

BovineHD0100029054	1	101894131	$7.78x10^{-6}$	0.37
Hapmap27299-BTC-035816	6	40327335	$7.82x10^{-6}$	0.34
BovineHD0800026931	8	90586297	$7.83x10^{-6}$	0.02
BovineHD2900011558	29	38244678	$7.91x10^{-6}$	0.08
BovineHD2900011562	29	38273578	$7.91x10^{-6}$	0.08
BovineHD2900011563	29	38274429	$7.91x10^{-6}$	0.08
ARS-BFGL-NGS-22225	10	7136112	$7.91x10^{-6}$	0.02
BovineHD2000019456	20	67224771	$8.05x10^{-6}$	0.23
BovineHD0400015524	4	56860693	$8.32x10^{-6}$	0.25
BovineHD1900008536	19	29007477	$8.42x10^{-6}$	0.17
BovineHD1000031149	10	15400267	$9.11x10^{-6}$	0.08
BovineHD1900010745	19	37033960	$9.39x10^{-6}$	0.06

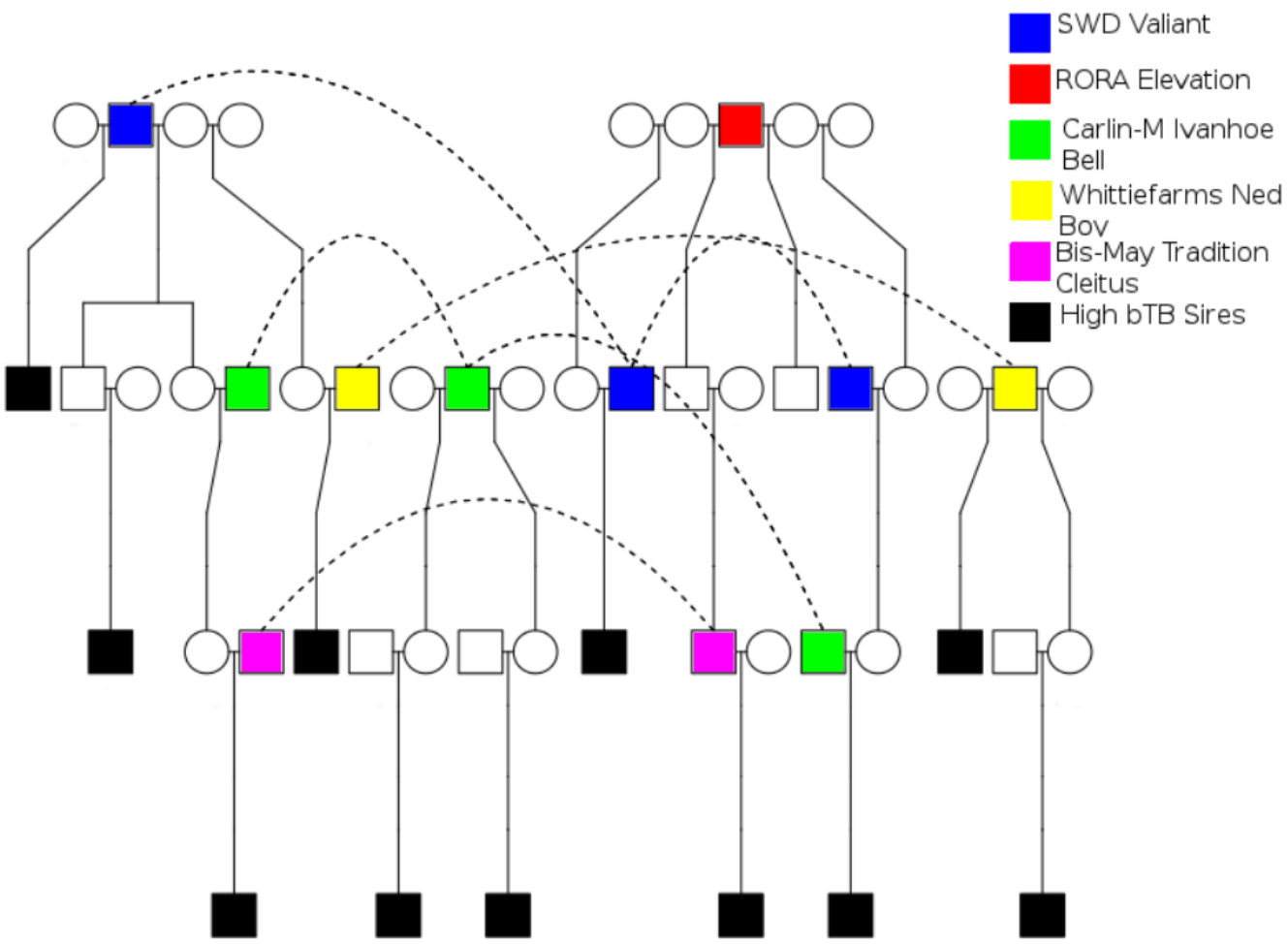
Supplementary table 2 All single nucleotide polymorphisms estimated to have Bayes factors > 20 from the Bayesian approach to GWAS

SNP	CHR	Genetic Position	Bayes Factor	MAF
BovineHD1700021001	17	72127337	285.557	0.3234
BovineHD1400000249	14	1892559	134.167	0.4875
BovineHD1000005406	10	16259113	83.45747	0.3276
BovineHD0800016027	8	53328810	64.89776	0.2842
BovineHD0800016025	8	53320264	60.8345	0.2842
BovineHD0100019801	1	69548446	57.58382	0.3377
BovineHD1400000243	14	1868636	54.68632	0.4875
BovineHD0600016172	6	58798792	51.47319	0.4233
BovineHD0600016181	6	58811643	49.7088	0.4334
BovineHD0600018068	6	65222126	47.62127	0.1046
BovineHD1700003641	17	12565967	47.29242	0.151
Hapmap25183-BTC-049425	14	6910008	46.85427	0.4721
BovineHD0300000514	3	1943906	45.86978	0.3175
BovineHD0800013179	8	44132612	45.43283	0.4792
BovineHD1400001861	14	6916008	43.68865	0.4721
BovineHD2000019681	20	67766491	41.9503	0.3115
BovineHD1700020903	17	71812992	41.625	0.2658
BovineHD1000002985	10	9147391	41.29991	0.2467
BovineHD4100010820	14	7121488	40.54214	0.4287
BovineHD0100029059	1	101899126	39.89351	0.2527
BovineHD1800007767	18	25182569	39.5695	0.2836
BovineHD2000019478	20	67268185	39.13779	0.4869
BovineHD0100029053	1	101890992	38.49091	0.3317
BovineHD0500022394	5	78922786	38.49091	0.4958
Hapmap42104-BTA-121232	2	81159639	38.16777	0.434
BovineHD1400000246	14	1880378	37.41457	0.4875
BovineHD4100010821	14	7123792	37.41457	0.4287

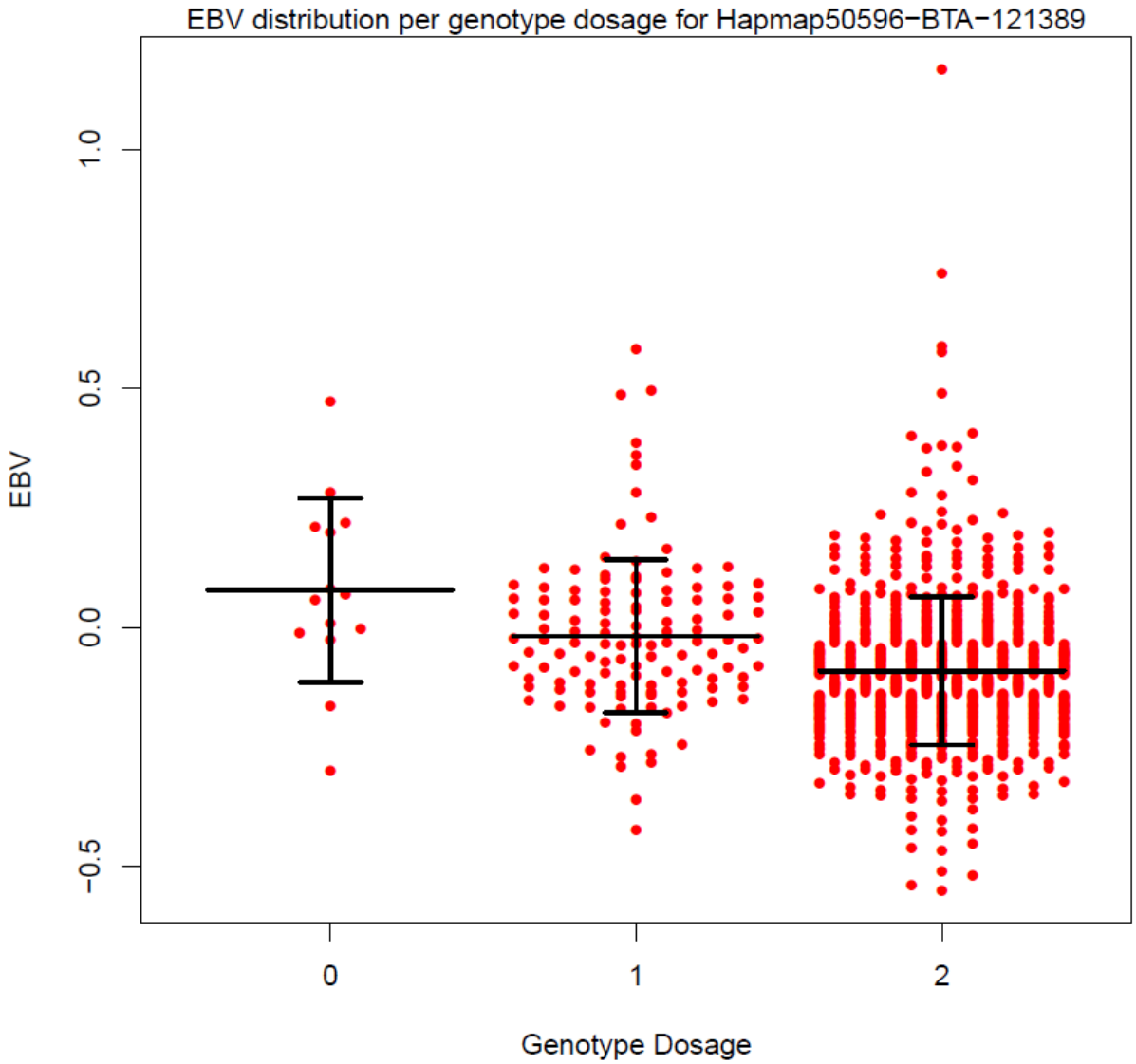
BovineHD2100002996	21	12032208	37.0921	0.3811
BovineHD2900004743	29	15963477	36.01865	0.1558
BovineHD0600024174	6	88436000	35.37565	0.3407
BovineHD0800016023	8	53311768	34.62649	0.2842
BovineHD0600016168	6	58794068	33.66487	0.4233
BovineHD2100005105	21	18142505	32.27903	0.4025
BovineHD2200007807	22	27217496	32.27903	0.3395
ARS-BFGL-NGS-79733	21	18140689	31.53435	0.4025
BovineHD2900004486	29	15223649	31.53435	0.3627
BovineHD0200040051	2	136633667	30.89691	0.3853
BovineHD0600029843	6	106313924	30.89691	0.1516
BovineHD2000019600	20	67513641	30.89691	0.2669
BovineHD4100003190	4	101767563	30.57848	0.4156
BovineHD0200023336	2	81172420	29.83625	0.434
ARS-BFGL-NGS-10346	17	12567376	29.20091	0.1492
BovineHD1700020906	17	71815400	28.88353	0.2658
BovineHD1700020907	17	71818985	28.88353	0.2658
BovineHD1800014284	18	48409309	28.46066	0.264
BovineHD2000019220	20	66569397	28.46066	0.4144
BovineHD2200007811	22	27222160	28.46066	0.3395
BovineHD0100029057	1	101896647	28.14374	0.2919
BovineHD0700021106	7	71544562	28.14374	0.2461
BovineHD2100009958	21	34694368	28.14374	0.3258
BovineHD4100018893	29	16796803	28.14374	0.126
BovineHD1100010579	11	35768623	27.82701	0.4643
ARS-BFGL-NGS-19947	6	58814212	27.51048	0.4334
BovineHD0600016173	6	58800548	27.19414	0.4233
BovineHD1100001671	11	4677803	27.19414	0.3014
BovineHD4100010816	14	7091599	27.19414	0.4304

BovineHD0100029060	1	101899991	26.77267	0.2539
BovineHD0300025132	3	87754159	26.77267	0.4168
BovineHD0600016170	6	58797218	26.77267	0.4233
BovineHD1400000271	14	2002126	26.77267	0.4822
Hapmap53234-rs29020933	1	101892420	26.45679	0.371
BovineHD1900005938	19	20698361	26.45679	0.1171
BTA-63683-no-rs	28	18663296	26.45679	0.4239
BovineHD1400001940	14	7164704	26.1411	0.4857
BovineHD2200007813	22	27229051	26.1411	0.3312
BovineHD2900002648	29	9063897	26.1411	0.4501
BovineHD0100033514	1	118726309	25.82561	0.3835
BovineHD0600034733	6	65241397	25.82561	0.1046
ARS-BFGL-NGS-64857	13	10926940	25.82561	0.299
BovineHD1300002984	13	10930589	25.09021	0.299
Hapmap38675-BTA-43020	18	33891066	25.09021	0.4828
BovineHD0100019699	1	69368803	24.77536	0.3329
BovineHD0200022581	2	78562397	23.72727	0.2842
BovineHD0800016030	8	53335899	23.72727	0.283
BovineHD0900007608	9	28195131	23.72727	0.2729
BovineHD2000019231	20	66597321	23.72727	0.4144
BovineHD0100041198	1	143234212	23.41326	0.1213
BovineHD0600000969	6	3712141	23.09945	0.4667
BovineHD0300024222	3	84790051	22.78582	0.396
BovineHD0600016166	6	58791807	22.78582	0.4233
BovineHD1300004762	13	16685980	22.78582	0.321
BovineHD2100005103	21	18139688	22.78582	0.4025
BovineHD2100009944	21	34672830	22.78582	0.3228
BovineHD0600016165	6	58789925	22.36796	0.4233
BovineHD0700021105	7	71543464	22.36796	0.2455

BovineHD1700020908	17	71819729	22.36796	0.2658
BovineHD2900005002	29	16799468	22.36796	0.126
BovineHD0400005469	4	18250653	22.05478	0.3056
BovineHD0600016176	6	58805297	22.05478	0.4233
BovineHD0600016178	6	58807281	22.05478	0.4233
BovineHD0800000084	8	397230	22.05478	0.3954
BovineHD1100010578	11	35767585	21.7418	0.4643
BovineHD1700020929	17	71853112	21.7418	0.2907
BovineHD2000019456	20	67224771	21.7418	0.2396
Hapmap50596-BTA-121389	23	9591806	21.7418	0.1361
BovineHD0600016162	6	58784831	21.42901	0.4233
BovineHD4100010844	14	7340004	21.42901	0.3496
BovineHD1700003642	17	12566669	21.42901	0.151
BovineHD0800013359	8	44720739	21.01225	0.4738
BovineHD1100005201	11	16583170	21.01225	0.2759
BovineHD0200022592	2	78603840	20.69991	0.2782
ARS-BFGL-NGS-114804	3	103336589	20.69991	0.3948
BovineHD0600017122	6	62131291	20.69991	0.2164
BovineHD2000019229	20	66595576	20.69991	0.4144
BovineHD0200022582	2	78566256	20.38776	0.2842
BovineHD0300026216	3	91110923	20.38776	0.2545
BovineHD0500022395	5	78923634	20.38776	0.497
BovineHD0400005437	4	18184158	20.07579	0.2533
BovineHD0600016163	6	58786220	20.07579	0.4233
ARS-BFGL-NGS-2712	18	23966401	20.07579	0.09631
BovineHD2200007809	22	27219995	20.07579	0.3312



Supplemental figure 1 Pedigree tree for sires with a high daughter prevalence of bTB infection (black squares) used as a case group for bTB susceptibility in identity by descent analysis.



Supplementary Figure 2 Distribution of estimated breeding values for bTB infection per genotype dosage

References

1. Irish Department of Agriculture, F.a.t.M. *Annual Review and Outlook for Agriculture, Food and the Marine*. 2015; Available from: http://www.agriculture.gov.ie/media/migration/publications/2015/ARO20141523061 5.pdf.
2. Federation, I.C.B. *Cattle breeding in Ireland*. 2007; Available from: http://www.icbf.com/publications/files/eaap_cattle_breeding_booklet.pdf.
3. Berry, D.P., et al., *Temporal trends in bulk tank somatic cell count and total bacterial count in Irish dairy herds during the past decade.* J Dairy Sci, 2006. **89**(10): p. 4083-93.
4. Crosson, P., et al., *The development of a mathematical model to investigate Irish beef production systems*. 2006, Elsevier Ltd.: Agricultural Systems. p. 349-370.
5. McCarthy, J. *Bull beef; Need to get basics right*. 2010 05/11/2010]; Available from: http://www.agriculture.gov.ie/media/migration/agri-foodindustry/agri-foodindustrypublications/2020Foodharvest190710.pdf.
6. Berry, D., et al. *The Economic Breeding Index: a generation on*. 2007 29/11/10]; Available from: www.icbf.com/publications/files/economic_breeding_index.pdf.
7. Nebel, R. and M. McGilliard, *Interactions of high milk yield and reproductive performance in dairy cows.* Journal of Dairy Science, 1993. **76**(10): p. 3257-3268.
8. Oltenacu, P. and D. Broom, *The impact of genetic selection for increased milk yield on the welfare of dairy cows.* Animal Welfare, 2010. **19**(1): p. 39-49.
9. Berry, D., et al., *Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows.* Journal of Dairy Science, 2003. **86**(6): p. 2193-2204.
10. Mrode, R.A. and R. Thompson, *Linear Models for the Prediction of Animal Breeding vales*. 2nd Edition ed. 2005: CABI Publishing.
11. *Introduction to Quantitative Genetics - Falconer,Ds.* Population, 1962. **17**(1): p. 152-153.
12. Falconer, D.S., *Introduction to quantitative genetics*. 1960: DS Falconer.

13. Mrode, R.A. and R. Thompson, *Linear Models for the Prediction of Animal Breeding Values, 2nd Edition*. Linear Models for the Prediction of Animal Breeding Values, 2nd Edition, 2005: p. 1-344.
14. Groeneveld, E., *PEST user's manual.* Institute of Animal Husbandry and Animal Behaviour, FAL, Germany, 1990.
15. Gilmour, A.R., Gogel, B.J., Cuillis, B.R., and Thompson, R., *ASReml User Guide Release 3.0*. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK, 2009.
16. Sorensen, D. and B. Kennedy, *The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments.* Theoretical and Applied Genetics, 1983. **66**(3-4): p. 217-220.
17. Bermingham, M.L., et al., *Genetics of tuberculosis in Irish Holstein-Friesian dairy herds.* Journal of dairy science, 2009. **92**(7): p. 3447-56.
18. Ramirez-Valverde, R., I. Misztal, and J. Bertrand, *Comparison of threshold vs linear and animal vs sire models for predicting direct and maternal genetic effects on calving difficulty in beef cattle.* Journal of animal science, 2001. **79**(2): p. 333-338.
19. Kadarmideen, H., R. Thompson, and G. Simm, *Linear and threshold model genetic parameters for disease, fertility and milk production in dairy cattle.* BSAS OCCASIONAL PUBLICATION, 2000: p. 83-84.
20. Bermingham, M., et al., *Genetics of tuberculosis in Irish Holstein-Friesian dairy herds.* J Dairy Sci, 2009. **92**(7): p. 3447-56.
21. Garrick, D.J., J.F. Taylor, and R.L. Fernando, *Deregressing estimated breeding values and weighting information for genomic regression analyses.* Genetics Selection Evolution, 2009. **41**.
22. Gondro, C., J. Van der Werf, and B. Hayes, *Genome-wide association studies and genomic prediction*. 2013: Humana Press.
23. Pritchard, J.K. and M. Przeworski, *Linkage disequilibrium in humans: models and data.* The American Journal of Human Genetics, 2001. **69**(1): p. 1-14.
24. Luo, Z., *Detecting linkage disequilibrium between a polymorphic marker locus and a trait locus in natural populations.* Heredity, 1998. **80**(2): p. 198-208.

25. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing*. Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.
26. Fernando, R., et al., *Controlling the proportion of false positives in multiple dependent tests*. Genetics, 2004. **166**(1): p. 611-619.
27. Weller, J.I., et al., *A new approach to the problem of multiple comparisons in the genetic dissection of complex traits*. Genetics, 1998. **150**(4): p. 1699-1706.
28. Browning, B.L. and S.R. Browning, *A fast, powerful method for detecting identity by descent*. The American Journal of Human Genetics, 2011. **88**(2): p. 173-182.
29. Browning, S.R. and B.L. Browning, *Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering*. American journal of human genetics, 2007. **81**(5): p. 1084-97.
30. Gusev, A., et al., *Whole population, genome-wide mapping of hidden relatedness*. Genome research, 2009. **19**(2): p. 318-326.
31. Browning, B.L. and S.R. Browning, *Improving the accuracy and efficiency of identity-by-descent detection in population data*. Genetics, 2013. **194**(2): p. 459-471.
32. Grapes, L., et al., *Optimal haplotype structure for linkage disequilibrium-based fine mapping of quantitative trait loci using identity by descent*. Genetics, 2006. **172**(3): p. 1955-1965.
33. Zhao, H., R.L. Fernando, and J.C. Dekkers, *Power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci*. Genetics, 2007. **175**(4): p. 1975-1986.
34. Hayes, B., et al., *Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle*. Genetical research, 2007. **89**(04): p. 215-220.
35. Calus, M., A. De Roos, and R. Veerkamp, *Accuracy of genomic selection using different methods to define haplotypes*. Genetics, 2008. **178**(1): p. 553-561.
36. Browning, S.R. and E.A. Thompson, *Detecting rare variant associations by identity-by-descent mapping in case-control studies*. Genetics, 2012. **190**(4): p. 1521-1531.

37. Maher, B., *Personal genomes: The case of the missing heritability.* Nature, 2008. **456**(7218): p. 18-21.
38. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-753.
39. Yang, J., et al., *Common SNPs explain a large proportion of the heritability for human height*. Nature genetics, 2010. **42**(7): p. 565-569.
40. Onteru, S., et al., *Whole-genome association analyses for lifetime reproductive traits in the pig.* Journal of animal science, 2011. **89**(4): p. 988-995.
41. Hayes, B.J., et al., *Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits.* PLoS Genet, 2010. **6**(9): p. e1001139.
42. Fan, B., et al., *Genome-wide association study identifies loci for body composition and structural soundness traits in pigs.* PloS one, 2011. **6**(2): p. e14726.
43. Hayes, B. and M. Goddard, *Prediction of total genetic value using genome-wide dense marker maps.* Genetics, 2001. **157**(4): p. 1819-1829.
44. Stephens, M. and D.J. Balding, *Bayesian statistical methods for genetic association studies.* Nat Rev Genet, 2009. **10**(10): p. 681-90.
45. Habier, D., et al., *Extension of the Bayesian alphabet for genomic selection.* BMC bioinformatics, 2011. **12**(1): p. 186.
46. Fernando, R. and D. Garrick, *GenSel-User manual for a portfolio of genomic selection related analyses.* Animal Breeding and Genetics, Iowa State University, Ames, 2008.
47. Bouwman, A.C., L.L. Janss, and H.C. Heuven, *A Bayesian approach to detect QTL affecting a simulated binary and quantitative trait.* BMC proceedings, 2011. **5 Suppl 3**: p. S4.
48. Raftery, K.A., R. Smithcoggins, and A.H. Ghen, *Gender-Associated Differences in Emergency Department Pain Management*. Annals of Emergency Medicine, 1995. **26**(4): p. 414-421.
49. Schaeffer, L., *Strategy for applying genome-wide selection in dairy cattle.* Journal of Animal Breeding and Genetics, 2006. **123**(4): p. 218-223.

50. VanRaden, P., *Efficient methods to compute genomic predictions.* Journal of dairy science, 2008. **91**(11): p. 4414-4423.
51. Habier, D., R. Fernando, and J. Dekkers, *The impact of genetic relationship information on genome-assisted breeding values.* Genetics, 2007. **177**(4): p. 2389-2397.
52. Berry, D.P., F. Kearney, and B.L. Harris, *Genomic selection in Ireland.* Interbull Bulletin, 2009(39): p. 29.
53. Garnier, T., et al., *The complete genome sequence of Mycobacterium bovis.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(13): p. 7877-7882.
54. Abernethy, D.A., et al., *Bovine tuberculosis trends in the UK and the Republic of Ireland, 1995-2010.* The Veterinary record, 2013. **172**(12): p. 312.
55. Perry, B.D., Randolph, T.F., McDermott, J.J., Sones, K.R. and Thornton, P.K., *Investing in animal health research to alleviate poverty.* ILRI (International Livestock Research Institute), Nairobi, Kenya. , 2002: p. 148.
56. O'Reilly, L. and C. Daborn, *The epidemiology of Mycobacterium bovis infections in animals and man: a review.* Tuber Lung Dis, 1995. **76 Suppl 1**: p. 1-46.
57. de la Rua-Domenech, R., *Human Mycobacterium bovis infection in the United Kingdom: Incidence, risks, control measures and review of the zoonotic aspects of bovine tuberculosis.* Tuberculosis (Edinb), 2006. **86**(2): p. 77-109.
58. Centre, H.P.S. *TB cases Notified in Ireland in 2008, Provisional data (as of 31st August 2009).* January 2010 [cited 2010 02/11/2010].
59. Garnier, T., et al., *The complete genome sequence of Mycobacterium bovis.* Proc Natl Acad Sci U S A, 2003. **100**(13): p. 7877-82.
60. Brosch, R., et al., *A new evolutionary scenario for the Mycobacterium tuberculosis complex.* Proceedings of the national academy of Sciences, 2002. **99**(6): p. 3684-3689.
61. Blanco, F., et al., *Differential transcriptome profiles of attenuated and hypervirulent strains of Mycobacterium bovis.* Microbes Infect, 2009. **11**(12): p. 956-63.

62. Neill, S., D. Bryson, and J. Pollock, *Pathogenesis of tuberculosis in cattle.* Tuberculosis, 2001. **81**(1): p. 79-86.
63. Phillips, C., et al., *The transmission of Mycobacterium bovis infection to cattle.* Research in veterinary science, 2003. **74**(1): p. 1-15.
64. Cousins, D., *Mycobacterium bovis infection and control in domestic livestock*. Revue scientifique et technique (International Office of Epizootics), 2001. **20**(1): p. 71-85.
65. de Lisle, G.W., C. Mackintosh, and R. Bengis, *Mycobacterium bovis in free-living and captive wildlife, including farmed deer.* Revue scientifique et technique (International Office of Epizootics), 2001. **20**(1): p. 86-111.
66. Denny, G. and J. Wilesmith, *Bovine tuberculosis in Northern Ireland: a case-control study of herd risk factors.* The Veterinary Record, 1999. **144**(12): p. 305-310.
67. More, S.J. and G. McGrath, *Randomised Badger Culling Trial: interpreting the results.* Veterinary Record, 2015. **177**(5): p. 128-129.
68. Bourne, F., et al., *Bovine TB: the scientific evidence.* A science base for a sustainable policy to control TB in cattle. An epidemiological investigation into bovine tuberculosis: Final Report of the Independent Scientific Group on Cattle TB, 2007.
69. Jenkins, H.E., R. Woodroffe, and C.A. Donnelly, *The effects of annual widespread badger culls on cattle tuberculosis following the cessation of culling.* International Journal of Infectious Diseases, 2008. **12**(5): p. 457-465.
70. Jenkins, H.E., R. Woodroffe, and C.A. Donnelly, *The duration of the effects of repeated widespread badger culling on cattle tuberculosis following the cessation of culling.* PLoS One, 2010. **5**(2): p. e9090.
71. Sheridan, M., et al., *The impact of an integrated wildlife and bovine tuberculosis eradication program in Ireland.* Zoonotic Tuberculosis: Mycobacterium bovis and Other Pathogenic Mycobacteria, Third Edition, 2014: p. 323-340.
72. Noonan, N.L., et al., *Wildlife as a possible reservoir of bovine tuberculosis*. 1975: Irsh Vet J. p. 1.

73. Griffin, J., et al., *The impact of badger removal on the control of tuberculosis in cattle herds in Ireland*. Preventive veterinary medicine, 2005. **67**(4): p. 237-266.
74. Griffin, J.M., et al., *Tuberculosis in cattle: the results of the four-area project*. Irish veterinary journal, 2005. **58**(11): p. 629.
75. Kelly, G.E., et al., *A long-term observational study of the impact of badger removal on herd restrictions due to bovine TB in the Irish midlands during 1989–2004.* Epidemiology and Infection, 2008. **136**(10): p. 1362-1373.
76. Corner, L.A., et al., *Oral vaccination of badgers (Meles meles) with BCG and protective immunity against endobronchial challenge with Mycobacterium bovis.* Vaccine, 2010. **28**(38): p. 6265-6272.
77. Aznar, I., et al., *Trial design to estimate the effect of vaccination on tuberculosis incidence in badgers.* Veterinary microbiology, 2011. **151**(1): p. 104-111.
78. More, S.J. and M. Good, *The tuberculosis eradication programme in Ireland: a review of scientific and policy advances since 1988.* Veterinary microbiology, 2006. **112**(2-4): p. 239-51.
79. More, S.J., *What is needed to eradicate bovine tuberculosis successfully: An Ireland perspective.* Veterinary Journal, 2009. **180**(3): p. 275-8.
80. Richardson, I.W., et al., *Variance components for susceptibility to Mycobacterium bovis infection in dairy and beef cattle.* Genet Sel Evol, 2014. **46**(1): p. 77.
81. Brotherstone, S., et al., *Evidence of genetic resistance of cattle to infection with Mycobacterium bovis.* Journal of dairy science, 2010. **93**(3): p. 1234-42.
82. Bermingham, M., et al., *Genetic correlations between measures of Mycobacterium bovis infection and economically important traits in Irish Holstein-Friesian dairy cows.* J Dairy Sci, 2010. **93**(11): p. 5413-22.
83. Tsairidou, S., et al., *Genomic prediction for tuberculosis resistance in dairy cattle.* PloS one, 2014. **9**(5).
84. Clegg, T.A., M. Good, and S.J. More, *Future risk of bovine tuberculosis recurrence among higher risk herds in Ireland.* Prev Vet Med, 2015. **118**(1): p. 71-9.
85. Allen, A.R., et al., *Bovine tuberculosis: the genetic basis of host susceptibility.* Proceedings. Biological sciences / The Royal Society, 2010. **277**(1695): p. 2737-45.

86. Good, M., *Bovine Tuberculosis Eradication in Ireland.* Irish Veterinary Journal 2006. **59**(3).
87. Payne, A., et al., *Bovine tuberculosis in "Eurasian" badgers (Meles meles) in France.* European Journal of Wildlife Research, 2013. **59**(3): p. 331-339.
88. Zanella, G., et al., *Modelling Transmission of Bovine Tuberculosis in Red Deer and Wild Boar in Normandy, France.* Zoonoses and public health, 2012. **59**: p. 170-178.
89. Aranaz, A., et al., *Bovine tuberculosis (Mycobacterium bovis) in wildlife in Spain.* Journal of clinical microbiology, 2004. **42**(6): p. 2602-2608.
90. Schoning, J.M., et al., *Surveillance of bovine tuberculosis and risk estimation of a future reservoir formation in wildlife in Switzerland and Liechtenstein.* PloS one, 2013. **8**(1): p. e54253.
91. Monaghan, M.L., et al., *The tuberculin test.* Veterinary microbiology, 1994. **40**(1-2): p. 111-24.
92. de la Rua-Domenech, R., et al., *Ante mortem diagnosis of tuberculosis in cattle: a review of the tuberculin tests, gamma-interferon assay and other ancillary diagnostic techniques.* Research in Veterinary Science, 2006. **81**(2): p. 190-210.
93. O'Keeffe, J.J. and M.J. Crowley, *Tuberculosis in cattle: Classifying breakdown episodes as a basis for decision making in eradication programmes.* Society for Veterinary Epidemiology and Preventive Medicine, Proceedings, 1998: p. 28-37.
94. VanRaden, P.M. and A.H. Sanders, *Economic merit of crossbred and purebred US dairy cattle.* J Dairy Sci, 2003. **86**(3): p. 1036-44.
95. Dempster, E.R. and I.M. Lerner, *Heritability of Threshold Characters.* Genetics, 1950. **35**(2): p. 212-36.
96. Berry, D.P., *Genetics - A tool to improve productivity and profitability.* International Journal of Dairy Technology, 2008. **61**(1): p. 30-35.
97. Berry, D.P., et al., *Genetics of animal health and disease in cattle.* Irish Veterinary Journal, 2011. **64**.
98. Berry, D.P., et al., *Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models.* Journal of dairy science, 2003. **86**(11): p. 3704-17.

99. McHugh, N., Evans, R.D., Amer, P.R., Fahey, A.G., Berry, D.P., *Genetic parameters for cattle price and body weight from routinely colelcted data at livestock auctions and commercial farms.* J. Anim. Sci., 2011. **89**: p. 29-39.
100. Pabiou, T., et al., *Genetic parameters for carcass cut weight in Irish beef cattle.* Journal of animal science, 2009. **87**(12): p. 3865-76.
101. Berry, D.P., et al., *Genetic variation in serological response to Mycobacterium avium subspecies paratuberculosis and its association with performance in Irish Holstein-Friesian dairy cows.* Livestock Science, 2010. **131**(1): p. 102-107.
102. Gonda, M.G., et al., *Genetic variation of Mycobacterium avium ssp paratuberculosis infection in US Holsteins.* Journal of dairy science, 2006. **89**(5): p. 1804-1812.
103. Mortensen, H., S.S. Nielsen, and P. Berg, *Genetic variation and heritability of the antibody response to Mycobacterium avium subspecies paratuberculosis in Danish Holstein cows.* Journal of dairy science, 2004. **87**(7): p. 2108-2113.
104. Sweeney, R.W., *Transmission of paratuberculosis.* The Veterinary clinics of North America. Food animal practice, 1996. **12**(2): p. 305-12.
105. Ameni, G., et al., *High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia.* Clinical and Vaccine Immunology, 2007. **14**(10): p. 1356-1361.
106. Schutz, M.M., P.M. Vanraden, and G.R. Wiggans, *Genetic-Variation in Lactation Means of Somatic-Cell Scores for 6 Breeds of Dairy-Cattle.* Journal of dairy science, 1994. **77**(1): p. 284-293.
107. Baird, L.G., et al., *Effects of breed and production system on lameness parameters in dairy cattle.* Journal of dairy science, 2009. **92**(5): p. 2174-2182.
108. Mattiello, S., et al., *Short communication: Breed differences affecting dairy cattle welfare in traditional alpine tie-stall husbandry systems.* Journal of dairy science, 2011. **94**(5): p. 2403-2407.
109. Snowder, G.D., et al., *Influence of breed, heterozygosity, and disease incidence on estimates of variance components of respiratory disease in preweaned beef calves.* Journal of animal science, 2005. **83**(6): p. 1247-61.

110. Berry, D.P., et al., *Associations among body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle.* J Dairy Sci, 2007. **90**(2): p. 637-48.
111. Heringstad, B., G. Klemetsdal, and J. Ruane, *Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries.* Livestock Production Science, 2000. **64**(2-3): p. 95-106.
112. Heringstad, B., G. Klemetsdal, and J. Ruane, *Responses to selection against clinical mastitis in the Norwegian cattle population.* Acta Agriculturae Scandinavica Section a-Animal Science, 2001. **51**(2): p. 155-160.
113. Bishop, S.C. and J.A. Woolliams, *On the genetic interpretation of disease data.* PLoS One, 2010. **5**(1): p. e8940.
114. Banos, G. and G.E. Shook, *Genotype by Environment Interaction and Genetic Correlations among Parities for Somatic-Cell Count and Milk-Yield.* Journal of dairy science, 1990. **73**(9): p. 2563-2573.
115. van Hulzen, K.J., et al., *Effect of herd prevalence on heritability estimates of antibody response to Mycobacterium avium subspecies paratuberculosis.* Journal of dairy science, 2011. **94**(2): p. 992-7.
116. Seifert, G.W., *Variations between and within breeds of cattle in resitance to field infestations of the cattle tick (Boophilus microplus).* Australian journal of agriculture, 1971. **22**: p. 159-68.
117. Stalhammar, H. and J. Philipsson, *Sex-specific genetic parameters for weaning and post-weaning gain in Swedish beef cattle under field conditions.* Acta Agriculturae Scandinavica Section a-Animal Science, 1997. **47**(3): p. 138-147.
118. N. Mc Hugh, R.D.E., P. R. Amer, A. G. Fahey, and D. P. Berry, *Genetic parameters for cattle price and body weight from routinely collected data at livestock auctions and commercial farms.* J. Anim. Sci., 2011. **89**: p. 29-39.
119. Calus, M.P.L., L.L.G. Janss, and R.F. Veerkamp, *Genotype by environment interaction for somatic cell score across bulk milk somatic cell count and days in milk*. Journal of dairy science, 2006. **89**(12): p. 4846-4857.

120. Mackintosh, C.G., et al., *Genetic resistance to experimental infection with Mycobacterium bovis in red deer (Cervus elaphus).* Infection and immunity, 2000. **68**(3): p. 1620-1625.
121. Bermingham, M.L., et al., *Evidence for genetic variance in resistance to tuberculosis in Great Britain and Irish Holstein-Friesian populations.* BMC proceedings, 2011. **5 Suppl 4**: p. S15.
122. Conlan, A.J.K., et al., *Estimating the Hidden Burden of Bovine Tuberculosis in Great Britain.* Plos Computational Biology, 2012. **8**(10).
123. Karolemeas, K., et al., *Estimation of the relative sensitivity of the comparative tuberculin skin test in tuberculous cattle herds subjected to depopulation.* PloS one, 2012. **7**(8): p. e43217.
124. Settles, M., et al., *A whole genome association analysis identifies loci associated with Mycobacterium avium subsp. paratuberculosis infection status in US holstein cattle.* Animal genetics, 2009. **40**(5): p. 655-62.
125. Minozzi, G., et al., *Genetic loci involved in antibody response to Mycobacterium avium ssp. paratuberculosis in cattle.* PloS one, 2010. **5**(6): p. e11117.
126. Minozzi, G., et al., *Meta-analysis of two genome-wide association studies of bovine paratuberculosis.* PloS one, 2012. **7**(3): p. e32578.
127. Finlay, E.K., et al., *A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle.* PloS one, 2012. **7**(2): p. e30545.
128. Bermingham, M.L., et al., *Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis.* Heredity (Edinb), 2014. **112**(5): p. 543-51.
129. Kassahun, Y., et al., *Admixture mapping of tuberculosis and pigmentation-related traits in an African–European hybrid cattle population.* Frontiers in genetics, 2015. **6**.
130. Kassahun, Y., et al., *Admixture mapping of tuberculosis and pigmentation-related traits in an African-European hybrid cattle population.* Front Genet, 2015. **6**: p. 210.

131. Moller, M., E. de Wit, and E.G. Hoal, *Past, present and future directions in human genetic susceptibility to tuberculosis.* FEMS immunology and medical microbiology, 2010. **58**(1): p. 3-26.
132. Kramnik, I., et al., *Genetic control of resistance to experimental infection with virulent Mycobacterium tuberculosis.* Proc Natl Acad Sci U S A, 2000. **97**(15): p. 8560-5.
133. Barthel, R., et al., *Pathologic findings and association of Mycobacterium bovis infection with the bovine NRAMP1 gene in cattle from herds with naturally occurring tuberculosis.* American journal of veterinary research, 2000. **61**(9): p. 1140-1144.
134. Bellamy, R., et al., *Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans.* New England Journal of Medicine, 1998. **338**(10): p. 640-644.
135. Kadarmideen, H., et al., *Polymorphisms of the SLC11A1 gene and resistance to bovine tuberculosis in African Zebu cattle.* Animal genetics, 2011. **42**(6): p. 656-658.
136. Browning, B.L. and S.R. Browning, *A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals.* American journal of human genetics, 2009. **84**(2): p. 210-23.
137. Berry, D., M. McClure, and M. Mullen, *Within-and across-breed imputation of high-density genotypes in dairy and beef cattle from medium-and low-density genotypes.* Journal of Animal Breeding and Genetics, 2014. **131**(3): p. 165-172.
138. Meyer, K., *WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML).* Journal of Zhejiang University. Science. B, 2007. **8**(11): p. 815-21.
139. Meyer, K. and B. Tier, *"SNP Snappy": a strategy for fast genome-wide association studies fitting a full mixed model.* Genetics, 2012. **190**(1): p. 275-7.
140. Kizilkaya, K., R. Fernando, and D. Garrick, *Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes.* Journal of animal science, 2010. **88**(2): p. 544-551.

141. Kass, R.E. and A.E. Raftery, *Bayes Factors*. Journal of the American Statistical Association, 1995. **90**(430): p. 773-795.
142. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
143. Breuer, K., et al., *InnateDB: systems biology of innate immunity and beyond-recent updates and continuing curation.* Nucleic Acids Research, 2013. **41**(D1): p. D1228-D1233.
144. Browning, B.L. and S.R. Browning, *A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals.* Am J Hum Genet, 2009. **84**(2): p. 210-23.
145. Daetwyler, H.D., et al., *Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle.* Nat Genet, 2014. **46**(8): p. 858-65.
146. Bhuachalla, N., et al., *The role of badgers in the epidemiology of Mycobacterium bovis infection (tuberculosis) in cattle in the United Kingdom and the Republic of Ireland: current perspectives on control strategies.* Veterinary Medicine: Research and Reports, 2015. **6**: p. 27-38.
147. Yang, J., et al., *Genomic inflation factors under polygenic inheritance.* European Journal of Human Genetics, 2011. **19**(7): p. 807-812.
148. Baughman, G., et al., *Tissue distribution and abundance of human FKBP51, and FK506-binding protein that can mediate calcineurin inhibition.* Biochem Biophys Res Commun, 1997. **232**(2): p. 437-43.
149. Flanagan, S.E., et al., *Next-generation sequencing reveals deep intronic cryptic ABCC8 and HADH splicing founder mutations causing hyperinsulinism by pseudoexon activation.* Am J Hum Genet, 2013. **92**(1): p. 131-6.
150. Agrawal, A., et al., *An intronic ABCA3 mutation that is responsible for respiratory disease.* Pediatr Res, 2012. **71**(6): p. 633-7.
151. Harland, M., et al., *A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees.* Hum Mol Genet, 2001. **10**(23): p. 2679-86.
152. Berry, D.P., et al., *Associations between the K232A polymorphism in the diacylglycerol-O-transferase 1 (DGAT1) gene and performance in Irish Holstein-Friesian dairy cattle.* Irish Journal of Agricultural and Food Research, 2010: p. 1-9.

153. Khor, C.C. and M.L. Hibberd, *Host-pathogen interactions revealed by human genome-wide surveys.* Trends Genet, 2012. **28**(5): p. 233-43.
154. Cassidy, J.P. and A.R. Martineau, *Innate resistance to tuberculosis in man, cattle and laboratory animal models: nipping disease in the bud?* J Comp Pathol, 2014. **151**(4): p. 291-308.
155. Purfield, D.C., et al., *Genome-wide association study for calving traits in Holstein-Friesian dairy cattle.* Animal, 2014. **8**(2): p. 224-35.
156. Vithana, E.N., et al., *Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma.* Nat Genet, 2012. **44**(10): p. 1142-6.
157. Meredith, B.K., et al., *Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland.* BMC Genet, 2012. **13**: p. 21.
158. Consortium, W.T.C.C., *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
159. Winkler, C.A., G.W. Nelson, and M.W. Smith, *Admixture mapping comes of age*.* Annual review of genomics and human genetics, 2010. **11**: p. 65-89.
160. Gautier, M. and M. Naves, *Footprints of selection in the ancestral admixture of a New World Creole cattle breed.* Molecular Ecology, 2011. **20**(15): p. 3128-3143.
161. Qanbari, S., et al., *A genome-wide scan for signatures of recent selection in Holstein cattle.* Animal genetics, 2010. **41**(4): p. 377-389.
162. Sabeti, P.C., et al., *Detecting recent positive selection in the human genome from haplotype structure.* Nature, 2002. **419**(6909): p. 832-837.
163. Nielsen, R., *Molecular signatures of natural selection.* Annu. Rev. Genet., 2005. **39**: p. 197-218.
164. Voight, B.F., et al., *A map of recent positive selection in the human genome.* PLoS biology, 2006. **4**(3): p. 446.
165. Kim, E.-S. and M.F. Rothschild, *Genomic adaptation of admixed dairy cattle in East Africa.* Frontiers in genetics, 2014. **5**.
166. Bahbahani, H., et al., *Signatures of positive selection in East African Shorthorn Zebu: A genome-wide single nucleotide polymorphism analysis.* Sci Rep, 2015. **5**: p. 11729.

167. Kemper, K.E., et al., *Selection for complex traits leaves little or no classic signatures of selection.* BMC genomics, 2014. **15**(1): p. 246.
168. Zhao, F., et al., *Detection of selection signatures in dairy and beef cattle using high-density genomic information.* Genetics Selection Evolution, 2015. **47**(1): p. 1-12.
169. Qanbari, S., et al., *Classic selective sweeps revealed by massive sequencing in cattle.* PLoS Genet, 2014. **10**(2): p. e1004148.
170. Patterson, N., et al., *Methods for high-density admixture mapping of disease genes.* The American Journal of Human Genetics, 2004. **74**(5): p. 979-1000.
171. McKeigue, P.M., *Prospects for admixture mapping of complex traits.* The American Journal of Human Genetics, 2005. **76**(1): p. 1-7.
172. Kopp, J.B., et al., *MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis.* Nature genetics, 2008. **40**(10): p. 1175-1184.
173. Loftus, R.T., et al., *Evidence for two independent domestications of cattle.* Proceedings of the National Academy of Sciences, 1994. **91**(7): p. 2757-2761.
174. Cartwright, T., *Prognosis of Zebu cattle: Research and application.* Journal of Animal Science, 1980. **50**(6): p. 1221-1226.
175. Grigson, C., *An African origin for African cattle?—some archaeological evidence.* African Archaeological Review, 1991. **9**(1): p. 119-144.
176. McTavish, E.J. and D.M. Hillis, *A genomic approach for distinguishing between recent and ancient admixture as applied to cattle.* Journal of Heredity, 2014. **105**(4): p. 445-456.
177. Gravel, S., *Population genetics models of local ancestry.* Genetics, 2012. **191**(2): p. 607-619.
178. Harris, K. and R. Nielsen, *Inferring demographic history from a spectrum of shared haplotype lengths.* 2013.
179. Patterson, N., et al., *Ancient admixture in human history.* Genetics, 2012. **192**(3): p. 1065-1093.
180. Loh, P.-R., et al., *Inferring admixture histories of human populations using linkage disequilibrium*. Genetics, 2013. **193**(4): p. 1233-1254.

181. Lawson, D.J., et al., *Inference of population structure using dense haplotype data.* PLoS Genet, 2012. **8**(1): p. e1002453-e1002453.
182. Hellenthal, G., et al., *A genetic atlas of human admixture history.* Science, 2014. **343**(6172): p. 747-751.
183. Clutton-Brock, J., *A natural history of domesticated mammals*. 1999: Cambridge University Press.
184. Freeman, A., et al., *Admixture and diversity in West African cattle populations.* Molecular Ecology, 2004. **13**(11): p. 3477-3487.
185. Decker, J.E., et al., *Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.* Proceedings of the National Academy of Sciences, 2009. **106**(44): p. 18644-18649.
186. Bollongino, R., et al., *Modern taurine cattle descended from small number of Near-Eastern founders.* Molecular biology and evolution, 2012. **29**(9): p. 2101-2104.
187. Ho, S.Y., et al., *Correlating Bayesian date estimates with climatic events and domestication using a bovine case study.* Biology Letters, 2008. **4**(4): p. 370-374.
188. Murray, C., et al., *Cattle demographic history modelled from autosomal sequence variation.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2010. **365**(1552): p. 2531-2539.
189. Achilli, A., et al., *The multifaceted origin of taurine cattle reflected by the mitochondrial genome.* PLoS One, 2009. **4**(6): p. e5753.
190. Helmer, D., et al., *Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism*. 2005: na.
191. Edwards, C., J. Baird, and D. MacHugh, *Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data.* Animal Genetics, 2007. **38**(5): p. 520-524.
192. Freeman, A.R., et al., *Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the Near East using X chromosome haplotype mosaicism*. Genetics, 2006. **173**(3): p. 1503-1510.
193. Clason, A., *Late Bronze Age-Iron Age zebu cattle in Jordan?* Journal of Archaeological Science, 1978. **5**(1): p. 91-93.
194. Felius, M., *Cattle breeds: an encyclopedia*. 1995: C Misset bv.

195. Zheng, X., et al., *A high-performance computing toolset for relatedness and principal component analysis of SNP data.* Bioinformatics, 2012. **28**(24): p. 3326-3328.

196. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome research, 2009. **19**(9): p. 1655-1664.

197. Gower, J.C., *Some distance properties of latent root and vector methods used in multivariate analysis.* Biometrika, 1966. **53**(3-4): p. 325-338.

198. Price, A.L., et al., *Sensitive detection of chromosomal segments of distinct ancestry in admixed populations.* PLoS Genet, 2009. **5**(6): p. e1000519.

199. Wilkins, J., *Criollo cattle of the Americas.* Animal Genetic Resources Information, 1984. **2**: p. 1-19.

200. Gautier, M. and R. Vitalis, *rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure.* Bioinformatics, 2012. **28**(8): p. 1176-1177.

201. Pan, D., et al., *Genome-wide detection of selective signature in Chinese Holstein.* PloS one, 2013. **8**(3).

202. Seabury, C.M., et al., *Diversity and evolution of 11 innate immune genes in Bos taurus taurus and Bos taurus indicus cattle.* Proceedings of the National Academy of Sciences, 2010. **107**(1): p. 151-156.

203. McDermott, C. and B. Fenwick, *Neutrophil activation associated with increased neutrophil acyloxyacyl hydrolase activity during inflammation in cattle.* American journal of veterinary research, 1992. **53**(5): p. 803-807.

204. McTavish, E.J., et al., *New World cattle show ancestry from multiple independent domestication events.* Proceedings of the National Academy of Sciences, 2013. **110**(15): p. E1398-E1406.

205. Briggs, H.M., *Modern breeds of livestock*. Modern breeds of livestock., 1969.

206. Gelhaus, A., et al., *YAC/BAC contig spanning the MHC class III region of cattle.* Cytogenetic and genome research, 2006. **115**(1): p. 45-50.

207. Workman, A., et al., *A protein (ORF2) encoded by the latency-related gene of bovine herpesvirus 1 interacts with Notch1 and Notch3.* Journal of virology, 2011. **85**(6): p. 2536-2546.

208. Pittayakhajonwut, D., D. Sinani, and C. Jones, *A Protein (ORF2) Encoded by the Latency-Related Gene of Bovine Herpesvirus 1 Interacts with DNA*. Journal of virology, 2013. **87**(10): p. 5493-5501.

209. González-Rodríguez, A. *Genome-wide Analysis of Genetic Diversity in Autochthonous Spanish Populations of Beef Cattle*. in *10th World Congress on Genetics Applied to Livestock Production*. 2014. Asas.

210. Rothammer, S., et al., *A genome-wide scan for signatures of differential artificial selection in ten cattle breeds*. BMC genomics, 2013. **14**(1): p. 908.

211. Grisart, B., et al., *Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition*. Genome research, 2002. **12**(2): p. 222-231.

212. Zhang, Z., et al., *[Polymorphisms of DGAT2 gene and its associations with several growth traits in Nanyang cattle]*. Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji, 2007. **29**(8): p. 945-950.

213. Mao, H., et al., *[Association of polymorphisms of DGAT2 gene with growth traits in Jiaxian red cattle]*. Yi chuan= Hereditas/Zhongguo yi chuan xue hui bian ji, 2008. **30**(3): p. 329-332.

214. Karlengen, I., et al., *The effect of excess cobalt on milk fatty acid profiles and transcriptional regulation of SCD, FASN, DGAT1 and DGAT2 in the mammary gland of lactating dairy cows*. Journal of animal physiology and animal nutrition, 2012. **96**(6): p. 1065-1073.

215. Gutiérrez-Gil, B., et al., *Identification of quantitative trait loci affecting cattle temperament*. Journal of heredity, 2008. **99**(6): p. 629-638.

216. Tizioto, P.C., et al., *Genome scan for meat quality traits in Nelore beef cattle*. Physiological genomics, 2013. **45**(21): p. 1012-1020.

217. Mancini, G., et al., *Signatures of selection in five Italian cattle breeds detected by a 54K SNP panel*. Molecular biology reports, 2014. **41**(2): p. 957-965.

218. Kim, E.-S., et al., *Effect of artificial selection on runs of homozygosity in US Holstein cattle*. 2013.

219. Meade, K.G., et al., *Innate gene repression associated with Mycobacterium bovis infection in cattle: toward a gene signature of disease.* BMC genomics, 2007. **8**(1): p. 400.

220. Wu, H., et al., *TALE nickase-mediated SP110 knockin endows cattle with increased resistance to tuberculosis.* Proc Natl Acad Sci U S A, 2015. **112**(13): p. E1530-9.