Animal (2017), **11:1**, pp 15–23 © The Animal Consortium 2016
doi:10.1017/S1751731116001099

animal

Inference of population structure of purebred dairy and beef cattle using high-density genotype data

M. M. Kelleher[1,2], D. P. Berry[1†], J. F. Kearney[3], S. McParland[1], F. Buckley[1] and D. C. Purfield[1]

[1]*Animal & Grassland Research and Innovation Centre, Teagasc, Moorepark, Co. Cork, Ireland;* [2]*School of Agriculture, Food Science, University College Dublin, Belfield, Dublin 4, Ireland;* [3]*Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland*

(Received 17 April 2015; Accepted 6 April 2016; First published online 22 June 2016)

Information on the genetic diversity and population structure of cattle breeds is useful when deciding the most optimal, for example, crossbreeding strategies to improve phenotypic performance by exploiting heterosis. The present study investigated the genetic diversity and population structure of the most prominent dairy and beef breeds used in Ireland. Illumina high-density genotypes (777 962 single nucleotide polymorphisms; SNPs) were available on 4623 purebred bulls from nine breeds; Angus (n = 430), Belgian Blue (n = 298), Charolais (n = 893), Hereford (n = 327), Holstein-Friesian (n = 1261), Jersey (n = 75), Limousin (n = 943), Montbéliarde (n = 33) and Simmental (n = 363). Principal component analysis revealed that Angus, Hereford, and Jersey formed non-overlapping clusters, representing distinct populations. In contrast, overlapping clusters suggested geographical proximity of origin and genetic similarity between Limousin, Simmental and Montbéliarde and to a lesser extent between Holstein, Friesian and Belgian Blue. The observed SNP heterozygosity averaged across all loci was 0.379. The Belgian Blue had the greatest mean observed heterozygosity ($H_O = 0.389$) among individuals within breed while the Holstein-Friesian and Jersey populations had the lowest mean heterozygosity ($H_O = 0.370$ and 0.376, respectively). The correlation between the genomic-based and pedigree-based inbreeding coefficients was weak (r = 0.171; P < 0.001). Mean genomic inbreeding estimates were greatest for Jersey (0.173) and least for Hereford (0.051). The pair-wise breed fixation index (F_{st}) ranged from 0.049 (Limousin and Charolais) to 0.165 (Hereford and Jersey). In conclusion, substantial genetic variation exists among breeds commercially used in Ireland. Thus custom-mating strategies would be successful in maximising the exploitation of heterosis in crossbreeding strategies.

Keywords: genetic diversity, population structure, fixation index, phylogenetic

Implications

The strong use of artificial selection to increase the frequency of favourable alleles at the loci affecting phenotypic performance (e.g. milk production), coupled with intense selection that featured heavily in some breeds, has resulted in reduced genetic diversity within breed and has affected the extent of genomic homozygosity. In this study high-density genotypic data on a large number of individuals of different breeds was available to make inferences of population structure for the different breeds used in temperate cattle production systems.

Introduction

Domestic cattle (*Bos taurus* and *Bos indicus*) originated from populations of the wild extinct aurochs ~10 000 years ago (Diamond, 2002). The complex origins that led to the domestication of the modern cow have led to numerous different cattle breeds specialised for different traits, such as milk and meat products, disease and pest resistance, drafting ability and religious beliefs. The differentiation among breeds is a result of natural selection, geographical variability due to drift, and more recently, strong artificial selection for traits of economic importance. Consequently, modern cattle breeds, of which there are more than 1000 recognised breeds worldwide (Felius, 2007), display extensive phenotypic variety depending on the economic and cultural goals.

Pioneering studies have quantified cattle population genetic structure and variation primarily using polymorphic microsatellite markers (MacHugh *et al.*, 1997; MacHugh *et al.*, 1998; Canon *et al.*, 2001; Beja-Pereira *et al.*, 2003). The advent of affordable high-throughput genotyping technologies (Hayes *et al.*, 2009; Calus, 2010) can however provide greater insight into the effect of selection on the distribution of genetic variation and evolution of domestic

† E-mail: Donagh.berry@teagasc.ie

cattle. Because selection alters the allele frequencies of markers that are in close proximity to the selected mutation (Bamshad and Wooding, 2003), the availability of tens of thousands of single nucleotide polymorphism (SNP) information overcomes the previous constraints where only markers of limited density were available. This technology facilitates the detection of breed-specific signatures of selection (Xu *et al.*, 2015; Zhao *et al.*, 2015), and can be used to broadly cluster cattle breeds into distinct groups as well as estimate the genetic diversity within and among breeds. Such information may be useful for producers in the design of herd mating programmes, especially those exploiting crossbreeding. Heterosis, which is the added increase in performance when animals are crossbred, is associated with an increase in genomic heterozygosity levels. Therefore, genomic information has the potential to provide more accurate estimates of the breed composition and genomic distance among breeds compared to conventionally recorded pedigree data. Although previous studies have investigated population substructure and genetic diversity using high-density genotype technology (Matukumalli *et al.*, 2009; Decker *et al.*, 2014; Edea *et al.*, 2015), larger sample sizes per population may be required to demonstrate more confidence in the inferences made (Pruett and Winker, 2008).

The objective of the present study was to differentiate a population of dairy and beef purebred cattle used in Ireland into different strata of breeds using high-density SNP data on a large number of individuals. All individuals were sires used in either commercial artificial insemination programmes or natural mating bulls in Ireland and are therefore representative of germplasm used in Irish dairy and beef breeding programmes.

Material and methods

Genotypic data

Illumina high-density genotypes (777 962 SNPs) were available on 4623 purebred artificial insemination and natural service bulls of different breeds used in Ireland; the exception was the Holstein and Friesian breeds which were grouped as Holstein-Friesian with animal 'breed' in the present study dictated by the recorded major breed proportion. All animals had a genotype call rate of >0.95. Breed composition of all animals for K ancestral populations was verified using the ADMIXTURE 1.23 software (Alexander *et al.*, 2009). An unsupervised ADMIXTURE analysis was used to estimate individual admixture proportions from the SNP data using a maximum likelihood method without prior information of individual ancestry. The programme ran for 38 iterations with a specified number of clusters ($K = 10$) assumed to reflect the major genetic components of the data set. Breeds represented included Angus ($n = 430$), Belgian Blue ($n = 298$), Charolais ($n = 893$), Friesian ($n = 174$), Hereford ($n = 327$), Holstein ($n = 1087$), Jersey ($n = 75$), Limousin ($n = 943$), Montbéliarde ($n = 33$) and Simmental ($n = 363$) bulls; all animals with the exception of Holstein and Friesian were predicted to be >87.5% pure. The pedigree of all animals was traced back to founder animals. Mendelian inconsistencies were used to validate animal identification through parentage assessment but also to discard autosomal SNPs that did not adhere to Mendelian inheritance. SNPs with <0.02 minor allele frequency and SNPs that deviated from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-9}$) within breed were also discarded. After edits a total of 624 179 autosomal SNPs remained.

Analysis

Effective population size. For data quality control, the effective population size was investigated using the SNeP tool described by Barbato *et al.* (2015). The method examines the relationship between the variance in linkage disequilibrium and effective population size in the presence of mutation.

$$N_{T(T)} = \left(4f\left(c_{(t)}\right)^{-1}\left(E\left[r^{|ct}_{adj}\right]^{-1} - \alpha\right)\right)$$

where N_T is the effective population size T generations ago calculated as $T = 2f\left(c_{(t)}\right)^{-1}$ (Hayes *et al.*, 2003), c_t the recombination rate for specific physical distance between markers calculated by the SNeP tool using default values, r_{adj} the linkage disequilibrium value adjusted for sample size, and $\alpha = \{1, 2, 2.2\}$ is the correction for the occurrence of mutation (Ohta and Kimura, 1971).

Principal component analysis. Population substructure for all animals was determined using principal component analysis (PCA) in the EIGENSOFT package (Price *et al.*, 2006). This algorithm uses a computationally efficient variant of eigenanalysis to determine the probability of population substructure according to Tracy–Widom distribution. The Tracy–Widom distribution describes the density of the largest eigenvalue of a random Hermitian matrix (Tracy and Widom, 2009). It is assumed that nonzero eigenvalues of the matrix are within the Tracy–Widom distribution and was used in the present study to test the statistical significance of whether an eigenvector was a significant principal component or not of the matrix (Patterson *et al.*, 2006). PCA plots were constructed using the first four components from the analysis.

Genetic diversity. Genetic diversity analyses were performed using PLINK v1.07 software (Purcell *et al.*, 2007). For each population, genetic diversity was measured as the observed heterozygosity (H_O) averaged over loci, and the expected heterozygosity (H_E) under the assumptions of Hardy–Weinberg equilibrium. Expected heterozygosity was calculated using the following equation:

$$H_E = \frac{\sum_{k=1}^{r} h_k}{r}$$

where $h_k = \frac{2n\left(1 - \sum x_i^2\right)}{(2n-1)}$

where h_k is the population heterozygosity at the kth locus, r the total number of loci studied, n the number of individuals per locus, and x_i the frequency of the ith allele at locus x in a sample population n (Nei, 1978). Genomic inbreeding coefficients (F) were computed based on the observed *v.*

the expected number of homozygous genotypes (Purcell *et al.*, 2007):

$$F = f_i + (1 - f_i)(p^2 + q^2)$$

where f_i is the probability that individual i is homozygous by decent, $(1 - f_i)$ is the probability that individual i is homozygous by chance, for a particular SNP with known allele frequencies p and q (Purcell *et al.*, 2007). Pedigree-based inbreeding coefficients for each animal were also estimated using the Meuwissen and Luo (1992) algorithm as described by McParland *et al.* (2007b) where the pedigree was traced back to the founders.

The pair-wise genetic differentiation among breeds was calculated using the F_{st} statistic. The F_{st} was calculated as:

$$F_{st} = \frac{s^2}{\bar{p}(1 - \bar{p})}$$

where s^2 is the variance of allele frequency among breeds and $\bar{p}$ the mean allele frequency across breed (Weir and Cockerham, 1984). A phylogenetic tree was computed based on the breed pair-wise F_{st} using the APE package in R software (Paradis *et al.*, 2004). The software provides functions to estimate phylogenetic trees with distance-based DNA information to facilitate comparative and diversification analyses.

The within-breed allele frequencies for each SNP were calculated as pair-wise breed correlations of the allele frequencies per SNP (Kuehn *et al.*, 2011). The delta value (Δ_{ij}) was calculated as the breed pair-wise absolute differences in allele frequency (p_A) between each population i and j, averaged over n SNPs:

$$\Delta_{ij} = \frac{\sum |p_{A_i} - p_{A_j}|}{n}$$

Results and discussion

Knowledge of the genomic distance between cattle breeds is particularly useful in the identification of distinct cattle breeds that might be exploited to maximise heterosis in crossbreeding programmes. The primary rationale for adopting a crossbreeding strategy is the enhanced phenotypic performance observed in the crossbred offspring, relative to their parental mean performance. The superior performance accruing from heterosis is attributable to the increased genomic heterozygosity due to the crossing of different breeds (Falconer and Mackay, 1996). Although pure-breeding in dairy has traditionally dominated, results from our study reveal that the effective population size has rapidly decreased in recent years (Supplementary Figure S1) which is in agreement with previous studies in international dairy populations (de Roos *et al.*, 2008; Gibbs *et al.*, 2009). Preliminary analyses revealed little genomic contribution from the founder animals; the standard deviation in genomic relationships varied from 0.035 (Limousin) to 0.150 (Holstein). The observed pattern in effective population size is possibly due to bottlenecks associated with intense selection and domestication (de Roos *et al.*, 2008) thereby intensifying the interest in crossbreeding strategies internationally (e.g. Sørensen *et al.*, 2008; Buckley *et al.*, 2014; LIC, 2015). Beef production, on the contrary, has had a strong tradition of crossbreeding in many countries (Kahn and Cottle, 2014). For example, Williams *et al.* (2010) reported individual heterosis estimates ranging between 0.63 kg (Continental × Continental) to 2.43 kg (British × Zebu), 3.47 kg (Continental × Continental) to 25.93 kg (Continental × Zebu), and 1.49 kg (Continental × Zebu) to 14.68 kg (British × Zebu), for increased birth weight, weaning weight, and post weaning BW gain in beef populations, respectively. The greater the genetic distance between breeds, the greater, on average, the extent of heterosis observed (Ehiobu *et al.*, 1990). Therefore, further investigation into the genomic differences between breeds is a very pertinent and relevant question in aiding the decision on what breeds to include in a crossbreeding strategy. This is especially true where breed effective population sizes have undergone rapid reductions possibly due to intense within breed selection.

Principal component analysis

PCA is based on a mathematical algorithm that reduces the dimensionality of the number of possibly correlated variables (e.g. SNPs) into a fewer number of uncorrelated variables (i.e. principal components), while retaining most of the variability in the data set under investigation (Jolliffe, 2002). Principal component analysis has previously been used to visualise the genomic relationships among cattle breeds (Gautier *et al.*, 2010; Lewis *et al.*, 2011; Harris *et al.*, 2014; Edea *et al.*, 2015).

The principal component analysis in the present study was successful in separating out breed clusters based on the genotypic data. The first and second principal components accounted for 37.20% and 27.22% of the variation, respectively (Figure 1). The third and fourth principal components

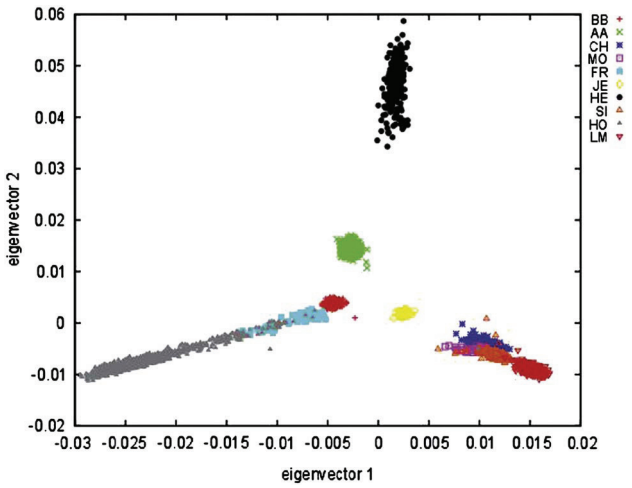

Figure 1 Principal component analysis of the purebred animals (AA = Angus; BB = Belgian Blue; CH = Charolais; FR = Friesian; HE = Hereford; HO = Holstein; JE = Jersey; LM = Limousin; MO = Montbéliarde; SM = Simmental) distributed across the first two principal components explaining 64.42% of the variation.

accounted for 21.10% and 14.48% of the variation, respectively. Pair-wise plots (Supplementary Figure S2) and the cumulative proportion of variance (Supplementary Figure S3) were plotted for each principal component.

PCA revealed that the three cattle breeds, Angus, Jersey and Hereford formed clear, non-overlapping clusters that represented separate populations. The first principal component depicted a distinct pattern of separation between these British Isle breeds and the remaining breeds in the study. The original stock of (Aberdeen) Angus developed from native polled and predominately black cattle of remote counties of Aberdeenshire and Angus in Northeast Scotland in the early 19th century (http://www.ansi.okstate.edu/breeds/cattle). The Jersey breed originated from Jersey Island in the English Channel off the coast of France. Effectively, the Jersey breed has developed in genetic isolation due to strict importing and breeding practices for over 200 years (MacHugh *et al*., 1997). The Hereford breed is an older breed than the Angus and Jersey, and formed in the 18th century in county Herefordshire in England. At this time, there was a growing food demand owing to the British Industrial Revolution which resulted in farmers selecting for high carcass yield and efficiency of production from grass and grain. The original Hereford was a much larger animal (i.e. mature bull weight of 1300 kg or more) than the modern Hereford today but gradually reduced in size in order to prioritise meat eating quality (http://www.ansi.okstate.edu/breeds/cattle). The second principal component axis of the PCA analysis revealed a clear distance between the Hereford and the other British Isle breeds as well as all other breeds in the current study. These results corroborate MacHugh *et al.* (1994) in that the Hereford does not exhibit a great deal of affinity to other breeds and may be as a result of northern Europe influence before the establishment of the breed in the mid-19th century.

The Holstein-Friesian and Belgian Blue populations each formed distinct clusters but were in close proximity to each other, with some Friesian individuals positioned within the Holstein breed cluster. The Holstein and Friesian clusters were widely distributed along the first principal component and were overlapping, characterising the relatedness between these dairy black and white breeds. The genetic similarity between these two breeds is in agreement with previous studies (Harris *et al*., 2014; Kim and Rothschild, 2014). Both breeds originated from two Northern provinces, North Holland and West Friesland, in the Netherlands which are located on either side of the shallow bay of the Zuider Zee (Del Bo *et al*., 2001). Approximately 2000 years ago the original black and white cattle were bred to generate the high producing Holstein-Friesian. The differences between the Holstein and Friesian became prominent as a consequence of differing selection objectives and the subsequent ban on importations of livestock in the late-19th century as a precaution against the outbreak of Foot and Mouth in Europe. For instance, the North American Holstein was established when Holstein-Friesians were exported from the Netherlands into the United States in the mid-19th century (Porter, 1991). Intense selection for high milk yield to satisfy the growing market demands has resulted in the separation of the original Holstein-Friesian into Holstein and Friesian. The Friesian, primarily residing in United Kingdom, the Netherlands and New Zealand, was selected as a dual purpose breed and tends to be of smaller stature, of better carcass conformation and produce less milk compared to the Holstein (Harris and Kolver, 2001; Horan *et al*., 2005). There was some degree of overlap between the Holstein and Friesian in the PCA. The pedigree-based and genotype-based information for assignment of breed for these animals were, however, not always in agreement, as depicted by overlapping clusters between both breeds on both principal component axes. This may be as a result of errors in differentiation of breed at the time of registration of the animal (i.e. Holstein and Friesian animals look very similar) but also due to the difficulty arising from accurately identifying the primary breed of the animal since Mendelian sampling during meiosis can influence what proportion of each breed of a crossbred parent has been inherited. The Belgian Blue cluster resided in close proximity to the Friesian on the PCA plot. The Belgian Blue may share common ancestry to the Friesian as a result of the crossing of Dutch Black Pied (which are similar to Friesian) and Shorthorn (Blott *et al*., 1998). The Belgian Blue breed is well known for their exceptional muscular development. Originally, the breed was selected as a dual purpose breed. Selection for increased muscular hypertrophy intensified as a result of premiums paid for double-muscled carcasses. The uptake in use of artificial insemination also contributed to the fixation of the breed's defining features in less than 20 years, between the 1960s and 1980s.

The French and Swiss breeds, Limousin, Charolais, Simmental and Montbéliarde formed clusters in close proximity and overlapping with each other on the first two principal components (Figure 1). However the third principal component clearly shows that the Charolais does not overlap with the remaining French and Swiss breeds (Supplementary Figure S2) similar to that documented by Lewis *et al.* (2011). The establishment of both the Limousin and Charolais breeds is not known, but thought to have occurred centuries ago, in relatively isolated parts of central and southwest France which reduced genetic interference from other cattle breeds. The Simmental is also thought to be amongst the oldest and most widely distributed of breeds worldwide. These dual purpose red and white cattle originated from western Switzerland (http://www.ansi.okstate.edu/breeds/breeds/cattle). The Montbéliarde breed (also known as the French Dairy Simmental) bears a close visual resemblance and coat markings to that of the Simmental. The Montbéliarde originated in a neighbouring region to the Simmental, in the Haute Saône-Doubs region of east France. Overall, the continental breeds appear to have some similarities in their gene pools and may be a result of several factors such as less intensive selection, absence of geographic barriers, and increased admixture over the years.

Heterozygosity

Heterozygosity is a measure of genetic diversity and is sensitive to the frequency of alleles (Hawksworth, 1995).

The observed heterozygosity is the observed occurrence of heterozygous individuals (at a locus). The expected heterozygosity, on the other hand, is the expected heterozygosity level under random mating conditions (Melka and Schenkel, 2012). Similar observed and expected heterozygosity ($H_O = 0.379$ and $H_E = 0.381$, respectively) for all SNPs were estimated in the present study, corroborating previously documented statistics in similar breeds where SNP average heterozygosity values were estimated as 0.30, 0.30 and 0.38 (Matukumalli *et al.*, 2009; Gautier *et al.*, 2010; Melka and Schenkel, 2012, respectively). Greater levels of heterozygosity indicate that the population in question could have greater adaptive genetic variation and fitness in comparison to populations of lower levels of heterozygosity. The heterozygosity statistics however reported in the present study, and elsewhere, must take cognisance of the selection of SNPs for inclusion in the SNP arrays and thus the associated ascertainment bias. Only a small number of individuals from a selected population are used to develop the SNP array panels, and therefore may not be representative of other populations (Matukumalli *et al.*, 2009; Albrechtsen *et al.*, 2010). Although only European breeds were considered in the present study, not all breed sequences are represented on the SNP arrays used and therefore may underestimate genetic diversity differences between breeds to make breeds appear more homogeneous or more different from each other than they really are.

Across all breeds, the differences between the observed heterozygosity and expected heterozygosity values were either nominally smaller or not different at all from the expected values ($\chi^2 = 0.0015$, $P > 0.05$; Table 1). Montbéliarde had the greatest difference between the observed and expected heterozygosity values, albeit the observed differences may be an artefact of the small sample size ($n = 33$) in the present study (Pruett and Winker, 2008). The Belgian Blue population had the greatest mean heterozygosity of all breeds ($H_O = 0.389$) while the Holstein-Friesian population had the lowest mean heterozygosity ($H_O = 0.370$). Globally, the dairy breeding objective for Holstein-Friesians focussed traditionally on a narrow selection of elite sires propagated throughout the world (Miglior *et al.*, 2005). The loss of genetic diversity (i.e. high genetic uniformity), as well as economic concerns associated with compromised long-term biological functionality of breeds are a real concern particularly with the use of advanced reproductive technologies that may accelerate the problem further (Thompson *et al.*, 2000). Genetic uniformity reduces the fitness and sustainability of breeds because individuals have a reduced ability to react to perturbations such as climate change (Kristensen *et al.*, 2015).

Inbreeding

Mean genomic-based inbreeding and mean pedigree-based inbreeding of each breed is in Table 1. The level of genomic inbreeding ranged from 0% and 36% across all breeds. Median genomic inbreeding estimates per breed were greatest for Jersey ($f_{genomic} = 0.173$) and least for Hereford ($f_{genomic} = 0.051$). A weak spearman correlation ($r = 0.171$; $P < 0.001$) existed between the genomic-based and pedigree-based inbreeding coefficients. The intercept (0.074; SE = 0.0005) of the regression of the genomic inbreeding coefficient on the pedigree inbreeding coefficient was greater than zero suggesting that the pedigree inbreeding coefficient may be underestimating the true level of inbreeding by ~7 percentage units. This statistic corroborates similar findings by Purfield *et al.* (2012) where runs of homozygosity (ROH) were used to determine the extent of inbreeding within individuals. Purfield *et al.* (2012) documented that the pedigree-based inbreeding coefficient underestimated the genomic inbreeding, as measured by ROH, by ~9 percentage units. Only considered in the present study was actual genomic inbreeding at a global level; it is however possible to quantify locus-specific, or region-specific level of inbreeding (Pryce *et al.*, 2014).

Inbreeding depression is the opposite of crossbreeding and reduces the phenotypic performance of the inbred offspring as compared to the average of its parents (Weigel, 2001). The rate of inbreeding in Ireland is increasing by 0.10% per annum in dairy populations and up to 0.13% in some beef populations (McParland *et al.*, 2007b). Greater accumulation of inbreeding exacerbates the risk of homozygous recessive deleterious alleles that affect phenotypic performance (Thompson *et al.*, 2000). Resulting offspring are more susceptible to compromised performance particularly in

Table 1 *Number of animals, mean (median in parentheses) observed heterozygosity (H_O), mean (median in parentheses) expected heterozygosity (H_E), as well as the mean genomic ($f_{genomic}$) and pedigree inbreeding ($f_{pedigree}$) coefficients of each breed*

Breed	*n*	H_O	H_E	$f_{genomic}$	$f_{pedigree}$
Angus	430	0.385 (0.421)	0.387 (0.430)	0.115	0.002
Belgian Blue	298	0.389 (0.420)	0.390 (0.432)	0.074	0.001
Charolais	893	0.381 (0.419)	0.381 (0.422)	0.065	0.004
Hereford	327	0.388 (0.421)	0.392 (0.433)	0.051	0.008
Holstein-Friesian	1261	0.370 (0.408)	0.372 (0.412)	0.075	0.034
Jersey	75	0.376 (0.387)	0.390 (0.435)	0.173	0.020
Limousin	943	0.379 (0.417)	0.380 (0.420)	0.075	0.003
Montbéliarde	33	0.379 (0.367)	0.398 (0.444)	0.120	0.015
Simmental	363	0.384 (0.417)	0.386 (0.428)	0.096	0.008

reproductive ability (Wall *et al.*, 2005; McParland *et al.*, 2007a) and vigour (Thompson *et al.*, 2000; Weigel, 2001). The Jersey was the most inbred breed in the present study corroborating results from Stachowicz *et al.* (2011) where the percentage inbreeding in the Jersey breed has increased since the 1980s to 14% between the years 2000 and 2007.

Fixation index

The F_{st} is a measure of genetic differentiation, measured as the reduction in heterozygosity of subpopulations relative to the total population (Weir and Hill, 2002). The F_{st} is calculated as the correlation between a pair of random alleles in the subpopulation and randomly sampled alleles from the entire population. Values for F_{st} range between 0 and 1. Low F_{st} values among subpopulations indicate low levels of genetic divergence in the population, where a value of 0 implies no subdivision between the populations. High F_{st} values indicate greater genetic differentiation between subpopulations, implying that individuals within subpopulations are more related to each other compared to individuals between subpopulations. A F_{st} value of 1 indicates complete isolation of the subpopulation from the total population; in other words, the genomic structure of an entire population can be explained by the subpopulation.

The mean F_{st} value across all loci for all individuals in the population was 0.108. This result corroborates similar levels of differentiation between lineages of 0.112 and 0.107 reported in microsatellite marker studies of European cattle by MacHugh *et al.* (1998) and Kantanen *et al.* (2000), respectively. McKay *et al.* (2008) estimated a mean F_{st} value of 0.100 from a whole genome SNP panel of *Bos taurus* cattle populations. The lower F_{st} for the French and Swiss breeds suggests a greater genetic differentiation within these breeds compared to the other breeds in the study, a conclusion that is supported by the previous genetic differentiation analyses in the present study (Table 1). The high F_{st} for Jersey is further supported by previous studies that have documented a reduced genetic variation in Jersey (MacHugh *et al.*, 1994; Chikhi *et al.*, 2004). The positive moderate correlation ($r = 0.48$) between the mean breed F_{st} and mean genomic-based inbreeding values indicates that genetically isolated breeds (i.e. high F_{st} values) are more likely to be more inbred (i.e. high genomic inbreeding coefficient).

The breed pair-wise F_{st} was used to estimate the degree of population genetic differentiation among breeds (Table 2). The measure of genetic differentiation between breed populations relative to the total population (i.e. pair-wise F_{st}) ranged from 0.049 (Limousin and Charolais) to 0.165 (Jersey and Hereford). The observed genetic distance between the cattle breeds using pair-wise F_{st} estimates is further illustrated in an unrooted phylogenetic tree in Figure 2. Populations originating from the same branch depict closeness in genetic relationships between breeds (i.e. Belgian Blue and Holstein-Friesian; Montbéliarde and Simmental). The French and Swiss breeds (Charolais, Limousin, Montbéliarde and Simmental) resided in close proximity, similar to that seen in Figure 1, suggesting that the location of origin was central in determining relatedness (Gautier *et al.*, 2010). The phylogenetic tree complements the PCA (Figure 1). The British Isles breeds (i.e. Angus, Jersey and Hereford) had the longest branches of the phylogeny and

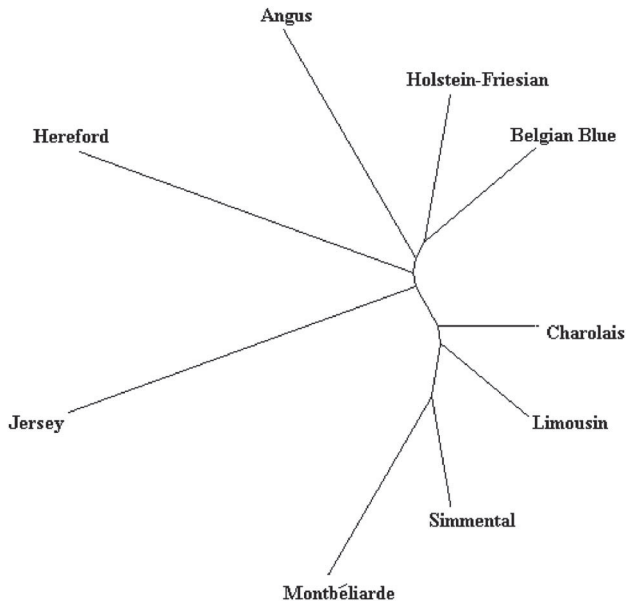

Figure 2 Genetic distance between breeds based on pair-wise F_{st} estimates for Irish cattle breeds.

Table 2 *Mean* F_{st} *values among breeds*

	Angus	Belgian Blue	Charolais	Hereford	Holstein-Friesian	Jersey	Limousin	Montbéliarde	Simmental
Angus		0.094	0.095	0.141	0.100	0.152	0.108	0.142	0.121
Belgian Blue			0.070	0.123	0.067	0.131	0.085	0.117	0.097
Charolais				0.119	0.074	0.119	0.049	0.088	0.066
Hereford					0.124	0.165	0.124	0.155	0.136
Holstein-Friesian						0.126	0.082	0.112	0.093
Jersey							0.120	0.152	0.133
Limousin								0.086	0.063
Montbéliarde									0.073
Simmental									

Table 3 *Mean delta values (below diagonal) and correlations in allele frequencies per single nucleotide polymorphism (above diagonal) among breeds*

	Angus	Belgian Blue	Charolais	Hereford	Holstein-Friesian	Jersey	Limousin	Montbéliarde	Simmental
Angus		0.769	0.760	0.608	0.744	0.647	0.724	0.661	0.698
Belgian Blue	0.151		0.820	0.643	0.825	0.691	0.777	0.715	0.751
Charolais	0.153	0.129		0.647	0.798	0.720	0.874	0.793	0.838
Hereford	0.198	0.184	0.181		0.625	0.568	0.633	0.580	0.612
Holstein-Friesian	0.158	0.126	0.135	0.186		0.698	0.777	0.721	0.754
Jersey	0.196	0.182	0.172	0.216	0.179		0.720	0.662	0.695
Limousin	0.165	0.144	0.106	0.185	0.142	0.172		0.801	0.847
Montbéliarde	0.190	0.171	0.145	0.208	0.169	0.194	0.142		0.842
Simmental	0.174	0.154	0.123	0.194	0.153	0.181	0.119	0.127	

appeared quite distinctive from the other breed groups. Although F_{st} values were used on a genome-wide basis in the present study, it is also possible to quantify F_{st} values for each individual locus or genomic region; locus-specific F_{st} values could be useful to make inferences about selection signatures. Zhao *et al.* (2015) used such an approach of using regional F_{st} values on a subset of the data used in the present study and documented 704 individual SNPs suggestive of selection signatures.

In a genome-wide analysis of sheep breeds, Kijas *et al.* (2012) highlighted that short branches on the phylogenetic tree were associated with greater breed heterozygosity whereas long branches were associated with low breed heterozygosity. The Jersey had the longest and isolated branch from all of the other studied breeds. In general, the phylogenetic tree separated the individuals in the study by geographical origin. Furthermore, breeds that have branched from a common source or breeds in close placement (e.g. the relationship between Holstein-Friesian and the Belgian Blue) depict the admixed nature of populations and closer genomic relationships compared to longer branches that suggest the breed diverged a long time ago (e.g. Hereford). These findings, combined with the PCA results, suggest that breeds of considerable genomic distance have the potential to result in greater levels of heterosis, relative to more related breeds (Ehiobu *et al.*, 1990). For instance, Penasa *et al.* (2010) reported that the first generation Holstein-Jersey crossbred dairy cows yielded 477 kg more milk, 25.3 kg more fat and 17.4 kg more protein, compared to the parental average. However, the heterosis effects for the Holstein-Friesian crossbred were approximately half the size for the Holstein-Jersey crossbreds (Penasa *et al.*, 2010).

Delta and allele frequency

Mean allele frequencies per population can be used as a measure of the genetic diversity. Inferences of the population substructure were similar, irrespective of whether they were based on the delta statistic (Δ) or the correlations between breed pair allele frequencies (r_{allele}; Table 3). The Limousin and Charolais were most similar genetically ($\Delta = 0.106$ and $r_{allele} = 0.874$) and the Jersey and Hereford were the most genetically different ($\Delta = 0.216$ and $r_{allele} = 0.568$). These findings further corroborate the PCA and the F_{st} results. Breed pairs with a strong interchange of germplasm had a low genetic differentiation and small differences in allele frequencies (i.e. Δ). It is thought that the use of breed herd books over time has accentuated the separation of breeds by reducing the degree of genetic exchange with other breeds (MacHugh *et al.*, 1997).

Conclusions

The alternative measures of genetic diversity in the present study are complementary to each other and, in general, support the conclusion that the geographic origin of breed has had a strong influence on the current genetic makeup of the alternative breeds. Use of the high-density SNP chip data can be used to quantify the genetic structure among cattle breeds and to reduce errors in pedigree recording. The most genetically isolated breeds were the British Isle breeds, whereas the French and Swiss breeds represented a strong genetic flow due to common origin and admixture over time. Genetically uniform breeds are at a greater risk of homozygous recessive allele expression and reduced fitness and performance. Some genetic variability could be recovered through the use of crossbreeding thus maximising performance. Although the analyses performed in the present study were at a genome-wide level, similar approaches such as region-specific genomic inbreeding (Pryce *et al.*, 2014), runs of homozygosity (Purfield *et al.*, 2012) or selection signatures (Zhao *et al.*, 2015) may also be used to describe interesting breed-specific differentiation patterns.

Acknowledgement

Financial support from the Irish Department of Agriculture Stimulus Research Fund Multi GS and Science Foundation Ireland Precision Breeding is gratefully acknowledged.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1751731116001099

References

Albrechtsen A, Nielsen FC and Nielsen R 2010. Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution 27, 2534–2547.

Alexander DH, Novembre J and Lange K 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19, 1655–1664.

Bamshad M and Wooding SP 2003. Signatures of natural selection in the human genome. Nature Reviews Genetics 4, 99–111.

Barbato M, Orozco-terWengel P, Tapio M and Bruford MW 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics 6, 1–12.

Beja-Pereira A, Alexandrino P, Bessa I, Carretero Y, Dunner S, Ferrand N, Jordana J, Laloe D, Moazami-Goudarzi K, Sanchez A and Canon J 2003. Genetic characterization of southwestern European bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. Journal of Heredity 94, 243–250.

Blott SC, Williams JL and Haley CS 1998. Genetic relationships among European cattle breeds. Animal Genetics 29, 273–282.

Buckley F, Lopez-Villalobos N and Heins BJ 2014. Crossbreeding: implications for dairy cow fertility and survival. Animal 8, 122–133.

Calus MP 2010. Genomic breeding value prediction: methods and procedures. Animal 4, 157–164.

Canon J, Alexandrino P, Bessa I, Carleos C, Carretero Y, Dunner S, Ferran N, Garcia D, Jordana J, Laloe D, Pereira A, Sanchez A and Moazami-Goudarzi K 2001. Genetic diversity measures of local European beef cattle breeds for conservation purposes. Genetic Selection Evolution 33, 311–332.

Chikhi L, Goossens B, Treanor A and Bruford MW 2004. Population genetic structure of and inbreeding in an insular cattle breed, the Jersey, and its implications for genetic resource management. Heredity 92, 396–401.

Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, Babar ME, Correia de Almeida Regitano L, Yildiz MA, Heaton MP, Liu WS, Lei CZ, Reecy JM, Saif-Ur-Rehman M, Schnabel RD and Taylor JF 2014. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. PLoS Genetics 10, e1004254.

De Roos APW, Hayes BJ, Spelman RJ and Goddard ME 2008. Linkage disequilibrium and persitence of phase in Holsetin-Friesian, Jersey and Anjus cattle. Genetics 179, 1503–1512.

Del Bo L, Polli M, Longeri M, Ceriotti G, Looft C, Barre-Dirie A, Dolf G and Zanotti M 2001. Genetic diversity among some cattle breeds in the Alpine area. Journal of Animal Breeding and Genetics 118, 317–325.

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

Edea Z, Bhuiyan MS, Dessie T, Rothschild MF, Dadi H and Kim KS 2015. Genome-wide genetic diversity, population structure and admixture analysis in African and Asian cattle breeds. Animal 9, 218–226.

Ehiobu NG, Goddard ME and Taylor JF 1990. Prediction of heterosis in crosses between inbred lines of *Drosophila melanogaster*. Theoretical and Applied Genetics 80, 321–325.

Falconer DS and Mackay TFC 1996. Introduction to quantitative genetics, 4th edition. Pearson Education LTD, Essex, UK.

Felius M 2007. Cattle breeds: an encyclopedia. Trafalgar Square Books, North Pomfret, United States.

Gautier M, Laloë D and Moazami-Goudarzi K 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One 5, e13038.

Gibbs RA, Taylor JF, van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Sonstegard TS, Williams JL, Diallo B, Hailemariam L, Martinez ML, Morris CA, Silva LO, Spelman RJ, Mulatu W, Zhao K, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mwakaya J, Plastow G, Sim W, Smith T, Thomas MB, Valentini A, Williams P, Womack J, Woolliams JA, Liu Y, Qin X, Worley KC, Gao C, Jiang H, Moore SS, Ren Y, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Patil S, San Lucas A, Fu Q, Kent MP, Vega R, Matukumalli A, McWilliam S, Sclep G, Bryc K, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Goddard ME, Hayes BJ, Bradley DG, Barbosa da Silva M, Lau LP, Liu GE, Lynn DJ, Panzitta F and Dodds KG 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528–532.

Harris BL and Kolver ES 2001. Review of Holsteinization of intensive pastoral dairy farming in New Zealand. Journal of Dairy Science 84 (suppl.), E56–E61.

Harris BL, Winkelman AM and Johnson DE 2014. Across-breed genomic prediction in dairy cattle. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Livestock Improvement Corporation, Hamilton, New Zealand.

Hawksworth DL 1995. Biodiversity: measurement and estimation. Chapman and Hall, London, UK.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009. Invited review: genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science 92, 433–443.

Hayes BJ, Visscher PM, McPartlan HC and Goddard ME 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research 13, 635–643.

Horan B, Dillon P, Faverdin P, Delaby L, Buckley F and Rath M 2005. The interaction of strain of Holstein-Friesian cows and pasture-based feed systems on milk yield, body weight and body condition score. Journal of Diary Science 88, 1231–1243.

Jolliffe IT 2002. Principal component analysis Springer. Springer-Verlag, New York.

Kahn L and Cottle D 2014. Beef cattle production and trade. CSIRO Publishing, Victoria, Australia.

Kantanen J, Olsaker I, Holm L-E, Lien S, Vilkki J, Brusgaard K, Eythorsdottir E, Danell B and Adalsteinsson S 2000. Genetic diversity and population structure of 20 north European cattle breeds. Journal of Heredity 91, 446–457.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J and Dalrymple B, other members of the International Sheep Genomics C 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biology 10, e1001258.

Kim E and Rothschild MF 2014. Genomic adaptation of admixed dairy cattle in East Africa. Frontiers in Genetics 5, 443–453.

Kristensen TN, Hoffmann AA, Pertoldi C and Stronen AV 2015. What can livestock breeders learn from conservation genetics and *vice versa*? Frontiers in Genetics 6, 1–12.

Kuehn LA, Keele JW, Bennett GL, McDaneld TG, Smith TP, Snelling WM, Sonstegard TS and Thallman RM 2011. Predicting breed composition using breed frequencies of 50 000 markers from the US Meat Animal Research Center 2000 Bull Project. Journal of Animal Science 89, 1742–1750.

LIC 2015. Livestock Improvement Corporation. Retrieved on 17 April 2015 from www.lic.co.nz

Lewis J, Abas Z, Dadousis C, Lykidis D, Paschou P and Drineas P 2011. Tracing cattle breeds with principal components analysis ancestry informative SNPs. PLoS One 6, e18007.

MacHugh DE, Loftus RT, Bradley DG, Sharp PM and Cunningham P 1994. Microsatellite DNA variation within and among European cattle breeds. Proceedings of the Royal Society: Biological Sciences 256, 25–31.

MacHugh DE, Loftus RT, Cunningham P and Bradley DG 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. Animal Genetics 29, 333–340.

MacHugh DE, Shriver MD, Loftus RT, Cunningham P and Bradley DG 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146, 1071–1086.

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS and Van Tassell CP 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, e5350.

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Wang Z, Van Tassell CP, Williams JL, Taylor JF and Moore SS 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics 9, 37.

McParland S, Kearney JF, Rath M and Berry DP 2007a. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. Journal of Dairy Science 90, 4411–4419.

McParland S, Kearney JF, Rath M and Berry DP 2007b. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. Journal of Animal Science 85, 322–331.

Melka MG and Schenkel FS 2012. Analysis of genetic diversity in Brown Swiss, Jersey and Holstein populations using genome-wide single nucleotide polymorphism markers. BMC Research Notes 5, 161.

Meuwissen THE and Luo Z 1992. Computing inbreeding coefficients in large populations. Genetic Selection Evolution 24, 305–313.

Miglior F, Muir BL and Van Doormaal BJ 2005. Selection indices in Holstein cattle of various countries. Journal of Dairy Science 88, 1255–1263.

Nei M 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89, 583–590.

Ohta T and Kimura M 1971. Linkage disequilibrium between two segregating nucleotide sites under the steady flux of mutations in a finite population. Genetics 68, 571–580.

Paradis E, Claude J and Strimmer K 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Patterson N, Price AL and Reich D 2006. Population structure and eigenanalysis. PLoS Genetics 2, 2074–2093.

Penasa M, López-Villalobos N, Evans RD, Cromie AR, Dal Zotto R and Cassandro M 2010. Crossbreeding effects on milk yield traits and calving interval in spring-calving dairy cows. Journal of Animal Breeding and Genetics 127, 300–307.

Porter V 1991. Cattle – a handbook to the breeds of the world. Facts on File Inc, New York, USA.

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

Pruett CL and Winker K 2008. The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. Journal of Avian Biology 39, 252–256.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

Pryce JE, Haile-Mariam M, Goddard ME and Hayes BJ 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. Genetics Selection Evolution 46, 71.

Purfield DC, Berry DP, McParland S and Bradley DG 2012. Runs of homozygosity and population history in cattle. BMC Genetics 13, 70–80.

Sørensen MK, Norberg E, Pedersen J and Christensen LG 2008. Invited review: cross-breeding in dairy cattle: a Danish perspective. Journal of Dairy Science 91, 4116–4128.

Stachowicz K, Sargolzaei M, Miglior F and Schenkel FS 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. Journal of Dairy Science 94, 5160–5175.

Thompson JR, Everett RW and Hammerschmidt NL 2000. Effects of inbreeding on production and survival in Holsteins. Journal of Dairy Science 83, 1856–1864.

Tracy CA and Widom H 2009. The distributions of random matrix theory and their applications. In New trends in mathematical physics (ed. V Sidoravicius), pp. 753–765. Springer, The Netherlands.

Wall E, Brotherstone S, Kearney JF, Woolliams JA and Coffey MP 2005. Impact of nonadditive genetic effects in the estimation of breeding values for fertility and correlated traits. Journal of Dairy Science 88, 376–385.

Weigel KA 2001. Controlling inbreeding in modern breeding programs. Journal of Dairy Science 84, E177–E184.

Weir BS and Cockerham CC 1984. Esti mating F-statistics for the analysis of population structure. Evolution 38, 1358–1370.

Weir BS and Hill WG 2002. Estimating F-statistics. Annual Review of Genetics 36, 721–750.

Williams JL, Aguilar I, Rekaya R and Bertrand JK 2010. Estimation of breed and heterosis effects for growth and carcass traits in cattle using published crossbreeding studies. Journal of Animal Science 88, 460–466.

Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Tassell CPV, Sonstegard TS and Liu GE 2015. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. Molecular Biology and Evolution 32, 711–725.

Zhao F, McParland S, Kearney JF, Du L and Berry DP 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genetics Selection Evolution 47, 49.