AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

Assessing the genetic differentiation of Holstein cattle herds in the Leningrad region using F_{st} statistics

Michael G. Smaragdov[1], Andrei A. Kudinov[1,2] and Pekka Uimari[2]

[1]Russian Research Institute of Farm Animals Genetics and Breeding, Saint-Petersburg, Russia

[2]Department of Agricultural Sciences, University of Helsinki, FI-00014 Helsinki, Finland

e-mail: mik7252@yandex.ru

Holstein bulls and semen have been imported to Russia from Western countries since the 1970s. The objective of our study was to examine the effect of this introgression on genetic diversity between various commercial Holstein herds in the Leningrad region. A total of 803 Holstein cows from 13 herds were genotyped using the Illumina BovineSNP50 v.2 array. The pairwise Hudson's estimator of F_{st} values between 13 herds varied from 0.002 to 0.015, which is less than values usually obtained between dairy cattle breeds (> 0.1). The mean of these pairwise F_{st} values revealed differences between herds depending, mainly, on the proportion of common sires shared between the herds. In addition, we investigated the cause of negative F_{st} values. Based on our results, these negative values could be interpreted as an excess of within-herd genetic diversity over the between-herds genetic diversity. Our results show that introgressions of Holstein genes into Russian Black and White cattle of the Leningrad region have created genetic separation between herds similar with those for Jersey cows in USA, Australia and New Zealand.

Key words: Wright's F_{st}, dairy cattle, herds, genetic diversity

Introduction

Minimizing the rate of inbreeding and maintaining genetic diversity are important goals of all successful dairy cattle breeding operations. Pedigree information is very helpful in the management of population genetic diversity. With the new single nucleotide polymorphism (SNP) array technology, massive SNP data can be used for more accurate estimation of genetic diversity both within and between herds·

The Leningrad region is the leading dairy farm region in Russia, with approximately 60 000 cows of holsteinized Black and White cattle. Dutch, Danish, and Swedish Black and White bulls and heifers were imported to Russia during the 1930s. The Black and White breed was officially approved in Russia in 1959. To improve the genetic potential of Black and White cattle, local breeders have used Holstein bulls and semen imported from USA (since 1978) and the Netherlands (since 2002). Because of the importation of Holstein semen, the current commercial Russian Black and White cattle population can be considered Holstein. Understanding the effect on genetic diversity of the introgressions of Holstein genetics into Russian Black and White cattle at the herd level is therefore interesting.

Several tools have been developed to provide estimates of genetic diversity among populations. One of them, Wright's F_{st} fixation index, is a widely used measurement of population diversity. The obtained degree of differentiation within and between populations depends on the power of F_{st} statistics. The methods and issues associated with the estimation of genetic diversity using F_{st} statistics have been covered in many excellent reviews (e.g. Weir and Cockerham 1984, Weir and Hill 2002, Holsinger and Weir 2009, Meirmans and Hedrick 2011, Leinonen et al. 2013). Several F_{st} estimators are currently available, making the choice difficult (Bhatia et al. 2013). The most commonly used estimators are those presented by Weir and Cockerham (1984) and Nei (1986). The first one is sensitive to sample size and the second one consistently overestimates F_{st} (Bhatia et al. 2013). An advantage of the above mentioned F_{st} estimators is that they are easy to calculate and do not require any assumptions concerning the shape of the distribution, other than mean and variance. Another example, the Hudson's estimator (Hudson et al. 1992), is not sensitive to sample size ratios, does not systematically overestimate F_{st}, and is accurate and stable under various ascertainment schemes (Hudson et al. 1992, Bhatia et al. 2013). The Hudson's estimator of F_{st} and the corresponding block-jackknife estimator for the standard error (SE) of F_{st} have been implemented in the EIGENSOFT 6.0.1 software package (Patterson et al. 2006).

Our objective was to examine the effect of the introgressions of Holstein genes into Russian Black and White cattle on genetic diversity measured as F_{st} between commercial dairy herds within the Leningrad region.

AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

Material and methods

Animals and SNP genotyping

Thirteen dairy farms with Russian Holstein cattle from the Leningrad region with the highest milk production (milk yield from 6700 kg to 11300 kg) were selected for our study. From each herd, $44 - 85$ randomly selected cows were genotyped with either the Illumina BovineSNP50 v.2 chip (526 cows born in 2013) or with a custom made Illumina BovineSNPIDB v.3 chip (303 cows born in 2011 – 2013). In total, $8.5 - 15\%$ of the cows were genotyped from each herd. All genotyping was performed in Ireland (Weatherbys Co. UK). PLINK 1.9 software (Pursell 2007) was used for quality control. Imputation of the BovineSNP50 v.2 chip data to BovineSNPIDB v.3 chip data was carried out with the Beagle software (Browning and Browning 2016). In the first quality control step, SNPs with quality scores (gencall in the Illumina's Genome Studio) less than 0.7 were removed. Next, all SNPs with more than 5% of missing genotypes, SNPs that deviate from the Hardy-Weinberg equilibrium at $p < 0.0005$, and SNPs with minor allele frequency less than 0.01 were removed from the data. Only autosomal chromosomes were considered. After imputing, 41 593 SNPs remained for the analysis. The requirement for the sample (genotyped cow) call rate was 0.995, resulting in the rejection of 26 samples, and thus the final number of genotyped cows in the data was 803.

The full data, later called Sample 1, included all 803 cows from the 13 herds. The proportion of cows with a common sire varied from 0% to 35% in the set of 78 pairwise comparisons (hereafter called Pairwise set) made between these 13 herds. In Sample 2, cows with a common sire were removed from the Pairwise set. Comparing the results from Samples 1 and 2 allowed us to determine the sensitivity of Hudson's F_{st} statistics. Moreover, the construction of these two sample sets enabled us to assess the effect that cows with common bulls had on between-herd F_{st} values.

F_{st} values were estimated with EIGENSOFT 6.0.1 software (Patterson et al. 2006). The Hudson's estimate of F_{st} (Bhatia et al. 2013) for a single SNP is:

$$F_{st} = \frac{(\tilde{p}_1 - \tilde{p}_2)^2 - \frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1 - 1} - \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2 - 1}}{\tilde{p}_1(1-\tilde{p}_2) + \tilde{p}_2(1-\tilde{p}_1)} \qquad \text{(Equation 1)},$$

where n_i is the sample size and $\tilde{p}_i$ is the sample allele frequency in herd i (i $\in$ [1, 2]). The final pairwise F_{st} estimates between herds were averaged across 41 593 SNP. The standard errors of the Hudson's F_{st} estimate were calculated using a block-jackknife approach implemented in EIGENSOFT. The standard error of the mean F_{st}^i for herd i was

calculated as $= MSE_i = \frac{1}{n-1} \sqrt{\sum_{j=1}^{n-1} SE_{i,j}^2}$, where n = 13, and i≠j.

Results

Pairwise comparisons between herds

The pairwise F_{st} values calculated between 13 herds for Samples 1 and 2 are shown in Table 1 and the summary of these comparisons in Table 2. Pairwise F_{st} values between the herds in Sample 1 varied from 0.002 to 0.011, and from 0.002 to 0.015 in Sample 2. On average, the F_{st} values were 0.0014 higher in Sample 2 (mean F_{st} = 0.0066) than in Sample 1 (mean F_{st} = 0.0052) after the exclusion of cows with a common sire in both herds (Table 2). As expected, the removal of these cows increased the between-herds genetic differences up to 0.005 points.

The pairwise comparisons of herd 4 with the other 12 herds showed the largest F_{st} values compared to all other herd comparisons (Table 1). We assume that this is caused by the heavy use of bulls from the Netherlands between $2000 - 2007$ in herd 4, while bulls imported mainly from the USA and Canada were with more recent ancestry used in the other 12 herds (Table 2). On the other hand, the pairwise comparisons of herd 3 (in Sample 1) with all other herds revealed the smallest F_{st} values (Table 1). This relates to the high correlation between the proportion of cows with the same sire in different herds and the F_{st} values (Table 2). This is confirmed by the strong negative correlation (-0.64) between the proportion of cows with the same sire and pairwise herd F_{st} values for herd 3 (Table 2).

Herd 7 is characterized by a very small proportion (5 of them had zero proportion) of cows with common sires in other herds (Table 1). Herd 7 therefore, had the smallest difference between the mean pairwise F_{st} values between Samples 1 and 2 (0.0003) among the studied herds (Table 2). On the contrary, herd 8 had a large number

AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

of cows with common sires in other herds (Table 1). However, this led to only a moderate increase in the mean F_{st} value (0.008) after the exclusion of these cows (Table 2). In Sample 1, herd 8 is the second most similar to the other 12 herds after herd 3 (mean F_{st} = 0.0038 for herd 3 vs. mean F_{st} = 0.0039 for herd 8), and it remains the most similar to the other herds after the exclusion of cows with common sires (mean F_{st} = 0.0047, Sample 2, Table 2).

Table 1. Estimates of F_{st} statistics between herds

Herd	1	2	3	4	5	6	7	8	9	10	11	12	13
1		0.005[a]	0.005	0.005	0.006	0.003	0.006	0.004	0.003	0.005	0.005	0.004	0.004
		(0.241)[b]	(0.115)	(0.277)	(0.131)	(0.264)	(0)	(0.152)	(0.197)	(0.13)	(0.074)	(0.176)	(0.254)
2	0.006		0.003	0.011	0.006	0.006	0.006	0.003	0.002	0.004	0.003	0.002	0.008
			(0.233)	(0.1)	(0.154)	(0.05)	(0.09)	(0.214)	(0.316)	(0.093)	(0.32)	(0.233)	(0.132)
3	0.006	0.005		0.009	0.003	0.004	0.005	0.002	0.002	0.003	0.003	0.002	0.005
				(0.134)	(0.242)	(0.21)	(0.022)	(0.277)	(0.3)	(0.322)	(0.129)	(0.27)	(0.1)
4	0.008	0.013	0.013		0.009	0.004	0.01	0.008	0.009	0.01	0.011	0.012	0.006
					(0.262)	(0.353)	(0)	(0.218)	(0.058)	(0.179)	(0.109)	(0.139)	(0.324)
5	0.007	0.008	0.005	0.013		0.005	0.006	0.004	0.004	0.004	0.006	0.004	0.006
						(0.241)	(0)	(0.167)	(0.082)	(0.284)	(0.179)	(0.14)	(0.153)
6	0.005	0.007	0.006	0.008	0.007		0.007	0.004	0.005	0.006	0.005	0.005	0.004
							(0.062)	(0.154)	(0.085)	(0.204)	(0.126)	(0.162)	(0.316)
7	0.006	0.006	0.006	0.01	0.006	0.007		0.004	0.006	0.006	0.007	0.005	0.007
								(0.017)	(0)	(0.019)	(0)	(0.156)	(0.032)
8	0.004	0.005	0.003	0.009	0.004	0.005	0.005		0.003	0.004	0.003	0.004	0.004
									(0.189)	(0.23)	(0.289)	(0.322)	(0.131)
9	0.004	0.005	0.003	0.01	0.004	0.005	0.006	0.002		0.004	0.003	0.004	0.006
										(0.166)	(0.283)	(0.129)	(0.12)
10	0.008	0.005	0.009	0.015	0.007	0.008	0.007	0.005	0.005		0.005	0.004	0.007
											(0.082)	(0.149)	(0.073)
11	0.006	0.005	0.005	0.013	0.007	0.006	0.007	0.004	0.004	0.006		0.005	0.006
												(0.187)	(0.091)
12	0.006	0.005	0.005	0.015	0.006	0.007	0.006	0.006	0.004	0.005	0.006		0.008
													(0.116)
13	0.006	0.009	0.006	0.009	0.005	0.006	0.007	0.004	0.006	0.008	0.008	0.009	

[a] F_{st} values for Sample 1 are above the diagonal and F_{st} values for Sample 2 are below the diagonal. SE of the estimates varied from 0.0002 to 0.0005; [b] Proportions of cows born from common bulls are in parentheses.

In Sample 1, all other 78 pairwise combinations of the mean F_{st} between 13 herds were statistically significant ($p \leq 0.05$), except between herd 1 and herds 2, 6, and 12, between herd 2 and herds 6 and 12, between herds 3 and 8, between herd 5 and herds 10 and 11, between herds 6 and 12, and between herds 10 and 11 (Table 2). The insignificant comparisons relate well with the proportion of common sires shared between the herds (compare insignificant data for Sample 1 and Sample 2).

In Sample 2, statistically insignificant ($p > 0.05$) F_{st} values among 78 pairwise combinations of the mean F_{st} were obtained between herds 1 and 3, between herd 2 and herds 5, 6, 7, 11, and 12, between herd 5 and herds 6, 7, 11, and 12, between herd 6 and herds 7 and 11, between herd 7 and herds 11 and 12, between herds 8 and 9, and between herds 12 and 13 (Table 2). Some of these pairwise F_{st} values were statistically significant in Sample 1 (herd 1 vs. 3; herd 2 vs. 5, 7, and 11; herd 5 vs. 6, 7, and 12; herd 6 vs. 7 and 11; herd 7 vs. 11 and 12; herd 8 vs. 9; and herd 12 vs. 13). Thus, the exclusion of common sires can both increase and decrease the genetic differences between the herds.

AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

Table 2. Mean of F_{st} values between a particular herd and other herds using either Sample 1 (F_{ST}^{1}) or Sample 2 (F_{ST}^{2})

Herd	Russia[a]	USA	Canada	NL	F_{st}^{1b}	F_{st}^{2}	$F_{st}^{2} - F_{st}^{1}$	ρ^{c}
1		54	13	33	0.0047	0.0060	0.0013	-0.49 (0.1)
2	3	69	15	13	0.0049	0.0066	0.0017	-0.67 (0.01)
3	13	56	9	22	0.0038	0.0060	0.0022	-0.64 (0.04)
4				100	0.0087	0.0113	0.0026	0.02 (0.95)
5	16	21	21	42	0.0053	0.0066	0.0013	0.02 (0.99)
6	28	16	3	53	0.0048	0.0064	0.0011	-0.68 (0.01)
7	38	31	24	7	0.0063	0.0066	0.0003*	-0.28 (0.6)
8	14	43	10	33	0.0039	0.0047	0.008	-0.14 (0.6)
9		67	20	13	0.0043	0.0048	0.0005*	-0.8 (0.002)
10	10	52	5	33	0.0052	0.0073	0.0021	-0.33 (0.29)
11	7	70	4	19	0.0052	0.0064	0.0012	-0.58 (0.05)
12	10	64	3	33	0.0049	0.0067	0.0018	-0.37 (0.2)
13	8	11	12	69	0.0059	0.0069	0.001	-0.50 (0.1)

[a] Country of origin of the sires of the genotyped cows, NL = the Netherlands; [b] MSE varied from 0.00006 to 0.0001 depending on herd; [c] Pearson's correlation between the proportion of cows with the same sire and pairwise herd F_{st} values for Sample 1; * $p \leq 0.001$, other differences between F_{st} values had $p \leq 0.0001$

Limits related to F_{st} estimates

In theory, F_{st} value should be zero if a population (here a herd) is compared to itself. However, all the F_{st} values in Sample 1 were negative (Table 3). To find an explanation for the negative values, let us assume that $p_1 = p_2$ and $n_1 = n_2 = n$ in equation (1), then $F_{st} = -1/(n-1)$ (Equation 2). In equation (2), F_{st} depends on the number of animals n. The limit value of F_{st} is 0, when n reaches infinity. Thus, Hudson's estimator is not defined (sample size independent) when calculating F_{st} values for one population.

Table 3. F_{st} values when a herd is compared to itself and when herd 4 is compared to herd 2 with different proportions of cows from herd 4

Herd	F_{st}	SE[a]	Herds	%[b]	F_{st}	SE
1 vs. 1	-0.0087	0.0021	2 vs. 2	0	-0.006	0.0007
2 vs. 2	-0.0058	0.0007	2 vs. 4a	17.6	-0.004	0.0007
3 vs. 3	-0.0067	0.0005	2 vs. 4b	35	-0.001	0.0004
4 vs. 4	-0.0111	0.001	2 vs. 4c	53	0.003	0.0002
5 vs. 5	-0.0086	0.0065	2 vs. 4d	73	0.01	0.0003
6 vs. 6	-0.0092	0.0056	2 vs. 4e	88	0.017	0.0005
7 vs. 7	-0.0084	0.0052				
8 vs. 8	-0.0087	0.0037				
9 vs. 9	-0.0065	0.0012				
10 vs. 10	-0.0113	0.0008				
11 vs. 11	-0.0076	0.0011				
12 vs. 12	-0.0078	0.0021				
13 vs. 13	-0.0075	0.007				

[a]Standard error; [b]Percent of herd 4 cows in herd 2

To further investigate the cause of the negative F_{st} values, we carried out a hypothetic simulation where two herds were compared with differing proportions of common cows. We selected herd 2 as one of the herds because it had the largest number of cows (85 cows). Herd 4 was selected as the other herd because it differed genetically the most from the other herds (Table 2). First, we calculated the F_{st} value of herd 2 with itself (Table 3).

AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

Then, step by step, we replaced 17.6%, 35.3%, 53.0%, 71.0%, and up to 88.2% of the cows in herd 2 with the cows from herd 4, which resulted in five artificial herds (4a – 4e in Table 3). Each replacement was performed with the same sample of 15 cows, i.e. 17.6% corresponds to one set of 15 cows, 35.3% corresponds to two sets of the same 15 cows etc. We should remember that Hudson's estimator can be written as $F_{st} = 1-(H_w / H_b)$, where H_w is the mean number of differences within populations, and H_b is the mean number of differences between populations (Bhatia et al. 2013). Firstly, if a population (here a herd) is compared to itself then $F_{st} = 1-(H_w / 0)$, that F_{st} is indefinite. Secondly, mathematically the result of the simulation can be proposed through equation 3.

$$F_{st} = 1 - \frac{(\Delta H_2 + \Delta H_4)}{H_b} \qquad \text{(Equation 3)}$$

, where ΔH_2 is the herd 2 residual genetic difference in artificial herds 4a – 4e, ΔH_4 is the herd 4 added genetic differences in artificial herds 4a – 4e and H_b is between herd 2 and artificial herds 4a – 4e genetic differences. Based on the results in Table 3, we observed a threshold, below which Hudson's estimator calculates negative F_{st} values. In our case, the threshold was reached when both herds shared approximately 50% of the same cows. Thus, Hudson's estimator before this threshold estimates within genetic difference larger than between herds genetic difference and F_{st} values become negative. In other words, the estimation of F_{st} values depend on both the within herds differences, and between herds differences.

Thus, if the average difference within a population is greater than the average difference between populations, the F_{st} values will be negative. As soon as H_w is equal to H_b, the F_{st} value becomes zero, with a threshold reached when H_w is equal to H_b. In our case, all pairwise F_{st} values are positive. This means that within-herd genetic differences were smaller than between the herds.

Discussion

Numerous studies have been carried out to estimate the genetic diversity between cattle breeds and populations using F_{st} (e.g. McKay et al. 2008, Gautier et al. 2010, Wilkinson et al. 2011, Edea et al. 2013, Rothammer et al. 2013, Cauas–Alvarez et al. 2015, Kelleher et al. 2016). These studies have observed varying F_{st} values, from very small between different populations of the same breed, e.g. the F_{st} value between Guernseys bulls and cows from the UK and the islands was 0.006 (Cooper et al. 2016), to relatively large values between dairy and beef breeds, e.g. the F_{st} value between Braunvieh and Galloway breeds was 0.16 (Rothammer et al. 2013). Jersey cattle is another example of the genetic differences between different populations of the same breed (Howard et al. 2015). The mean F_{st} value for Australian Jersey cows versus US Jersey cows was 0.008 and the average F_{st} value for US Jersey cows versus New Zealand jersey cows, and Australian Jersey cows versus New Zealand Jersey cows were 0.029 and 0.009, respectively.

In our study, between-herd F_{st} values (on average 0.006) were lower than values between different populations of Jersey cows (Howard et al. 2015) and considerably lower than values between different breeds. Thus, we can conclude that herds in the Leningrad region are less diverse than various populations of the same breed globally. This is a consequence of using the same bulls or bulls with the same origin in different herds.

Negative F_{st} values are obtained occasionally. Certain authors do not report such values. However, for example Wilkinson et al. (2011) claimed that the F_{st} estimator by Weir and Cockerham (1984), can get negative values if alleles, drawn at random from within a population, are less similar to one another than those drawn from different populations. Negative F_{st} values were reported in our study, and based on Equation 2 and a simple simulation (Table 3) we show that there is a limit value depending on proportion of shared genetics after which negative of F_{st} values can be expected.

Conclusions

F_{st}-statistics was carried out to reveal genetic diversity of Holstein cows between herds in the Leningrad region of Russia. The observed between-herd F_{st} values ranged from 0.002 to 0.015. These estimates are smaller than those obtained between various populations of the same breed, and far smaller than those obtained between breeds. In addition, the absence of negative F_{st} values indicates that within-herd genetic differences are smaller than between herds. The exclusion of cows with common sires slightly increased the pairwise F_{st} values, but its effect was small. We believe that the methodology can be used to monitor the between-herd genetic diversity of cattle populations within and across countries. This approach can also be applied to other farm animals and is useful in breed conservation.

AGRICULTURAL AND FOOD SCIENCE

M.G. Smaragdov et al. (2018) 27: 96–101

Acknowledgments

References

Bhatia, G., Patterson, N., Sankararaman, S. & Price, A.L. 2013. Estimating and interpreting F_{st}: The impact of rare variants. Genome Research 23: 1514–1521. https://doi.org/10.1101/gr.154831.113

Browning, B.L. & Browning, S.R. 2016. Genotype imputation with millions of reference samples. American Journal of Human Genetics 98: 116–126. https://doi.org/10.1016/j.ajhg.2015.11.020

Cauas–Alvarez, J.J., Gonzalez-Rodriguez, A., Munilla, S., Varona, L., Díaz, C., Baro, J.A., Altarriba, J., Molina, A. & Piedrafita, J. 2015. Genetic diversity and divergence among Spanish beef breeds assessed by a bovine high-density SNP chip. Journal of Animal Science 93: 5164–5174. https://doi.org/10.2527/jas.2015-9271

Cooper, T.A., Eaglen, S.A.E., Wiggans, G.R., Jenko, J., Huson, H.J., Morrice, D.R., Bichard, M., de L. Luff, W.G. & Woolliams, J.A. 2016. Genomic evaluation, breed identification, and population structure of Guernsey cattle in North America, Great Britain, and the Isle of Guernsey. Journal of Dairy Science 99: 5508–5515. https://doi.org/10.3168/jds.2015-10445

Edea, Z., Dadi, H., Kim, S.W., Dessie, T., Lee, T., Kim, H., Kim, J.-J. & Kim, K.-S. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Frontiers in Genetics 4: 1–9. https://doi.org/10.3389/fgene.2013.00035

Gautier, M., Laloe, D. & Moazami-Goudarzi, K. 2010. Insight into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS One 5: e13038. https://doi.org/10.1371/journal.pone.0013038

Holsinger, K.E. & Weir, B.S. 2009. Genetics in geographically structured populations: defining, estimating and interpreting F_{st}. Nature Reviews Genetics 10: 639–650. https://doi.org/10.1038/nrg2611

Howard, J.T., Maltecca, C., Haile-Mariam, M., Hayes, B.J. & Pryce, J.E. 2015. Characterizing homozygosity across United States, New Zealand and Australian Jersey cow and bull populations. BMC Genomics 16:187. https://doi.org/10.1186/s12864-015-1352-4

Hudson, R.R., Slatkin, M. & Maddison, W.P. 1992. Estimation of level of gene flow from DNA sequence data. Genetics 132: 583–589.

Kelleher, M.M., Berry, D.P., Kearney, J.F., McParland, S., Buckley, F. & Purfield, D.C. 2016. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. Animal 2: 1–9. https://doi.org/10.2527/jas2016.94supplement4103x

Leinonen, T., Scott McCairns, R.J., O' Hara, R.B. & Merilä, J. 2013. Qst – F_{st} comparisons: evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics 14: 179–190. https://doi.org/10.1038/nrg3395

McKay, S.D., Schnabel, R.D., Murdoch, B.M., Matukumalli, L.K., Aerts, J., Coppieters, W., Crews, D., Neto, E. D., Gill, C.A., Gao, C., Mannen, H., Wang, Z., Van Tassell, C.P., Williams, J.L., Taylor, J.F. & Moore, S.S. 2008. An assesment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics 9:37. https://doi.org/10.1186/1471-2156-9-37

Meirmans, P.G. & Hedrick, P.W. 2011. Assessing population structure: F_{st} and related measures. Molecular Ecology Resources 11: 5–18. https://doi.org/10.1111/j.1755-0998.2010.02927.x

Nei, M. 1986. Defenition and estimation of fixation indices. Evolution 40: 643–645. https://doi.org/10.1111/j.1558-5646.1986.tb00516.x

Patterson, N., Price, A.L. & Reich, D. 2006. Population structure and eigenanalysis. PLoS Genetics 2:e190. https://doi.org/10.1371/journal.pgen.0020190

Pursell, S. 2007. PLINK: a tool set for whole genome association and population based linkage analyses. American Journal of Human Genetics 81: 559–575. https://doi.org/10.1086/519795

Rothammer, S., Seichter, D., Forster, M. & Medugorac, I. 2013. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. BMC Genomics 14:908. https://doi.org/10.1186/1471-2164-14-908

Weir, B.S. & Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358–1370.

Weir, B.S. & Hill, W.G. 2002. Estimating F-statistics. Annual Review of Genetics 36: 721–750. https://doi.org/10.1146/annurev.genet.36.050802.093940

Wilkinson, S., Wiener, P., Archibald, A.L., Law, A., Schnabel, R.D., McKay, S.D., Taylor, J.F. & Ogden, R. 2011. Evaluation of approaches for identifying population informative markers from high density SNP chips. BMC Genetics 12:45. https://doi.org/10.1186/1471-2156-12-45