Genomic data reveals large similarities among Canadian and French maternal pig lines

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Animal Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>CJAS-2017-0103.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>27-Jan-2018</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Boré, Raphael; Institut de la Filière Porcine Brito, Luiz; University of Guelph, Department of Animal Biosciences Jafarikia, Mohsen; Canadian Centre for Swine Improvement, ; University of Guelph, Animal and Poultry Science Bouquet, Alban; Institut de la Filière Porcine maignel, Laurence; Canadian Center for Swine Improvement, Sullivan, Brian; Canadian Centre for Swine Improvement Schenkel, Flavio; University of Guelph, Animal and Poultry Science</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Genetic diversity, linkage disequilibrium, pig, persistence of gametic phase, genomic selection</td>
</tr>
</tbody>
</table>
Genomic data reveals large similarities among Canadian and French maternal pig lines

Raphael Boré¹, Luiz F. Brito², Mohsen Jafarikia²,³, Alban Bouquet¹, Laurence Maignel³, Brian Sullivan³, Flávio S. Schenkel²

¹Institut de la Filière Porcine, La Motte au Vicomte, BP 35104, Le Rheu, France.
²Centre for Genetic Improvement of Livestock, University of Guelph, 50 Stone Road East, Guelph, ON, Canada.
³Canadian Centre of Swine Improvement, Central Experimental Farm, Building #75, 960 Carling Avenue, Ottawa, ON, Canada.

Corresponding author: Alban Bouquet. E-mail: alban.bouquet@ifip.asso.fr

Postal address: IFIP Institut du Porc - Pôle Génétique, La motte au Vicomte - BP 35104 – 35651, Le Rheu, France.

ABSTRACT

Combining reference populations from different countries and breeds could be an affordable way to enlarge the size of the reference populations for genomic prediction of breeding values. Therefore, the main objectives of this study were to assess the genetic diversity within and between two Canadian and French pig breeds (Landrace and Yorkshire) and the genomic relatedness among populations in order to evaluate the feasibility of an across-country reference population for pig genomic selection. A total of 14,756 pigs were genotyped on two SNP chip panels (~65K SNPs). A principal component analysis clearly discriminated Landrace and Yorkshire breeds, and also, but to a lesser extent, the Canadian
and French purebred pigs of each breed. Linkage disequilibrium (LD) between adjacent SNPs was similar within Yorkshire populations. However, levels of LD were slightly different for Landrace populations. The consistency of gametic phase was very high between Yorkshire populations (0.96 at 0.05 Mb) and high for Landrace (0.88 at 0.05 Mb). Based on consistency of gametic phase, Canadian and French pig maternal lines are genetically close to each other. These results are promising, as they indicate that the accuracy of estimated genomic breeding values may increase by combining reference populations from the two countries.

**RÉSUMÉ**

Combiner des populations de référence provenant de différents pays ou de différentes races peut être un moyen abordable d’agrandissement de la taille de la population de référence pour les prédictions génomiques. Par conséquent, les principaux objectifs de cette étude sont d’évaluer la diversité génomique entre et au sein des deux races porcines françaises et canadiennes (Landrace et Yorkshire) ainsi que l’apparentement des populations afin d’évaluer la faisabilité de combiner les populations de référence des deux pays en une population de référence commune pour la sélection génomique porcine. Un total de 14,756 animaux ont été génotypés sur deux puces à ADN commerciales (~ 65K SNPs). L’analyse en composantes principales discrimine clairement les deux races Landrace et Yorkshire, et dans une moindre mesure les populations de chacun des deux pays. Le déséquilibre de liaison (LD) entre les SNPs adjacents est similaire dans les populations Yorkshire. En revanche, les niveaux de LD sont légèrement différents pour les populations Landrace. La persistance de phase gamétique entre les populations Yorkshire est très élevée (0.96 à une distance de 0.05 Mb) et élevée entre les populations Landrace (0.88 à une distance de 0.05 Mb). Ces persistances de phase gamétique élevées suggèrent que les lignées maternelles canadiennes et françaises sont génétiquement proches les unes des autres. Ces résultats sont prometteurs et indiquent que la
 précision des valeurs génomiques estimées pourrait augmenter avec une population de référence commune entre le Canada et la France.

**Key words:** Genetic diversity, linkage disequilibrium, pig, persistence of gametic phase, genomic selection

**INTRODUCTION**

In genomic selection schemes, the size of the reference population has a major impact on the accuracy of Genomic Estimated Breeding Values (GEBVs) for selection candidates and, hence, on achieved genetic progress (Hayes et al., 2003). Cost effective alternatives to enlarge reference populations are sharing genotyping data between countries and combining reference populations of different breeds (VanRaden et al., 2009). However, the feasibility of this strategy depends on the degree of genetic relatedness among populations. In Holstein cattle, consortia between countries greatly improved achieved reliabilities of GEBVs when using an international reference population (Lund et al., 2011; Mäntysaari et al., 2014). Gains in prediction accuracy were lower when populations were less genetically related (Brøndum et al., 2011; Carillier et al., 2013) and even null when reference populations of distant breeds were merged together (Olson et al., 2012).

In pig breeding programs, genomic selection has played a major role on increasing genetic gains for a variety of economically important traits. However, genotyping price is still a constraint for a wider use of this technology compared to other livestock species (Tribout et al., 2012; Abell et al., 2014). Furthermore, although similar breeds are genetically selected around the world, breeding nuclei are of small to moderate size and fragmented between companies and countries, making it difficult to gather large numbers of genotyped animals.
In the last 15 years, French and Canadian pig breeding companies have regularly exchanged genetic material of maternal breeds (i.e. Landrace and Yorkshire). The main goal of this collaboration was primarily to increase the genetic diversity in these populations that were selected roughly for the same breeding objectives. The exchange of genetic resources between both countries can now be potentially used as an advantage for implementing genomic selection more efficiently by sharing genotypes and phenotypes (joint reference populations). Therefore, the objective of this study was to evaluate the persistence of gametic phases of Landrace and Yorkshire breeds selected in Canada and France using a large genomic dataset to assess the feasibility of multi-breed and/or across-country genomic evaluation.

MATERIAL AND METHODS

Ethics statement

The animals included in this study were managed in accordance with the “Code of practice for the care and handling of pigs” in Canada (National Farm Animal Care Council, 2014) and the “Code of practice for hygiene and application of HACCP principles in pig production” in France (Journal Officiel de la République Française, 2012). All the samples were collected from breeding farms and the animal owners agreed to be involved in the project through their respective producers’ associations. Samples were collected by well-trained staff following industry best practices.

Animals

From a total of 14,756 pigs included in this study, 4,055 were Canadian Landrace (LA_CAN), 2,943 French Landrace (LA_FRA), 5,528 Canadian Yorkshire (YO_CAN) and 2,230 French Yorkshire (YO_FRA). French animals were sampled from purebred herds
affiliated to two French breeding companies participating in the national breeding schemes, but also from an experimental herd rearing purebred Yorkshire pigs. Canadian animals were sampled from three major genetic companies located in Ontario, Quebec and Saskatchewan.

**SNP genotyping and data editing**

Animals were genotyped using either the Illumina Porcine SNP60 BeadChip (Illumina Inc. San Diego, CA) or the GeneSeek SNP70 BeadChip (Neogen, Lincoln, NE) containing 64,232 and 68,516 Single Nucleotide Polymorphisms (SNPs), respectively. For the following analyses, the 40,753 SNPs shared by these two SNP panels were considered. A genotyping quality control (QC) was performed using the snp1101 software Version 1.0 (Sargolzaei, 2014) to remove SNPs and/or animals that could bias the estimation of population parameters. SNPs with Minor Allele frequency (MAF) lower than 1% were excluded prior to estimation of LD to prevent low MAF loci inflating LD estimates. MAF was defined as the allele frequency of the less common allele. SNPs were also removed if the call rate was lower than 90%, if they displayed an excess of heterozygosity greater than 15% compared to proportions expected under Hardy-Weinberg equilibrium (Wiggans et al., 2009), were located on sexual chromosomes or had unknown or duplicated positions. The SNP physical positions were obtained from the pig genome assembly 10.2 (Sscrofa10.2), (Martien Groenen, Wageningen University, data downloaded from the AnimalGenome.org data repository (http://www.animalgenome.org/repository/pig/). Individual animals with a genotyping call rate lower than 90% were also excluded. This QC step led to exclusion of 32 LA_CAN, 11 LA_FRA, 31 YO_CAN and 10 YO_FRA animals. The number of SNPs excluded during the QC for each analysis is shown in Table 1, as well as the remaining numbers of SNPs and genotypes for each population.
Genetic diversity metrics

Average minor allele frequency and SNP distribution along the genome

Minor allele frequencies were calculated separately for each country for Landrace and Yorkshire. The number of SNPs remaining on each chromosome as well as the density in markers was evaluated for each breed and country.

Gene origins and structures between populations

To quantify the amount of genetic material exchanged between Canada and France for Landrace and Yorkshire breeds, a pedigree analysis was carried out using the PEDIG software (Boichard, 2002). Analyses were performed considering all recorded pedigrees available in each country. The total genetic contribution from Canadian, French and other foreign founders was assessed in the Canadian and French genotyped individuals. In the French Yorkshire population, 557 experimental animals whose pedigree was not available on the maternal side were excluded from the analysis. Then, a principal component analysis (PCA) was carried out on SNP genotype data to examine population sub-structure and genetic relationships between breeds. Principal components were calculated from the genomic relationship matrix \( (G) \) using snp1101 software (Sargolzaei, 2014) according to VanRaden method (VanRaden 2008).

Extent of linkage disequilibrium

The extent of LD was calculated using the \( r^2 \) measure (Hill and Robertson 1968), implemented in snp1101 software. The \( r^2 \) measure is the squared correlation between alleles at two loci and was calculated as

\[
r^2 = \frac{D^2}{f(A)f(a)f(B)f(b)},
\]

where \( D = f(AB) - f(A)f(B) \) and \( f(AB), f(A), f(a), f(B) \) and \( f(b) \) are the frequencies of haplotype AB and alleles A, a, B and b.
respectively. Haplotypes were not phased and the Hill and Robertson D was replaced by a D estimate as suggested by Lynch and Walsh (1998):

\[ D = \frac{n}{n-1} \left( \frac{4n_{AABB} + 2(n_{AABB} + n_{AaBB}) + n_{AaBb}}{2n} - 2 \times f(A) \times f(B) \right), \]

where \( n \) is the total number of individuals and \( n_{AABB}, n_{AaBB}, n_{AaBb} \) and \( n_{AaBb} \) are the numbers of individuals for each genotype combination. LD value was calculated for each pair of syntenic SNPs to determine firstly the LD between adjacent SNPs and secondly the LD over physical distances. Then average LD was estimated for groups of SNPs based on pairwise marker distances defined every 0.01 Mb from 0 to 1 Mb.

**Consistency of gametic phase**

In each breed, the signed \( r \)-value was derived as the square root of \( r^2 \) for all pairs of markers assigning the appropriate sign based on D value. The consistency of gametic phase value was calculated as the Pearson correlation between the signed \( r \)-values estimated between two populations for inter-marker distance bins defined every 0.01 Mb. The breakdown in the consistency of gametic phase between breeds and countries was plotted over physical distances.

**RESULTS**

**Allele frequencies and SNP distribution across the genome**

The distribution of MAF was similar between populations (Figure S1). The French populations displayed a slightly larger proportion of SNPs with low MAF (< 5%) than the Canadian populations. SNPs considered in the analyses covered the whole range of MAF values. This was expected since both Yorkshire and Landrace breeds were included in the design of the SNP chip panels used in this study (Ramos et al. 2009; http://genomics.neogen.com/pdf/slicks/ag284_ggp_porcline.pdf). Correlations estimated
between allele frequencies observed within French and Canadian populations were 0.4 and 0.8 in Landrace and Yorkshire breeds, respectively.

The SNP density was similar across breeds but varied between chromosomes. Statistics about distribution of SNPs along the genome were reported for each breed in Tables S1 and S2. The largest gaps between adjacent markers for Landrace breed were observed on SSC1, SSC2, SSC3, SSC8 and SSC15, with length of 2.147 Mb, 2.628 Mb, 2.818 Mb, 2.349 Mb and 2.845 Mb, respectively. For the Yorkshire breed the largest gaps were observed on SSC1 (2.274 Mb), SSC3 (2.818 Mb) and SSC15 (2.845 Mb).

Gene origins and principal component analysis

The genetic relatedness between Canadian and French populations varied depending on the breed. The gene flow between France and Canada was the largest for the Yorkshire breed (Figure 1a). Indeed, YO_CAN genotyped animals had on average about 67% of French Yorkshire genes. On the other hand, based on the data analyzed in this study, no Yorkshire genetic material was imported into France from Canada (Figure 1b). For the Landrace breed, the amount of gene flow was comparable, with about 10-12% of genotyped animals having genetic contribution from founders born in the other country.

Individuals from the four different populations were plotted according to the first two principal components in Figure 2, and the first and the third principal components in Figure S2. The first component explained 23.2% of the genomic variation observed and clearly discriminated Yorkshire and Landrace breeds. The second principal component explained 11.4% of the variation and revealed a clear clustering of Canadian and French Landrace. The Canadian Landrace individuals grouped with the French Landrace were actually French boars exported and used in Canada. The third component allowed to discriminate the Canadian and French Yorkshire populations and explained 8.6% of the variation. Similarly, the Canadian
Yorkshire individuals grouped with the French Yorkshire were actually French purebred boars exported and used in Canada. It could be argued that animals born in France, but that were exported to breeding farms in Canada, should still be considered as part of the French pig population. However, as these animals are contributing to the Canadian genetic pool and leaving descendants in Canada we have considered them as part of the Canadian pig population.

**Extent of linkage disequilibrium**

As shown in Table 2, the average $r^2$ between adjacent SNP was very similar among breeds (0.39 for Canadian and French Yorkshire, and, 0.37 and 0.35 for Canadian and French Landrace, respectively). The detailed LD statistics per chromosome were reported in Tables S1 and S2. Small differences in LD between adjacent SNPs were observed for the same chromosome in different breeds. This was expected, given that there was not much variation in the average distance between adjacent SNP pairs (~ 0.07 Mb).

The proportion of SNP pairs with $r^2 > 0.2$ and $r^2 > 0.3$ was around 54% and 45%, respectively. French Landrace had the lowest proportion for the two thresholds of $r^2$ (51% and 42%), while Yorkshire population (French and Canadian) had the highest ones (56% and 47%). The decay of LD over distance between SNPs (Figure 3) was similar between breeds, with a slightly lower LD level in the Landrace breed both on short and long distance.

**Consistency of gametic phase**

The consistency of gametic phase over distance was estimated between the four populations (Figure 4). Within breed, the highest consistency of phase was found between Canadian and French Yorkshire (0.96 at 0.05 Mb). The gametic phase consistency between Landrace population from both countries was lower than between Yorkshire lines (0.88 at
0.05 Mb), but it was still strong. The consistency of gametic phase between Landrace and Yorkshire breeds was much lower, regardless the country considered (around 0.77 at 0.05 Mb).

**DISCUSSION**

It was known that Yorkshire and Landrace populations bred in France and Canada were genetically connected to some degree due to the exchange of breeding boars in the early 2000s to increase genetic diversity in both populations (CCSI, 2001). Results of the present study, using SNP genotype data, support that these populations are relatively close genetically, especially in the case of Yorkshire populations.

In this study, a large dataset with genomic information was analyzed and enabled a fine description of genetic diversity and genetic relatedness between populations. The gaps observed in the physical map could have been shortened by imputing all missing marker information on each one of the panels to gain extra SNP information (Cleveland and Hickey, 2013). However, we considered that information provided by over 35,000 SNPs was already sufficient to accurately describe the genetic diversity and genetic relatedness of these populations. Denser SNP panels would enable to investigate levels of linkage disequilibrium and consistency of gametic phase at shorter distances between markers, however, the SNP panel used in this study already covered very short distances and presented desirable levels of linkage disequilibrium to perform accurate genomic predictions.

Linkage disequilibrium was high (> 0.35) at short distances and similar to estimates reported in other selected pig populations (Uimari and Tapio 2011; Badke et al., 2012; Veroneze et al., 2013; Grossi et al., 2017). This high level of LD was important for the successful implementation of within country, within breed genomic selection in these populations. The LD decay over physical distance between markers was also similar among
populations. This suggests that the change in effective population size of these four populations might have been similar, at least in the recent past (Hayes et al., 2003).

To our knowledge, estimates of consistency of gametic phase are the first reported between two pig populations of the same breed in different countries. Consistency of gametic phase had already been estimated between Landrace and Yorkshire breeds within the same country. Our results were slightly lower than those reported by Badke et al., (2012) on US pig breeds and very similar to Grossi et al. (2017) for the same Canadian breeds. Veroneze et al. (2014) evaluated the consistency of gametic phase across five pure pig lines, one F1 cross and two commercial finishing crosses, and observed that the crossbred animals showed a high consistency of gametic phase with their parental lines. In addition, the consistency of gametic phase across purebred lines varied considerably between the different line comparisons; however, correlations were above 0.8 for all line comparisons when marker distances were smaller than 50 kb. These results are in agreement with our findings for some population scenarios, while for other scenarios they were slightly lower.

Within a breed, the genomic similarity between two populations depends on the amount of genetic material exchanged between countries, but also on the genomic similarity of founders in each subpopulation. Both aspects might explain the difference in terms of genomic similarity observed between the Yorkshire and Landrace populations. The analysis of founder gene origins as well as the estimates of consistency of gametic phase indicated that Yorkshire populations in Canada and France were more genetically related than Landrace populations. This result was consistent with expectations given the genetic material exchanges between these populations. The gene flow between Canadian and French Yorkshire populations was important during the last two decades. In 2000, a French purebred nucleus was created in Canada and was regularly provided with new breeding animals. Meanwhile, a few French Yorkshire breeding animals were also used in the Canadian Yorkshire nucleus.
herds. As a result, the consistency of gametic phase was very strong between Yorkshire populations in Canada and France. This level of consistency of gametic phase is comparable to that estimated between Holstein populations in the Netherlands and Australia (De Roos et al., 2009) or China and the Nordic Countries (Zhou et al. 2013). De Roos et al. (2009) showed by simulation that such high estimates of consistency of gametic phase should enable improving the accuracy of genomic predictions, even for low heritability traits. Zhou et al. (2013) using real data estimated an improvement of genomic prediction reliabilities for milk, fat and protein yields for Chinese bulls (from 0.16 to 0.45) and cows (from 0.14 to 0.21) using a single reference population including Nordic Holstein breeding animals. This is also in agreement with the gains in GEBV accuracy reported by Lund et al. (2011) when using a joint Holstein reference population within the Eurogenomics consortium (http://www.eurogenomics.com/) for genomic evaluations. Hence, our results suggest that a combined reference population for the Yorkshire breed between France and Canada may increase the accuracy of GEBVs in both countries.

In the Landrace breed, gene flow was moderate between Canada and France. About 10% of genes of genotyped individuals in France originated from Canada and vice-versa. Furthermore, the PCA analysis indicated that the genetic distance between Canadian and French Landrace was not much bigger than between Canadian and French Yorkshire. Hence, the consistency of gametic phase estimated between Canadian and French Landrace was relatively high, though lower than values estimated between the Yorkshire populations. With similar levels of consistency of gametic phase (0.88 at 0.05 Mb distance), Carillier et al. (2013) observed that combining reference populations of two dairy goat breeds could increase GEBV accuracy. However, the genomic relationships between reference populations in different countries also affects the accuracy of genomic predictions achieved by merging different reference populations. For instance, Fangmann et al. (2015) reported no gain in
GEBV prediction accuracy on the trait “number of piglets born alive” when performing genomic evaluations with a reference population that combined German, Austrian and Swiss Large White datasets. They concluded that genetic connectedness between these Large White subpopulations was too low. The size of their reference and validation sets was also small (Fangmann et al., 2015).

A common reference population across Landrace and Yorkshire populations is unlikely to increase the accuracy of genomic selection because those breeds are genetically distant, as supported by the moderate consistency of gametic phase estimated at a short as well as a long distance. Hidalgo et al. (2015) had already reported limited advantages in terms of gains in GEBVs accuracy for litter size traits when Landrace and Yorkshire reference populations were combined.

Results of the present study suggest that performing genomic evaluations based on across-country reference populations might increase the reliability of genomic predictions for Landrace and Yorkshire breeds. However, to implement such genomic evaluations, a critical issue may be related to the construction of the genomic relationship matrix used for predicting genomic breeding values. This is particularly the case for the Landrace breed in which the correlation between allele frequencies observed in populations from both countries was low (0.40). Another evidence for this is presented on the PCA plots, in which Canadian and French populations clustered as different populations. Strategies to account for stratification in the reference population have already been proposed, but with low impact on GEBV accuracy (Makgahlela et al., 2013). Constructing the genomic relationship matrix using the meta-founder approach as recently proposed by Legarra et al. (2015) may be an alternative to avoid this limitation. Indeed, their approach explicitly accounts for the genomic differentiation between populations and thus considers that subpopulations may have different MAF. Lourenco et al. (2016) investigated different breed-specific G matrices for evaluation
of crossbreds when genotypes were available for 2 purebred breeds and their crosses and observed that in the evaluation of crossbreds, accounting for breed-specific allele frequencies promoted changes in $G$ that were not influential enough to improve accuracy of GEBV. Therefore, the next step of our research will be to investigate strategies to combine across-country (e.g., Canada and France) reference populations for genomic predictions.

**CONCLUSIONS**

Evaluation of genetic diversity and genetic relatedness within and between Landrace and Yorkshire pig populations bred in Canada and France was carried out using a large number of genotyped pigs. Canadian and French populations were shown to be genetically related and present great similarities in their genomic composition, in particular a high consistency of gametic phase along the genome, supporting the sharing genotypic and phenotypic data for genomic evaluations in both countries. These results may have practical implications to the pig industry in both countries, as a joint reference population may increase the accuracies of GEBVs for traditional and novel economically important traits. Strategies to combine reference populations will be investigated next.

**ACKNOWLEDGEMENTS**

The authors would like to express their gratitude to all Canadian and French pig breeders, scientists, funding agencies and other organizations who contributed their valuable time, financial support and provided data for conducting this study. On the French side, authors would like to thank French breeding companies, Axiom and Nucleus, as well as Canadian breeding companies Fast Genetics, Olymel, Alliance Genetics Canada and other users of the Canadian Swine Improvement Program for providing genomic data. Dr. Juliette Riquet (INRA UMR GenPhySE, Toulouse, France), Dr. Claire Rogel-Gaillard and Dr. Jordi
Estelle (INRA UMR GABI, Jouy-en-Josas, France) are also acknowledged for kindly providing some extra genomic data obtained through research projects partly funded by the French National Research Agency (ANR), namely the projects DELISUS (ANR-07-GANI-0001), IMMOPIG (ANR-06-GANI-0008), SUSFLORA (ANR-2011-GENOM-BTV-016) and SWAN (ANR-08-GENM-040).

REFERENCES


Sargolzaei M. 2014. snp1101 User’s Guide. Version 1.0, University of Guelph (Contact: msargol@uoguelph.ca).


### Table 1. Quality control (QC) description.

<table>
<thead>
<tr>
<th>Breed/Country</th>
<th>Number of valid genotypes</th>
<th>Unk/Dup</th>
<th>MAF&lt;0.01</th>
<th>CR&lt;0.90</th>
<th>Het</th>
<th>Remaining SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA_CAN</td>
<td>4,023</td>
<td>2,646</td>
<td>1,309</td>
<td>660</td>
<td>7</td>
<td>36,154</td>
</tr>
<tr>
<td>LA_FRA</td>
<td>2,932</td>
<td>2,646</td>
<td>1,046</td>
<td>285</td>
<td>2</td>
<td>36,786</td>
</tr>
<tr>
<td>YO_CAN</td>
<td>5,497</td>
<td>2,646</td>
<td>1,292</td>
<td>655</td>
<td>4</td>
<td>36,181</td>
</tr>
<tr>
<td>YO_FRA</td>
<td>2,220</td>
<td>2,646</td>
<td>2,139</td>
<td>243</td>
<td>7</td>
<td>35,747</td>
</tr>
<tr>
<td>LA_all</td>
<td>6,955</td>
<td>2,646</td>
<td>564</td>
<td>609</td>
<td>3</td>
<td>36,943</td>
</tr>
<tr>
<td>YO_all</td>
<td>7,717</td>
<td>2,646</td>
<td>1,071</td>
<td>651</td>
<td>4</td>
<td>36,408</td>
</tr>
</tbody>
</table>

LA_CAN, LA_FRA, YO_CAN, YO_FRA, LA_all, and YO_all represents Canadian Landrace, French Landrace, Canadian Yorkshire, French Yorkshire, Landrace from both countries, and Yorkshire from both countries, respectively.

Unk/Dup

MAF = Minor Allele Frequency.

CR = Call rate.

Het = Excess of heterozygosity (> 0.15).

SNP = Single Nucleotide Polymorphism.

Note: Number of common SNPs between 60K and 70K SNP chips = 40,753.
Table 2. Mean linkage disequilibrium ($r^2$) ± standard deviation (SD), and mean and maximum distance in Mb between adjacent SNP pairs per breed within country and overall within each breed.

<table>
<thead>
<tr>
<th>Breed/country</th>
<th>N pairs</th>
<th>$r^2$±SD</th>
<th>Mean dist (Mb)</th>
<th>Max dist (Mb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA_CAN</td>
<td>36,136</td>
<td>0.373 ± 0.354</td>
<td>0.067 ± 0.096</td>
<td>2.845</td>
</tr>
<tr>
<td>LA_FRA</td>
<td>36,768</td>
<td>0.350 ± 0.345</td>
<td>0.066 ± 0.095</td>
<td>2.845</td>
</tr>
<tr>
<td>YO_CAN</td>
<td>36,163</td>
<td>0.392 ± 0.363</td>
<td>0.067 ± 0.096</td>
<td>2.845</td>
</tr>
<tr>
<td>YO_FRA</td>
<td>35,729</td>
<td>0.390 ± 0.355</td>
<td>0.068 ± 0.099</td>
<td>3.904</td>
</tr>
<tr>
<td>LA_all</td>
<td>36,925</td>
<td>0.365 ± 0.353</td>
<td>0.066 ± 0.094</td>
<td>2.845</td>
</tr>
<tr>
<td>YO_all</td>
<td>36,390</td>
<td>0.386 ± 0.362</td>
<td>0.068 ± 0.095</td>
<td>2.845</td>
</tr>
</tbody>
</table>

*LA_CAN, LA_FRA, YO_CAN, YO_FRA, LA_all, and YO_all represents Canadian Landrace, French Landrace, Canadian Yorkshire, French Yorkshire, Landrace from both countries, and Yorkshire from both countries, respectively.*

*Number of SNP pairs.*

*Mean distance in mega base pairs.*

*Maximum distance in mega base pairs.*
FIGURE CAPTIONS

**Figure 1a.** French Genes Flow between Canada and France (LA_CAN: Canadian Landrace, YO_CAN: Canadian Yorkshire).

**Figure 1b.** Canadian Genes Flow between France and Canada (LA_FRA: French Landrace, YO_FRA: French Yorkshire).

**Figure 2.** Plot of the first two principal components using genotypes for all animal coloured by breed/country (LA_CAN, LA_FRA, YO_CAN and YO_FRA represents Canadian Landrace, French Landrace, Canadian Yorkshire and French Yorkshire, respectively).

**Figure 3.** Average $r^2$ at given distances for Landrace and Yorkshire per country. (Mb = Mega Base, $r^2$ = measure of linkage disequilibrium, LA_CAN, LA_FRA, YO_CAN and YO_FRA represents Canadian Landrace, French Landrace, Canadian Yorkshire and French Yorkshire, respectively).

**Figure 4.** Consistency of gametic phase over between countries and breeds. (Mb= Mega Base; LA_CAN, LA_FRA, YO_CAN, YO_FRA, LA_all, and YO_all represents Canadian Landrace, French Landrace, Canadian Yorkshire, French Yorkshire, Landrace from both countries, and Yorkshire from both countries, respectively).
Figure 1a. French Genes Flow between Canada and France (LA_CAN: Canadian Landrace, YO_CAN: Canadian Yorkshire).
Figure 1b. Canadian Genes Flow between France and Canada (LA_FRA: French Landrace, YO_FRA: French Yorkshire).
Figure 2. Plot of the first two principal components using genotypes for all animal coloured by breed/country (LA_CAN, LA_FRA, YO_CAN and YO_FRA represents Canadian Landrace, French Landrace, Canadian Yorkshire and French Yorkshire, respectively).
Figure 3. Average $r^2$ at given distances for Landrace and Yorkshire per country. (Mb = Mega Base, $r^2$ = measure of linkage disequilibrium, LA_CAN, LA_FRA, YO_CAN and YO_FRA represents Canadian Landrace, French Landrace, Canadian Yorkshire and French Yorkshire, respectively).
Figure 4. Consistency of gametic phase over between countries and breeds. (Mb = Mega Base; LA_CAN, LA_FRA, YO_CAN, YO_FRA, LA_all, and YO_all represents Canadian Landrace, French Landrace, Canadian Yorkshire, French Yorkshire, Landrace from both countries, and Yorkshire from both countries, respectively).