Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds

P. Uimari and M. Tapio

Agrifood Research Finland, MTT, Biotechnology and Food Research, FI-36100 Jokioinen, Finland

ABSTRACT: The extent of linkage disequilibrium (LD) and effective population size in Finnish Landrace and Finnish Yorkshire pig populations were studied using a whole genome SNP panel (Illumina PorcineSNP60 BeadChip) and pedigree data. Genotypic data included 86 Finnish Landrace and 32 Finnish Yorkshire boars. Pedigree data included 608,138 Finnish Landrace and 554,237 and Finnish Yorkshire pigs, and on average 15 ancestral generations were known for the reference animals, born in 2005 to 2009. The breeding animals of the 2 populations have been kept separate in the breeding programs. Based on the pedigree data, the current effective population size for Finnish Landrace is 91 and for Finnish Yorkshire 61. Linkage disequilibrium measures (D’ and r²) were estimated for over 1.5 million pairs of SNP. Average r² for SNP 30 kb apart was 0.47 and 0.49 and for SNP 5 Mb apart 0.09 and 0.12 for Finnish Landrace and Finnish Yorkshire, respectively. Average LD (r²) between adjacent SNP in the Illumina PorcineSNP60 BeadChip was 0.43 (57% of the adjacent SNP pairs had r² > 0.2) for Finnish Landrace and 0.46 (60% of the adjacent SNP pairs had r² > 0.2) for Finnish Yorkshire, and average r² > 0.2 extended to 1.0 and 1.5 Mb for Finnish Landrace and Finnish Yorkshire, respectively. Effective population size estimates based on the decay of r² with distance were similar to those based on the pedigree data: 80 and 55 for Finnish Landrace and Finnish Yorkshire, respectively. Thus, the results indicate that the effective population size of Finnish Yorkshire is smaller than of Finnish Landrace and has a clear effect on the extent of LD. The current effective population size of both breeds is above the recommended minimum of 50 but may get smaller than that in the near future, if no action is taken to balance the inbreeding rate and selection response. Because a moderate level of LD extends over a long distance, selection based on whole genome SNP markers (genomic selection) is expected to be efficient for both breeds.

Key words: effective population size, genome, linkage disequilibrium, pedigree, pig

INTRODUCTION

Management of the genetic diversity of a breed is essential for the sustainable use of its genetic resources and its genetic improvement in the future (Burrow, 1993; Falconer and MacKay, 1996), which is why significant effort is made to control inbreeding through breeding programs (Meuwissen and Woolliams, 1994; Hill, 2000). Traditionally, inbreeding has been evaluated on the basis of pedigree information (Wright, 1922), and the inbreeding rate is often converted to effective population size (Nₑ), which is seen as a general indicator of the risk of genetic erosion (FAO, 2000). Pedigree-based estimates of Nₑ are, however, dependent on the completeness of the available pedigree. A promising approach that is independent of the completeness of the pedigree is to estimate past effective population size on the basis of genomic data. Such methods are based on linkage disequilibrium (LD) patterns (Sved, 1971; Hill, 1981; Hayes et al., 2003). Genome-wide SNP genotyping became possible in pigs in 2009 (Ramos et al., 2009). This new SNP technology gives useful tools to study the structure of pig populations and allows for a more detailed comparison between populations than earlier pedigree-based approaches.

Finnish Landrace and Finnish Yorkshire are the 2 domestic pig breeds in Finland. The selection target for both breeds is the same: good growth rate with efficient feed conversion rate and increased carcass quality. The size of their active purebred breeding populations has also been similar over the last decades and is currently approximately 2,000 sows in both breeds. Moreover, the same genetic evaluation system and selection intensity have been applied for both breeds, and both underwent a strong selection against halothane sensitivity.

©2011 American Society of Animal Science. All rights reserved.

doi:10.2527/jas.2010-3249

609
in the 1980s and 1990s. The major difference in selection practices between the 2 breeds is a strong marker-assisted selection against the immotile short-tail sperm defect (Sironen et al., 2006), which has been applied for Finnish Yorkshire since 2001.

The objective of this study is to characterize LD in the 2 domestic Finnish pig breeds and to assess temporal changes in their effective populations. Both pedigree and LD-based methods were applied in estimating $N_e$. Genomic data were also used for pedigree structure analysis. The genomic analysis in this study is based on the Illumina PorcineSNP60 BeadChip (Illumina, San Diego, CA) containing 62,163 SNP.

**MATERIALS AND METHODS**

Institutional Animal Care and Use Committee approval was not obtained for this study because already existing semen and hair samples were used for genotyping. Only semen and hair samples were collected from animals.

**Pedigree Data and Analysis**

Pedigree data were obtained from Faba Pig, the national pig breeding organization, and included 608,138 Finnish Landrace and 554,237 Finnish Yorkshire pigs. The oldest records were from the early 1970s, but the majority of animals in the data were born after 1980. Effective population size was estimated based on the average inbreeding rate ($\Delta F$) of the reference animals born in 2005 to 2009, using the equation $N_e = 1/(2\Delta F)$ (Gutiérrez et al., 2008, 2009). The individual increase in inbreeding was calculated for each animal $i$ with the equation proposed by Gutiérrez et al. (2009):

$$\Delta F_i = 1 - \sqrt[2t]{1 - F_i},$$

where $t_i$ is the number of equivalent complete generations (Maingel et al., 1996). Inbreeding coefficients were calculated for all animals in the data, but $N_e$ estimates were based only on animals with a pedigree completeness value ($PEC$; MacCluer et al., 1983) of 0.7 or greater. To validate the robustness of the $N_e$ estimates, a smaller reference population (animals born in 2007 to 2009) and a lower limit for $PEC$ were also used. The PEC calculation was based either on 5 or 10 ancestral generations. The individual inbreeding rate and effective population size were estimated using the RelaX2 program (Strandén and Vuori, 2006).

**DNA Samples and Genotyping**

For this study we analyzed the DNA samples of 86 Finnish Landrace and 32 Finnish Yorkshire boars born in 2004 to 2008. The average inbreeding coefficient of the genotyped boars was 0.07 for Finnish Landrace and 0.10 for Finnish Yorkshire. These values are the same as the average inbreeding coefficients of all registered animals in these breeds born in 2004 to 2008. The average coefficient of relationship of the genotyped Finnish Landrace boars was 0.16, whereas the population average of all registered Finnish Landrace animals born in 2004 to 2008 was 0.14. The genotyped boars were thus slightly more related than randomly selected animals. In the case of the genotyped Finnish Yorkshire boars, the average coefficient of relationship was 0.22, which is very close to the population average of 0.21 (animals born in 2004 to 2008).

The DNA was extracted either from semen after phenol/chloroform extraction or from hair follicles using a DNeasy Blood and Tissue kit (Qiagen, Helsinki, Finland). For each sample, 20 μL of extracted DNA with target DNA concentrations of 50 (hair) and 100 (semen) ng/μL in TE buffer was provided for the genotyping laboratory. Genotyping was conducted at the Institute for Molecular Medicine Finland (FIMM) using the PorcineSNP60 BeadChip and the protocol provided by the manufacturer.

**Population Analysis Using SNP Genotypes**

The population structure of the 2 breeds was examined with the multidimensional scaling algorithm implemented in PLINK (Purcell et al., 2007). Only SNP that had a minor allele frequency $\geq$0.05 and Hardy-Weinberg equilibrium test statistics ($\chi^2 < 24$ corresponding $P$-value $\sim 10^{-6}$), and that had been mapped on pig genome build 9 (Sscrofa9, http://www.ensembl.org) were used in the calculations. A total of 35,090 SNP in the Finnish Yorkshire and 35,243 SNP in the Finnish Landrace data met all 3 requirements. Pairwise haplotype frequencies were estimated from the genotype data using the EM algorithm (Weir, 1996). The estimated pairwise haplotype frequencies were used to calculate the standardized LD value ($D'$) and the squared correlation coefficient between the 2 loci ($r^2$).

For each SNP, pairwise LD was calculated for adjacent SNP that were less than 5 Mb apart. The SNP quality and distance requirements resulted in approximately 1.5 million SNP pairs distributed over the genome.

The relationship between LD and effective population size can be expressed by the equation:

$$r^2 = 1/(4cN_e + 1),$$

where $c$ is the genetic distance between 2 SNP expressed in morgans (Sved, 1971). Effective population size $N_e$ was estimated based on all autosomal chromosomes, and the physical distances between SNP (Sscrofa9, http://www.ensembl.org) were converted to genetic distances with the simple assumption of 1 cM $\sim$1 Mb. Past effective population size $N_0$, at generation $t$, where $t = 1/2c$, was estimated for each SNP pair using the above equation. The final point estimates of $N_e$ were obtained by averaging the estimated $N_e$ values within $t$. Because $r^2$ has a limited parameter space and the point estimates of $N_e$ are infinite at $r^2 = 0$ and zero at $r^2 = 1$, only values of $0.01 < r^2 < 0.99$ were used for estimating $N_e$. 


RESULTS

Rate of Inbreeding and Effective Population Size Based on Pedigree Information

Average PEC of the reference population (animals born in 2005 to 2009) was 0.97 and 0.99 for Finnish Landrace and Finnish Yorkshire, respectively, using 10 ancestral generations. The average number of equivalent complete generations in this data set reached 15 for both breeds by the end of 2009. The estimated rate of inbreeding per generation was 0.55% for Finnish Landrace and 0.8% for Finnish Yorkshire. These correspond to an effective population size of 91 and 61 for Finnish Landrace and Finnish Yorkshire, respectively (Table 1). Overall the estimates were not sensitive to a reduction in the base data; a reduced reference population and a decreased PEC limit (0.7 vs. 0.4) had only a minor effect on the estimates (Table 1). Also, the use of 5 ancestral generations instead of 10 in the PEC calculations did not affect the $N_e$ estimates greatly. The largest differences between the $N_e$ estimates were obtained for Finnish Landrace when moving from the larger to the smaller reference population. One explanation for the observed difference between the 2 breeds is that in Finnish Landrace $N_e$ may still decrease, whereas in Finnish Yorkshire it has stabilized. The use of the smaller reference data set (animals born in 2007 to 2009) resulted in a greater SE than observed for the larger reference data set (Table 1).

Population Structure

The population structure of the genotyped data is visualized in the multidimensional scaling plot in Figure 1. The 2 studied breeds are clearly separate populations: the genotyped animals were purebred with a single possible outlier that stood apart from both clusters. This animal most probably had a crossbred parent because it had a similar profile as Finnish Landrace $\times$ (Finnish Landrace $\times$ Finnish Yorkshire) backcrosses that have been genotyped before this study. Sample mixing or DNA contamination with Finnish Yorkshire DNA is improbable because the DNA of the 2 breeds were extracted at different times. The outlier individual was excluded from subsequent analysis.

Average LD

Average LD ($r^2$) between adjacent SNP in the PorcineSNP60 BeadChip was 0.43 for Finnish Landrace and 0.46 for Finnish Yorkshire. Overall, in Finnish Landrace, 49% of adjacent SNP pairs had $r^2 > 0.3$ and 57% had $r^2 > 0.2$. The corresponding percentages for Finnish Yorkshire were 52 and 60%. Table 2 shows the average LD for each chromosome for distances of 30 kb (median distance between markers in the PorcineSNP60 BeadChip) and 3 Mb between SNP. Although the aim of this study was not to compare LD in different chromosomes, we observed some variation in the extent of LD in different chromosomes in both populations. Chromosomes 10 and 12 had the least average LD and chromosomes 1, 13, and 14 the greatest average LD. The average LD in different autosomal chromosomes for SNP 30 kb apart varied from $D' = 0.85$ to $D' = 0.93$ and $r^2 = 0.39$ to $r^2 = 0.55$. Even larger differences were found between different chromosomes when SNP farther apart (3 Mb) were considered.

Figure 2 plots the average LD across all autosomal chromosomes. Overall LD was stronger in Finnish Yorkshire compared with Finnish Landrace, their average $r^2 > 0.2$ extending to 1.5 and 1.0 Mb, respectively. This is in a good agreement with the pedigree-based estimates of effective population size, which gave a smaller $N_e$ for Finnish Yorkshire than for Finnish Landrace. The decrease in LD with an increasing distance between SNP was quite moderate, and even SNP as far as 5 Mb apart

Table 1. Average effective population size of Finnish Landrace and Finnish Yorkshire populations based on reference animals born in 2005 to 2009 and in 2007 to 2009 with different pedigree completeness (PEC) limits and numbers of ancestral generations used for PEC calculation

<table>
<thead>
<tr>
<th>Item</th>
<th>5 generations</th>
<th>10 generations</th>
<th>5 generations</th>
<th>10 generations</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEC limit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005 to 2009</td>
<td>0.4</td>
<td>0.7</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.16)</td>
</tr>
<tr>
<td>2007 to 2009</td>
<td>89.01</td>
<td>88.98</td>
<td>88.98</td>
<td>88.98</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
</tr>
</tbody>
</table>
| Numbers in parentheses are SE of the estimates.
showed relatively large LD: $D' = 0.49$ and $r^2 = 0.09$ for Finnish Landrace and $D' = 0.59$ and $r^2 = 0.12$ for Finnish Yorkshire.

**Effective Population Size Based on Genomic Data**

Figure 3 plots the estimated effective population size $N_e$ at $t$ generations ago. Based on the genomic data, the current effective population size of Finnish Landrace is approximately 80 animals and of Finnish Yorkshire approximately 55. The estimates of effective population sizes calculated from the genomic data were in good agreement with the $N_e$ values calculated from pedigree data. Figure 3 also indicates that the reduction in $N_e$ in both populations was stronger during the past 10 generation than before that. The estimate of past $N_e$ (50 generations ago) for Finnish Landrace was approximately 600 and for Finnish Yorkshire approximately 500.

**DISCUSSION**

We investigated the extent of LD as well as the current and past effective population size $N_e$ of the 2 commercial Finnish pig breeds, Finnish Landrace and Finnish Yorkshire, based on pedigree and whole genome SNP data. The observed LD, measured as $r^2$, extended for a long distance in the genome for both breeds, as $r^2 > 0.2$ extended to 1.0 and 1.5 Mb. Moreover, our study shows that the 2 $N_e$ estimation approaches give concordant results, thus supporting the usability of genomic estimates when pedigrees are not available.

In agreement with the pedigree-based estimates of effective population sizes, LD was generally stronger in the Finnish Yorkshire population ($r^2 = 0.16$ for SNP 3 Mb apart) than in Finnish Landrace ($r^2 = 0.12$ for SNP 3 Mb apart). These LD values are similar to the average value presented by Du et al. (2007) for 6 commercial pig lines ($r^2 = 0.11$ for markers 3 cM apart), which is the only SNP-based study on pig genome LD known to the authors. Overall LD in pigs seems to be stronger

**Table 2. Average linkage disequilibrium (LD) in different autosomal chromosomes for SNP 30 kb and 3 Mb apart**

<table>
<thead>
<tr>
<th>Item</th>
<th>30 kb between SNP</th>
<th>3 Mb between SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Finnish Landrace</td>
<td>Finnish Yorkshire</td>
</tr>
<tr>
<td>Chromosome</td>
<td>n</td>
<td>$D'$</td>
</tr>
<tr>
<td>1</td>
<td>1,142</td>
<td>0.93</td>
</tr>
<tr>
<td>2</td>
<td>511</td>
<td>0.91</td>
</tr>
<tr>
<td>3</td>
<td>369</td>
<td>0.93</td>
</tr>
<tr>
<td>4</td>
<td>703</td>
<td>0.92</td>
</tr>
<tr>
<td>5</td>
<td>360</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>394</td>
<td>0.91</td>
</tr>
<tr>
<td>7</td>
<td>667</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>414</td>
<td>0.88</td>
</tr>
<tr>
<td>9</td>
<td>539</td>
<td>0.87</td>
</tr>
<tr>
<td>10</td>
<td>256</td>
<td>0.85</td>
</tr>
<tr>
<td>11</td>
<td>302</td>
<td>0.90</td>
</tr>
<tr>
<td>12</td>
<td>240</td>
<td>0.85</td>
</tr>
<tr>
<td>13</td>
<td>577</td>
<td>0.92</td>
</tr>
<tr>
<td>14</td>
<td>855</td>
<td>0.93</td>
</tr>
<tr>
<td>15</td>
<td>391</td>
<td>0.91</td>
</tr>
<tr>
<td>16</td>
<td>278</td>
<td>0.90</td>
</tr>
<tr>
<td>17</td>
<td>245</td>
<td>0.92</td>
</tr>
<tr>
<td>18</td>
<td>214</td>
<td>0.93</td>
</tr>
<tr>
<td>All</td>
<td>8,457</td>
<td>0.91</td>
</tr>
</tbody>
</table>

1n is the total number of SNP pairs used to calculate the average values. $D'$ is the standardized LD value, and $r^2$ is the squared correlation coefficient between SNP.

2Average values across all autosomal chromosomes. The SE of the estimates are between 0.01 and 0.03 for individual chromosomes and between 0.003 and 0.006 across chromosomes (All).
than in cattle, where SNP-based $r^2 > 0.2$ extends from 40 to 80 kb (de Roos et al., 2008; Khatkar et al., 2008; Flury et al., 2010). Information on other livestock species is less extensive, particularly as different marker types are not directly comparable (Pritchard and Przeworski, 2001).

The ancestors of most animals in the reference population were known up to 15 generations, which made it possible to estimate the inbreeding coefficients and effective population sizes quite reliably. The pedigree-based effective population size for Finnish Landrace was 91 and for Finnish Yorkshire 61. These $N_e$ values are similar to those reported for American pig breeds: 74 for Landrace, 77 for Berkshire, 109 for Hampshire, and 113 for Duroc and Yorkshire (Welsh et al., 2010). Breeding value estimation and selection intensity have been similar for the 2 Finnish breeds for several decades. Thus, the greater inbreeding rate and smaller effective population size of the Finnish Yorkshire population compared with Finnish Landrace is most likely due to an uneven use of top sires and sire lines. Evidence for this is the proportion of genes inherited from the top 10 AI boars, which in Finnish Yorkshire increased from 44 to 70% during the last 2 decades, whereas the corresponding increase in Finnish Landrace has been more moderate (from 37 to 56%; Uimari et al., 2010). Also, a strong marker-assisted selection against the immotile short-tail sperm defect (Sironen et al., 2006) has been applied in the Finnish Yorkshire population since 2001, which has restricted the pool of breeding animals to those that do not carry the defect. However, the rate of inbreeding was stronger in Finnish Yorkshire than in Finnish Landrace even before 2001, and the proportion of the Finnish Yorkshire gene pool explained by the main AI boar, which was also a carrier of the defect, has not decreased after the marker-assisted selection was introduced; the proportion was 14% for pigs born in 2001 and 18% for pigs born in 2009 (calculated with the algorithm described by Boichard et al., 1997, using RELAX2). The recent strong marker-assisted selection against the short-tail sperm defect, thus, does not explain the difference in effective population sizes between the 2 breeds.

Linkage disequilibrium over long distances reflects the recent population history, whereas LD over short distances depends on the effective population size many generations ago (Hill, 1981; Hayes et al., 2003). Estimating the current and past effective population sizes using the equation of Sved (1971) raises difficulties in handling values that are in the limits of the parameter space, because if $r^2 = 0.0$, the estimate of $N_e$ is infinite, and if $r^2 = 1.0$, it is zero. On the other hand, limiting the range of $r^2$ between 0.01 and 0.99, as in our approach, may create some bias in the estimated values. However, the results from this simplified approach yielded quite similar estimates of $N_e$ as did the pedigree-based approach. Another concern related to the relationship between estimated LD and the distance between markers lies in the accuracy of the genome assembly (Sscrofa9) used in this study. Eventually, the order and distance between markers in the commercial Illumina PorcineSNP60 BeadChip will be refined. However, it can be assumed that the bias from a wrong order of markers or wrong distance between markers will be diluted by the large number of SNP pairs used in this study, and also by the fact that some distances will be overestimated and some underestimated. Additional challenges are created by the unknown genetic distance between SNP. In this article we used the simple relationship of 1 cM ~ 1 Mb. It is well known that the relationship between genetic and physical distances varies across chromosomes and chromosomal regions. If LD is small and the assumed genetic distance between SNP is too small, then $N_e$ will be underestimated. Thus, more detailed estimates of genetic distances between SNP will yield more reliable $N_e$ estimates. Finally, a limited sample size can bias the estimates of $r^2$. In this study, correcting the estimates of $r^2$ for sample size $n$ ($r^2 - 1/2n$, Weir and Hill, 1980) and then using the Sved (1971) equation had only a minor effect on our primary interest: the estimates of $N_e$ of the recent population history of Finnish Landrace and Finnish Yorkshire.

The current effective population size of Finnish Yorkshire is close to 50, which is the minimum number of breeding animals recommended by the UN Food and Agriculture Organization (FAO, 2000). Meuwissen (2009), however, suggested this recommendation as the lower limit for probable critical population size, proposing that the actual critical size should be somewhere between 50 and 100. To maintain sufficient genetic variation in the population as well as an acceptable inbreeding rate without compromising genetic gain in commercially important traits in the future, it is necessary to apply appropriate selection methods that maximize the selection response with a fixed rate of inbreeding (Meuwissen, 1997; Colleau and Tribout, 2008) or methods that optimize the use of the genetic resources of the parental generation (Sánchez et al., 2003). Moreover, it has been argued that genomic selection (Meuwissen et al., 2001), an effective new method for estimating breeding values, leaves a decreased rate of
inbreeding compared with current BLUP-based breeding value estimation (Daetwyler et al., 2007). Yet, if the effective population size continues to be very small or continues to decrease, it is highly important to give more emphasis in breed management to the avoidance of inbreeding. Although this may have an effect on short-term genetic gain, it is essential for securing the long-term genetic variability of the breed and, thus, its future genetic gain and competitiveness. Imported germplasm can be used to reduce inbreeding, but there is a risk that a small population will gradually lose its own genetic characteristics and the foreign germplasm will dominate. Alternatively, breeds with a similar ancestral background and selection objectives might be combined and treated as a single breeding population. Again, this would increase the local effective population size momentarily, but monitoring methods would also be needed in the long run to control the use of animals globally to avoid an unintended reduction in the global effective population size. The best way to maintain a sustainable global effective size of a breed is to make sure that its production populations maintain a sufficient $N_e$. Efficient monitoring and management are, therefore, imperative.

**LITERATURE CITED**


