Norwegian University
of Life Sciences

Master's Thesis 2018 60 ECTS

Faculty of biosciences, department Animal and Aquacultural Sciences (IHA)
Hanne Fjerdingby Olsen,
Gunnar Klemetsdal
Dag Inge Våge

Utilizing shared segments for build-up genomic relationship between the Norwegian and the Swedish Fjord Horse

Saija Tenhunen

Animal breeding and genetics
Faculty of biosciences, department Animal and Aquacultural Sciences (IHA)

Contents

Abstract 1

Introduction 2

Material and Methods 5

- Pedigree data 5
- Genomic data 5
- Inbreeding calculations 6
- Coancestry calculations 7
- Effective population size based on changes on inbreeding and coancestry 9
- Statistical calculations 9

Results 10

- Pedigree data and population structure 10
- Estimations of inbreeding and coancestry 10
- Effective population size 12
- Multidimensional scaling plot 14

Discussion 14

Sources 19

Tables and figures 25

Abstract

Challenges in conservation of native breeds, like the Norwegian Fjord Horse, require an evaluation of the relationship between animals in different countries. Traditionally pedigree data have been used to calculate relationships between animals, but pedigree information is not comparable in most cases between different countries or breeding organizations. Methods based on genomic relationships are therefore a good option. In this study we looked at 413 samples from Fjord Horses, of which 311 were Norwegian and 102 were Swedish. Individual inbreeding and coancestry were evaluated based on pedigree (PED), molecular homozygosity (HOM) and individual runs of homozygosity (ROH) or shared genomic segments (SEG) between two horses. Effective population size (Ne) was calculated from the increase of inbreeding (ΔF) or coancestry (Δf) in a generation. These methods were tested with 13 different SNP densities from 485,918 SNPs to 33,420 SNPs. With Fjord Horses pedigrees are well known and the complete generation equivalent for Norwegian genotyped horses was 13.7 and for Swedish genotyped horses was 12.4. For the Norwegian genotyped horses average PED inbreeding was 0.077 and coancestry was 0.082, and for the Swedish PED inbreeding was on average 0.052 and coancestry was 0.065. HOM methods were stable when SNP density was thinned but these were not of the same scale as PED and ROH inbreeding or SEG coancestry. ROH and SEG methods to calculate inbreeding and coancestry performed best when regressed on PED. When SNP densities were thinned down below 65k SNPs ROH and SEG methods decreased significantly and gave inaccurate estimates for inbreeding and coancestry. With ROH and SEG, small segment sizes (100kb and 500kb) detected more inbreeding and coancestry than medium segment sizes (1.5 Mb and 2 Mb). A large segment size of 5 Mb underperformed in all calculations and was ruled out. Ne calculated from PED ΔF for Norwegian horses was 71 and for Swedish Ne was 269. When calculating Ne from ΔF, ROH and HOM methods gave similar values for Ne as what was gained from PED with Norwegian horses. Results were unreliable for Swedish horses due to the low number of genotyped horses in the timeline. Ne based on ΔF was tested with three different timelines; pairs born per year, pairs born in two years and pairs born in same ten-year cohort. With HOM coancestry there was a low variation in results which lead to a shallow regression slope and therefore higher values for Ne with more standard error than Ne from SEG. Ne from SEG, when used ten-year cohort timeline, was 63 and was close to PED Ne. Ne SEG had the lowest standard error and most narrow confidence interval as compared to molecular methods used to calculate Ne. When using methods like optimal

contribution selection, SEG could be the more beneficial method based on our results and previous research. This needs to be studied more deeply with horse populations where breeding schemes are very different than in commercial livestock breeding.

Introduction

The Norwegian Fjord Horse, or Fjord Horse for short, is believed to be one of the oldest horse breeds existing. The breed was developed after the last ice age in the mountainous western Norway region and it was domesticated over four thousand years ago (Lynghaug, 2009). Fjord Horses share close genetic relatedness with Mongolian and Asian wild Przewalski horses, and it has been suggested that Mongolian horses had influence on the development of Fjord Horses (Bjørnstad et al., 2003 & McCue et al., 2012). A studbook for Fjord Horses in Norway was established in 1909 (Bjørnstad et al., 2000), and in Sweden a breed specific studbook was established in 1961 (Fjord Horse International, 2018). The breed has gone through several genetic bottlenecks in its history because of isolated locations, wars and agricultural industry change, when work horses were not needed anymore (Olsen et al., 2018). Currently the Fjord Horse population in Norway is around 5,000 horses, with less than 150 foals born per year, and the number of foals born has been decreasing for the past decades (Norwegian Equine Centre, NEC, 2018). In 1975 there were still over a thousand recorded matings for mares (Nestaas, 2010) and by 2015 the number of registered foals had dropped to a little bit over 200 (NEC, 2018).

Fjord Horses have gained popularity around the globe and, for example, in the United States of America the breed had 1659 born Fjord Horses between years 2000-2009 (Bhatnagar et al., 2011). Fjord Horse International (FjHI) also keeps records on foals born in each country (FjHI, 2018) and offers international cooperation between Fjord Horse breeders. Based on FjHI records there are substantial Fjord Horse populations outside of the Nordic countries in Germany and in the Netherlands. In 2016 there were almost 300 registered Fjord Horses in Germany, and in the Netherlands there were close to 150. These registration numbers surpass the number of foals registered in Norway, the country of origin of Fjord Horses. In 2016 FjHI only recorded 129 foals in Norway. Even though the Fjord Horse population has spread all over the world, it is still considered to be globally at risk for extinction (Domestic Animal Diversity Information System DAD-IS, 2018). One way to estimate a population's risk of extinction is to calculate effective population size (Boettcher et al., 2013). When estimating a population's survivability without signs of inbreeding depression, the effective

population size rule 50/500 has been used (Franklin, 1980), where 50 is the minimum limit for effective population size in the short term and 500 is the minimum limit in the long term. In 2011, the Fjord Horse's effective population size was estimated to be between 107-118 based on pedigree data (NEC, 2018). But this was calculated based on the number of animals bred, and this method tends to overestimate the effective population size (Leroy et al. 2013). Because the Fjord Horse population is in decline in Norway, it is important to measure current genetic diversity. It is critical that the breed be protected by conservation programs, and population management for genetic variation is needed for future survival of the breed.

Optimal contribution selection (OCS) is widely accepted as an efficient method in conservation breeding (Woolliams et al., 2015). OCS is based on restricting the increase of inbreeding through average coancestry between the selected animals when maximizing the genetic gain in a population (Sonesson et al. 2010). Coancestry is traditionally calculated from pedigree data, but in the past years molecular methods have been replacing the use of pedigrees (Cara et al., 2011, Gómez-Romano et al., 2013, Clark et al., 2013, Toro et al., 2014, Cara et al., 2013a). In most studies molecular coancestry has been calculated on a SNP-by-SNP basis but this method does not penalize for deleterious mutations when used with OCS. When using SNP-by-SNP coancestry the OCS will try to maintain all possible variations of alleles, which includes the deleterious mutations. In this strategy high genetic diversity will be sustained but fitness in a population will decrease (Cara et al., 2013b). Lately, methods based on shared genomic segments between two animals have been suggested for calculating molecular coancestry to avoid this issue (Cara et al., 2013a, Bosse et al., 2015, Rodríguez-Ramilo et al., 2015, Gómez-Romano et al., 2016). In this method genomes from two different animals are compared and the fraction of shared genomic segments is calculated (Cara et al. 2013a). It is thought that deleterious mutations are located in the longer runs of homozygosity (Szpiech et al., 2013) and when deleterious areas in the genome are unknown it would be advantageous to use coancestry management with shared segments because this method will aim to decrease the number and length of shared segments. Simulations have shown that loss of fitness is reduced when segment-based coancestry is used with OCS (Bosse et al. 2015).

With horses, breeding is based on traits like speed, endurance or color and not on production of food as with cattle (Olsen, 2006). Dairy cattle breeding schemes have been based on estimated breeding values (EBV) of quantitative traits for decades (Hayes et al., 2009). It

was noted that increasing the accuracy of EBV tended to increase the inbreeding in a population because of the preference for related animals with high EBVs. With genomic selection the increase of inbreeding in dairy cattle populations has slowed, but they are still losing genetic diversity (Daetwyler et al., 2007). There are no such strict breeding schemes in Fjord Horses as there are in dairy cattle breeding and breeders are mainly hobby breeders (Avlsplan for den norske fjordhesten, 2018). In Fjord Horses, qualitative traits like color have been under the selection in the past. These differences on breeding schemes and goals can lead to different genetic structures in a population because in one method you are making selection of many genes controlling different traits (quantitative) and in the other you have only few single genes under selection (qualitative).

Different molecular coancestry methods have been under research with commercial livestock breeds, but not at all on hobby animals such as horses. Over species, there are more studies about runs of homozygosity inbreeding (Peripolli et al. 2017) in a population than there are about shared segment coancestry. However, individual inbreeding based on runs of homozygosity has been studied in different horse populations (Metzger et al., 2015, Khanshour, 2013, Kamiński et al., 2017 & Druml et al., 2017). Metzger et al. (2015) focused on distribution and amount of runs of homozygosity in horses and how these might influence the genetic diversity with selection. Kamiński et al. (2017) studied genomic inbreeding based on runs of homozygosity in Polish Konik horses with EquineSNP60 BeadChip (65k SNPs), but they did not research how well their results correlated with a pedigree inbreeding coefficient. Druml et al. used Equine BeadChip 670k and studied genetic diversity with runs of homozygosity in Haflingers, Norikers, Bosnian Mountain Horses, Gidrans, Shagaya Arabians and Purebreed Arabians but they only had pedigree information for the Haflingers from Austria (N=78) and Noriker (N=190). They found low correlations between ROH and pedigree inbreeding.

Because Norway is the country of origin of Fjord horses and the population has been in decline for the past decades, it is important to consider Fjord Horse populations in other countries as possible breeding stock to increase and maintain genetic diversity in the breed. In the past there have been imports from Denmark to Norway, and these horses have contributed to the current population. Further, it is difficult to analyze how closely related horses from different countries are because pedigrees are not comparable between the different organizations carrying out registration. For a small and endangered population, it is crucial to be able to use animals from different countries and subpopulations for breeding

and conservation. Therefore, it is important to study methods where relationships are calculated from molecular data that will bring about a better understanding of relationships between Fjord Horses around the globe. In this research we will study different methods to calculate inbreeding, coancestry and effective population size in the genotyped Fjord Horse populations from Norway and Sweden. The aim of this study was to identify which methods and parameters will perform best when aimed at calculating molecular inbreeding and coancestry in the Fjord horse population.

Material and Methods

Pedigree data

Through a research project on the Fjord horse population (Olsen et al. 2018), data was made available for this study. In this project pedigree data had been organized for the Fjord horse populations in Norway and Sweden. The data was acquired from the Norwegian Trotting Association and the Swedish Trotting Association. There were 26,446 animals in the Norwegian pedigree data and 14,595 animals in the Swedish pedigree data. Pedigree data for both populations was included from the founding of the studbook to the year 2015. Because of differences in individual numbering between countries, combining pedigree data from different registration associations was not possible. Thus, all calculations with pedigree data were done separately for each population.

EVA (Berg 2006) was used to calculate complete generation equivalent (CGE) for both populations. When CGE is calculated with EVA the animal itself is the first generation. Swedish pedigree data was lacking information of birth years and therefore data was re-ordered by generation and not by birth year. Re-ordering was done with Rstudio (RStudio Team, 2015) and Pedigreemm (Vazquez et al., 2010).

Genomic data

Animals sampled were selected for the largest possible age distribution. For practical reasons samples were mainly from the Eastern and Western parts of Norway since most of the Fjord Horses are located in these areas. Samples from Sweden were collected from horses which were accessible.

In total, 432 samples were genotyped with an Axiom Equine Genotyping Array chip (Schaefer et al., 2017) at the Centre for Integrative Genetics (CIGENE), Norway. The array includes 670,796 SNP markers. Quality control (QC) for samples was first done with using

Axiom Analysis Suite v2.0.0.35 software. This included two filters; Dish QC (threshold 0.82) and QC Call Rate (call rate >97%), done respectively. Dish QC is done to measure the number of non-polymorphic loci with clean signals to avoid contaminated and low-quality DNA or processing problems. QC Call Rate calculates the percent of genotypes reported from an assigned subset of SNPs. If this percentage is under the chosen threshold, it can indicate quality problems in DNA. In total 423 samples and 505,601 SNPs passed these QC.

A second QC was done with Plink 1.9 software (Purcell et al., 2007, Chang et al., 2015) to prune out possible genotyping errors in SNPs. The following filter parameters were used for QC: variants which have Hardy-Weinberg equilibrium exact test p-value under the threshold (0.0001), missing call rate (threshold 0.05) and maximum missing call rate per sample (threshold 0.05). Sex chromosomes were filtered out from the datasets used. Minor allele frequencies (MAF) were not pruned out, to keep estimations of inbreeding as accurate as possible (Hillestad, 2015). Six horses were removed from the dataset because they had missing pedigree information and four were removed because they were duplicates (same horses tested twice by error). After the QCs, the full dataset included 485,918 SNPs from 413 samples, whereof 311 were Norwegian and 102 were Swedish. Of all genotyped animals 198 were male and 215 were female. Most of the genotyped horses were born between 1999-2015 but the oldest animals were born as early as 1985 in the Swedish population and in 1988 in the Norwegian population (Figure 1).

Plink 1.9 was used to create subsets with different SNP densities. This was done to see how different inbreeding and coancestry methods would perform when SNP densities were decreased. This would give better understanding of what might be the optimal or minimal number of SNPs to calculate molecular inbreeding and coancestry. Subsets with decreased SNP densities were done by thinning the datasets in every run by 20% (Table 1). In total 13 different SNP datasets were made.

Inbreeding calculations

Individual inbreeding estimates were calculated with three different methods: 1) pedigree based inbreeding coefficient, F_{ped}; 2) molecular homozygosity, $F_{snp;}$ and 3) runs of homozygosity. F_{roh} . EVA was used to calculate F_{ped} for the genotyped horses with an algorithm described by Meuwissen and Luo (1992). Plink 1.9 was used to calculate F_{snp} and F_{roh}.

F_{snp} was calculated for the genotyped animal as

$F_{snp} = (O_{hom} - E_{hom})/(N - E_{hom})$

where O_{hom} is the number of homozygeous SNP, E_{hom} is the expected number of homozygeous SNP and N is the number of non-missing SNP for the individual. For this the --het command, was used, which computes observed and expected homozygous genotype count from each sampled animal. When using --het it is important to use the command --non-founders to declare that the genotypes are from a non-founder population. If this is not done, --het will use only founder genotypes for the calculation.

F_{roh} was calculated as

$F_{roh} = \Sigma L_{roh}/\Sigma L_{auto}$

where ΣL_{roh} was the total ROH length for individual and ΣL_{auto} was the length of the autosomal equine genome, 2243.06 Mb (McQuillan et al., 2008). Information regarding the length of the equine genome was from the Genome database in the National Center for Biotechnology Information (NCBI).

Runs of homozygosity were detected using a sliding window 50 kb and assuming the parameters in Table 2 for, 1.) minimum number of homozygote SNPs, 2.) maximum gap between homozygote runs and 3.) missing SNPs allowed per sliding window. The parameters were chosen based on the previous research (Peripolli et al., 2017, Hillestad, 2015). Heterozygote calls were not allowed in any runs to avoid inaccurate ROH calls (Ferenčaković et al., 2013 & Hillestad, 2015). The gap was chosen to be 100 kb in high density datasets and was increased to 1000 kb with low density. This was done to increase the accuracy on finding the ROHs in low density datasets (Purfield et al., 2012). More missing SNPs were allowed in datasets with high density because when you have more SNPs included there is higher chance that the genotyping results are missing information from some SNPs (Hillestad, 2015). ROH segment lengths, tested with all SNP densities, were; 100 kb, 500 kb, 1.5 Mb, 2 Mb and 5 Mb.

Coancestry calculations

Pairwise coancestry estimates were calculated with three different methods; 1.) pedigree based coancestry, $f_{ped,}$ 2.) molecular homozygosity, f_{hom}, and 3.) shared genomic segments, $f_{seg.}$ In total, in the coancestry calculations there were 85,180 animal pairs, where 48,205 were Norwegians, 5,151 were Swedish and 31,824 were pairs with one Norwegian and one

Swedish horse. RStudio package Kinship2 (Sinnwell et al., 2014) was used to calculate f_{ped},for horses from each population. f_{hom}, was calculated with pairwise distances IBS (identity by state) and Hamming distance with Plink 1.9 software for all genotyped horse pairs. IBS population clustering was visualized by Multidimensional scaling (MDS) plot by Plink 1.9 --mds-plot and Rstudio. MDS calculates dimensional IBS distances between genotyped pairs, to identify population structure.

Coancestry based on shared segments, f_{seg}, was calculated with an IBD (Identity by descent) detection algorithm in BEAGLE 4.1 software (Browning & Browning, 2013). BEAGLE 4.1 uses refined IBD detection which is done in two steps. In the first step shared haplotypes are identified with GERMLINE algorithm (Gusev et al., 2009) and in the second step candidate segments are refined with a probabilistic approach to asses evidence of IBD. With IBD calculations default settings were used, except for effective population size (Ne) and minimum logarithm of the odds (LOD). The Ne parameter was set to 83, based on previous research on Ne_{hom} in Norwegian Fjord Horses (Olsen et al., 2018). On default BEAGLE 4.1 uses Ne 1,000,000, which is more suitable for human populations than more inbred animal populations. BEAGLE 4.1 uses the LOD score 3 by default. This prunes out those shared segments between two animals which are not common in the population. But in this study, we are more focused on the two individuals and shared segments between them. Therefore, the LOD score parameter was changed to 0.1 to find most of the shared segments. This was the lowest value that the software accepted without issues on running the data. In the output file BEAGLE 4.1 creates a list of pairs and their shared segments which can be used to calculate lengths of shared segments.

f_{seg} (Cara et al., 2013) was calculated as

$$f_{seg} = \Sigma_k \Sigma_{a_i=1}^{2} \Sigma_{b_j=1}^{2} (L_{seg_k}(a_i, b_j))/4L_{auto}$$

where $L_{segk}(a_i, b_i)$ is the length of the k-th shared segment seg_k over the homologue a of individual i and homologue b of individual j, and L_{auto} is the length of the autosomal genome. Shared segment sizes tested with all SNP densities were 100 kb, 500 kb, 1,5 Mb, 2 Mb and 5 Mb.

Effective population size based on changes on inbreeding and coancestry

Effective population size, (Falconer & Mackay, 1996) for genotyped animals were calculated from the regression slopes in rates of change in inbreeding (ΔF) and coancestry (Δf) per generation. Natural logarithm was used to linearize the results and computed as ln(1-F) and ln(1-f) for every genotyped horse or horse pair in the timeline used. The regression model was the following:

$$y_i = \mu + \beta_1 * X_i + e_i$$

where μ is the constant (intercept), β_1 is the regression coefficient associated with the regressor X_i, which is the birth year of an individual or pair *i* and e_i is the random error.

ΔF and Δf (Hillestad, 2015) were calculated as

ΔF and Δf =$1 - e^{\beta_1}$ and the confidence interval of regression slope was calculated as
ΔF and Δf =$1 - e^{\beta_1 \pm 1.96 * SE}$

where β_1 is the regression slope and SE is the standard error. Effective population size was calculated as

Ne= 1/2ΔFL and Ne=1/2ΔfL

where L is the generation interval. Generation intervals for Norwegian and Swedish populations were obtained from a previous study by Olsen et al. (2018) (Table 3).

Δf was tested with three different ways to create a timeline for the regression slope. 1.) genotyped horse pairs born in the same year 2.) genotyped horse pairs born in neighboring two years where birth year of the younger horse was used to place it in the timeline and 3.) genotyped horse pairs born in same ten-year cohort where horses were given a pseudo birth year based on the average birth year of the pair. The two latter approaches increased the number of coancestries that could be used to calculate Δf, and the latter maximized it. Horse and horse pair distributions by year are shown in Figure 1. These latter calculations were done for molecular and pedigree data (F_{ped}, F_{snp},F_{roh} , f_{ped}, f_{hom}, f_{seg}).

Statistical calculations

Statistical calculations were performed with RStudio package R Commander (Fox, 2005). This package was used to calculate intercept (*a*), regression coefficient (*b*), coefficient of determination (R^2), mean, variance, coefficient of variation (CV), standard error (SE) and

confidence interval (CI). More specific information about formulas used in the package can be found in the documentation from Fox (2016). The following regression model was used

$$\ln(1 - F_y) = y_i = \mu + \beta * \ln(1 - F_x)_i + e_i$$

where F_y refers molecular homozygosity (F_{SNP} or f_{hom}), runs of homozygosity inbreeding (F_{roh}) or shared genomic segment coancestry (f_{seg}), μ is constant, β is the regression coefficient and e is the error. To test the regression the null hypothesis was set to H_0 $\beta = 1$ against the alternative $H_1 \neq 1$.

Results

Pedigree data and population structure

In Table 3. the complete generation equivalent for Norwegian genotyped horses was 13.7 and for Swedish 12.4. Average inbreeding was 2.56% and the average coancestry was 1.74% higher in the Norwegian genotyped population than in the Swedish. These values are indicated in Figure 2, with a horizontal line.

Estimations of inbreeding and coancestry

For all SNP densities used, average inbreeding based on molecular homozygosity, $F_{SNP,}$ gave stable results without great variation (Figure 2 A & B). With Swedish horses the F_{SNP} value was negative and lower than in Norwegian horses (Figure 2B). Results for average inbreeding from runs of homozygosity, F_{roh}, gave higher inbreeding on the Norwegian genotyped animals than in the Swedish (Figure 2 A&B). There was a difference on F_{roh} segment sizes in how much inbreeding each segment size used could detect. Small segment sizes (100 kb and 500 kb) detected more inbreeding with higher SNP densities than middle-sized segments (1.5 Mb and 2 Mb) or a large segment size of 5 Mb. Average F_{roh} with medium size segments was closer to average pedigree inbreeding, Fped, than the F_{roh} from small or large segment sizes. When the SNP density was decreased below 65K SNPs all F_{roh} results had a significant drop in average inbreeding and the lowest SNP density could only detect few runs of homozygosity.

Average coancestry values calculated from molecular homozygosity, f_{hom}, were much higher than results for pedigree coancestry, $f_{ped,}$ or for shared segments coancestry, f_{seg}, so the average results for f_{hom} did not fit in Figure 2. Actually, performance of f_{hom} in all SNP densities used, was very similar to F_{SNP}; it gave stable results without much variation even at the lowest SNP densities used. With Norwegian horses f_{hom} ranged between 0.785-0.786,

with Swedish horses this was 0.781-0.782 and with pairs between the two populations, where one of the coancestry pair was Norwegian and the other Swedish, f_{hom} ranged between 0.778-0.779. Similarly, to average F_{roh} the smallest segment sizes gave the highest average values for average f_{seg} in both populations and pairs between them (Figure 2 C, D & E). However, pedigree coancestry could not be calculated for pairs between Norwegian and Swedish populations so this value was set to 0 (Figure 2E). When SNP density was decreased, average coancestry results from small segment sizes started to level out with the results from medium sized segments. When the SNP density decreased to around 100k SNPs in thin dataset 7, average coancestry results were almost the same for small and medium segment sizes. The large segment size of 5 MB was underperforming compared to other segment sizes when calculating inbreeding and coancestry and was left out from further calculations (Figure 2 A, B, C, D & E).

When molecular inbreeding and coancestry methods are regressed on pedigree inbreeding or coancestry, the regression coefficient (*b*) should be 1. In Table 4 the regression coefficient is calculated for inbreeding and coancestry results when using largest SNP density. Regression coefficient (Table 4, *b*) for F_{SNP} was close to one and this method had the highest coefficient of determination (R^2) from molecular inbreeding methods (Table 4, R^2). Small segment sizes had values closer to 1 for the regression coefficient than medium segment sizes and these also had the highest R^2 values. But R^2 was still relatively low with the inbreeding methods. The intercept for F_{SNP} when regressed on F_{ped} was positive when with other methods this was negative. When a natural logarithm was used to linearize inbreeding values, all average results were negative except F_{SNP}. This indicates that molecular homozygosity inbreeding is in its own scale compared to F_{ped} and F_{roh}.

When f_{seg} was regressed on f_{ped}, regression coefficient (*b*) was almost equal to 1 with small segment sizes and medium sized segments gave lower values. Small segment sizes with f_{seg} gave higher values for R^2 than medium sized segment used on f_{seg} and therefore small segments could explain more of the variation on the results. Coancestry based on molecular homozygosity, f_{hom}, had very high intercept and low regression coefficients when regressed on f_{ped}. This indicates that f_{hom} is on a different scale from f_{ped} and f_{seg}. Further to support this, the regression coefficient intercept, when f_{seg} was regressed on f_{hom}, had very high values, far from the ideal 1 value. On the contrary, with inbreeding results when different F_{roh} segment sizes were regressed on F_{SNP}, the regression coefficient was closer to 1. Medium sized segments had lower values for regression coefficient and R^2. Small segments could

explain almost all variation between F_{roh} and F_{SNP} giving R^2 values of 0.921 in the Norwegian population and 0.886 in the Swedish.

Norwegian and Swedish results for intercepts, regression coefficients and R^2 (Table 4) performed in a similar way. When f_{seg} was regressed on f_{hom}, R^2 values were higher in the Swedish population than in Norwegian population. Pairs between these two populations had very low R2. Different segment sizes used with f_{seg} performed in the same way in all tested methods. Small segment sizes could explain more of the variation than medium sized segment sizes.

Molucular homozygosity methods to calculate inbreeding, F_{SNP}, and coancestry, f_{hom}, gave mean results for inbreeding and coancestry, which were widely different from the results of methods based on pedigree, runs of homozygosity or shared segments (Table 5). With F_{SNP} mean values for Norwegian horses were almost zero and with Swedish horses the mean value was even negative. Lowest values were -7.5% inbreeding with Norwegians, and with Swedish this was even lower: -9.1%. Inbreeding values should be between 0 to 1 and this makes the F_{SNP} method to be on a different scale than other inbreeding calculation methods. Coefficient of variation (Table 5, CV) gave unrealistic values for F_{SNP} because of these negative values. Molecular homozygosity coancestry, f_{hom}, gave, on the contrary, very high results for coancestry with very low variance compared to the other methods.

Medium segment size runs of homozygosity inbreeding, $F_{roh1.5Mb}$ and F_{roh2Mb} had lower range values than small segment sizes, $F_{roh100kb}$ and $F_{roh500kb}$ (Table 5). Medium segment F_{roh} values had more variance than small segment F_{roh} and the former had higher values for coefficient of variation (CV) than the latter. Pairs between Norwegian and Swedish horses had a lower mean for coancestry and they also had less variance in their results which gave much lower CV values than with Norwegian or Swedish populations.

Because small segment sizes had higher regression coefficient and could explain more of the variation, when regressed on both pedigree and molecular homozygosity methods, than medium sized segments, it was decided to use segment size of 100kb to test out effective population size calculation with different timelines.

Effective population size

Low variance with molecular homozygosity coancestry, f_{hom}, in the Table 5 gave a low value for the regression coefficient when increase in coancestry (Δf) was calculated for effective population size (Table 6, *b*). Regression coefficient for f_{hom} was very low on all timelines

and populations tested, expect with timeline of Swedish horses born per year. But this was not seen in other timelines with Swedish horses and was most likely due to the small numbers of genotyped Swedish horses. Standard error (Table 6, SE) was relatively high (0.0008) for this result compared to other timelines (0.00024 and 0.00043). Timelines with increased numbers of coancestry pairs did not have an effect on the f_{hom} regression coefficient. With pairs between two populations (NxS) f_{hom} gave infinite confidence intervals for the effective population size (Table 6, CI), except with the 'pairs born per 10 years' timeline. With pairs born per year with all genotyped horses f_{hom} also gave infinite CI for the effective population size.

With Norwegian horses molecular inbreeding, F_{SNP}, and runs of homozygosity inbreeding, F_{roh}, gave similar results (70 and 74) for effective population size (Table 6, Ne) to what was calculated from full pedigree data, F_{ped} (71). With F_{ped} Ne had a very narrow confidence interval (Table 6, CI) but this was larger with F_{roh} and F_{SNP}, where the latter gave the larger CI result for Ne of the two. In the Swedish population, standard error (Table 6, SE) was so high that CI for effective population size (Table 6, CI) went to infinite. The same was seen with Swedish coancestry results for Ne. This indicates that there were not enough genotyped horses from the Swedish population to create a reliable timeline. Even though with coancestry and an increased number of pairs in the timeline, results for Swedish population Ne confidence interval went to infinite and were unreliable.

Shared segment coancestry, f_{seg}, gave reliable results for effective population size with reasonable confidence intervals (Table 6) in all timelines used with Norwegian, NxS and all genotyped horses. When the coancestry pairs in the timeline were increased by including more years, f_{seg} gave Ne results with a smaller standard error and therefore the confidence interval was also smaller (Table 6, SE & CI). The same happened to the effective population size of Norwegian, NxS and all genotyped horses. With a ten-year timeline CI was the smallest and very close to Ne from F_{ped}. With Swedish horses even f_{seg} gave an infinite CI with all timelines for Ne, which strengthens the conclusion that there were not enough genotyped horses from the Swedish population.

Norwegian effective population size results based on f_{seg} were very close to the Ne calculated from the increase of inbreeding (ΔF) from full pedigree, F_{ped}, results. With F_{ped} the effective population size was 71 and with f_{seg} the Ne ranged between 52 — 63 with three different timelines used. Effective population size for Swedish horses was 269 and the larger Ne might be because of frequent use of foreign stallions in the breeding scheme. Swedish Ne values

from f_{seg} method were unreliable and were between 52 - 1136 with high standard error which lead to an infinite confidence interval. With pairs between Norwegian and Swedish horses Ne from f_{seg} was between 125 -137 and interestingly with this group Ne decreased when more results were included in the timeline. This might be because the number of related animals was also increased. When all genotyped horses were included to the effective population calculation the f_{seg} Ne had values between 75 - 87. Effective population size for all genotyped horses was therefore larger than in the Norwegian population alone.

The confidence interval (CI) was smallest when we used a timeline of pairs born in ten-year cohorts to calculate effective population (Ne) size from shared segment coancestry, f_{seg} (Table 6). For the Norwegian horses, Effective population size from f_{seg} with this timeline was 63 and confidence interval was only 57-77. Timeline of ten-year cohorts included the most results; with Norwegian horses there were 20,865 results, 2,242 Swedish, 13,867 results from pairs between Norwegian and Swedish horses and in total there were 63,674 results in the timeline. This increase in number of results in the timeline gave a significantly smaller standard error (SE) compared to other timelines with Norwegian, NxS and all genotyped horses. With the Swedish population the increase of results in the timeline did not improve the standard error and the confidence interval still gave infinite results for effective population size.

Multidimensional scaling plot

Relationships between Norwegian and Swedish horses were compared with a multidimensional scaling plot (MDS) in Figure 3. From this we can see that these two populations cluster together with some overlapping. There is variation between Norwegian and Swedish Fjord Horses, but most of the variation in the Swedish can be also seen in the Norwegians. This variation most likely explains results for effective population size in Table 6 where Ne was larger when both populations were included in the calculation than when Ne was calculated with Norwegians only.

Discussion

In this study, different methods were examined for to calculate inbreeding, coancestry and effective population size. The results show that methods based on coancestry can give precise estimates of Ne in rather small sample sizes.

The hypothesis (Kirin et al., 2010, Szpiech et al., 2013 & Curik et al., 2014) is that shorter runs of homozygosity (ROH) and shared segments (SEG) indicate old inbreeding many

generations ago and longer ROHs and SEG mean more recent inbreeding, where there have not been so many generations of recombination between the individual and its common ancestor. Higher inbreeding levels were detected for small segment sizes than for medium or large segments. Small segment sizes performed better than medium sized when ROH inbreeding and SEG coancestry results were regressed on pedigree results. Part of the inbreeding in Fjord Horses is therefore seen only in small ROHs and this indicates that the inbreeding in the populations is partly ancient and had not happened in recent generations. This might be because of the small number of founders and genetic bottlenecks in the breed history. Based on our results, it seems that Fjord Horse breeders have tried to avoid inbreeding when making their mating choices, which has lead to situation where runs of homozygosity and shared genomic segments in genotyped horses are relatively short on average. Kirin et al. (2010) noticed with the human Oceanian population that they had large number of short ROHs and only few longer ROHs. This population had reduced effective population size in their past (genetic bottleneck), but inbreeding was avoided in the recent generations. This is similar to the Fjord Horse population, which has had historical bottlenecks because of wars, isolation and popular sires and for the past decades inbreeding has been avoided in mating choices mostly due to the import of horses from Denmark to Norway, likely introducing longer segments.

When the SNP density was decreased below 65k SNPs, the runs of homozygosity inbreeding, F_{roh}, had significant decrease for average inbreeding. Similar decrease was seen with the shared segments coancestry, f_{seg}, but this was not so drastic as with runs of homozygosity inbreeding, F_{roh}. This indicates imprecise inbreeding and coancestry calculations based on ROHs or shared segments are made when the SNP density is too low. But the SNP thinning in our study was done completely randomly and results for F_{roh} and f_{seg} could be more reliable when used with specifically developed low density SNP BeadChips for horses.

With cattle there have been reliable results with ROH and shared segments even with low SNP densities. With cattle, in these low-density datasets larger segment sizes like 4 Mb (Ferenčaković et al., 2013) or 5 Mb (Purfield et al., 2012) have performed best, but with the Fjord Horses we could see that larger segment sizes couldn't detect the inbreeding or coancestry that well. This might be because in our analyses we randomly thinned the SNP density to desired numbers of SNP, but in cattle they have used designed SNP chips where SNPs are more carefully chosen for their purposes. Breeding systems are also very different

with hobby animals like horses as compared to cattle. In dairy cattle the breeding system is more goal orientated, and in recent years also based on genomic selection (Meuwissen, 2007), while Fjord Horse breeding is based on the subjective judgements of different hobby breeders.

Previous reports on shared segment coancestry have not shown as high pedigree complete generation equivalent, CGE – values as we found here. Rodríguez-Ramilo et al. (2015) had CGE of 6.14 with Spanish Holsteins and Gómez-Romano et al. (2016) had CGE of 4.12, 6.03 and 6.48 in their three genotyped cattle populations (Pinzgauer, Brown Swiss and Tyrolean Grey). The first study had 10,569 genotyped animals and the later one had 219 to 465 genotyped animals. Both researches used Illumina Bovine SNP50 BeadCHip (54k SNPs) to calculate molecular coancestry and later included genotypes from Illumina BovineHD BeadChip (786k SNPs), but they only used SNPs which were included in both BeadChips. A strength in our study is in the higher CGE and the higher number of SNPs used. It seems as though the inbreeding that has occurred in more recent generations tends to improve the correlation between ROH and pedigree inbreeding (McQuillan et al. 2008, Gómez-Romano et al., 2016). Druml et al. (2017) estimated a CGE of 9 in both horse breeds for which they had pedigree data. These breeds also had the lowest ROH inbreeding in their tested breeds. Correlations between pedigrees and ROH were relatively low compared to other studies (Peripolli et al. 2016) and they concluded that this could be due to short ROHs, which also was seen in our Fjord Horses results. Weak correlations in Druml et al. might also be because they pruned out minor allele frequencies (MAF) and didn't do LD or HWE SNP pruning on their data which lead to more SNPs included in the data used. When we tested how different parameters influenced the ROHs we noticed that all these factors decreased correlations between ROH and pedigree (results not included). Even though Druml et al. used same SNP BeadChip that we used in this study there were differences in SNPs-sets used for the calculations; Druml used 589,172 SNPs while we used 485,918 at the highest SNP density.

Previous studies using coancestry IBD matrices from BEAGLE software have used the default parameters for LOD (Rodríguez-Ramilo et al., 2015 & Gómez-Romano et al., 2016). We found that the default LOD parameter gave highly underestimated values for shared segments, and the f_{seg} correlated less with f_{ped} when a default LOD of 3.0 was used. Therefore, the LOD score parameter was set as low as possible in the program. Using higher

LOD score parameters might function better in cattle because of the previously mentioned differences in breeding schemes. It is possible that cattle have longer shared segments in their population compared to horses.

When calculating effective population size, Ne, increase of coancestry (Δf) based on shared segments coancestry, f_{seg}, was the most robust method. With coancestry we can utilize significantly more coancestry results in the timeline as compared to individual inbreeding, and this has the potential to reduce the standard error (SE). This was tested with three different timelines, and the timeline with most results for pairs per ten years performed the best when calculating Ne. But this increase of results in the timeline was not enough with the Swedish population, because we still got a high SE which lead to an infinite confidence interval (CI) for Ne. This means that we need to have more equally distributed coancestry results in the timeline, based on more than few genotyped horses. Methods to calculate Ne from SNP linkage disequilibrium, LD, (Barbato et al., 2015) in genotyped animals might work better than calculating the increase of molecular inbreeding/coancestry over time when the number of genotyped animals is limited.

Effective population size, Ne, based on shared segments for all genotyped horses was 87 with a ten- year timeline. This was larger than the Ne of 63 from the Norwegian population. This indicates that there is genetic variation in the Swedish population that is not seen in the Norwegian. It is possible that other countries exhibit additional variation, comparable to the Swedish population. This might increase the effective population size for the breed. Still, effective population size for Fjord Horses remains low, even with all genotyped horses included. The recommended minimum effective population size in the short term, to avoid issues from inbreeding depression, is 50-100 (Boettcher et al., 2013 & Frankham et al., 2014). It is crucial for future Fjord Horse population survivability that there be a focus on finding genetically different horses for breeding and expanding the breeding population. It would be beneficial to identify genetic differences between Nordic Fjord Horses and Fjord Horses in other countries such as United States of America. Most of the foundation stock was imported to North America between the 1950s and 1960s from Norway (Bhatnagar et al., 2011). The population in North America might therefore represent a valuable genetically different subpopulation due to its isolated history. There are imports between the continents, but the last large import to the U.S. was back in 1988 when 40 Fjord Horses were imported from different bloodlines in Norway.

It would be recommended to use shared segments coancestry, f_{seg} in the breeding programs with Fjord Horses to conserve the existing genetic diversity and to understand how related Fjord Horses from different countries are and what genetic variance could be found from foreign horses. When analyzing the population situation through effective population size, shared segment coancestry can give more precise estimates when the number of genotyped horses is relatively low. This can be improved by including more results in the timeline, as we did with the ten-year cohorts. Practically, f_{seg} would be easier to adapt for use in Fjord Horse breeding, because breeders are used to calculating inbreeding coefficients from pedigree and f_{seg} is in the same scale with the pedigree coancestry. Genomic results would be easier to understand for breeders when they are similar to what they have been using to calculate inbreeding coefficient. There are still practical issues which need to be solved before this can be adapted to everyday use for breeders. Genotyping horses with HD SNP chip, which was used in this study, is very expensive and it is a question whether breeders or stallion owners would be willing to pay the cost for genotyping their horses. Currently there are only two low density SNP chips available for equines; Illumina EquineSNP50 (54k SNPs) which was further developed to Equine SNP70 (65k SNPs) and one high density SNP BeadChip that was used in this study. It would be beneficial for a breeding program to test low density SNP panels if these could be used to calculate shared segments coancestry efficiently or whether the accuracy will be too low, as we saw in our results. There is a chance that a more carefully built low density SNP BeadChip could give more accurate results for calculating coancestry from shared segments than what was seen in our study.

When it is possible to calculate relationships between breeding animals across the world, it would be beneficial for the breed to start adapting OCS in their breeding program. This could help to conserve existing genetic diversity in the whole population. Currently the Norwegian Equine Centre is working on using EVA software to do OCS analyzing for their breeding populations of all Norwegian horse breeds, including Fjord Horse (NEC, 2018). In the future this could be expanded to include breeding animals outside of Norway by using a relationship matrix from shared segment coancestry.

Sources

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in genetics, 6, 109.

Berg, P., Nielsen, J., & Sørensen, M. K. (2006). EVA: Realized and predicted optimal genetic contributions. In Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006 (pp. 27-09). Instituto Prociência.

Bhatnagar, A. S., East, C. M., & Splan, R. K. (2011). Genetic variability of the Norwegian Fjord horse in North America. Animal Genetic Resources/Resources génétiques animales/Recursos genéticos animales, 49, 43-49.

Bjørnstad, G., Gunby, E., & Røed, K. H. (2000). Genetic structure of Norwegian horse breeds. Journal of Animal Breeding and Genetics, 117(5), 307-317.

Bjørnstad, G., Nilsen, N. Ø., & Røed, K. H. (2003). Genetic relationship between Mongolian and Norwegian horses?. Animal Genetics, 34(1), 55-58.

Boettcher, P., Martin, J. F., Gandini, G., Joshi, B. K., & Oldenbroek, J. K. (2013). In vivo conservation of animal genetic resources (Vol. 14, No. 14). FAO.

Bosse, M., Megens, H. J., Madsen, O., Crooijmans, R. P., Ryder, O. A., Austerlitz, F., ... & de Cara, M. A. R. (2015). Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Research, 25(7), 970-981.

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics, 194(2), 459-471.

Cara, M. Á. R., Fernández, J., Toro, M. A., & Villanueva, B. (2011). Using genome-wide information to minimize the loss of diversity in conservation programmes. Journal of animal Breeding and Genetics, 128(6), 456-464.

Cara, M. Á. R., Villanueva, B., Toro, M. Á., & Fernández, J. (2013a). Using genomic tools to maintain diversity and fitness in conservation programmes. Molecular ecology, 22(24), 6091-6099.

Cara, M. Á. R., Villanueva, B., Toro, M. Á., & Fernández, J. (2013b). Purging deleterious mutations in conservation programmes: combining optimal contributions with inbred matings. Heredity, 110(6), 530.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(1), 7.

Clark, S. A., Kinghorn, B. P., Hickey, J. M., & van der Werf, J. H. (2013). The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution, 45(1), 44.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. Livestock Science, 166, 26-34.

Domestic Animal Diversity Information System DAD-IS [Internet]. Fjordhest / Norway (Horse) – [cited 2018 May 5]. Available from: – http://www.fao.org/dad-is/browse-by-country-and-species/en/

Druml, T., Neuditschko, M., Grilz-Seger, G., Horna, M., Ricard, A., Mesarič, M., ... & Brem, G. (2017). Population networks associated with runs of homozygosity reveal new insights into the breeding history of the Haflinger horse. Journal of Heredity, esx114.

Falconer, D. S., & Mackay, T. F. C. (1996). Heritability. Introduction to quantitative genetics, 192, 65-71.

Ferenčaković, M., Sölkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution, 45(1), 42.

Fjord Horse International [Internet]. Presentation of Sweden; 2014 – [cited 2018 February 24]. Available from: http://www.fjordhorseinternational.org/FjHI/index.php/sweden/23-members

Fjord Horse International [Internet]. International Breeding Statistics, 2017 – [cited 2018 May 11]. http://www.fjordhorseinternational.org/FjHI/index.php/the-fjord-horse/international-statistics

Fox, J. (2005). The R Commander: a basic statistics graphical user interface to R. J Stat Softw.

Fox, J. (2016). Using the R Commander: A Point-and-click Interface for R. CRC Press.

Franklin, I.R. (1980) Evolutionary change in small populations. In Conservation Biology: an Evolutionary–Ecological Perspective (Soule´, M.E. and Wilcox, B.A., eds), pp. 135–150, Sinauer Associates

Genome [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information(NCBI); 2004 – [cited 2018 January 6]. Available from: https://www.ncbi.nlm.nih.gov/genome/?term=equine

Gómez-Romano, F., Villanueva, B., de Cara, M. Á. R., & Fernández, J. (2013). Maintaining genetic diversity using molecular coancestry: the effect of marker density and effective population size. Genetics Selection Evolution, 45(1), 38.

Gómez-Romano, F., Villanueva, B., Sölkner, J., Cara, M. A. R., Mészáros, G., Pérez

O'Brien, A. M., & Fernández, J. (2016). The use of coancestry based on shared segments for maintaining genetic diversity. Journal of Animal Breeding and Genetics, 133(5), 357-365.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L. & Pe'er, I. (2009). Whole population, genome-wide mapping of hidden relatedness. Genome research, 19(2), 318-326.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of dairy science, 92(2), 433-443.

Hillestad, B. (2015). Inbreeding determined by the amount of homozygous regions in the genome. (doctoral dissertation). Norwegian University of Life Sciences. Thesis number 2015:41. ISSN: 1894-6402. ISBN: 978-82-575-1287-3

Kamiński, S., Hering, D. M., Jaworski, Z., Zabolewicz, T., & Ruść, A. (2017). Assessment of genomic inbreeding in Polish Konik horses. Polish journal of veterinary sciences, 20(3), 603-605.

Khanshour, A. M. (2013). Genetic Diversity and Population Structure of the Arabian Horse Populations from Syria and other Countries (Doctoral dissertation).

Kirin, M., McQuillan, R., Franklin, C. S., Campbell, H., McKeigue, P. M., & Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. PloS one, 5(11), e13996.

Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E., & Danchin-Burge, C. (2013). Methods to estimate effective population size using pedigree data: examples in dog, sheep, cattle and horse. Genetics Selection Evolution, 45(1), 1.

Lynghaug, F. (Ed.). (2009). The Official Horse Breeds Standards Guide: The Complete Guide to the Standards of All North American Equine Breed Associatio. Voyageur Press (MN).

McCue, M. E., Bannasch, D. L., Petersen, J. L., Gurr, J., Bailey, E., Binns, M. M., ... & Leeb, T. (2012). A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLoS genetics, 8(1), e1002451.

Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., ... & Distl, O. (2015). Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC genomics, 16(1), 764.

Meuwissen, T. & Luo, Z. (1992). Computing inbreeding coefficients in large populations. Genetics Selection Evolution, 24(4), 305.

Meuwissen, T. (2007). Genomic selection: marker assisted selection on a genome wide scale. Journal of animal Breeding and genetics, 124(6), 321-322.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., & MacLeod, A. K. (2008). Runs of homozygosity in European populations. The American Journal of Human Genetics, 83(3), 359-372.

Nestaas, T. Nestaas (2010). Fjordhesten gjennom tidene, Norges Fjordhestlag (In Norwegian)

Norwegian Equine Centre [Internet]. Status for de nasjonale hesterasene – [cited 2018 April 25]. Available from: http://www.nhest.no/Nyheter/2016/August/Status-for-de-nasjonale-hesterasene/

Norwegian Equine Centre [Internet]. Slektskapsmessig avlsrådgivning – [cited 2018 May 6]. Available from: http://www.nhest.no/Nyheter/2017/April/Slektskapsmessig-avlsradgivning/

Olsen, S. L. (2006). Early horse domestication on the Eurasian steppe. Documenting domestication: new genetic and archaeological paradigms, 24

Olsen, H.F., Dolvik, N.I. & Klemetsdal, G. (2018). The populations of the Norwegian and the Swedish Fjord horse: Comparison of rate of inbreeding from observed homozygosity and from pedigree analysis. (unpublished)

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. Animal genetics, 48(3), 255-271.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559-575.

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. Bmc Genetics, 13(1), 70.

Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Hernández, D., & Villanueva, B. (2015). Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. PLoS One, 10(4), e0124157.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K., ... & Finno, C. J. (2017). Developing a 670k genotyping array to tag~ 2M SNPs across 24 horse breeds. BMC genomics, 18(1), 565.

Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 R package for pedigree data. Human heredity, 78(2), 91-93.

Sonesson, A. K., Woolliams, J. A., & Meuwissen, T. H. E. (2010, August). Maximising genetic gain whilst controlling rates of genomic inbreeding using genomic optimum contribution selection. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, 1 (No. 892).

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zöllner, S., Rosenberg, N. A., & Li, J. Z. (2013). Long runs of homozygosity are enriched for deleterious variation. The American Journal of Human Genetics, 93(1), 90-102.

Toro, M. A., Villanueva, B., & Fernández, J. (2014). Genomics applied to management strategies in conservation programmes. Livestock Science, 166, 48-53.

Vazquez, A. I., Bates, D. M., Rosa, G. J. M., Gianola, D., & Weigel, K. A. (2010). An R package for fitting generalized linear mixed models in animal breeding1. Journal of animal science, 88(2), 497-504.

Woolliams, J. A., Berg, P., Dagnachew, B. S., & Meuwissen, T. H. E. (2015). Genetic contributions and their optimization. Journal of Animal Breeding and Genetics, 132(2), 89-99.

Tables and figures

Table 1 SNP densities in different datasets

Dataset	SNP density
Full	485,918
Thin1	389,573
Thin2	311,416
Thin3	249,044
Thin4	199,470
Thin5	159,644
Thin6	127,823
Thin7	102,151
Thin8	81,695
Thin9	65,369
Thin10	52,274
Thin11	41,826
Thin12	33,420

Table 2 Plink 1.9 parameters used in different datasets to find runs of homozygosity

Dataset	Min. homozygous SNP	Max. Gap lenght	Missing SNPs allowed per window
Full	50	100	3
Thin1	50	100	3
Thin2	50	100	3
Thin3	50	100	3
Thin4	25	500	2
Thin5	25	500	2
Thin6	25	500	2
Thin7	15	1000	1
Thin8	15	1000	1
Thin9	15	1000	1
Thin10	15	1000	1
Thin11	15	1000	1
Thin12	15	1000	1

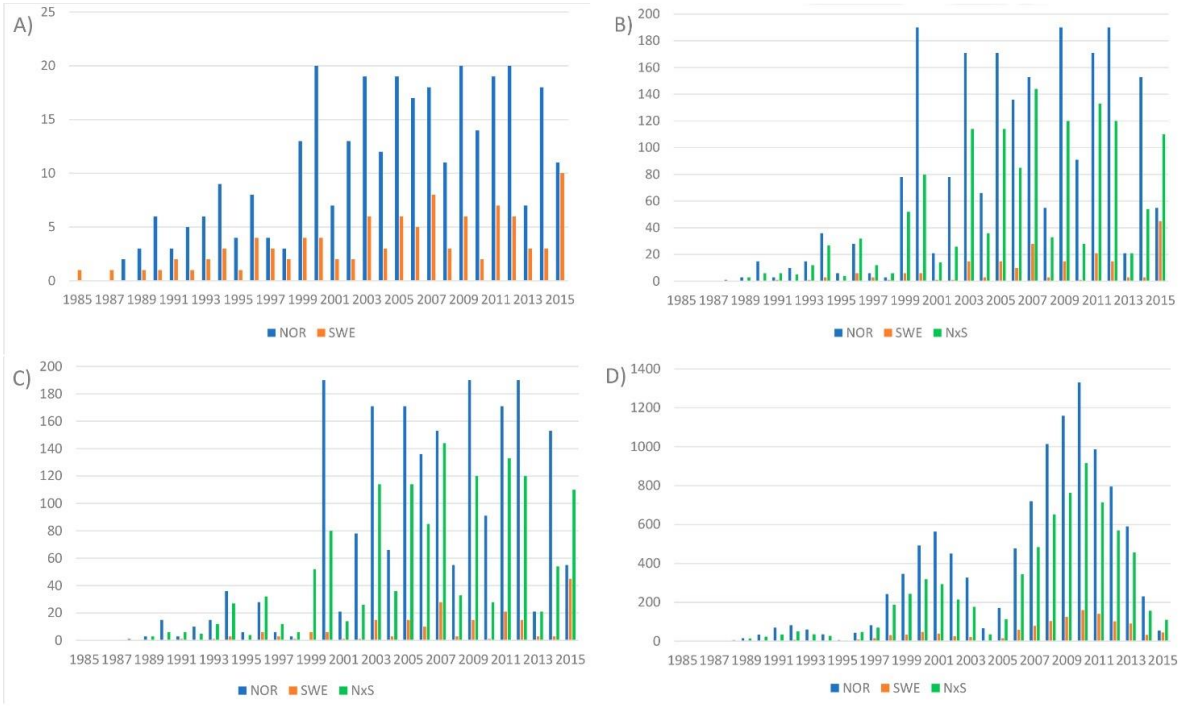

Figure 1 Genotyped A) horses per year, B) horse pairs per year, C) horse pairs per two years and D) horse pairs in ten-year cohorts

Table 3 Pedigree based Complete Generation Equivelant (CGE), average generation interval (L), average inbreeding (F) and coancestry (f)*
**Olsen et al., 2018*

	Norway	**Sweden**
CGE	13.7	12.4
L*	9.08	11.60
F_{ped}	0.077207	0.051667
f_{ped}	0.082029	0.064608

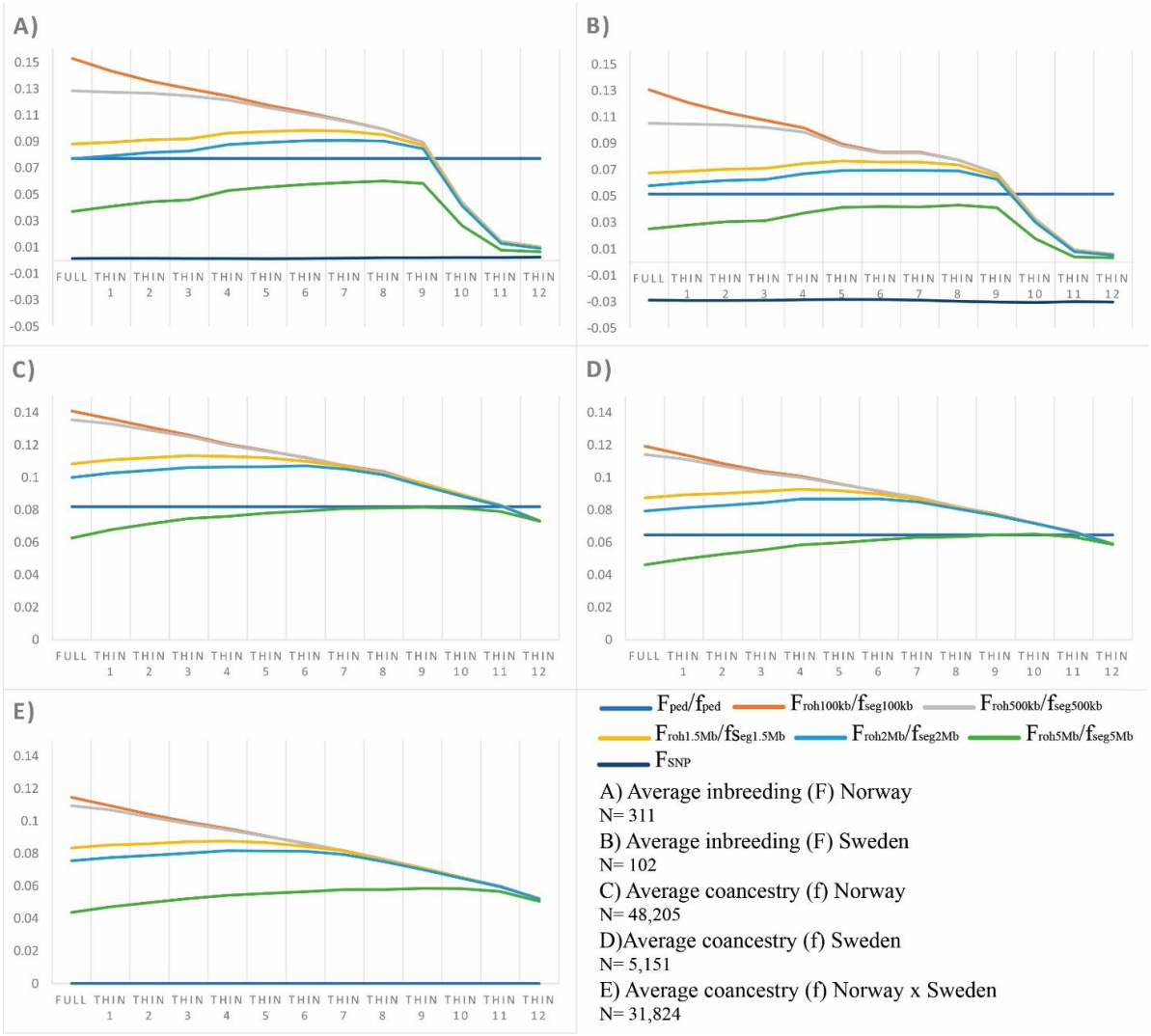

Figure 2 Inbreeding (F) for horses in Norway and Sweden (Figures A and B, respectively) and coancestry (f) between horses in the two countries (Figures C, D and E, respectively as calculated by various method. F_{ped} = pedigree inbreeding, f_{ped}= pedigree coancestry, F_{roh} = runs of homozygosity inbreeding with different segment sizes used, f_{seg} = shared genomic segments coancestry with different segment sizes used and F_{snp} = Molecular homozygosity inbreeding.

Table 4 Intercept (a), regression coefficient (b) and coefficient of determination (R2) between different inbreeding (F) and coancestry (f) calculation methods with Norwegian (No) and Swedish (Swe) genotyped horses and genotyped pairs between the two populations (NxS). F_{snp} = Molecular homozygosity inbreeding, F_{roh} = runs of homozygosity inbreeding with different segment sizes used, F_{ped} = pedigree inbreeding, f_{hom} = molecular homozygosity coancestry, f_{seg} = shared genomic segments coancestry with different segment sizes used and f_{ped}= pedigree coancestry.

Regression of			a			b			R^2		
			Nor	Swe	NxS	Nor	Swe	NxS	Nor	Swe	NxS
F_{snp}	on	F_{ped}	0.063	0.062	-	0.801	0.644	-	0.293	0.151	-
$F_{roh100kb}$	on	F_{ped}	-0.105	-0.107	-	0.761	0.623	-	0.287	0.137	-
$F_{roh500kb}$	on	F_{ped}	-0.080	-0.080	-	0.723	0.603	-	0.255	0.127	-
$F_{roh1.5Mb}$	on	F_{ped}	-0.046	-0.042	-	0.579	0.525	-	0.169	0.103	-
F_{roh2Mb}	on	F_{ped}	-0.038	-0.033	-	0.533	0.505	-	0.152	0.106	-
$F_{roh100kb}$	on	F_{snp}	-0.165	-0.167	-	0.921	0.958	-	0.921	0.896	-
$F_{roh500kb}$	on	F_{snp}	-0.136	-0.138	-	0.911	0.948	-	0.889	0.870	-
$F_{roh1.5Mb}$	on	F_{snp}	-0.091	-0.094	-	0.817	0.844	-	0.741	0.751	-
F_{roh2Mb}	on	F_{snp}	-0.079	-0.082	-	0.762	0.770	-	0.687	0.691	-
f_{hom}	on	f_{ped}	-1.417	-1.419	-	1.410	1.531	-	0.703	0.866	-
$f_{seg100kb}$	on	f_{ped}	-0.068	-0.062	-	0.983	0.952	-	0.636	0.819	-
$f_{seg500kb}$	on	f_{ped}	-0.063	-0.057	-	0.973	0.950	-	0.630	0.817	-
$f_{seg1.5Mb}$	on	f_{ped}	-0.038	-0.031	-	0.897	0.886	-	0.560	0.784	-
f_{seg2Mb}	on	f_{ped}	-0.032	-0.024	-	0.863	0.858	-	0.533	0.774	-
$f_{seg100kb}$	on	f_{hom}	0.838	0.792	0.784	0.644	0.603	0.601	0.770	0.891	0.562
$f_{seg500kb}$	on	f_{hom}	0.834	0.797	0.777	0.637	0.603	0.593	0.764	0.890	0.551
$f_{seg1.5Mb}$	on	f_{hom}	0.794	0.762	0.735	0.591	0.560	0.546	0.687	0.848	0.459
f_{seg2Mb}	on	f_{hom}	0.768	0.740	0.704	0.568	0.540	0.520	0.652	0.829	0.422

Table 5 Mean, Range, variance (var) and coefficient of variation (CV) for different methods used to calculate inbreeding (F) and coancestry (f) for Norwegian, Swedish and pairs from both populations (NxS). F_{snp} = Molecular homozygosity inbreeding, F_{roh} = runs of homozygosity inbreeding with different segment sizes used, F_{ped} = pedigree inbreeding, f_{hom} = molecular homozygosity coancestry, f_{seg} = shared genomic segments coancestry with different segment sizes used and f_{ped}= pedigree coancestry.

	Mean			**Range**			**Variance**			**CV**		
	Nor	Swe	NxS	Nor	Swe	NxS	Nor	Swe	NxS	Nor	Swe	NxS
F_{ped}	0.077	0.052	-	0.005	0.191	0.008	0.129	-	-	0.00042	0.00022	-	0.264	0.285	-
F_{snp}	0.002	-0.029	-	-0.075	0.129	-0.091	0.049	-	-	0.00105	0.00066	-	21.224	NA	-
$F_{roh100kb}$	0.153	0.131	-	0.087	0.257	0.074	0.197	-	-	0.00070	0.00049	-	0.173	0.169	-
$F_{roh500kb}$	0.129	0.105	-	0.059	0.238	0.042	0.172	-	-	0.00075	0.00052	-	0.213	0.217	-
$F_{roh1.5Mb}$	0.088	0.068	-	0.006	0.197	0.004	0.135	-	-	0.00080	0.00052	-	0.320	0.338	-
F_{roh2Mb}	0.077	0.058	-	0.002	0.187	0.001	0.121	-	-	0.00077	0.00048	-	0.359	0.378	-
f_{ped}	0.082	0.065	-	0.030	0.344	0.017	0.311	-	-	0.00063	0.00104	-	0.306	0.499	-
f_{hom}	0.785	0.782	0.778	0.761	0.885	0.763	0.873	0.760	0.868	0.00009	0.00015	0.00003	0.012	0.015	0.007
$f_{seg100kb}$	0.141	0.119	0.115	0.060	0.439	0.055	0.379	0.047	0.396	0.00085	0.00120	0.00036	0.207	0.291	0.094
$f_{seg500kb}$	0.136	0.114	0.110	0.055	0.427	0.049	0.375	0.038	0.390	0.00085	0.00120	0.00036	0.215	0.304	0.094
$f_{seg1.5Mb}$	0.108	0.087	0.083	0.016	0.401	0.016	0.337	0.011	0.366	0.00088	0.00119	0.00039	0.274	0.394	0.099
f_{seg2Mb}	0.100	0.079	0.076	0.012	0.389	0.008	0.327	0.005	0.354	0.00088	0.00115	0.00039	0.296	0.427	0.099

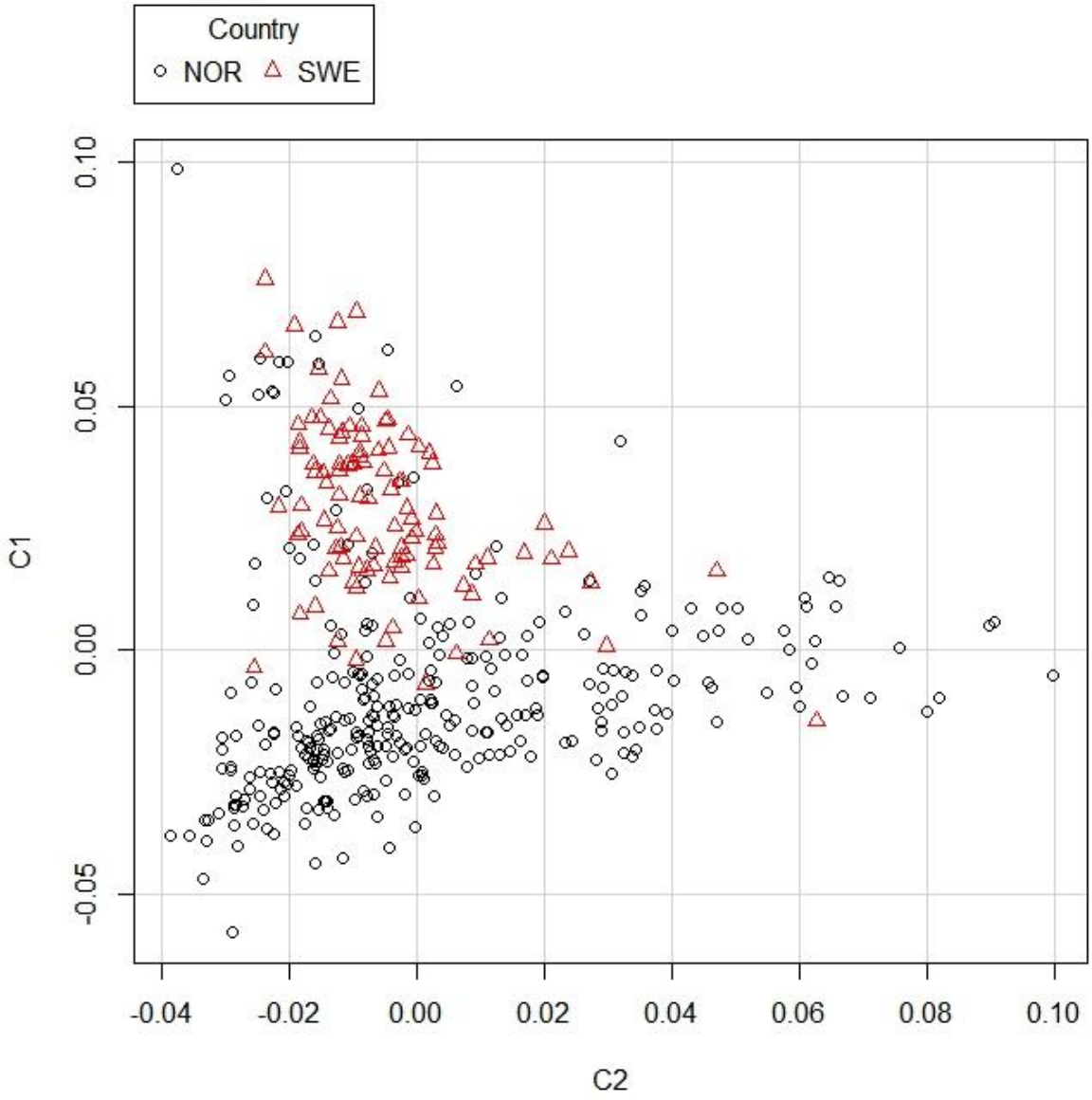

Figure 3 Multidimensional scaling plot based on IBS average distances between genotyped horses

Table 6 Regression coefficient (b), standard error (SE), effective population size (Ne) and confidence interval (CI) calculated for Norwegian (No), Swedish (Swe) and pairs between these two populations (NxS) and for all genotyped horses. F_{ped}= Inbreeding from full pedigree data, $F_{roh100kb}$ = runs of homozygosity inbreeding with 100kb segment size, F_{snp} = Molecular homozygosity inbreeding, $f_{seg100kb}$ = shared genomic segments coancestry with 100kb segment size and f_{hom} = molecular homozygosity coancestry.

	Regression Coefficient (b)				**Standard Error (SE)**			
	Nor	Swe	NxS	All	Nor	Swe	NxS	All
F_{ped}	-0.0007719	-0.0001603	-	-	0.0000053	0.0000031	-	-
N	26446	14677	-	-	-	-	-	-
$F_{roh100kb}$	-0.00078	-0.00028	-	-0.00064	0.00026	0.00034	-	0.00023
F_{snp}	-0.00075	-0.00031	-	-0.00062	0.00027	0.00034	-	0.00024
N	311	102	-	413	311	102	-	413
Pairs per year	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	-0.00105	-0.00082	-0.00035	-0.00065	0.00017	0.00051	0.00010	0.00012
f_{hom}	-0.00049	-0.00089	-0.00009	-0.00026	0.00024	0.00080	0.00012	0.00016
N	2116	206	1397	3719	2116	206	1397	3719
Two years	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	-0.00090	-0.00017	-0.00034	-0.00061	0.00010	0.00029	0.00006	0.00007
f_{hom}	-0.00040	-0.00023	-0.00006	-0.00024	0.00013	0.00043	0.00007	0.00009
N	5821	565	3839	10225	5821	565	3839	10225
10 years	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	-0.00087	-0.00004	-0.00039	-0.00056	0.00005	0.00017	0.00003	0.00004
f_{hom}	-0.00043	0.00007	-0.00016	-0.00023	0.00007	0.00024	0.00004	0.00005
N	20865	2242	13867	36974	20865	2242	13867	36974
	Effective population size (Ne)				**Confidence Interval (CI)**			
	Nor	Swe	NxS	All	Nor	Swe	NxS	All
F_{ped}	71	269	-	-	70—72	259—279	-	-
N	26446	14677	-	-	26446	14677	-	-
$F_{roh100kb}$	70	155	-	76	43—201	30— ∞	-	45—250
F_{snp}	74	141	-	78	43—255	31— ∞	-	45—317
N	311	102	-	413	311	102	-	413
Pairs per year	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	52	52	137	75	40—76	24— ∞	89—302	55—116
f_{hom}	113	48	525	184	57—3759	18— ∞	144— ∞	84— ∞
N	2116	206	1397	3719	2116	206	1397	3719
Two years	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	61	260	141	79	51—77	58— ∞	105—214	65—101
f_{hom}	139	189	792	202	84—401	40— ∞	233— ∞	117—731
N	5821	565	3839	10225	5821	565	3839	10225
10 years	Nor	Swe	NxS	All	Nor	Swe	NxS	All
$f_{seg100kb}$	63	1136	125	87	57—72	119— ∞	107—152	77—100
f_{hom}	129	∞	293	209	98—187	106— ∞	194—602	149—347
N	20865	2242	13867	36974	20865	2242	13867	36974