

## **Sluttrapport H-14-47-015**

### *Introduction*

The Fjord horse and its areas of use has the last decades been changing from working purposes to sports and leisure purposes, also facing the strong competition from imported, specialized horse breeds in these market segments. The greatest challenge for the modern Fjord horse, as with other small, national horse breeds, is to improve the competitive edge and develop the demand in the market. This is essential to increase the population size, which is necessary to decrease the pressure on the genetic resources in the breed.

Norway is the origin country for this breed, and the Norwegian population of the Fjord horse consists only of about 5000 animals, with less than 150 foals born per year (Norwegian Equine Centre, 2016), which in itself requires surveillance and management of the genetic variation to ensure a future live population of this national breed. However, there are several small populations of the Fjord horse in different countries all over the world, which all together make a rather large population. There is no organized breeding of Fjord horses across countries today, and the breeding goals in the different populations are not coordinated. There is also different rules for registration, which make it difficult to trace the pedigree on individuals to calculate the relationship *between* populations.

To be able to manage the total population of Fjord horse in a sustainable way, there is a need to gain knowledge of the subpopulations' genetic variation and their relationship to other subpopulations, and in addition strengthen the international organization structure to ensure the possibilities for an international information base. Fjord Horse International (FHI) is an international forum where member countries can exchange information about breeding, sport and education regarding the purebred Fjord horse worldwide (FHI, 2017). FHI has today 13 member countries, where Norway holds the president of the board. The limited formalized cooperation between the countries is a drawback for the development of the breed, but FHI will play an important role in development of the cooperation and the organization structure.

An aim for an international database must be to provide possibilities for a common breeding goal for the global population with establishment of an international BLUP evaluation and exchange of genetic information across the subpopulations of the Fjord horse. Based only on pedigree information, this task will be very challenging, if not impossible, due to the different registration routines and different id-systems. Therefore, it is necessary to explore the possibilities by utilizing genomic information, based on Single Nucleotide Polymorphisms (SNPs). The Norwegian and the Swedish population of the Fjord horse were chosen as example populations for this project, and the aim was to calculate the genetic variation in these populations with different methods, to prepare for development towards an international studbook and common breeding value estimation.

### *Material and methods*

Pedigree data on all registered Norwegian Fjord horses was received from the Norwegian Equine Centre spring 2016, and consisted of 26,462 individuals born in the period 1857 to 2015. Of these, 206 individuals (<1%) were originally registered in a foreign country, of which 63% were originally registered in Denmark. Others were from either Sweden (15%), Germany (15%), Netherlands (6%) or others (1%).

In the Norwegian studbook, the foreign horses exist because they are imported or they have progeny registered in Norway. Foreign horses are automatically registered with 5 generations of pedigree information, if available. The original file was edited for obvious or logical errors, and individuals only occurring as parents in the pedigree were added with individual information, resulting in a file of 25,364 individuals. Missing (<0.5%) or erroneous birth years were re-constructed either by use of external information, if available, or by making fictive birth years. The fictive birth years were set to one year older than the eldest offspring of the individual. Then, average inbreeding coefficient per birth year were calculated for all registered animals.

A data file consisting of 14,595 registered Swedish Fjord horses was received from the Swedish Horse Breeding Society in September 2016. As much as 27 % of these individuals had unknown birth year, and thus there were not made any attempt to replace these with fictive birth years or calculate average coefficient of inbreeding per year. Of all registered Swedish animals, over 30% originated from another country than Sweden. Of these, 17% were originally from Norway and 13% were from Denmark. The total pedigree file were further used to trace pedigree information of the test animals.

In both populations, the inbreeding coefficients were calculated for all the individuals in the pedigree of the test animals.

Blood samples from 365 Norwegian and 103 Swedish Fjord horses were collected in the period December 2015 to March 2016. The samples were either; collected by local veterinarians, who froze the samples and sent them to the university (NMBU) by mail, or collected by a team from the project, travelling around in the eastern part of Norway. Of these, 327 samples were analysed for genotype, whereof 224 Norwegian and 103 Swedish samples. Affymetrix' high-density Axiom™ Equine Genotyping Array was used for the analyses, featuring 670,796 markers. The SNP genotypes were called using the software Axiom Analysis Suite (v. 2.0.0.35). A quality control (QC) was accomplished with two levels of filtering; i) Dish QC, measuring the number of non-polymorphic loci (threshold 0.82; rejecting three Norwegian samples), and ii) QC Call Rate, calculating the percent of genotypes assigned to a subset of SNPs (threshold 0.97; rejecting five Norwegian samples), passed 328 samples (157 males and 171 females) and 506,128 SNPs. In addition, a third filter was applied, where SNPs were discharged due to the following criteria: i) those deviating from Hardy-Weinberg equilibrium ( $p \leq 10^{-5}$ ; rejecting 18,425 markers) and ii) those failing missingness test ( $GENO > 0.05$ ; rejecting 5,389 SNPs), resulting in total 482,314 SNPs after the three steps of quality controls.

The inbreeding coefficients based on homozygosity,  $F_{\text{snp}}$ , were calculated using PLINK 1.9 (Purcell et al., 2007), and the function *-het*.  $F_{\text{snp}}$  was calculated as:

$$F_{\text{snp}} = \frac{O(\text{HOM}) - E(\text{HOM})}{N(\text{NM}) - E(\text{HOM})}$$

where O(HOM) is the observed homozygosity for the individual, N(NM) is the number of non-missing genotypes for the individual and E(HOM) is the expected homozygosity for the individual.

One of the Norwegian horses was apparently not related to the rest of the population, due to inbreeding coefficient of zero, and was omitted from further analysis, giving 224 Norwegian horses.

The pedigree of the Norwegian and the Swedish Fjord horses was traced back to the founders. A founder was defined as an animal with both parents unknown. Animals with only one parent missing were defined as half-founders, with the unknown parent considered as an unknown founder (Boichard et al., 1997). The accumulated data file from the reference population was used to calculate the complete generation equivalent (CGE):

$$CGE = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^{n_j} \frac{1}{2^{g_{ij}}} ,$$

where N is the number of individuals in the reference population,  $n_j$  is the number of ancestors generated for animal j and  $g_{ij}$  is the number of generations between individual j and its ancestor i (Boichard et al., 1997). CGE can be interpreted as the number of generations in a comparable complete pedigree, illustrating the depth of the pedigree data.

Then, duplicates were removed and the pedigree file of the reference population was renumbered. Individual inbreeding coefficients were calculated using SAS® Software version 9.4.

To quantify the rate of genetic drift, the rate of inbreeding ( $\Delta F$ ) in the two test populations was calculated as:

$$\Delta F = 1 - e^{\beta_1 \pm s.e.} ,$$

where  $\beta_1$  is the regression coefficient from a log regression on birth year when the response variable,  $y_i$ , is either based on the individual inbreeding coefficient from pedigree information,  $F_{ped}$  (1), or from SNP information,  $F_{hom}$  (2):

- 1)  $\ln(1 - F_{ped_i}) = y_i = \beta_0 + \beta_1 \cdot X_i + e_i$
- 2)  $\ln(1 - F_{hom_i}) = y_i = \beta_0 + \beta_1 \cdot X_i + e_i$

where  $\beta_0$  is an unknown constant,  $\beta_1$  is the regression coefficient associated with the regressor  $X_i$ , which is the birth year of individual  $i$ , and  $e_i$  is the random error term.

To estimate the relationship between the  $F_{ped}$  and the  $F_{hom}$ , a simple linear regression model was used:

$$\ln(1 - F_{hom}) = y_i = \beta_0 + \beta_1 \cdot X + e_i ,$$

where  $\beta_0$  is an unknown constant,  $\beta_1$  is the regression coefficient associated with the regressor  $X_i$ , which is the log-transformation of the individual inbreeding coefficient based on pedigree information and  $e_i$  is the random error term.

### *Results and discussion*

Figure 1 shows the average coefficient of inbreeding for each year in the period 1857 to 2016 in the Norwegian Fjord horse population, based on registered animals. From the early 1990's the level of inbreeding is flattening out, though with some fluctuations the last half decade, compared to the relatively rapid increase from the 1960's onwards. The flattening of the level of inbreeding is concurrent with the import of Danish fjord horses in this breed. These imports has their origin from Norway, but

mostly there are more than five generations back to the Norwegian ancestors, making the imports founders to the population, and thus affecting the level of inbreeding in a positive way.

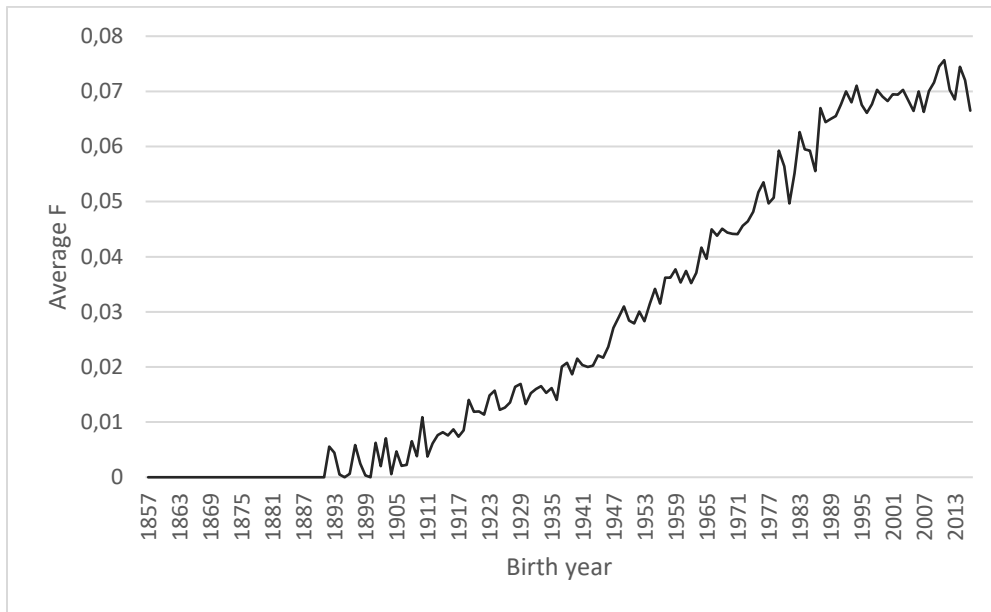


Figure 1: Average coefficient of inbreeding per birth year for all registered Norwegian Fjord horses in the period 1857 to 2016.

In the Swedish Fjord horse studbook, too many individuals lacked information of birth year to be able to calculate the average inbreeding coefficient per year of birth.

Table 1 shows the quality of the pedigree of the test animals in the two populations. Both the Norwegian and the Swedish Fjord horse seemingly have quite complete pedigree information, with 99.1% and 98.6% known ancestors in the fifth generation, respectively.

Table 1: Pedigree depth in the nested pedigree of the Norwegian and the Swedish individuals in the test population.

	NOR	SWE
# animals in reference population	224	103
Average number of ancestors per animal	79,864	44,344
% of known ancestors in generation		
1	1	1
3	0,9978	0,9976
5	0,9905	0,9860
7	0,9644	0,9505
10	0,8601	0,7386
15	0,2722	0,1629
20	0,0057	0,0026
Max number of generations generated	27	27
Complete generation equivalent (CGE)	12.64	11.48

The complete generation equivalent shows that the Norwegian pedigree contains information equivalent to 12.6 generations, whereas the Swedish covers 11.5 generations (Table 1). The complete pedigrees should make a solid foundation to calculate the level of inbreeding within population, but the challenge is the imports in both populations. The imports are to some extent in relation to the existing population, but the pedigree link is often too many generations back to be able to trace the pedigree back to the real, Norwegian founders. This will potentially underestimate the level of inbreeding in both populations.

The age distribution of the test animals in the two populations are given in Figure 2. The birth years of the test animals stretched out over a time span of 30 years, and with a natural decrease in the presence of the eldest birth years. The Swedish population had though a quite large fraction of animals born in the latest years, but with exception of the some of the eldest years, all years were present in the test material of both populations, giving a good foundation for the analysis of the genetic markers.

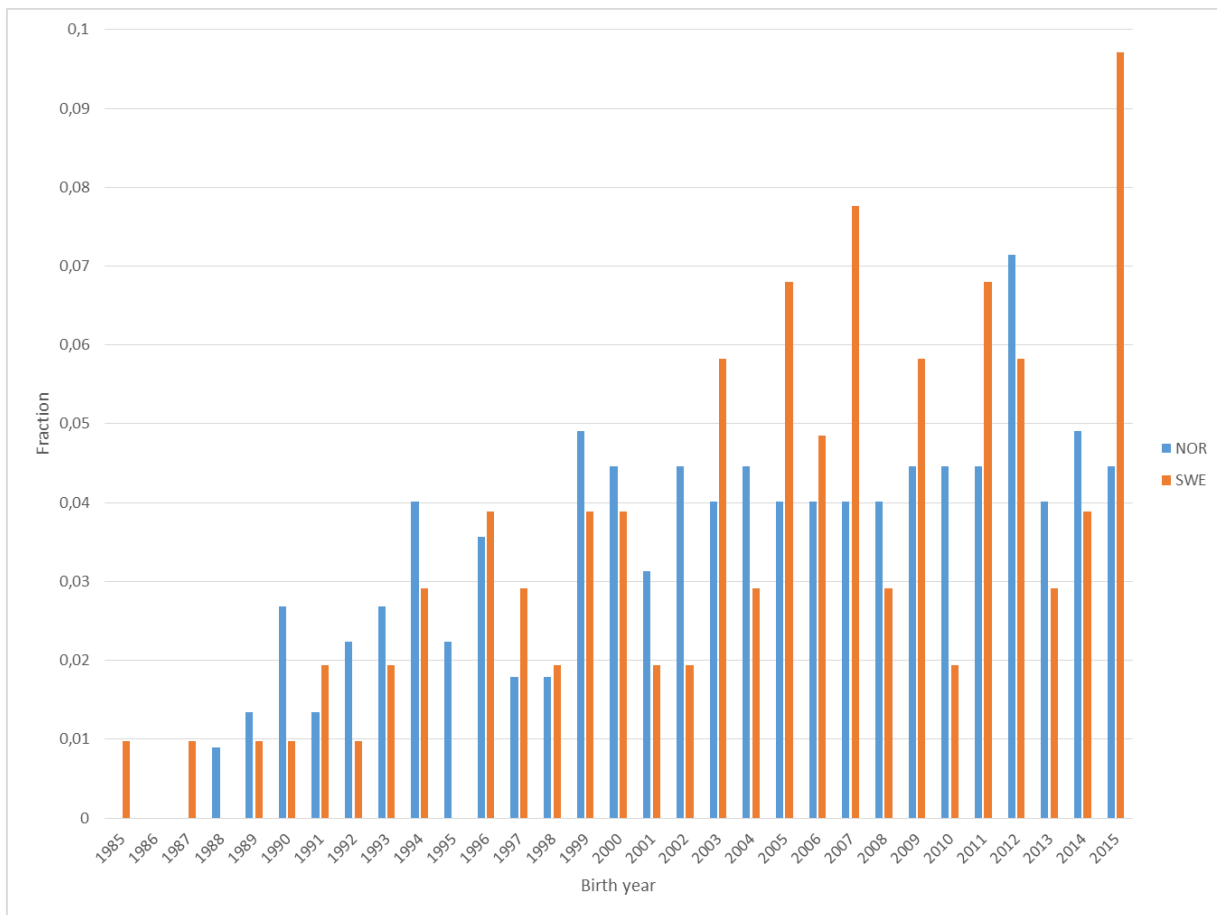


Figure 2: The age distribution of the test animals, shown through number of animals born per year in the Norwegian (N=224) and the Swedish Fjord horse population (N=103).

There were over twice as many test animals from the Norwegian population than from the Swedish, due to practical and economic reasons regarding sampling the blood tests (Table 2). Although the average number of ancestors per animal in the pedigree of the Swedish test animals is almost half of the corresponding number in the Norwegian horses (Table 1), the total number of unique animals in the pedigree is larger in the Swedish population than in the Norwegian (Table 2). The reason for this is probably that the Norwegian Fjord horse has a higher level of inbreeding, and thus has fewer unique individuals in the pedigree.

*Table 2: Number of animals, average coefficient of inbreeding (F), both pedigree based ( $F_{ped}$ ) and based on molecular information ( $F_{hom}$ ), and generation interval (L) in the test populations of Norwegian and Swedish Fjord horses.*

	<b>NOR</b>	<b>SWE</b>
# test animals	224	103
# animals in pedigree	2,659	2,673
$F_{ped} \pm$ s.d.	0.0760 $\pm$ 0.0205	0.0514 $\pm$ 0.0146
$F_{hom} \pm$ s.d.	0.0035 $\pm$ 0.0330	-0.0265 $\pm$ 0.0259
L	9.08	11.60

The Norwegian population has approximately 2.5 years shorter generation interval than the Swedish (Table 2). In Norway, there has been a focus on the importance of using young stallions, as an attempt to even out the progeny groups and perhaps retire some of the old stallions with large genetic contributions to the population. This initiative could be an explanation of the shorter generation interval. Still, it is important not to focus on shorter generation interval in itself, as this can lead to loss of genetic variation due to fast generation shifts, but rather as a tool to include more stallions in breeding.

*Table 3: The parameter estimates with belonging standard errors and t-values from the regression  $\ln(1-F_{ped,i}) = \beta_0 + \beta_1 * (\text{birth year})_i + e_i$ , where  $F_{ped,i}$  is the individual inbreeding coefficient based on pedigree.*

<b>Variable</b>	<b>DF</b>	<b>Parameter Estimate</b>	<b>Standard Error</b>	<b>t-value</b>	<b>Pr &gt;  t </b>	<b>R<sup>2</sup></b>
<b>NOR (#2659)</b>						
Slope	1	-0.0007275	0.0000147	-49.29	<0.0001	0.478
<b>SWE (#2673)</b>						
Slope	1	-0.0001748	0.0000092	-19.00	<0.0001	0.119

In Table 3 and Table 4 the parameter estimates constituting the basis for the calculation of the effective population size based on either pedigree information or information from genetic markers and level of homozygosity (Table 5). The low coefficient of determination for the Swedish horses in Table 3, says that the data is not very close to the fitted regression line. This is supported by Figure 3, showing the fit plot for  $Y = \ln(1-F_{ped}) = \beta_0 + \beta_1 * (\text{birth year})_i + e_i$  for both the Norwegian (a) and the Swedish (b) horses.

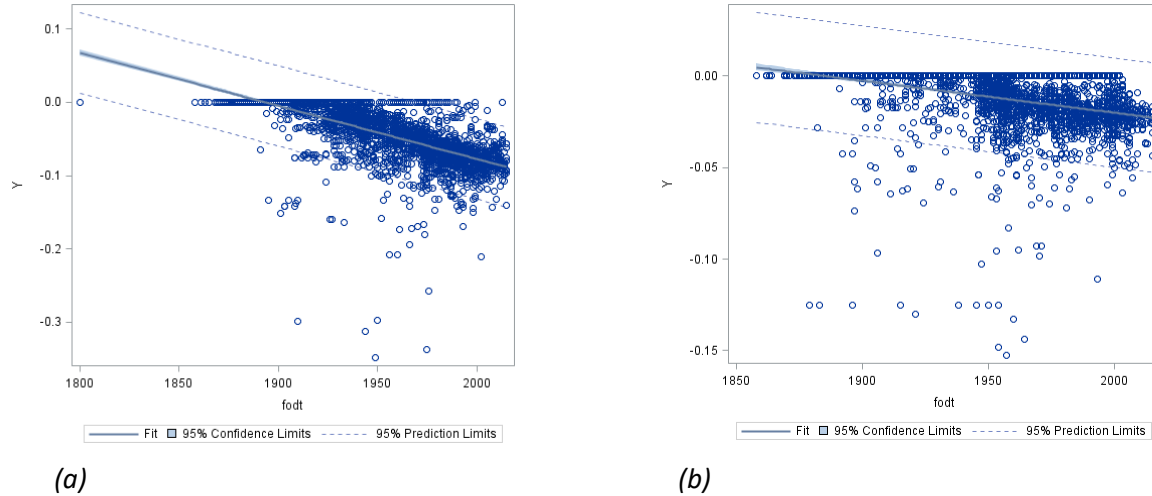


Figure 3: Fit plot of the regression  $Y = \ln(1-Fped_i) = \beta_0 + \beta_1 * (\text{birth year})_i + e_i$ , where  $Fped_i$  is the individual inbreeding coefficient based on pedigree.

The Swedish test population has a larger spread in the observations, and several individuals with much higher inbreeding coefficients than the average for the age class, causing the low  $R^2$ -value.

Table 4: The parameter estimates with belonging standard errors and t-values from the regression  $\ln(1-Fsnp_i) = \beta_0 + \beta_1 * (\text{birth year})_i + e_i$ , where  $Fsnp_i$  is the individual inbreeding coefficient based on homozygosity.

Variable	DF	Parameter Estimate	Standard Error	t-value	Pr >  t	R <sup>2</sup>
<b>NOR (#224)</b>						
Slope	1	-0.0006629	0.0002970	-2.23	0.0266	0.2488
<b>SWE (#103)</b>						
Slope	1	-0.0002942	0.0003389	-0.87	0.3874	0.1562

With basis in the parameter estimates from Table 3 and Table 4, the rate of inbreeding, both based on pedigree and on homozygosity, was calculated together with the corresponding effective population size (Table 5). The low standard error on the method based on pedigree information, the range of the effective population size is limited, although the Swedish population shows a wider range than the Norwegian does, due to less data. With the method based on homozygosity, the range of the effective population size is far wider, especially in the Swedish population.

Table 5: Rate of inbreeding ( $\Delta F = 1 - e^{\beta \pm 1.96 * s.e. (\beta)}$ , where  $\beta$  is the slope from the regression used in Table 3 and 4), both based on pedigree ( $F_{ped}$ ) and based on homozygosity ( $F_{snp}$ ), and the respective effective population size ( $N_e$ ), including the maximum and minimum range, amongst the test populations of Norwegian and Swedish Fjord horses.

	NOR	SWE
$\Delta F_{ped}$	0.0007242	0.0001748
$N_e$ (ped-min)	73	224
$N_e$ (ped)	76	246
$N_e$ (ped-max)	79	275
$\Delta F_{hom}$	0.0006627	0.0002942
$N_e$ (hom-min)	44	45
$N_e$ (hom)	83	147
$N_e$ (hom-max)	682	$\infty$

Comparing the results in Table 5, with information based on pedigree and information based on homozygosity, the effective population size with the two methods in the Norwegian Fjord horse is quite equal; around 80. The Norwegian pedigree is quite deep and well structured, giving small sources of errors. In addition, the impact of imports with lacking data to the Norwegian population are minor in the big context (< 1% of foreign origin), giving comparable results with the two methods. On the other hand, the results in the Swedish population have larger standard errors, due to less data, increasing the range of the effective population size and increases the distance between the results based on pedigree and homozygosity. The lack of information in pedigree seems to contribute to a solid overestimation of the effective population size based on pedigree. Nevertheless, there are undoubtedly more genetic variation in the Swedish Fjord horse population than in the Norwegian, which is in correspondence with the level of imports in the two populations.

Table 6: The parameter estimates with belonging standard errors and t-values from the regression  $\ln(1-F_{hom_i}) = \beta_0 + \beta_1 * \ln(1-F_{ped_i}) + e_i$ , where  $F_{hom_i}$  is the individual inbreeding coefficient based on homozygosity and  $F_{ped_i}$  is the individual inbreeding coefficient based on pedigree.

Variable	DF	Parameter Estimate	Standard Error	t-value	Pr >  t	R <sup>2</sup>
<b>NOR (#224)</b>						
Slope	1	0.7471	0.0871	8.58	<0.0001	0.2488
<b>SWE (#103)</b>						
Slope	1	0.6452	0.1492	4.32	<0.0001	0.1479

Table 6 shows a comparison between the two methods through a regression analysis, which corresponds to a correlation of approximately 0.5 in the Norwegian population and 0.3 in the Swedish population. This is not a very strong relationship, pointing out the possibilities that there are different factors affecting the two methods, such as for instance old homozygosity when using information based



on genetic markers. In addition, there are relatively few samples in the study, exhibiting the sensitivity in the method with genetic markers, according to number of records. The Swedish sample was less than half of the Norwegian sample, and the small sample size gave a large range of the effective population size. The sampling and chipping of the SNP's is a quite expensive procedure, which will be a challenge in further work including several populations of the Fjord horse, given the sensitivity of the method regarding number of records. To follow up the weak relationship between the methods using pedigree information and observed homozygosity, we are now exploring a method using linkage disequilibrium, which will strengthen the possibility for getting the study published. In the long term, these results points towards the need of a simulation study, which would be a natural next step for further research on this topic.

In 2017, the sample was expanded with another chip, including 96 more individuals to the study. This data set was used to explore the use of runs of homozygosity in calculations of inbreeding and coancestry within and between the Norwegian and the Swedish population. The preliminary results show that the inbreeding estimated with runs of homozygosity is close to the inbreeding calculated as probabilities utilizing pedigree information ('identity-by-descent'), when using segment lengths of size 1.5Mb to 2Mb. The preliminary results also show that calculating the genomic relationship matrix is feasible by using coancestry based on shared segments. Some more quality checks are necessary on the last part of this study for final conclusions.

### *Conclusion*

The Norwegian and the Swedish Fjord horse has different impact of imports, giving different levels of genetic variation, where the Swedish population has a larger effective population size than the Norwegian one. There is relatively low relationship between the methods based on pedigree data and by using genetic markers, but still the estimates of the effective population size points in the same direction. The method using genetic markers is clearly sensitive for number of records, which addresses the need for the expansion with a third method using linkage disequilibrium. The Swedish population has a high impact of the Norwegian Fjord horse, and the final part of this work will lead to a recommendation for use of method to assess the coancestry between sub-populations of the Fjord horse, and such give the basis for development of a common global maintenance of the genetic variation in the Fjord horse.

### *Publication plan*

Paper 1: "The Norwegian and the Swedish Fjord horse: Comparison of rate of inbreeding from observed homozygosity, linkage disequilibrium and from pedigree analysis", submitting to Journal of Animal Breeding and Genetics within end of April 2018.

Paper 2: "Utilizing runs of homozygosity for build-up of genomic relationship between the Norwegian and the Swedish Fjord horse", submitting to Journal of Animal Breeding and Genetics within medio June 2018.

Master thesis: M. Høiseth, “Genetic variation and colour genetics of the Norwegian Fjord horse”, finished May 2017.

Master thesis: S. Tenhunen, “Use of genome-wide association study for estimations of inbreeding and coancestry in the Norwegian and Swedish Fjord horse population”, finished May 2018.

Bachelor thesis: M.M. Regland, “Colour preferences on the Norwegian Fjord horse”

Popular science 1: “Innavl hos fjordhest. En sammenlikning av innavl beregnet ved analyse av frekvensen homozygote områder på genomet og slektskapsdata beregnet på norske og svenske fjordhester”. Will be submitted to Fjordhesten, the official organ for the Norwegian Fjord Horse Association, and on their web: [www.fjordhest.no](http://www.fjordhest.no)

Popular science 2: “Inavel hos fjordhäst. En beräkning av inavelsgrad grundad på analys av frekvensen homozygota områden på genomet samt släktskapsdata beräknad på stamtavlor för fjordhästar i Norge och Sverige. UNSI tidningen Lill-Blakken, medlemstidning för Svenska fjordhästföreningen, och på Svenska fjordhästföreningens hemsida [www.fjordhastforeningen.se](http://www.fjordhastforeningen.se)