Genotyping and reference population

Genotyping *Reference population*	• Strategy for minimizing relationship of genotyped animals - Relationship between genotyped animals - Genotyped animals/bull • Plan for genotyping in the herds with high qulity phenotypes - Percentage of high quality phenotypes that are genotyped • Cost per genotype
Reference population *Size*	• Number of genotyped animals in the reference population (females and bulls) for all breeding goal traits • Number of animals contributing information to the reference population in the future total merit - Percentage of the population • Accuracy of all traits • Number of animals (females and bulls) added/year
Reference population *Composition*	• Relationship between animals - Production population, breeding animals and reference population • Number of tested daughters/sire • Variation in numbers of tested daughters/sire
Reference population *Regularly renewed*	• Age of reference population - Age structure in years per trait (females and males) • Average accuracy of all traits • Renewed regularly - Track how/when/why - Effect on accuracy
Reference population *Combining outside the Nordic countries*	• Yearly evaluated - Effect on accuracy - Genetic gain/cost • Yearly check recording strategies changes in different countries

Genotyping *Reference population*	
Strategy for minimizing relationship of genotyped animals	Strategy for minimizing relationship of genotyped animals.
Relationship between genotyped animals	Track relationship with best available method between animals in the reference population.
Genotyped animals/bull	Track how many animals are genotyped per bull. What variation between bulls?
Percentage of best phenotypes that are genotyped	What percentage of high quality phenotype animals are also genotyped? (High quality phenotype animals)/ (High quality phenotype animals genotyped).
Cost per genotype	Cost for a genotype in the reference population.
Yearly evaluate for best payment solutions	Yearly check how to best combine genotyping for reference population and other usage of genotyping.

Reference population *Size*	
Number of animals in the reference population (females and bulls) for all breeding goal traits	Number of animals in the reference population (females and bulls) for all breeding goal traits - All animals included - Animals with genotype - Animals included without genotype
Number of animals contributing information to the reference population in the future total merit - Percentage of the population	Number of animals contributing information to the reference population in the future total merit (Percentage of the population).
Accuracy of all traits	Accuracy of all traits in the breeding goal.

Number of animals (females and bulls) added/year	The number of added animals in the reference population - All animals included - Animals with genotype - Animals included without genotype
Reference population *Composition*	
Relationship between animals	Relationship between animals - Production population, breeding animals and reference population
Number of tested daughters/sire	Number of tested daughters per sire.
Variation in numbers of tested daughters/sire	Variation in numbers of tested daughters per sire.
Reference population *Regularly renewed*	
Age of reference population - Age structure in years per trait (females and males)	Age of reference population - Age structure in years per trait (females and males)
Average accuracy of all traits	Track the average accuracy of all interesting traits.
Renewed regularly - Track how/when/why - Effect on accuracy	Track how, when and why the reference population is renewed. What was the effect on the accuracy?

Breeding scheme *Evaluation*	
Genetic trends for total merit	Track trends for total merit.
Check genetic trends for all breeding goal traits - Within and across countries	Track trends for all breeding goal traits. Are there differences between countries?
Phenotypic trends in main breeding goal traits	Phenotypic trends in main breeding goal traits over time to check breeding scheme.
Cost per dose of semen - Sold and produced	Could in cases be a good KPI to evaluate efficiency of the breeding scheme. Check both sold and produced.
Yearly track law changes - Impact on breeding scheme	Check law changes that could have impact on the breeding scheme. For example: You have to register a certain phenotype correct which could affect quality.
Action plan for ethical aspects	Plan for ethical aspects to be able to argue for decisions. Some examples: gen editing, cloning.

Breeding scheme *Evaluation*	• Genetic trends for total merit - Check genetic trends in main breeding goal traits - Within and across countries • Phenotypic trends in main breeding goal traits - Within and across countries - Females and bulls • Cost per dose of semen - Sold and produced • Yearly track law changes - Impact on breeding scheme • Action plan for ethical aspects - For example: gen editing, cloning

Comments:

Genotyping strategies - Literature review

Family structure and the design of the reference population may influence the accuracy of genomic predictions (Pszczola et al., 2012). Pszczola et al. (2012) estimated the effect of different relationship of evaluated animals within the reference population on the reliability of direct genomic breeding values (DGV). The reference population consisted of highly, moderately, lowly related animals, the selection was done through selection of paternal half-sib families of decreasing size. In addition a complete random reference population was evaluated. The randomly chosen animals had the lowest relationship within the reference population. The results showed the randomly chosen reference population achieved the highest reliability. For a trait with a heritability of 0.3 the reliability was 0.53 in the highly related reference population compared to 0.61 in the randomly selected reference population. In addition, the reliabilities increase as the relationship between the selection candidates and the animals in the reference population increases. Also the reliabilities decrease as the number of generations between the selection candidates and the animals in the reference population increases (Pszczola et al., 2012).

Plieschke et al. (2016) evaluated systematic genotyping of cows in a try to improve genomic breeding values. The reference set consisted of only genotyped bulls at start and was thereafter extended with a fixed number of first-crop daughters of each sire in a regular and balanced system of genotyping. The results showed that even with a relative low number of genotyped daughters per sire it led to improvements in the reliability of the predicted breeding values. In the most extended design with 200 genotyped daughters per sire, reliabilities of 80% or more were achieved for traits with moderate to high heritability. In this case SNP effects were estimated from a reference population of 420,000 cows and 4200 daughters (Plieschke et al., 2016).

Genotyping for reference population - Literature review

Jiménez-Montero et al. (2012) evaluated different female-selective genotyping strategies for improving predictive accuracy of genomic breeding values in small dairy cattle populations. The study simulated a population of 996 progeny-tested sires and 40,000 dams. The results showed that strategies that genotyped females from the top of the yield deviations and breeding values rankings gave the most biased prediction. However, genotyping animals from both the top and bottom rankings resulted in increased reliability in small reference populations (Jiménez-Montero et al., 2012).

Boligon et al. (2012) used stochastic simulation to evaluate the quality of breeding values with different selective genotyping strategies. The reference population and the generation were fixed at 2500 animals each. Five genotyping strategies were used to find 500 animals in each generation for the reference population: 1) random selected animals, 2) top animals with the largest yield deviation values, 3) bottom animals with the lowest yield deviation values, 4) extreme animals which consisted of the animals with the 250 largest and the 250 lowest yield deviation values, 5) less genetically related animals. Overall the extreme genotyping strategy gave the best predictive ability of breeding values. The authors concluded that the extreme animals were the most informative when training genomic selection models (Boligon et al., 2012).

Chip density and imputation - Literature review

The density of SNP markers affects the accuracy of GEBV (Meuwissen, 2009, Habier et al., 2009). In theory, a higher density should lead to a higher accuracy but it also leads to an increased cost for genotyping (Peipei, et al 2013). Some countries have genotyped bulls with a 777,000-markers high density chip (777K; high-density, HD), with the purpose of increasing the accuracy (Su et al., 2012). In addition, low density chips with 6,900-markers and 2,900-markers (BovineLD and Bovine3K) have been developed, those should be more suitable for a large scale and have a lower genotype cost (Boichard et al., 2012).

When several chips are used in genomic selection it is important to make use of all available marker data by imputation of missing genotypes. Imputation is also useful to increase the call rate of genotyped animals when the same chip is used (Peipei, et al 2013). Imputing from 3K to 54K gave lower imputation accuracies than imputing from 54K to 777K, 93.5 to 97.1% compared to 97.1 to 99.3% (Peipei, et al 2013).

Reference population - Literature review

The first steps when predicting GEBV are to divide the genome into small segments and thereafter estimate the marker effects in a reference population where animals are both genotyped and phenotyped. This further means that the next generation can be genotyped for the markers and thereafter the sum of the effects determine their predicted GEBV (Meuwissen et al., 2001). The accuracy of genomic predictions depends on the number of phenotypic records that the marker effects were estimated on. The required size of the reference population also depends of the heritability; with a higher heritability a lower amount of phenotypic records is required (Meuwissen et al., 2001). Lund et al. (2010) evaluated the increased reliability archived when combining Holstein reference populations. The four organizations were UNCEIA (France), DHV-VIT (Germany), CRV (The Netherland, Flanders) and VikingGenetics (Denmark, Sweden, Finland). Each country contributed with 4,000 genotyped bulls. The average increase in reliability for all traits was 11%-units with a combine reference population. Further, the increase in reliability between traits and countries ranged from 2%-units to 19%-units (Lund et al., 2010).

Females in the reference population - Literature review

During the early years of genomic selection mainly bulls were genotyped, but when the costs decreased genotyping of heifers became more interesting. For the dairy producer, higher accuracy when selecting replacement animals and better mating plans are some of the main arguments for genotyping heifers (Pryce & Hayes 2012). It has also been shown that expanding the reference population with cows were an efficient way to increase reliability of GEBV (Thomasen et al., 2014b). Koivula et al. (2014) studied the effect of including 5,593, 3,111 or 0 genotyped Nordic Red Dairy cows in the reference population on the gain in accuracy of the genomic predictions. In all evaluations 4,188 genotyped bulls were used. The extra gain in accuracy from cows in the reference population varied from 0.8% to 2.6%-units (Koivula et al., 2014). Wiggans et al. (2011) included 3,559 Holstein females and 388 Jersey females in the reference population. The results showed extra gains in reliabilities of 3.5%-units for Holstein and 0.9%-units for Jersey. Further, Pryce et al. (2012b) demonstrated an increase of 8%-units reliabilities when 10,000 cows were added to a reference population of 3,000 bulls.

Thomasen et al. (2014b) evaluated the effect of including cows in the reference population for a small dairy cattle population. Three scenarios in a period of 15 years were investigated regarding maintaining and updating the reference population. The first scenario included the number of progeny tested bulls which was tested at four levels; 15, 40, 60 100. In the second scenario, 2,000 first lactation cows each year were randomly selected and genotyped. In the third scenario, an additional of 2,000 first lactation cows were selected and genotyped of the first two years. The study evaluated a juvenile breeding scheme and a Hybrid breeding scheme. The breeding goal consisted of a production trait with a heritability of 0.3 and a functional trait with a heritability of 0.04. The study used a stochastic approach as model to the different strategies and the evaluation criteria were AMGG, rate of inbreeding, reliability of genomic predictions and variance of response. The results showed that including cows in the reference population increased AMGG and decreased inbreeding compared to updating the reference population with 60 progeny tested bulls annually. The additional 2,000 cows during the two first years provided extra AMGG. The juvenile breeding scheme outperformed the hybrid breeding scheme regarding AMGG due to the shorter generation intervals. The variance of response which reflected the risk was in general higher using genomic bulls in juvenile schemes compared to using progeny tested bulls, due to the lower reliabilities. The study concluded that genotyping cows are a fast way to increase reliabilities of genomic predictions in small dairy cattle populations (Thomasen et al., 2014b).

Female genomic information - Literature review

Hugh et al. (2011) investigated the complete effect of genomic female information in juvenile schemes. The study used a stimulation program and the population consisted of 100 males and 100 females and a Fisher-Wright population model was used. The study showed that including females in genomic breeding programs could triple the genetic gain compared with a conventional progeny testing scheme. The main reasons for the extra genetic gain were increased accuracies and shorter generation intervals. The generation interval of males could be decreased by 3.8 years without any reasonable change in inbreeding. In addition the accuracy of the selected males was increased by 73% in the final 3 years in the genomic breeding scheme compared to a traditional BLUB breeding scheme (Hugh et al., 2011).

Genotyping for management - Literature review

One other use of genotyped heifers is to find the best heifers for replacement on herd level (Pryce & Hayes 2012). The study assumed a herd of 100 cows where the heifers available for selection varied from 20 to 50. The replacement rate varied from 15% to 30%. Three different costs of genotyping were assumed; 5 Australian dollar (AU$5), AU$50 or AU$100. Comparison of genomic selection with parental average (PA) information or no PA information was made. Genotyping heifers became profitable when the price of genotyping was AU$50 with no PA information and at AU$5 when PA information was included. The largest benefit was achieved with a high number of candidates for a few replacement spots. However, their comparison of costs and benefits of genotyping heifers did not take marketing into account. It might be profitable to market heifers or embryos from heifers with breeding values at birth with up to 60% reliability. Other advantages of genotyping heifers may be more optimal mating plans and keeping recessive alleles under control (Pryce & Hayes 2012).

Hjortø et al., (2015) investigated if genotyping females could help to reduce genetic lag on herd level. Genomic test results of females were used in combination with sexed semen or a high management level, which result in a reproductive surplus in the herd. The use of sexed semen was also combined with beef semen to produce crossbred slaughter animals. The study used 2 stochastic simulation programs, the first (SimHerd) estimated economic effects of different strategies and also the distribution of the dams of heifers calves. The second simulation program (ADAM) estimated genetic merit in different scenarios. The annual net return per slot was calculated as the sum of operational return and value of genetic lag minus the cost for genomic test divided by the total number of slots. The results showed that the use of genomic tests for decision making could reduce genetic lag by as much as 0.14 genetic standard deviation units of the breeding goal. In addition, if genomic test were used in combination with strategies for increasing and using a reproducing surplus the genetic lag decreased up 0.30 genetic standard deviation units. However, the current price (€50) for a LD chip genotyping in the Nordic countries was too high to bring any extra profitability (Hjortø et al., 2015).

Genotyping strategies – Optimal breeding scheme

The most value out of genotypes is acquired when they are used for several purposes (VanRaden et al., 2014). *Figure 1* shows different purposes of genotypes that the literature suggests. In a breeding scheme perspective it would easiest if all animals were genotyped. However, with the current prices for genotyping there are split opinions if it is profitable or not to genotype all females for management purposes alone (Pryce & Hayes 2012a; Hjortø et al., 2015). Therefore, it is not sure that genotyping all animals is the most economical optimal breeding scheme in a farmer perspective. Instead, to get the most value out of a genotype it is probably necessary with collaborations between farmers and breeding companies, this will split the cost on several stakeholders. There might also be different animals of interest. For example, genotypes for further breeding might not be the right genotypes for the reference population (Plieschke et al., 2016; Pszczola et al., 2012; Pryce & Hayes 2012). This could make it more difficult for farmers and breeding companies to spit the cost fairly.

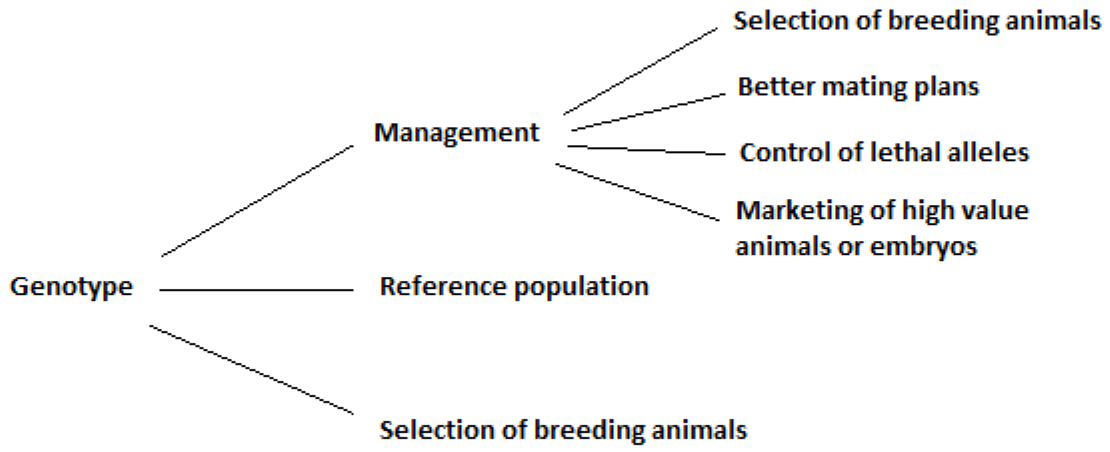

Figure 1. Different purposes of genotyping. Sources: (Pryce & Hayes 2012; Boligon et al., 2012; Pszczola et al., 2012; Koivula et al., 2014; Wiggans et al., 2011).

It is hard to know how the future will look regarding the availability of genotypes. The price for genotyping has decreased and if the genotyping price continues to fall, genotypes might be more accessible for breeding schemes. The best thing from an optimal breeding scheme perspective then might be to improve phenotype recording strategies and thereby increase the accuracies of breeding values. However, as of 2016, the genotype situation requires attention to achieve as optimal breeding scheme as possible. In a situation where genotypes are limited, it is also important to make sure as far as possible that the animals with the most accurate and well-defined phenotypes are genotyped. Further, genotyping should be were the most traits are recorded and thereby is each genotype utilized to its maximum.

Reference population - Optimal breeding scheme

The required size of the reference population depends of the heritability; with a higher heritability a lower amount of animals are required to reach certain accuracy (Meuwissen et al., 2001). With a broad breeding goal like in the Nordic countries the heritability of the different traits varies a lot and thereby the optimal size of reference population varies depending on trait. However, there are also other optimizations, for example the family structure in the reference population may influence the accuracy of genomic selection (Pszczola et al., 2012). This will affect what animals you want to include in the reference population. Reliabilities increase when the average relationship within the reference population decreases. In practice this means that you want to minimize relationship within the reference population. Further, the reliabilities increased as the relationship between the selection candidates and the animals in the reference population increases.

Moreover, two tailed selection (top and bottom ranked) on parental average or yield deviation have been shown to give the highest accuracies (Jiménez-Montero et al., 2012). This means that you want genomic information from the most informative animals and thereby the reference population represents the total genetic variation. If you become more and more selective among genotypes and phenotypes, the renewal and composition of the reference population will be important objectives for research.

Combining reference populations with genotyped bulls through cooperation between countries has been shown to be an effective way to increase the reliability (Lund et al., 2010). However, those animals have to be representative and phenotypes have to be well-defined and accurately recorded. Further, it is not sure that the breeding goals include the same traits. For small breeds like the Jersey and the Nordic Red it might also be hard to find cooperation partners due to their less widespreadness compared to Holstein.