

# Genetic Selection on Animals using Pedigree, Phenotypic, and Genomic Information

Shogo TSURUTA\*

---

**Abstract:** Genetic selection of quantitative traits in animals has been performed for a century in the United States. Especially, in the last two decades, field data analysis of dairy cattle, beef cattle, swine, and chicken has contributed to livestock industries. We have analyzed genetic components in animal production, reproduction, and disease traits using our own computer programs (so-called BLUPF90 family). Those programs have been used by breed associations and breeding companies in routine genetic evaluations. Recently, we have developed a method called single-step genomic BLUP (ssGBLUP) to predict genomic breeding values by combining pedigree, phenotypes, and genotypes (SNP markers).

With ssGBLUP, we found that the accuracy of genomic breeding values can be increased in dairy and beef cattle, pigs, and chickens by 10-30% compared with traditional breeding values. This methodology has been expanded to use a large number of genotyped animals (> 1 million). The application programs are available on our website at <http://nce.ads.uga.edu/projects/programs/~>. In this paper, we explain general BLUP and ssGBLUP algorithms applicable to livestock and fish breeding. Also, we show evidence of genetic progress by genetic selection and results from genomic analyses in US Holsteins.

Genomic selection has a greater advantage when more genotyped animals are available because predictions are more accurate and more complex models can be applied. When animals are genotyped in the earlier stages of life, genetic progress can be accelerated and genetic gains can be magnified. The ssGBLUP methodologies are applicable to other organisms, such as farmed fish, shellfish, anadromous fish (salmon), honey bees, and even plants.

**Key words:** Genetic selection, BLUP, genomic information

---

## Introduction

Genetic selection of quantitative traits in US Holsteins using genetic evaluation has been performed for 80 years in the United States (USDA-ARS Animal Improvement Program 2013). In the beginning, artificial selection was conducted using only phenotypes. Later, pedigree information was added from parent-offspring relationships, siblings, and families to all available information among related animals. Around the same time, advanced statistical methods were developed (e.g., regression

analysis, least square analysis, and best linear unbiased prediction (BLUP) with an inverse of the relationship matrix) to estimate genetic parameters and to predict breeding values for economically important traits in animals. As a result, animal productions have dramatically increased in the last half-century.

Genetic progress can be equated to a function of accuracy of the selection, genetic variation (standard deviation), selection intensity, and generation interval. To accelerate genetic progress for quantitative traits, we need to increase the first three parameters and

decrease the last one, ignoring all possible interactions. Accuracy can be increased by adding genomic information into traditional breeding values estimated with phenotypes and pedigrees. We conducted several studies with genomic information using a single step genomic BLUP (ssGBLUP) to increase accuracy in genomic breeding values (Aguilar *et al.*, 2010; Tsuruta *et al.*, 2011; Lourenco *et al.*, 2015; Fragomeni *et al.*, 2015). Genetic variation can be increased by outbreeding and mutation. Selection intensity can be increased by increasing the population size. Generation interval can be reduced by evaluating an animal's performance at the early stages of its life, using genomic information.

Unfortunately, in the US and in Japan, genetic selection in fish and shellfish has not garnered much popularity compared to other livestock. The main reasons may include 1) recording phenotypes and identifying individuals were expensive, 2) farming was much more expensive than fishing, and 3) as a result, genetic analyses were not well conducted. However, these conditions have been changing over the years. As seafood gains popularity, overfishing becomes a problem for popular species, and the balanced ocean ecosystem could collapse. Fortunately, a complete breeding cycle technology in aquaculture for some species has been successfully developed. In addition, low-cost single nucleotide polymorphism (SNP) marker information is now available for genomic selection.

The objective of this paper was to show how genetic selection has been efficiently conducted in livestock production in the US and to explain its potential for improving marine species as well.

### Materials and Methods

For 34,506 US Holstein bulls, 42,503 SNP markers were available to predict genomic breeding values and SNP marker effects for 305-day (d) milk yield and cow mortality, using ssGBLUP. Details are described in Tsuruta *et al.* (2015).

In general, there are two options in genetic evaluation processes: 1) estimating genetic parameters (heritability and genetic correlation) and breeding values simultaneously or 2) estimating genetic parameters first and then breeding values

(empirical BLUP), assuming that those variance components are known. Using matrix notation, a general mixed model equation (MME) in animal breeding and genetics can be written as

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{A}^{-1}\frac{1}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

where  $\mathbf{b}$  and  $\mathbf{a}$  are unknown vectors and  $\mathbf{X}$  and  $\mathbf{Z}$  are design matrices for fixed and random effects, respectively. The additive genetic variance for  $\mathbf{a}$  (breeding values) and the residual variance for  $\mathbf{e}$  are described as

$$\text{Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_u^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

where  $\mathbf{A}$  is a pedigree or numerator relationship matrix;  $\mathbf{I}$  is an identity matrix;  $\sigma_u^2$  and  $\sigma_e^2$  are variance components for additive genetic and residual effects, respectively. Therefore, heritability can be calculated as  $\sigma_a^2 / (\sigma_e^2 + \sigma_a^2)$ . To obtain solutions for  $\mathbf{b}$  (BLUE) and  $\mathbf{a}$  (BLUP) from MME, the preconditioning conjugate gradient (PCG) algorithm is widely used owing to its stable convergence property (Strandén and Lidauer, 1999; Tsuruta *et al.*, 2001).

For ssGBLUP (Legarra *et al.*, 2009; Aguilar *et al.*, 2010), simply replace  $\mathbf{A}^{-1}$  by  $\mathbf{H}^{-1}$  as described below:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{H}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tau\mathbf{G}^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}$$

where  $\mathbf{G}$  is a genomic relationship matrix defined as follows (VanRaden, 2008):

$$\mathbf{G} = \frac{\mathbf{WDW}'}{2\sum_{i=1}^n p_i(1-p_i)}$$

where  $p_i$  = allele frequency at locus  $i$ ;  $\mathbf{D}$  = weight for each locus ( $\mathbf{I}$  if assuming the same variance);  $\mathbf{W}$  = a design matrix as follows:

$$w_{ii} = \begin{cases} 0 - 2p_j & \text{homozygous} \\ 1 - 2p_j & \text{heterozygous} \\ 2 - 2p_j & \text{homozygous} \end{cases}$$

For a genome-wide association study (GWAS), SNP marker effects and SNP variances can be estimated by solving  $\mathbf{a}=\mathbf{W}\mathbf{u}$  where  $\mathbf{u}$  is a vector of SNP marker effects (Strandén and Garrick, 2009). The BLUPF90 family programs (Misztal *et al.*, 2002) for predicting genomic breeding values and estimating variance components are available at <http://nce.ads.uga.edu/~>.

## Results and discussion

### Selection index

A selection index is widely used when selecting animals on multiple traits. Table 1 shows various selection indexes used in US Holsteins (VanRaden and Cole, 2014). A selection index can be weighted differently for each trait, depending on the selection goal. Each dairy farm has a unique strategy to improve their animals in that environment (e.g., focusing on milk production or cheese production).

### Genetic trend

Phenotypic and environmental trends of 305-d milk yield in US Holsteins (Fig. 1) were calculated based on cow's EBV published in 2015 from Council on Dairy Cattle Breeding (CDCB 2015). The difference between phenotypic and environmental trends,

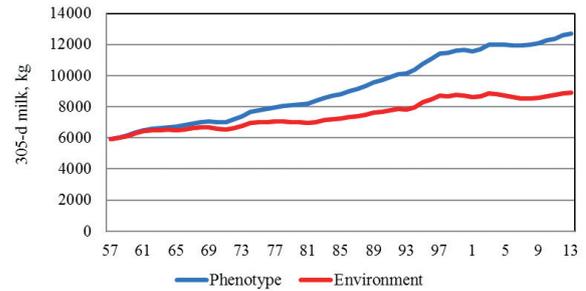


Fig. 1. Phenotypic and environmental trend for 305-d milk yield in US Holsteins

which started in 1957 assuming no genetic trend at first, can be distinguished as genetic progress or gain.

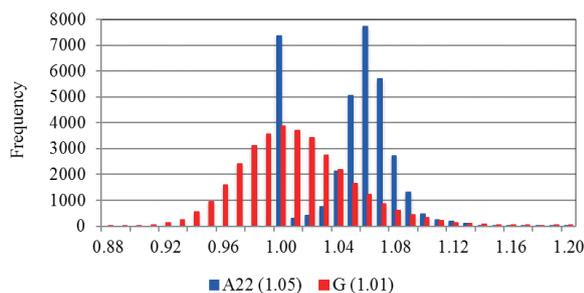
### Genomic selection

Since the USDA-ARS started genomic evaluation using SNP markers as genotypes in 2008, genomic evaluation is gaining popularity in other breeds and species. As previously mentioned in  $H^{-1}$ , MME contains three different relationship matrices:  $\mathbf{A}$ ,  $\mathbf{G}$ , and  $\mathbf{A}_{22}$ . Due to different definitions and bases on  $\mathbf{A}$  (or  $\mathbf{A}_{22}$ ) and  $\mathbf{G}$ , distributions of the matrix elements look inconsistent (Fig. 2) when 42,503 SNP markers for 34,506 US Holstein bulls were analyzed (Tsuruta *et al.*, 2015). The inverse matrices ( $\mathbf{A}_{22}^{-1}$  and  $\mathbf{G}^{-1}$ ) can be consistent, but the base adjustment could still be required for a large population with many generations.

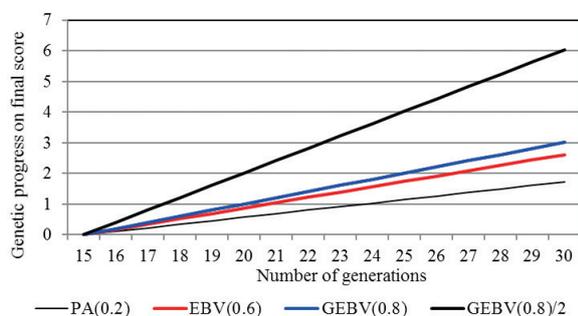
In Fig. 3, according to the genetic progress formula

Table 1. Relative weights (%) on four different selection indexes for US Holsteins

Trait	Relative weight (%)			
	Net Merit \$	Cheese Merit \$	Fluid Merit \$	Grazing Merit \$
Protein	20	24	0	18
Fat	22	19	23	20
Milk	-1	-9	23	-1
Productive life	19	16	20	10
Somatic cell score	-7	-7	-3	-6
Udder	8	6	8	8
Feet/legs	3	2	3	3
Body size	-5	-4	-5	-4
Daughter pregnancy rate	7	6	7	19
Heifer conception rate	2	1	2	3
Cow conception rate	1	1	2	5
Caving ability (\$)	5	4	5	5



**Fig. 2.** Distribution of diagonals of genomic relationship (G) and pedigree relationship ( $A_{22}$ ) matrices (mean values > 1.0 in parentheses indicate inbreeding) in US Holsteins

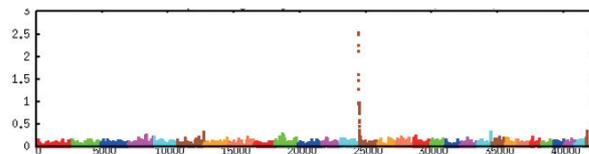


**Fig. 3.** Future prediction of genetic trend for final score in US Holsteins

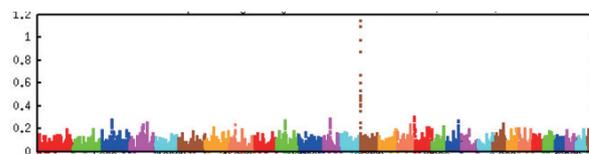
described before, future genetic trends for final score with heritability (0.31) in US Holsteins were calculated with several conditions that assume that reliability in GEBV is 0.2 higher than that in EBV and reliability in EBV is 0.4 higher than that in parent average (PA). The solid thick line is the highest genetic trend when making the generation interval a half, indicating that reducing generation interval will increase genetic gains significantly. This becomes possible by genotyping and evaluating animals in the early stages of their life before sexual maturity.

### GWAS

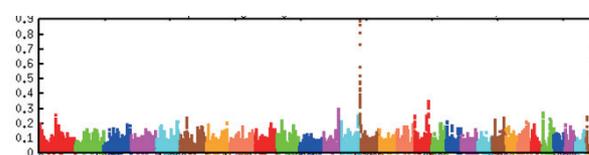
Using the ssGBLUP algorithm, SNP marker effects and variances were estimated for 305-d milk yield and cow mortality in US Holsteins (Tsuruta *et al.*, 2015). Figs. 4(a-d) show genetic variance (%) explained by SNP markers on each chromosome for 305-d milk yield in the first lactation and cow mortality in the first three lactations for US Holsteins. The Manhattan plots indicated that the SNP variance on chromosome



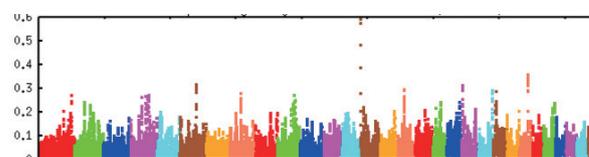
**Fig. 4a.** Genetic variance (%) explained by SNP markers for 305-d milk yield in first lactation



**Fig. 4b.** Genetic variance (%) explained by SNP markers for cow mortality in first lactation



**Fig. 4c.** Genetic variance (%) explained by SNP markers for cow mortality in second lactation



**Fig. 4d.** Genetic variance (%) explained by SNP markers for cow mortality in third lactation.

14 was significantly large (2.5%) for milk yield and the variance for cow mortality decreased as the number of lactations increased, suggesting that younger cows died for a reason more genetically related to milk production than older cows.

### Genetic selection in aquaculture

Norway has been successfully conducting genetic selection in Atlantic salmon for decades. Table 2 shows four different breeding programs that select animals for four to 11 traits.

The FAO published results for “genetic improvement” in fish (FAO Fisheries and Aquaculture Department 2015), showing some genetic progress in fish and shellfish. Selective breeding improved growth rate 50% after 10

**Table 2.** Traits presently selected in the four breeding programs for Atlantic salmon in Norway (Thomassen *et al.* 2007)

Trait	Breeding program			
	A	B	C	D
Growth in freshwater	✓	✓	✓	✓
Body weight at slaughter	✓	✓	✓	✓
Early sexual maturity	✓	✓		
Furunculosis	✓	✓		
Infectious salmon anaemia	✓	✓		
Infectious pancreatic necrosis	✓	✓	✓	
Dressing percentage	✓	✓		
Fillet yield	✓	✓		
Fillet color	✓	✓	✓	✓
Fillet fat	✓	✓		
Deformity	✓	✓		
Body shape	✓			

generations in coho salmon, 20% per generation in gilthead sea bream, 10 – 13% live weight in oyster, and 60% in tilapia. Maturity and time of spawning, physiological tolerance (stress), and disease and pollutant resistances were also improved.

If genetic selection can successfully be applied, fish farming may become the most profitable option in the fish industry. If sufficient facilities and resources (ocean, river, or lake) are available, strong selection pressure can be put on breeding lines because fish and shellfish produce a large number of eggs in a single spawning (that could be many full sibs and half sibs). A superior individual within full sibs can be identified by using genotypes. If genetic selection is systematically conducted in a complete breeding cycle by recording pedigrees (at least selection lines), phenotypes, and genotypes, the genetic progress in fish farming will be maximized. Genetic selection on economically important traits for farmed fish, shellfish, and anadromous fish is environmentally friendly and sustainable because no gene or gamete manipulation is involved (i.e., only gene frequencies are changed). It also has the capacity to multiply seafood production while maintaining product quality.

### Conclusions

Genomic selection has a greater advantage over genetic selection when young animals are genotyped

because predicted breeding values are more accurate than the traditional breeding values and the generation interval can be reduced. Genetic and genomic selection can be applied to any quantitative trait for marine animals, such as farmed fish, shellfish, and anadromous fish, if accurate pedigree information, sufficient phenotypes, and/or genotypes are available. Aquaculture has great potential in food supply all over the world, and fish production can be significantly increased by sustainable genetic selection.

### Acknowledgements

This paper was made possible by numerous studies in our animal breeding and genetics group at the University of Georgia, including current faculties (Ignacy Misztal and Daniela Lourenco), postdocs (Yutaka Masuda and Breno Fragomeni), a former postdoc (Andres Legarra), and a former graduate student (Ignacio Aguilar). We thank the livestock industry and USDA for their continued support.

### References

- Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S., and Lawlor T. J., 2010: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **93**, 743-752.
- Council on Dairy Cattle Breeding, 2015: Dairy cattle evaluation summaries: genetic and phenotypic trend. <https://www.cdcb.us/eval/summary/trend.cfm>. Accessed on March 9, 2016.
- FAO Fisheries and Aquaculture Department, 2015: Genetic improvement. [http://webcache.googleusercontent.com/search?q=cache:D0lDT6aHz-UJ:www.fao.org/fi/website/FIRetrieveAction.do%3Fdom%3Dtopic%26fid%3D14828%26xp\\_%26lang%3Den+&cd=1&hl=en&ct=clnk&gl=us](http://webcache.googleusercontent.com/search?q=cache:D0lDT6aHz-UJ:www.fao.org/fi/website/FIRetrieveAction.do%3Fdom%3Dtopic%26fid%3D14828%26xp_%26lang%3Den+&cd=1&hl=en&ct=clnk&gl=us). Accessed on March 9, 2016.
- Fragomeni B. O., Lourenco D. A. L., Tsuruta S., Masuda Y., Aguilar I., Legarra, A., Lawlor T. J., and Misztal I., 2015: Use of genomic recursions in single-step genomic BLUP with a large number of genotypes. *J. Dairy Sci.* **98**, 4090-4094.
- Legarra A., Aguilar I., and Misztal I., 2009: A

- relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **92**, 4656-4663.
- Lourenco, D. A. L., Tsuruta S., Fragomeni B. O., Masuda Y., Aguilar I., Legarra A., Bertrand J. K., Moser D., and Misztal I., 2015: Genomic evaluation using single-step genomic BLUP in American Angus. *J. Anim. Sci.* **93**, 2653-2662.
- Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., and Lee D., 2002: BLUPF90 and related programs (BGF90). Proc. 7th WCGALP, Montpellier, France.
- Strandén I. and Lidauer M., 1999: Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* **82**, 2779-2787.
- Strandén I. and Garrick D. J., 2009: Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J. Dairy Sci.* **92**, 2971-75.
- Thomassen M., Gugging R., Norberg B., and Jørgensen L. (ed.), 2007: Aquaculture Research: From Cage to Consumption. The Research Council of Norway. Oslo, Norway, pp.268-284.
- Tsuruta S., Strandén I., and Misztal I., 2001: Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.* **79**, 1166-1172.
- Tsuruta S., Misztal I., Aguilar I., and Lawlor T. J., 2011: Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in Holsteins. *J. Dairy Sci.* **94**, 4198-204.
- Tsuruta S., Lourenco D. A. L., Misztal I., and Lawlor T. J., 2015: Genotype by environment interactions on culling rates and 305-d milk yield of Holstein cows in three US regions. *J. Dairy Sci.* **98**, 5796-5805.
- USDA-ARS Animal Improvement Program, 2013: History of USDA dairy evaluations. [http://www.aipl.arsusda.gov/aipl/history/hist\\_eval.htm](http://www.aipl.arsusda.gov/aipl/history/hist_eval.htm). Accessed on March 9, 2016.
- VanRaden P. M., 2008: Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* **91**, 4414-4423.
- VanRaden P.M. and Cole J. B., 2014: Net merit as a measure of lifetime profit: 2014 revision. <http://aipl.arsusda.gov/reference/nmcalc-2014.htm>. Accessed on March 9, 2016.

### Annotated bibliography

- (1) Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., and Lee D., 2002: BLUPF90 and related programs (BGF90). Proc. 7th WCGALP, Montpellier, France.
  - (2) Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S., and Lawlor T. J., 2010: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* **93**, 743-752.
  - (3) Tsuruta S., Misztal I., Aguilar I., and Lawlor T. J., 2011: Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in Holsteins. *J. Dairy Sci.* **94**, 4198-4204.
  - (4) Lourenco D. A. L., Tsuruta S., Fragomeni B. O., Masuda Y., Aguilar I., Legarra A., Bertrand J. K., Moser D., and Misztal I., 2015: Genomic evaluation using single-step genomic BLUP in American Angus. *J. Anim. Sci.* **93**, 2653-2662.
  - (5) Fragomeni B. O., Lourenco D. A. L., Tsuruta S., Masuda Y., Aguilar I., Legarra A., Lawlor T. J., and Misztal I., 2015: Use of genomic recursions in single-step genomic BLUP with a large number of genotypes. *J. Dairy Sci.* **98**, 4090-4094.
- The authors have been working on estimation of genetic parameters and prediction of breeding values for economically important traits in animals. They have recently developed a method called single-step genomic BLUP (ssGBLUP) to predict genomic breeding values by combining pedigree, phenotypic, and genomic information and have implemented the method to their existing computer (BLUPF90) programs (e.g., BLUP, REML, and Gibbs Sampler) to use SNP marker information. Recently, they published several studies of ssGBLUP to maximize accuracy and minimize bias in genomic breeding values, applying this methodology to dairy cattle, beef cattle, pigs, and chickens. Currently, they have been working on how to include a large number of genotyped animals (> 1 million) in genomic evaluation. Breed associations and breeding companies that have been using their software are now planning to introduce ssGBLUP to routine evaluations. The same or similar methodology can be applied to fish breeding. The BLUPF90 family of programs and the manual are available at <http://nce.ads.uga.edu/~>.