



**HAL**  
open science

## Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation

Panya Sae-Lim, Antti Kause, Marie Lillehammer, Han A. Mulder

### ► To cite this version:

Panya Sae-Lim, Antti Kause, Marie Lillehammer, Han A. Mulder. Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation. *Genetics Selection Evolution*, 2017, 49 (1), pp.33. 10.1186/s12711-017-0308-3 . hal-01484835

**HAL Id: hal-01484835**

**<https://hal.science/hal-01484835>**

Submitted on 7 Mar 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

Open Access



# Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation

Panya Sae-Lim<sup>1\*</sup>, Antti Kause<sup>2</sup>, Marie Lillehammer<sup>1</sup> and Han A. Mulder<sup>3</sup>

## Abstract

**Background:** In farmed Atlantic salmon, heritability for uniformity of body weight is low, indicating that the accuracy of estimated breeding values (EBV) may be low. The use of genomic information could be one way to increase accuracy and, hence, obtain greater response to selection. Genomic information can be merged with pedigree information to construct a combined relationship matrix (**H** matrix) for a single-step genomic evaluation (ssGBLUP), allowing realized relationships of the genotyped animals to be exploited, in addition to numerator pedigree relationships (**A** matrix). We compared the predictive ability of EBV for uniformity of body weight in Atlantic salmon, when implementing either the **A** or **H** matrix in the genetic evaluation. We used double hierarchical generalized linear models (DHGLM) based either on a sire-dam (sire-dam DHGLM) or an animal model (animal DHGLM) for both body weight and its uniformity.

**Results:** With the animal DHGLM, the use of **H** instead of **A** significantly increased the correlation between the predicted EBV and adjusted phenotypes, which is a measure of predictive ability, for both body weight and its uniformity (41.1 to 78.1%). When log-transformed body weights were used to account for a scale effect, the use of **H** instead of **A** produced a small and non-significant increase (1.3 to 13.9%) in predictive ability. The sire-dam DHGLM had lower predictive ability for uniformity compared to the animal DHGLM.

**Conclusions:** Use of the combined numerator and genomic relationship matrix (**H**) significantly increased the predictive ability of EBV for uniformity when using the animal DHGLM for untransformed body weight. The increase was only minor when using log-transformed body weights, which may be due to the lower heritability of scaled uniformity, the lower genetic correlation of transformed body weight with its uniformity compared to the untransformed traits, and the small number of genotyped animals in the reference population. This study shows that ssGBLUP increases the accuracy of EBV for uniformity of body weight and is expected to increase response to selection in uniformity.

## Background

In aquaculture, selection to increase economically important traits such as growth is one of the main breeding goals. However, fish producers show interest to improve not only the mean but also the variance of traits [1]. Uniformity of growth is preferable because more uniform

growth allows a more uniform product, harvest of a larger proportion of the population at market size, and reduction of size grading and multiple harvests [2–4]. More uniform growth may also reduce competitive interactions between animals, which contributes to reduce feed monopolization and dominant behaviour, and thus improve well-being of fish [5]. Uniformity is also important for traits that have an intermediate optimal trait value [6], such as fillet lipid%, body shape, and condition

\*Correspondence: panya.sae-lim@nofima.no

<sup>1</sup> Nofima Ås, Osloveien 1, P.O. Box 210, 1431 Ås, Norway

Full list of author information is available at the end of the article

factor in the aquaculture industry. A fish whose growth is sensitive to non-measurable environmental factors, known as micro-environments, shows micro-environmental sensitivity, which results in high environmental variance and consequently contributes to increased phenotypic variation, leading to increased size variation within a group of fish. A number of empirical studies in terrestrial and aquatic species show that uniformity is partly determined by genetic factors [4, 7–16]. Thus, selective breeding can open up one avenue to improve uniformity of fish traits.

Atlantic salmon (*Salmo salar* L.) is a farmed fish that is of major economic importance. Heritability for uniformity of body weight has been estimated in Atlantic salmon [14], rainbow trout (*Oncorhynchus mykiss* Walbaum) [4, 8], and Nile tilapia (*Oreochromis niloticus*) [15, 16]. In general, heritability for uniformity ( $h_v^2$ ) is low in livestock and aquaculture species ( $h_v^2 < 0.05$ ), indicating that the prediction accuracy of breeding values for uniformity may be low [17, 18]. However, the coefficient of genetic variation (*GCV*) of uniformity of body weight is high in fish species (median *GCV* = 34.0%: min = 17.4% and max = 64.0%), which indicates high potential for response to selection [4, 8, 14, 16, 19]. One way to increase response to selection for uniformity is to increase the accuracy of estimated breeding values (EBV) for uniformity [20].

In aquaculture, full- and half-sib family sizes are usually large and thus the accuracy of EBV based on full-sibs, half-sibs and own performance is high for body weight, but not for uniformity due to its low heritability [8]. One approach to increase the accuracy of EBV is to use genomic information [21]. With genomic selection, genomic estimated breeding values (GEBV) can be obtained for the selection candidates that are genotyped, even when they have no phenotype records. One reason why genomic selection results in higher accuracy of selection is the more accurate estimation of the Mendelian sampling genetic effects through realized additive genetic relationships among animals [22]. Consequently, individual squared residuals, which is the phenotype for uniformity in a double hierarchical generalized linear model (DHGLM), may also be more accurately estimated when using genomic information.

In many cases, combining numerator pedigree and genomic information in genomic evaluations is implemented in multiple steps, which may introduce bias and need some calculations to combine with pedigree-based EBV [23, 24]. Single step genomic best linear unbiased prediction (ssGBLUP) avoids this, and genomic and pedigree information are combined in one step [23], which may lead to less bias and is less prone to double counting of information compared to genomic evaluation methods

that are performed in multiple steps. The ssGBLUP augments the numerator relationship (**A**) matrix by the genomic relationship (**G**) matrix in conventional genetic evaluation using BLUP [24]. This combined numerator and genomic relationship matrix is known as the **H** matrix [25]. In fish breeding, combining pedigree and genomic information allows exploiting the large full- and half-sib families and the more accurate relationships of the genotyped animals, and may yield a higher accuracy of selection for uniformity than the use of the **A** matrix.

To date, the use of ssGBLUP for uniformity has not been studied. Furthermore, according to a previous study, the sire-dam model, but not the animal model, implemented within the framework of DHGLM provided unbiased (co)variance component estimates [14]. However, an animal DHGLM is expected to perform better than a sire-dam DHGLM for genetic evaluation because the animal DHGLM uses full relationships between animals rather than only among sires and dams. This is particularly important for uniformity, which is quantified by the residuals of individuals, which in the animal model do not contain the Mendelian sampling term. Moreover, for genetic evaluation, the animal DHGLM uses all phenotypic information and, for most breeding programs, at least part of the selection candidates, e.g. females for sex-linked traits, have phenotypes available at the time of selection. Use of the animal DHGLM with ssGBLUP for uniformity has not been tested.

In this study, we implemented ssGBLUP for predicting GEBV for uniformity in Atlantic salmon. Specifically, our aim was to compare the predictive ability of EBV for uniformity of body weight when implementing either BLUP with the **A** matrix or ssGBLUP with the **H** matrix. The (co)variance components were estimated from the sire-dam DHGLM with either **A** or **H** and compared prior to genetic evaluation.

## Methods

### Data

The data used in this study originated from the experiment conducted by Nofima AS and the breeding company SalmoBreed in Norway. The experiment followed all the regulations of animal ethical practice and was approved by the Norwegian Research Animal Committee (ID 6489). In 2013, 234 full-sib families were established from the mating of 131 sires to 234 dams (Table 1) during four weeks. Forty-seven percent of the parents were from year class 2009 and the rest from year class 2010. After hatching, fingerlings from each family were held in a 180-L family tank until tagging size (at mean body weight of 50 g). Each animal was tagged using passive integrated transponder (PIT) tags (Satpos AS, Norway). During tagging, a fin sample for genotyping was collected

**Table 1 Population structure of Atlantic salmon**

Population structure	
Sires, dams	131, 234
Sires per dam, mean (range)	1.0 (1.0)
Dams per sire, mean (range)	1.78 (1–3)
Full-sib families	234
Fish per full-sib family, mean (range)	15.4 (4–54)
Number of fish with records	3595
Genotyped animals	
Full-sib families	50
Fish per full-sib family, mean (range)	28.3 (21–38)
Number of fish with records	1416

from 21 to 38 sibs of each of 50 full-sib families. Thereafter, all fish were randomly allocated to three experiment tanks and grown for 11 months. At the average age of 16 months, all fish were challenged with sea lice using a co-habitat challenge, and at the end of the challenge test, final body weight (g) was measured for all 3595 fish with an electronic balance. A total of 1416 offspring (39% of all offspring) and the 131 sires and 234 dams were genotyped using the 31 K Affymetrix single nucleotide polymorphism (SNP) chip for Atlantic salmon developed by Nofima. Quality control of SNPs was performed in PLINK v1.9 [26] based on the following criteria: SNPs were removed if (1) their call rate was lower than 90%, (2) they deviated from Hardy–Weinberg equilibrium with a  $P$  value cut-off of  $10^{-15}$ , and (3) their minor allele frequency (MAF) was lower than 0.01. After quality control, 921 of 31,013 SNPs were removed (2.9%) and, thus 30,092 SNPs remained to create the genomic relationship.

### Relationship matrix

The numerator relationship ( $\mathbf{A}$ ) matrix with 814 ancestors in four generations was prepared based on pedigree information using ASReml [27]. The combined numerator and genomic relationship ( $\mathbf{H}$ ) matrix was defined as [23]:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{A}_{12} + \mathbf{A}_{22}^{-1}(\mathbf{G} - \mathbf{A}_{22})\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{G} \\ \mathbf{G}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{G} \end{bmatrix},$$

where  $\mathbf{A}_{11}$  is the pedigree relationship matrix between non-genotyped animals,  $\mathbf{A}_{12}$  and  $\mathbf{A}_{21}$  are pedigree relationship matrices between genotyped and non-genotyped animals,  $\mathbf{A}_{22}$  is the pedigree relationship matrix between genotyped animals, and  $\mathbf{G}$  is the genomic relationship matrix between genotyped animals. The  $\mathbf{G}$  matrix was computed as [28]:  $\mathbf{G} = \frac{\mathbf{W}\mathbf{W}'}{N}$ , where  $\mathbf{W}$  is the matrix of the scaled SNP genotypes for all loci and  $N$  is the total number of SNPs (30,092). The elements of  $\mathbf{W}$  were calculated as:

$$w_{ij} = \frac{(x_{ij} - 2p_j)}{\sqrt{2p_j(1 - p_j)}},$$

where  $x_{ij}$  is the SNP genotype (coded 0, 1, or 2) for the  $i$ th individual at SNP  $j$  and  $p_j$  is the allele frequency of the homozygous genotype coded as 2.

However, Aguilar et al. [29] and Christensen and Lund [30] showed that the inverse of the  $\mathbf{H}$  matrix can be computed as:

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

which is less computational demanding and more simple than preparing and subsequently inverting the  $\mathbf{H}$  matrix. The  $\mathbf{H}^{-1}$  was prepared by using the Calc\_grm computer software [31], which prepares both  $\mathbf{A}^{-1}$  and  $\mathbf{G}^{-1}$  internally before computing  $\mathbf{H}^{-1}$ .

### Statistical analysis

#### Analysis of residuals

Uniformity can be quantified by squared residuals from a BLUP mixed model equation [32]. The use of genomic information to construct realised relationships between animals, especially for full-sibs, is expected to increase the accuracy of residual estimates due to a greater accuracy of EBV for body weight. Therefore, we investigated the effect of ssGBLUP and traditional BLUP on individual residual estimation. Furthermore, we investigated sire-dam and animal models because residual estimates from a sire-dam model contain not only the unexplained environmental effects but also Mendelian sampling genetic effects. Residual estimates from an animal model do not contain the latter when EBV are estimated with an accuracy of 1. In total, residuals from four models were compared, i.e. the sire-dam or animal model with either  $\mathbf{A}$  or  $\mathbf{H}$ .

The animal mixed model was:

$$y_{iklmn} = \mu + \beta age_k + t_l + yc_m + a_i + c_n + e_{iklmn}, \quad (1)$$

where  $y_{iklmn}$  is the observation (body weight) of the  $i$ th individual,  $\mu$  is the overall mean,  $age$  is the fixed covariate effect due to different levels of age of the fish, calculated from the start feeding date until the date of measurement (day),  $\beta$  is the fixed linear regression coefficient on age,  $t$  is the  $l$ th fixed communal tank effect,  $yc$  is the  $m$ th fixed effect of year class of the parents,  $a_i$  is the random additive genetic effect,  $\mathbf{a} \sim [0, \mathbf{A}\sigma_a^2]$ , where  $\mathbf{A}$  is the numerator relationship matrix, or  $\mathbf{a} \sim N[0, \mathbf{H}\sigma_a^2]$ , where  $\mathbf{H}$  is the combined genomic and pedigree relationship matrix,  $N$  is the normal distribution, and  $\sigma_a^2$  is the additive genetic variance for body weight,  $c_n$  is the random common effect for full-sibs,  $\mathbf{c} \sim N[0, \mathbf{I}\sigma_c^2]$ , where  $\mathbf{I}$  is the identity

matrix and  $\sigma_c^2$  is the common environmental variance of body weight, and  $e_{iklmn}$  is the random residual effect,  $\mathbf{e} \sim N[0, \mathbf{I}\sigma_e^2]$ , where  $\sigma_e^2$  is the residual variance of body weight assumed to be homogeneous. For the sire-dam model, the term  $a_i$  in Eq. (1) was replaced by the random sire-dam ( $u_i$ ) effect,  $\mathbf{u} \sim N[0, \mathbf{A}\sigma_u^2]$  or  $\mathbf{u} \sim N[0, \mathbf{H}\sigma_u^2]$ . The same  $\mathbf{A}$  and  $\mathbf{H}$  matrices were used for the sire-dam and the animal models.

**Estimation of genetic parameters for uniformity**

To estimate genetic parameters for body weight and its uniformity, the sire-dam DHGLM was used [33, 34] because it is expected to provide unbiased (co)variance components for uniformity [14]. Body weight records were treated in two different ways. First, observed body weight was standardized to a mean of 0 and variance of 1, which facilitates convergence of the DHGLM. Second, we used either the natural log or the Box-Cox transformation to account for possible scale effects, because variances typically increase with increasing trait means [35, 36]. For the Box-Cox transformation, each observation was computed as  $\frac{y_i^\lambda - 1}{\lambda}$ , where  $\lambda$  is the transformation parameter, which was estimated based on Eq. (1) without the random effects [37] by maximum likelihood using the MASS package in R software [38]. The estimate of  $\lambda$  was close to 0 (0.076), indicating that the Box-Cox transformation is very similar to log-transformation, which sets  $\lambda$  equal to 0. Therefore, the Box-Cox transformed body weight was not used further. The standardized body weight and natural logarithm body weight are abbreviated as stdWT and lnWT, respectively.

To estimate genetic parameters, standardized and transformed body weights were modelled using sire-dam DHGLM in ASReml [32]:

$$\begin{bmatrix} \mathbf{y} \\ \Psi \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_v \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{b}_v \end{bmatrix} + \begin{bmatrix} (\mathbf{Z}_s + \mathbf{Z}_d) & \mathbf{0} \\ \mathbf{0} & (\mathbf{Z}_s + \mathbf{Z}_d)_v \end{bmatrix} \times \begin{bmatrix} \mathbf{u} \\ \mathbf{u}_v \end{bmatrix} + \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_v \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{c}_v \end{bmatrix} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_v \end{bmatrix}, \tag{2}$$

where  $\mathbf{y}$  is the vector of stdWT or lnWT records for the  $i$  th individual;  $\Psi$  is the vector of response variables for the residual variance, where  $\psi_i = \log(\hat{\sigma}_{e_i}^2) + \frac{\hat{e}_i^2}{1-h_i} - \frac{\hat{\sigma}_{e_i}^2}{\hat{\sigma}_{e_i}^2}$ , which was linearized using a Taylor series approximation in ASReml [34],  $\hat{e}_i^2$  is the squared residual of the  $i$ th body weight record,  $h_i$  is the diagonal element in the hat-matrix of  $\mathbf{y}$  [39], and  $\hat{\sigma}_{e_i}^2$  is the estimated residual variance of the  $i$  th observation in the previous iteration of ASReml;  $\mathbf{X}$  and  $\mathbf{X}_v$  are incidence matrices of the fixed effects described in Eq. (1) for the trait mean and its uniformity, respectively;  $\mathbf{b}$  ( $\mathbf{b}_v$ ) is the solution vector for the corresponding fixed

effects;  $\mathbf{Z}_s$  and  $\mathbf{Z}_d$  are incidence matrices for the random sire (s) and dam (d) effects;  $\mathbf{u}$  ( $\mathbf{u}_v$ ) is the vector of additive genetic effects of sire-dam on the weight (uniformity), which was assumed to follow a normal distribution for the  $\mathbf{A}$  matrix:

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{u}_v \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{4} \begin{bmatrix} \sigma_a^2 & \sigma_{a,a_v,exp} \\ \sigma_{a,a_v,exp} & \sigma_{a_v,exp}^2 \end{bmatrix} \otimes \mathbf{A}\right),$$

and for the  $\mathbf{H}$  matrix:

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{u}_v \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{4} \begin{bmatrix} \sigma_a^2 & \sigma_{a,a_v,exp} \\ \sigma_{a,a_v,exp} & \sigma_{a_v,exp}^2 \end{bmatrix} \otimes \mathbf{H}\right),$$

where the  $\frac{1}{4}$  accounts for the fact that the sire and dam each explain only a quarter of the additive genetic variance;  $\mathbf{Q}$  ( $\mathbf{Q}_v$ ) is the incidence matrix for the random common effects to full-sibs;  $\mathbf{c}$  ( $\mathbf{c}_v$ ) is the vector of common effects to full-sibs:

$$\begin{bmatrix} \mathbf{c} \\ \mathbf{c}_v \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_c^2 & \sigma_{c,c_v,exp} \\ \sigma_{c,c_v,exp} & \sigma_{c_v,exp}^2 \end{bmatrix} \otimes \mathbf{I}\right).$$

The residuals of  $\mathbf{y}$  ( $\mathbf{e}$ ) and  $\Psi$  ( $\mathbf{e}_v$ ) were assumed to be independently normally distributed as follows:

$$\begin{bmatrix} \mathbf{e} \\ \mathbf{e}_v \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{W}^{-1}\sigma_\epsilon^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_v^{-1}\sigma_{\epsilon_v}^2 \end{bmatrix}\right),$$

where  $\mathbf{W} = \text{diag}(\hat{\Psi}^{-1})$  and  $\mathbf{W}_v = \text{diag}\left(\frac{1-h}{2}\right)$ , and  $\sigma_\epsilon^2$  ( $\sigma_{\epsilon_v}^2$ ) is a scaled variance that was expected to be 1. The sire-dam DHGLM was fitted iteratively to update  $\Psi$ ,  $\text{diag}(\mathbf{W})$  and  $\text{diag}(\mathbf{W}_v)$  until the log-likelihood converged [34].

**Calculation of genetic parameters**

In the sire-dam DHGLM, the estimated variance for sires was set equal to the estimated genetic variance for dams and equal to one quarter of the additive genetic variance. Hence, the additive genetic variance for body weight ( $\sigma_a^2$ ) and its uniformity ( $\sigma_{a_v,exp}^2$ ) were equal to  $4\sigma_u^2$  and  $4\sigma_{u_v,exp}^2$ , respectively. Estimates for  $\sigma_{u_v,exp}^2$  and  $\sigma_{c_v,exp}^2$  for uniformity of body weight were on the exponential scale ( $exp$ ) and were converted to an additive scale ( $\sigma_{u_v}^2$  and  $\sigma_{c_v}^2$ ) using the extension of the equations of Mulder et al. [17], as derived by Sae-Lim et al. [8]. The additive genetic variance for uniformity of body weight on the additive scale was equal to  $4\sigma_{u_v}^2$ . Phenotypic variance ( $\sigma_p^2$ ) of body weight was equal to  $2\sigma_u^2 + \sigma_c^2 + \sigma_e^2$ , where  $\sigma_c^2$  is the variance component for the effect common to full-sibs and  $\sigma_e^2$  is the residual variance of body weight. Heritability for body weight ( $h^2$ ) was calculated as  $\sigma_a^2/\sigma_p^2$ . Heritability for uniformity of body weight ( $h_v^2$ ) on the additive scale was calculated as  $\frac{\sigma_{a_v}^2}{2\sigma_p^4 + 3(\sigma_{a_v}^2 + \sigma_{c_v}^2)}$  [8, 40]. Similarly, the common environmental effect was calculated as  $c^2 = \sigma_c^2/\sigma_p^2$  for body weight and as  $c_v^2 = \frac{\sigma_{c_v}^2}{2\sigma_p^4 + 3(\sigma_{a_v}^2 + \sigma_{c_v}^2)}$  for uniformity of body weight

[8]. The genetic coefficient of variation for uniformity of body weight (*GCV*) was calculated as  $\sqrt{\sigma_{a_v,exp}^2}$ . Standard errors of  $h_v^2$  and *GCV* were approximated using the equations presented by Mulder et al. [41].

#### Genetic evaluation and cross-validation

Two genetic evaluations, i.e., BLUP with **A** and ssGBLUP with **H**, were performed in a 10-fold cross-validation using the genetic parameters estimated based on the sire-dam DHGLM and !BLUP option in ASReml. In total, four models were used in the 10-fold cross-validation, i.e. animal DHGLM with either **A** or **H** on stdWT and lnWT.

The 10-fold cross-validation was performed on standardized and transformed body weight data as follows:

- Adjusted phenotypes for body weight ( $y_i^*$ ) and its uniformity ( $\psi_i^*$ ) were calculated as  $y_i^* = \hat{a}_i + \hat{c}_i + \hat{e}_i$  and  $\psi_i^* = \hat{a}_{v_i} + \hat{c}_{v_i} + \hat{e}_{v_i}$  using the solutions from the analysis with Eq. (2) on the full dataset.
- In a modified dataset, approximately 10% of observed phenotypes ( $y_i$ ) of animals from each family were masked (=10% of the full dataset). All phenotypes had an equal chance to be masked, but the animals that were masked in the previous fold were not masked again in the next fold.
- The genetic analysis with Eq. (2) was run on the modified dataset using the **A** and **H** matrices and EBV for body weight and its uniformity were predicted for the masked animals.
- For each fold, two measurements were computed:
  - The predictive ability of EBV was calculated as the Pearson correlation of adjusted phenotypes (step 1) with the corresponding EBV ( $\hat{a}^*$ ) (step 3) for the masked animals that were genotyped, i.e.,  $\text{cor}(y_i^*, \hat{a}_i^*)$  for body weight and  $\text{cor}(\psi_i^*, \hat{a}_{v_i}^*)$  for uniformity. Kendall and Spearman correlations were also calculated for uniformity because  $\psi_i^*$  was exponentially rather than normally distributed.
  - To measure the degree of bias and accuracy of EBV or GEBV of the masked records, the mean square error prediction (MSEP) was calculated as  $\frac{\sum_i^n (\hat{a}_i^* - y_i^*)^2}{n}$  for body weight and  $\frac{\sum_i^n (\hat{a}_{v_i}^* - \psi_i^*)^2}{n}$  for uniformity of body weight, where  $n$  is the number of masked records in each fold. The MSEP was scaled by the variance of the adjusted phenotypes of the corresponding trait.
- Steps (1) to (4) were repeated for each of the 10 folds.

Finally, average Pearson, Kendall, and Spearman correlations, MSEP and their standard error (SE) over the 10

folds were calculated. A 95% confidence interval of the difference ( $d$ ) in the predictive ability from different models with either **A** or **H** was constructed using  $d \pm 1.96 \times SE_d$ , where the  $SE_d = \sqrt{\frac{SD_{animal\ DHGLM}^2 + SD_{sire-dam\ DHGLM}^2}{number\ of\ folds}}$ . When 0 was not within the 95% confidence interval, the predictive abilities of two models were considered statistically different ( $P < 0.05$ ).

## Results

### Residual estimates

Individual residuals estimated from using the **A** (BLUP) and **H** matrices (ssGBLUP) were plotted against each other to examine their relationship. As expected, the range of residual estimates from the animal models was lower than that from the sire-dam model since residual estimates from the sire-dam model included the entire Mendelian sampling term (Fig. 1).

For the sire-dam model, the use of **H** instead of **A** did not affect estimated residuals of genotyped animals since the regression coefficient of the estimated residuals using **H** on the estimated residuals using **A** and the Pearson correlation between the two were equal to 0.999, which was very similar to the regression coefficient of non-genotyped animals (0.998). The Pearson correlations between estimated residuals using **H** and **A** were the same as regression coefficients of estimated residuals using **A** on estimated residuals using **H** for genotyped animals (0.999) and non-genotyped animals (0.998).

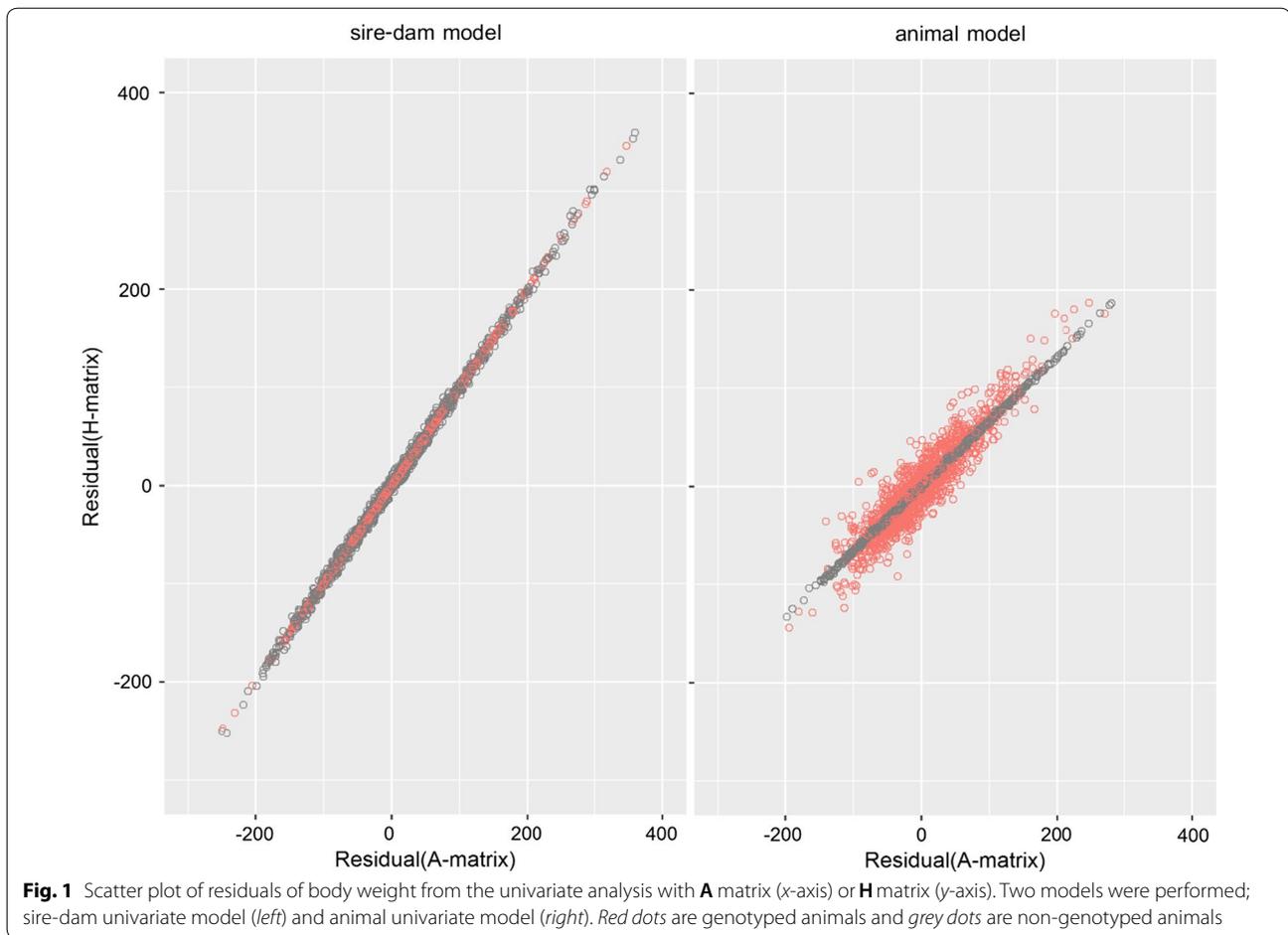
In contrast, the use of **H** in the animal model affected residual estimates of genotyped animals since their distribution was much more scattered (Fig. 1). The slope of estimated residuals using **A** on estimated residuals using **H** was lower than 1 and slightly steeper for genotyped animals (regression coefficient = 0.7025) than for non-genotyped animals (regression coefficient = 0.6798). The Pearson correlations between estimated residuals using **H** and **A** were equal to 0.922 for genotyped animals and 0.966 for non-genotyped animals.

When using the sire-dam model, the difference in estimated residuals with **H** and **A** was small and ranged from -10.8 to 10.0. When using the animal model, this difference was larger and ranged from -95.3 to 104.5.

### Genetic parameters of body weight and its uniformity

For body weight, estimates of additive genetic variances from the sire-dam DHGLM with either **A** or **H** were similar (Table 2). Likewise, estimates of  $h^2$  were similar with **A** and **H** for both traits: 0.266 and 0.296, respectively for stdWT and 0.325 and 0.346, respectively for lnWT.

When using **A**, the estimate of  $h_v^2$  was higher for uniformity of stdWT (0.036) than for uniformity of lnWT (0.015), while the use of **H** did not affect the magnitude



of  $h_v^2$  for uniformity. Standard errors of  $h_v^2$  estimates were, however, high (Table 2).

Although the estimates of  $h_v^2$  were low, estimates of *GCV* were high for uniformity of stdWT (48.0% for **A** and 52.3% for **H**), which indicates substantial genetic potential for response to selection. After accounting for scale effects, estimates of *GCV* for uniformity of lnWT were reduced to 30% (for both **A** and **H**), which supports the existence of genetic variation for uniformity beyond the scale effects.

Estimates of  $c^2$  for stdWT and lnWT were moderate and similar for **A** (0.103 to 0.117) and **H** (0.103 to 0.111), which suggests that part of the phenotypic variation was explained by non-genetic effects that are common to full-sibs. Instead, the estimates of  $c_v^2$  for uniformity of stdWT and lnWT were very low and ranged from 0.001 to 0.022 for **A** and 0.002 to 0.019 for **H** (Table 2).

The estimate of the genetic correlation between stdWT and its uniformity was close to 1, using either **A** (0.952) or **H** (0.951), which shows the high dependency between mean and variance of body weight. However, the estimate of the genetic correlation between

lnWT and its uniformity was reduced to  $-0.093$  with **A** and to  $0.024$  with **H**, which suggests that after accounting for the scale effects, the mean and variance became independent.

#### Cross-validation

The use of **H** instead of **A** with the animal DHGLM resulted in more variation of the within-family GEBV for stdWT and its uniformity, compared to within-family EBV (Fig. 2; Additional file 1: Figure S1), for sire-dam DHGLM).

The average correlation of adjusted phenotypes with predicted breeding values for stdWT and its uniformity was significantly higher with **H** (stdWT = 0.443; uniformity = 0.217 to 0.317) than with **A** (stdWT = 0.372; uniformity = 0.128 to 0.192). However, after accounting for scale effects using log-transformation, the average Pearson, Kendall and Spearman correlations of adjusted phenotypes with predicted breeding values for lnWT and their uniformity were only slightly higher with **H** than with **A**, and not significantly different from each other ( $P > 0.05$ ).

**Table 2 Estimates of variance components and genetic parameters of body weight and its uniformity based on the sire-dam double hierarchical generalized linear model when using pedigree (A) or combined pedigree and genomic relationships (H) and standard or log-transformed phenotypes**

Trait/parameter	Standardization		Logarithm	
	A	H	A	H
Body weight				
$\sigma_p^2$	0.843	0.856	0.131	0.132
$\sigma_a^2$	0.216	0.243	0.043	0.046
$\sigma_c^2$	0.095	0.091	0.013	0.014
$h^2$	0.266 <sup>0.095</sup>	0.296 <sup>0.102</sup>	0.325 <sup>0.102</sup>	0.346 <sup>0.107</sup>
$c^2$	0.117 <sup>0.037</sup>	0.111 <sup>0.037</sup>	0.103 <sup>0.038</sup>	0.103 <sup>0.038</sup>
Uniformity of body weight				
$\sigma_{a_v,exp}^2$	0.2303 <sup>0.1094</sup>	0.2732 <sup>0.1211</sup>	0.0896 <sup>0.0569</sup>	0.0885 <sup>0.0598</sup>
$\sigma_{a_v}^2$	0.0612	0.0677	0.0005	0.0005
$\sigma_{c_v}^2$	0.0360	0.0306	0.0000	0.0001
GCV	0.480 <sup>0.114</sup>	0.523 <sup>0.116</sup>	0.299 <sup>0.095</sup>	0.298 <sup>0.100</sup>
$h_v^2$	0.036 <sup>0.019</sup>	0.038 <sup>0.020</sup>	0.015 <sup>0.014</sup>	0.014 <sup>0.013</sup>
$c_v^2$	0.022	0.019	0.001	0.002

**A** = pedigree based relationship matrix; **H** = combined genotyped and non-genotyped relationship matrix;  $\sigma_p^2$  = phenotypic variance ( $2\sigma_a^2 + \sigma_c^2 + \sigma_e^2$ ), where  $\sigma_e^2$  is the residual variance for body weight;  $\sigma_a^2$  and  $\sigma_{a_v}^2$  = additive genetic variance for body weight and its uniformity, respectively;  $\sigma_c^2$  = common environmental variance; GCV = coefficient of additive genetic variance for uniformity ( $\sqrt{\sigma_{a_v,exp}^2}$ );  $h^2$  = heritability for body weight;  $c^2$  = common environmental effect due to full-sib tanks;  $h_v^2$  = heritability for uniformity;  $c_v^2$  = same as  $c^2$  but for uniformity of body weight. Superscripts are SE of the estimates

The average MSEP for uniformity from the animal DHGLM (0.608 to 0.944) were lower than those from the sire-dam DHGLM (0.973 to 1.112), suggesting that the use of an animal DHGLM increases the accuracy and may reduce bias in predicting breeding values for uniformity (Table 3; Additional file 2: Table S1). However, the average MSEP for uniformity of stdWT and lnWT obtained with **H** (0.608 to 0.944) were not notably different from those obtained with **A** (0.625 to 0.936).

The predictive ability of EBV of uniformity was sensitive to the type of correlation used, i.e. Pearson, Kendall and Spearman (Table 3). Spearman correlations were 39.1 to 49.0% higher than Kendall correlations. Predictive abilities of EBV and GEBV for uniformity of lnWT differed more from each other based on Kendall and Spearman correlations, albeit not significant at  $P < 0.05$ , than based on Pearson correlations. However, the SE of Kendall correlations were approximately 50% lower than the SE of Pearson and Spearman correlations, suggesting that Kendall correlations provide a more reliable estimate of predictive ability than Pearson and Spearman correlations.

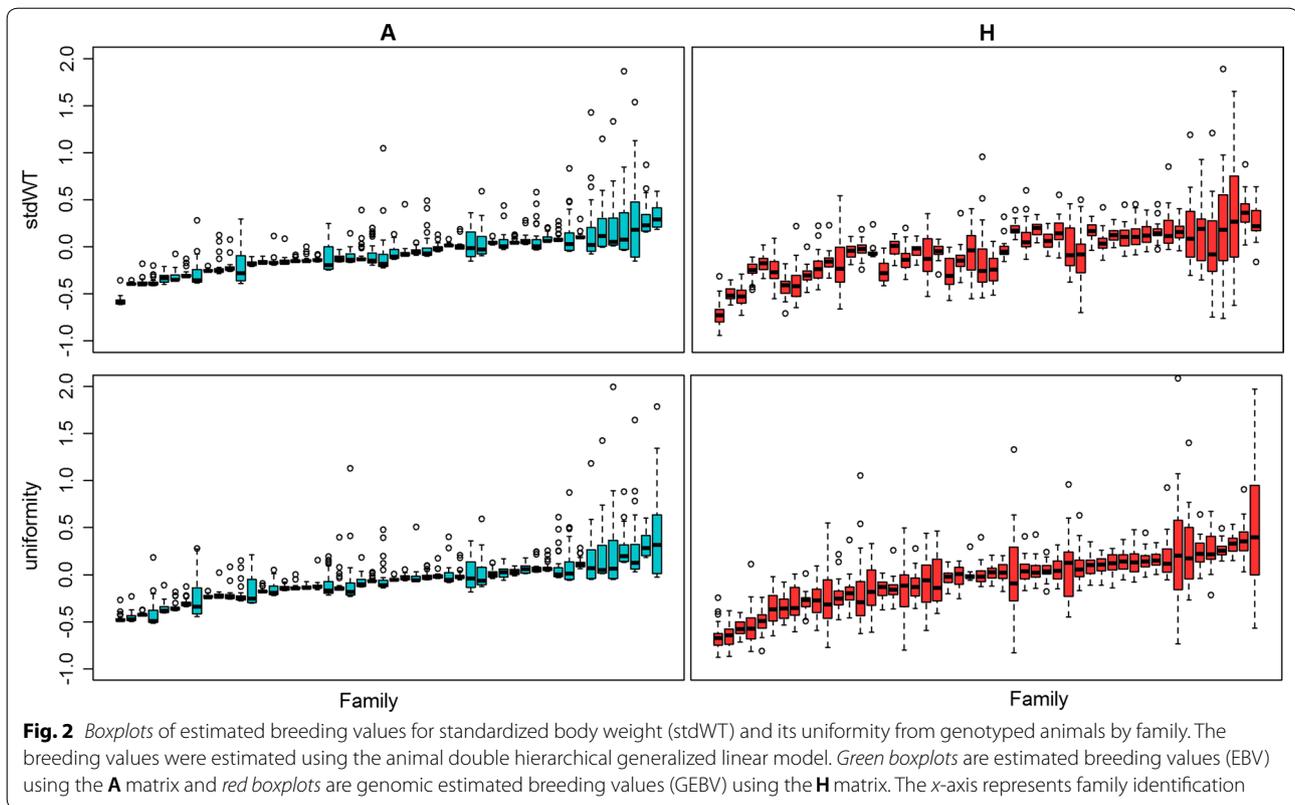
## Discussion

To the best of our knowledge, this is the first study that compares the use of the numerator relationship (**A**) and a combined genomics and numerator relationship (**H**) matrix for estimating genetic parameters and predicting breeding values for body weight and its uniformity. The use of the animal DHGLM with **H** significantly improved the predictive ability of GEBV for uniformity of body weight (stdWT) but not for scale-adjusted uniformity.

### Genetic parameters

The estimate of heritability for uniformity of stdWT from sire-dam DHGLM with **A** was low ( $h_v^2 = 0.036$ ) but higher than estimates of  $h_v^2$  obtained in previous studies on rainbow trout [4, 8] and Nile tilapia [15, 16] ( $\bar{h}_v^2 = 0.016$ ; min = 0.010; max = 0.024). However, after accounting for scale effects by logarithm transformations, the estimate of  $h_v^2$  decreased to 0.014 to 0.015, which is in line with the previous reports that also used transformations [4, 8, 14–16].

Estimates of  $h_v^2$  for stdWT and lnWT using the sire-dam DHGLM with **H** did not differ from those with **A**, which is in line with estimates of  $h_v^2$  for uniformity of piglet birth weight obtained using either **A** or only the genomic relationship matrix (**G**) [42], while lower estimates were reported for environmental variance of somatic cell score in dairy cattle when using **G** compared to **A** [43]. The similarity of the estimates of  $h_v^2$  obtained by using **A** or **H** in this study can be explained by the very similar estimated residuals (proxy of uniformity) between non-genotyped and genotyped animals when using the sire-dam model with **A** and **H**. The sire-dam model only exploits relationships between sires and dams and does not exploit the full potential of the genotype-based relationships between animals, and especially between full-sibs. In contrast, residuals of genotyped animals estimated by using the animal model were more differentiated when either **A** or **H** was used, and likely more accurate than estimates of residuals for non-genotyped animals. However, in a DHGLM analysis, the sire-dam model provides less biased (co)variance components than the animal model [14], likely because of the dependency between estimates of the breeding value and residual of an individual, which are obtained from the same phenotype of body weight. The use of genomic relationships combined with numerator relationships is expected to reduce the dependency between EBV and estimated residuals because the EBV are more accurate. Therefore, we performed the animal DHGLM with **H** but the model did not converge when the variance components were estimated, which may be due to (1) the dependency between EBV and estimated residuals for body weight remaining high, or (2) the difficulty to disentangle genetic



**Table 3** Average Pearson, Kendall and Spearman correlations and mean square error prediction from a 10-fold cross-validation based on the sire-dam double hierarchical generalized linear model<sup>a</sup> when using pedigree (**A**) or combined pedigree and genomic relationships (**H**) and standard or log-transformed phenotypes

Transformation	Relationship	Body weight		Uniformity of body weight			
		Pearson	MSEP	Pearson	Kendall	Spearman	MSEP
Standardized	<b>A</b> matrix	0.372 <sup>0.013</sup>	0.724 <sup>0.021</sup>	0.192 <sup>0.033</sup>	0.128 <sup>0.021</sup>	0.178 <sup>0.030</sup>	0.625 <sup>0.086</sup>
	<b>H</b> matrix	0.443 <sup>0.017</sup>	0.682 <sup>0.021</sup>	0.271 <sup>0.018</sup>	0.217 <sup>0.017</sup>	0.317 <sup>0.025</sup>	0.608 <sup>0.082</sup>
Logarithm	<b>A</b> matrix	0.396 <sup>0.019</sup>	0.823 <sup>0.029</sup>	0.378 <sup>0.032</sup>	0.182 <sup>0.016</sup>	0.263 <sup>0.023</sup>	0.936 <sup>0.085</sup>
	<b>H</b> matrix	0.440 <sup>0.016</sup>	0.813 <sup>0.028</sup>	0.383 <sup>0.026</sup>	0.203 <sup>0.014</sup>	0.294 <sup>0.020</sup>	0.944 <sup>0.085</sup>

<sup>a</sup> The variance components from the sire-dam double hierarchical generalized linear model were converted to the animal double hierarchical generalized linear model and were used in the 10-fold cross-validation. Relationship = relationship matrix, where **A** refers to pedigree-based relationship matrix and **H** refers to combined genotyped and non-genotyped relationship matrix. The predictability was calculated as the Pearson, Kendall and Spearman correlations between marked phenotype and predicted breeding value. MSEP was scaled by the phenotypic variance of corresponding traits

effects from the common environmental effects for uniformity of body weight.

The standard errors of  $h_v^2$  estimates were high, which may be due to the large variation in family size (4 to 54). According to Hill and Mulder [30], large family sizes or repeated measurements are recommended for estimating the genetic heteroscedasticity of traits. The optimal full-sib family size is 39 with a *GCV* of 39% and an  $h^2$  of 0.36 [18]. The use of **H** did not affect the standard errors of the  $h_v^2$  estimates, which does not agree with previous studies, for example Veerkamp et al. [44] reported lower standard

errors of  $h^2$  estimates for dry matter intake, milk yield, and body weight of heifers when using genomic relationships with an animal model. One possible explanation could be that the benefit of using the genomic relationship matrix may be limited when the sire-dam DHGLM is applied, since the variance of genomic relationships between sires (0.02) was very similar to the variance of numerator relationships between sires (0.01). In contrast, the variance in genomic relationships between animals (0.01) was much larger than the variance of numerator relationships between animals (0.004). Hence, the

SE of  $h_v^2$  estimates may be lower when using an animal DHGLM with a genomic relationship matrix, compared to the numerator relationship matrix. Nevertheless, in our study, it was not possible to investigate this phenomenon since the log-likelihood did not converge for the animal DHGLM with **H**.

The *GCV* of uniformity for stdWT was substantial (48.0%), which indicates high potential for response to selection. This result is in the upper range of previous findings in fish species (17.4 to 64.0%) [4, 8, 14–16, 19] and in terrestrial animals (10.0 to 58.0%) [6, 9–13, 42]. After accounting for scale effects by logarithm transformations, the *GCV* for uniformity was reduced but still substantial (29.5 to 29.9%), which was also reported in previous studies on Atlantic salmon [14], rainbow trout [8], rabbit, and pig [45]. Thus, scale effects affect estimates of genetic parameters for uniformity of body weight considerably, but there is genetic variation for uniformity beyond the scale effect.

Genomic information slightly increased the *GCV* of uniformity of stdWT (from 48.0 to 52.3%). In contrast, genomic information did not influence the *GCV* for uniformity of lnWT (29.8%). Since estimates of genetic parameters obtained with **A** and **H** were similar, the *GCV* for body weight remained similar, which is in agreement with the previous comparison between **A** (*GCV* = 11.0 to 12.0%) and **G** (*GCV* = 10.0 to 11.0%) for uniformity of birth weight of piglets using a dam model [42].

### Genetic and genomic predictions

In this study, we used the Pearson correlation of EBV and GEBV with adjusted phenotype as the measure of predictive ability. The use of **H** instead of **A** in the animal DHGLM significantly improved the ability to predict breeding values for stdWT (19%) and its uniformity (41.1 to 78.1%). Furthermore, the use of the animal DHGLM instead of the sire-dam DHGLM significantly increased the predictive ability of EBV and GEBV for uniformity (see Additional file 2: Table S1), as expected.

Our findings indicate that ssGBLUP with an animal DHGLM can increase the accuracy of EBV for uniformity substantially compared to pedigree-based BLUP. However, after accounting for scaling effects by using log transformations, the use of **H** compared to **A** only slightly improved the correlation (1.6 to 13.9%) and MSEP between GEBV and adjusted phenotypes, and the improvement was not significant. There are two main reasons why these results differed between uniformity of stdWT and lnWT. First, log-transformation substantially reduced the genetic correlation of stdWT with its uniformity. As a result, any increases in predictive ability of GEBV of lnWT when using **H** instead of **A** (15.7%) did not positively influence the predictive ability of GEBV

of its uniformity. Second, the lower additive genetic variance and  $h_v^2$  for uniformity after accounting for scale effects reduces the accuracy of EBV for uniformity. Consequently, MSEP increased from 0.63 (stdWT) to 0.94 (lnWT) with **A** and from 0.61 (stdWT) to 0.94 (lnWT) with **H** in the animal DHGLM.

The accuracy of genomic selection is expected to increase when the number of genotyped animals in the reference population increases for any trait [46] but in particular for lowly heritable traits, such as uniformity as shown by Sell-Kubiak et al. [42] and somatic cell score by Mulder et al. [43]. In this study, uniformity of lnWT had an even lower  $h_v^2$  than uniformity of stdWT. The number of genotyped animals in the reference population used for cross-validation was on average equal to 1274, which may have limited the benefit of using genomic information in ssGBLUP. A future empirical study should investigate the effect of the number of genotyped animals in the reference population on the ability to predict breeding values for uniformity to validate our findings and conclusions.

### Pearson or rank correlations?

Squared residuals or adjusted phenotype for uniformity ( $\psi_i^*$ ) are exponentially rather than normally distributed, which may not justify quantifying predictive ability using a Pearson correlation. Thus, we also calculated distribution-free rank correlations (Kendall and Spearman) and indeed found estimation of the predictive ability for uniformity to be sensitive to the type of correlation used.

Although not significantly different, Kendall and Spearman correlations explained differences in predictive ability of EBV and GEBV for uniformity of lnWT slightly better than the Pearson correlation. Hence, the conclusion that the benefit of using genomic relationship for computing EBV for uniformity of logarithm transformations is limited remained the same when using Kendall and Spearman correlations. Colwel and Gillett [47] showed that, in general, estimates of Kendall correlations are similar to estimates of Spearman correlations, but in some cases, the magnitude of Spearman correlations can be 50% greater than the magnitude of Kendall correlations [47]. This is in line with our findings since Spearman correlations were 42.2 to 49.0% and 39.1 to 46.1% greater than Kendall correlations for the sire-dam DHGLM (see Additional file 2: Table S1) and the animal DHGLM, respectively. Nevertheless, the SE of Kendall correlations were notably lower by approximately 50% than the SE of Pearson and Spearman correlations, which indicates that the Kendall correlation may be a more reliable estimate of predictive ability than the Pearson and Spearman correlations. Hence, it is recommended to use Kendall instead of Pearson correlations when studying predictive ability for uniformity.

### Selection for uniformity

For fish breeding, major goals are to increase mean body weight and reduce variability (more uniformity) of body weight. Nevertheless, definitions of uniformity of stdWT and lnWT are not the same. From a biological point of view, genetic variation for environmental canalization can be quantified after the scale effect is accounted for. However, from an animal breeding point of view, uniformity on the observed scale explains the actual range of fish sizes that are processed by aquaculture industries.

Selection for body weight and uniformity may be challenging because the genetic correlation between body weight and its variability is high and positive, and sometimes approaches 1. A general observation is that the genetic correlation between log-transformed body weight and its variability is zero or even negative, allowing selection to simultaneously increase transformed body weight and reduce variability. Therefore log-transformed body weight and its variability could be included in a selection index. This would require knowledge of the genetic correlation between variability of stdWT and variability of lnWT, which is not available. Hence, we used the sire-EBV and sire-GEBV, obtained from BLUP and ssGBLUP with the animal DHGLM, to calculate the Pearson correlations between EBV for the trait and its variability. Pearson correlations between EBV for variability of stdWT and EBV for variability of lnWT were positively moderate with BLUP (0.48) and ssGBLUP (0.68). The Pearson correlation between EBV for stdWT and EBV for its variability was close to 1 with either BLUP (0.96) or ssGBLUP (0.97), while the Pearson correlation between EBV of lnWT and its variability was  $-0.24$  for BLUP and  $-0.005$  for ssGBLUP. Not surprisingly, the Pearson correlation between EBV for lnWT and EBV for stdWT was highly positive (0.82). These Pearson correlations suggest that variability of lnWT should be included in the selection index because the GEBV for variability of lnWT are positively correlated with GEBV for variability of stdWT, which indicates that selection against variability of lnWT will indirectly reduce variability on the observed scale. Furthermore, GEBV for variability of lnWT are not correlated to GEBV for lnWT and thus selection against variability of lnWT will not indirectly reduce lnWT.

To elucidate responses to selection for body weight and its variability, we performed truncation selection on sires based on their GEBV from the animal DHGLM with ssGBLUP. The breeding goal could include body weight and variability on the observed scale and their economic values ( $v$ ):  $v_1 BV_{stdWT} - v_2 BV_{variability_{stdWT}}$ , while, based on the Pearson correlations discussed above, the selection index ( $I$ ) could include lnWT and its variability with their relative weighting factors ( $b$ ):

$$I = b_1 GEBV_{lnWT} - b_2 GEBV_{variability_{lnWT}}.$$

Selecting the 10% best sires on an index with  $b_1$  of 0.3, i.e.,  $I = 0.3 * GEBV_{lnWT} - 0.7 * GEBV_{variability_{lnWT}}$  provides almost no genetic gain in variability of stdWT ( $-0.001$ ) but positive genetic gain in stdWT (3.62% of mean body weight in g). In contrast, a selection index, based on breeding goal traits:  $I = 0.52 * GEBV_{stdWT} - 0.48 * GEBV_{variability_{stdWT}}$ , provides zero genetic gains for both stdWT and its variability, showing no possibility to achieve genetic gain on body weight while maintaining stable phenotypic variability. Nevertheless, the genetic gain in stdWT was much greater (17.32% of mean body weight in g) when variability was not included in the selection index. Therefore, although it is possible to increase body weight while keeping variability constant, there is a trade-off in genetic gain for body weight when selecting for reduced variability.

### Conclusions

The use of the animal DHGLM instead of the sire-dam DHGLM substantially increased the predictive ability for breeding values of uniformity, because the animal DHGLM fully exploits the relationships between full- and half-sibs. When using the animal DHGLM, the use of a combined numerator and genomic relationship matrix significantly increased the predictive ability for breeding values of uniformity of body weight, but only a slight and non-significant increase was observed after accounting for the scale effects by using transformed body weights. The small increase in predictive ability with transformed body weights may be due to lower heritability for uniformity of transformed body weight, a lower genetic correlation between transformed body weights and their uniformities, and/or a small number of genotyped animals in the reference population. The use of a Kendall correlation provided the lowest SE of predictive ability for uniformity and provided a more accurate estimate of predictive ability for uniformity over Pearson and Spearman correlations. In conclusion, the use of ssGBLUP increases the accuracy of breeding values for uniformity of harvest weight, which is expected to increase response to selection in uniformity.

### Additional files

**Additional file 1: Figure S1.** Estimated breeding values for standardized body weight and its uniformity based on sire-dam DHGLM. The data provided represent the boxplot of estimated (genomic) breeding values of genotyped animals based on sire-dam DHGLM when using pedigree (A) or combined pedigree and genomic relationships (H).

**Additional file 2: Table S1.** Cross-validation based on sire-dam DHGLM. The data provided represent the results from 10-fold cross-validations based on sire-dam DHGLM.

**Authors' contributions**

PSL analyzed the data. HAM, AK, and ML provided theoretical support for genomic DHGLM and cross-validation. HAM, AK, and ML contributed to the discussion of the results. PSL drafted the manuscript. HAM, AK, and ML improved the manuscript. All authors read and approved the final manuscript.

**Author details**

<sup>1</sup> Nofima Ås, Osloveien 1, P.O. Box 210, 1431 Ås, Norway. <sup>2</sup> Biometrical Genetics, Natural Resources Institute Finland, 31600 Jokioinen, Finland. <sup>3</sup> Animal Breeding and Genomics Centre, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

**Acknowledgements**

This study is a part of the research project entitled STABLEFISH funded by Norwegian Research Council (NRC: 234144/E49). We would like to thank SalmoBreed for providing data for this study. Matthew Baranski genotyped the animals and generated genotype data file for this study. Arthur Gilmour is acknowledged for help in implementing genomic DHGLM in ASReml v4. PSL would like to thank Mario Calus for his guidance on Cal\_grm computer software, Solomon Antwi Boison and Sergio Vela Avitúa for a fruitful discussion during the drafting of this manuscript.

**Competing interests**

The authors declare that they have no competing interests.

Received: 29 September 2016 Accepted: 28 February 2017

Published online: 07 March 2017

**References**

- Sae-Lim P, Komen H, Kause A, van Arendonk JAM, Barfoot AJ, Martin KE, et al. Defining desired genetic gains for rainbow trout breeding objective using analytic hierarchy process. *J Anim Sci*. 2012;90:1766–76.
- Gilmour KM, DiBattista JD, Thomas JB. Physiological causes and consequences of social status in salmonid fish. *Integr Comp Biol*. 2005;45:263–73.
- Janhunen M, Kause A, Järvisalo O. Costs of being extreme - Do body size deviations from population or sire means decrease vitality in rainbow trout? *Aquaculture*. 2012;370–371:123–9.
- Janhunen M, Kause A, Vehviläinen H, Jarvisalo O. Genetics of micro-environmental sensitivity of body weight in rainbow trout (*Oncorhynchus mykiss*) selected for improved growth. *PLoS One*. 2012;7:e38766.
- Baras E, Jobling M. Dynamics of intracohort cannibalism in cultured fish. *Aquacult Res*. 2002;33:461–79.
- Mulder HA, Bijma P, Hill WG. Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet Sel Evol*. 2008;40:37–59.
- Mulder H, Hill W, Vereijken A, Veerkamp R. Estimation of genetic variation in residual variance in female and male broiler chickens. *Animal*. 2009;3:1673–80.
- Sae-Lim P, Kause A, Janhunen M, Vehviläinen H, Koskinen H, Gjerde B, et al. Genetic (co) variance of rainbow trout (*Oncorhynchus mykiss*) body weight and its uniformity across production environments. *Genet Sel Evol*. 2015;47:46.
- Ros M, Sorensen D, Waagepetersen R, Dupont-Nivet M, SanCristobal M, Bonnet JC, et al. Evidence for genetic control of adult weight plasticity in the snail *Helix aspersa*. *Genetics*. 2004;168:2089–97.
- Rowe S, White IM, Avendano S, Hill WG. Genetic heterogeneity of residual variance in broiler chickens. *Genet Sel Evol*. 2006;38:617–35.
- Wolc A, White IM, Avendano S, Hill WG. Genetic variability in residual variation of body weight and conformation scores in broiler chickens. *Poult Sci*. 2009;88:1156–61.
- Ibáñez-Escriche N, Moreno A, Nieto B, Piqueras P, Salgado C, Gutiérrez JP. Genetic parameters related to environmental variability of weight traits in a selection experiment for weight gain in mice; signs of correlated canalised response. *Genet Sel Evol*. 2008;40:279–93.
- Ibáñez-Escriche N, Varona L, Sorensen D, Noguera JL. A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal*. 2008;2:19–26.
- Sonesson A, Ødegaard J, Ronnegård L. Genetic heterogeneity of within-family variance of body weight in Atlantic salmon (*Salmo salar*). *Genet Sel Evol*. 2013;45:41.
- Marjanovic J, Mulder H, Khaw H, Bijma P. Genetic parameters for uniformity of harvest weight in the gift strain of Nile tilapia estimated using double hierarchical generalized linear models. In: Proceedings of the international symposium on genetics in aquaculture XII, 21–27 June 2015; Santiago de Compostela; 2015. <http://isga2015.acuigen.es/isga-2015-Abstract-Book.pdf>.
- Khaw HL, Ponzoni RW, Yee HY, bin Aziz MA, Mulder HA, Marjanovic J, et al. Genetic variance for uniformity of harvest weight in Nile tilapia (*Oreochromis niloticus*). *Aquaculture*. 2016;451:113–20.
- Mulder HA, Bijma P, Hill WG. Prediction of breeding values and selection response with genetic heterogeneity of environmental variance. *Genetics*. 2007;175:1895–910.
- Sae-Lim P, Gjerde B, Nielsen HM, Mulder H, Kause A. A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Rev Aquacult*. 2015;8:369–93.
- Marjanovic J, Mulder HA, Khaw HL, Bijma P. Genetic parameters for uniformity of harvest weight and body size traits in the GIFT strain of Nile tilapia. *Genet Sel Evol*. 2016;48:41.
- Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. London: Pearson; 1996.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001;157:1819–29.
- Goddard M, Hayes B, Meuwissen THE. Genomic selection in farm animal species—lessons learnt and future perspectives. In: Proceedings of the 9th world congress on genetics applied to livestock production, 1–6 August 2010; Leipzig. 2010.
- Misztal I, Aggrey SE, Muir WM. Experiences with a single-step genome evaluation. *Poult Sci*. 2013;92:2530–4.
- Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci*. 2009;92:4648–55.
- Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*. 2009;92:4656–63.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
- Gilmour AR, Gogel BJ, Cullis BR, Thompson R. ASReml User Guide Release 4.0. Hemel Hempstead: VSM International Ltd; 2012.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
- Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score 1. *J Dairy Sci*. 2010;93:743–52.
- Christensen OF, Lund MS. Genomic prediction when some animals are not genotyped. *Genet Sel Evol*. 2010;42:2.
- Calus MPL, Vandenplas J. calc\_grm—a program to compute pedigree, genomic, and combined relationship matrices. Animal Breeding and Genomics Centre: Wageningen; 2015.
- Hill WG, Mulder HA. Genetic analysis of environmental variation. *Genet Res (Camb)*. 2010;92:381–95.
- Rönnegård L, Felleki M, Fikse F, Mulder H, Strandberg E. Genetic heterogeneity of residual variance—estimation of variance components using double hierarchical generalized linear models. *Genet Sel Evol*. 2010;42:8.
- Felleki M, Lee D, Lee Y, Gilmour AR, Rönnegård L. Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models. *Genet Res (Camb)*. 2012;94:307–17.
- Lande R. On comparing coefficients of variation. *Syst Zool*. 1977;26:214–7.
- Box GE, Cox DR. An analysis of transformations. *J R Stat Soc Ser B Stat Methodol*. 1964;26:211–52.
- Sakia R. The Box–Cox transformation technique: a review. *Statistician*. 1992;41:169–78.
- R Development Core Team. R: a language and environment for statistical computing. Vienna: The R Foundation for Statistical Computing; 2011.
- Hoaglin DC, Welsch RE. The hat matrix in regression and ANOVA. *Am Stat*. 1978;32:17–22.

40. Felleki M, Lundeheim N. Genetic control of residual variance for teat number in pigs. *Proc Assoc Advmt Anim Breed Genet*. 2013;20:538–41.
41. Mulder HA, Visscher J, Fablet J. Estimating the purebred–crossbred genetic correlation for uniformity of eggshell color in laying hens. *Genet Sel Evol*. 2016;48:39.
42. Sell-Kubiak E, Wang S, Knol EF, Mulder HA. Genetic analysis of within-litter variation in piglets' birth weight using genomic or pedigree relationship matrices. *J Anim Sci*. 2015;93:1471–80.
43. Mulder HA, Crump RE, Calus MPL, Veerkamp RF. Unraveling the genetic architecture of environmental variance of somatic cell score using high-density single nucleotide polymorphism and cow data from experimental farms. *J Dairy Sci*. 2013;96:7306–17.
44. Veerkamp RF, Mulder HA, Thompson R, Calus MPL. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *J Dairy Sci*. 2011;94:4189–97.
45. Yang Y, Christensen O, Sorensen D. Analysis of a genetically structured variance heterogeneity model using the Box–Cox transformation. *Genet Res (Camb)*. 2011;93:33–46.
46. Daetwyler HD, Villanueva B, Woolliams JA. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*. 2008;3:e3395.
47. Colwell DJ, Gillett JR. 66.49 Spearman versus Kendall. *Math Gaz*. 1982;66:307–9.

Submit your next manuscript to BioMed Central  
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

