



Norwegian University  
of Life Sciences

Master's Thesis 2016 30 Credits  
Department of Animal and Aquacultural Sciences (IHA)

**Statistical model comparison in  
genetic analysis of challenge test  
data on *Streptococcus agalactiae*  
resistance in Nile tilapia  
(*Oreochromis niloticus*)**

Meng Lin  
Master programme in Aquaculture

**Statistical model comparison in genetic  
analysis of challenge test data on  
Streptococcus agalactiae resistance in Nile  
tilapia (*Oreochromis niloticus*)**

**By**

**Meng Lin**

**Supervisor**

**Professor Hans Magnus Gjøen**

Department of Animal and Aquaculture Sciences  
Norwegian University of Life and Sciences (NMBU)

June 2016

# Contents

<b>List of tables</b> .....	ii
<b>List of figures</b> .....	iii
<b>Acknowledgement</b> .....	iv
<b>Abstract</b> .....	v
1. Introduction .....	1
2. Material and Methods .....	2
2.1 Experiment.....	2
2.1.1 Fish preparation and experimental design.....	2
2.1.2 Tag and challenge.....	4
2.1.3 Data recording and collection .....	5
2.2 Statistical analysis.....	5
2.2.1 Trait definition .....	5
2.2.2 Statistical model.....	7
2.2.3 Heritabilities .....	10
2.2.4 Model comparison - Cross validation and Pearson correlation coefficients ...	11
3. Results.....	13
3.1 Overview of survival.....	13
3.2 Variance components and heritability .....	15
3.3 model comparison by Pearson correlation coefficients.....	16
4. Discussion.....	16
5. Conclusion.....	18
<b>Reference</b> .....	19
<b>Appendix</b> .....	21
Appendix 1. Variance components and heritability calculation in <b>AsReml</b> .....	21
Appendix 2. Pearson correlation coefficient between family EBV's for different models in <b>R Studio</b> .....	25
Appendix 3 Pearson correlation within models calculation (Cross Validation).....	28
Appendix 3.1 First split the data into two part evenly in <b>R Studio</b> .....	28
Appendix 3.2 Rerun the separated data one by one correspond to the same model in AsReml by just Adding '! CONTINUE !BLUP' after the data line. (e.g.in BTH_Sire dam model).....	28
Appendix 3.3 Mid-parent full-sib family EBVs and Pearson correlation coefficients between the predicted family EBV's. Calculated in R studio (Eg. for BTH_Sire dam model) .....	29

## List of tables

<b>Table 1.</b> Mating design; one sire mated to one dam. ....	3
<b>Table 2.</b> Design of experiment. Fish from different families and treatments were identifiable by tags. Fish that was not injected with <i>Streptococcus agalactiae</i> , were challenged by cohabitants. The <i>Streptococcus agalactiae</i> challenge isolate code was AL20109-4, which was diluted to 1:5000 and an injection of 0.05 ml was used. ....	3
<b>Table 3. Overall challenge materials</b> .....	5
<b>Table 4.</b> Basic statistical data as input parameters.....	13
<b>Table 5.</b> Variance components and heritability with standard errors for five models.....	15
<b>Table 6.</b> Pearson correlation coefficients between family EBV's of different models .....	15
<b>Table 7.</b> Pearson correlation coefficients within models (between estimated family EBV's from two sub set data) and accuracy of selections ( $r\tau = rEBV$ ). ....	15

## List of figures

<b>Figure 1.</b> Survival rate during the 18 test period of cohabitant families in G8 group.....	13
<b>Figure 2.</b> Survival rate during the 18 test period of cohabitant families in G35 group.....	14
<b>Figure 3.</b> Survival rate during the 18 test period of IP-injected families in G8 group.....	14
<b>Figure 4.</b> Survival rate during the 18 test period of IP-injected families in G35 group.....	14

## **Acknowledgement**

I would express my sincerely thank to my supervisor Professor Hans Magnus GjØen for guiding me through the whole period of my thesis work with statistical model building, data reprocessing and idea for model comparison. Without his help, I would not have been able to finish this thesis as expected.

Secondly I would also express my sincerely thank to Associate professor JØrgen Ødegård for his professional help in trying different models and performing them in AsRepl.

Thirdly, I would deeply thank to PhD student Xiaodong Wang for his professional help with R studio.

Lastly I would like to thank my dearest mom for her grate love to me throughout my two-year master life in Norway.

## Abstract

Five statistical models were used to analysis genetic variations, heritability and the ability to predict survival of Nile tilapia based on challenge test data. 12 full-sib Nile tilapia families (1870 individuals) with an average weight of 8g in *young group (G8)* and 35g in *older group (G35)* were tested. Sire-dam binary threshold (BTH\_sire-dam) model and Animal model (BTH\_Animal) were using test-period survival as the univariate trait definition. And sire-dam survival score (SS) model was applied to test-day survival (one record per fish per day). Also a family mean survival (FMS) model and a family mean area (FMA) model were tested with the mean survival rate and integral area per family as trait definitions, respectively. Highest heritability was found with the FMS model ( $0.56 \pm 0.42$ ), followed by the BTH\_sire-dam model ( $0.45 \pm 0.17$ ) and the FMA model ( $0.42 \pm 0.34$ ). The BTH\_Animal model got a relatively much lower value of heritability ( $0.27 \pm 0.11$ ), whereas the SS model ( $0.07 \pm 0.03$ ) was distinctly the lowest. The highest Pearson correlation coefficients between predicted breeding values for the various models were 0.97 for the BTH\_sire-dam model and the BTH\_Animal model, 0.96 for the SS model, 0.87 for the FMA model and the FMS model had the lowest prediction power with a correlation of 0.77. In conclusion, BTH\_sire-dam model was the best fitted model with a high heritability and the highest selection accuracy among these five models.

## 1. Introduction

Tilapia is one of the most important farmed fish species in the world, only second to carp (Linnaeus, 1758) and among the tilapia species, Nile tilapia (*Oreochromis niloticus*) is the preferred one to culture. This African omnivore fish is well known for rapid growth, high tolerance for intensive culture and high disease resistance. However, this hardiness is of course not without limitations, according to the long time experience from fish farming industries and lots of experiment done in the lab. Fish farmers and scientists have found that disease may happen to be the first and major difficulty that need to be overcome in the future tilapia farming industry (<http://www.thefishsite.cn/articles/190/streptococcus-in-tilapia/>).

Non-haemolytic group B *Streptococcus agalactiae* (GBS) is a streptococcal disease (Eldar et al., 1995) which has a wide range of host (Joyce et al., 2009). It can not only influence human beings to get meningitis but also lead to meningoencephalitis in fish and mastitis in cattle (Joyce et al., 2009). It is one of the main causes to streptococcosis in Nile tilapia, which can lead to high mortality. Fish can get abnormal behavior, abscess, skin hemorrhage, less feed intake, septicaemia and peritonitis after being infected by GBS. This specific pathogen has a significant effect on fish central nerve system (<http://www.thefishsite.cn/articles/190/streptococcus-in-tilapia/>).

The experiment was primarily designed to develop a *Streptococcus agalactiae* challenge model by intraperitoneal (IP) injection and cohabitation in Nile tilapia. But the recorded challenge test data is here used for later statistical genetic variance analysis.

Many models can be found to perform analysis of genetic variance of challenge test data in aquatic species, i.e. their resistant to various diseases, dependent on various trait definitions (Ødegård et al., 2006). An animal threshold model usually give biased heritability in genetic analysis of binary traits. However, problems can be solved with a new Gibbs sampling algorithm especially when individual records exist on parents (Ødegård et al., 2010). A linear model and a threshold model were used for the test



period survival data, i.e. dead or alive (usually scaled 0 or 1) at the end of the test period. Linear repeatability model was used for test-day survival and proportional hazards Weibull model were used in time until death occasion (Ødegård et al., 2006, Gitterle et al., 2006). Different models were using different information obtained from the challenge test. That to find a best fitted model for a specific disease resistance challenge test in aquatic species was of significant important to predict the breeding value for candidates.

The aim of this thesis was to apply different statistical models for genetic analysis on *Streptococcus agalactiae* resistance in Nile tilapia and make comparison of the model's prediction ability for the survival trait. Animal and sire-dam threshold linear models and family mean models were tested and compared.

## **2. Material and Methods**

### **2.1 Experiment**

#### **2.1.1 Fish preparation and experimental design**

All tilapia families used in this experiment originated from Genomar's facilities in the Philippines, and they arrived at ABT Innovia (an independent testing facility located in Malta) (<http://www.abtinnovia.com/>) on 28 May 2015, with an average weight of ca 0.05g. Different families were separately held in a Recirculating Aquaculture system (RAS), and were raised at a temperature of around  $28\pm 1$  °C. This study was designed to be a repeated experiment, with the fish size as the only difference between the two treatments, i.e. the type of challenge tests. The specific fish sizes were 8g for *G8* and 35g for *G35*, respectively. Each treatment consisted of 12 families, with 80 individuals per family. After a period of on-growing, the two challenge test were commenced on the 4<sup>th</sup> and 25<sup>th</sup> of August, 2015, respectively (AquaBioTech Group, 2015).

The mating design is shown in Table 1, and the experimental design is shown in Table 2. The 80 individuals were separated into 30 fish being intraperitoneally injected (*IP*) with *Streptococcus agalactiae* (pathogen) and 50 not being injected (*Naïve*). In each of

the two challenge options, fish were divided evenly into two groups and put into different tanks. It means that two challenge ways were duplicated, i.e. 4 batches per family in total (AquaBioTech Group, 2015).

**Table 1.** Mating design; one sire mated to one dam.

Family	Sire_ID	Sire_Family	Dam_ID	Dam_Family
F01	447636	G24-158	18221	G24-113
F02	447203	G24-148	15608	G24-127
F03	447566	G24-151	401281	G24-117
F04	446943	G24-161	407580	G24-101
F05	447554	G24-151	17615	G24-121
F06	448934	G24-147	18887	G24-117
F07	448346	G24-138	16081	G24-124
F08	448919	G24-147	16753	G24-116
F09	448264	G24-162	403290	G24-107
F10	447543	G24-134	16263	G24-105
F11	447379	G24-140	18304	G24-105
F12	448148	G24-148	18283	G24-127

**Table 2.** Design of experiment. Fish from different families and treatments were identifiable by tags. Fish that was not injected with *Streptococcus agalactiae*, were challenged by cohabitants. The *Streptococcus agalactiae* challenge isolate code was AL20109-4, which was diluted to 1:5000 and an injection of 0.05 ml was used.

Family	Tank No.	No. of fish	Batch	Challenge type
F01	T1	25	1	Cohabitant
F01	T7	25	2	Cohabitant
F02	T1	25	1	Cohabitant
F02	T7	25	2	Cohabitant
F03	T2	25	1	Cohabitant
F03	T8	25	2	Cohabitant
F04	T2	25	1	Cohabitant
F04	T8	25	2	Cohabitant
F05*	T3	25	1	Cohabitant
F05*	T9	25	2	Cohabitant
F06	T3	25	1	Cohabitant
F06	T9	25	2	Cohabitant
F07	T4	25	1	Cohabitant
F07	T10	25	2	Cohabitant
F08	T4	25	1	Cohabitant
F08	T10	25	2	Cohabitant
F09	T5	25	1	Cohabitant
F09	T11	25	2	Cohabitant

F10	T5	25	1	Cohabitant
F10	T11	25	2	Cohabitant
F11	T6	25	1	Cohabitant
F11	T12	25	2	Cohabitant
F12	T6	25	1	Cohabitant
F12	T12	25	2	Cohabitant
F01	T1	15	3	Injected
F01	T7	15	4	Injected
F02	T1	15	3	Injected
F02	T7	15	4	Injected
F03	T2	15	3	Injected
F03	T8	15	4	Injected
F04	T2	15	3	Injected
F04	T8	15	4	Injected
F05*	T3	15	3	Injected
F05*	T9	15	4	Injected
F06	T3	15	3	Injected
F06	T9	15	4	Injected
F07	T4	15	3	Injected
F07	T10	15	4	Injected
F08	T4	15	3	Injected
F08	T10	15	4	Injected
F09	T5	15	3	Injected
F09	T11	15	4	Injected
F10	T5	15	3	Injected
F10	T11	15	4	Injected
F11	T6	15	3	Injected
F11	T12	15	4	Injected
F12	T6	15	3	Injected
F12	T12	15	4	Injected

\* Family F05 in G35 (with 35g average fish weight) has only 48 fish available before challenge. So only 30 fish were used in IP injection and no fish for Naïve

### 2.1.2 Tag and challenge

Materials for challenge was 1 ml vial with *Streptococcus agalactiae* which was stored at -80°C with isolate coded: AL 20109-4 ( $7.3 \times 10^9$  cell/ml). 50 ml challenge dilution was used (phosphate buffer saline (PBS)) with a density of 1:5000 ( $1.46 \times 10^6$  cell/ml)(AquaBioTech Group, 2015).

After 24 hours of starvation, all fish were anaesthetized with MS222 (Tricaine Pharmaq, Pharmaq Ltd, UK). Then each fish was tagged with a visible implant elastomer (VIE) (Northwest Marine Technology Inc., USA); using the red and blue tags left for *Naïve* and orange and green tags right for *IP*. In *IP* group, each fish was injected with 0.05 ml challenge dilution ( $7.3 \times 10^4$  cell/fish) between the pelvic fins (AquaBioTech Group, 2015).

### 2.1.3 Data recording and collection

Dead fish were picked, labeled and stored at  $-20^{\circ}\text{C}$ . Records for dead fish were done with tank number, no. of fish died, date and tag. The total duration of the experiment was 18 days. For the whole experimental period, three of the newly dead fish were randomly selected to have a post mortem kidney test to confirm that for the cause of death is infection with *Streptococcus agalactiae*.

**Table 3. Overall challenge materials**

Challenge material	<i>Streptococcus agalactiae</i>
Challenge dilution	Phosphate buffer saline (PBS)
Amount of challenge material	1ml
Amount of challenge dilution	50ml
Cell density of challenge material	$7.3 \times 10^9$ cell/ml
Density of challenge dilution	1:5000 ( $1.46 \times 10^6$ cell/ml)
Store temperature for challenge material	$-80^{\circ}\text{C}$
Anaesthetic	MS222
Challenge dilution IP injected per fish	0.05ml ( $7.3 \times 10^4$ cell/fish)

## 2.2 Statistical analysis

### 2.2.1 Trait definition

Four different kinds of trait definitions were made in this study in order to test different models:

i. Binary test-period survival (BTS) per individual (Gianola et al., 1983)

Fish was scored 0 when it dies before the end of the whole experimental period, and 1 if it stayed for alive.

ii. Binary test-day survival (BSS) per individual (Veerkamp et al., 2001)

Status of each fish was recorded every day until it died. That means the number of record per fish is equal to the number of survival day of fish in the experiment. The score method is the same as for BTS above, i.e. 0 is dead and 1 is alive. For instance, when a fish died at day 4, the survival score should be [1 1 1 0] and when fish survived for the whole 18 days of experimental period, the survival score should be [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]. Each animal thus gets a record for each day it survives, plus eventually the day it dies (except for the fish who survived during the whole test period).

iii. Survival rate (SR) per family

The number of dead fish were recorded every day, and the survival rate is thus equal to the number of fish died divide by total fish number:

$$SR = \frac{No. fish died}{No. total fish}$$

Hence, we can calculate the survival rate for each batch and day within the 12 families. The test day with the highest variance among average survival rates were eventually chosen to be the records to be used for this model.

iv Integral area (IA) of survival rate curve per family

Survival curves fit well to a reversed logistic curve. The corresponding formula is as follow:

$$Y = \frac{e^z}{1+e^z} , \text{ where } z = b_0 + b_1 * X$$

Y: Survival rate

X: Experiment day

$b_0$  and  $b_1$ : Regression coefficients which were estimated in R studio.

From the above formula, we can get the relationship function between X and Y:

$$Y = \frac{e^{b_0+b_1*X}}{1 + e^{b_0+b_1*X}}$$

IA is then the integral, i.e. the area under the curve for the whole experiment period here in this case from day 1 to day 18.

$$IA = \int_1^{18} \frac{e^{b_0+b_1*X}}{1 + e^{b_0+b_1*X}} dX$$

### 2.2.2 Statistical model

Different models were applied in order to make comparison and find the best fitted model for challenge test data of *streptococcus agalacitae* resistance in Nile tilapia. Variance components and heritabilities were calculated among the models shown as follow:

- a. Sire-dam binary threshold (BTH\_sire-dam) model (Gianola et al., 1983):

This model is based on the first trait definition, BTS, which means that all fish are scored 0 for dead and 1 for alive for the duration of the whole test period. Here an unobserved normal underlying liability variable,  $l$ , (Ødegård et al., 2006) is assumed to determine the observed value  $m$  ( $m=0,1$ ) of the trait  $Y_{ijk}$ . When  $l_{ijk} \leq 0$ ,  $Y_{ijk}$  is equal to 0 and when  $l_{ijk} > 0$ ,  $Y_{ijk}$  is equal to 1. This is the standard probability threshold model:

$$\Pr(Y_{ijk} = m) = \Pr(l_{ijk} > 0)^m \Pr(l_{ijk} \leq 0)^{1-m} = \Phi(l_{ijk})^m [1 - \Phi(l_{ijk})^{1-m}]$$

Where  $\Phi(\cdot)$  is the cumulative distribution function.

If  $m=0$ , which means fish died during the test period

$$\Pr(Y_{ijk} = 0) = \Pr(l_{ijk} \leq 0) = 1 - \Phi(l_{ijk})$$

If  $m=1$ , which means fish survived during the test period

$$\Pr(Y_{ijk} = 1) = \Pr(l_{ijk} > 0) = \Phi(l_{ijk})$$

And the model for underlying liability variable  $l$  is as follow:

$$l_{ijk} = \mathbf{Xb} + (\mathbf{Z}_s + \mathbf{Z}_d)\mathbf{u} + \mathbf{Wc} + \mathbf{e}$$

Where

- l<sub>ijk</sub>** underlying liability of  $Y_{ijk}$
- X** incidence matrix for fixed effects
- b** vector of fixed effects
- Z<sub>s</sub>** incidence matrix for sire
- Z<sub>d</sub>** incidence matrix for dam
- u** random additive genetic effects of sires and dams
- W** incidence matrix for common environmental effect
- c** random common environmental effect
- e** random residual effect

b. Binary threshold animal model (BTH\_Animal):

The BTH\_Animal model is using the BTS trait definition as threshold model, as the BTH\_sire-dam model above. Differences between animal model and sire-dam model is that EBVs here are calculated for each animal, instead of for only the parents, i.e. sire and dam. Variance of sire is equal to variance of dam and is 1/4 of the additive genetic variance in sire-dam model.

$$Y_{ijk} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wc} + \mathbf{e}$$

- Z:** incidence matrix for individual animal

**a:** random animal additive genetic effect

Other variable components are as described above

c. Sire-dam survival score (SS) model:

The SS model is similar to the linear repeatability model (Ødegård et al., 2006) that use the binary test-day survival trait definition (BSS, described above) as input data:

$$Y_t = \sum_{p=0}^4 \beta_p Z_p(t) + (Z_s + Z_d)u + Wc + e$$

**Y<sub>t</sub>** individual animal survival (0 for dead and 1 for alive) at day t of experiment.

**Z<sub>p</sub>(t)** pth order orthogonal polynomial of t

**β<sub>p</sub>** pth order regression coefficient

Other variable components are as described above

d. Family mean survival (FMS) model:

FMS model was using the SR trait definition described above, with family mean survival rates as the trait. Since there were four batches per family in both the G8 and G35 groups, there are four family mean survival rates available in each family per group. The model is expressed as follow:

$$y = \mu + P + (Z_{sf} + Z_{df})u + W_f c + e$$

**Where**

**y** is the mean survival rate per family

**μ** is a vector of overall mean survival rates per batch

**P** is the fixed effect of weight, batch and challenge type

**Z<sub>sf</sub>** incidence matrix for sire of family mean

**Z<sub>df</sub>** incidence matrix for dam of family mean



$W_f$  incidence matrix for common environmental effect of family mean

Other variable components are as described above

e. Family mean area (FMA) model:

FMA model was using the IA trait definition described above, with the integral area of the survival curve as the trait. The model is otherwise the same as the FMS model above.

For the respective models above, the random additive genetic effects of sires and dams were assumed to have  $\mathbf{u} \sim N(0, A\sigma_u^2)$ , the random animal additive genetic effect were assumed to have  $\mathbf{a} \sim N(0, A_1\sigma_a^2)$ , the random common environmental effect were assumed to be  $\mathbf{c} \sim N(0, I\sigma_c^2)$ , the random residual variance were assumed to be  $\mathbf{e} \sim N(0, I\sigma_e^2)$  in the FMS and FMA models, 1 in the BTH\_sire- dam and BTH\_Animal models and finally  $\pi^2/3 \sim 3.3$  in the SS model (logistic distribution). Where A is the genetic relationship matrix, and I is an identity matrix. All the models were analyzed in ASREML 4 (Gilmour et al., 2015).

### 2.2.3 Heritabilities

Calculation of heritability for all models:

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

Where

$\sigma_G^2$ : Genetic variance

$\sigma_P^2$ : Phenotypic variance

and

$\sigma_P^2 = \sigma_a^2 + \sigma_c^2 + \sigma_e^2$  in animal model

$\sigma_P^2 = \sigma_s^2 + \sigma_d^2 + \sigma_c^2 + \sigma_e^2$  in sire-dam model

$\sigma_a^2$ : Additive genetic variance of animals

$\sigma_c^2$ : Common enviromental familiy variance

$\sigma_e^2$ : Residual variance, which is stricted to 1 in the threshold model

$\sigma_s^2$ : Additive genetic sire variance

$\sigma_d^2$ : Additive genetic dam variance

The difference of phenotypic variance between animal model and sire-dam model lead to a different heritability calculation.

### **Animal model (BTH\_Animal)**

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}$$

Residual variance is strict to 1

### **Sire-dam linear model (FMS and FMA)**

$$h^2 = \frac{4\sigma_u^2}{4\sigma_u^2 + \sigma_c^2 + \sigma_e^2}$$

$\sigma_u^2$ : Additive genetic sire dam variance, is equal to both sire and dam variance,  $\sigma_u^2 = \sigma_s^2 = \sigma_d^2 = 1/4\sigma_a^2$ , which means that additive genetic sire-dam effect express 1/4 of the total additive genetic variance.

### **Sire-dam threshold model (BTH\_sire-dam and SS)**

$$h^2 = \frac{4\sigma_u^2}{2\sigma_u^2 + \sigma_c^2 + \sigma_e^2}$$

Residual variance is restricted to 1 in the BTH\_sire-dam model and  $\pi^2/3$  in SS model (Ødegård et al., 2010).

#### **2.2.4 Model comparison - Cross validation and Pearson correlation coefficients**

A direct model comparison for the above five models can for some aspects be done through the estimated heritabilities. However, these models are using three different survival trait definitions, thus the models cannot be compared well directly through the

estimated heritabilities (Ødegård et al., 2006). An alternative method is to use predicted full-sib family (mid-parent) EBV's estimated in independent subsamples, and compare the prediction power by means of cross validation. The main idea is to randomly split the dataset into two sub sets (here performed in R; program given in Appendix 3.1) and run the same model (performed in AsReml 4) again using the variance components obtained from entire data set as posterior variances, and then estimate BVs in one iteration (e.g. Bangera et al., 2014). The purpose is to get the predicted breeding values of each family from two sub sets of the data and then make comparison. The mid-parent full-sib family EBV's are calculated as follow:

$$EBV_{ij} = (Z_{sij} + Z_{dij})/2$$

$EBV_{ij}$  vector of family EBV's in model  $i$  (1,2, ...,5) and subset  $j$  (1,2)

$Z_{sij}$  vector of predicted sire EBV's in model  $i$  and subset  $j$

$Z_{dij}$  vector of predicted dam EBV's in model  $i$  and subset  $j$

Pearson correlation coefficients,  $r_{EBVi}$  for each model is in this case the correlation between two sub sets of the predicted family EBVs ( $EBV_{i1}$  and  $EBV_{i2}$ ). In selection index theory, the Pearson correlation coefficients are close to the accuracy of selection's square in models (Gitterle et al., 2006).

$$r_{EBV} \approx r_T^2$$

Thus the predicted mid-parent family EBV's are an independent and unbiased estimator of the true breeding value of parents because there had a similar amount of information per full-sib family (Gitterle et al., 2006). Hence the predict ability of the models can be compared by their accuracy of selections (square root of the estimated  $r_{EBV}$ ). Furthermore, the Pearson correlation coefficients between full-sib family EBV's (obtained from the entire data set per model) were also calculated in order to test the similarity in ranking of families among different methods (Bangera et al., 2014).

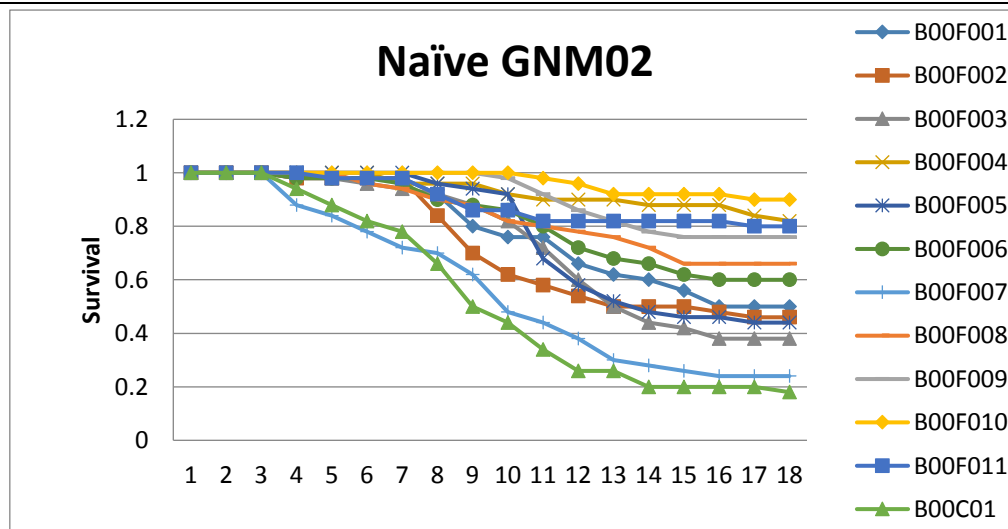
### 3. Results

#### 3.1 Overview of survival

Summary statistics of the dataset used is given in Table 3. The overall survival of fish was 44.3% at the end of the experiment. Nile tilapia survival rate per family were calculated and plotted for both challenge methods (IP-injected and cohabitation) in *G8* (8g average body weight) and *G35* (35g average body weight) groups. Details are shown in figure 1 to figure 4 below. It is obvious that cohabitant families had a much higher survival than intraperitoneal injected families in both groups. For cohabitant families, the overall mean survival rate was 56.2% in the *G8* group and 72.2% in the *G35* group. However, for the IP-injected families the overall mean survival rate was 16.7% in the *G8* group and 33.9% in the *G35* group.

**Table 4.** Basic statistical data as input parameters

Item	
Total fish, no.	1870
Full-sib families, no.	12
Sire, no.	12
Dam, no.	12
Fish per family, no.	80
Average mortality (%)	55.7
Length of recording period for both trails, days	18
Average weight at the beginning of the test day ( <i>G8</i> )	8g
Average weight at the beginning of the test day ( <i>G35</i> )	35g



**Figure 1.** Survival rate during the 18 test period of cohabitant families in *G8* group



### 3.2 Variance components and heritability

Variance components and heritabilities with standard errors for the different models above are showed in Table 5. The linear family mean models have a relative higher heritability with the FMS model  $h^2 = 0.56 \pm 0.42$  and the FMA model  $h^2 = 0.42 \pm 0.34$ . The BTH\_sire-dam model also gave a high heritability  $0.45 \pm 0.17$ , followed by the BTH\_Animal model ( $0.27 \pm 0.11$ ) and the SS model ( $0.07 \pm 0.03$ ).

**Table 5.** Variance components and heritability with standard errors for five models

Parameters	Models				
	BTH_sire-dam	BTH_Animal	SS	FMS	FMA
$h^2 \pm SE$	$0.45 \pm 0.17$	$0.27 \pm 0.11$	$0.07 \pm 0.03$	$0.56 \pm 0.42$	$0.42 \pm 0.34$
$\sigma_c^2 \pm SE$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.05 \pm 0.24$	$0.003 \pm 0.16$
$\sigma_a^2 \pm SE$	$0.58 \pm 0.28$	$0.06 \pm 0.03$	$0.26 \pm 0.12$	$0.04 \pm 0.05$	$8.8 \pm 10.25$
$c^2 \pm SE$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.00 \pm 0.00$	$0.003 \pm 0.01$	$0.07 \pm 3.28$

**Table 6.** Pearson correlation coefficients between family EBV's of different models

Models	BTH_sire-dam	BTH_Animal	SS	FMS	FMA
BTH_sire-dam	-	$\approx 1.00^a$	0.97	0.88	0.82
BTH_Animal	$\approx 1.00^a$	-	0.96	0.87	0.80
SS	0.97	0.96	-	0.96	0.92
FMS	0.88	0.97	0.96	-	0.95
FMA	0.82	0.80	0.92	0.95	-

<sup>a</sup> The Pearson correlation coefficient is very close to 1, but not the exact value 1.

**Table 7.** Pearson correlation coefficients within models (between estimated family EBV's from two sub set data) and accuracy of selections ( $r_\tau = \sqrt{r_{EBV}}$ ).

Models	$r_{EBV}$	$r_\tau$
BTH_sire-dam	0.97	0.98
BTH_Animal	0.97	0.98
SS	0.96	0.98
FMS	0.77	0.88
FMA	0.87	0.93

### 3.3 model comparison by Pearson correlation coefficients

Pearson (rank) correlation coefficients were calculated both between and within family EBV's, for the five statistical models. Results are showed in Table 6 and Table 7 above.

## 4. Discussion

The BTH\_sire-dam and the BTH\_Aniaml models are using the same trait definition, and the estimated heritabilities were in this case similar. Higher heritability in the BTH\_sire-dam model indicates a better ability to get unbiased additive genetic variances with the sire-dam model compared to the animal model. For a cross sectional animal threshold model, applied to a small data set, Ødegård et al., (2010) found large biases, especially when combined with an unfavorable fix effect structure. The heritability value 0.07 for the SS model is the lowest among the five models. However, it is reasonable compared to the heritability estimates of disease resistance to bacteria disease found in other aquatic species with the same model. For instance,  $h^2 = 0.024$  was estimated in Atlantic salmon (*Salmo salar*) for resistance to furunculosis using a linear repeatability model, which is similar to the SS model (Ødegård et al., 2006).

Large standard errors of the heritability estimates occurred in Table 5, especially for the FMS (0.42) and FMA (0.34) models, and they ranged from 0.03 to 0.42 among the five models. The main reason for that could be the small data set, with only 12 full-sib families. Although the total number observations were quite high, with 1870 individuals recorded (i.e. 80 individuals per family). To increase the number of full-sib families is likely to lower the standard errors considerably. For the FMS and FMA models, the average survival rate and the integral area were made as dependent variables per batch (4 batches in one family), respectively.

No common environmental effects were found in the BTH\_sire-dam, BTH\_Animal and SS models. Although different fish families were separately held in different tanks until tagging, the environmental conditions were kept almost the same, and the common

environmental variances were thus ignorable. As for the common environmental effects, which occurred in the FMS ( $0.003 \pm 0.01$ ) and FMA ( $0.07 \pm 3.28$ ) models, preliminary analysis of the original data that making batch mean survival rate and the integral of survival curve should account for the deviations.

Table 6 gives the correlation between family EBV's of the entire data set for pairwise comparison of the five specific models. Overall, high correlations (ranging from 0.80 to  $\approx 1.0$ ) were obtained for all the models. The highest rank coefficient was found between the BTH\_sire-dam and BTH\_Animal models, which was close to 1. Also, high rank coefficients were found between the SS and BTH\_sire-dam models (0.97), and the SS and BTH\_Animal models (0.96), respectively. This indicate that these three models had a very similar ranking of families (Gitterle et al., 2006). The result is coherent with the fact that the BTH\_sire-dam and BTH\_Animal models were using test-period survival as input trait definition, while the SS model was using test-day survival. Likewise, rank coefficient between FMS and FMA models was 0.95, which is logical because they were using the same type of family mean data.

Table 7 shows the Pearson correlation coefficients between estimated family (mid-parent) EBV's ( $r_{EBV}$ ) within each model and also the corresponding selection accuracy ( $r_{\tau}$ ) for each model. The high  $r_{\tau}$  (0.88 to 0.98) obtained, indicate that predictive ability (accuracy) of the model is 88% to 98% of the theoretical maximum family selection accuracy (Bangera et al., 2014). The ranking of models were comparable for both the BTH\_sire-dam and the BTH\_Animal models (0.98), followed by the SS model (0.98), the FMA model (0.93) and the FMS model (0.88).

Such high selection accuracies were not expected, especially with the limited data available. As the data sub-set for cross validation was randomly separated into two even parts, it could possibly be more precise to perform a 5 fold or even more cross validation analysis.



## **5. Conclusion**

All models were performed relatively well, both by the perspective of heritability for resistance to *Streptococcus agalactiae* in Nile tilapia and by the perspective of predictive ability (accuracy of selection). For the data structure that we have investigated, the Binary threshold sire-dam (BTH\_sire-dam) model is the best choice among the tested models, with a heritability of 0.45 and selection accuracy of 0.98. More full-sib families and individual per families were recommended to be involved in this genetic analysis in order to get more accurate calculated EBVs.

## Reference

- AquaBio Tech Group. 2015. GNM03 Report 29oct15.docx.
- Bangera, R., Ødegård, J., Mikkelsen, H., et al, 2014. Genetic analysis of francisellosis field outbreak in Atlantic cod (*Gadus morhua* L.) using an ordinal threshold model. *Aquaculture* 420-421, S50-S56.
- Eldar, A., Bejerano, Y., Livoff, A., et al, 1995. Experimental streptococcal meningoencephalitis in cultured fish. *Veterinary Microbiology*. 43, 33-40.
- Gianola, D., & Foulley, J. L. 1983. Sire evaluation for ordered categorical data with a threshold model. *Genetique, selection, evolution*, 15(2), 201-224.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R., Welham, S.J., Thompson, R., 2015. ASReml User Guide Release 4.1 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK [www.vsn.co.uk](http://www.vsn.co.uk)
- Gitterle, T., Ødegård, J., Gjerde, B., Rye, M., Salte, R., 2006. Genetic parameters and accuracy of selection for resistance to White Spot Syndrome Virus (WSSV) in *Penaeus (Litopenaeus) vannamei* using different statistical models. *Aquaculture* 251, 210-218.
- Gjedrem, T., Gjøen, H.M. and Salte, R., 1991. Genetic variation in red cell membrane fragility in Atlantic salmon and rainbow trout. *Aquaculture* 98, 349-354.
- Joyce, J.E., Phillip H.K., David J.P., John, F.B., 2009. Human Streptococcus agalactiae Isolate in Nile Tilapia (*Oreochromis niloticus*). *Emerging Infectious Disease*. 15, 5.
- Linnaeus, 1758. Cultured aquatic species information programme *Oreochromis niloticus*. ([http://www.fao.org/fishery/culturedspecies/Oreochromis\\_niloticus/en#tcNA0050](http://www.fao.org/fishery/culturedspecies/Oreochromis_niloticus/en#tcNA0050))
- Ødegård, J., Olesen, I., Gjerde, B., Klemetsdal, G., 2006. Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (*Salmo salar*): Prediction of field survival. *Aquaculture* 259, 116-123.

Ødegård, J., Meuwissen, T.H., Heringstad, B., Madsen, P., 2010. A simple algorithm to estimate genetic variance in an animal threshold model using Bayesian inference. *Genetics Selection Evolution* 42:29.

Veerkamp, R. F., Brotherstone, S., Engel, B., and Meuwissen, T. H. E. 2001. Analysis of censored survival data using random regression models. *British Society of Animal Science* 72: 1-10.

## Appendix

### Appendix 1. Variance components and heritability calculation in **AsReml**

#### a) **BTH\_sire dam**

Sire & Dam binary threshold model

individual !A !P

sire !A !P

dam !A !P

weight

tankno

batch 8

challenge

BTH

challcode !!

#pedigree

pedigree.csv !SKIP 1 !ALPHA !MAKE

#data

BTH.csv !SKIP 1 !SAVE

BTH !BIN !PROBIT ~ tankno.batch weight challcode !r -individual sire and(dam,1) fac(sire,dam)

Residual units

Predict challcode

VPREDICT !DEFINE

F VarA sire \* 4

F VarC fac(sire,dam)

F VarP VarC + VarA\*0.5 + Residual

H h2 VarA VarP

H C2 VarC VarP

#### b) **BTH\_Animal**

Univariate BTH Animal model

individual !A !P

sire !A !P

dam !A !P

weight

tankno

batch !!

challenge

```

BTH
challcode !!
#pedigree
pedigree.csv !SKIP 1 !ALPHA !Make
#data
BTH.csv !SKIP 1 !SAVE !EXTRA 4 !MAXIT 100
BTH ~ tankno.batch weight challcode !r individual fac(sire,dam)

```

```

VPREDICT !DEFINE
F VarA individual
F VarC fac(sire,dam)
F VarP VarA + VarC + Residual
H h2 VarA VarP
H c2 VarC VarP

```

### c) **SS\_Sire dam**

```

Survival score (SS) model
individual !A !P
sire !A !P
dam !A !P
weight
tankno
batch !!
challenge
ss
day
challcode !!
pedigree.csv !SKIP 1 !ALPHA !Make
#data
SS.csv !SKIP 1 !MAXIT 200
ss !BIN !PROBIT ~ tankno.batch weight challcode pol(day,-4) !r sire and(dam,1) fac(sire, dam)

```

```

VPREDICT !DEFINE
F VarA sire*4
F VarE Residual*3.3
F VarC fac(sire,dam)
F VarP VarA*0.5 + VarC + VarE
H h2A VarA VarP

```

### d) **FM\_Survival**

```

family mean model
family !A !P
sire !A !P
dam !A !P

```

```

batch !!
mean_SR
No_obs !!
weight !!
chal_T !!
area

#pedigree
pedigree1.csv !SKIP 1 !ALPHA !MAKE

#data
family mean.csv !SKIP 1

mean_SR !WT No_obs ~ mu weight batch chal_T !r sire and(dam,1) fac(sire,dam)

```

```

VPREDICT !DEFINE
F VarA sire * 4
F VarC fac(sire,dam)
F VarP VarA + VarC + Residual
H h2 VarA VarP
H C2 VarC VarP

```

#### e) FM\_Area

```

family mean model
family !A !P
sire !A !P
dam !A !P
batch !!
mean_SR
No_obs !!
weight !!
chal_T !!
area

#pedigree
pedigree1.csv !SKIP 1 !ALPHA !MAKE

#data
family mean.csv !SKIP 1
area !WT No_obs ~ mu weight batch chal_T !r sire and(dam,1) fac(sire,dam)

```

```

VPREDICT !DEFINE
F VarA sire * 4
F VarC fac(sire,dam)

```

F VarP VarC + VarA + Residual

H h2 VarA VarP

H C2 VarC VarP

## Appendix 2. Pearson correlation coefficient between family EBV's for different models in **R Studio**

```
EBV <- read.table("new4.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV4 <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV4)
write.csv(FamilyEBV,'Pro4.csv')

EBV <- read.table("new7b.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
```



```

FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV7b <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV7b)
write.csv(FamilyEBV,'Pro7b.csv')

```

```

EBV <- read.table("new9b.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

```

```

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

```

```

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV9b <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV9b)
write.csv(FamilyEBV,'Pro9b.csv')

```

```

EBV <- read.table("new10.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

```

```

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

```

```

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}

```

```

FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV10 <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV10)
write.csv(FamilyEBV,'Pro10.csv')

EBV <- read.table("new10b.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV10b <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV10b)
write.csv(FamilyEBV,'Pro10b.csv')

Correlation <- cbind(MeanEBV4,MeanEBV7b,MeanEBV9b,MeanEBV10,MeanEBV10b)
Cor <- cor(Correlation, use='complete.obs',method = 'pearson')
write.csv(Cor,'Correaltion.csv')

```

## Appendix 3 Pearson correlation within models calculation (Cross Validation)

### Appendix 3.1 First split the data into two part evenly in **R Studio**

```
BinTH <- read.csv("BTH1.csv")
library(pls)
CVs <- cvsegments(1870,2,type="random")
CV1 <- as.vector(CVs[[1]])
CV2 <- as.vector(CVs[[2]])
as.factor(BinTH$sire)
as.factor(BinTH$dam)
BinTh1 <- BinTH[CV1,]
BinTh2 <- BinTH[CV2,]
write.csv(BinTh1,"CV1_BTH.csv")
write.csv(BinTh2,"CV2_BTH.csv")
# same process for SS.csv data
```

Appendix 3.2 Rerun the separated data one by one correspond to the same model in AsReml by just Adding ‘! CONTINUE !BLUP’ after the data line. (e.g.in BTH\_Sire dam model)

```
Sire & Dam binary threshold model
individual !A !P
sire !A !P
dam !A !P
weight
tankno
batch 8
challenge
BTH
challcode !!

#pedigree
pedigree.csv !SKIP 1 !ALPHA !MAKE

#data
BTH.csv !SKIP 1 !SAVE !CONTINUE !BLUP
```

```

BTH !BIN !PROBIT ~ tankno.batch weight challcode lr -individual sire and(dam,1) fac(sire,dam)
Residual units
Predict challcode
VPREDICT !DEFINE
F VarA sire * 4
F VarC fac(sire,dam)
F VarP VarC + VarA*0.5 + Residual
H h2 VarA VarP
H C2 VarC VarP

```

Appendix 3.3 Mid-parent full-sib family EBVs and Pearson correlation coefficients between the predicted family EBV's. Calculated in R studio (Eg. for BTH\_Sire dam model)

```

EBV <- read.table("new4 set1.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV1 <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV1)
write.csv(FamilyEBV,'Pro4 set1.csv')

EBV <- read.table("new4 set2.sln",header = T)
PED <- read.csv('Pedigree_ind.csv')
EBVsire <- rep(0,1870)
EBVdam <- rep(0,1870)

```

```

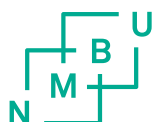
EBVlevel <- as.character(EBV$Level)
PEDsire <- as.character(PED$sire)
PEDdam <- as.character(PED$dam)
for (i in 1:1870){
  idx <- which(EBVlevel==PEDsire[i])
  EBVsire[i] <- EBV$Effect[idx]
}

for (i in 1:1870){
  idx <- which(EBVlevel==PEDdam[i])
  EBVdam[i] <- EBV$Effect[idx]
}
FamilyEBV <- cbind(PED,EBVsire,EBVdam)
MeanEBV2 <- (EBVdam+EBVsire)/2
FamilyEBV <- cbind(FamilyEBV,MeanEBV2)
write.csv(FamilyEBV,'Pro4 set2.csv')

Correlation <- cbind(MeanEBV1,MeanEBV2)

COR <- cor(Correlation, use='complete.obs',method = 'pearson')

```



Norges miljø- og biovitenskapelig universitet  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway