Aquaculture 310 (2010) 66–73

Contents lists available at ScienceDirect

Aquaculture

journal homepage: www.elsevier.com/locate/aqua-online

Bias and precision of estimates of genotype-by-environment interaction: A simulation study

Panya Sae-Lim *, Hans Komen, Antti Kause

Animal Breeding and Genomics Centre, Wageningen University, P.O. Box 338, 6700 AH, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 17 May 2010
Received in revised form 7 September 2010
Accepted 13 October 2010

Keywords:
Breeding programme
Genetic correlation
Genotype by environment interaction
Optimal design
Population structure
Simulation

ABSTRACT

Re-ranking of genotypes across environments is a form of genotype-by-environment (G×E) interaction with serious consequences for breeding programmes. The degree of such G×E interaction can be estimated using the genetic correlation (r_g) between measurements in two environments for a given trait. When r_g is lower than 0.8, G×E interaction is commonly considered to be biologically significant. Here a stochastic simulation was used to study the impact of population structure on bias and precision of genetic correlation estimates between two environments. Simulated populations resulted from a nested mating design (1 sire to 2 dams). Simulated r_g was 0.0, 0.5, or 0.8. A trait with heritability (h^2) of either 0.3 or 0.1 in both environments was simulated. Simulation results show that genetic correlation estimates are biased downward especially when the simulated r_g is 0.8, heritability is 0.1, and family size is less than 10. A downward biased genetic correlation estimate incorrectly suggests the existence of G×E interaction. This can lead to the erroneous conclusion that a multi-environment breeding programme is needed. The optimal design with the lowest mean square error for r_g for a trait with low h^2 requires a large family size (20–25) and a low number of families (100–80 or 50–40 for population size fixed to 2000 and 1000 animals, respectively). For traits with moderate h^2, the optimal family size is 10 with 200 or 100 families for population size fixed to 2000 and 1000, respectively. We also studied the effect of selective mortality on G×E estimates. However, schemes with unequal family sizes due to differences between families in survival produced similar results for the optimum design as schemes with equal family sizes. Equal-family-size design can thus be used to determine the optimal design for estimating G×E interaction. Our study can be used as a guideline for estimating a genetic correlation for practical breeding programmes.

1. Introduction

Many breeding programmes distribute animal material across diverse production environments, sometimes even at a global scale. Selection within a nucleus broodstock may lead to lower-than-expected genetic gains in other production environments when genotype-by-environment (G×E) interaction exists but it is not introduced in the selection criteria.

G×E interaction is defined as a phenomenon that genotypes respond differently to an environment gradient (Falconer and Mackay, 1996). There are two main types of G×E interaction: scaling effects and re-ranking. A scaling effect means that the amount of genetic variation in two environments differs. Re-ranking means that ranking of genotypes changes across different environments (Lynch and Walsh, 1998). Re-ranking in particular is a challenge for breeding because genotypes in one environment are not necessarily the best ones in other environments. Re-ranking across environments can be estimated using a genetic correlation between measurements in two environments for a given trait (Falconer, 1952). G×E interaction is commonly considered to be biologically significant when genetic correlation is lower than 0.8 (Robertson, 1959b).

In aquaculture, a number of studies on G×E interaction have been conducted under diverse management practices. The published studies on genetic correlations between environments have used family sizes and family numbers ranging from tens to several hundred (e.g. Sylvén et al. 1991; Fishback et al., 2002; Kause et al., 2003, 2004; Saillant et al., 2006; Quinton et al., 2007; Dupont-Nivet et al., 2008; Pierce et al., 2008; Vehviläinen et al., 2008; Khaw et al., 2009).

To accurately estimate a genetic correlation between environments, an optimal design needs to be established; an experimental design which produces a precise and unbiased result while using minimum testing capacity. Enlarging population size typically increases the power of a design but simultaneously increases costs. In contrast, too small population size or suboptimal population structure (number of families, family size, and mating design) may potentially result in biased and inaccurate estimates. Furthermore, differences in family size caused by differential survival or differences in parental contributions to the whole population size will result in unequal family sizes. The resulting

* Corresponding author. Tel.: +31 317482207; fax: +31 317483929.
E-mail address: panya.sae@wur.nl (P. Sae-Lim).

0044-8486/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.aquaculture.2010.10.020

population structure is unbalanced which may influence the bias and precision of genetic correlation estimates.

To our knowledge, no study has been conducted to assess bias and precision of estimates of G×E interaction. The present study describes the use of a stochastic simulation to construct an optimal population structure promoting precise and unbiased estimation of a genetic correlation between environments. The simulations employed here are divided into three scenarios. Firstly, various population sizes were simulated. Secondly, varying combinations of family size and family number were used to find an optimal population structure under a fixed population size. Thirdly, in practice, an experimental design is unintentionally challenged with between-family variation in survival leading to an unbalanced design. Unequal family sizes may result in larger sampling variance compared to equal family sizes (Hammersley, 1949; Tallis, 1959). Therefore, this scenario was used to study the influence of unequal family sizes on the bias and precision of the estimation.

2. Materials and methods

In the simulation, three different population structures were constructed, and (co)variance components were estimated.

2.1. Population construction

The simulated population structure was a split-family design with two environments, where the offspring generation had trait records and their parents only contributed to the pedigree. In each environment, phenotype of an individual was calculated as $y = 0.5a_s + 0.5a_d + m + e$, where a_s and a_d are additive genetic values of sire and dam, respectively, m is Mendelian sampling term, and e is environmental effect. Additive genetic values were sampled from a bivariate normal distribution of environments A and B: $Var(a) = \begin{bmatrix} \sigma^2_{a,A} & \sigma_{a,AB} \\ \sigma_{a,AB} & \sigma^2_{a,B} \end{bmatrix}$, Mendelian sampling terms from $Var(m) = \begin{bmatrix} 1/2\sigma^2_{a,A} & 1/2\sigma_{a,AB} \\ 1/2\sigma_{a,AB} & 1/2\sigma^2_{a,B} \end{bmatrix}$, and environmental effects from $Var(e) = \begin{bmatrix} \sigma^2_{e,A} & 0 \\ 0 & \sigma^2_{e,B} \end{bmatrix}$. Each of these effects had a mean of zero. Phenotypic variance (σ^2_P) was set to 1. Additive genetic variance (σ^2_a) was calculated as $\sigma^2_P h^2$ and environmental variance (σ^2_e) was calculated as $\sigma^2_P(1-h^2)$. Genetic covariance between measurements of a trait in two environments ($\sigma_{a,AB}$) determined the degree of family re-ranking, and was sampled from a simulated value (described below). No environmental covariance was simulated between the two environments because each animal inhabited only one environment. The population construction was done in *R* (R Development Core Team, 2008).

2.2. Simulated scenarios

A population was simulated with a genetic correlation (r_g) of 0.8, 0.5, or 0.0 between environments. A value of 0.8 is often considered a threshold value for G×E interaction to be significant for a breeding programme (Robertson, 1959b), whereas genetic correlations of 0.5 and 0.0 mean that strong re-ranking occurs. A trait with heritability of either 0.3 or 0.1 in both environments was used in all scenarios. Number of sires, dams, and offspring were constructed following three population design scenarios. For all scenarios, the mating design was one sire mated to two different dams (paternal nested design). The paternal nested mating designs are used, e.g. in GIFT and Troutlodge breeding programmes.

2.2.1. Varied population size (scenario A)

Family size is one important factor that determines the amount of bias, standard error and mean square error. Therefore, this scenario was to evaluate the impact of family size on precisions and bias. The simulated population had a fixed family number of 100 but family size ranged from 3 to 75 within each environment. Note that with the increase of family size, also the population size increases, e.g. family size of 3×100 families = 300. The range of family sizes is given in Table 1.

2.2.2. Fixed population size (scenario B)

An experiment typically has a limit for the maximum number of fish reared, tagged or genotyped. The results from scenario A showed that estimates of r_g were unbiased for traits with both low (0.1) and moderate (0.3) heritabilities when population size was larger than 2000 (100 families×20 individuals). Therefore, the starting point for this simulation was a fixed population size of 2000 in both environments. In this scenario, both family size and family number were varied. Given a fixed population size, this means that increasing family size results in decreasing family number. The results from population size of 2000 were compared to the bias and precision of the r_g estimates when simulating a fixed population size of 1000, i.e. when the number of animals was decreased to 50%. Table 1 summarizes the used family sizes and number of families for population sizes of 2000 and 1000.

2.2.3. Unequal family size (scenario C)

In this scenario, the effect of unequal family size on the bias and precision of the estimate of r_g was studied. The initial population size was 2000 and survival was 50%, meaning that the population size at harvest trait recording was reduced to 1000. To generate differences between families in size, each individual was assigned a trait record for survival (0 = alive, 1 = died). Survival was not correlated with the traits recorded in two environments, and was not analysed as a correlated trait in the genetic analyses. Survival was modelled as a binary threshold trait with the following underlying liability scale phenotypic and genetic parameters. Phenotypic variance for survival was assumed to be one, and thus additive genetic variance is equal to heritability for survival (h^2_{surv}). To generate different degrees of between-family variation in survival for the population construction, three alternative sets of parameters were used for survival: $h^2_{surv} = 0.00$ and $c^2_{surv} = 0.3$; $h^2_{surv} = 0.15$ and $c^2_{surv} = 0.1$; $h^2_{surv} = 0.30$ and $c^2_{surv} = 0.0$, where c^2_{surv} is the ratio of variance for common environment of full-sibs to phenotypic variance. These represent realistic estimates for rainbow trout (Kanis et al., 1976; Vehviläinen et al., 2008, 2010).

Scenario C was performed for a trait with h^2 of 0.1 and 0.3 with a simulated genetic correlation of 0.8 between two environments. Results from scenario A show that this is the most difficult scenario to estimate genetic correlation correctly.

2.3. Estimation of (co)variance components

The simulated data were analysed using a bivariate animal model in which the same trait in two environments was treated as two different traits. The model fitted was:

$$y_{ij} = \mu_i + a_{ij} + e_{ij}$$

where y_{ij} represents a trait measured in one of two environments ($i = 1, 2$) for an individual j (j = number of individuals); μ_i is the overall mean of the trait i; a_{ij} is the random additive genetic effect of individual j; and e_{ij} is the random residual effect. Due to only one observation for each individual, residual covariance was fixed to zero. Estimated genetic correlation between two environments ($\hat{r}_g$), its standard error, and heritabilities with their standard errors were estimated using restricted maximum likelihood (REML) in ASReml software (Gilmour et al., 2006). The (co)variance matrix was constrained to be positive definite.

2.4. Summarising output from the simulation

Each population structure alternative was simulated 500 times. For each alternative, the results were summarized using: (*i*) average,