Electronic supplementary material to

Genetic growth potential, rather than phenotypic size, predicts migration phenotype in Atlantic salmon (doi: 10.1098/rspb.2020-0867)

Paul V. Debes[1,3], Nikolai Piavchenko[1], Jaakko Erkinaro[4], and Craig R. Primmer[1,2]

[1]Organismal & Evolutionary Biology Research Program, and [2]Institute of Biotechnology, University of Helsinki, Helsinki, 00014, Finland

[3]Department of Aquaculture and Fish Biology, Hólar University College, Sauðárkrókur, 551, Iceland

[4]Natural Resources Institute Finland (Luke), Oulu, 90014, Finland

Author for correspondence: Paul V. Debes; email: debes@holar.is

Contents:

Table S1. Model mean posterior estimates (on the liability scale), lower and upper 95% credible intervals, and number of effective samples (N_{eff}) for the reduced univariate generalized animal model on migrant phenotype binaries of 663 Atlantic salmon individuals from 32 half-sib families, including the effects of a locus with major effects on age at sexual maturation (*vgll3*, alleles: E = early, L = late maturation). Results for the binary response of migrant (1) or resident (0) are based on models not controlling for phenotypic length (model 1) or controlling for phenotypic length (model 2; length = length in late summer; mean centred and variance scaled).

Term	mean	lower	upper	N_{eff}
model 1				
model intercept	-0.338	-1.271	0.551	14516
temperature (cold - warm)	1.701	1.025	2.439	14564
Vgll3 (EE - EL)	-0.139	-0.780	0.512	14000
Vgll3 (EE - LL)	-0.276	-1.328	0.806	14000
model 2				
model intercept	1.287	0.616	1.913	7000
length (continuous)	2.946	2.435	3.507	7000
temperature (cold - warm)	-0.984	-1.660	-0.305	7000
temperature:length	-1.192	-1.864	-0.532	6317
Vgll3 (EE - EL)	0.164	-0.299	0.642	7000
Vgll3 (EE - LL)	0.213	-0.489	0.883	7000

Description of the general linear mixed to test for feed restriction treatment effects

To test for effect of the feed treatment (see methods for details) on growth rate, we fitted a general linear animal model under restricted residual maximum likelihood (REML) using asreml-R [1], to the (log) of length before and after the temporary feed restriction treatment. To estimate specific growth rate (i.e., proportional increase of length per unit time), we included date of measurement as a continuous covariate (date.integer; first measurement date taken as 0, days have a value of 1). We allowed the intercept (at the first measurement) and the growth rates to vary by temperature (cold, warm), feed restriction (full, restricted), and by migrant phenotype (resident, migrant). We also fitted all interactions among these terms. To account for the randomisation of treatments to tanks, we fitted random regression effects for tanks (i.e., a 2x2 covariance matrix for tank intercepts and date slopes with covariance between them). We accounted for the non-independence of, and among, individual data by including animal effects with additive genetic variance estimated via the inverse of the relationship matrix [2]. Using likelihood ratio tests between nested models, we determined that variances for animal effects (X_1^2 = 17.79, P < 0.001) and residuals (X_1^2 = 17.79, P < 0.001), but not tanks effects (X_3^2 = 1.97, P = 0.580), differed between temperature environments. We thus fitted these former two effect terms conditional on temperature environment, and with between-temperature covariance for the genetic effects. We found that the feed restriction had affected the specific growth rate of both residents and migrants and in both the cold and the warm environments (**table S2, figure S1**).

Table S2. ANOVA table for model terms of a univariate general animal model on body length phenotype in late summer of 663 Atlantic salmon individuals from 32 half-sib families. Intercept effects were estimated at the first measurement date (date.integer = 0). Statistical significance has been estimated based on *F*-tests with denominator degrees of freedom, DDF, approximated according to [3].

term	DF	DDF	*F*	*P*
model intercept	1	51.6	6162.0	< 0.001
migrant phenotype (resident, migrant)	1	518.8	448.4	< 0.001
temperature (cold - warm)	1	22.7	146.1	< 0.001
feed restriction (full - restricted)	1	11.8	7.6	0.018
date.integer (continuous)	1	11.3	904.1	< 0.001
migrant phenotype:temperature	1	206.0	8.8	0.003
migrant phenotype:feed restriction	1	550.9	0	0.968
temperature:feed restriction	1	14.7	0.1	0.782
migrant phenotype:date.integer	1	1302.0	430.9	< 0.001
temperature:date.integer	1	1302.0	9.7	0.002
feed restriction:date.integer	1	1302.0	122.0	< 0.001
migrant phenotype:temperature:feed restriction	1	407.8	0.3	0.607
migrant phenotype:temperature:date.integer	1	1302.0	8.1	0.004
migrant phenotype:feed restriction:date.integer	1	1302.0	38.3	< 0.001
temperature:feed restriction:date.integer	1	12.0	0.1	0.757
migrant phenotype:temperature:feed restriction:date.integer	1	409.3	10.8	0.001

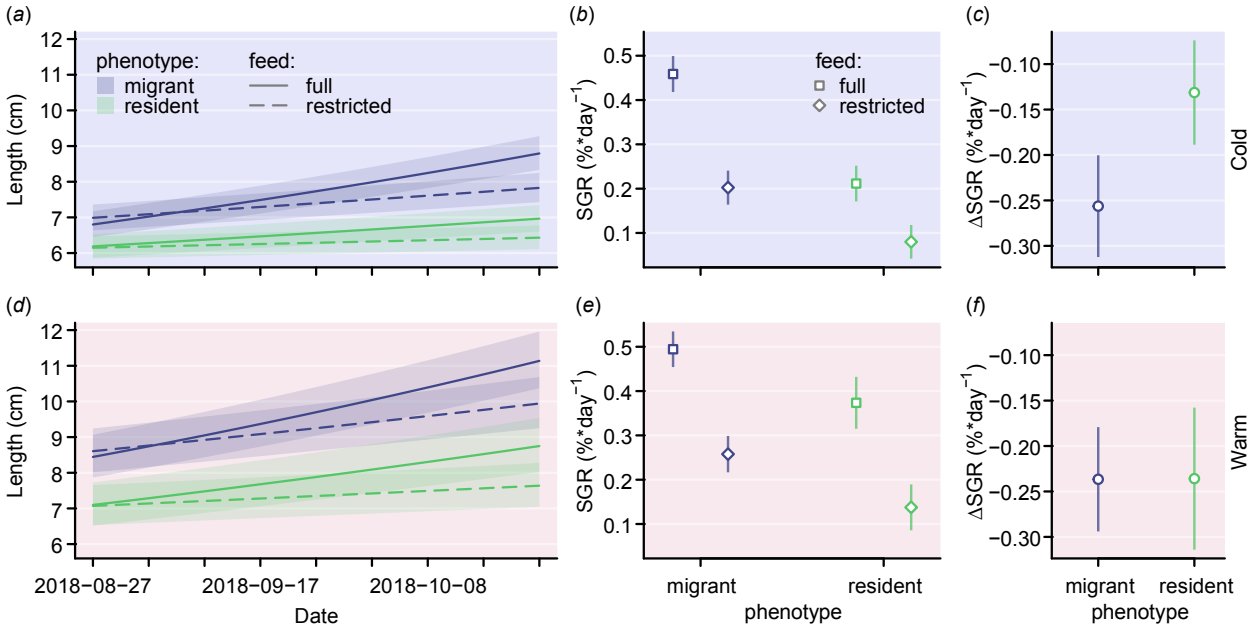

Figure S1. Model predicted average size trajectories (*a*, *d*) and associated specific growth rates (SGR; *b*, *e*) for prospective migrants or residents that were either fully fed or temporarily restrictedly fed in either a cold (upper row) or a 2°C warmer (lower row) environment, and the corresponding contrasts for SGR between the feeding treatments (ΔSGR; *c*, *f*). Estimates refer to the model as in **table S2**.

Description of the generalised linear mixed model to estimate male maturation rate

To estimate male maturation rate, we fit a generalised linear animal model with probit-link function to maturation binaries recorded during spawning time. We only fitted one overall mean effect (intercept). To account for the randomisation to tanks, we fitted random effects for tanks and to account for the non-independence among individual data, we including animal effects with additive genetic variance estimated via the inverse of the relationship matrix [2]. We used priors following [4] and methods as described for univariate Bayesian models in the main manuscript. We found that among tank effect variance was absent or negligible but that additive genetic variance was present (although methodologically inflated; **table S2**) and predicted a marginal overall maturation rate of males in the warm environment of 0.19 (95% CI: 0.08-0.33; **table S2**).

Table S3. Model mean posterior estimates (on the liability scale), lower and upper 95% credible intervals, and number of effective samples (N_{eff}) for the univariate generalized animal model on the sexual maturation phenotype binaries (0 = immature, 1 = mature) during the first year of 114 Atlantic salmon individuals from 30 half-sib families. Estimates were obtained for only males (no female matured) and only in the warm environment (no maturation occurred in the cold environment). Residual variance was fixed to 1, resulting in scaling of all components relative to the fixed residual variance. Variances different from zero (approximated by credible interval not including zero) are in bold.

Term	mean	lower	upper	N_{eff}
mean terms				
model intercept	-1.647	-2.615	-0.758	10000
variance terms				
Tank (V_C)	0.417	0.000	1.449	10210
Animal (V_A)*	**2.127**	**0.600**	**3.828**	10000

*Estimate is inflated because it is based on the coefficients of relatedness among individuals, but genotypes for the locus with major effects on maturation rate (*vgll3*) did not vary within families (leading to an absence of the expecting Mendelian sampling variance within families) due to the breeding design of using only *vgll3* homozygous parents as reported in [5].

Prior sensitivity analysis

We also assessed how prior specifications influenced the results. We focussed on the binary trait (MIG) because Bayesian heritability estimates for LEN closely matched estimates by residual maximum likelihood (REML), although the genetic correlation between environments was higher under REML (**figure S2**), whereby REML is the recommend method for animal models of continuous responses [4]. We compared the bivariate results (responses: MIG, LEN; including GxE) with univariate results (MIG per environment) obtained with recommended priors for binary animal models (following a X_1^2 distribution) [4]. We also separately assessed the effects, relative to the univariate models, of either including GxE for MIG or also fitting LEN per temperature environment. For the bivariate model, we varied prior specifications that resulted in different prior distributions for the variances and covariances. For the variances, we varied the parameter expansion variance scale by up to two magnitudes higher and lower than used (100, 10, 1, 0.1, 0.01), resulting in increasing prior densities for the proportions of the phenotypic variances towards one and zero, respectively. For the covariances, and thus correlations, we specified prior distributions that were either relatively flat for correlations (in *MCMCglmm*: nu = dimension of C, G, or R + 1) or showed higher densities towards -1 and 1 (in *MCMCglmm*: nu = dimension of C, G, or R). We found all investigated modelling variants to affect model estimates, but none strong enough to compromise our major inferences (**figures S2-S4**).

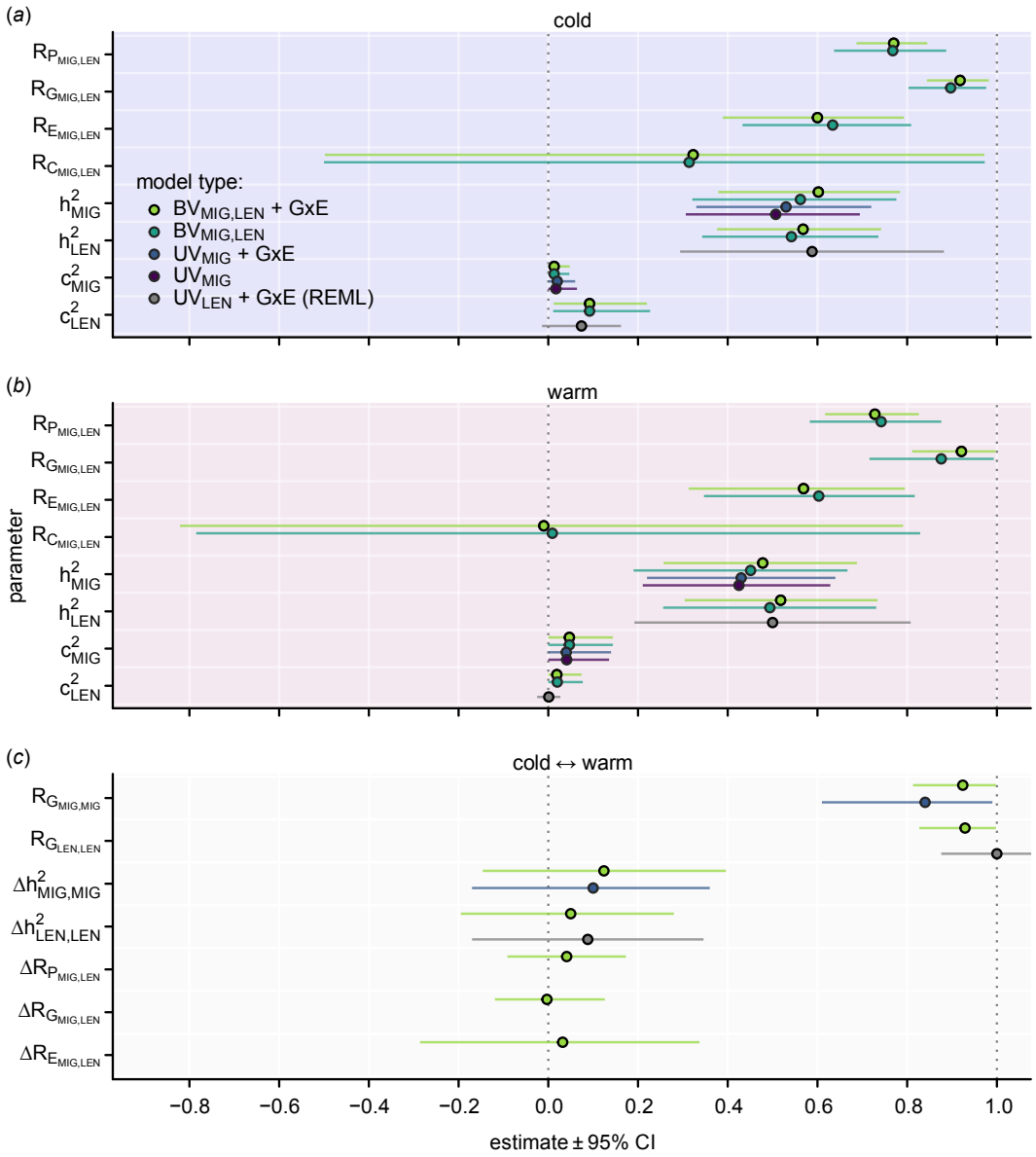

Figure S2. Comparison of the Bayesian model estimates for (co)variance components with increasing model complexity for the binary (MIG) and with REML estimates for the continuous trait (LEN). Heritability estimates for MIG increased in the following order: univariate single environment (UV) < univariate two environments (UV + GxE) < bivariate single environment (BV) < bivariate two environments (BV + GxE). Unfortunately, it is not clear whether model estimates in the absence of an explicitly modelled between-environment correlation (GxE) for the additive genetic effects should equal estimates in the absence of such modelled GxE. Furthermore, it is not clear whether model estimates for a single trait should equal those for estimates when a correlated second traits is modelled. Error bars show 95% credible intervals or approximate confidence intervals based on the delta method for REML estimates.

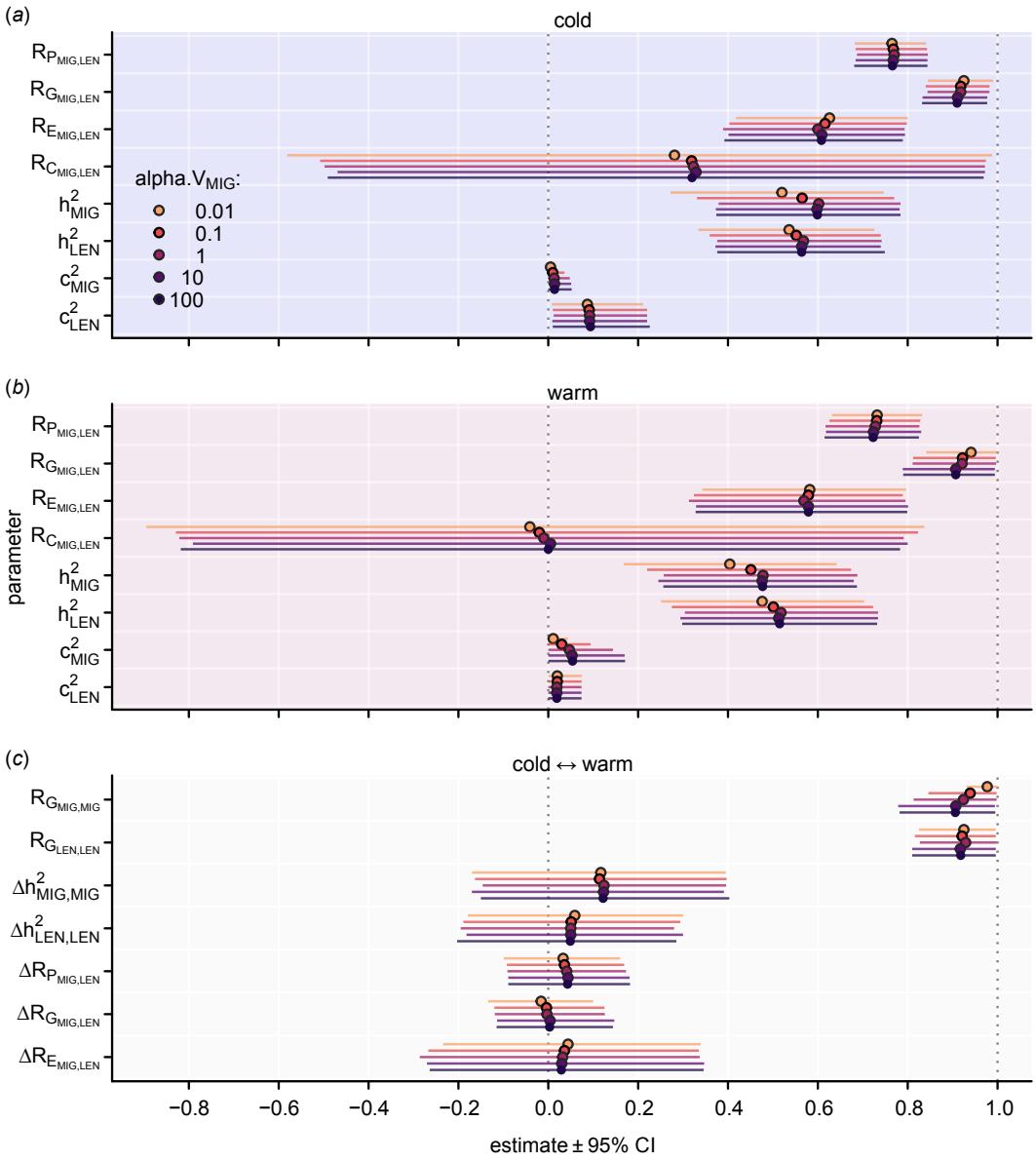

Figure S3 Comparison of the Bayesian model estimates for (co)variance components under different prior specifications for the parameter expansion variance scale (in *MCMCglmm* specified as "alpha.V") for the genetic and common environmental covariance matrices of the binary trait (MIG). For MIG heritability (h^2) and phenotypic proportion of the common environmental variance (c^2), the priors specify an increasing density towards 0 (alpha.V specified as either 100, 10, 1, 0.1, or 0.01). The parameter expansion variance scale specification for the continuous trait (LEN) was kept constant (alpha.V = 1). It could be noted that a smaller variance scale resulted in smaller heritability (h^2) and proportion of common environmental variance (c^2) of MIG when alpha.V fell < 1. Similar effects on h^2, but not c^2, were also observed for LEN.

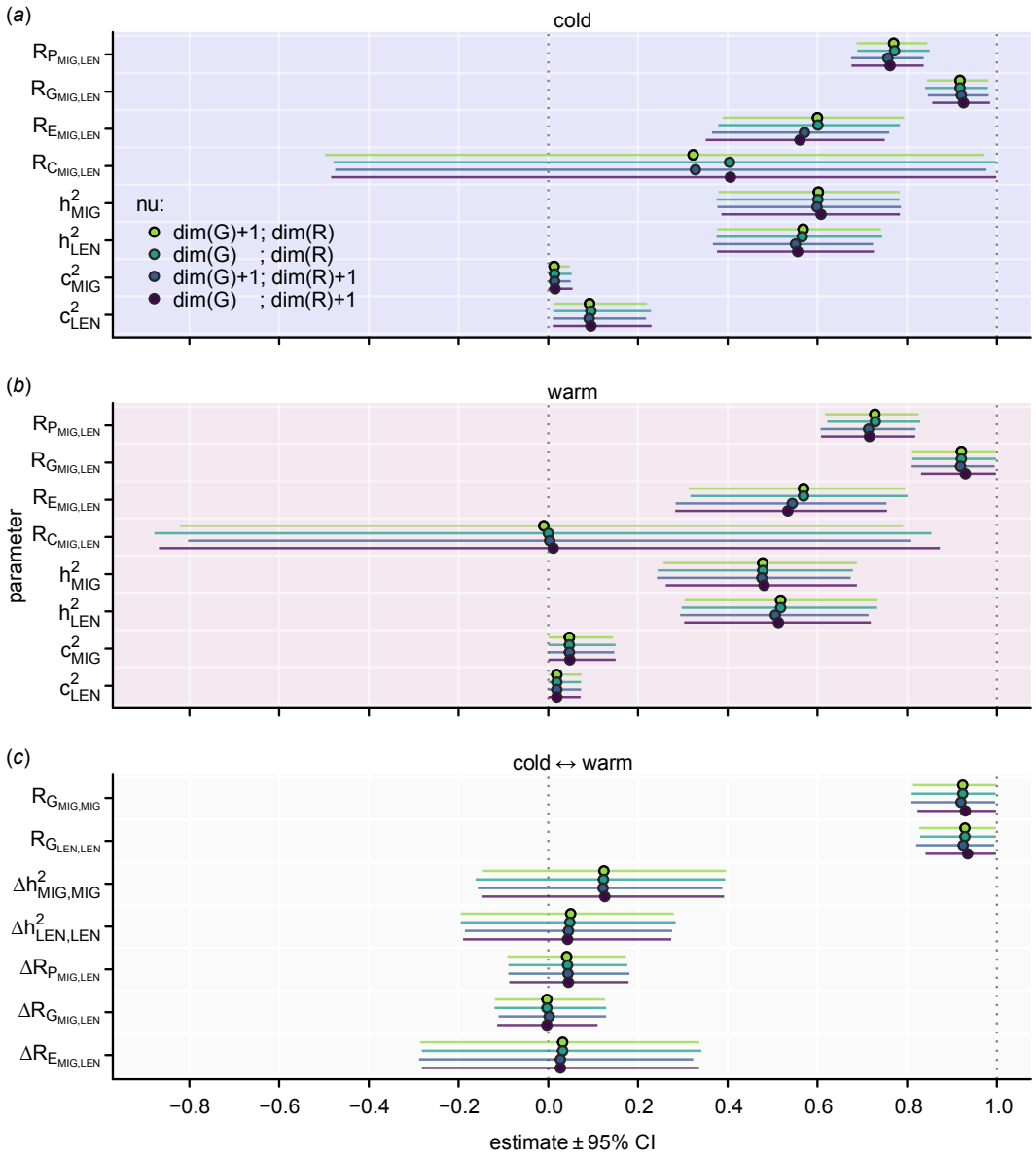

Figure S4. Comparison of the MCMC bivariate model estimates for (co)variance components varying the prior specification for the degree of believe parameter (nu) for combinations of the genetic (G, also applied for the environmental C) and residual (R) covariance matrices. The different priors result in either relatively flat distributions for the correlations (in *MCMCglmm*: nu = dimension [dim] of G or R + 1) or distributions with higher densities towards 0 and 1 (in *MCMCglmm*: nu = dimension [dim] of G or R). We noted an effect on the between-trait correlation for the common environmental effects within the cold environment ($R_{C_{MIG,LEN}}$; ***a***), which is unsurprising because common environmental effects were inferred as absent for one trait (c^2_{MIG}) but present for the other (c^2_{LEN}), thus lacking data information on correlations. As a result, we

refrained from making inferences about $R_{C_{MIG,LEN}}$, and effects of $R_{C_{MIG,LEN}}$ on $R_{P_{MIG,LEN}}$ were marginal because of the low or absent c^2. A second effect could be noted on the between-trait covariance for the residual environmental effects within both temperature environments ($R_{E_{MIG,LEN}}$; ***a***, ***b***). Given the prior knowledge of a suspected causal relationship between the two traits and for each trait between environments, but uncertainty whether this correlation is stronger at the environmental or genetic level, it is unclear which prior is more appropriate (i.e., whether assuming a between-trait correlation closer to zero or unity). However, this difference affects inferences only marginally and genetic correlation estimates appeared unaffected.

References

1. Butler D.G., Cullis B.R., Gilmour A.R., Gogel B.J. 2009 *Mixed models for S language environments ASReml-R reference manual*. Brisbane, Australia, Queensland Department of Primary Industries and Fisheries, NSW Department of Primary Industries; 144 p.
2. Henderson C.R. 1973 Sire evaluation and genetic trends. *J Anim Sci* **1973**(Symposium), 10-41. (doi:/1973.1973Symposium10x).
3. Kenward M.G., Roger J.H. 1997 Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**(3), 983-997. (doi:10.2307/2533558).
4. de Villemereuil P., Gimenez O., Doligez B., Freckleton R. 2013 Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits. *Methods Ecol Evol* **4**(3), 260-275. (doi:10.1111/2041-210x.12011).
5. Verta J.-P., Debes P.V., Piavchenko N., Ruokolainen A., Ovaskainen O., Moustakas-Verho J.E., Tillanen S., Parre N., Aykanat T., Erkinaro J., et al. 2019 Regulatory divergence in *vgll3* underlies variation in age at maturity in male Atlantic salmon. *bioRxiv*. (doi:10.1101/777300).