Random Regression Analysis Of Disease Tolerance Under Natural Disease Infection: A Simulation Study

A. Kause[*]

Introduction

Defence mechanism against pathogens and parasites can be divided into two: resistance and tolerance. Disease resistance is the host trait that prevents infection or reduces performance of the enemy on a host, both factors reducing pathogen burden within an individual. Disease tolerance is the ability of the host to limit the impact of a given pathogen burden on host performance (Simms (2000)). In animal science, tolerance is often termed 'resilience' (Bisset and Morris (1996)). A great deal is known about the genetics of resistance but there is a lack of work on disease tolerance (Råberg et al. (2007); Schneider and Ayres (2008)).

Tolerance needs to be defined as a change in host performance in response to increasing pathogen burden to ensure that there is a causal relation between the two (Simms (2000); figure 1). Thus, measuring tolerance follows the methodology of reaction norm analysis. Genetic variance in tolerance can be estimated as the genetic variance in regression slope of host performance along a gradient of increasing pathogen burden. Furthermore, a genetic cost of tolerance can be measured as a genetic correlation between tolerance slope and host performance in a pathogen-free environment (i.e., the intercept). A genetic trade-off occurs if a tolerant genotype has low fitness in a pathogen-free environment (Núñz-Farfán et al. (2007)). Contrary to reaction norm analysis, analysing tolerance as host performance change from before to after pathogen attack confounds both natural temporal variation in host performance (e.g. growth curves) and impact of pathogen burden on the host.

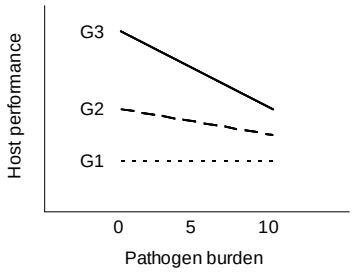

Figure 1: Reaction norms of host performance for three genotypes along a pathogen burden gradient. Flat slope means high tolerance (G1), and steep slope low tolerance (G3). Note that neither 'average host performance level' nor a trait calculated as 'performance divided by pathogen burden' reflects the level tolerance

Using a simulation, I assessed the potential of analysing tolerance using random regressions, when a host population is under a natural enemy attack. Many animal breeding data sets include data on natural exposure to diseases or parasites. Random regressions have not been applied for tolerance analysis before. Two challenges to be encountered when analyzing data were considered. Firstly, the genetic analysis of tolerance uses in particular within-family information (fullsibs and halfsibs, and more distant relatives) to estimate the tolerance slope. Thus, family size is a crucial parameter influencing estimation accuracy. Secondly, under

[*] Wageningen University and Research Centre, Animal Breeding and Genomics Centre, The Netherlands

natural infection, pathogen burden (i.e. resistance) is typically non-randomly distributed across individuals. When pathogen burden and host performance before infection are initially correlated, biased estimates for tolerance variation can be potentially obtained. The degree of bias and precision in estimates of genetic (co)variance for slope and intercept were evaluated in alternative scenarios.

Material and methods

Simulation of data. Full-sib data were simulated. The parental population consisted of 200 unrelated animals, 100 males and 100 females, without phenotypes. Without a loss of generalization, genetic variance here refers to full-sib family variance. In offspring population, phenotypic values of pathogen burden ($Pburden_i$), tolerance slope (b_{1i}), and intercept (b_{0i}) of an individual i were simulated as the sum of genetic and environmental effects randomly sampled from a multitrait distribution. A genetic value of an individual was simulated as the average of parents breeding values plus a Mendelian sampling term (equal to half genetic variance). Genetic and environmental effects were randomly sampled from ~N(4,0.3) and ~N(4,0.7) for pathogen burden, from ~N(100,120) and ~N(100,280) for intercept, and from ~N(0,9) and ~N(0,21) for tolerance slope. Thus, simulated h^2 was 0.3 for all three traits. In the base scenario, covariances between pathogen burden, slope and intercept were simulated to be zero. A phenotypic value (y) of host performance of an individual i at its own pathogen burden phenotype level was calculated as: $y_i = b_{0i} + b_{1i} Pburden_i$.

The simulated values result in phenotypic coefficient of variation (CV_P) of 20% and heritability of 0.3 for host performance in a pathogen-free environment. This represents a normal growth trait. The simulated values result in CV_P of 30% for host performance under pathogen burden, and h^2 for host performance remains constant 0.3 along parasite burden gradient.

Alternative scenarios studied. Firstly, to assess the effect of sample size, family size was set to either 10, 30, 50, 100, or 200. Secondly, to assess the effect of non-zero environmental correlation (r_E) between intercept and pathogen burden, r_E was simulated to be either -0.5, -0.3, 0, 0.3, or 0.5, and family size was fixed to 100. The simulation was repeated 500 times for each alternative scenario.

Genetic analysis of simulated data. Tolerance is the slope of host performance when regressed against individuals' own pathogen burden phenotype (Simms (2000)). Thus, data on individual's pathogen burden and its performance value at that burden were used. Data points for intercept were not utilized and thus each individual had only one host performance observation. A full-sib random regression model using ASReml software was applied: $y_{ij} = \mu + b_{0j} + b_{1j} Pburden_{ij} + \varepsilon_{ij}$ [1], where y_{ij} is host performance of an individual i from family j at its parasite burden, μ is the population mean, b_{0j} is the random genetic effect of intercept for a family j, b_{1j} is the random genetic effect of regression slope for a family j, $Pburden_{ij}$ is pathogen burden of an individual i for family j, and ε_{ij} is the random error term. Heterogeneous error variance was modelled within six classes along the x-axis. When non-zero environment correlation between intercept and pathogen burden was simulated, an additional genetic model was used in which the model [1] was upgraded with a covariate of

phenotypic values of host performance in a pathogen-free environment. Estimated genetic variance was calculated as two times the full-sib family variance.

Results and discussion

Effect of family size on genetic variance estimates. Low family size resulted in increased, not decreased, estimates for tolerance genetic variance (table 1). For instance, with family size of 10, the estimated genetic variance for slope was 2.0 times higher than the simulated value. To obtain unbiased estimates for tolerance genetic variance, family sizes of 100 were needed (table 1). The intercept showed a pattern similar to the slope (table 1).

The tolerance slope is estimated within each family and thus family size is a crucial parameter influencing estimation accuracy. With decreasing family size, it is increasingly difficult to accurately estimate the true slope. When a small number of individuals is sampled for each family, the sample no longer is representative of the true distribution and single observations have strong impact on the slope estimate. For some families the slope is underestimated, for others overestimated, and thus genetic variance estimate for slope is artificially increased. This result is similar to one by Knap and Su (2008).

Table 1: Estimated genetic parameters (± s.d.) for a scenario with varied family size

		Family size				
Parameter estimated	Simulated value	10	30	50	100	200
		Estimated values				
Slope V_G	9.0	17.7±17.9	11.5±7.80	10.2±5.57	9.22±3.79	9.13±2.47
Intercept V_G	120	228±247	148±110	130±83.3	116±50.7	120±34.4
$r_{Slope\text{-}Intercept}$	0.0	-0.51±0.47	0.06±0.60	0.04±0.53	0.08±0.42	0.03±0.22

V_G: Genetic variance; *r*(Slope-Intercept): Genetic correlation between slope and intercept.

Table 2: Genetic parameters (± s.d.) for a scenario with varied environmental correlation (r_E) between slope and intercept, estimated with a statistical model either including or excluding host performance in a pathogen-free environment as a covariate

		r_E				
Parameter estimated	Simulated value	*-0.5*	*-0.3*	*0*	*0.3*	*0.5*
		Estimated values				
		Statistical model without the covariate				
Slope V_G	9.0	140±12.2	56.7±8.35	9.40±3.59	57.2±7.97	142±11.7
Intercept V_G	120	905±245	409±127	117±51.6	413±121	921±251
$r_{Slope\text{-}Intercept}$	0.0	-0.89±0.03	-0.76±0.07	0.07±0.41	-0.76±0.06	-0.89±0.03
		Statistical model with the covariate				
Slope V_G	9.0	9.35±1.93	9.32±1.66	9.48±1.78	9.33±1.82	9.44±1.82

V_G: Genetic variance; *r*(Slope-Intercept): Genetic correlation between slope and intercept.

Environmental correlation between pathogen burden and host performance. With non-zero environmental correlation between pathogen burden and intercept either the initially smallest (negative r_E) or biggest (positive r_E) of each family are more prone to diseases. When r_E was changed to be either negative or positive, genetic variance for tolerance slope

increased symmetrically (table 2). For example, when environment correlation was -0.5 (or 0.5), the estimated genetic variance for slope was 15.6 times higher than the simulated value. This occurs because pathogen burden is non-randomly distributed within a family, and part of initial variation in host performance is translated to tolerance variation. Yet, genetic variance of slope was estimated without bias when host performance in pathogen-free environment was included as a covariate in the statistical model (table 2). It is well established that individuals with initially different production or life-history trait levels are differently exposed to diseases, parasites and production diseases. Consequently, it is expected that individuals with initially different growth and life-history traits receive differential pathogen burden, confounding the cause-and-effect relation between pathogen burden and reduction in host performance. In field data, this issue can be reduced by including initial host performance as a covariate in the statistical model. This is convenient for traits that can be repeatedly recorded from the same animals.

Cost of tolerance. In the above scenarios, zero genetic correlation between intercept and slope was simulated, i.e. no genetic trade-off was assumed between the two. A decrease in family size lead to strongly negative genetic correlation between slope and intercept (table 1). When slope h^2 of 0.05 was simulated, negative genetic correlation was obtained with family sizes of 10-50. Similarly, with increasing or decreasing environmental correlation between pathogen burden and performance in pathogen-free environment, the genetic correlation between slope and intercept became highly negative (table 2). The negative correlation occurs because when the slope for a family is overestimated, the intercept for the family will go down (and vice versa). The results imply that the analysis may falsely indicate the presence of a genetic trade-off (i.e., cost of tolerance).

Conclusion

Random regressions provide powerful means for analysing disease tolerance genetics. Yet, analysing tolerance genetics from field data sets is a challenge. A care should be taken to assess potential for bias due to small sample size and correlations between initial host performance and pathogen burden. Separating resistance and tolerance is a challenge but possible using the approach presented here.

References

Bisset, S.A. and Morris, C.A. (1996). *Int. J. Par*., 26:857–868.

Knap, P.W. and Su, G. (2008). *Animal*, 2:1742–1747.

Núñz-Farfán, J., Fornoni, J., and Valverde, P.L. (2007). *Annu. Rev. Ecol. Evol. Syst.*, 38:541–566.

Råberg, L., Sim, D., and Read, A.F. (2007). *Science*, 318:812–814.

Schaeffer, L.R. (2004). *Livestock Prod. Sci.*, 86:35–45.

Schneider, D.S. and Ayres, J.S. (2008). *Nature Rev. Imm.*, 8:889–895.

Simms, E.L. (2000). *Evol. Ecol.*, 14:563–570.