A quantitative genetic theory for infectious diseases

P. Bijma*, A. D. Hulst and M. C. M. de Jong

Wageningen University & Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands
*piter.bijma@wur.nl

Abstract
We integrated quantitative genetics and epidemiology to develop a quantitative genetic theory of the prevalence of endemic infectious diseases. Results show that infectious diseases respond very differently to selection than common non-communicable traits, and strongly suggest that the genetic variance determining the potential response of prevalence to selection must be much larger than currently believed. Moreover, heritable variation and response to selection increase significantly when the prevalence of the infection decreases, ultimately leading to local extinction of the infection due to herd immunity. These results change our perspective on the prospects of genetic selection against infectious diseases.

Introduction
A reduction of the prevalence of infectious diseases in livestock is highly desirable. Individual disease status, however, measured as $y = 0$ or 1 indicating non-infected vs infected, typically shows low heritability (h^2). Moreover, the classical threshold model predicts that h^2 goes to zero when a disease becomes rare. Hence, genetic selection against infectious diseases seems difficult. This perspective, however, ignores that infectious diseases rely on transmission between individuals. Here we present a quantitative genetic theory for response to selection and heritable variation in the prevalence of endemic infectious diseases.

Methods & Results
The endemic prevalence (P) of an infectious disease is the average fraction of the population that is infected, and is equal to the mean of individual disease status (P = avg(y)). Standard epidemiological theory shows that the endemic prevalence follows from the basic reproduction number (R_0; Weiss and Dishon 1971; Figure 1),

$$P = 1 - 1/R_0 \quad (1)$$

R_0 is the number of individuals that gets infected by a single typical infected individual in a naive population. For example, with $R_0 = 3$, an infected individual would on average infect three others. However, if 2/3 of the population is infected already, then only a single new infection will occur on average. Hence, prevalence reaches an equilibrium at $P = 1 - 1/3 = 2/3$. To reduce prevalence, therefore, breeders should lower R_0. At low R_0, prevalence is very sensitive to changes in R_0 (Fig. 1), which has significant implications for response to selection.

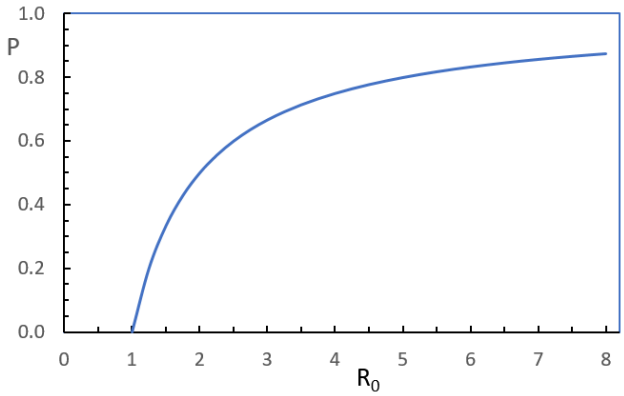

Figure 1. Prevalence as a function of R_0 (from Bijma et al. 2022).

Breeding values for R_0, prevalence and disease status. R_0 is the product of the contact rate between individuals, the *sus*ceptibility of individuals to become infected, the propensity of individuals to infect others (*inf*ectivity) and the duration of the infectious period (*t*). Because R_0 is a product and because it takes strictly positive values, we define an additive model with normally distributed genetic effects for the logarithm of R_0,

$$A_{ln(R0)} = A_{ln(sus)} + A_{ln(inf)} + A_{ln(t)} \quad (2)$$

We assume that additive genetic (co)variances are constant on this log scale, *i.e.*, are independent of the level. (This implies that additive genetic variance in R_0 decreases as R_0 decreases). Connecting Equations 1 and 2 leads to a breeding value for individual disease status (A_y) and a breeding value for prevalence (A_P; Bijma et al. 2022; here assuming no genetic variation in infectivity),

$$A_y = P(1-P)\, A_{ln(R0)} \quad (3a)$$

$$A_P = (1-P)\, A_{ln(R0)} \quad (3b)$$

A_y is the ordinary ("direct") breeding value for the disease status of the individual itself, while A_P is a total breeding value that measures the total effect of an individual's genes on the prevalence in the population. Hence, response of prevalence to selection equals the per generation change in mean A_P. Note that A_P is a factor 1/P greater than A_y. Moreover, for a constant $A_{ln(R0)}$, A_y has a maximum for P = 0.5, while A_P increases when P decreases. Thus, the common h^2 of individual disease status (y = 0,1) has a maximum at P = 0.5, similar to the threshold model (Robertson 1950), but the additive genetic variance that breeders can use for response to selection increases significantly as prevalence falls (Figure 2). This result is a consequence of the increasing slope of the relationship between P and R_0 at lower values of R_0 (Figure 1).

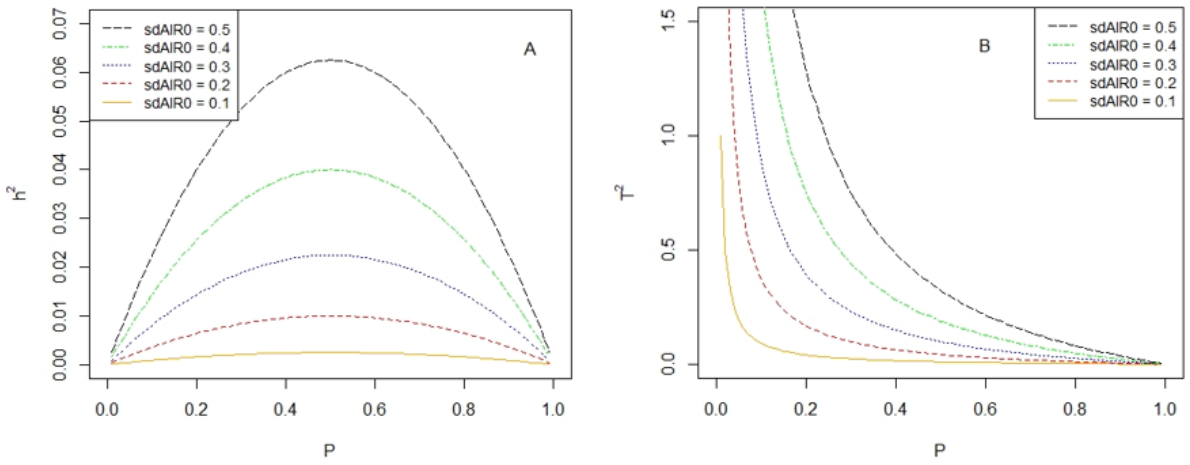

Figure 2. Heritability of individual disease status (h^2, panel A), and additive genetic variance in the endemic prevalence as a fraction of phenotypic variance (T^2, panel B), as a function of prevalence (P). For different additive genetic standard deviations in ln(R_0), and without genetic variation in infectivity (from Bijma et al. 2022).

Figure 2 shows that common heritabilities of disease status correspond to a large additive genetic variance in prevalence, particularly when prevalence is small. For example, for P = 0.2 and sd($A_{ln(R0)}$) = 0.4, the common observed-scale heritability of binary disease status is only 0.026, while the additive genetic variance that can be used for response to selection is 64% of phenotypic variance (T^2 = 0.64). Hence, for this scenario, the additive genetic standard deviation in prevalence is sqrt[0.64 x 0.2 x (1-0.2)] = 0.32, which is very large.

Response to selection. Figure 2B shows a strong increase in additive genetic variance as prevalence falls, suggesting that response to selection for lower prevalence should accelerate over generations. Figure 3 shows response to selection observed in simulations following standard methods in epidemiology, not using any of the above theory. Figure 3 also shows the mean breeding values for prevalence (A_P) and for individual disease status (A_y), expressed as deviations from prevalence in the first generation. As expected based on Figure 2, the response accelerates as prevalence decreases. In the final generations (t = 16, and t = 70) the infection disappears due to herd immunity in the local population, not because individuals are fully resistant to infection. The mean breeding value for prevalence (Eqn. 3b) closely matches the observed response, while the mean breeding value for individual disease (Eqn. 3a) deviates more and more from the observations as prevalence decreases. Hence, response in Figure 3 is significantly greater than the common breeding values for individual disease status suggest.

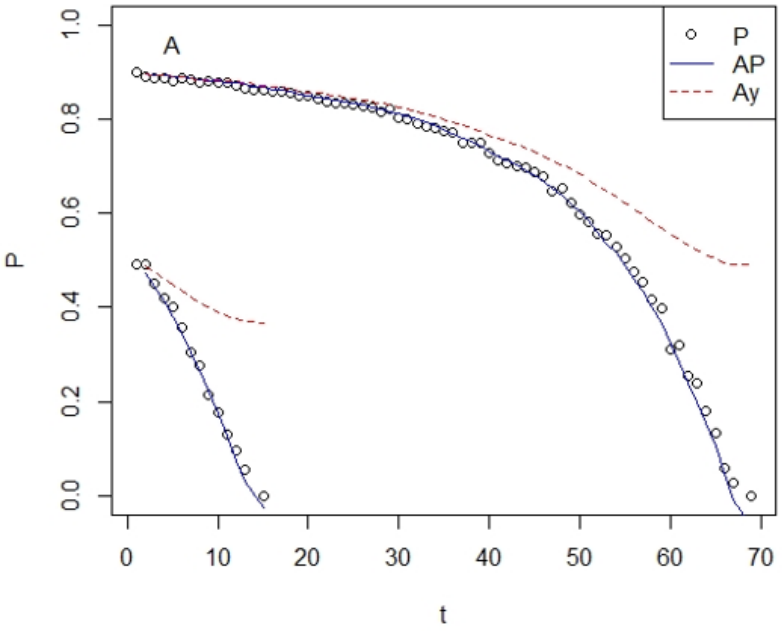

Figure 3. Response to mass selection against an infectious disease; observed response (small circles), mean breeding value for prevalence in (A_P, solid blue line) and mean breeding value for individual diseases status (A_y, dashed red line). For two populations, one starting at a prevalence of 50%, and the other at 90%. With genetic variation in susceptibility only, with $var(A_{ln(R0)}) = var(A_{ln(sus)}) = 0.3^2$, and a selected proportion of 0.5. The common observed-scale heritability for individual disease status in any generation can be read from Figure 2A, using an x-axis value corresponding to the prevalence in that generation, and does not exceed 0.022 (from Bijma et al. 2022).

Discussion

We developed a quantitative genetic theory of infectious diseases, focussing on R_0 and on the endemic prevalence of the infection. The breeding value for the logarithm of R_0 is central to the theory, and is the sum of the breeding values for the logarithms of susceptibility, infectivity and duration of the infectious period. From the breeding value for $ln(R_0)$ we derived a breeding value for individual binary disease status and a breeding value for the prevalence of the infection in the population. The first relates to ordinary h^2 of individual disease status, while the second relates to response to selection. Without genetic variation in

infectivity, the breeding value for prevalence is a factor $1/P$ greater than the breeding value for individual disease status. This is a general result, and does not rely on the assumption of an additive model on the log scale. For this reason, both the additive genetic variance for prevalence and the response of prevalence to selection increase when prevalence goes down (Figure 2B, 3).

The large difference between the h^2 of individual disease status and the total additive genetic variance in prevalence as a fraction of phenotypic variance (T^2) shown in Figure 2 indicates the presence of considerable indirect genetic effects (IGE). The h^2 reflects the variance of only the direct genetic effect of individuals on their own disease status. The T^2 reflects the variance of the full effect, *i.e.*, of the sum of the effects on the disease status of the individual itself and on the disease status of its herd mates. Diseases that rely on transmission generate considerable IGE because an individual who does not become infected itself also cannot infect other individuals (simply because it is not infected). Hence, the direct effects for susceptibility and for duration of the infectious period are fully correlated to the corresponding indirect effects, resulting in positive feedback and an increase in the total additive genetic variance. This positive feedback effect increases strongly when prevalence decreases, which causes IGE to make up an increasing proportion of the total effect, and explains the difference between h^2 and T^2 in Figure 2 A vs B.

The results shown here apply to endemic microparasitic infections, and where the pathogen relies on the host individual (*i.e.*, the livestock) for its replication. Hence, we assume absence of an external reservoir of pathogens, such as a second host species (for example badgers in the case of bovine Tb). Note that temporary survival of the pathogen in the environment does not violate this assumption. This assumption implies that a reduction in, *e.g.*, susceptibility translates fully into reduced exposure in the local population (because of fewer infected individuals). Example traits include mastitis, infectious claw disorders, respiratory infections in young animals (young replacement stock, meat calves, fattening pigs), faecal-oral transmitted infections causing gastro-intestinal diseases, and several zoonosis.

In conclusion, our results show that the quantitative genetics of infectious diseases is very different from the quantitative genetic theory for ordinary, non-communicable, traits, and that prospects for genetic improvement are much better than currently believed, as was also proposed by Bishop and Stear (1997). These findings, coupled with the high importance of infectious diseases for animal health and welfare, and also for human health in the case of zoonoses, suggest that breeders should become familiar with basic concepts in epidemiology, such as R_0. Together with recent advances in sensor technology and AI, which will enable us to collect large-scale high-quality data on individual disease status, these findings facilitate artificial selection against infectious diseases.

References

Bijma, P., Hulst, A. D., and de Jong, M. C. M. (2022) Genetics iyab141. https://doi.org/10.1093/genetics/iyab141
Bishop, S. C., and Stear, M. J. (1997) Anim. Sci. 64(3):469-478.
Robertson, A. (1950) Genetics 35:234-36.
Weiss, G. H., and Dishon, M. (1971) Math. Biosci. 11(3-4):261-265.