Definition and utilization of among hosts heritable variation in reproduction ratio R_0 for infectious diseases

Mahlet T. Anche[*,†], **Mart C. M. de Jong**[†], **P. Bijma**[*]

[*]Animal Breeding and Genomics Centre, Wageningen University, The Netherlands
[†]Quantitative Veterinary Epidemiology Group, Wageningen University, The Netherlands

Abstract: Objective of this study was to show that an individual's breeding value for basic reproduction ratio, R_0 is a function of its own allele frequencies for susceptibility and infectivity and of population average susceptibility and infectivity. When interacting group mates are unrelated, selection for individual disease status captures heritable variation in susceptibility only, yielding limited response in R_0. With related group mates, however, selection also captures heritable variation in infectivity and additional variation in susceptibility, yielding substantial response in R_0. This shows that also genetic variation in susceptibility represents an Indirect Genetic Effect. Furthermore, a Generalized Linear Model (GLM) with complementary log-log link function applied to final size from simulated data was developed to estimate relative gene effects for susceptibility and infectivity.

Keywords: R_0; Relatedness; GLM

Introduction

Infectious diseases are imposing a worldwide concern for the livestock sector, due to their impact on the welfare and productivity of livestock. In order to contain the threat imposed by infectious diseases, different control strategies have been implemented widely. However, the evolution of resistance by bacteria and viruses, and undesirable environmental impacts of antibiotic treatment put these strategies under question (Gibson and Bishop, 2005). Thus, there is a need to investigate additional control strategies, so as to extend the repertoire of possible interventions.

Several studies have demonstrated the existence of genetic variation for susceptibility (resistance) to disease. Such studies, focus on capturing heritable variation in susceptibility (resistance) to disease (Lipschutz-Powell et al. (2012)). However, there also exists (phenotypic) variation in infectivity as can be seen from the occurrence of superspreaders (Lloyd-Smith et al. (2005)).

The basic reproduction ratio, R_0, is the key parameter determining the risk and size of an epidemic. R_0 is the average number of secondary cases produced by a typical infectious individual during its entire infectious life time, in an otherwise naïve population (Diekmann et al. (1990)). Breeding strategies to reduce the risk and prevalence of an infectious disease should aim at reducing R_0, preferably to below a value of 1, because if it is below 1, the disease will die out; else a major outbreak can occur. Genetic improvement aiming to reduce R_0 should ideally be based on the effects of an individual's genes on R_0, which would require defining individual breeding values for R_0. Moreover, defining a breeding value for R_0 would also allow defining heritable variation in R_0, which would give an indication of the prospects for genetic improvement with respect to R_0. In this study, we show how to define breeding value and heritable variation for R_0 for a genetically heterogeneous host population, where individuals differ for susceptibility and infectivity. We also examined mechanisms determining the utilization of heritable variation in R_0, focusing on the effects of kin selection on response in R_0, and in susceptibility and infectivity. Furthermore, we showed that a Generalized Linear Model (GLM) with a complementary log-log link function and a binomial error function applied to the final size of an epidemic can be used to estimate the relative gene effects for susceptibility and infectivity.

Breeding value for R_0: We model genetic heterogeneity in a diploid population using two bi-allelic loci, one locus for susceptibility effect (γ), and the other locus for infectivity effect (φ). The susceptibility locus has alleles G and g, with susceptibility values γ_G and γ_g respectively, and infectivity locus has alleles F and f, with infectivity values φ_F and φ_f, respectively. Assuming additive allelic effects without dominance, genotypic values are given by $\gamma_{GG} = 2\gamma_G$, $\gamma_{gg} = 2\gamma_g$, and $\gamma_{Gg} = \gamma_G + \gamma_g$, for susceptibility, and $\varphi_{FF} = 2\varphi_{FF}$, $\varphi_{ff} = 2\varphi_{ff}$ and $\varphi_{Ff} = \varphi_F + \varphi_f$ for infectivity. Since we assumed additive gene action, average susceptibility in the population is given by,

$$\bar{\gamma} = 2p_g\gamma_g + 2(1 - p_g)\gamma_G, \qquad (1)$$

and average infectivity is given by

$$\bar{\varphi} = 2p_f\varphi_f + 2(1 - p_f)\varphi_F, \qquad (2)$$

where p_f is the frequency of the f allele, and p_g the frequency of the g allele, and the "2" arises because each individual carries two alleles. Note that $\bar{\gamma}$ and $\bar{\varphi}$ are average susceptibility and average infectivity over individuals, not average over allele effects. The objective here is to find R_0 for a heterogeneous population. For that purpose we constructed the Next Generation Matrix (NGM) that describes the number of infectious individual of each type in the next generation of the epidemic, produced by infectious individuals of each type in the current generation. R_0 was then calculated as dominant eigenvalue of the NGM. Under the assumption of separable mixing R_0 can be obtained as the trace of the NGM (Diekmann et al. (2010)). The trace then can be simplified to:

$$R_0 = \bar{\gamma}\bar{\varphi} \, c/\alpha, \qquad (3)$$

where c is the contact rate and α the recovery rate.

We use results from the field of Indirect Genetic Effects (IGEs) to define breeding value for R_0. An IGE is heritable effect of an individual on the trait value of another individual (Griffing 1967; Griffing 1976; Griffing 1981; Moore et al. (1997); Wolf et al. (1998); Muir 2005). Hence,

infectivity is an IGE, since an individual's infectivity affects the disease status of its contacts. Bijma (2011) shows how the approach can be generalized to any trait, including traits that are an emerging property of a population, such as R_0. For an emergent trait, there is only a single trait value for the entire population, and the average effects of alleles on that trait follow from the partial derivatives of the trait value with respect to allele frequency, which will give individual breeding value for R_0 as follows,

$$A_{R_0,i} = 2\left[\overline{\varphi}(\gamma_g - \gamma_G)p_{g,i} + \overline{\gamma}(\varphi_f - \varphi_F)p_{f,i}\right]\frac{c}{\alpha} \quad (4)$$

where $p_{g,i}$ and $p_{f,i}$ refer to the allele frequencies in individual i, thus taking values of 0, ½ or 1.

Heritable variation in R_0: In the following, heritable variation strictly refers to the potential of a population to respond to selection, and may differ from the classical additive genetic variance in a trait. R_0, for example, has no classical additive genetic variance, since there exist no individual phenotypes for R_0. From the above, it follows that heritable variation in R_0 equals the variance in breeding value for R_0 among individuals in the population. Taking the variance of Equation 4, assuming linkage equilibrium, shows that heritable variation in R_0 equals,

$$var\left(A_{R_0}\right) = 2\begin{pmatrix} p_g(1-p_g)\overline{\varphi}^2\left(\gamma_g - \gamma_G\right)^2 + \\ p_f(1-p_f)\overline{\gamma}^2(\varphi_f - \varphi_F)^2 \end{pmatrix}\left(\frac{c}{\alpha}\right)^2 \quad (5)$$

where $var\left(A_{R_0}\right)$ is the variance among individuals in breeding value for R_0. Hence, Equation 5 shows how heritable variation in R_0 depends on the susceptibility and infectivity effects of alleles and on the allele frequencies in the population.

Utilization of heritable variation in R_0: To investigate mechanisms affecting response in R_0, a simulation study was performed on a population sub-divided into 100 groups of 100 individuals each, with discrete generations. (Group size did not affect R_0). A within-group epidemic was started by a single randomly infected individual in each group. At the end of an epidemic, only those individuals that escaped the infection were selected from each group to be parent of the next generation. Hence, selection intensity depended on the severity of the epidemic. The size and the number of groups were kept constant throughout the generations and no transmission of an infection was assumed between groups. Each group was set up in such a way that group mates were genetically related to each other with some degree of relatedness r. Relatedness was defined as the correlation between the genotypes of group mates. The G-allele at the susceptibility locus and the F-allele at the infectivity locus were considered to have favourable effects relative to the g- and f-allele. Four different scenarios were simulated. First, a scenario with heritable variation in both susceptibility and infectivity and groups created randomly with respect to relatedness r among group mates. No LD and a recombination rate θ of 0.5 between both loci were assumed. In this case, a response to selection was observed only at the susceptibility locus, where the G-allele became

fixed after an average of 100 generations (Figure 1). At the infectivity locus, in contrast, only a random fluctuation of allele frequency was observed. As a result, in the final generation, the response in R_0 was limited. Second, varying degrees of relatedness were used, which were the same at both loci. In this case, the population became fixed for the G-allele at susceptibility locus and for F-allele at the infectivity locus (Figure 2). As a result, selection resulted in a greater reduction of R_0 than in the first scenario (Figure 2 vs 1). As relatedness among group mates increased, response was much faster in both traits (Figure 2). As it was also faster at the susceptibility locus, this suggested that also the susceptibility locus showed an IGE. To verify this IGE in susceptibility in the third scenario, we chose to have variation in susceptibility only. Also in this case, the response at the susceptibility locus increased substantially when relatedness among group mates increased. Finally, to investigate the potential effect of relatedness on response in R_0 in the case where there is strong negative LD between both loci and no recombination, a relatedness of either 0 or 0.1 at both loci was simulated. When relatedness $r_\gamma = r_\varphi = 0$, selection fixed the G-allele irrespective of the linked allele at the infectivity locus. As a consequence, selection increased the frequency of f-allele possibly yielding an increase rather than decrease of R_0. When relatedness $r_\gamma = r_\varphi = 0.1$, was used, however, selection caused fixation of the GF-haplotype, resulting in a decrease in R_0. This result shows that selection among related group mates can prevent a maladaptive response to selection.
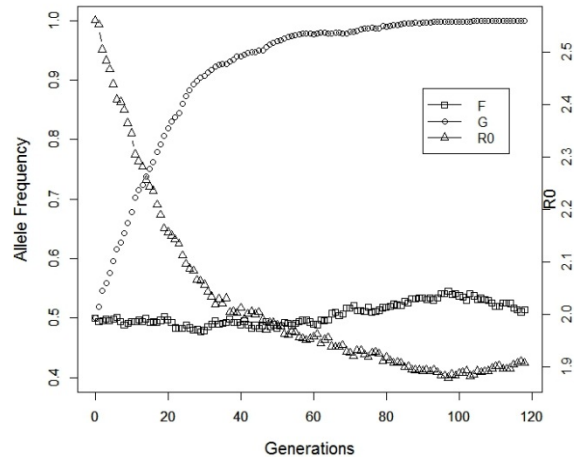

Figure 1. Allele frequency (F) at infectivity locus, allele frequency (G) at susceptibility locus and R_0 as a function of time (generations) when there is no relatedness among group.

Estimation of relative gene effects: The need to estimate the relative gene effect for susceptibility and infectivity motivates the search for a statistical tool that can estimate susceptibility and infectivity from disease data. In this section, we show that a GLM can be fitted to disease data using an equation that describes the number of infected individuals at the end of an epidemic (Andreasen 2011; Lipschutz-Powell et al. (2013). From (Andreasen 2011; Lipschutz-Powell et al. (2013)) the probabilities to escape

from an infection for individuals with (geno)types i (i=1,2 ... n) is given by,

$$\sigma_i = e^{-\gamma_i \sum_{k=1}^{n} \varphi_k n_k (1-\sigma_k)/N} \qquad (6)$$

where σ_i is the fraction of individuals of type i that escape the infection, γ_i is susceptibility of an individual of type i, φ_k is infectivity of individual type k, n_k is fraction of individuals of type k at the beginning of the epidemics, and the summation is over n, i.e. the number of different types. Then a GLM was fitted to simulated outbreak data generated by a SIR model (counts of infected/not-infected at the end of the epidemic) using complementary log-log transformed form of $(1-\sigma_i)$ from equation 6 (Velthuis et al. (2003)). The GLM for a haploid genetic model will then be,

$$cloglog\, E\left[\frac{y_i}{n_i}\right] = intercept + log(\gamma_G)index_G + $$
$$log\varphi_F inf_F + log\frac{inf_{tot}}{N} \qquad (7)$$

where the $cloglog$ is applied to the expectation of y_i which is the number of infected individuals of type i, $index_G$ is a 0/1 dummy variable for the susceptibility type of y_i. inf_F is fraction of infected individuals of type F in the population and $log\frac{inf_{tot}}{N}$ is fraction of infected individuals in the whole population and it was set as an offset in the GLM. The geometric mean, instead of arithmetic mean approximation was applied to linearize the $log\varphi_F inf_F$ part of equation (7). Note that Equation (7) assumes $\gamma_g = 1$ and $\varphi_f = 1$.

The first results from a haploid genetic model suggest that estimates for the relative allele effects for susceptibility and infectivity seem to be unbiased or have a small bias and have small variance.
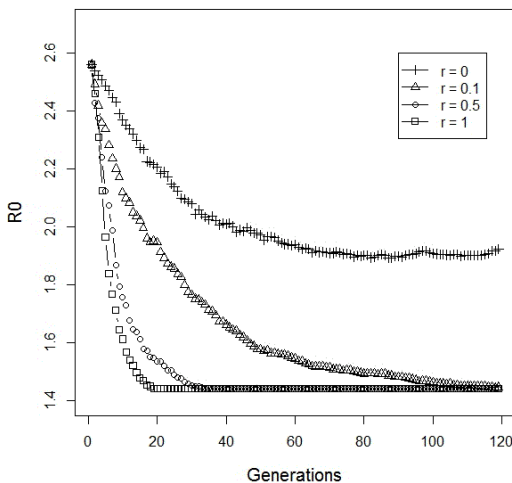

Figure 2. R_0 as function of time (generations) for different values of relatedness (r).

Discussion

Our result showed that relatedness among group members increased response in R_0 in two ways. First, it increases response in R_0 by capturing the heritable variation in infectivity, which is fully an IGE of the individual. Second, relatedness among group mates increases response in susceptibility. This occurs because an individual that carries the favourable allele for susceptibility on-average has fewer infected group mates, which increases its probability of escaping the epidemic and being selected. These results show that not only infectivity, but also susceptibility exhibits an IGE and this genetic variance is utilized by relatedness among group mates. The net result of both mechanisms is a strong increase in response to selection in R_0 when relatedness among group mates increases. Finally, we developed the method to estimate gene effects for susceptibility and infectivity from the final outcome of an epidemic, using a GLM with a complementary log-log link function.

Conclusion

Our results show that the use of related groups can greatly accelerate response to selection in R_0, because it utilizes the indirect genetic variance in both susceptibility and infectivity in the host population. For any actual case, the potential impact of kin selection will of course depend critically on the magnitude of this indirect genetic variance. Furthermore, complementary log-log transformation to the final size equation is required to fit a GLM to a simulated data for estimation of relative gene effects for susceptibility and infectivity.

Literature cited

Andreasen, V., (2011). Bull. Math. Biol. 73(10)**:** 2305-2321.

Bijma, P., (2011). Genetics 189(4)**:** 1347-1359.

Bijma, P., Wade, M., (2008). J. Evol. Biol. 21(5)**:** 1175-1188.

Diekmann, O., Heesterbeek, J., Metz, J. A., (1990). J. Math. Biol. 28(4)**:** 365-382.

Diekmann, O., Heesterbeek, J., Roberts, M,. et al. (2010). J. R. Soc. Interface 7(47)**:** 873-885.

Gibson, J. R., Bishop, S. C., (2005). Rev. Sci. Tech. Oie 24(1)**:** 343-353.

Griffing, B., (1967). Aust. J. Biol. Sci. **20**(1)**:** 127-142.

Griffing, B., (1976). Genetics 82(4)**:** 703-722.

Griffing, B., (1981). J. Theor. Biol. 89(4)**:** 659-677.

Lipschutz-Powell, D., Woolliams J. A., Bijma, P., et al. (2012). PloS one 7(6)**:** e39551.

Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., et al. (2005). Nature 438(7066)**:** 355-359.

Lipschutz-Powell, D., Woolliams, J. A., Wilson, A. B. (2014) Genet. Sel. Evol. 46:15

Moore, A. J., Brodie, III E. D., Wolf J. B., et al. (1997). Evolution 51**:** 1352-1362.

Muir W. M. (2005). Genetics 170(3)**:** 1247-1259.

Velthuis, A., De Jong, M. C. M., Kamp, E., et al. (2003). Prev. Vet. Med. 60(1)**:** 53-68.

Wolf, J. B., Brodie, III E. D., Cheverud, J. M., et al. (1998). Trends Ecol. Evol. 13(2)**:** 64-69.