The Contribution of Genetic Diversity to the Spread of Infectious Diseases in Livestock Populations

A. J. Springbett,[1] K. MacKenzie,[2] J. A. Woolliams and S. C. Bishop

Roslin Institute (Edinburgh), Midlothian EH25 9PS, United Kingdom

Manuscript received May 2, 2003
Accepted for publication July 28, 2003

ABSTRACT

This article uses stochastic simulations with a compartmental epidemic model to quantify the impact of genetic diversity within animal populations on the transmission of infectious disease. Genetic diversity is defined by the number of distinct genotypes in the population conferring resistance to microparasitic (*e.g.*, viral or bacterial) infections. Scenarios include homogeneous populations and populations composed of few (finite-locus model) or many (infinitesimal model) genotypes. Genetic heterogeneity has no impact upon the expected value of the basic reproductive ratio (the primary description of the transmission of infection) but affects the variability of this parameter. Consequently, increasing genetic heterogeneity is associated with an increased probability of minor epidemics and decreased probabilities of both major (catastrophic) epidemics and no epidemics. Additionally, heterogeneity *per se* is associated with a breakdown in the expected relationship between the basic reproductive ratio and epidemic severity, which has been developed for homogeneous populations, with increasing heterogeneity generally resulting in fewer infected animals than expected. Furthermore, increased heterogeneity is associated with decreased disease-dependent mortality in major epidemics and a complex trend toward decreased duration of these epidemics. In summary, more heterogeneous populations are not expected to suffer fewer epidemics on average, but are less likely to suffer catastrophic epidemics.

THERE is substantial evidence that resistance to infectious disease in animals has a genetic component and it has often been shown that there are genetic differences in response to various infectious challenges (summarized for livestock species by Office International des Epizooties 1998; Axford *et al.* 2000; Bishop *et al.* 2002). The implication of this observation is that genotypes for resistance to a particular pathogen in a host population will influence the transmission of that pathogen through the population and hence the likely disease impact. Both the mean level of resistance of the population and the variability of resistance, *i.e.*, the genetic heterogeneity, may have an impact upon the transmission of the infection. The effects of genetic heterogeneity are potentially important with respect to livestock management strategies because genetic heterogeneity and its maintenance are associated with the structure and genetic management of the population (*e.g.*, effective population size).

Several authors have discussed the effect of host heterogeneity on the ability of an infection to establish itself in a population (Hethcote 1977; May and Anderson 1989; Adler 1992; Dushoff and Levin 1994). This discussion is usually centered upon variation in the value of the basic reproductive ratio R_0, which is the expected number of secondary infections arising directly from an initial infection. In general a pathogen will invade a homogeneous population only if $R_0 \geq 1$. When $R_0 < 1$ no epidemic is expected. In populations consisting of several groups there may be distinct values of R_0 for each group. This heterogeneity can occur in a variety of ways and may arise from environmental, behavioral, or genetic factors. The predicted impact upon the course of an infection depends upon the nature of the assumptions that are made but some useful conclusions of general validity have been reported. For example, Adler (1992) considered the impact of nonrandom mixing between groups caused by geographical location. He found that estimated values of R_0 based on averages over a population tend to be biased downward, producing over-optimistic predictions of the likelihood of avoiding epidemics. Dushoff and Levin (1994) consider the case where there is random mixing of groups with differing values of R_0 and find that essentially the same result holds for a heterogeneous population as for a homogeneous population; *i.e.*, an epidemic will occur only if the average population value of $R_0 \geq 1$. May and Anderson (1989) look at heterogeneity in sexual activity during an AIDS epidemic and find that the predicted proportion of infected individuals decreases as the variation in sexual activity increases. In all of these articles the results presented are based upon expectations from deterministic models.

[1]*Corresponding author:* Department of Genetics and Biometry, Roslin Institute, Roslin, Midlothian EH25 9PS, United Kingdom. E-mail: anthea.springbett@bbsrc.ac.uk

[2]*Present address:* BioSS, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, United Kingdom.

The overall impact of host genetic heterogeneity on the transmission of infection is a noticeable gap in the literature. In situations where population management is possible, such as in domestic or zoo populations, the potential disease risks associated with various genetic management strategies need to be investigated and brought to the attention of geneticists. Risks are a function both of mean epidemic outcomes and of the variability of outcomes. Quantifying the variability of outcomes may be achieved by using stochastic rather than deterministic epidemic models.

In this article we use a stochastic epidemic model to examine the spread of microparasitic infections in livestock populations of varying genetic diversity and to draw conclusions about the relationship between genetic heterogeneity and the impact of disease. This impact is measured not merely in terms of expectations, as has usually been the case in the literature, but as a full range of possible outcomes. In addition to summarizing epidemics in terms of R_0, the stochastic model also provides detailed information on the proportion of the population that becomes infected, the duration of the epidemic, and disease-dependent mortality as a function of genetic heterogeneity.

METHODOLOGY

Host population: The host population was assumed to consist of a number (n) of groups of animals. Within each group all animals shared the same genotype for susceptibility to a notional infection and hence were equally susceptible (or resistant) to infection. However, the different genotypes of the different groups conferred varying degrees of susceptibility to this pathogen between groups.

The susceptibilities to infection of the different genotypes were selected at random by sampling from statistical distributions, as described below. It was assumed that susceptibility to infection was the sum of contributions from several genes of which none had an overriding effect, corresponding to a finite locus model with several genes or, when n is large, an infinitesimal model. However, when n is small this model is also consistent with a single gene controlling resistance, with genotypes defined by the combination of alleles at one locus. For example, five common alleles at the PrP locus lead to 15 distinct genotypes for scrapie resistance in sheep.

For simplicity, full contact between groups was assumed with random mixing of all animals. The population was also considered to be static, with no births, migration, or deaths, except for those induced by infection (described below).

Infection model: A susceptible, infected, recovered (SIR) compartmental model was assumed to describe the infection dynamics (Anderson and May 1992); however, this was extended in some cases to include infection-induced mortality. Infection was assumed to be transmitted only by direct contact between hosts. Initially, the population was assumed to be immunologically susceptible, *i.e.*, had never previously been exposed to the notional pathogen. Potential epidemics were then triggered by exposing the population to the notional pathogen, through the introduction of a single individual. The time courses of the epidemics were then quantified by means of stochastic epidemic models, as described below.

Standard theoretical results for deterministic SIR models describing the relationships between total proportion of the population infected during the course of an epidemic (I), the basic reproductive ratio (R_0), the size of the population (N), the transmission parameter (β), and the recovery rate (γ) apply to the populations. The transmission parameter (b) is the expected number of new infections per infectious individual per susceptible individual per day and the recovery rate (γ) is the inverse of the infectious period of the disease. The asymptotic relationship between I and R_0 for a single genotype, $n = 1$, is

$$I = 1 - \exp(-IR_0), \quad \text{where } R_0 = \beta N/\gamma. \qquad (1)$$

There is no general analytical solution for n genotypes. This has been shown by Hethcote (1977), who derived an asymptotic result for the numbers of infected individuals in each group in the form of a set of n simultaneous equations for a population consisting of n homogeneous groups (see appendix a). The equations are functions of the numbers of individuals in the groups, their reproductive rates, and their recovery rates.

When n is greater than one and variability is expected in R_0, variability will also be observed in I. An approximate value can be generated for the variance of I as a function of the variance of R_0 by using a Taylor series expansion for $\ln(1 - I)$ (see appendix b):

$$\mathrm{Var}(I) = \mathrm{var}(R_0)J^2I^4/[I + J\ln J]^2,$$
$$\text{where } J = 1 - I,\ 0 < I < 1. \qquad (2)$$

This gives a measure of the variation to be expected in I across epidemics with differing values of R_0. It is useful for comparisons between the predicted values of I derived from the deterministic model and results obtained for I from the stochastic model described below.

Stochastic simulation: A stochastic setting enables exploration of the impact of variability more easily than does a deterministic model. In particular, it allows variability in both pathogens and host populations, particularly in the basic reproductive ratio. In this article R_0 describes the expected basic reproductive ratio of a particular pathogen across all potential host populations. R describes the basic reproductive ratio in a particular population; hence, it is a function of the host population genotype.

TABLE 1

Parameter values used for stochastic simulation of epidemics

Parameter	Values or distribution
N	1000
R_0	Gamma: $(\alpha, \theta) = (2.5, 0.6)$ or $(20, 0.075)$
R	Lognormal: mean R_0, CV 0.75 or 1.5
n	1, 2, 10, 100
m	0 or 0.08 or 0.16 deaths/infected animal/day
γ^{-1}	14 or 28 days

We present results for stochastic simulations of epidemics with and without disease-dependent mortality and with variation between subgroups in susceptibility to infection. The parameters of the model were the number of genotypes in the population (n), the susceptibilities of the genotypes (β_i, $i = 1, \ldots, n$), the contact rate between genotypes, and the infectious period of the disease (γ^{-1}). Both the contact rate and the recovery rate were assumed to be independent of genotype. Disease-dependent mortality was also assumed to be constant across genotypes and had expected values of 0.00, 0.08, or 0.16 deaths per infected individual per day. It was assumed that there was no mortality from other causes. The chosen population of size $N = 1000$ was divided into n genotypes, each having the same number of individuals. The population-specific basic reproductive rate, R, is a function of N, γ^{-1}, and β_i. The contact among animals was assumed to be random with equal mixing, and the infectious period was fixed at either 14 or 28 days for all genotypes. The choice of recovery rate is not critical because it has little impact upon the pattern of the results. It is essentially a scaling factor. The parameter values used in the simulations are all listed in Table 1.

The epidemic was initiated by the introduction of a single infected individual into the susceptible population. The model then simulated the occurrences of three types of events: infection of a susceptible animal, recovery of an infected animal, and death of an infected animal, and the time at which these events took place. The epidemic terminated either when no more susceptible animals were left in the population or on the death or recovery of the last infected animal. It is described in detail in Mackenzie and Bishop (2001).

The expected value of R, R_0, was sampled from a gamma distribution. Two gamma distributions were chosen, with equal means and different variances, to enable a comparison of results for different degrees of variation in R_0. Following standard distribution theory the parameters of the gamma distribution were α and θ; thus the mean of the distribution was $\alpha\theta$ and the variance was $\alpha\theta^2$. The chosen distributions were (i) $\alpha = 2.5$, $\theta = 0.6$, $\alpha\theta = 1.5$, $\alpha\theta^2 = 0.9$; and (ii) $\alpha = 20.0$, $\theta = 0.075$, $\alpha\theta = 1.5$, $\alpha\theta^2 = 0.11$. Assuming an average value of 1.5 for R_0, with an associated variance of 0.9 or 0.11, implies that the full range of outcomes is possible, ranging from no epidemic up to an epidemic such that the entire population becomes infected. The values of R_i, $i = 1, \ldots, n$ for each genotype were sampled from a lognormal distribution with mean R_0 and coefficient of variation (CV) of either 0.75 or 1.5. The value of 0.75 for the CV was chosen to be typical of variation among animals in disease resistance data (*e.g.*, Stear *et al.*, 1995) and 1.5 was used for comparative purposes.

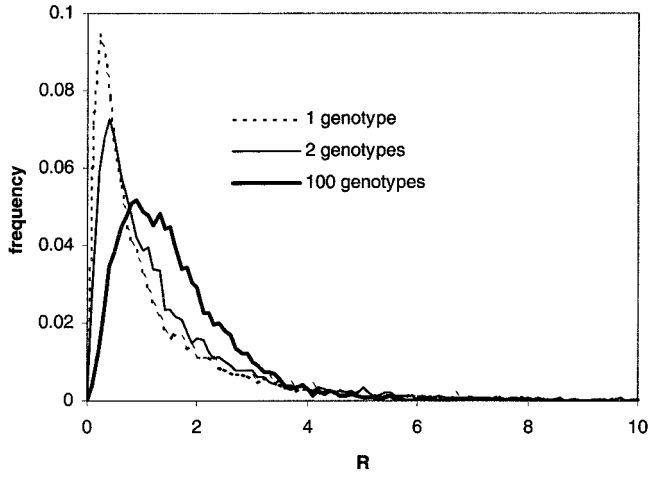

Figure 1.—Frequency distribution for average observed reproductive ratio (R).

Ten thousand simulations were run for each choice of gamma distribution and for populations consisting of $n = 1$, 2, 10, and 100 genotypes. The results provided estimated distributions for the average realized basic reproductive rate, the proportion of animals infected, and disease-dependent mortality and information about the probabilities of severe or mild epidemics or no epidemic. For ease of computation, we defined a minor epidemic as one in which $<10\%$ of the population became infected and a major epidemic as one in which at least 10% became infected. This choice can be justified on the grounds that epidemics that die out quickly generally result in $<10\%$ of the population becoming infected. Comparisons concerning the frequency and severity of minor and major epidemics can be made between homogeneous and heterogeneous populations for different degrees of variation in the distribution of the expected basic reproductive ratio, R_0.

RESULTS

Distribution of basic reproductive rate: Expected reproductive ratios (R_0) were sampled from the two chosen gamma distributions (see Table 1 for parameter values). As expected, the mean was equal to 1.5 for both distributions and the variances were 0.9 and 0.11, respectively. Figure 1 shows the distribution of the average observed reproduction rate when R_0 is drawn from

TABLE 2

Summary statistics for the distribution of average observed reproductive ratio, *R*

Gamma distribution	No. of genotypes	$P(R \leq 1)$	Average	Maximum	Variance
(2.5, 0.6)	1	0.495	1.50	21.8	2.63
	2	0.436	1.50	14.1	1.68
	10	0.367	1.50	10.5	1.05
	100	0.346	1.50	7.64	0.90
(20.0, 0.075)	1	0.417	1.50	24.2	1.59
	2	0.323	1.50	12.8	0.83
	10	0.146	1.50	4.93	0.25
	100	0.067	1.50	3.34	0.13

R is sampled from a lognormal distribution with coefficient of variation 0.75 and mean R_0, which in turn is sampled from a gamma distribution with parameters $(\alpha, \theta) = (2.6, 0.6)$ or (20.0, 0.075).

the first of the gamma distributions, with parameters $\alpha = 2.5$ and $\theta = 0.6$. The reproductive rates for the genotypes are then sampled from a lognormal distribution with mean R_0 and coefficient of variation 0.75. The distribution is shown both for a homogeneous population with one genotype and for populations with increasing heterogeneity. The mean is equal to 1.5 in all cases but the variance in *R* decreases with increasing population heterogeneity.

Summary statistics for *R* are shown in Table 2 and it is apparent that the degree of genetic heterogeneity will affect expected epidemic outcomes. For a homogeneous population the probability that $R \leq 1$ is greater than for a heterogeneous population. For example, if R_0 is sampled from a gamma (20.0, 0.075) and 100 genotypes are simulated, the probability that $R \leq 1$ is 0.067. This means that an epidemic has the possibility of occurring on almost every occasion that an infected individual is introduced into the population. By contrast, if R_0 is sampled from the same distribution but only one genotype is simulated, *i.e.*, the population is homogeneous, the probability that $R \leq 1$ is 0.417. In other words, an epidemic has the possibility of occurring on only 58% of occasions following an initial infection. Similarly, the maximum observed value of *R* is greatest for the homogeneous population. This implies that homogeneous populations are more likely than heterogeneous populations to suffer very serious epidemics; however, the incidence of such epidemics will be low.

Probabilities of no epidemic, minor, and major epidemics: Epidemics can be classified in terms of the proportion of the population that is infected. There is no epidemic if the initial infection gives rise to no secondary infections. In this article, a minor epidemic is defined as one in which <10% of the population is infected; otherwise the epidemic is classified as major. Table 3 shows the probabilities that the introduction of an infected animal results in the occurrence of either no epidemic or a minor or a major epidemic, for populations of differing heterogeneity. The probability of either a major epidemic or no epidemic decreases as the heterogeneity in the population increases. Correspondingly, the probability of a minor epidemic becomes greater. This result is consistent for both gamma distributions. When the CV of the lognormal distribution for *R* is increased from 0.75 to 1.5 the probabilities of either no epidemic or a major epidemic both increase with increasing genetic heterogeneity for all populations except the most diverse. When there are 100 genotypes the probability of a major epidemic decreases when the CV rises to 1.5. This effect occurs because the curve for the total proportion infected (*I*) as a function of *R* is very much lower and flatter for very heterogeneous populations (described below). Thus the increased variation does not produce a corresponding increase in the number of values for *I* above the 10% threshold for a major epidemic.

Proportion of population infected: In addition to the classification of epidemic type, extra insight can be gained by considering the proportion of the population that becomes infected during the course of the epidemic. Figure 2, A–D, shows the total proportion of animals infected *vs.* the average observed basic reproductive rate for populations with 1, 2, 10, and 100 genotypes, respectively, for the more variable gamma distribution (2.5, 0.6) and a CV of 0.75 for *R*. Results for the gamma with lower variance are similar and are not presented here. Figure 2A, for $n = 1$, closely follows the theoretical expected result for a homogeneous population, $I = 1 - \exp(-IR_0)$. The values on or close to the base of the figure represent cases where there are minor or no epidemics, both of which occur with probability $1/(R_0 + 1)$ in a fully mixed population when $R_0 \geq 1$ (Bishop and Mackenzie, 2003), assuming an SIR model. Figure 2B shows extensive deviation of observed values below this average prediction. Figure 2, C and D, shows a systematic departure from it. When the number of genotypes is 100 the total proportion infected in the simulations never reaches one.

Approximate variances for the predicted value of *I*,

TABLE 3

Probabilities of no epidemic, minor, and major epidemics when mortality is 8% and recovery time is 14 days

CV	No. of genotypes	Epidemic type: Gamma (2.5, 0.6) None	Gamma (2.5, 0.6) Minor	Gamma (2.5, 0.6) Major	Gamma (20.0, 0.075) None	Gamma (20.0, 0.075) Minor	Gamma (20.0, 0.075) Major
0.75	1	0.66	0.26	0.077	0.64	0.31	0.055
	2	0.64	0.29	0.068	0.62	0.34	0.043
	10	0.63	0.32	0.056	0.61	0.38	0.016
	100	0.62	0.33	0.053	0.59	0.40	0.006
1.5	1	0.70	0.21	0.083	0.69	0.23	0.072
	2	0.68	0.25	0.077	0.65	0.29	0.068
	10	0.63	0.30	0.063	0.61	0.36	0.031
	100	0.62	0.33	0.048	0.59	0.40	0.005

R is sampled from a lognormal distribution with coefficient of variation 0.75 or 1.5 and mean R_0, which in turn is sampled from gamma distributions with $(\alpha, \theta) = (2.5, 0.6)$ or $(20.0, 0.075)$.

in homogeneous populations, can be calculated using the theoretical expectation (2) above, assuming $R \geq 1$. These can be then compared with the empirical variances calculated from the simulations, shown in Figure 3A for $n = 1, 2, 10, 100$. The empirical variances for $R \geq 1$ were calculated from data excluding values of $I < 0.01$, *i.e.*, trivial or nonepidemics corresponding to the "foot" along the *x*-axis seen in Figure 2, A–D. We are

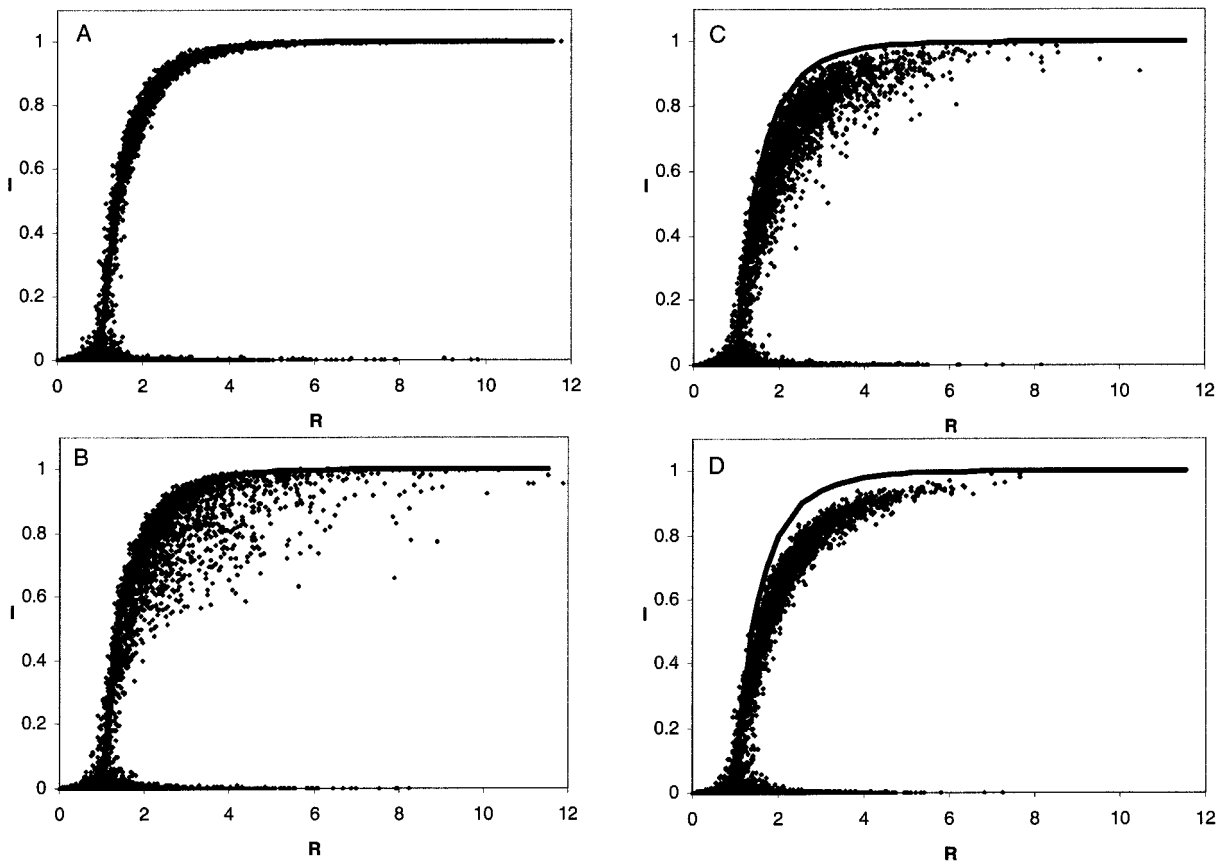

FIGURE 2.—(A) Total proportion of population infected (I) *vs.* average observed reproductive ratio (R): $n = 1$, $m = 0$, $\gamma^{-1} = 14$ days, CV = 0.75. (B) Total proportion of population infected (I) *vs.* average reproductive ratio (R): $n = 2$, $m = 0$, $\gamma^{-1} = 14$ days, CV = 0.75. (C) Total proportion of population infected (I) *vs.* average reproductive ratio (R): $n = 10$, $m = 0$, $\gamma^{-1} = 14$ days, CV = 0.75. (D) Total proportion of population infected (I) *vs.* average reproductive ratio (R): $n = 100$, $m = 0$, $\gamma^{-1} = 14$ days, CV = 0.75.

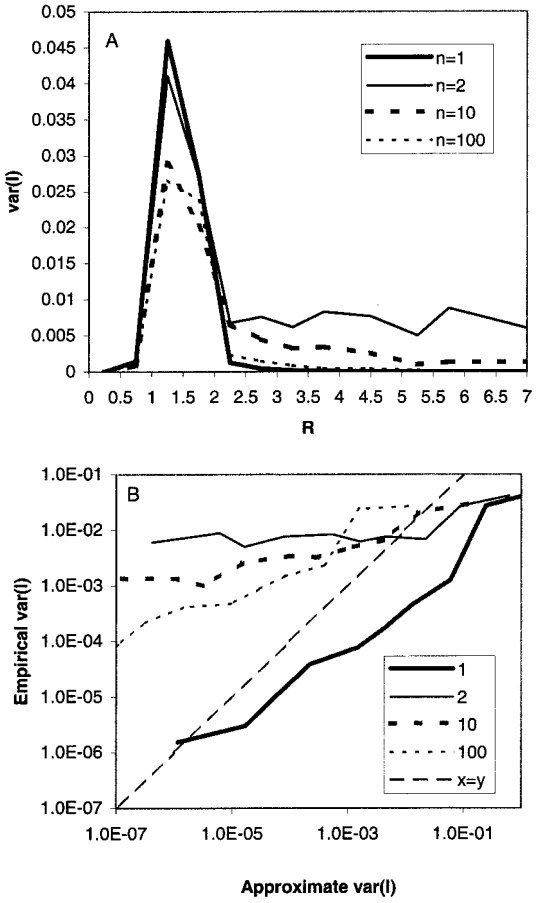

FIGURE 3.—(A) Empirical variance of total proportion of population infected (I) *vs.* average reproductive ratio (R). (B) Empirical *vs.* approximate variance of total proportion of population infected (I).

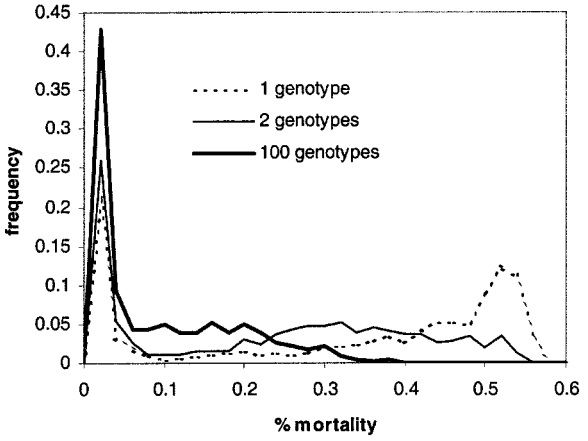

FIGURE 4.—Mortality in nontrivial epidemics (total proportion of population infected > 0.01): average observed reproductive ratio (R) from gamma(2.5, 0.6), lognormal CV = 0.75, $m = 0.08$, $\gamma^{-1} = 14$ days.

interested in the variation in the main body of values of I and not in the extreme values within the foot, which are present for all values of n and, if included, have a large effect upon the estimated variance in all cases. Figure 3A shows that the empirical variance of I is high for all values of n when R lies between 1 and 2. This is the region where the observed slope of the curve for I *vs.* R is greatest. The largest slopes correspond to $n = 1, 2$. The slopes for $n = 10$, 100 are lower (see Figure 2, A–D) and this is reflected in the heights of the peaks for the variance of I. As R increases the variance of I for $n = 1$, 2 falls rapidly and continues at a low level for all $R \geq 2.5$. The variance of I for $n = 2$ does not fall so rapidly and stabilizes at a higher value. The variance for $n = 10$ is intermediate. Figure 3B shows the empirical variances plotted against the approximate theoretical variances. The agreement for $n = 1$ is good. However, when $n > 1$ empirical variances are generally considerably greater than expected variances in homogeneous populations, especially when the expected variances are small. This reflects the departure of the empirical relationship between I and R from the theoretical expectation for $n > 1$, and the contrast between empirical and expected variances is generally greatest when $n = 2$. Thus, the empirical variances of I observed in heterogeneous populations show complex but systematic departures from those expected in homogeneous populations.

Disease-dependent mortality: Figure 4 shows observed disease-dependent mortality during minor and major epidemics (but excluding trivial epidemics for which $I < 0.01$) for populations with 1, 2, or 100 genotypes when R_0 is drawn from a gamma distribution with parameters $\alpha = 2.5$ and $\theta = 0.6$. The distribution for $n = 10$ is very similar to that for $n = 100$ and is not shown. The mortality in trivial epidemics ($I < 0.01$, not shown) is effectively zero. Results are shown for disease-dependent mortality of 8% of infected individuals per day; however, 16% mortality gave a similar pattern of results. Expected mortality decreases with increasing heterogeneity. For the data shown in Figure 5 average mortality is 0.32 (0.006) for $n = 1$, 0.22 (0.005) for $n = 2$, and 0.085 (0.003) for $n = 100$. This is due to the lower number of animals from heterogeneous populations that are infected in major epidemics. The main difference between distributions for $n = 1$, 2, 10, and 100 is that there is a small peak between 50 and 60% mortality for $n = 1$, which is not present in the distributions for the heterogeneous populations. All distributions have a sharp peak at ~20% mortality. This is highest for the most heterogeneous populations. The frequency distributions are a function of the total proportion of the population infected during the course of an epidemic; however, it should be noted that when mortality is greater than zero the profile of infection is different from those shown in Figure 2, A–D. The curve for I is shifted to the right so that the total proportion infected for a given R is equivalent to that for a lower value of R in the absence of mortality. This effect is explained by the fact that death removes an infective individual

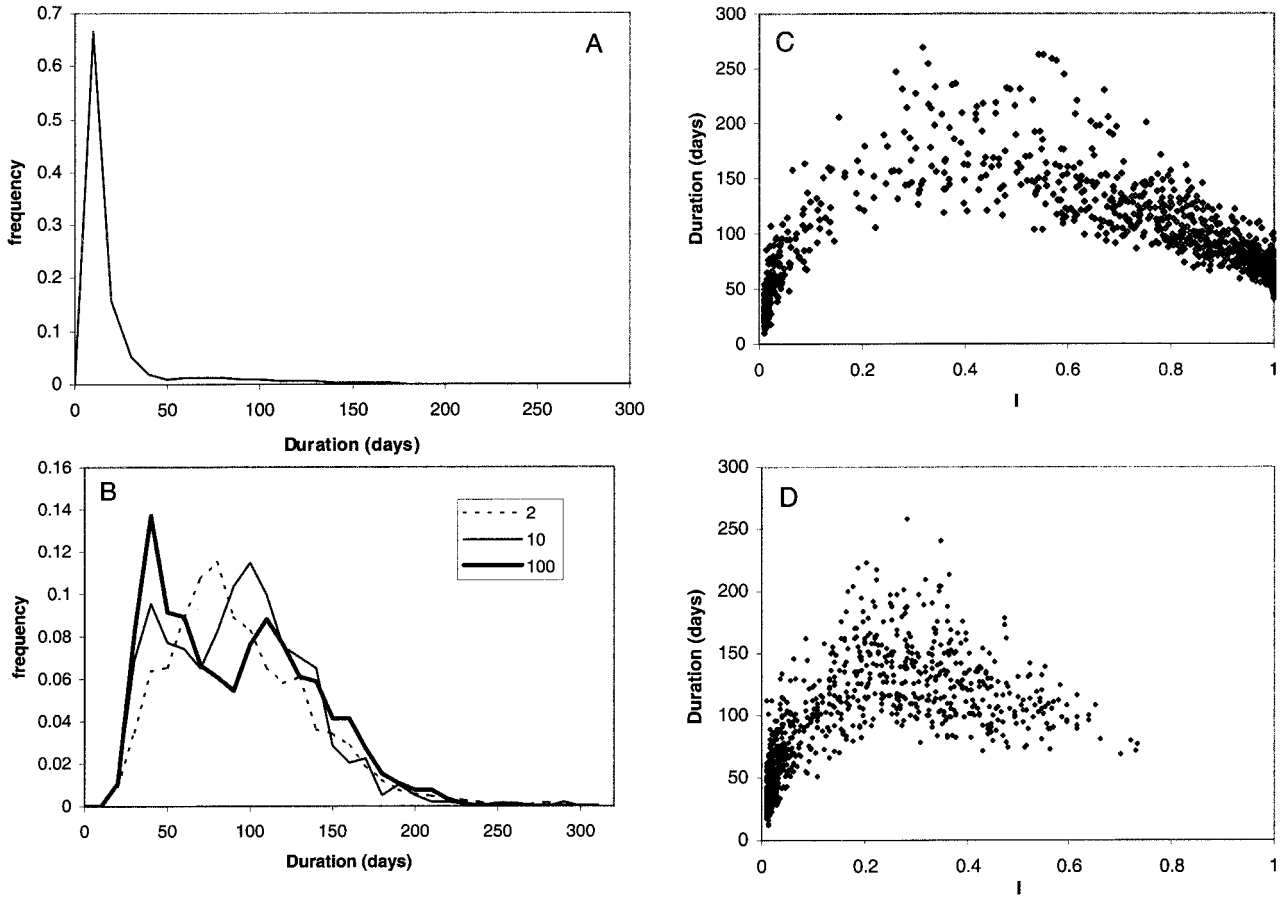

FIGURE 5.—(A) Frequency distribution of duration of all epidemics: average observed reproductive ratio (R) from gamma(2.5, 0.6), lognormal CV = 0.75, m = 0.08, γ^{-1} = 14 days. (B) Frequency distribution of duration of nontrivial epidemics (total proportion of population infected >0.01): average observed reproductive ratio (R) from gamma(2.5, 0.6), lognormal CV = 0.75, m = 0.08, γ^{-1} = 14 days. (C) Duration of nontrivial epidemics (total proportion of population infected >0.01) *vs.* total proportion of population infected (I) for n = 1. (D) Duration of nontrivial epidemics (total proportion of population infected >0.01) *vs.* total proportion of population infected (I) for n = 100.

from the population before it has recovered and so reduces the overall probability that it infects others. This effect is consistent across both homogeneous and heterogeneous populations.

Epidemic duration: Genetic diversity appears to have little impact upon the overall duration of epidemics when all are considered together. However, for minor and major epidemics (excluding trivial epidemics for which $I < 0.01$) there is a trend for the average duration to decrease with increasing heterogeneity. This is associated with a change in the distribution of durations as the number of genotypes rises. The distribution of all epidemics is shown in Figure 5A and the distribution of epidemics with $I > 0.01$ is shown in Figure 5B for R_0 sampled from gamma (2.5, 0.6) (mortality rate of 0). The distribution for n = 1 is very similar to that for n = 2 and is not shown. Figure 5A shows no apparent effect of diversity because the majority of the distribution arises from minor epidemics where the differences are negligible. However, when nontrivial epidemics with $I >$ 0.01 are considered there is a noticeable alteration in the duration distribution as n increases from 2 to 10, as seen in Figure 5B. The distributions for n = 1 and n = 2 are both unimodal. The distributions for n = 10 and n = 100 are bimodal. The means (with standard errors in parentheses) for the distributions for n = 1, 2, 10, and 100 are 92.2 (1.4), 88.7 (1.3), 87.4 (1.3), and 86.9 (1.5) days, respectively. The reason for the change from unimodality to bimodality can be seen in Figure 5, C and D. These show duration plotted as a function of the total proportion infected (I) for nontrivial epidemics with n = 1 (Figure 5C) and n = 100 (Figure 5D). The longest epidemics, with durations >200 days, occur at intermediate values of I, between 0.2 and 0.7. Epidemics in which either <10% or >90% of the population become infected do not last for > ~150 days. The epidemic either dies out very quickly or takes hold and passes through most of the population very quickly. The frequency distributions for n = 1 and n = 2 are unimodal because the majority of epidemics occur at extreme values of I and have similar durations. When the population is more diverse, n = 10 and n = 100,

there are no epidemics with values of I in excess of 0.9 and relatively more with intermediate values of I. This produces a second peak at a higher duration than that associated with very low values of I.

DISCUSSION

The aim of this study was to estimate the effect of genetic variability on both the probability of occurrence of an epidemic and its potential severity following the introduction of a microparasitic infection into a susceptible population. Our main results may be summarized as follows. Genetic heterogeneity, with random mixing between genotypes, has no impact upon the mean observed R_0; however, it does affect the variability in R_0 values. The consequence of this is that increased genetic heterogeneity is associated with an increased probability of minor epidemics and decreased probabilities of both major epidemics and no epidemics. Additionally, heterogeneity *per se* is associated with a breakdown in the expected relationship between R_0 and I developed for homogeneous populations, with epidemics generally infecting fewer animals than expected given the mean population value of R_0 (using arguments developed for homogeneous populations). The joint effect of these two factors is that increased heterogeneity is associated with decreased disease-dependent mortality in nontrivial epidemics and a complex trend toward decreased duration of these epidemics. It is important to note that for a homogeneous population our results, in terms of epidemic type and outcome, are as anticipated from the deterministic theory.

The general pattern of the effects of genetic diversity upon the outcome of epidemics was consistent for all choices of the mortality rate, for both of the gamma distributions used for generating R_0 and for both choices of the coefficient of variation for the lognormal distribution used to generate R. Increased mortality simply decreases the effective values of R_0 and R independently of the number of genotypes in the population. Similarly, changes in the variance of the distribution of R_0 and the coefficient of variation of the distribution of R alter the variation observed in the results but do not change the pattern or the general interpretation of the results.

A considerable body of published theory exists, based on deterministic models of epidemics describing the impacts of various specific types of heterogeneity. However, this theory generally provides results only for the expected outcome of epidemics and does not deal fully with the impact of variation (Hethcote 1977; May 1987; May and Anderson 1989; Adler 1992; Dushoff and Levin 1994). Moreover, this literature does not make extensions to genetic heterogeneity or give results that may be directly and easily interpreted by geneticists.

The study with the greatest analogy to ours is that of May and Anderson (1989) who describe a situation in which contact rates among subpopulations vary according to a gamma distribution. They present results in which the total proportion of infected individuals in the population (I) depends not only upon the expected value of R_0 but also upon the CV of the distribution of R_0. In general terms, the I values that May and Anderson (1989) present for a given R_0 decrease as the CV increases, with values for heterogeneous populations always being less than those for homogeneous populations. In principle this result is consistent with ours although presented in a different setting. However, in terms of the specific results, our I values converge to an asymptote similar to those expected for homogeneous populations for high values of R, whereas the I values presented by May and Anderson (1989) converge to much lower values. It is difficult to tell whether the two sets of results are consistent with each other, first because of the differences between the models and the choice of distributions and second because May and Anderson (1989) do not provide an analytical derivation of their results.

Despite the uncertainty of the comparison with May and Anderson (1989), the interpretation of their result is instructive. Their interpretation, derived in the context of a model for AIDS, is that the epidemic burns itself out in highly active subgroups of the population when the coefficient of variation is high. An example to support this argument is given in May (1987), which shows that the predicted proportion of individuals infected in a population divided into six categories classified by sexual activity is directly proportional to the level of activity. The equivalent interpretation in the context of our model is that the disease spreads quickly among highly susceptible genotypes but dies out because of the relatively high number of less susceptible genotypes in a heterogeneous population. This effect is magnified when the coefficients of variation of the distributions for R_0 and R are increased. The overall conclusion from both studies is that heterogeneity, in whatever form, does alter the expected relationship between R_0 and I.

The results discussed here are based on a model that assumes equal subgroup sizes. This is a simplifying assumption chosen to illustrate the impact of heterogeneity and is unlikely to hold exactly in either natural or selected populations. Equal subgroup sizes will maximize the influence of heterogeneity. For example, we have investigated populations with two subgroups where the ratio of subgroup sizes varied from 1:1 (*i.e.*, subgroups of equal size) to 19:1. As the inequality in subgroup sizes increased, both the probability of a major epidemic and the total number of individuals infected increased toward the values for a homogeneous population; *i.e.*, the limit as the size of the smaller subgroup tends to zero. The beneficial effect of heterogeneity is at a maximum when subgroup sizes are equal or approximately equal. This pattern generalizes to populations with any number of subgroups but the impact of varying subgroup sizes becomes smaller and more difficult to

quantify satisfactorily as the number of subgroups increases.

A comment is warranted on the additional insight gained from using a stochastic modeling approach. Deterministic approaches remain important for providing elegant insights into expected outcomes of biological processes. However, in the case of complex and nonlinear processes deterministic solutions may prove difficult to obtain and stochastic simulation may more easily yield solutions. Additionally, the stochastic approach used here has yielded additional information on variability of outcomes. In particular, the variability of the relationship between R_0 and I and the mortality distributions yield insight that is novel and unlikely to have been obtained by a deterministic approach.

The main implication for geneticists of the results presented in this article is that heterogeneity in disease resistance is potentially a useful characteristic for protecting a population from very serious epidemics. Homogeneous populations may have fewer epidemics on average, but are more likely to suffer catastrophic epidemics. Published theory has shown that the spread of an epidemic through a homogeneous population can be represented accurately by a nonlinear system. Thus inappropriate averaging of parameters within the system can produce misleading predictions, and theoretical results that ignore variation in contact rates, transmission rates, and susceptibility produce expressions for the expected outcomes that are inaccurate or inappropriate in certain circumstances. The variation about the expected values for total proportion infected and other parameters of interest must also be considered. As we have shown above this variation may be considerable in practice. May and Anderson (1989) suggest that a strategy for tackling AIDS is to attempt to limit its spread in those groups with the highest contact rates where an epidemic is likely to take hold. Similarly, in a population with varying degrees of susceptibility to a microparasitic pathogen, the equivalent strategy would be to attempt to vaccinate or treat those genotypes known to be most susceptible. Thus heterogeneity not only may help to protect a population from the spread of infection but also can provide a clear means of protecting against potential epidemics.

This work was funded by a grant from the Biotechnology and Biological Sciences Research Council Mathematics and Modelling of Agricultural and Food Systems initiative.

LITERATURE CITED

Adler, F. R., 1992 The effects of averaging on the basic reproduction ratio. Math. Biosci. **111:** 89–98.

Axford, R. F. E., S. C. Bishop, F. W. Nicholas and J. B. Owen, 2000 *Breeding for Disease Resistance in Farm Animals*, Ed. 2. CABI Publishing, Oxon, UK.

Bishop, S. C., and K. Mackenzie, 2003 Genetic management strategies for controlling infectious disease in livestock populations. Genet. Sel. Evol. **35** (Suppl.): 3–17.

Bishop, S. C., J. Chesnais and M. J. Stear, 2002 Breeding for disease resistance: issues and opportunities. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production, Communication 13-01.

Dushoff, J., and S. Levin, 1994 The effects of population heterogeneity on disease invasion. Math. Biosci. **128:** 25–40.

Hethcote, H. W., 1977 An immunization model for a heterogeneous population. Theor. Popul. Biol. **14:** 338–349.

Mackenzie, K., and S. C. Bishop, 2001 Developing stochastic epidemiological models to quantify the dynamics of infectious diseases in livestock. J. Anim. Sci. **79:** 2047–2056.

May, R. M., 1987 Nonlinearities and complex behavior in simple ecological and epidemiologic models. Ann. NY Acad. Sci. **504:** 1–15.

May, R. M., and R. M. Anderson, 1989 The transmission dynamics of human immunodeficiency virus (HIV), pp. 265–311 in *Applied Mathematical Ecology*, edited by S. A. Levin, T. G. Hallam and L. J. Gross. Springer-Verlag, Berlin.

Office International des Epizooties, 1998 Genetic resistance to animal diseases. Rev. Sci. Tech. **17:** 1–391.

Stear, M. J., S. C. Bishop, M. Doligalaska, J. L. Duncan, P. H. Holmes *et al.*, 1995 Regulation of egg production, worm burden, worm length and worm fecundity by host responses in sheep infected with *Ostertagia circumcincta*. Parasite Immunol. **17:** 643–652.

Communicating editor: S. W. Schaeffer

APPENDIX A

Hethcote (1977) gives the following analytical result for the spread of an infectious disease in a population consisting of n subgroups, each of which is homogeneous with respect to resistance to the infection. The subgroup sizes are N_i $(I = 1, \ldots, n)$ and the proportions of infected, susceptible, and recovered individuals in each subgroup at time t are $I_i(t)$, $S_i(t)$, and $R_i(t)$. The recovery rates are γ_i. The contact rates between the ith and jth subgroups are λ_{ij} $(i, j = 1, \ldots, n)$.

The proportions of the subgroups that have been infected at infinity are given by the n simultaneous equations:

$$I_i(\infty) = 1 - S_i(0)\exp\left[-\sum_{j=1}^{n} \frac{\lambda_{ij}N_j}{N_i}\left(\frac{1 - S_j(\infty) - R_j(0)}{\gamma_j}\right)\right], \quad i = 1, \ldots, n.$$

However, there is no analytical solution.

APPENDIX B

Derivation of approximate variance for the total proportion infected (I) as a function of the variance of the reproductive rate (R).

We have the asymptotic result for a SIR epidemic:

$$I = 1 - e^{-IR} \quad \text{for } R \geq 1 \text{ and } 0 \leq I \leq 1.$$

This can be rearranged as a function of R:

$$R = -\log_e(1 - I)/I.$$

Using a Taylor series expansion for the rhs about I_0,

$$\begin{aligned} -\log_e(1 - I)/I &\approx -\log_e(1 - I_0)/I_0 + (I - I_0)[I_0^{-1}(1 - I_0)^{-1} + I_0^{-2}\log_e(1 - I_0)] \\ &= -I[I_0^{-1}(1 - I_0)^{-1} + I_0^{-2}\log_e(1 - I_0)] - (1 - I_0)^{-1} - 2I_0^{-1}\log_e(1 - I_0). \end{aligned}$$

Putting $R = -\log_e(1 - I)/I$ and $J_0 = 1 - I_0$ and simplifying,

$$I \approx -RJ_0I_0^2/[I_0 + J_0 \log_e J_0]^2 - f(I_0, J_0),$$

where $f(I_0, J_0)$ is a function solely of I_0 and J_0 and is a constant for given I_0 and J_0. Thus,

$$\text{Var}(I) \approx \text{var}(R)J_0^2I_0^4/[I_0 + J_0 \log J_0]^2.$$