

# Estimation of Rates of Recombination and Migration in Populations of Plant Pathogens

J. K. M. Brown

Cereals Research Department, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, United Kingdom.  
Accepted for publication 3 January 2000.

## ABSTRACT

Brown, J. K. M. 2000. Estimation of rates of recombination and migration in populations of plant pathogens. *Phytopathology* 90:320-323.

A method of estimating the frequency of recombination in field populations of ascomycete fungi is presented. This is a development of a technique described by J. Zhan, C. C. Mundt, and B. A. McDonald, in which field plots were inoculated with isolates of *Mycosphaerella graminicola* of known genotypes and then, later, the proportions of the fungal populations in the plots that originated as recombinant progeny of matings between the inoculated isolates or as immigrants from the local population were estimated. It is argued that J. Zhan, C. C. Mundt, and B. A. McDonald's method is inappropriate in terms of its approach to statistical inference

and that the lack of confidence intervals for rates of recombination ( $r$ ) or immigration ( $m$ ) is a further weakness. An improvement to their method would involve the use of inoculated isolates that can be clearly distinguished from the local population, for instance, by the use of DNA markers. There should also be sufficient markers for there to be a negligible probability that progeny of matings between inoculated isolates will have the same genotype as either of their parents. With this experimental design, estimates of  $r$  and  $m$  can be found by a standard maximum likelihood method, while confidence intervals for the parameters can be estimated by a simple bootstrap procedure. The method appears to be about an order of magnitude more efficient than are the existing ways of estimating the frequency of recombination in ascomycete fungi.

Sexual reproduction plays a significant role in evolution, because it causes genes to become recombined, thereby generating new genotypes on which natural selection may act. For many fungal pathogens, sex is believed to be important in providing the genetic variation that enables them to adapt to host plants and to diverse environments, while sexual spores are important as sources of inoculum of many plant diseases (3,6,11). In diploid and dikaryotic organisms, such as oomycete or basidiomycete fungi, the extent to which genotype frequencies are in Hardy-Weinberg equilibrium within loci can be used to measure the importance of recombination. However, it has proven to be much harder to find ways of quantifying rates of recombination in haploid species such as ascomycete fungi.

A recent paper by J. Zhan, C. C. Mundt, and B. A. McDonald (13) throws fresh light on this difficult problem by suggesting a new way of quantifying the frequency of sex in ascomycete plant pathogens. The basis of their method was that field trial plots were inoculated with several isolates of a fungus with genotypes defined by a set of DNA markers (inoculants). Later, the numbers of isolates in those plots with novel genotypes that could have arisen either by recombination between the inoculants (recombinants) or by immigration from the local population (immigrants) were compared. This method may be especially useful because, in principle, it could be applied to any ascomycete fungus. The fungus studied by Zhan et al. (13), *Mycosphaerella graminicola*, the causal agent of Septoria tritici leaf blotch of wheat, has a sexual stage that occurs throughout the year (9,10), while its population structure suggests that sex occurs at a reasonably high frequency (7). However, the relative importance of ascospores, produced by sexual reproduction, and pycnidiospores, produced asexually, is not known.

Despite the potential value of the method of Zhan et al. (13), the statistical methods they used to quantify rates of recombination and immigration had several significant deficiencies. As a result, their estimates of those rates cannot be considered reliable. Al-

though one purpose of this letter is to explain why Zhan et al.'s method was unsound, I also explain how the principles of the experimental design used by Zhan et al. can be adapted, so as to develop a simple method of estimating population genetic parameters of ascomycete fungi. In particular, a method of estimating recombination rates is derived that is an order of magnitude more efficient than are previously published methods.

I consider five aspects of Zhan et al.'s (13) approach: (i) their method of estimating rates of recombination ( $r$ ) and immigration ( $m$ ); (ii) the need for information about the isolates' mating types; (iii) genotypes that could be used to maximize the power of the experiments; (iv) the calculation of confidence intervals for  $r$  and  $m$ ; and (v) the number of isolates that need to be sampled to obtain reliable estimates of those rates.

**Method of estimation.** Zhan et al. (13) tested whether an isolate with a novel genotype, not one carried by an inoculant, originated by recombination or by immigration by estimating what the frequencies of that genotype would be in populations generated by each of those two processes, recombination or immigration, acting alone. The observed frequency of allele  $i$  at locus  $j$  in the population outside the trial plot, i.e., in the population of potential immigrants, is  $p_{ijM}$ . The observed frequency of that allele in the plot itself, i.e., in the population that gives rise to recombinants, is  $p_{ijR}$ . These observed allele frequencies are equal to the expectations of the actual frequencies in the populations in question. The expected frequencies of a genotype  $G$  defined by all loci studied in the potential immigrant and recombinant populations are, therefore,

$$M_G = \prod_{j=1}^{\lambda} p_{ijM} \text{ and } R_G = \prod_{j=1}^{\lambda} p_{ijR} \quad (1)$$

respectively, in which the alleles  $i$  are those present in genotype  $G$ . Equation 1 assumes that all alleles are in linkage equilibrium both inside the plot and in the local population. (This is an important assumption that, in general, is not necessarily true in plant pathogen populations [11].) Zhan et al. (13) assumed that an isolate with a novel genotype originated by immigration if  $M_G/(M_G + R_G) > 10$  and by recombination if  $R_G/(M_G + R_G) > 10$ . Genotypes that had intermediate values of  $M_G$  and  $R_G$ , such that neither inequality was true, were

Corresponding author: J. K. M. Brown; E-mail address: James.Brown@bbsrc.ac.uk

Publication no. P-2000-0203-01O

© 2000 The American Phytopathological Society

assigned to an “uncertain” category and were not used to estimate the frequencies of recombination or immigration. (Note that I have used a different notation from that of Zhan et al., for the sake of clarity.)

This procedure is not appropriate for two reasons. First, it discards a large part of the data, from genotypes in the “uncertain” class (including the majority of isolates in one of Zhan et al.’s samples; Fig. 1B in literature citation 13). This is unnecessarily wasteful because information from all isolates sampled should be used to estimate  $m$  and  $r$ . Below, I describe how this can be done. Furthermore, it makes the implicit assumption that the ratio of immigrants to recombinants among the isolates with genotypes placed in the “uncertain” class was equal to the ratio of the numbers of isolates in the two extreme classes, i.e., with high probabilities of being immigrants or recombinants. There is no reason at all why this should be so. Consequently, estimates of  $m$  and  $r$  calculated using only the most extreme data cannot be considered reliable.

Second, Zhan et al.’s approach is unconventional in terms of statistical inference.  $M_G$  is the expectation of what the frequency of genotype  $G$  would be if the entire *M. graminicola* population in the trial plot at the second time of sampling were the progeny of immigrant spores (i.e.,  $m = 1$ ). Similarly,  $R_G$  is the expected frequency of  $G$  if that population were wholly made up of immigrant spores and their descendants (i.e.,  $r = 1$ ). There is no theoretical justification for using genotype frequencies in two hypothetical, extreme situations as the basis for a method of estimating population genetic parameters. Such an approach treats the two probabilities,  $M_G$  and  $R_G$ , as if they were fixed biological properties of an isolate. However, had additional loci been studied, or a completely different set of loci, the ratio of  $M_G$  to  $R_G$  would be different, thereby generating a different classification of isolates as immigrants, recombinants, or “uncertain.”

A more conventional approach to estimation involves finding the values of  $m$  and  $r$  that maximize the total likelihood, summed across all isolates sampled. The likelihood of a hypothesis  $H$ , given data  $D$ , is the probability that  $D$  would be observed if  $H$  were true. Here, we wish to find values of  $m$  and  $r$  that maximize the probability of observing the various allele and genotype frequencies that were actually detected.

Consider an experiment similar to that of Zhan et al., in which a plot is inoculated with several isolates and from which samples are subsequently collected on two occasions. In addition, samples are taken from an uninoculated plot in order to estimate allele frequencies in the local pathogen population. The frequency of genotype  $G$  is  $V_G$  in the first sample from the inoculated plot and  $M_G$  in the sample from the uninoculated plot (eq. 1). In the second sample from the inoculated plot, an isolate with genotype  $G$  has three possible origins: as an asexual descendant of the population in the inoculated plot at the first time of sampling, as an immigrant, or as a recombinant progeny of the population in the inoculated plot. At the second sampling time, the proportions of individuals in the plot that have those three origins are  $1 - m - r$ ,  $m$ , and  $r$ , respectively. The frequency of  $G$  at the second sampling time is, therefore,

$$W_G = (1 - m - r)V_G + mM_G + rR_G \quad (2)$$

in which  $R_G$  is calculated by equation 1, using allele frequencies  $p_{ijR}$  calculated from the genotype frequencies  $V_G$  in the first sample from the inoculated plot. The numbers of isolates with genotype  $G$  in the first and second samples from the inoculated plot and in the sample from the uninoculated plot are  $n_{G1}$ ,  $n_{G2}$ , and  $n_{Gu}$ , respectively. The probability of obtaining  $g$  individuals with a genotype  $G$  in a sample, size  $n$ , from a population in which the frequency of  $G$  is  $P_G$ , is the binomial probability  $\binom{n}{g} P_G^g$ , so that the joint probability of obtaining the genotype frequencies detected in the three samples is the product of the binomial probabilities for that genotype in the three samples, i.e.,

$$f_G = k_G V_G^{n_{G1}} [(1 - m - r)V_G + mM_G + rR_G]^{n_{G2}} M_G^{n_{Gu}} \quad (3)$$

in which  $k_G$  is a constant, multinomial coefficient. Likelihoods are almost always used in the form of their logarithms, to make the calculations easier. The probabilities  $f_G$  for the various genotypes must be multiplied, so the total log-likelihood, summed across genotypes, is

$$L_1 = \sum_G \log f_G \quad (4)$$

$$= \sum_G \left\{ n_{G1} \log V_G + n_{G2} \log [(1 - m - r)V_G + mM_G + rR_G] + N_{Gu} \log M_G \right\}$$

(ignoring the constants  $k_G$  that do not affect the solution of eq. 5). The maximum likelihood estimates (MLEs) of the parameters are those at which  $L_1$  has its maximum value, which is found by setting

$$\frac{\partial L_1}{\partial m} = \frac{\partial L_1}{\partial r} = \frac{\partial L_1}{\partial V_G} = \frac{\partial L_1}{\partial p_{ijM}} = 0 \quad (5)$$

and solving for  $m$ ,  $r$ ,  $V_G$ , and  $p_{ijM}$ .

Note that all second differentials of  $L_1$  are negative, so the curve of  $L_1$  is concave with respect to all parameters.  $L_1$ , therefore, has a single maximum, and equation 5 has a single, unique solution.

However, equation 5 involves a very large number of parameters. The two of genuine interest are  $m$  and  $r$ . The other parameters to be estimated are all possible genotype frequencies  $V_G$  at the first sample time (constrained to a sum of 1) that are used to calculate the allele frequencies  $p_{ijR}$  needed to estimate  $R_G$ , and all allele frequencies  $p_{ijM}$  in the uninoculated plot (again, summing to 1 for each locus) from which  $M_G$  are estimated. Another complication is that the differential of  $L_1$  with respect to any one of the parameters also involves all the other parameters, because the differential of the middle part of  $L_1$  with respect to any of the parameters involves a term in  $[(1 - m - r)V_G + mM_G + rR_G]^{-1}$ . In principle, therefore, it is not possible to estimate any of the parameters separately.

To describe such an estimation problem as formidable is perhaps an understatement. Moreover, it is not possible to do this using the data reported by Zhan et al. (13), because none of the genotype frequencies  $V_G$  are given, while some allele frequencies  $p_{ijM}$  are not provided. (A further difficulty in estimating  $m$  and  $r$  from Zhan et al.’s data, in relation to information about mating types, is discussed below.) Computer programs for maximizing functions such as equation 4 are available in libraries of mathematical routines, such as the NAG Fortran Library (NAG Ltd., Oxford, England), and in some statistical packages, such as Genstat 5 (NAG Ltd.). A brief introduction to numerical analysis of statistical models involving numerous parameters may be found in section 8.7 of literature citation 12.

In summary, therefore, the problem posed by Zhan et al. (13) is rather more complicated than it might appear at first sight. Below, I show how the genotypes of inoculated isolates used in experiments such as these can be chosen so as to simplify the calculations considerably.

**Mating types.** The estimates of  $m$  and  $r$ , found from equation 5, are not accurate if the mating types of the inoculated isolates are unknown. Zhan et al. (13) calculated the expected frequencies of recombinant genotypes,  $R_G$  (eq. 1), by assuming that there was free reassortment of alleles of all genotypes in the trial plot. However, only half of all pairs of individuals of *M. graminicola*, at most, those with opposite mating types, can actually mate and form recombinant progeny, because this fungus has a bipolar mating system and is probably heterothallic (10). Hence, about half the genotypes that might be generated by recombination of the inoculants’ genotypes cannot, in fact, be produced by sexual reproduction. If such genotypes are found in a plot, they must have originated as immigrants. A lack of information about mating types therefore means that genotypes which must be immigrants, so that  $R_G = 0$  in equation 4, are, in fact, estimated to have been produced partly by immigration but also partly by recombination (i.e.,  $R_G > 0$ ). This problem is most severe when the number of inoculant genotypes is small, so the variance in proportions of the two mating types among the inoculants is high. When 10 inoculants are used, as by Zhan et al. (13), the 95% confidence interval for the numbers of the two mating types includes such an extreme distribution as two of one type and eight of the other.

If mating types are not identified, therefore, the estimate of  $r$  will be biased upwards, possibly by as much as a factor of 2, while the estimate of  $m$  will consequently be biased downwards. This is partly a result of not being able to distinguish immigrants and recombinants completely. The lack of information about mating types is a further reason why I have not attempted to reanalyze the data presented by Zhan et al. (13) using the maximum likelihood methods described above.

**The choice of genotypes of inoculated isolates.** Although analysis of experiments of the kind described by Zhan et al. (13), using equations 4 and 5, is undoubtedly complex, it would be considerably simpler if one could distinguish genotypes as those of inoculants, recombinants, and immigrants with certainty or at least with a very high degree of confidence. This could be achieved by ensuring that all isolates used as inoculants carry markers that are not present, or at least have extremely low frequencies, in the local population of the fungus at the trial site or that the inoculants all lack a marker that is more-or-less fixed in the local population. (Ideally, both inoculants and potential immigrants should have markers that allow them to be positively identified.)

Furthermore, all the inoculants should differ from one another by enough markers so that sexual crosses between inoculants give rise to progeny with genotypes indistinguishable from those of the parents at a negligibly low rate. If, as I suggest below, samples collected at each time should consist of at least 500 individuals, the inoculants should all differ from one another by at least  $\log_2 500 \approx 9$  markers (if each marker locus has two alleles) in order to be fairly certain that all progeny of matings between inoculants will be distinguishable from both parents. Nowadays, it is relatively straightforward to generate a large number of markers, from which a suitable subset can be chosen, by polymerase chain reaction (PCR)-based methods such as amplified fragment length polymorphism, simple-sequence repeat (microsatellite) polymorphism, or single-strand conformation polymorphism. Note that it is now not necessary to know the mating types of the inoculants, since the procedure I describe below relies on knowing simply whether an isolate has a recombinant, immigrant, or inoculant genotype.

Consider an experiment to estimate how much immigration and recombination contribute to a pathogen population in a plot over a certain period of time. Two samples are collected, at different times, and in each case the experimenter determines, by analysis of genetic markers, how many isolates were produced (i) by clonal reproduction of the inoculants, having genotypes identical to those of the inoculants; (ii) by sexual reproduction between inoculants, having all the markers that distinguish inoculants from the local, natural population, but with genotypes different from those of the inoculants; and (iii) by immigration from the surrounding population, having markers that distinguish the local population from the inoculants (this category includes progeny of matings between immigrants and inoculants).

The expected frequencies and observed numbers of the three types at the two times of sampling are given in Table 1. The log-likelihood is

$$L_2 = n_{c1} \log(1 - p_m - p_r) + n_{m1} \log p_m + n_{r1} \log p_r + \quad (6)$$

$$n_{c2} \log(1 - p_m - p_r - m - r) + n_{m2} \log(p_m + m) + n_{r2} \log(p_r + r)$$

(again, ignoring a constant, as in eq. 4). With only four parameters, equation 6 is much simpler than equation 4. This is essentially because one now need only consider whether genotypes are those of inoculants, immigrants, or recombinants, while it is not necessary to estimate the frequencies of individual alleles or genotypes.

The frequencies of immigration ( $m$ ) and recombination ( $r$ ) estimated in this way are the fractional contributions of those two processes to the population in the trial plot between the two times of sampling (as, indeed, were the parameters estimated by Zhan et al. [13]). They are not the same as the rates of immigration and recombination over a single generation, which may be quite different for a polycyclic pathogen.

The first derivatives of  $L_2$  with respect to  $p_r$  and  $r$  are

$$\frac{\partial L_2}{\partial p_r} = \frac{-n_{c1}}{1 - p_m - p_r} + \frac{n_{r1}}{p_r} - \frac{n_{c2}}{1 - p_m - p_r - m - r} + \frac{n_{r2}}{p_r + r} \quad (7)$$

and

$$\frac{\partial L_2}{\partial r} = \frac{-n_{c2}}{1 - p_m - p_r - m - r} + \frac{n_{r2}}{p_r + r}$$

$\partial L_2 / \partial p_m$  and  $\partial L_2 / \partial m$  have forms similar to those of  $\partial L_2 / \partial p_r$  and  $\partial L_2 / \partial r$ , respectively. The MLEs of  $p_m$ ,  $p_r$ ,  $m$ , and  $r$  ( $\hat{p}_m$ ,  $\hat{p}_r$ ,  $\hat{m}$ , and  $\hat{r}$ , respectively) are found by solving  $\partial L_2 / \partial p_m = \partial L_2 / \partial p_r = \partial L_2 / \partial m = \partial L_2 / \partial r = 0$ , as in equation 5. It follows, as one would expect intuitively, that

$$\hat{p}_m = \frac{n_{m1}}{n_{T1}}, \hat{p}_r = \frac{n_{r1}}{n_{T1}}, \hat{m} = \frac{n_{m2}}{n_{T2}} - \frac{n_{m1}}{n_{T1}}, \text{ and } \hat{r} = \frac{n_{r2}}{n_{T2}} - \frac{n_{r1}}{n_{T1}} \quad (8)$$

This basic approach can be adapted to more complex situations. For instance, allele frequencies at three of the loci studied by Zhan et al. (13) changed significantly during the course of the field trial. Coefficients of natural selection acting on individual alleles or genotypes could be added to the model in Table 1 and their MLEs found along with those of the migration and recombination rates.

Note that if one simply wishes to know how much of the pathogen population in a plot, at a single time, originates from the three sources, one can use equations 6 to 8, setting  $p_r$ ,  $p_m$ , and  $n_{T1}$  all to zero. From equation 8,  $\hat{p}_m = n_{m2} / n_{T2}$  and  $\hat{p}_r = n_{r2} / n_{T2}$ .

**Confidence intervals.** Zhan et al. (13) did not calculate confidence intervals for their estimates of recombination and migration rates. This limits the value of their conclusions considerably because, without a confidence interval, one cannot know how reliable an estimate of a parameter really is. (Zhan et al. [13] used a simulation model to test the reliability of their classification of genotypes as immigrant, recombinant, or uncertain, but this does not address the issue of constructing confidence intervals for  $m$  and  $r$ . Moreover, since the method of classifying genotypes is itself unsound, the reliability of the classification is not really a significant issue.)

The construction of confidence intervals has been a flourishing area of statistical research in recent years, and reliable confidence intervals can now be calculated for almost any statistic of practical importance in genetics by bootstrapping or by related, numerical methods. The theory of the bootstrap and its application to calculating confidence intervals is well described in a recent book written for nonstatisticians (8).

The variances of  $\hat{m}$  and  $\hat{r}$  in equation 8 and their covariance can be found analytically by a standard likelihood method. This is done using the second derivatives of  $L_2$ , evaluated at the MLEs of the four parameters, so that  $[\text{Var}(\hat{r})]^{-1} = -\partial^2 L_2 / \partial r^2 = n_{T2}^2 (n_{r2}^{-1} + n_{c2}^{-1})$  and  $[\text{Var}(\hat{m})]^{-1} = n_{T2}^2 (n_{m2}^{-1} + n_{c2}^{-1})$ . However, there are non-zero covariances between all four parameters of equation 6, e.g.,  $[\text{Cov}(\hat{m}, \hat{r})]^{-1} = -\partial^2 L_2 / \partial m \partial r = n_{T2}^2 n_{c2}^{-1}$ , so one cannot simply use the variance of one statistic alone to calculate its confidence interval.

Nevertheless, it is quite simple to construct a confidence interval for each individual parameter using the bootstrap. To do this, (i) use computer randomization to sample, with replacement, the data for  $n_{T1}$  individuals from the original set of  $n_{T1}$  individuals collected in the first sample; (ii) do likewise, sampling  $n_{T2}$  individuals, with replace-

TABLE 1. Model of recombination and immigration of a plant pathogen in a field plot between two times of sampling

Origin	First sample		Second sample	
	Frequency	Number sampled	Frequency	Number sampled
Clonal progeny of inoculants	$1 - p_m - p_r$	$n_{c1}$	$1 - p_m - p_r - m - r$	$n_{c2}$
Immigrants	$p_m$	$n_{m1}$	$p_m + m$	$n_{m2}$
Sexual progeny of inoculants	$p_r$	$n_{r1}$	$p_r + r$	$n_{r2}$
Total	1	$n_{T1}$	1	$n_{T2}$

ment, from the second sample; (iii) calculate estimates,  $m^*$  and  $r^*$ , of  $m$  and  $r$ , for this bootstrap sample; (iv) do steps i to iii many ( $n_B$ ) times (10,000 bootstrap samples were used in the calculations in Table 2); and (v) the standard error (SE) of  $r$  is estimated by the standard deviation (SD) of  $r^*$ . There are several ways of calculating the confidence interval using the bootstrap (8), but the simplest and often most accurate is to define the  $1 - 2\alpha$  confidence interval as the range of  $r^*$  that excludes the lowest  $\alpha n_B$  and the highest  $\alpha n_B$  values of  $r^*$ .

The effect of sample sizes on confidence intervals for  $r$  and  $m$ , as in the model of Table 1, is shown in Table 2. The parametric SE of  $r$ , calculated as  $(-\partial^2 L_2 / \partial r^2)^{-0.5}$ , depends only on the size of the second sample,  $n_{T2}$ . The bootstrap SE, calculated as the SD of  $r^*$ , however, depends on the size of both samples, although it is affected more by  $n_{T2}$  than by  $n_{T1}$ . The bootstrap SE depends on  $n_{T1}$  because the accuracy with which  $p_r$  and  $p_m$  are estimated affects the precision with which  $r$  and  $m$  can be estimated; the parametric SE ignores this source of statistical error.

If it is not possible to find markers that discriminate inoculants, recombinants, and immigrants completely, the SEs of parameter estimates calculated using equation 4 may be found by bootstrapping (provided one is not daunted by the problem of estimating  $m$  and  $r$  from eq. 5). In Zhan et al.'s situation, bootstrap samples should be taken from the data for both samples from the inoculated plot and from the sample from the uninoculated plot, and  $r^*$  and  $m^*$  calculated, by equation 5, for each set of bootstrap samples.  $SE(r)$  should not be calculated by the standard, parametric formula,  $(-\partial^2 L_2 / \partial r^2)^{-0.5}$ , because this ignores the effect on the estimated value of  $r$  of the many other parameters, which are themselves subject to statistical error. This exemplifies the power of bootstrapping to find reliable confidence intervals for parameters even of very complex population genetic models.

**Sample sizes.** From Table 2, it can be seen that if one wishes to estimate  $r$  or  $m$  to within  $\pm 5\%$ , given true values of  $r$  and  $m$  of 0.15, and if it is possible to discriminate inoculants, recombinants, and immigrants completely, two samples of about 500 individuals need to be collected. Zhan et al. (13) used a sample of just 45 individuals from uninoculated plots to estimate allele frequencies in the local population, as the basis for calculating migration rates, and three samples of 582, 183, and 600 individuals, in the early, middle, and late parts of the season, to estimate genotype frequencies in the inoculated plots. The sample from the uninoculated plot is clearly so small that, even if an appropriate method of estimation had been used (e.g., eq. 4), the SEs of estimated allele frequencies in the

local population would be very high, thereby inflating the SEs of  $r$  and  $m$ . The midseason sample from the inoculated plot is probably also too small to allow  $r$  and  $m$  to be estimated accurately.

On the positive side, the principle underlying Zhan et al.'s method, that of using genetic markers to discriminate genotypes arising from different sources, has the potential to provide a considerably more powerful way of estimating recombination rates of ascomycetes than has been available hitherto. Methods of estimating recombination rates from changes in frequencies of mating types (1,4) or in linkage disequilibrium (5) have been described, but they require sample sizes of several thousand individuals in order to get reasonably accurate estimates (2,4). The modified version of Zhan et al.'s method described in this paper is a considerable improvement, as it only requires samples sizes of several hundred (Table 2).

**General conclusion.** The population genetics of plant pathogens, as much as those of other organisms, can now be studied in great detail using molecular markers. Powerful methods of detecting genetic variation, using PCR-based techniques, can be used to generate a huge number of genetic markers and to characterize variation in those markers very accurately. This is a huge advance on the restriction-fragment length polymorphism and allozyme systems that, in turn, were great improvements over morphological markers. The precision with which population genetics can now be studied challenges biologists to quantify the parameters that describe the structure and evolution of populations, instead of merely describing population structure in general terms. To do this, we need to apply the range of statistical techniques that are now available as effectively as we use molecular genetic markers.

## ACKNOWLEDGMENTS

This work was supported by the Ministry of Agriculture, Fisheries and Food.

## LITERATURE CITED

- Brändle, U. E., Haemmerli, U. A., McDermott, J. M., and Wolfe, M. S. 1997. Interpreting population genetic data with the help of genetic linkage maps. Pages 157-171 in: *The Gene-for-Gene Relationship in Plant-Parasite Interactions*. I. R. Crute, E. B. Holub, and J. J. Burdon, eds. CAB International, Wallingford, United Kingdom.
- Brown, J. K. M. 1994. Chance and selection in the evolution of barley mildew. *Trends Microbiol.* 2:470-475.
- Brown, J. K. M. 1999. The evolution of sex and recombination in fungi. Pages 73-95 in: *Structure and Dynamics of Fungal Populations*. J. Worrall, ed. Kluwer, Dordrecht, the Netherlands.
- Brown, J. K. M., Foster, E. M., and O'Hara, R. B. 1997. Adaptation of powdery mildew populations to cereal varieties in relation to durable and non-durable resistance. Pages 119-138 in: *The Gene-for-Gene Relationship in Plant-Parasite Interactions*. I. R. Crute, E. B. Holub, and J. J. Burdon, eds. CAB International, Wallingford, United Kingdom.
- Brown, J. K. M., and Wolfe, M. S. 1990. Structure and evolution of a population of *Erysiphe graminis* f.sp. *hordei*. *Plant Pathol.* 39:376-390.
- Chamberlain, M., and Ingram, D. S. 1997. The balance and interplay between asexual and sexual reproduction in fungi. *Adv. Bot. Res. Incorporat. Adv. Plant Pathol.* 24:71-87.
- Chen, R. S., and McDonald, B. A. 1996. Sexual reproduction plays a major role in the genetic structure of populations of the fungus *Mycosphaerella graminicola*. *Genetics* 142:1119-1127.
- Efron, B., and Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Hunter, T., Coker, R. R., and Royle, D. J. 1999. The teleomorph stage, *Mycosphaerella graminicola*, in epidemics of Septoria tritici leaf blotch in UK winter wheat. *Plant Pathol.* 48:51-57.
- Kema, G. H. J., Verstappen, E. C. P., Todorova, M., and Waalwijk, C. 1996. Successful crosses and molecular tetrad and progeny analysis demonstrate heterothallism in *Mycosphaerella graminicola*. *Curr. Genet.* 30:252-258.
- Milgroom, M. G. 1996. Recombination and the multilocus structure of fungal populations. *Annu. Rev. Phytopathol.* 34:457-477.
- Rothamsted Experimental Station, Genstat 5 Committee. 1993. *Genstat 5 Release 3 Reference Manual*. Oxford University Press, Oxford, U.K.
- Zhan, J., Mundt, C. C., and McDonald, B. A. 1998. Measuring immigration and sexual reproduction in field populations of *Mycosphaerella graminicola*. *Phytopathology* 88:1330-1337.

TABLE 2. Standard errors and 95% confidence intervals, estimated by bootstrapping, for rates of recombination ( $r$ ) and immigration ( $m$ ) of a plant pathogen between two times of sampling<sup>a</sup>

$n_{T1}^b$	$n_{T2}^b$	SE( $r$ ) <sup>c</sup>	Bootstrap SE <sup>d</sup>		95% confidence interval	
			$r$	$m$	$r$	$m$
50	500	0.0183	0.0468	0.0465	0.054-0.234	0.056-0.236
100	500	0.0183	0.0357	0.0357	0.076-0.216	0.078-0.218
200	500	0.0183	0.0292	0.0284	0.092-0.205	0.093-0.205
500	500	0.0183	0.0236	0.0235	0.104-0.196	0.104-0.196
1,000	500	0.0183	0.0218	0.0215	0.108-0.192	0.108-0.192
2,000	500	0.0183	0.0204	0.0202	0.111-0.190	0.110-0.190
500	50	0.0573	0.0623	0.0617	0.024-0.266	0.024-0.264
500	100	0.0408	0.0457	0.0453	0.062-0.240	0.064-0.240
500	200	0.0289	0.0337	0.0335	0.084-0.216	0.086-0.215
500	500	0.0183	0.0238	0.0240	0.104-0.196	0.104-0.198
500	1,000	0.0129	0.0192	0.0190	0.113-0.188	0.113-0.186
500	2,000	0.0091	0.0166	0.0167	0.118-0.182	0.117-0.183

<sup>a</sup> The model used is that of Table 1. In the first sample, 10% of the isolates have recombinant genotypes and 10% have immigrant genotypes. In the second sample, 25% are recombinants and 25% are immigrants. The true values of  $r$  and  $m$  are, therefore, both 0.15.

<sup>b</sup>  $n_{T1}$ ,  $n_{T2}$ : total number of isolates in first and second samples, respectively.

<sup>c</sup> Standard error (SE) of  $r$ , calculated as  $(-\partial^2 L_2 / \partial r^2)^{-0.5}$ , = SE( $m$ ) in this example ( $L_2$  defined in eq. 6).

<sup>d</sup> 10,000 bootstrap samples used in each case. Calculations were done using Genstat for Windows, 4th edition.