6 Analysis of virulence diversity in populations of plant pathogens

J.V. GROTH and A.P. ROELFS

Department of Plant Pathology, University of Minnesota and United States Department of Agriculture Cereal Rust Laboratory, respectively, St Paul, MN 55108, USA

Introduction

The quantitative measurement of genetic diversity in populations of plant pathogenic fungi is essential in our goal of understanding the dynamics of virulence shifts as they influence the durability of resistance. Host-specific (intraspecific) virulence genes have been the most abundant, and hence the most commonly employed, markers in attempts to describe either overall virulence diversity (Roelfs & Groth, 1980; Lebeda, 1982) or some aspect of the distribution of virulence genes or races in populations (Browder & Eversmeyer, 1977; Simons *et al.*, 1979; Groth & Roelfs, 1982; Wolfe, this volume; Crute, this volume). Other kinds of markers, particularly isozymes, must be and are being employed (Burdon *et al.*, 1982) in obtaining a more complete and more objective picture of pathogen diversity. However, the direct importance and increasing abundance of virulence markers will ensure that they will continue to serve a key role in describing plant pathogen populations. Markers of all kinds will be needed to fully describe genetic structure of populations of plant pathogens and subsequently ecological and evolutionary events of importance to economic plant pathology.

The use of ecological diversity indexes at an intraspecific level has only recently been advocated and used as a means of combining into one easily understood variable two important aspects of genetic diversity: the number of races (race abundance) and the homogeneity/heterogeneity of the frequencies of different races (race frequency evenness) (Lebeda, 1982; Groth & Roelfs, 1982). A number of different indexes are available for use (Whittaker, 1972; Lebeda, 1982). The Shannon index is perhaps the most popular and simple to use, either in its pure form or, for finite populations, as the Brillouin's Index (Pielou, 1977). Since we are not describing sampled populations in most of the following development, we use the Shannon index:

$$D = \sum_i p_i \log_e (p_i) \qquad [1]$$

Where D = the Shannon index, p_i = the frequency of the ith phenotype or race. For simplicity $\log_e$ = ln. The use of this index as a means of quantitatively linking various distributional properties of virulence genes will become clear in the description to follow.

Wolfe M.S. & Caten C.E. (1987) *Populations of Plant Pathogens: their Dynamics and Genetics.* Blackwell Scientific Publications, Oxford.

We have been interested in the means by which differential cultivars of a host species can be used to detect genetic diversity for virulence in plant pathogens. There are several aspects of the distribution of either resistance genes in the host differentials, or of virulence genes in the pathogen, which influence how much diversity a differential host set can reveal in a particular pathogen population. These influences must be quantitatively accounted for in order to present as complete a picture as possible of how differential host lines function. We will attempt to deal with the important aspects in this chapter.

Differential hosts divide the pathogen population into races (distinct virulence phenotypes). Each differential line that is added is capable of doubling the number of races. Because of its logarithmic nature, the Shannon index is linearly related to the number of differential lines when all detected virulences are polymorphic in the pathogen population, and at the same frequency. This linearity is not found for the other common index, the Simpson index of similarity, whose complement or inverse are used as diversity indexes. Because of this linearity and the potential value of the hierarchical use of the Shannon index, whereby amount of difference between phenotypes can be partly accounted for (Pielou, 1977, p. 311), we chose to use it exclusively in this analysis. Briefly, we use the index as a means of quantifying and directly comparing the effects of three different constraints on the capacity of a differential set to detect population diversity in the pathogen. These constraints are non-independence of resistance genes in differential lines, near fixation of virulence genes and associations of virulence genes.

The distribution of resistance genes

Differential host lines fall into two types, 'Single-gene' differentials, and those for which the number of resistance genes (and corresponding virulence genes) is not known. In a few instances the latter kind of differential may unknowingly be of the single-gene type. What is important is the number of *detectable* resistance genes, that is, the number that will be detected because the pathogen population possesses some phenotypes which are avirulent. On the other hand, if a resistance gene is unmatched by virulence, not only will the particular cultivar and gene be of no use in revealing genetic diversity of the pathogen, but also, because of the epistatic masking of their effects, any other genes in the cultivar will be rendered undetectable. Frequencies of virulences in the pathogen population corresponding to the masked resistance will have no effect on this loss of information. At times this loss may be reduced because the masking resistance gives a different, identifiable reaction from that of those which are masked. Usually, however, the masking of one gene by another will be complete. Qualitatively stated, the result is that for a given number of differentials used to measure diversity in a population, the best possible case is where each differential contains one gene only.

Any degree of clustering in distribution of the resistance genes will detract from the capacity of the differential host set to detect diversity. The worst case is an

absurd one where one host line contains all of the resistance genes while all others contain none. In such an instance, if all virulences are independent and at 0.5 frequency, only two races would be detected, and only 0.5^n of the population would appear as a different race, where n is the number of genes for resistance and virulence.

Intermediate cases between the best and worst can be of different types. Taking a simple case, where $n = 4$, the number of resistance genes in each host line can be placed in four classes of decreasing degree of clustering (the number of *functioning* differential lines is in parentheses): (a) all four genes in one line and none in the other three (1 line); (b) three genes in one line and one in one of the other three, or two genes in each of two lines (2 lines); (c) two genes in one line and one in each of the other two lines (3 lines); and (d) one resistance gene in each line (4 lines). Figure 6.1 shows the effects of clustering on the Shannon index. There is an almost linear decrease in the measured diversity as the resistance genes are concentrated in fewer host lines, until only a tenth of the diversity is being detected. Maximum diversity is detected (or minimum information is lost) in the single-gene case (d). If we can generalize from this simple example, we must conclude that single-gene differentials are the most effective at detecting diversity. This is admittedly no great

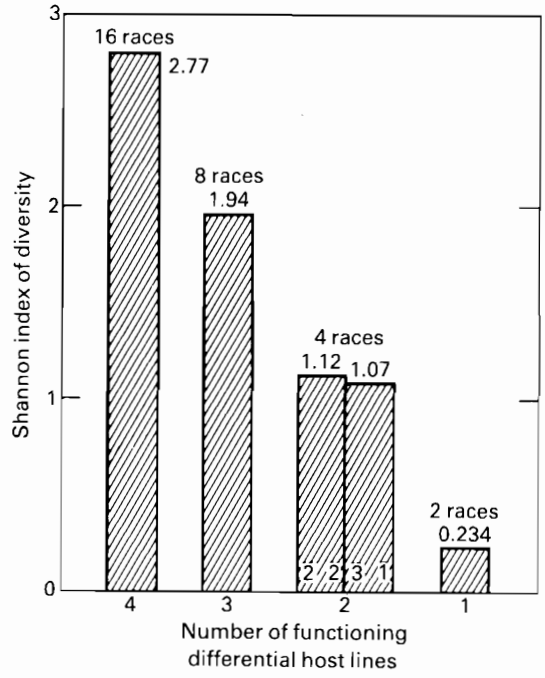

Figure 6.1. Potential diversity for virulence, as measured by the Shannon diversity index, in a pathogen population that is divided using differential host lines containing four resistance genes, where the genes are, from left to right, increasingly non-independent in distribution, being included in 4, 3, 2, and 1 of the lines, respectively. All corresponding virulences are assumed to be independent with a 0.5 frequency.

revelation as a qualitative fact. Nevertheless, it is worthwhile to begin to express how much is being lost as this assumption is relaxed more and more.

Virulence properties – polymorphism and independence

Using the Shannon index of diversity, we will look separately at the two properties of virulence that are responsible for loss of information or loss of efficiency of a set of differential lines in detecting diversity: (1) degree of polymorphism of virulence, and (2) association of virulences. For simplicity in this initial analysis, only the case involving single-gene differentials will be explored.

With respect to the two components of diversity, i.e. race abundance and race-frequency evenness, that are considered by the diversity indexes, maximum diversity (Figure 6.2) occurs in a hypothetical pathogen population that has (a) each virulence at 0.5 frequency and (b) each virulence independent of every other, i.e. no epistasis, allelism, or linkage disequilibrium (including association due to vegetative propagation of a small number of races). It follows then that departures from the above conditions provide the only two ways of decreasing diversity.

The effect of virulence frequency on diversity

The first property, virulence frequency, is the easiest to account for because, unlike virulence association, frequency is a property solely of a single virulence. In the analysis, we must first examine the ideal case for a set of differential lines, in which the effects of adding differential lines are examined cumulatively. Frequency of virulence is most efficiently expressed as departure from 0.5, which we call d. Retaining the independence assumption, the relationship between the Shannon index D and the number of differential lines for values of d decreasing in 0.1 increments is expressed as a set of lines (Figure 6.2). The assumption is that frequency of each virulence is the same in these linear cases. If this is not so, each added differential line increases the value of D differently, and the slope of each is described by:

$$\Delta D_i = -((0.5 + d_i) \ln (0.5 + d_i) + (0.5 - d_i) \ln (0.5 - d_i)) \qquad [2]$$

where ΔD_i = the change in Shannon index by addition of the ith differential, d_i = the departure from frequency of 0.5 of virulence of the ith differential.

This will reduce to $-\ln (0.5)$ if $d_i = 0$.

Equation [2] allows a simple measurement of the influence on diversity of departure from 0.5 of each virulence, so that the information needed to account for this constraint is easily incorporated into the analysis. If the only source of loss in diversity is departure from 0.5 the total diversity is simply the sum of all ΔD_i's.

The effect of virulence associations on diversity

The other effect to consider is that of associations (in coupling or repulsion)

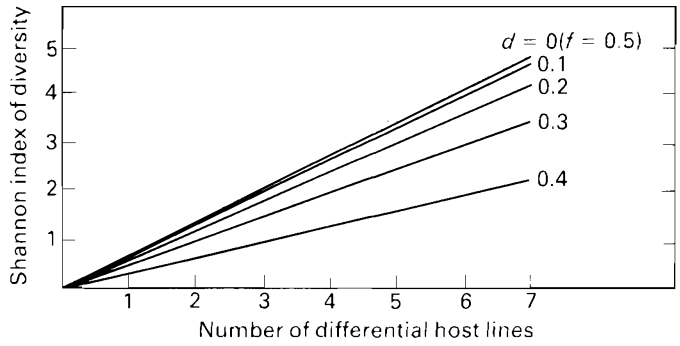

Figure 6.2. The relationship between the number of differential host lines and increase in Shannon diversity index for five virulence frequencies. The departure (d) of the virulence frequency from 0.5 is shown above each line and is the same for each virulence. All virulences are independent.

between virulences, whether they be due to lack of recombination, linkage, or something more obscure. This constraint on variation is not simply expressed algebraically, in fact we will not completely detail it in this treatment. The problem is that each virulence can be associated with some or all of the others in different degrees. At the most complex level, there exist extensive networks or matrices of associations. The magnitude of associations may or may not be mathematically related, depending on the biological basis of the association. For example, the linearity and approximate additivity of distances between genes in linkage groups might provide a mathematical relationship between associations in a few cases, provided that genetic linkage is the basis of association. Likewise, knowledge of the complete phenotypes of prevalent races with respect to pertinent virulences may allow associations to be related in pathogens that are predominantly asexual and possess a relatively small number of races. Confounded with real associations in such a case is the effect of selection on diversity. For some populations, association may also include various kinds of non-random combinations of virulence and background genotypes, all resulting in reduced diversity (Wolfe & Knott, 1982).

Two special cases can be explored with respect to associations. The first is the case (considering all virulences at 0.5 frequency) where each of the identified virulence genes is equally associated with every other. The effect of this case on the increment of diversity is defined by the relationship:

$$\Delta D = \frac{-\left((1 + a) \ln \left(\frac{1 + a}{2}\right) + (1 - a) \ln \left(\frac{1 - a}{2}\right)\right)}{2} \qquad [3]$$

where ΔD = the change in Shannon index due to association when a differential line is added and a = the coefficient of association such that the products of frequencies of two virulences are multiplied by $1 + a$ if associated in coupling or $1 - a$ if associated in repulsion. This case can be described graphically as a series of lines in the same way as levels of departure from 0.5 frequency (Figure 6.3).

The second special case is where only one pair of the virulences that are being

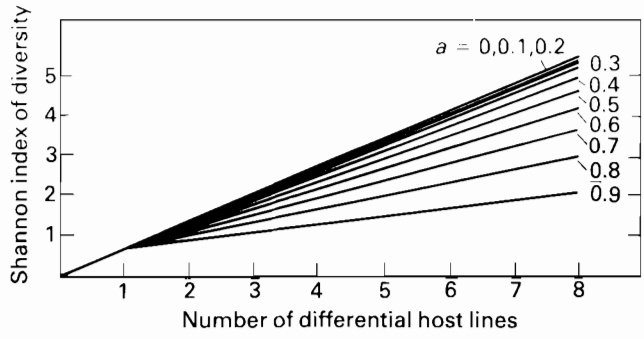

Figure 6.3. The relationship between the number of differential host lines and increase in Shannon diversity index for ten different degrees of association or disassociation of the corresponding virulences. The degrees of association are shown by the coefficients a above each line. All virulence frequencies are 0.5.

detected is associated. Mathematically, the increment of diversity is generally described as:

$$\Delta D = \frac{-2\left(\ln\left(\frac{1}{2^n}\right) + (1+a)\ln\left(\frac{1+a}{2^n}\right) + (1-a)\ln\left(\frac{1-a}{2^n}\right)\right)}{2^n} \qquad [4]$$

where ΔD = the change in Shannon index when a differential line is added, a = the coefficient of association, and n = the number of differential lines.

The case when all associations are different is not so clear, and no simply stated relationship for increments of Shannon index has been developed as yet. All of the myriad associations will have to be accounted for as each virulence is added, resulting in a cumbersome algebraic statement if all associations are explicitly included.

Since there are only two factors (virulence frequency and virulence associations) which influence diversity in our development, the lack of ability to generalize about many different associations is not a major problem. Direct calculation of the change in Shannon index when a new differential host line is added requires only that race numbers and frequencies be known both before and after adding the new line. The value obtained, which is the 'actual' line in Figure 6.4, will nearly always be less than the maximum possible, which is a slope of $-\ln(0.5)$ per differential line. Explicitly, the part of the difference between actual and maximum increment (a in Figure 6.4) that is due to virulence frequency of the newly detected virulence can be calculated using equation [2]. The value calculated is an increment or slope that will fall somewhere between, or rarely equal, either the actual or the maximum increments. The difference between $-\ln(0.5)$ and the actual calculated increment is the total effect of both virulence frequency (which is explicit) and associations (whose overall effects are known only implicitly). The difference (b in Figure 6.4) between the increment allowing for virulence frequency alone and the actual calculated slope

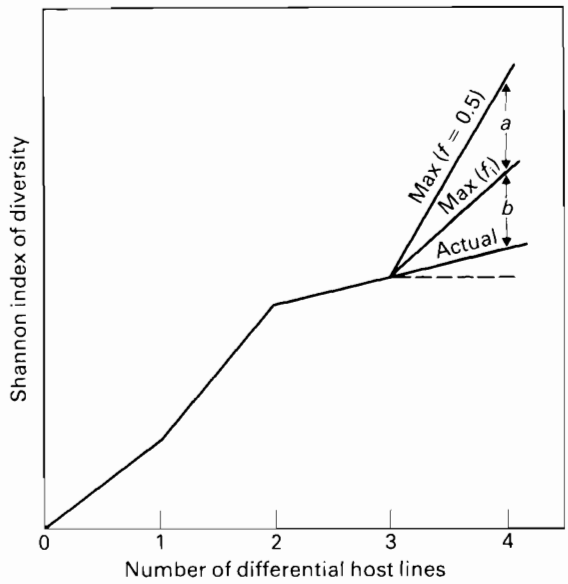

Figure 6.4. The increment of virulence diversity obtained when a hypothetical 4th host line is added to a hypothetical differential set, each member of whose increments is shown. Actual = the increment observed; Max ($f = 0.5$) = the estimated increment assuming the frequency of matching virulence is 0.5; Max (f_i) = the estimated increment assuming the actual frequency of matching virulence. Segment *a* represents the portion of diversity (information) lost due to the departure of the virulence frequency from 0.5. Segment *b* represents the portion of diversity lost because of non-random association of virulences.

must then be due to associations. So, while the associations cannot be separated, their overall influence can be determined. In so doing, we can make quantitative comparisons of the relative importance of degree of association and virulence frequency. The above analyses can be made as each differential host line is added. The linear relationship between the number of differential lines and the value of the Shannon index ensures that results are simply additive.

Application of the theory

To illustrate the uses to which these analyses can be put, we will use wheat stem rust data from 1975 collections made in the United States. The overall expectation that the amount of diversity detected is a function of the differential set used is best illustrated by an already published example of Young & Prescott (1977) for leaf rust of wheat, *Puccinia recondita* f. sp. *tritici*. Using two different sets of differential host lines, they obtained very different numbers and frequencies of races for each of two populations. The differences were illustrated and Shannon indexes calculated in a later analysis by Groth & Roelfs (1982).

Analyses were carried out on the 1975 sexual (Washington and Idaho) and

asexual (east of the Rocky Mountains) populations, where 16 single-gene differential lines of wheat were used to characterize virulence diversity (Roelfs & Groth, 1980). Since virulence was not detected in either population to *Sr*13, this differential line was not included. Two other resistance genes, *Sr*5 and *Sr*16, added nothing to diversity in the asexual population because virulence to both was fixed. They were retained, however, because virulence to neither was fixed in the sexual population.

The sexual population was more diverse, with 39% of the maximum potential diversity (D_{max} = 10.4) being accounted for, as opposed to only 12% in the asexual population (Figure 6.5). As also might have been expected, a larger percentage of the maximum possible diversity (39%) was not detected due to virulence associations in the asexual population than in the sexual population (14%). The remaining diversity loss, due to virulence frequency was nearly equal, being 47% and 49% for the sexual and asexual populations, respectively (Figure 6.5).

In stem rust, we have examined selected sets of four differential host lines. For both asexually and sexually reproducing populations of the fungus two sets of four single-gene differential lines each were selected (Table 6.1). Set one includes four *Sr* genes whose corresponding virulences were as near to fixation as was possible without being actually fixed. Set two includes four *Sr* genes selected because the matching virulence in the rust population is polymorphic, and as close to 0.5 frequency as was available in the 16 single-gene differentials examined. An additional case, case 3, was included for the asexual population; the four genes were selected on the basis that they appeared to detect a larger number of races than other sets of four. Table 6.1 shows maximum possible and actual diversity indexes, as well as calculated loss in diversity due to degree of near-fixation of the four genes in each of the five cases. These analyses are presented graphically in Figures 6.6 and 6.7. The percentage of potential diversity represented by actual diversity ranges from 19% in the less polymorphic sexual case (Set 1) to 83.1% in the more polymorphic sexual case (Set 2) (Figure 6.7). These losses are further split into loss due to

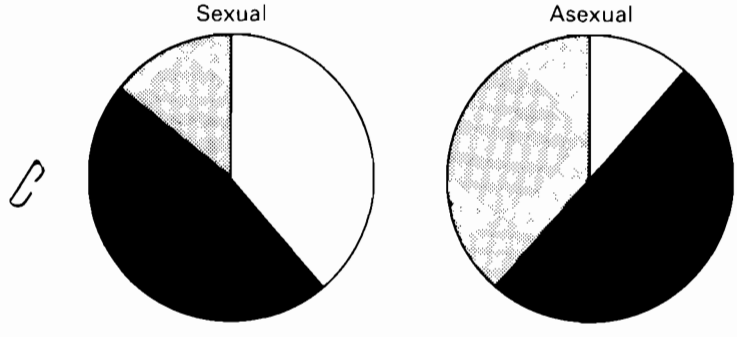

Figure 6.5. Components of virulence diversity of two 1975 populations of stem rust of wheat in the US. The sexual population occurs where the alternate host, barberry, is common while the asexual population effectively lacks barberry. White = detected diversity (on 15 single-gene differential host lines); black = loss in diversity due to departure of virulence from 0.5 frequency; stippled = loss in diversity due to associations of virulences.

Table 6.1. Values of components of virulence diversity for five sets of four single-gene differential wheat lines used to subdivide two 1975 populations of wheat stem rust in the US

Set	*Sr* gene series[d]	Shannon index of diversity[e]	
		Actual *D*	Loss due to departure from 0.5 frequency
Asexual			
Set 1[a]	*9d*(2363), *3c*(2148), *9b*(293), *6*(254)	.967	1.706
Set 2[b]	*Tmp*(1860), *9e*(1855), *17*(463), *7b*(414)	.682	0.768
Set 3[c]	*11*(1980), *6*(254), *8*(2077), *9a*(400)	1.171	1.150
Sexual			
Set 1[a]	*6*(4), *9b*(5), *10*(42), *15*(399)	.489	2.275
Set 2[b]	*5*(296), *9d*(296), *8*(92), *9a*(270)	2.234	0.364

[a] Four wheat lines whose corresponding virulences were less polymorphic.
[b] Four wheat lines whose corresponding virulences were more polymorphic.
[c] Four wheat lines that subdivided the pathogen population into a large number of races.
[d] Number of virulent isolates indicated in brackets.
[e] Maximum possible diversity (D) for each each set is 2.773.

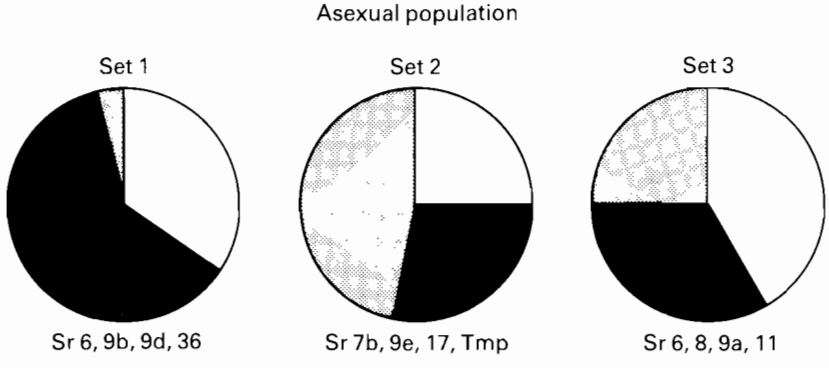

Figure 6.6. Components of virulence diversity of an asexually reproducing population of stem rust of wheat from the US sampled in 1975. Each circle is based on a different set of four single-gene differential host lines (see Table 6.1). White = detected diversity; black = loss in diversity due to departure of virulence from 0.5 frequency; stippled = loss in diversity due to associations of virulences.

virulence frequency, which ranged from 14% of the total possible diversity in set 2 sexual (Figure 6.7) to 82% of the total in set 1 sexual, and loss due to virulence associations ranging from 1% of total possible diversity in set 1 sexual to 48% in set 2 asexual. Both total amount of diversity detected and proportion of diversity loss due to frequency vs. association must be considered in order to obtain a complete picture of diversity.

Similar analyses can be done on single differentials. The amount of information

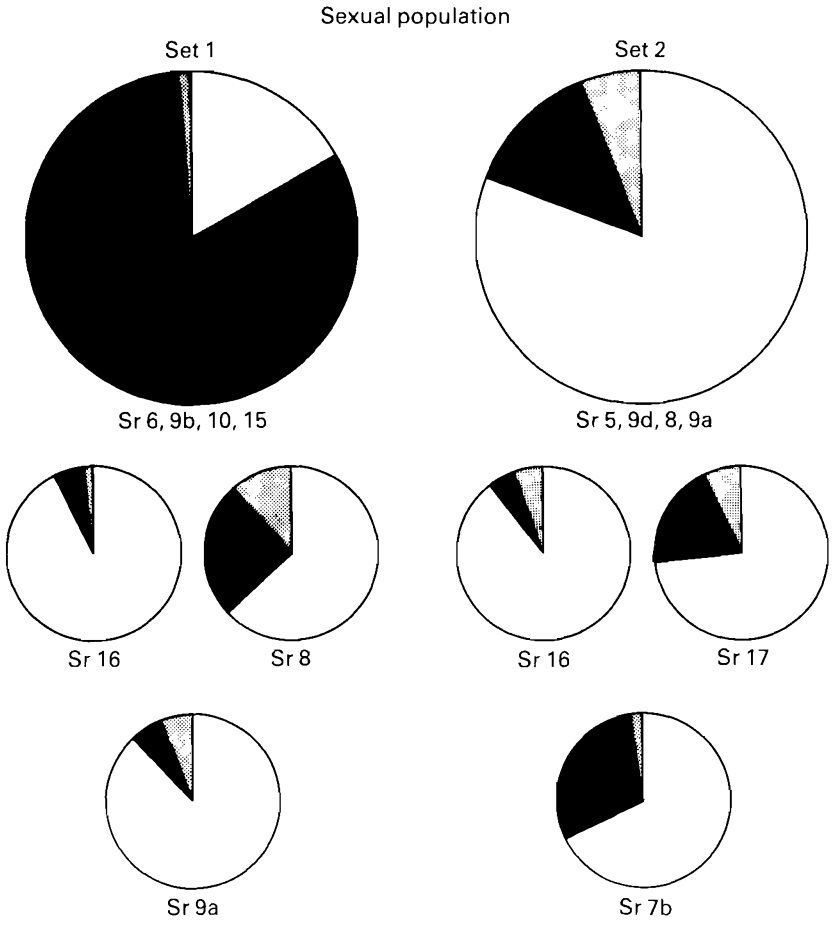

Figure 6.7. Components of virulence diversity of a sexually reproducing population of stem rust of wheat from the US sampled in 1975. Each large circle is based on a different set of four single-gene differential host-lines (see Table 6.1). The smaller circles each represent a fifth line which is being added to the set above. White = detected diversity; black = loss in diversity due to departure of virulence from 0.5 frequency; stippled = loss in diversity due to associations of virulences.

lost due to near-fixation is independent of the identities of the resistance genes already present in the differential host set to which a line is being added. This is not true for either the loss in information due to associations or the increment in diversity when the line is included; these will be influenced by the genes already present. To illustrate the effects of the addition of different genes to existing four-gene sets, separate analyses were made of the addition in diversity when new genes are added, singly, to each of the two sets of the sexual population (Table 6.2 and Figure 6.7). The proportion of loss in detected diversity due to departure from 0.5 virulence frequency of the newly-introduced resistance gene is directly proportional to the magnitude of this departure (Table 6.2), which is the single independent variable of equation [2]. The loss of diversity thus obtained is independent of

Table 6.2. Values of components of virulence diversity when single-gene differential lines are added, as the fifth differential host, to two sets of four lines that were used to divide a 1975 sexual population of stem rust of wheat in the US. Maximum possible diversity added by a single line is 0.693

		Shannon index of diversity		
Added gene	Virulence frequency	Actual increment	Loss due to departure from 0.5 frequency	Number of races added
Set 1[a]				
*Sr*16	.63	.647	.036	4
*Sr*8	.22	.440	.171	4
*Sr*9a	.63	.617	.036	3
Set 2[b]				
*Sr*16	.63	.630	.036	13
*Sr*17	.76	.513	.137	12
*Sr*7b	.19	.474	.207	13

[a] Four single-gene lines whose corresponding virulences were less polymorphic; set 1 identified 6 races.
[b] Four single-gene lines whose corresponding virulences were more polymorphic; set 2 identified 14 races.

the specific four genes to which the fifth is being added, as seen by the identical sizes of the black slices for the addition of *Sr*16 to the two different four-line sets (Figure 6.7). In addition, *Sr*9a, which shows the same virulence frequency as *Sr*16, shows an identical proportion of lost diversity as *Sr*16. As has been indicated more generally earlier, the loss in diversity due to association is dependent on the specific genes already present in a differential host set, when a new gene is added. This can most clearly be seen for *Sr*16 where the size of the stippled slice is larger when this gene is added to set 2 than when it is added to set 1 (Figure 6.7).

The remaining pie charts illustrate that various relative and absolute amounts of detected diversity are possible when new genes are added, and that the loss in diversity can be more the result of departure of virulence frequencies from 0.5 (*Sr*7b), association (*Sr*16 added to set 2), or both (*Sr*9a) (Figure 6.7). As previously pointed out, association was not as important a cause of loss in the sexual as in the asexual population (Figure 6.5, and cf. Figure 6.6 with Figure 6.7). A degree of non-random association of virulences in the sexual population has been shown to exist using other methods (Alexander *et al*, 1984).

Conclusions

The intent of this development and demonstration is to suggest a way in which binomial expectations can be used in conjuction with a diversity index to permit a more complete and objective analysis of how differential host lines operate. A number of changes or improvements in the analysis are possible. One problem is

that the correspondence between the ideal cases outlined in the first part of the chapter, and the real cases for which such an analysis is intended, is not perfect. We have not addressed sampling problems here; small sample sizes and small numbers of races are not conducive to analysis using the above methods. Asexual populations, in particular, may show very large association effects. We have not tried explicitly to define virulence associations alone, much less the effects of selection when virulences are 'trapped' in specific genetic backgrounds. The contribution of association therefore is, and probably will contine to be, the most elusive component of diversity. A more analytical treatment of associations will provide further insight into phenomena important in the adaptation of plant pathogen populations to the use of host resistance. Integration of background genotype and race-specific virulence complement, as well as other interactions between virulence and natural selection, are examples of such phenomena.

Eventually such analyses should be applied to differential host lines that do not contain a single resistance gene each. It seems likely to us, however, that rather extensive knowledge of the genetics of resistance will be necessary if the model is to be valid.

Acknowledgements

This work was supported in part by USDA Competitive Grant 59-2271-1-1-687-0, and by the Minnesota Agricultural Experiment Station.

References

Alexander H.M., Roelfs A.P. & Groth J.V. (1984) Pathogenicity associations in *Puccinia graminis* f. sp. *tritici* in the United States. *Phytopathology* **74**, 1161–6.

Browder L.E. & Eversmeyer M.G. (1977) Pathogenicity associations in *Puccinia recondita tritici*. *Phytopathology* **67**, 766–71.

Burdon J.J., Marshall D.R., Luig H.H. & Gow D.J.S. (1982) Isozyme studies on the origin and evolution of *Puccinia graminis* f. sp. *tritici* in Australia. *Australian Journal of Biological Sciences* **35**, 231–8.

Groth J.V. & Roelfs A.P. (1982) Effect of sexual and asexual reproduction on race abundance in cereal rust fungus populations. *Phytopathology* **72**, 1503–7.

Lebeda A. (1982) Measurement of genotypic diversity of virulence in populations of phytopathogenic fungi. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz* **89**, 88–95.

Pielou E.C. (1977) *Mathematical Ecology*. Wiley, New York.

Roelfs A.P. & Groth J.V. (1980) A comparison of virulence phenotypes in wheat stem rust populations reproducing sexually and asexually. *Phytopathology* **70**, 855–62.

Simons M.D., Rothman P.G. & Michel L.J. (1979) Pathogenicity of *Puccinia coronata* from buckthorn and from oats adjacent to and distant from buckthorn. *Phytopathology* **69**, 156–8.

Whittaker R.H. (1972) Evolution and measurement of species diversity. *Taxon* **21**, 213–51.

Wolfe M.S. & Knott D.R. (1982) Populations of plant pathogens: some constraints on analysis of variation in pathogenicity. *Plant Pathology* **31**, 79–90.

Young H.C. & Prescott J.M. (1977) A study of race populations of *Puccinia recondita* f. sp. *tritici*. *Phytopathology* **67**, 528–32.