Consideration of Nonparametric Approaches for Assessing Genotype-by-Environment ($G \times E$) Interaction with Disease Severity Data

L. V. Madden, P. A. Paul, and P. E. Lipps, Department of Plant Pathology, The Ohio State University, Ohio Agricultural and Research Development Center (OARDC), Wooster 44691

ABSTRACT

Madden, L. V., Paul, P. A., and Lipps, P. E. 2007. Consideration of nonparametric approaches for assessing genotype-by-environment ($G \times E$) interaction with disease severity data. *Plant Dis.* 91:891-900.

Determination of host genotype-by-environment ($G \times E$) interaction is needed to assess the stability of cultivar traits such as plant disease resistance and to reveal differences in aggressiveness or virulence of pathogen strains among locations. Here we explored the use of rank-based methodology to quantify the concordance (or discordance) of disease responses of host genotypes across environments, based on the Kendall coefficient of concordance (W) and ancillary test statistics, in order to determine the extent to which environment affected rankings of genotypes. An analysis of four data sets for disease severity of gray leaf spot of maize (with genotypes planted in as many as 11 locations in a given year) revealed highly significant concordance ($P \leq 0.001$) overall, indicating that genotypes varied little in within-environment rankings. This suggests that the $G \times E$ interaction was small or nonexistent (in terms of rankings). A novel rank-based method by Piepho was evaluated to further elucidate the interaction (if any) through a test for variance homogeneity. The Piepho test statistic was not significant ($P > 0.25$) for any of the gray leaf spot data sets, confirming the stability of genotypes across environments for this pathosystem; however, analysis of published data sets for other pathosystems indicated significant results. The relationship shown by Hühn, Lotito, and Piepho between the ratio of genotype and residual variances of the original data and the rank-based W statistic was evaluated using Monte Carlo simulations. A more general functional relationship was developed that is applicable over a wide range of number of genotypes and environments in the analyzed studies. This confirms previously shown linkages between rankings of genotypes within environments and variability in the original (unranked) data, thus permitting ease of interpretation of parametric and nonparametric results.

Additional keywords: *Cercospora zea-maydis*, coefficient of discordance, Nassar-Hühn rank variance statistic, rank-variance homogeneity

The study of host genotype-by-environment ($G \times E$) interactions is of major importance in plant pathology and other agricultural sciences (20,23). For instance, plant breeders routinely grow cultivars and other genotypes in multiple locations and years to determine whether or not environment affects the magnitude of specific quantitative traits of the host genotypes, such as yield and disease severity, as well as differences of the values of the traits among genotypes (25). For plant genotypes that potentially differ in susceptibility to a given pathogen, a $G \times E$ interaction could indicate differences in aggressiveness or virulence of the pathogen isolates among locations/years, and possi-

bly even the existence of races. In this context, environment represents, at least in part, the pathogen population. Thus, multiple environment studies can focus on characterizing properties of the genotypes or the locations. Given the large variability that can exist between host genotypes and environments, Yan and Falk (38) stated that "understanding the host-by-pathogen interaction patterns for a particular host-pathogen system can be difficult and challenging".

A specific problem involves gray leaf spot of maize in the United States (36). The disease, caused by the fungus *Cercospora zea-maydis* Tehon & Daniels, is a major constraint to maize production in many parts of the United States and other parts of the world, especially with minimum tillage or no-till cultivation. Resistance to gray leaf spot is a quantitatively inherited trait (8,30), and there is considerable variation in disease among locations or years (7). It is of interest to know whether local pathogen populations vary in aggressiveness so that the stability of resis-

tance in inbreds and hybrids can be assessed (6). For instance, in one study in North Carolina (7), there was a significant $G \times E$ interaction for disease severity when a set of maize genotypes were grown at four locations.

Numerous statistical methods have been developed to characterize $G \times E$ interactions (9,14,19,23,25,33,34), especially for the purpose of quantifying yield stability of plant genotypes. A typical approach (20) is to assume a normal distribution for the measured response and utilize the general linear model

$$X_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (1)$$

where X_{ij} is mean response of genotype i in environment (location or year) j , μ is a constant (mean), α_i and β_j are the main effects of the i th genotype and j th environment, respectively, on X , and e_{ij} is the residual. Here, e_{ij} is the sum of the $G \times E$ interaction effect and the average experimental error over the number of replicates per environment $[(\alpha\beta)_{ij} + \bar{\varepsilon}_{ij}]$. Many of the statistical methods centers on estimation and/or modeling of the e_{ij} term in order to determine the extent to which genotype effects depend on the environment (25,26,28). Sophisticated graphical analyses of e_{ij} have also been developed (*biplots*) to address this same issue (37–39).

Typically, it is assumed that genotype has a fixed effect on X , and that environment and the residual have random effects. However, genotype could also be considered to have a random effect. With a $G \times E$ interaction, the variance of the residuals, $\text{Var}(e_{ij}) = \sigma^2$, would be expected to be larger than 0, where σ^2 is the sum of the interaction variance $[\text{Var}(\alpha\beta)]$ and the experimental variance divided by the number of replications per environment. Or, the residual variance may vary with the genotype, so that $\text{Var}(e_{ij}) = \sigma_i^2$. The genotype-specific variance is known as the stability variance (25). One approach to test for $G \times E$ interaction is to determine if the σ_i^2 are equal (the null hypothesis), or if σ_i^2 vary (the alternative hypothesis) (25,33). However, tests for variances are generally more sensitive to lack of normality than are tests for fixed effects in mixed-model analyses (2). Moreover, all forms of interaction are not of equal interest in genotype responses to pathogen populations. Biologically,

Corresponding author: L. V. Madden
E-mail: MADDEN.1@osu.edu

Accepted for publication 13 February 2007.

interactions that result in change of the ordered ranking of genotypes within environments may be of most interest. For instance, with two genotypes, an interaction would occur in the original measurement scale if the difference in X between the genotypes increased (decreased) from environment 1 to 2; a rank interaction would only occur if the ordering of genotypes changed between the two environments. Thus, nonparametric approaches based on ranks (10,12,14,32,35) are advantageous for studying interactions. Rank-based methods also reduce the impact of differences in disease assessment approaches among environments and the influence of extreme values of X (outliers) on results.

Several nonparametric approaches have been developed or implemented for quantifying $G \times E$ interaction effects on crop yield and similar response variables (9,14,15,17,24,33,34), in the context of phenotype-stability characterization. Considerably less attention has been given to quantifying the interaction effect on disease severity (5), especially for the purpose of characterizing the pathogen population aggressiveness at multiple locations. Here we present a nonparametric analysis of disease severity of maize genotypes in response to natural infection by the gray leaf spot pathogen in multiple environments. The data were collected by members of the NCR-25 USDA North Central Region Technical Committee on Corn and Sorghum Diseases for the purpose of documenting the level of resistance of selected genotypes and assessing differential responses of genotypes across locations (pathogen populations and environments).

THEORY AND APPLICATION

Concordance. Let K represent the number of genotypes and N the number of environments (e.g., locations). Furthermore, let X represent the response variable being analyzed (disease severity in this case). We refer to X as the disease response variable, or more simply, the disease response. The data can be presented conveniently as a two-way table with K rows and N columns. X_{ij} is the disease response for

the i th genotype ($i = 1, \dots, K$) in the j th environment ($j = 1, \dots, N$). If there are multiple replications at a location, then X_{ij} is the mean of those replicate values in the ij th cell of the table. The mean X across all environments for the i th genotype is given by $\bar{X}_{i\cdot}$, and the mean across all genotypes for environment j is given by $\bar{X}_{\cdot j}$. Overall mean is $\bar{X}_{\cdot\cdot}$.

Ranks of the disease responses for the genotypes are performed within each environment. The rank of X_{ij} within the j th environment is represented by R_{ij} . The mean rank across all environments for the i th genotype is $\bar{R}_{i\cdot}$. The mean rank (across all genotypes) within an environment, $\bar{R}_{\cdot j}$, is $(K + 1)/2$. The mean of the K different $\bar{R}_{i\cdot}$ values, $\bar{R}_{\cdot\cdot}$, is also simply $(K + 1)/2$. Note that by ranking host genotypes separately for each environment, the mean rank in each environment is always the constant $(K + 1)/2$ because any main effect of environment (α_i) is removed by the within-environment rankings. The variance of the ranks for the i th genotype is given by:

$$S_i^{(2)} = \frac{\sum_{j=1}^N (R_{ij} - \bar{R}_{i\cdot})^2}{N-1} \quad (2)$$

Maximum concordance (i.e., agreement) occurs if the rankings of the genotypes are identical for all environments. That is, one of the genotypes would be ranked 1 in all environments, another one would be ranked 2, and so on. The $\bar{R}_{i\cdot}$ values would also be the same as the individual-environment rankings (the integers 1, 2, ..., K) with maximum concordance. The $S_i^{(2)}$ values would thus be 0 for all genotypes. $S_i^{(2)}$ is loosely analogous to the stability variance (σ_i^2) often calculated for normal-distribution models (25). When there is no evidence of interaction regarding the original measurements, σ_i^2 values will be about equal; when there is no rank interaction, $S_i^{(2)}$ values will all be about equal and also small (close to 0). $S_i^{(2)}$ would be small because it is determined by genotype effect and the residual; a strong genotype effect could lead to identical (or almost identical) rankings in each environment.

The genotype rank totals across all environments ($\sum R_{i\cdot}$) must consist of the integer values $1N, 2N, 3N, \dots, KN$, when there is perfect agreement in rankings. The sum of squares of the actual rank totals around the overall mean total $[N(K + 1)/2]$ for any realized data set is given by:

$$S = N^2 \sum_{i=1}^K \left(\bar{R}_{i\cdot} - \frac{K+1}{2} \right)^2 = \sum_{i=1}^K \left(N\bar{R}_{i\cdot} - \frac{N(K+1)}{2} \right)^2 \quad (3)$$

S will be a large number if there is high agreement in within-environment rankings (since the actual rank totals will cover a wide range). The maximum possible sum of squares of the totals for ranks (summing over all genotypes) is given by $S_{\max} = N^2 K(K^2 - 1)/12$ (18). Kendall's coefficient of concordance (W) is the ratio of the actual sum of squares to the maximum possible value:

$$W = \frac{S}{N^2 K(K^2 - 1)/12} \quad (4)$$

W has the value 1 when the ranks are identical within each environment and a value of 0 when there is absolutely no concordance in the rankings between environments (21).

W can also be defined based on the pairwise rank correlations of R_{ij} between environments (16). If $r_{jj'}$ is the Spearman rank correlation (12) between the pair of environments j and j' , then it can be shown that the mean Spearman rank correlation between all pairs of environments is equal to $\bar{r} = (NW - 1)/(N - 1)$ (18). It can be shown that W is equal to:

$$W = \frac{1 + (N - 1)\bar{r}}{N} = \bar{r} + \frac{1 - \bar{r}}{N} \quad (5)$$

The more consistent the ranks are between environments, the higher the $\bar{r}$, and thus the higher the W value.

Two hypothetical, and extreme, very small examples (with $K = 5$ and $N = 4$) are shown in Table 1. One would normally not

Table 1. Two theoretical examples of disease response (X_{ij}) of five genotypes ($i = 1, \dots, 5$) in four environments ($j = 1, \dots, 4$), ranking within each environment (R_{ij}), mean rank across environments ($\bar{R}_{i\cdot}$), and rank stability variance ($S_i^{(2)}$) for each genotype

Example	Genotype (<i>i</i>)	Environments (<i>j</i>)								Mean ($\bar{R}_{i\cdot}$)	Var. $S_i^{(2)}$
		Observations (X_{ij})				Ranks (R_{ij})					
		1	2	3	4	1	2	3	4		
A	1	5	20	2	64	1	1	1	1	1	0
	2	20	40	4	65	2	2	2	2	2	0
	3	25	60	6	70	3	3	3	3	3	0
	4	30	80	8	75	4	4	4	4	4	0
	5	60	100	10	80	5	5	5	5	5	0
B	1	5	100	2	80	1	5	1	5	3	5.33
	2	20	80	4	75	2	4	2	4	3	1.33
	3	25	60	6	70	3	3	3	3	3	0
	4	30	40	8	65	4	2	4	2	3	1.33
	5	60	20	10	64	5	1	5	1	3	5.33

use such a small data set for evaluating interactions because of low power of the statistical tests (see below), but the data are useful for demonstrating the methodology. For example A, although the X values varied considerably between locations (with very high disease severity values in environments 2 and 4, and very low values in environment 3), the rankings of the genotypes were identical across all locations. Mean ranks across the locations thus equaled the integers from 1 to 5, and rank totals equaled the values 4 (=4·1) to 20 (=4·5). The Spearman rank correlation was 1 for all pairs of environments, so that $\bar{r} = 1$. Furthermore, the variance of the ranks across locations, $S_i^{(2)}$, equaled 0 for all genotypes because of the identical rankings. With these values of K and N , the maximum possible sum of squares of totals is $S_{\max} = 4^2 \cdot 5 \cdot (5^2 - 1)/12 = 160$. The actual sum of squares of totals (S), based on equation 3, was also 160 here, giving $W = 1.0$ (see equation 4). Use of equation 5 also gave the same W value, as required.

In example B, the same X values are observed within each environment, but they corresponded to different genotypes in some cases (Table 1). Here the ranks within environments differed, and a wide range of Spearman rank correlations were obtained: either 1 (e.g., environments 1 and 3), or -1 (e.g., environments 1 and 2). With four environments, there are $4 \cdot (4 - 1)/1 = 6$ unique correlations. There were four correlations equal to -1 and two equal to +1. Mean rank correlation ($\bar{r}$) was -0.33. The mean ranks across environments were all equal to 3 (with a rank total of $4 \cdot 3 = 12$). S equaled 0 (compared to 160 for example A). Using either equation 4 or 5, W equaled 0.0, clearly showing a lack of concordance. Moreover, the rank variances ($S_i^{(2)}$) varied considerably, from 0 to 5.3, for the five genotypes (Table 1).

Statistical tests. The Friedman test can be used to evaluate the equality of treatment (including genotype) effects on a response variable, such as disease severity, where the data are collected in separate locations (blocks). The null hypothesis of equality of genotypes (in terms of disease response) is equivalent to the null hypothesis of random rankings of genotypes within environments (i.e., the lack of concordance of rankings) (12,21). It is easily shown that the Friedman test statistic, T , can be written as:

$$T = N(K-1)W = \frac{N(K-1)S}{N^2 K(K^2-1)} \quad (6)$$

Under the null hypothesis, and with a large number of environments ($N > 10$), T has a chi-square distribution with $K-1$ degrees of freedom (df). With small N , use of a chi-square test may be inaccurate for determining a significant effect because T may deviate from a chi-square variate

under the null hypothesis. A more accurate test is based on $F = (N-1)W/(1-W)$, which has an F distribution under the null hypothesis, with numerator df given by $df_N = (K-1) - (2/N)$, and denominator df given by $df_D = df_N \cdot (N-1)$ (21). As W goes to 1 in the limit, F goes to infinity (essentially becoming undefined with perfect concordance).

For example A (Table 1), $T = 16$ (its largest possible value). The achieved significance level for the chi-square test, with $5-1$ df is $P = 0.003$. Note that there is low power for testing hypotheses with only 5 genotypes and 4 locations. F is undefined because $W = 1$. This is interpreted as a highly significant result (27). For example B, $T = 0.0$, giving an achieved significance level of $P = 1.0$ with the chi-square test. For this example, F also equals 0 (with 3.5 and 10.5 df), giving a P value of 1.0. Clearly, the data set of example B exhibits no concordance of rankings.

It should be noted that the Friedman test, or its variations, is not used to test for an environment (location) main effect. This is because the separate rankings of genotypes within each location remove the environmental effect on the results.

Variance summation. As indicated above, the between-environment rank variances for the individual genotypes, $S_i^{(2)}$, will be small if the genotype ranks are consistent among environments. The sum of the $S_i^{(2)}$ values can also be used to test for concordance instead of using W or S . Let $\Psi = \sum_i^K S_i^{(2)}$. As shown by Hühn and Piepho (18) for data without ties,

$$\Psi = \frac{NK(K^2-1)}{12(N-1)}(1-W) = \frac{K(K^2-1)}{12}(1-\bar{r}) \quad (7)$$

The term in front of $(1-W)$, determined by the size of the two-way table and not the response values or ranks, can be considered a standardization factor in the relationship between Ψ , W , and $\bar{r}$. Note that $1-W$ is a discordance measure: the larger the discordance (or the smaller the concordance), the larger the value of Ψ . The Ψ values move in opposite direction of S with increasing discordance—large discordance results in a small value of S but large value of Ψ . For example A, $\Psi = 0$, and for example B, $\Psi = 13.33$. As pointed out by Hühn and Piepho (18), Ψ is a scaled measure of the difference between the Friedman T statistic and the maximum possible value that can be taken by T [$T_{\max} = N(K-1)$]. In other words, Ψ indicates how far the actual data set is from maximum concordance.

Based on the relationships of equation 7, a critical value of Ψ can be defined for significance (Ψ^*) at a given P value (e.g., 0.05). There is significant concordance if Ψ is less than Ψ^* (i.e., $\Psi < \Psi^*$), where the latter is defined as:

$$\Psi^* = \frac{NK(K^2-1)}{12(N-1)(1+\psi)} \quad (8)$$

in which ψ is a value based on the critical value of a chi-square or F distribution. For the former, ψ is given by:

$$\psi = \frac{\chi_{K-1}^2}{N(K-1) - \chi_{K-1}^2} \quad (9)$$

where χ_{K-1}^2 is the tabulated chi-square value at the chosen P with $K-1$ df. For the example ($K = 5$), $\chi_{K-1}^2 = 9.49$ at $P = 0.05$. This gives $\psi = 1.46$. With this value, one obtains $\Psi^* = 5.43$ for $P = 0.05$ and $K = 5$. In other words, any value of actual Ψ less than 5.43 is significant at $P = 0.05$. Example B does not exhibit significant concordance because $\Psi \approx 13$. The Ψ^* value based on the F distribution is given in Hühn and Piepho (18).

Genotype $\times$ environment (G $\times$ E) interaction. If there is maximum concordance (as in example A), one obtains $W = 1$, $T = T_{\max} = N(K-1)$, and $\Psi = 0$. This means there is no G $\times$ E interaction, in terms of ranks, because genotypes have identical ranks in every environment. However, there still may be an interaction in terms of the original observations. Even though the genotype X values may have the same rank order in each environment, the magnitude of differences of X values between genotypes may depend on the environment. If $W < 1$, there is less than perfect agreement in the ranks. This can be due to an interaction (where environment affects the rank order) or experimental variation (error or noise). The larger the W (but less than 1), the less likely there is an interaction, but the test for W (or for S or Ψ) is not a direct test for interaction. There may be an interaction in terms of the original X values even with a significant concordance. Large differences in means of the genotypes could mask or overshadow any G $\times$ E interaction.

Hühn and Nassar (17) and Nassar and Hühn (24) developed methods for quantifying interactions for two-way tables as in Table 1. Their approach involves calculating so-called "aligned variables" from the original X_{ij} values in order to remove the "nuisance" effect of genotype main effect (ranking within environments automatically removes the nuisance effect of environment). The aligned value is $X'_{ij} = X_{ij} - (\bar{X}_{i.} - \bar{X}_{..})$. Aligning of variables has been advanced by several authors for nonparametric analyses (e.g., 1,22). However, Piepho (27) has indicated some deficiencies with the Nassar-Hühn approach, because the aligned variables are not independent and test results depend on the original data distribution (thus, the tests are not nonparametric). Because of the former, the $S_i^{(2)}$ values are not independent when determined based on the aligned values, thus leading to problems in interpretation of test statistics.

Piepho (27) developed a new nonparametric procedure to test for $G \times E$ interaction (i.e., more specifically, rank-variance homogeneity) that does not have the drawbacks of the Hühn and Nassar (17) method. The basic concept is to calculate differences of observations *between* environments for each genotype. If the number of columns (environments) of the two-way table is an even number, then the columns are subdivided into $C = N/2$ pairs of new columns. Each of the new columns consists of the contrast Y_{il} ($i = 1, \dots, K$; $l = 1, \dots, C$), defined as:

$$Y_{il} = X_{ij'} - X_{ij} \quad (10a)$$

where j and j' represent a pair of the original columns. Equation 10a removes the genotype effect (α_i) on the disease response. This can be seen by considering contrasts for the model in the Introduction (equation 1): $X_{ij'} - X_{ij} = (\beta_{j'} - \beta_j) + (e_{ij'} - e_{ij})$. Both μ and α_i are eliminated by the differencing. This allows the researcher to focus on detection of interaction.

The method is demonstrated for examples A and B in Table 2. The original four columns of data (Table 1) become two columns of contrasts. For instance, the values for the four environments with the first genotype (A) were 5, 20, 2, and 64. The calculated Y values for the two new "environments" are 15 ($= 20 - 5$) and 62 ($= 64 - 2$).

If there is an odd number of environments, one of the new columns in the table is based on a contrast of three original columns (a "triplet"), and the rest of the columns are determined based on equation 10a. The triplet contrast is defined as:

$$Y_{il} = X_{ij} + X_{ij'} - 2X_{ij''} \quad (10b)$$

where j , j' , and j'' are three adjacent original columns. When N is odd, then the number of new columns is $C = 1 + (N - 3)/2$. Equation 10b could be used for either the first three columns or the last three columns, but we use it for the first three.

For example, suppose there were seven environments for a genotype, with X observations 20, 40, 10, 0, 25, 45, and 60. The three new Y observations would be 40 ($= 20 + 40 - 2 \cdot 10$), 25 ($= 25 - 0$), and 15 ($= 60 - 45$). All subsequent analyses are based on the Y values and their within-column ranks.

Let R_{il} represent the rank of the Y_{il} within the l th column. Just as equation 10a or 10b removes genotype effects, within-column ranking removes environment effects on the results. A rank variance for each genotype (across all new columns) can be defined as:

$$U_i = \frac{\sum_{l=1}^C \left(R_{il} - \frac{K+1}{2} \right)^2}{C} \quad (11)$$

The sum of the U_i^2 values (not the U_i values) over all genotypes ($i = 1, \dots, K$) is a measure of homogeneity of rank variances (of the Y values), and hence one measure of interaction (27). Unlike the $S_i^{(2)}$ values for the original ranks (equation 2), U_i values are based on the overall mean rank $[(K + 1)/2]$ and not genotype-specific rank mean. The test statistic is given by:

$$U = \left(\frac{180C}{K(K+1)(K^2-4)} \right) \sum_{i=1}^K (U_i)^2 - \left(\frac{5C(K^2-1)(K-1)}{4(K^2-4)} \right) \quad (12)$$

The null hypothesis tested is that the rank orders in the different (new) columns are stochastically independent, and each ordering of genotypes is equally likely. Under the null hypothesis, U has asymptotically (i.e., at large C) a chi-square distribution, with $K - 1$ df. Large U values indicate that one or more of the U_i variances are large, reflecting an interaction. That is, $\sum U_i^2$, and hence, U , will be large when there is not homogeneity of rank variances. The maximum possible U is $U_{\max} = C(K - 1)$. When N (and, thus, C) is not large, there are several improved statistical tests (27). For example, under the null hypothesis,

$$F_U = \frac{(C-1)U}{C(K-1)-U} \quad (13)$$

has, approximately, an F distribution with numerator df equal to $df_N = (K - 1) - (2/C)$, and denominator df equal to $df_D = (C - 1)(21, 27)$. When C is large, df_N and df_D are approximately $(K - 1)$ and $(K - 1)(N - 1)$, respectively. Other tests based on U are given in Piepho (27).

For example A, where there were identical ranks of the original X data for each environment (Table 1), the ranks of the Y contrasts were not identical in the two new columns (Table 2). However, there was a general increase in ranks from the first to the fifth genotype. Test statistics were $U = 2.86$ ($P = 0.582$, based on the chi-square test; $df = 4$) and $F_U = 0.56$ ($P = 0.708$, based on the F test; $df_N = 3$, $df_D = 3$). Thus, there was no evidence of a significant interaction. It should be pointed out that the tests are not powerful for detecting the interaction with only $C = 2$ and $K = 5$ (27).

For example B, although there was no concordance of the original X data (i.e., there were opposite rankings in some of the environments; see Table 1), the ranks of the Y contrasts were identical in the two new columns (Table 2). In the pairs of environments considered, the differences in X values declined with increasing genotype label. The chi-square statistic was $U = 8.0$ ($P = 0.092$; $df = 4$). This is the largest possible value of U with $K = 5$ and $C = 2$. With this very small size data set, it is not possible to reject the null hypothesis at $P = 0.05$ using the chi-square test, even when there is perfect agreement in rankings within the columns. As stated above, the chi-square test is not accurate with a small number of locations (N), and should not be used with such small data sets (but is given here for completeness). The F_U value is undefined here since the denominator of equation 13 is 0. This is interpreted as a highly significant result ($P \approx 0$) (27). Thus, the lack of significant concordance ($W = 0$ for the original four environments; see above) can be attributed in this example to an interaction of genotype and environment rather than to experimental variation.

When N and K are very small ($N \leq 6$, $K \leq 6$), alternative testing can be done that does not depend on any assumptions about the distribution of the test statistic (chi-square or F). So-called "exact" significance levels can be calculated by determining all possible arrangements of the within-column ranks for a given number of rows and columns (Appendix in [27]). For example B, the exact P was 0.033 with the realized value of U , in qualitative agreement with the conclusion based on F_U . For example A, the exact P was 0.73, very similar to that obtained with the F statistic. With $K > 5$ and $N > 5$, the exact significance levels are not required, and the use of the F_U statistic is accurate.

Table 2. Two theoretical examples of contrasts (differences) of disease response between pairs of environments ($Y_{il} = X_{ij'} - X_{ij}$) for five genotypes ($i = 1, \dots, 5$) in two columns ($l = 1, 2$; based on $j = 1, \dots, 4$ original environments), ranking of Y_{il} within each column (R_{il}), mean rank of contrasts across columns ($\bar{R}_{i\bullet}$), and rank variance (U_i) of contrasts for each genotype^a

Example	Genotype (<i>i</i>)	Contrasts (<i>Y_{il}</i>)		Rank of contrasts (<i>R_{il}</i>)		Mean (<i>R_{i•}</i>)	<i>U_i</i>
		Columns (<i>l</i>)		Columns (<i>l</i>)			
		1	2	1	2		
A	1	15	62	1	2	1.5	2.5
	2	20	61	2	1	1.5	2.5
	3	35	64	3	3	3	0
	4	50	67	5	4	4.5	2.5
	5	40	70	4	5	4.5	2.5
B	1	95	78	5	5	5	4
	2	60	71	4	4	4	1
	3	35	64	3	3	3	0
	4	10	57	2	2	2	1
	5	−40	54	1	1	1	4

^a Original values in Table 1.

It is important to distinguish between the U test statistic here for assessing the Y contrasts, and those based on the original X values. A large U (or F_U) indicates significant interaction (heterogeneity of the U_i variances); in contrast, a large S , T , or W , as well as a small Ψ , indicates a significant concordance (agreement) of the rankings of X , and an implied small (or no) interaction.

Computation. Many statistical programs calculate the Friedman chi-square T statistic. From T , it is straightforward to calculate W , S , and Ψ , as well as the F statistic for concordance, using the number of rows (K) and columns (N) in the data set and the formulae presented above. F and the other statistics can also be obtained from a two-way analysis of variance of the within-environment ranks, with genotype and environment as fixed effects. The df values from the standard ANOVA output [$K - 1$, $(K - 1)(N - 1)$] are (somewhat) different from the ones given here, except for large K and N . The stability statistics of Piepho (27) require some specific manipulations not automatically performed in commercial statistics programs in order to obtain the desired statistics. One must first calculate the Y_{ij} contrasts of pairs (or the triplet) of environments (equations 10a,b) and the within-column ranks of the Y_{ij} values (R_{ij}). If one calculates $R_{ij}^* = [R_{ij} - (K + 1)/2]^2$ as a response variable, the U and F_U statistic can then be easily obtained using standard programs, with R_{ij}^* used instead of R_{ij} . For instance, performing a two-way analysis of variance on R_{ij}^* as the response variable, with genotype and column as fixed-effects, directly gives F_U . The df calculations need to be done separately.

Other statistics can be readily obtained from the calculated F_U . We have written a macro in SAS to perform all required calculations.

The formulae above are based on the absence of ties. With ties, some of the exact equivalences between statistics are only approximations. For instance, with ties, the sum of the $S_i^{(2)}$ values (Ψ) is only approximately equal to the functions of W and $\bar{r}$ in equation 7. Also, the sum of the U_i^2 values only approximately gives the U (and, hence, F_U) statistic in equation 12. It is most convenient to calculate W , T , U , F_U , and related statistics directly using ANOVA or Friedman-based algorithms, and approximate the sum of the $S_i^{(2)}$ and U_i^2 values based on the presented formulae. One can easily adjust for very large numbers of ties by modifying the df values for the chi-square or F tests (21,27).

MATERIALS AND METHODS

Data collection on gray leaf spot severity. Ten different maize genotypes ($K = 10$) adapted to southern climates were planted in 11 different locations (environments; $N = 11$) in 1996 (Table 3). There was a history of gray leaf spot epidemics at all locations. The locations were: Washington, IN (S1); Union City, TN (S2); Marion, NC (S3); Keedysville, MD (S4); Moffitt, IA (S5); Doniphan Co., KS (S6); West Lafayette, IN (S7); Andrews, NC (S8); Conrad, IA (S9); Whitethorne Plantation, VA (S10); and Apple Creek, OH (S11). In 1997, the study was repeated, but only four locations were used. Ten genotypes adapted to northern climates were planted in 10 different locations in 1996. The locations were: Ithica, NE (N1); Wooster, OH

(N2); Moffitt, IA (N3); West Lafayette, IN (N4); Marion, NC (N5); Keedysville, MD (N6); Conrad, IA (N7); Apple Creek, OH (N8); Whitethorne Plantation, VA (N9); and Doniphan City, KS (N10) (Table 4). Some of the locations were the same as for the southern genotypes. The experiment was repeated in 1997 with northern-adapted genotypes at five locations.

At each location, there was a single row of each genotype, planted at standard density for the location. Severity of gray leaf spot was assessed at approximately growth stage R4–R5 (dough-dent) using a 0 to 100% continuous scale. Ten plants of each genotype were arbitrarily selected in each row, and severity was assessed by visually estimating the percentage of the ear leaf area covered with lesions. The X_{ij} values in Tables 3 and 4 consist of the means of the 10 plants.

Simulation. Hühn et al. (16) showed that there can be a strong numerical linkage between parametric measures of $G \times E$ interactions and the nonparametric coefficient of concordance, W . Based on analysis of crop yield data from multiple locations (with an assumed normal distribution, at least as an approximation), they found that W could be predicted from:

$$\hat{W} = a + b \left(1 + \frac{\hat{\sigma}_e^2}{\hat{\sigma}_a^2} \right)^{-1} \quad (14)$$

where $\hat{\sigma}_a^2$ and $\hat{\sigma}_e^2$ are the estimated variances for genotype and residual (sum of interaction variance and mean experimental error variance), based on the model in the Introduction (assuming here that genotype is a random effect), and a and b are parameters estimated from the variance

Table 3. Disease severity (X_{ij}) for gray leaf spot of maize, caused by *Cercospora zea-maydis*, for 10 southern-adapted maize genotypes (i) in 11 environments (locations; $j = S1, \dots, S11^a$), rank of severity within each environment (R_{ij}), mean severity ($\bar{X}_{i\bullet}$) and mean rank across environments ($\bar{R}_{i\bullet}$), and rank stability variance ($S_i^{(2)}$)

	Genotype (i)	Environment (j)										Mean	Var. $S_i^{(2)}$
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	
X_{ij}	P3394	27	22	35	30	47	27	23	100	19	24	62	37.8 ... ^b
	FR1064xLH185	36	21	43	23	37	27	15	100	14	24	33	33.9 ...
	P3163	17	14	20	14	26	13	4	48	7	8	35	18.7 ...
	LH195xLH212	15	8	21	13	20	19	8	81	5	21	26	21.5 ...
	LH195xLH210	7	8	23	11	31	17	10	82	3	22	19	21.2 ...
	LH235xLH185	17	7	25	13	27	9	8	83	6	16	27	21.6 ...
	LH200xLH216	14	4	19	11	21	18	7	92	3	16	20	20.4 ...
	LH200xLH172	3	4	19	6	27	4	6	82	2	19	34	18.7 ...
	LH195xNC258	2	2	11	6	6	5	2	13	1	19	18	7.7 ...
	DK683	3	3	5	10	6	2	4	18	1	5	9	6.0 ...
R_{ij}	P3394	9	10	9	10	10	9.5	10	9.5	10	9.5	10	9.7 0.16
	FR1064xLH185	10	9	10	9	9	9.5	9	9.5	9	9.5	7	9.1 0.65
	P3163	7.5	8	5	8	5	5	2.5	3	8	2	9	5.7 6.27
	LH195xLH212	6	6.5	6	6.5	3	8	6.5	4	6	7	5	5.9 1.95
	LH195xLH210	4	6.5	7	4.5	8	6	8	5.5	4.5	8	3	5.9 3.09
	LH235xLH185	7.5	5	8	6.5	6.5	4	6.5	7	7	3.5	6	6.1 2.00
	LH200xLH216	5	3.5	3.5	4.5	4	7	5	8	4.5	3.5	4	4.8 2.17
	LH200xLH172	2.5	3.5	3.5	1.5	6.5	2	4	5.5	3	5.5	8	4.1 4.05
	LH195xNC258	1	1	2	1.5	1.5	3	1	1	1.5	5.5	2	1.9 1.79
	DK683	2.5	2	1	3	1.5	1	2.5	2	1.5	1	1	1.7 0.52

^a See text for description of the 11 locations. Data collected in 1996.

^b Only applicable to the ranks, not to the actual severities.

and concordance data using least squares regression. Based on the tested genotype data sets, Hühn et al. (16) found that $a \approx 0$ and $b \approx 1$. We explored the generality of this relationship by simulating disease intensity data according to a normal distribution with the following choices for the genotype and residual variances: $\sigma_a^2 = 1, 4, 9$, and 16 ; and $\sigma_e^2 = 1, 4, 9$, and 16 . An environmental variance (σ_b^2) of 2 was used. Thus, there were 16 combinations of genotype and residual variances for any K and N values chosen. For $K = 10$ genotypes, N values of $4, 5, 10, 15, 20, 25, 30$, and 40 were utilized. Moreover, the following (K, N) pairs were utilized: $(4, 4)$, $(5, 5)$, $(10, 10)$, $(15, 15)$, $(20, 20)$, $(25, 25)$, $(30, 30)$, and $(40, 40)$.

For each combination of K and N , 100 independent data sets were simulated for each of the pairs of variances, using the random number generation facility of SAS. PROC MIXED of SAS was then used to estimate the genotype, environment, and residual variance for each generated data set, and the macro (described above) was used to calculate W for each of the data sets. Then ordinary least squares regression was used to fit equation 14 to the function of the estimated variances. Appropriateness of equation 14 was evaluated using standard statistical methods (31).

RESULTS

Gray leaf spot. Test statistics are given for all four studies (Table 5), but details are given primarily for the southern-adapted genotypes in 1996, and to a lesser extent, for the northern-adapted genotypes in the same year. Disease severity varied considerably among the southern locations in

1996 (see Table 3), with the highest overall severity in S8, and the lowest severities in S2, S6, and S7. Genotypes also varied considerably in severity, with P2294 and RF1064xLH185 having the highest severities and LH195xNC258 and DK683 having the lowest. The rankings of genotypes for disease severity were fairly consistent among locations, but not identical. Mean ranks ranged from 9.7 down to 1.7 . Based on the $S_i^{(2)}$ values, it was clear that the third genotype, P3163, varied the most in ranking, with $S_3^{(2)} = 6.27$ (Table 3). The rank of this genotype ranged from 2 to 9 across the 11 environments. LH195xLH210 and LH200xLH172 also exhibited

fairly large variation in ranks. There was little variation in the ranks for the first or last genotype.

Kendall's coefficient of concordance was $W = 0.75$. Tests based on the chi-square (T) and F statistics (F) indicated highly significant concordance ($P < 0.001$) (Table 5). These test results equivalently indicate that the genotypes significantly differed in terms of mean ranks and that there was general agreement (concordance) in rankings among environments. Based on the use of equation 5, mean Spearman rank correlation was $\bar{r} = 0.72$. Seventy-five percent of the individual r values were above 0.63 , and the full range

Table 5. Statistics for testing for concordance of ranks and rank homoscedasticity of disease severity of gray leaf spot, caused by *Cercospora zea-maydis*, for four data sets ($K = 10$ genotypes each) grown in N environments^a

Statistic ^b	South-1996	North-1996	South-1997	North-1997
N	11	10	4	5
W	0.75	0.67	0.82	0.61
T	74.00	61.06	29.22	27.29
df for T	9	9	9	9
P for T	<0.0001	<0.0001	0.0006	0.0013
F	29.59	18.99	12.93	6.16
df _N , df _D for F	8.8, 88.2	8.8, 79.2	8.5, 25.5	8.6, 34.4
P for F	<0.0001	<0.0001	<0.0001	<0.0001
Ψ	22.9	29.5	20.7	40.6
Ψ^*	75.2	74.4	58.3	64.4
U	10.77	9.91	9.26	4.29
df for U	9	9	9	9
P for U	0.2917	0.2202	0.4132	0.8914
F_U	1.26	1.13	1.06	0.32
df _N , df _D for F_U	8.6, 34.4	8.6, 34.4	8.0, 8.0	8.0, 8.0
P for F_U	0.2949	0.3694	0.4678	0.9398

^a Studies involved southern-adapted genotypes in 1996 (South-1996) and 1997 (South-1997), and northern adapted genotypes in 1996 (North-1996) and 1997 (North-1997).

^b See text for description of all terms.

Table 4. Disease severity (X_{ij}) for gray leaf spot of maize, caused by *Cercospora zea-maydis*, for 10 northern-adapted maize genotypes (i) in 10 environments (locations; $j = N1, \dots, N10^a$), rank of severity within each environment (R_{ij}), mean severity ($\bar{X}_{i\bullet}$) and mean rank across environments ($\bar{R}_{i\bullet}$), and rank stability variance ($S_i^{(2)}$)

		Environment (<i>j</i>)										Mean	Var. <i>S_i⁽²⁾</i>
Genotype (<i>i</i>)		N1	N2	N3	N4	N5	N6	N7	N8	N9	N10		
<i>X_{ij}</i>	LH74xLH61	50	18	30	15	50	43	25	60	60	49	40.0	... ^b
	LH202xLH163	47	18	55	14	46	43	16	64	43	56	40.2	...
	N4242	5	6	35	9	44	23	8	40	24	27	22.1	...
	LH227xLH176	7	8	26	3	34	22	5	55	48	18	22.6	...
	LH198xLH176	5	7	28	4	34	10	3	51	40	13	19.5	...
	LH74xLH51	4	8	27	15	38	18	4	34	33	26	20.7	...
	P3489	6	6	25	11	28	13	7	30	22	26	17.4	...
	LH198xLH172	5	5	26	2	27	15	1	35	33	11	16.0	...
	DK546	2	2	24	10	18	8	4	33	23	13	13.7	...
	LH198xLH172	2	5	19	1	35	5	3	33	27	6	13.6	...
<i>R_{ij}</i>	LH74xLH61	10	9.5	8	9.5	10	9.5	10	9	10	9	9.4	0.4
	LH202xLH163	9	9.5	10	8	9	9.5	9	10	8	10	9.2	0.6
	N4242	5	4.5	9	5	8	8	8	6	3	8	6.4	4.0
	LH227xLH176	8	7.5	4.5	3	4.5	7	6	8	9	5	6.2	3.8
	LH198xLH176	5	6	7	4	4.5	3	2.5	7	7	3.5	5.0	3.0
	LH74xLH51	3	7.5	6	9.5	7	6	4.5	4	5.5	6.5	6.0	3.5
	P3489	7	4.5	3	7	3	4	7	1	1	6.5	4.4	5.8
	LH198xLH172	5	2.5	4.5	2	2	5	1	5	5.5	2	3.4	2.8
	DK546	1.5	1	2	6	1	2	4.5	2.5	2	3.5	2.6	2.6
	LH198xLH172	1.5	2.5	1	1	6	1	2.5	2.5	4	1	2.3	2.7

^a See text for description of the 11 locations. Data collected in 1996.

^b Only applicable to the ranks, not to the actual severities.

of pair-wise correlations was 0.33 to 0.97. The sum of the $S_i^{(2)}$ values was $\Psi = 22.9$, considerably below the critical value for significance at $P = 0.05$ ($\Psi^* = 75.2$) (Table 5). Thus, there was considerable evidence that genotype had a strong effect on disease severity and that the effect was fairly consistent across environments.

With $W < 1$, there could still be a rank interaction of genotype and environment in addition to a genotype effect on disease, although the large W indicates that the genotype main effect is dominant. Even with $W = 1$, the magnitude of differences in disease severity among genotypes could still depend on environment (even though the ranks of the original observations would be invariant to environment). The U stability measure of Piepho (27), and related statistics, allows for the assessment of interactions.

The Y contrasts and their ranks for the data set in Table 3 are shown in Table 6. The clear differences among genotypes in terms of the original X values were eliminated when determining the contrasts. The rankings of the contrasts varied considerably among the columns. For instance, rank 8 corresponded to genotypes 9, 1, 6, 3, and 8 for columns 1, 2, 3, 4, and 5, respectively. There also was a fairly narrow range of mean ranks of the contrasts (over all columns). The genotype variances (U_i) were fairly similar, with the smallest and largest values for LH235xLH285 and DK683, respectively. The U statistic (=10.77) was not significant, based on a chi-square test ($P > 0.20$) (Table 5). The F_U statistic, determined using equation 13, was also not significant ($P > 0.20$). Thus, there was no substantial evidence of a $G \times E$ interaction, based on the ranks of the contrasts.

Disease severities followed a similar pattern in the other data sets as found for the southern-adapted genotypes in 1996. For example, the ranking of genotypes for the northern-adapted genotypes in 1996 was similar in the 10 environments, even though the environments varied considerably in disease severity (Table 4). However, a few genotypes (e.g., N4242) varied a few

positions in ranking. With $W = 0.68$ (Table 5), concordance was high ($P < 0.001$) based on the results of the chi-square and F test statistics. The sum of the rank variances ($\Psi = 29.5$) was 40% of the critical value (Ψ^*) for significance at $P = 0.05$, supporting the (overall) consistent ranking of genotypes. As with the southern-adapted genotypes, there was still the possibility of an interaction since concordance was not perfect. Nevertheless, the U stability measure was not significant for the northern 1996 data set ($P > 0.30$, based on F_U) (Table 5). Therefore, there was no evidence of an interaction of genotype and environment.

There was significant concordance ($P < 0.01$) for the two 1997 data sets, indicating the generally consistent rankings across environments (Table 5). The U stability measures were not significant ($P > 0.40$), indicating a lack of evidence of an interaction.

Concordance and random-effect variances. In agreement with empirical evidence (16), there was a strong linear relationship between estimated W based on rank agreement among environments and the function of the estimated genotype and residual variances [$f = (1 + (\hat{\sigma}_e^2 / \hat{\sigma}_a^2))^{-1}$] for the original normally distributed data (equation 14). An example for $K = 10$ and $N = 10$ is given in Figure 1A. The relationship between estimated W and the variance ratio ($\hat{\sigma}_e^2 / \hat{\sigma}_a^2$) was highly nonlinear (Fig. 1B), with a declining concordance coefficient with increasing ratio.

The estimated a and b parameters depended on N , but not on K (L. V. Madden, unpublished), and results are shown graphically for five N values for the cases where $K = N$ (Fig. 1C). The intercept parameter declined toward 0, and the slope parameter increased toward 1, as N increased. Thus, the approximation in Hühn et al. (16), that is, $a \cong 0$ and $b \cong 1$, primarily applies with large numbers of environments. The variability also declined as N increased (L. V. Madden, unpublished). Part of the high variability at small N could be attributed to poor precision of estimated

variances at small sample sizes (31). This could be responsible for the higher a and lower b at small N .

The effect of genotype and residual (including interaction) variances on predicted concordance can be seen by using equation 13 for selected $\hat{\sigma}_e^2 / \hat{\sigma}_a^2$ ratios. For ratios of 9/1, 1, and 1/9, one obtains predicted W values of 0.18, 0.52, and 0.87, respectively, at $K = N = 10$. Thus, as indicated by Hühn et al. (16), increasing magnitude of the genotype variance relative to the residual (= interaction + mean experimental error) variance leads to increasing agreement of the rankings of the observed data. When there is relatively little interaction effect on X (i.e., residual variance much smaller than genotype variance; $\hat{\sigma}_e^2 / \hat{\sigma}_a^2$ near 0), rankings will be almost identical and W will be near 1; in contrast, when there is relatively little genotype effect (i.e., genotype variance much smaller than residual variance; $\hat{\sigma}_e^2 / \hat{\sigma}_a^2$ considerably above 1), rankings will vary considerably among environments and W will be near 0.

After transforming X_{ij} to $\text{logit}(X_{ij})$ in order to obtain a more constant variance, PROC MIXED of SAS was used to estimate the genotype, environment, and residual variances for the four gray leaf spot data sets. The $\hat{\sigma}_e^2 / \hat{\sigma}_a^2$ ratios were 0.682, 0.640, 0.388, and 1.087 for southern-adapted genotype study in 1996, northern-adapted genotype study in 1996, southern-adapted genotype study in 1997, and northern-adapted genotype study in 1997, respectively. This resulted in predicted W values (based on equation 14) of 0.61, 0.62, 0.79, and 0.55 for the four studies. These values were fairly close to the directly calculated W s (Table 5), except for the southern-adapted genotype study of 1996. Assessment of the residuals from the linear model fit indicated, however, that the southern-1996 data were not normally distributed with the logit or other (e.g., angular) attempted transformations.

DISCUSSION

The $G \times E$ interaction can be quantified in several useful ways, using parametric and

Table 6. Contrasts of disease severity (Y_{ij}) for gray leaf spot of maize, caused by *Cercospora zea-maydis*, for 10 southern-adapted maize genotypes (i) in five columns of contrasts ($l = C1, \dots, C5$)^a, rank of severity contrast within each column (R_{il}), mean rank of contrasts across columns ($R_{i\bullet}$), and rank variance of contrasts (U_i)

Genotype	Contrasts (Y_{il})					Contrast ranks (R_{il})					Mean ($R_{i\bullet}$)	U_i
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5		
P3394	-21	17	-4	-81	38	5	8	6	3	10	6.4	6.6
FR1064xLH185	-29	14	-12	-86	9	3	6.5	1	2	6	3.7	8.0
P3163	-9	12	-9	-41	27	9	5	4	8	9	7.0	6.6
LH195xLH212	-19	7	-11	-76	5	7	3	2.5	7	5	4.9	4.0
LH195xLH210	-31	20	-7	-79	-3	1.5	9	5	5	1	4.3	9.8
LH235xLH185	-26	14	-1	-77	11	4	6.5	8	6	7	6.3	2.4
LH200xLH216	-20	10	-11	-89	4	6	4	2.5	1	3.5	3.4	7.2
LH200xLH172	-31	21	2	-80	15	1.5	10	9.5	4	8	6.6	12.2
LH195xNC258	-18	0	-3	-12	-1	8	2	7	10	2	5.8	10.6
DK683	-4	-4	2	-17	4	10	1	9.5	9	3.5	6.6	14.6

^a Y_{il} contrast for C1 determined from equation 10b (first three original environments), and contrasts for C2-C5 determined from equation 10a. Original observations in Table 3.

nonparametric methods, as well as by using biplots and other graphic approaches (9,11, 14,15,17,19,23–25,28,33,37–39). There are many possible forms to the interaction, and the different methods may quantify (or emphasize) different components of the interaction. Besides being robust to violations of statistical assumptions regarding

the distribution of the response variable (disease severity in this situation), and insensitive to outliers, nonparametric rank-based approaches are of value for elucidating certain meaningful ways that environments (in the broad sense) differentially affect (or not affect) the measured response (14). In particular, Kendall's coefficient of

concordance (W) and related statistics described in this article (S , T , Ψ) can identify the degree of agreement (concordance) of rank order of genotypes among environments. These statistics specifically quantify the magnitude of the genotype effect on rank order relative to the residual (interaction and experimental error) effect; by the nature of ranks, however, the stronger the effect of genotype, the weaker the effect (or measurable effect) of an interaction or experimental error on the rank orders. For the four data sets considered here, all exhibited significant concordance ($0.61 \leq W \leq 0.81$; Table 5), but none showed identical ranking within environments.

Finding significant concordance using global tests based on W (T , F , or Ψ) may not be the only interest of the investigator. It may be of equal importance for some studies to determine which individual genotypes are the most stable (or unstable) across environments in terms of rank order. The Nassar-Hühn $S_i^{(2)}$ statistic (17,24)—the component of the Ψ statistic—allows for easy appraisal of the consistency of genotype rankings (Tables 3 and 4). Because of the significant W for the data sets analyzed here, there were not many unusually large $S_i^{(2)}$ values. Approximate tests given in Nassar and Hühn (24) and Hühn (14) can be used to test whether these variances are different from their expected value under rank variance homogeneity. For gray leaf spot disease severity with the southern-adapted genotypes in 1996, P3163 (Table 3) had the largest $S_i^{(2)}$. However, the $S_i^{(2)}$ value for this genotype in 1997 was only the fourth largest of the 10 (L. V. Madden, *unpublished*). For the northern-adapted genotypes (Table 4), none of the $S_i^{(2)}$ was particularly large. This supports the global result that genotypes were stable across environments in terms of ranks.

Because $1 - W$ is the degree of discordance of the data, it can be seen from equation 7 that Ψ is a scaled measure of discordance, which means that the individual $S_i^{(2)}$ value represents the genotype's contribution to the discordance of a data set (18). $S_i^{(2)}$ is analogous to Wricke's ecovalence, which is sometimes calculated as a parametric measure of phenotype stability (14,23). This is because the sum of the ecovalence values equals the interaction sum of squares in a linear model, a measure of the differential effects of environments on the genotype effects (a form of concordance). The ecovalence values also determine the stability variance values (σ_i^2) of Shukla (33) discussed in the Introduction (see equation 2.2 in [25]), thus providing further analogy between the parametric and nonparametric approaches to the problem.

As shown by Hühn et al. (16), and in Figure 1, there is more than just analogy between the parametric and nonparametric

Fig. 1. Estimated and predicted nonparametric concordance coefficient (W ; equation 4) in relation to estimates of ratio of residual (interaction and error) to genotype variance from a random-effects linear model, based on 100 simulations of a normal distribution with 16 different prespecified combinations of the genotype and residual variance. **A**, W versus a nonlinear function (f) of the variance ratio (see equation 14), together with predictions based on equation 14, and the 95% prediction interval from regression analysis. **B**, W in relation to the variance ratio (on a log scale), with the same predictions and prediction interval as in **A**. **C**, Predictions of W from f , for five different numbers of environments, based on regression analysis.

approaches. That is, for normally distributed data, W is numerically related (in a nonlinear manner) to the ratio of the residual (including interaction) variance to the genotype variance (equation 13). Knowledge of the relative residual variance of the original observations allows one to predict the degree of concordance of the ranks, or knowledge of W allows one to predict, on a relative scale, the residual variance of the X values. Our simulation results confirm the general nature of equation 13, but showed for the first time that the slope (b) and intercept (a) both depend on number of locations (N), approaching 1 and 0, respectively, as N increased. As expected from the simulations, W increased with declining ratios of residual to genotype variances for the gray leaf spot data. At large N (e.g., >10), it would be reasonable to assume $a = 0$ and $b = 1$. Under these circumstances, one can write:

$$\frac{1-W}{W} \approx \frac{\hat{\sigma}_e^2}{\hat{\sigma}_a^2} \quad (15)$$

This means that the discordance ($1 - W$), expressed in units of concordance (W), is approximated simply by the ratio of the residual and genotype variances of the X values (14).

By nature of the linkage between W and the Friedman T statistic (12,21), significant concordance of ranks among environments indicates that genotype affects ranks, and that there are two or more genotypes with significantly different mean ranks. One can calculate a rank-based least significant difference (LSD) as:

$$\text{LSD} = z\sqrt{K(K+1)/(6N)} \quad (16)$$

where z is the standard normal statistic at the chosen P value ($z = 1.96$ for $P = 0.05$). With $K = 10$ genotypes, the LSD is 2.53 with $N = 11$ (Table 3) or 2.65 with $N = 10$ (Table 4). For instance, with the southern-adapted genotypes in 1996, two genotypes are significantly different if their mean ranks (across locations) are greater than 2.53; the first two genotypes in Table 3 are not significantly different, but both of these are different from the third genotype. An implicit assumption of the standard Friedman test is that the ranks within environments are equally correlated (4). This may be overly restrictive when the primary objective is to quantify the effects of genotypes on disease or other responses across a number of environments. The recently developed nonparametric methods of Brunner, Akritas, Puri, and colleagues (3,4,32) for analyzing data from factorial layouts are more general than the Friedman test methods, and allow for more complicated correlation structures. This so-called marginal effects analysis may be the preferred nonparametric approach whenever the primary objective is to test equality of mean ranks of genotypes (which was not our objective in this pa-

per). The $G \times E$ two-way layout here corresponds to the LD-F1 layout presented in Brunner et al. (3), where environment is equivalent to "subject", and genotypes (treatments) are randomized within each subject.

Piepho (27) provided an innovative extension of the concordance correlation approach with the development of a test for homoscedasticity in any two-way layout of data. Overcoming the limitations of the alignment methods (17,24), contrasts involving pairs (or triplets) of environments are calculated to remove the genotype effect. Within-column ranks of the contrast data provide a direct test of genotype stability across environments when the effect of genotype is removed (i.e., for testing the equality of the stability variances). This may be of special interest to the investigator whenever W is significant but not very close to 1. The U or F_U test is not very powerful statistically for detecting nonhomogeneity of variances (27) relative to parametric tests of the original X observations. This is partly because only interactions strong enough to result in consistent rankings of the Y contrasts in each column will produce large U or F_U values. However, this type of extreme interaction may be of primary interest to the investigator. An example of a situation with nonhomogeneity of variances is the barley net blotch data of Pinnschmidt and Hovmöller (29). We analyzed the disease severity data in Table 6 of their paper ($K = 7$, $N = 27$) and found a highly significant nonhomogeneity of variances using the Piepho (27) method ($U = 28.3$; $F_U = 6.85$, $df_N = 5.8$, $df_D = 70.2$, $P < 0.001$) (L. V. Madden, unpublished). Reducing the number of columns of the two-way table of observations from N to $C = N/2$ (for even number of environments) also results in low power for the variance homogeneity test discussed here. For these reasons, it may be desirable for other studies with $N < 10$ to test for significance at a critical P value of 0.10 rather than 0.05. The smallest achieved P value for the gray leaf spot data was 0.29, so none of the results were borderline for these studies.

In conclusion, using rank-based methods for concordance and homoscedasticity, there were generally consistent rankings of genotypes for gray leaf spot severity across environments, and no evidence that environment affected the rank order. Thus, the data analyzed here suggested that differences in pathogen populations (or aggressiveness) or weather conditions among locations were not great enough to affect the rank order of genotypes in terms of disease responses. Although Carson et al. (7), for example, did find a significant $G \times E$ interaction for gray leaf spot in North Carolina, based on a parametric (i.e., normal distribution assumption) analysis, their results are not inconsistent with ours. As stated by Carson et al. (7), "most of the

genotype-environment interaction appears to result from changes in the magnitude of differences among hybrids between locations, rather than changes in rankings." Gordon et al. (13) came to similar conclusions regarding $G \times E$ interaction for gray leaf spot across sites at Wooster, OH, and Cedara, Republic of South Africa. The rank-based methods serve as convenient tools to specifically detect situations where the ranks do change with environment. The methods discussed here can be used for any study where the same set of treatments (e.g., fungicides) are tested in each of N environments (different locations and/or years). A significant fungicide-by-environment ($F \times E$) interaction would suggest differential effectiveness of tested fungicides under different environments (broadly defined), possibly due to development of resistance in the pathogen population to one or more of the tested compounds.

There can be several interrelated causes for a $G \times E$ (or, more generally, treatment-by-environment) interaction, including differential effects of the physical environment (e.g., weather) on the plant-microbe interaction at the population scale, and differences in aggressiveness (or other interrelated traits) of the local pathogen population among environments. Use of ancillary information (e.g., results on diversity and population structure of the pathogen at individual locations, or local weather measurements) may give evidence for one cause over another, although it is unlikely for there to be a single cause for an interaction. Additional studies may be warranted when an interaction is found, to clarify whether or not there is a dominant factor responsible for the interaction. Controlling the physical environment (using growth-chamber studies) or the pathogen strain used in inoculating plants can lead to improved understanding of $G \times E$ interactions.

ACKNOWLEDGMENTS

We thank the following individuals for supplying gray leaf spot data: E. Stromberg, D. Dowden, B. Anderson, D. Jardine, A. Grybauskas, J. Partridge, L. Wysong, L. Dunkle, D. Scott, C. Martinson, G. Munkvold, and G. Shaner. Salaries and research support were provided by state and federal funds to the Ohio Agricultural Research and Development Center, The Ohio State University.

LITERATURE CITED

1. Beasley, T. M. 2002. Multivariate aligned rank test for interactions in multiple group repeated measures designs. *Multivar. Behav. Res.* 37:197-226.
2. Brindley, R. D., and Bradley, R. A. 1985. Some new results on Grubbs' estimators. *J. Am. Stat. Assoc.* 80:711-714.
3. Brunner, E., Domhof, S., and Langer, F. 2002. *Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. John Wiley & Sons, New York.
4. Brunner, E., and Puri, M. L. 2001. Nonparametric methods in factorial designs. *Stat. Pap.* 42:1-52.
5. Cardwell, K. F., and Wehrly, T. E. 1997. A rank

- test for distinguishing environmentally and genetically induced disease resistance in plant varieties. *Biometrics* 53:195-206.
6. Carson, M. L., and Goodman, M. M. 2006. Pathogenicity, aggressiveness, and virulence of three species of *Cercospora* associated with gray leaf spot of maize. *Maydica* 51:89-92.
7. Carson, M. L., Goodman, M. M., and Williamson, S. M. 2002. Variation in aggressiveness among isolates of *Cercospora* from maize as a potential cause of genotype-environment interaction in gray leaf spot trials. *Plant Dis.* 86:1089-1093.
8. Clements, M. J., Dudley, J. W., and White, D. G. 2000. Quantitative trait loci associated with resistance to gray leaf spot of corn. *Phytopathology* 90:1018-1025.
9. Eberhart, S. A., and Russell, W. A. 1966. Stability parameters for comparing varieties. *Crop Sci.* 6:36-40.
10. Garrett, K. A., Madden, L. V., Hughes, G., and Pfender, W. F. 2004. New applications of statistical tools in plant pathology. *Phytopathology* 94:999-1003.
11. Gauch, H. G. 1988. Model selection and validation for yield trials with interaction. *Biometrics* 44:705-715.
12. Gibbons, J. D. 1985. *Nonparametric Methods for Quantitative Analysis*. 2nd ed. American Sciences Press, Columbus, OH.
13. Gordon, S. G., Bartsch, M., Matthies, I., Gevers, H. O., Lipps, P. E., and Pratt, R. C. 2004. Linkage of molecular markers to *Cercospora zeae-maydis* resistance in maize. *Crop Sci.* 44:628-636.
14. Hühn, M. 1996. Nonparametric analysis of genotype $\times$ environment interactions by ranks. Pages 235-271 in: *Genotype by Environment Interaction*. M. S. Kang and H. G. Gauch, Jr., eds. CRC Press, Boca Raton, FL.
15. Hühn, M., and Leon, J. 1995. Nonparametric analysis of cultivar performance trials—Experimental results and comparison of different procedures based on ranks. *Agron. J.* 87:627-632.
16. Hühn, M., Lotito, S., and Piepho, H. P. 1993. Relationships between genotype $\times$ environment interactions and rank orders for a set of genotypes tested in different environments. *Theor. Appl. Genet.* 86:943-950.
17. Hühn, M., and Nassar, R. 1989. On tests of significance for nonparametric measures of phenotypic stability. *Biometrics* 45:997-1000.
18. Hühn, M., and Piepho, H. P. 1994. Relationships between Kendall's coefficient of concordance and a nonparametric measure of phenotypic stability with implications for the consistency in rankings as affected by variance components. *Biometr. J.* 36:719-727.
19. Hussein, M. A., Bjornstad, A., and Aastveit, A. H. 2000. SASG $\times$ ESTAB: A SAS program for computing genotype $\times$ environment stability statistics. *Agron. J.* 92:454-459.
20. Kang, M. S., and Gauch, H. G., Jr., eds. 1996. *Genotype by Environment Interaction*. CRC Press, Boca Raton, FL.
21. Kendall, M. G., and Gibbons, J. D. 1990. *Rank Correlation Methods*, 5th ed. Edward Arnold, London.
22. Kowalchuk, R., K., Keselman, H. J., and Algina, J. 2003. Repeated measures interaction test with aligned ranks. *Multivar. Behav. Res.* 38:433-461.
23. Lin, C. S., Binns, M. R., and Lefkovich, L. P. 1986. Stability analysis: Where do we stand? *Crop Sci.* 26:894-900.
24. Nassar, R., and Hühn, M. 1987. Studies on estimation of phenotypic stability: Test of significance for nonparametric measures of phenotypic stability. *Biometrics* 43:45-53.
25. Piepho, H. P. 1996. Analysis of genotype-by-environment interaction and phenotypic stability. Pages 151-174 in: *Genotype by Environment Interaction*. M. S. Kang and H. G. Gauch, Jr., eds. CRC Press, Boca Raton, FL.
26. Piepho, H. P. 1996. Analyzing genotype-environment data by mixed models with multiplicative effects. *Biometrics* 53:242-252.
27. Piepho, H. P. 1997. Distribution-free tests for one-way homoscedasticity in a two-way classification. *Biometrics* 53:340-351.
28. Piepho, H. P. 1999. Stability analysis using the SAS system. *Agron. J.* 91:154-160.
29. Pinnschmidt, H. O., and Hovmöller, M. S. 2002. Genotype $\times$ environment interactions in the expression of net blotch resistance in spring and winter barley varieties. *Euphytica* 125:227-243.
30. Saghai Maroof, M. A., Zue, Y. G., Xiang, Z. X., Stromberg, E. L., and Rufener, G. K. 1996. Identification of quantitative trait loci controlling resistance to gray leaf spot disease in maize. *Theor. Appl. Genet.* 93:539-546.
31. Schabenberger, O., and Pierce, F. J. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press, Boca Raton, FL.
32. Shah, D., and Madden, L. V. 2004. Nonparametric analysis of ordinal data in designed factorial experiments. *Phytopathology* 94:33-43.
33. Shukla, G. K. 1972. Some statistical aspects of partitioning genotype-environment components of variability. *Heredity* 29:237-245.
34. Shukla, G. K. 1982. Testing the heterogeneity of variances in a two-way classification. *Biometrika* 69:411-416.
35. Turechek, W. W. 2004. Nonparametric tests in plant disease epidemiology: Characterizing disease associations. *Phytopathology* 94:1018-1021.
36. Ward, J. M. J., Stromberg, E. L., Nowell, D. C., and Nutter, F. W., Jr. 1999. Gray leaf spot: A disease of global importance in maize production. *Plant Dis.* 83:884-895.
37. Yan, W. 2001. GGEbiplot—A Windows application for graphical analysis of multi-environment trial data and other types of two-way data. *Agron. J.* 93:1111-1118.
38. Yan, W., and Falk, D. E. 2002. Biplot analysis of host-by-pathogen data. *Plant Dis.* 86:1396-1401.
39. Yan, W., and Hunt, L. A. 2001. Interpretation of genotype $\times$ environment interaction for winter wheat yield in Ontario. *Crop Sci.* 41:19-25.