Sorting of Infection-Type Data Sets Toward the Gene-for-Gene Model: A Reply

L. E. Browder and M. G. Eversmeyer

Research plant pathologists, U.S. Department of Agriculture, Agricultural Research Service, Department of Plant Pathology, Kansas State University, Manhattan 66506.

Cooperative investigations of Agricultural Research Service, U.S. Department of Agriculture and the Kansas Agricultural Experiment Station, Department of Plant Pathology, Contribution 81-506-j.
Accepted for publication 23 September 1981.

Knott and Johnson (3) questioned the validity of our procedure (2) for sorting cereal rust infection-type (IT) data sets and suggested another procedure. They also questioned our analysis of a demonstration data set. Although they (3) listed our objectives, further discussion of these objectives may be useful. Our stated objective of determining the minimum number of corresponding gene pairs (CGPs) that would explain the IT variation, although listed first, probably was the least important. The more important objective was to compare IT spectra of lines with known and unknown reaction genotypes to develop hypotheses of similarities and differences. Comparison of IT spectra, at various levels of formal organization, is the only method that can be used to make such hypotheses.

In meeting the objectives of our study, we used 10 lines of wheat (*Triticum aestivum* L. emend. Thell.) known to have different genes for low reaction to *Puccinia recondita* Rob. ex. Desm. (*Lr* genes). We used the lines in different ways, depending on the objective. In demonstrating the minimum number of CGPs, the lines with known *Lr* genes were considered to be of unknown genotype. Knott and Johnson (3) correctly pointed out that two low ITs on the same host line produced by different cultures cannot be shown, by the diagonal check method, to result from the same CGP. Thus, a particular low IT used in the diagonal check can only be presumed to be effected by the known *Lr* gene in the line and its corresponding gene for low pathogenicity (*Lp* gene). We (2) did not presume that specific low ITs were effected by known CGPs in ascertaining minimum numbers of CGPs; such presumptions are not necessary at that level of analysis. Whether the host materials have known *Lr* genes or not is irrelevant to using the diagonal check method. In analyses to show similarities of host lines with known and unknown reaction genotype, we used characteristic low ITs (1) to indicate the various known CGPs.

The theoretical gene-for-gene models were constructed with premises based on the best information available from previous experimentation. Thus, they are based, at least indirectly, on real data. These models are complete representations of what would happen if all possible genotypes were present. We (2) used the Person/Habgood genotype arrangement proposed by Robinson (7) but added the premise that different low ITs are effected by different CGPs. This results in a model of IT data that can be sorted. The IT data in a real data set also can be sorted, but we would not expect to sort a real, incomplete data set to an exact theoretical model, only toward the model. If the model is based on correct premises, sorting a real data set will bring the set as close to the theoretical model as the real data set is complete and accurately coded. We sort real data sets toward a model to facilitate developing hypotheses about the host and parasite materials for further testing. A real data set cannot be sorted to either the Person model (6) or the Person/Habgood model (7) by sorting on the basis of number of high ITs in rows and columns. Neither a real data set nor a theoretical data set can be sorted to the Person model (6) by sorting two-class IT data per se, simply because there is no way to relate phenotypes to particular genotypes.

The essence of the gene-for-gene relationship is that a definitive host gene interacts with a definitive parasite gene to produce a definitive phenotype (4). This can be portrayed most simply in the cereal rust systems as:

		Parasite genotype	
		Lp	*Hp*
Host genotype	*Lr*	low IT	high IT
	Hr	high IT	high IT

in which *Lr* and *Hr* symbolize alternate host alleles for reaction at a locus, *Lp* and *Hp* symbolize alternate parasite alleles for pathogenicity at a matching locus, and low IT and high IT symbolize differences in host:parasite association phenotypes. In this case, *Lr*, *Lp*, and low IT are definitive; when two of the three are known, the other can be derived unequivocally. Neither *Hr* nor *Hp* can be detected without establishing *Lp* or *Lr* as a known constant in the other association member. This model is equivalent to Loegering and Burton's (5) box A, which Knott and Johnson (3) accept. *This is the gene-for-gene model*, showing the variation due to one CGP. The four isomers of this matrix are equivalent and, according to Loegering and Burton (5), "would not represent any basic difference in analysis." The isomers, listing the IT portion only, are:

Isomer A:		
	L	H
	H	H

Isomer B:		
	H	H
	L	H

Isomer C:		
	H	H
	H	L

Isomer D:		
	H	L
	H	H

Any sorting procedure that moves high ITs toward any corner of a data set and low ITs toward the opposite corner sorts the data toward the gene-for-gene model. We now believe that sorting high ITs toward the lower right corner (Isomer A) rather than the upper right is more logical, because all genetic analyses proceed from the demonstration of the definitive interaction phenotype; in this case, the low IT.

We believe that Knott and Johnson's (3) contention that we did not conclusively show five CGPs is the most important issue raised because their other questions seem to stem largely from this. Knott and Johnson do accept, and use as the basis for their data analysis, Loegering and Burton's box E to show that "two CGPs are operating." We also used Loegering and Burton's box E configuration and portrayed it in what we termed the diagonal check.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

This article is in the public domain and not copyrightable. It may be freely reprinted with customary crediting of the source. The American Phytopathological Society, 1982.

The diagonal check is logically derived from the gene-for-gene model. Restated, the diagonal check shows that two low ITs in a diagonal of a 2 × 2 matrix with a high IT in either of the other corners must result from the interaction of at least two CGPs, *regardless of whether the fourth IT is low, high, or missing*. We (2) showed a logical proof for this which Knott and Johnson did not question, but they contend our evidence that "the CGP for *Lr24*" was not complete "because all cultures gave a LIT on LR24(Agent)." We did not state that we had demonstrated *Lr24*; we concluded, in an analysis of our sorted table, that a low IT on LR24(Agent) produced by culture 6B-NA65-9, along with high ITs produced by that culture on the host lines having the first four *Lr* genes, demonstrated a fifth CGP. That is so because a series of diagonal checks shows that all five differ. The ITs on LR24(Agent) produced by cultures other than 6B-NA65-9 need not be known to draw our conclusions. Our analysis is in keeping with Knott and Johnson's own stated standard (3); so indeed, we demonstrated five CGPs.

Knott and Johnson (3) seem correct that, using their sorted Table 1, LR24(Agent) cannot be conclusively demonstrated not to carry a combination of *Lr* genes in LR2A(TC) and LR16(TC). The failure to show LR24(Agent) to be involved in a CGP in their analysis, however, is a failure to show that LR24(Agent) is involved in a sixth CGP rather than in a fifth CGP. A CGP had already been demonstrated to be associated with LR2A(TC) and with four other lines. The low IT produced by 6B-NA65-9 on any one of seven lines, including LR24(Agent), can be used to demonstrate the fifth CGP, if another has not already been used. There is no reason the first, on LR2A(TC), or any other must be used. Thus, we can readily demonstrate a CGP associated with LR24(Agent) after Knott and Johnson's sorting also. We accept Knott and Johnson's analysis of five CGPs as a minimum needed to explain the variation in our data set. That, indeed, was our conclusion. Different sets of five CGPs can be demonstrated, all correctly, depending on sorting method, original order, and criteria used to classify ITs.

In our sorting procedure we used the concept that two CGPs can effect different definitive ITs that are both low in relation to the 99P IT. One of the definitive ITs, however, can be low, the other high, in relation to one another. Numeric coding of ITs places them in relative classes. A 23X IT is high compared with 01C IT, but low compared with 99P IT. This seems obvious when stated in this way, but Knott and Johnson (3) apparently fail to recognize the validity and usefulness of this concept; even though they allude to it in their process of ordering lines and cultures by placing "those that have the highest LITs first" when lines and cultures had the same number of 99P ITs. Because 23X IT is high compared with 01C IT, an ascending sort will order them low to high. This ordering of ITs also systematizes the genotypes effecting them. Two 01C ITs can be shown to result from different CGPs if they are in a diagonal of a 2 × 2 matrix with a 23X IT in either other corner. For example, in the 2 × 2 data subset, Chicoro 'S':UN17-68A, Chicoro 'S':6B-NA65-9, and LR9(TC):6B-NA65-9 in Knott and Johnson's Table 1 (3) show that Chicoro 'S' and LR9(TC) are different and that UN17-68A and 6B-NA65-9 are different.

Although Knott and Johnson (3) correctly point out that our procedure was not general, their statement that the final order of host lines and cultures after sorting with our procedure "depends largely on the differences among the ITs in the last row and last column of the unsorted data set" is incorrect. Every value in every row and every column influences the final sorted order because all rows and all columns are sorted sequentially and "when several entries in one row and column have identical ITs, their order remains the same as it was in the preceding row or column." The order of cultures usually changes before host-line order is sorted and host-line order usually changes before culture order is resorted. Thus, their statement does not hold. Further, our previously published procedure sorts the first column of the data set last.

Knott and Johnson (3) suggest a procedure, based on Person's model (6), in which host lines and cultures are first ordered by the relative number of high ITs with which they are associated. Within that order, they sorted host lines and cultures associated with the same number of high ITs according to the coded value of the low ITs.

Three major problems arise with that procedure: (a) The method appears workable with small data sets with few CGPs operating and relatively few of the possible genotypes represented. Person's model (6), however, is arranged so that geometrically increasing numbers of lines and cultures occur in groups of more than one having the same number of high ITs when the number of CGPs is increased. In an eight-CGP model, 99.2% (254 ÷ 256) of lines and/or cultures occur in groups of more than one having the same number of high ITs. When sorting randomly ordered lines and cultures on the basis of numbers of high ITs only, there is no way to sort this 99.2% of the lines and cultures beyond random order within their groups. It is clear that the lines and cultures would remain in their original order within such groups. If a procedure does not sort permutations of a theoretical model it cannot adequately sort real data sets. (b) Knott and Johnson attempt to solve the problem of ordering lines and cultures within such groups by placing "those having the highest LITs first." Their Table 1 (3) indicates they did not use their rule algorithmicly because, if they had, Chicoro 'S' would occur before LR24(Agent), culture 6B-NA65-9 would occur before UN17-68A and 0967-1; and culture 0967-1 would occur before UN17-68A. The rule of placing ITs, not low ITs only, with the highest value first or last depending on the sort, is precisely the rule that we use algorithmicly. It is not logical to use such a rule in a subroutine applied to the low ITs of a two-class coding only; it should apply to all ITs. Knott and Johnson's suggested method (3) would be more difficult to program than ours. (c) Even if their procedure were programmed and applied algorithmicly, the first part of their procedure separates lines and cultures associated with similar low ITs when they differ in numbers of high ITs. Knott and Johnson's sorting procedure separated Waldron and LR2A(TC), which probably are related, and also separated Chicoro 'S' and LR1(TC), which probably are related.

Knott and Johnson (3) correctly point out that the PL/1 program that we detailed in our previous paper (2) does not return all permutations of a theoretical data set to the model form. Thus, our previously described program was, in fact, inadequate. We have, however, revised our program by simply extending the same kind of sorting operations with program loops. This revised program has not failed to return any of 30 randomly derived permutations of the theoretical model to an isomer of the Person/Habgood (7) genotype arrangement. The number of possible permutations of a two-dimensional matrix is very large; eg, 576 for a 4 × 4, two-CGP model. This precludes any proof of the correctness of a sorting procedure by simply testing it by a sample of different permutations. Confidence in the procedure must be based on the logic of the approach; ie, that if a low IT is epistatic to all higher ITs, then sorting of numerically coded data would systematize host and parasite materials associated with the ITs.

Knott and Johnson's analysis of our data in their Table 1 is correct insofar as they show that two different sets of five CGPs can be demonstrated, depending on whether one chooses to use LR1(TC):UN17-68A or LR17(TC):UN17-68A to demonstrate the third CGP. Neither of their sets is the same as the set of five CGPs which we demonstrated; we believe all three sets are correct. At least 56 different sets of culture:line combinations can be used to correctly demonstrate five CGPs, but never more or never less than five CGPs. Thus, neither we (2) nor Knott and Johnson (3) used the full power of the diagonal check. The fact that these 56 different sets can be found indicates that the cultures represented in our data set were inadequate to demonstrate more than five CGPs of the 10 CGPs known to occur in the materials, rather than invalidating either the diagonal check or the principle of sorting IT data.

We disagree with Knott and Johnson (3) that there is any significant probability that LR17(TC) has *Lr1* because if it did, the IT on LR17(TC) produced by cultures UN01-68B, UN01-68A, and UN17-68A, would have been 01C. Our conclusion assumes LR1(TC) has only *Lr1*. We disagree that there is any significant probability that LR16(TC) has *Lr2a*. If LR16(TC) has *Lr2a*, it would have had a 01C IT produced by cultures 65359-01, UN01-68A, and UN01-68B. Our conclusion assumes LR2A(TC) has only

Lr2a. We also disagree that Waldron is likely to have *Lr16* because of the 99P IT on Waldron produced by UN17-68A. If LR16(TC) has only *Lr16* and Waldron has *Lr16*, the IT produced by UN17-68A on Waldron would have been 34N or less.

In summary, we find that our previously published procedure does not bring all permutations of a theoretical model to one form. This, however, does not invalidate the concept of sorting coded IT data toward the gene-for-gene model. It has been adequately demonstrated that different CGPs effect different low ITs (1), and that low IT is epistatic (4). Accepting these as premises, it is elementary that systematically bringing like ITs together will bring like host genotypes together and like parasite genotypes together. We have revised our program to bring all permutations of a theoretical data set to the model form insofar as we have tested it. We do not see, for the reasons given earlier, that Knott and Johnson can develop such a procedure using the principles that they indicate in their paper (3).

Knott and Johnson obviously are approaching IT data analyses from different perspectives than our own, especially as to the importance of IT phenotype differences and as to what is required to demonstrate a CGP.

LITERATURE CITED

1. Browder, L. E. 1980. A compendium of information about named genes for low reaction to *Puccinia recondita* in wheat. Crop Sci. 20:775-779.
2. Browder, L. E., and Eversmeyer, M. G. 1980. Sorting of *Puccinia recondita:Triticum* infection type data sets toward the gene-for-gene model. Phytopathology 70:666-670.
3. Knott, D. R., and Johnson, R. 1981. The sorting and analysis of infection types from *Triticum aestivum*/*Puccinia recondita* interactions. Phytopathology 71:1010-1012.
4. Loegering, W. Q. 1978. Current concepts in interorganismal genetics. Annu. Rev. Phytopathol. 16:309-320.
5. Loegering, W. Q., and Burton, C. H. 1974. Computer-generated hypothetical genotypes for reaction and pathogenicity of wheat cultivars and cultures of *Puccinia graminis tritici*. Phytopathology 64:1380-1384.
6. Person, C. 1959. The gene-for-gene relationship in host:parasite systems. Can. J. Bot. 37:1101-1130.
7. Robinson, R. A. 1976. Plant Pathosystems. Springer-Verlag, Berlin, Heidelberg, and New York. 184 pp.