Theor Appl Genet (2005) 112: 21–29
DOI 10.1007/s00122-005-0084-2

ORIGINAL PAPER

Min-Jea Kim · Walter Federer · Martha A. Mutschler

Using half-normal probability plot and regression analysis to differentiate complex traits: differentiating disease response of multigenic resistance and susceptibility in tomatoes to multiple pathogen isolates

Received: 27 June 2005 / Accepted: 8 August 2005 / Published online: 20 October 2005

Abstract The need for a new analytical approach was encountered in the course of characterizing newly developed tomato lines resistant to late blight. Late blight resistant tomato lines were created in independent breeding programs using the accession *Solanum pimpinellifolium* L. (formerly *Lycopersicon pimpinellifolium* (L.) Miller) L3708 as the source of the resistance. However, initial field observation suggested that the late blight resistance in the lines produced by two independent breeding programs differed. Possible causes included a partial transfer of the late blight resistance derived from *S. pimpinellifolium* L3708 or the possibility of race specificity of this resistance. A crucial issue was determining the most appropriate and robust analytical method to use with data from laboratory analyses of the responses of nine tomato lines against five *P. infestans* isolates. Prior analysis by standard ANOVA revealed significant differences across tomato lines but could not determine whether the disease responses in the CLN-R lines were different from those of the heterozygous F_1 hybrids, created by crossing susceptible tomatoes with the fixed CU-R lines. A different analytical method was needed. Therefore, sporangia numbers/leaflet and diseased area data were analyzed using a half-normal probability plot and regression analysis. The results of this analysis show its utility for genetic or pathology studies. Considering only populations of the uniform tomato lines, this method confirms the results obtained by using a standard ANOVA, but provides a clearer demonstration of the distributions of the individuals within the populations and how this distribution impacts variance and the difference among the populations. This method also allows a joint analysis of the uniform lines with an additional population that is less uniform, because it is segregating. Such an analysis would be invalid using a standard ANOVA. The results of this joint analysis determined that the additional population was divergent from the fixed CU-R lines, and, against some isolates, against the CLN-R lines as well. Half-normal probability plot analysis method would be applicable more broadly beyond analysis of disease resistance data. It could be useful for data from populations that are not normally distributed, for traits which are affected by epistatic gene action, and could be useful for selection of extremes.

Keywords Resistance · Late blight · *Phytophthora infestans* · Tomato · *Lycopersicon esculentum* · Half-normal probability analysis

Communicated by H. C. Becker

M.-J. Kim · W. Federer · M. A. Mutschler (✉)
Department of Plant Breeding and Department of Biological Statistics and Computational Biology,
Cornell University, Ithaca, NY 14853, USA
E-mail: mam13@cornell.edu

Introduction

Late blight (caused by *Phytophthora infestans* (Mont.) de Bary), causes severe loss of tomato production when the environment is favorable to the pathogen. Incorporation of new late blight resistance could be a useful addition to an integrated late blight control strategy. Researchers at AVRDC (Asia Vegetable Research Development Center) found that *S. pimpinellifolium* accession L3708 (*a.k.a.* LA1269, NSL116890 and PI365957) is a source of strong resistance to late blight in tomatoes (AVRDC 1994; Chunwongse et al. 2002) and generously provided this accession to other breeding programs, each of which proceeded to transfer the resistance to tomato breeding lines. However, when the resulting late blight resistant lines were grown together under natural infestation, the degree of resistance appeared to differ among lines from the different programs (R. Gardner (personal communication).

Tomato lines with resistance derived from L3708 developed in two independent breeding programs (CU-R and CLN-R lines, respectively) were tested against a series of *P. infestans* isolates to test for differences in disease response among tomato lines across isolates.

Standard ANOVA analysis revealed that the lines produced by the two programs were significantly different, with the set of CU-R lines all resistant across the five isolates used and the set of CLN-R lines all showing resistance to high levels of the disease against only a subset of the five isolates (Kim 2003; Kim and Mutschler 2005). Heterozygous F_1 hybrids produced by crosses of susceptible lines with the homozygous CU-R lines also showed resistance to high levels of disease against a subset of the five isolates. That analysis, however, could not fully determine whether the disease responses differed among all of the populations of interest. Furthermore, segregation for resistance genes was a possible explanation for the lower level of late blight resistance in the CLN-R lines than that in highly resistant CU-R lines bred from the same resistance source (Kim 2003; Kim and Mutschler 2005), and use of standard ANOVA method for data analysis would not be appropriate if any of the lines/populations analyzed were segregating. For this reason, an alternative analytical method was needed for analysis of data including these populations.

Half-normal probability plot and regression analysis could be utilized to analyze a data set possibly including segregating populations. The half-normal probability plot method was conceived by Daniel (1959) and further developed by Birnbaum (1959) and Krane (1963). It is a procedure to determine whether a set of observations are members of a single distribution, or if there are outliers present. The n observations are ranked from 1 to n (highest). The ordered values of $P_k = (2k-1)/2n, k = 1, 2, ..., n$, are computed. Using half-normal probability plot graph paper, the values of P_k, as ordinate values, are plotted against the response Y_k as the abscissa values. The values of Y_k falling on a straight line are considered to belong to the same distribution. Those not falling on the line are omitted and the P_k values are recomputed for the reduced set of observations. The values are then replotted to determine whether additional observations will be considered as outliers. Finally, the linear fit of the remaining line is assessed by regression.

Originally, the half-normal probability plot was used to identify important versus unimportant factors on effect (expressed as orders) in singly replicated factorial design experiments (Daniel 1959). If the data are normally distributed, they will be on the line. If some combinations of factors contribute differently from the combined majority of the other factors, they are found off this common line. Thus, if the factors have different effects, data will be grouped differently, affecting the lines drawn.

To our knowledge, the half-normal probability plot and regression analysis has not previously been used in genetic or pathology studies. The goals of this work were to apply the half-normal probability plot and regression analysis to the sporangia number and disease area data in order to test the utility of this method on host–pathogen interaction data, and to attempt to characterize more completely the differences in disease response across isolates among the tomato lines tested against set of isolates of *P. infestans*.

Materials and methods

Tomato lines tested

Nine tomato lines were tested against five *Phytophthora infestans* isolates in this study. The late blight susceptible control lines used were the open-pollinated freshmarket tomato line NC215E (R. Gardner, North Carolina State University) and the open-pollinated processing tomato line E6203 (LA4024, available from C. M. Rick Tomato Genetics Resource Center, Davis, CA, USA).

Two late blight resistance lines CLN 2037 B and CLN 2037 E (developed and provided by Hanson and Black of the AVRDC Tainan, Taiwan), which also carry resistance from L3708 were also used in the replicated test. These are referred to as the CLN lines

Two late blight resistant lines bred at 993104-10 and 993111-7, (Kim 2003; Kim and Mutschler 2005) carrying the resistance from *L pimpinellifolium* L3708 (AVRDC 1994; Chunwongse et al. 2002) were also used in this analysis. These are referred to as the CU-R lines. The *P. infestans* isolates US-7 and US-17 were used in screening late blight resistance during the breeding program; the resulting fixed lines were uniformly resistant to both of these isolates (Kim 2003; Kim and Mutschler 2005).

Pollinating the susceptible lines, E6203 and NC215E, with pollen from the late blight resistant homozygous CU-R line 993104-10 produced experimental hybrids that were heterozygous for the late blight resistance gene(s) carried by 993104-10.

In the course of breeding the fixed late blight resistant CU-R lines, selections had also been made for plants that appeared to have lower levels of resistance to US-7 and US-17. Self-progeny of one of these selections, 982067-3 (designed Low-R), was also tested.

Pathogen isolates used

Phytophthora infestans isolates US-7 (940330), US-11 (980066), US-17 (970001), NC-1 (980003), and DR4B (DR990004) were obtained from W. Fry (Dept. of Plant Pathology, Cornell University) for use in these tests. US-7 was previously a dominant isolate in US, and US-11 is still a major isolate in California. NC-1 has been a dominant isolate in North Carolina, and US-17 was called "a tomato-specified isolate" in the southeast US. DR4B was collected from the Dominican Republic. Culture maintenance and inoculum preparation were as described in Kim (2003) and Kim and Mutschler (2005).

Inoculation and data collection

The detached leaflet droplet test method (Legard et al. 1995) was used to test resistance and susceptibility level. Six plants per line were tested except for the low

resistance selection population, for which 10 plants were tested. Thus, 58 plants in total were used. From each plant, five leaflets were detached and each leaflet was inoculated with one of five *P. infestans* isolates (one leaflet per isolate). These experiments were repeated three times with leaflets from same plants. In all, 870 leaflets were tested. Assays were performed, and diseased leaflet area and sporangia produced per leaflet data were collected as described in Kim (2003) and Kim and Mutschler (2005).

Data analysis

The collected spore number, diseased leaflet area data were analyzed both by half-normal probability plot and by regression analysis. Three-rep average values of individual plants were calculated and analyzed with half-normal probability plots. The n data values were ranked from lowest to highest and then P_k values were calculated k ($k = 1, ..., n$) by the following equation:

$$P_k = (2k - 1)/2n$$

The actual average sporangia number was plotted as the horizontal axis against P_k value as the vertical axis, and the diseased area was plotted as the horizontal axis against P_k value as the vertical axis. The regression lines were then determined for each tomato line.

Results and discussion

The trends are apparent in the differences among tomato lines for average sporangia number and diseased leaflet area (cm^2) by some of the lines against the five isolates (Tables 1 and 2). The average sporangia numbers indicate that the susceptible lines produced high-average sporangia numbers, while in contrast the average sporangia numbers of the two homozygous CU-R lines were very low, demonstrating that these lines were resistant to all of these pathogen isolates (Table 1). Thus, very little of the variation among the tomato lines or pathogen isolates in the experiment was generated from these homozygous lines. The average sporangia numbers of the heterozygous hybrids depended on the pathogen isolate used. They were similar to those of the homozygous fixed lines against US-11, but lay between those of the resistant fixed lines and susceptible controls against the isolates US-17, NC-1 and DR4B (Table 1). Furthermore, the heterozygous hybrids did not suppress sporangial production of US-7 although the parental homozygous fixed line was resistant to this isolate.

The CLN-R lines, CLN 2037 B and CLN 2037 E, which were also considered to be fixed for late blight resistance from L3708, produced very different average sporangia numbers against the five isolates from those of the two homozygous CU-R lines. The resistance of the CLN-R lines was not effective across all five isolates used. Indeed, the average sporangial numbers of the CLN-R lines against US-7 followed a pattern that was more similar to that of the heterozygous F_1s than of the fixed CU-R lines.

The average sporangia numbers of the low-R progeny against US-7 was also similar to that of the CLN-R lines (Table 1). This presentation of the data provides the means and a measure of the variance around the means, but is not informative regarding the distribution within a group.

The results for average diseased areas of the susceptible lines and the fixed CU-R lines were very similar to results for average sporangia numbers for these lines (Table 2). The susceptible tomato lines were all highly susceptible to all isolates, with diseased areas often extending throughout the entire leaflet, while the two homozygous CU-R lines showed similar strong resistance to all of the isolates. Thus, very little of the variability among tomato lines or pathogen isolates for disease area was generated from these late blight resistant entries (Kim 2003). The results for the average diseased area were different from the sporangia results for the heterozygous hybrids. The average diseased areas in the heterozygous hybrids were generally much closer to those of the susceptible lines than were sporangia

Table 1 Average sporangia numbers and standard errors for nine tomato genotypes tested with five *P. infestans* isolates

Tomato name and class		*P. infestans* isolate				
		US-11	US-17	DR4B	NC-1	US-7
E6203	S	67,812 ± 13,370	163,953 ± 49,056	256,078 ± 36,872	383,109 ± 86,844	683,655 ± 116,371
NC215E	S	125,891 ± 19,695	145,086 ± 43,315	196,279 ± 23,712	209,995 ± 31,532	518,186 ± 64,490
NC215E X 993104-10	F_1	1,172 ± 682	9,688 ± 7,034	18,984 ± 8,246	24,961 ± 8,865	158,273 ± 40,985
E6203 X 993104-10	F_1	312 ± 243	7,656 ± 2,540	4,063 ± 1,897	26,367 ± 7,973	186,832 ± 28,608
993104-10	CU-R	0	0	0	0	820 ± 661
993111-7	CU-R	0	0	0	0	0
CLN 2037 B	CLN-R	273 ± 273	9,648 ± 3,722	742 ± 510	9,766 ± 4,404	107,294 ± 25,924
CLN 2037 E	CLN-R	0	4,961 ± 2,439	0	1,211 ± 760	61,377 ± 21,843
992067-3	Low-R	0	2,273 ± 1,176	30,975 ± 13,812	2,391 ± 1,272	53,423 ± 15,924

$N = 30$ for 992067-3 and all others, $N = 18$
Class S: Susceptible control lines
CU-R: From CU-R series of late blight resistant tomato lines
F_1: Heterozygous F_1 hybrid between Susceptible lines X CU-R lines
CLN-R: From CLN series of late blight resistant tomato lines
Low-R: Low resistance selection from CU-R breeding program

Table 2 Average diseased areas (cm^2) and standard errors for nine tomato genotypes tested with five *P. infestans* isolates

Genotype name and class		*P. infestans* isolate				
		US-11	US-17	DR4B	NC-1	US-7
E6203	S	11.19 ± 0.54	13.74 ± 1.06	9.42 ± 0.64	12.66 ± 0.82	14.07 ± 0.94
NC215E	S	7.12 ± 0.69	9.90 ± 0.94	8.44 ± 0.78	10.97 ± 0.63	13.09 ± 0.88
NC215E X 993104-10	F_1	0.58 ± 0.33	1.92 ± 1.06	1.27 ± 0.42	3.98 ± 0.90	8.11 ± 0.90
E6203 X 993104-10	F_1	0.27 ± 0.20	4.20 ± 1.07	0.71 ± 0.29	4.78 ± 0.98	9.59 ± 0.75
993104-10	CU-R	0	0	0	0	0.42 ± 0.42
993111-7	CU-R	0	0	0	0	0
CLN 2037 B	CLN-R	0.21 ± 0.21	2.57 ± 0.90	0.66 ± 0.49	2.86 ± 0.99	9.76 ± 1.45
CLN 2037 E	CLN-R	0	2.39 ± 1.00	0	0.39 ± 0.19	3.80 ± 1.15
992067-3	Low-R	0	0.84 ± 0.40	1.99 ± 0.57	1.07 ± 0.56	4.67 ± 0.73

$N = 30$ for 992067-3 and all others, $N = 18$
Class S: Susceptible control lines
CU-R: From CU-R series of late blight resistant tomato lines
F_1: Heterozygous F_1 hybrid between Susceptible line X CU-R line
CLN-R: From CLN series of late blight resistant tomato lines
Low-R: Low resistance selection from CU-R breeding program

numbers, and particularly so when challenged with US-7 (Table 2).

The disease expression of the two CLN-R lines and the low-resistance population inoculated with the five isolates was similar to that of the heterozygous F_1s. The response pattern for the average diseased area results for heterozygous hybrids were similar to sporangia results for the CLN-R lines, as far as general ranking of the virulence of the isolates. When US-7 was used, the CLN-R lines, low-resistance selection population, and the heterozygous F_1s all showed disease expression closer to that of the susceptible lines (Table 2).

The averages of both disease measures presented in Tables 1 and 2, showed clearly that the CLN-R lines showed different responses against the five isolates than the late blight resistant CU-R lines. However, a statistical test of the data is required to show that the differences are significant. To choose the appropriate method of testing, one must consider the natures of the lines being tested and of the L3708-derived resistance as transferred into the CLN-R lines vs. the CU-R lines. If a single gene controls the L3708 resistance and this resistance was transferred to the fixed lines bred in both breeding programs, then these fixed lines should all perform similarly. Prior field observations with the CLN-R lines suggested that this might not be the case. One explanation is that the full resistance derived from L3708 is controlled by more than one gene and that CLN-R lines either are not homozygous for, or are missing at least one of, the resistance genes. If either the CLN-R lines or the low-resistance Cornell selection were heterozygous for a resistance gene and produced segregating progeny, then comparisons of averages and variations with the other non-segregating lines would be inappropriate. For this reason, we employed the half-normal probability plot and regression analysis (Birnbaum 1959; Daniel 1959; Krane 1963) instead of a more typical ANOVA analysis.

In analysis of these data, a partitioning of the degrees of freedom for each of the nine lines would have been possible and corresponding sums of squares could be computed to obtain an ANOVA. One could use a multiple comparisons procedure and the Behrens procedure if interest centered only in the means of the combinations rather than the distribution of responses. Degrees of freedom could be approximated by the Satterwaite or other procedure. However, there are several reasons why F-tests in this ANOVA would be inappropriate and invalid for this experiment. For instance, there is a problem of unequal variances for the 45 line-isolate combinations. Twelve of these have ZERO variances. No variance stabilizing transformation, e.g. $\log(1 + Y)$, would correct the variance heteroscedasticity problem.

Our goal was to study the distribution of responses for each of the 45 line-isolate combinations and determine which combinations have similar distributions. Hypothesis testing is not relevant here. The half normal probability plot graphical method was ideal for interpreting the responses for each combination. For the 45 line-isolate combination responses, we used the average of the three repetitions of experiment. Use of averages tends to make the distribution of responses more symmetrical (Central Limit Theorem) and hence closer to normality. This tends to validate the use of the half normal plot procedure for this experiment.

The half-normal probability plot method has been criticized because a precise rule for omitting observations has not been formulated (Daniel, 1959; Krane 1963). Usually, however, one omits only observations with large divergences from the line. The process is then repeated to determine which additional observations are likely outliers. The process of repeating the procedure until all remaining observations are approximately on the same line is sufficient to detect outliers. Often the experimenter will be able to assign reason for an outlier. In our situation, we minimized the problem since we were able to detect an entire group that was divergent, namely resistant versus susceptible. The procedure is also useful for detecting divergent observations in a segregating group. In addition, the procedure allows for an estimate of the experimental error variance. From the final set of n observations considered to have the same distribution, compute $m = 0.683n + 0.5$. The value of

Y_m is the estimated experimental variance. An eye-fitted line rather than a computed linear regression line is usually sufficient.

There are advantages in using the half-normal probability plot method for analysis to detect heterogeneous components. In a homogeneous fixed line population, each plant will be considered as a factor combination and will be in a commonly distributed data set. In a segregating population, the population is heterogeneous and each plant is considered as a different factor combination; susceptible plants and resistant plants will not be in the same distributed data set. However, if the contribution to an effect is similar, the data will be in a commonly distributed data set. If we expand the concept that resistance gene combinations are factor combinations, the contribution of the same resistance gene combinations will be in the same commonly distributed data set. In other words, the same resistance gene(s) combination will lie on the same line or one of similar slope. If the data of two populations were distributed in the same range, the slopes of regression lines would be similar because the P_k value would be similar too. If the range of the Y_k values of one population was small and the range of another population was wide, the slopes of the distributions would be different. With this method, we could also compare individual plants in a low-resistance population that might be segregating.

This procedure allows searching for outlying distributions even though it was originally proposed to find outlying observations. The procedure graphically shows the distribution of segregating combinations and where the responses fall relative to the parents of a cross. Line-isolate combinations for the nine lines do not fall on the same line. This means we are studying outlying distributions rather than outlying observations as is customarily done when using this procedure.

Data were grouped as fixed CU-R lines, the low-resistance selection population, the heterozygous F_1s, the CLN-R lines, and the susceptible line, and then analyzed by the half-normal method described. Rather than using half-normal probability graph paper, we used Microsoft Excel software to obtain the graphs. Excel uses equally spaced values of P_k, which has the effect of flattening the slopes. This, however, does not affect the ability to observe discrepant observations. The data points located on the vertical-axis are all zero, even though they have different rank and P_k values. For the purpose of graphing, when n points are tied at zero, they were given n consecutive ranks within the range, so the points would be visually distinguishable on the figure. This will not affect regressions of non-zero-containing data sets and regressions are not calculated on all-zero data sets. The sporangia number plots for five different isolates are summarized in Fig. 1 with the regression line slopes summarized in Table 3.

Sporangia half-normal probability results against US-11 indicated that the homozygous CU-R lines and the low-resistance selection population were both resistant against this isolate (Fig. 1a; Table 3) and so fall on a common line lying along the vertical axis. The heterozygous F_1s and the CLN-R lines also were resistant and fell along this axis, but due to a very few outlying points, which were probably due to experimental error, there is a very slight deviation from the vertical axis. The susceptible lines were a discrete group with a line that differed in its slope and placement from all of the other lines (Fig. 1a; Table 3).

Sporangia half-normal probability results against US-17, DR4B and NC-1 indicated that the fixed CU-R lines were a unique group with the greatest resistance. The low resistance selection population and the CLN-R lines grouped together and showed similar patterns of distribution and slopes indicating resistances lower than that of the fixed CU-R lines (Fig. 1b–d; Table 3). The heterozygous F_1s were in a group of their own between a group with CLN-R lines and the low resistance selection population and the group of susceptible lines.

Sporangia half-normal probability results using US-7 were different than those obtained with any of the other isolates (Fig. 1e; Table 3). The US-7 results separated the lines into three distinct grouping. Most resistant were the homozygous fixed CU-R lines, which had sporangia numbers of zero. The heterozygous F_1s, CLN-R lines and the low resistance selections all had similar slopes of regression lines and had sporangia number ranges greater than the homozygous CU-R lines but lower than that of the last group composed of the susceptible lines.

Against US-7, the group composed of the CLN-R lines, the heterozygous F_1s, and the low-resistance selection population were in same range of distribution and had similar regression line slopes. This result clearly suggested that the heterozygotes and CLN-R lines were not resistant to US-7 even though they were resistant to US-11. The results across isolates indicated that the CLN-R lines and the low-resistance selection were more susceptible to US-7 than to US-17, NC-1 and DR4B. These results indicate that the susceptibility level of the two CLN-R lines was similar to that of the heterozygous F_1s and the low resistance selection, rather than the fixed CU-R lines.

Considering the preceding results, it is unlikely that resistance transferred to the CLN-R and the CU-R lines is controlled by single completely dominant gene. The results of the heterozygous F_1s were clearly different from their fixed line parent against the four isolates other than US-11.

The alternative hypothesis, suggested by Chunwongse et al. (2002), is that resistance is due to a single incompletely dominant gene, and so lower levels of resistance could be attributed to the heterozygous condition. However, the data from the less resistant CLN-R and high resistant fixed CU-R lines across isolates do not support this hypothesis. The responses across isolates of the CLN-R and fixed CU-R lines are very different. The levels of resistance of the CLN-R lines for some isolates does have similarities to that of the heterozygous F_1 hybrid created using the fixed CU-R lines,

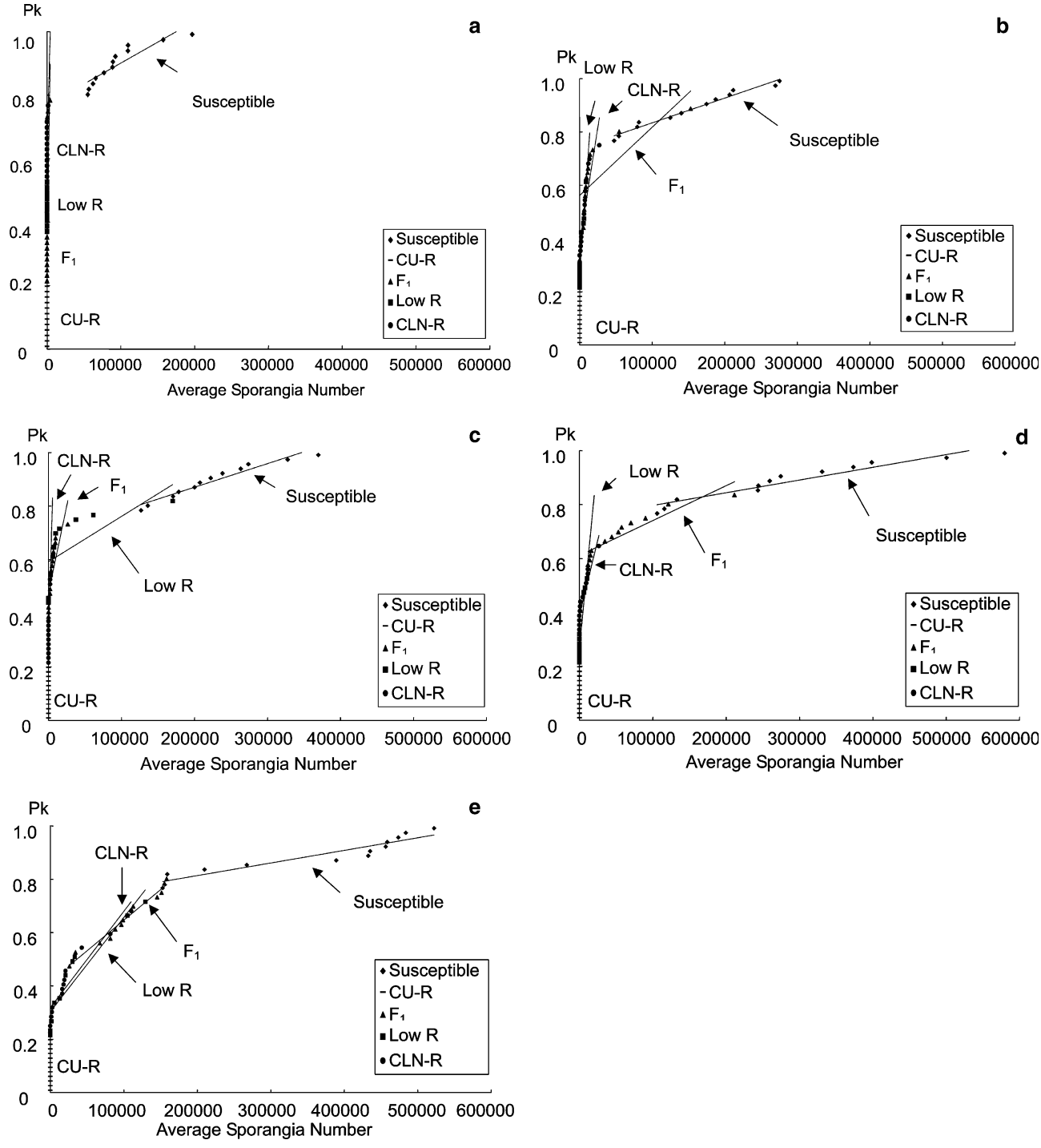

Fig. 1 Plot of average sporangia number/leaflet for nine tomato lines by P_k against five *P. infestans* isolates. **a** US-11, **b** US-17, **c** DR4B, **d** NC-1, **e** US-7. *Susceptible* Susceptible control lines (E6203 and NC215E) *CU-R* From CU-R series of late blight resistant tomato lines (993104-10 and 993111-7). F_1 Heterozygous F_1 hybrids between susceptible lines and CU-R lines (E6203 X 993104-10 and NC215E X 993104-10), *Low R* Low resistance selection from CU-R breeding program (982067-3), *CLN-R* From CLN series of late blight resistant tomato lines (CLN 2037 B and CLN 2037 E)

however there was no evidence that these less resistant CLN-R lines or their selfed progenies segregate for resistance. If only one incompletely dominant gene controlled the resistance, and a population is not fixed and uniform for the resistance (due to segregation and/or assortment), then the progeny of at least some of these lines should include plants with the higher resistance against all 5 isolates and/or plants that are fully susceptible to all 5 isolates. Such off-type plants were not observed in the progeny of the CLN-R lines. Therefore,

Table 3 Summary of regression and R^2 of P_k versus average sporangia numbers by genotypes

Isolate	Genotype							
	Susceptible		F_1		Low R		CLN-R	
	Regression	R^2	Regression	R^2	Regression	R^2	Regression	R^2
US-11	$y = 1.E{-}6\ x + 0.77$	0.82	$y = 1.E{-}4\ x + 0.33$	0.58	NA	NA	$y = 8.E{-}5\ x + 0.63$	0.35
US-17	$y = 9.E{-}7\ x + 0.74$	0.97	$y = 3.E{-}6\ x + 0.54$	0.38	$y = 3.E{-}5\ x + 0.30$	0.93	$y = 2.E{-}5\ x + 0.34$	0.71
DR4B	$y = 9.E{-}7\ x + 0.69$	0.94	$y = 2.E{-}5\ x + 0.40$	0.54	$y = 2.E{-}6\ x + 0.54$	0.33	$y = 6.E{-}5\ x + 0.38$	0.57
NC-1	$y = 5.E{-}7\ x + 0.75$	0.92	$y = 1.E{-}6\ x + 0.61$	0.86	$y = 2.E{-}5\ x + 0.34$	0.86	$y = 1.E{-}5\ x + 0.33$	0.69
US-7	$y = 5.E{-}7\ x + 0.72$	0.88	$y = 2.E{-}6\ x + 0.42$	0.98	$y = 3.E{-}6\ x + 0.31$	0.88	$y = 4.E{-}6\ x + 0.30$	0.78

Regression and R^2 for CU-R genotype could not be calculated.

one cannot attribute the lower levels of resistance to heterozygosity.

Diseased area data were analyzed with the half--normal probability plot method (Fig. 2; Table 4). The half-normal probability plot results of the diseased area data had similarities with and differences from the results of the sporangia number analysis. Diseased area half-normal probability results of the homozygous fixed CU-R lines against the isolates US-17, DR4B, NC-1 indicate that these lines grouped together and have strong resistance against all of these isolates (Fig. 2b–d; Table 4), and with the exception of a few points from the F_1 and CLN-R, they fall on a common line lying along the vertical axis. Comparatively, no other plant lines groups with these lines. These results were very similar to those concerning sporangia numbers. The US-11 diseased area results were also the same as US-11 sporangia results. Homozygous fixed lines, low-resistance selection population, F_1s, and CLN-R lines were all grouped together and resistant against US-11 (Fig. 2a; Table 4).

Diseased area half-normal probability results of F_1s, the low-resistance selection population and the two CLN-R lines against US-17, DR4B, NC-1 (Fig. 2b–d; Table 4) were very similar, and more similar to that of the susceptible than to the resistant CU-R lines. This stands in contrast to the sporangia number analysis, in which a group with CLN-R lines and the low resistance selection population were more resistant than the heterozygous F_1's, and closer to the CU-R lines.

Diseased area half-normal probability results using US-7 are perhaps the most extreme. Most resistant were the homozygous fixed CU-R lines, which are a discrete class. All the other lines fall in different, though overlapping, ranges on or very near a common line, indicating a lack of significant difference among these lines against US-7 (Fig. 2e; Table 4).

The combined sporangia results and diseased area results indicates that the CLN-R lines showed wider diseased area than sporangia production, like heterozygous F_1s. A model that would fit the data would postulate that the resistance is controlled by a major gene, which provides resistance to US-11, and in combination with one (or more) additional gene(s), provides the resistance to the other isolates. The resistant fixed CU-R line results were completely different from the CLN-R line results. This difference would be explained if other minor gene(s), fixed in the CU-R lines, was/were recessive and supported major gene action to provide the wider range of resistance. The existence of the low-resistance selection, which was derived from the same base population as the homozygous CU-R fixed lines, also fits the hypothesis that the full resistance, as expressed in the fixed CU-R lines, involves more than one gene. Progeny test results of CLN 2037 B and CLN 2037 E against US-17 support that the two CLN-R lines are homozygous and do not segregate (Kim 2003; Kim and Mutschler 2005). If all the populations are indeed fixed, one cannot explain the differences in response of the fixed CU-R lines and the CLN-R lines if resistance is controlled by single gene.

Considered together, the results of these experiments show that the resistance of L3708 can be transferred to create lines that have a full level of resistance, such as the CU-R lines. The results also show the possibility that even a very good breeding program could transfer only partial resistance due to isolates used in a selective screen. The weaker or partial resistance may be due to the absence of the other gene(s). Considering the impact of the choice of isolate on the expression of the resistance (Kim 2003; Kim and Mutschler 2005), a possible cause for the difference in the independently produced lines could be the type of isolates used for screening and selections in the course of breeding, rather than any difference in screening methods used.

Control of resistance by more than one gene is not unusual. In a study of resistance to late blight in potatoes, race-specific resistance required more than a single dominant R gene for expression of the dominant suppressor (El-Kharbotly et al. 1996). The interaction of more than one dominant gene for a fully functional expression of resistance has also been seen in studies of other host plant/disease systems. *Cf-2*, which is derived from *S. pimpinellifolium*, required the unlinked *Rcr3* gene to be fully functional. Interestingly, *Rcr3* is allelic to *Ne* gene, which is derived from *S. pimpinellifolium* and suppresses *Cf-2*–dependent autonecrosis conditioned by its *Solanum lycopersicum* L (formerly *Lycopersicon esculentum* Miller) allele *ne* (Kruger et al. 2002). *Mla-12*, race-specific resistant gene to powdery

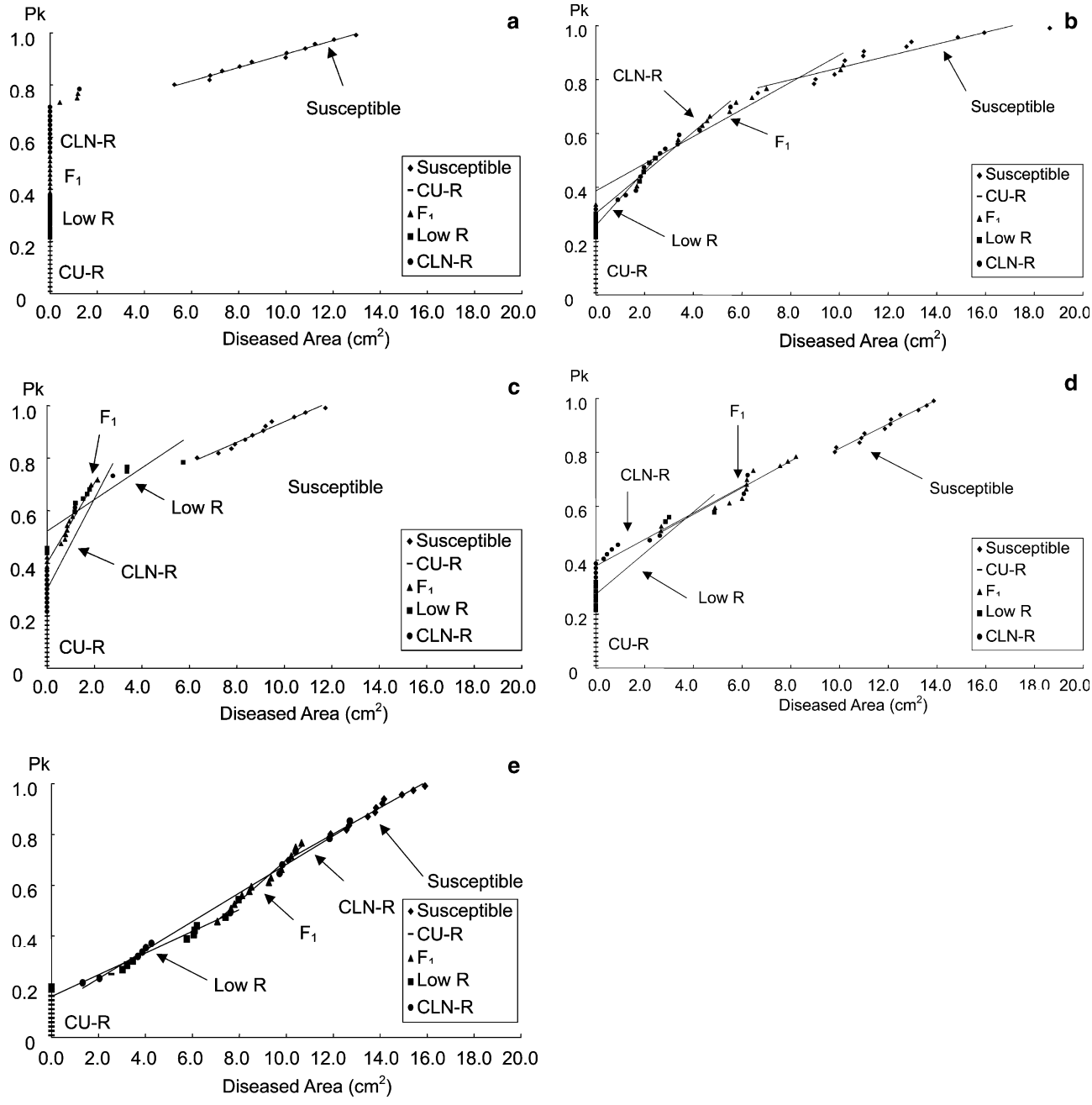

Fig. 2 Plot of average diseased area for nine tomato genotypes by P_k against five *P. infestans* isolates. **a** US-11, **b** US-17, **c** DR4B, **d** NC-1, **e** US-7. *Susceptible* Susceptible control lines (E6203 and NC215E), *CU-R* From CU-R series of late blight resistant tomato lines (993104-10 and 993111-7). F_1 Heterozygous F_1 hybrids between susceptible lines and CU-R lines (E6203 X 993104-10 and NC215E X 993104-10) *Low R* Low resistance selection from CU-R breeding program (982067-3). *CLN-R* From CLN series of late blight resistant tomato lines (CLN 2037 B and CLN 2037 E)

mildew in barley, also required *Nar-1* and *Nar-2* loci for full functionality (Freialdenhoven et al. 1994). Resistance for rice blast was found to be controlled by two dominant unlinked genes (Pan et al. 1996). A race non-specific resistance breeding effort is currently under way for rice blast resistance (Castano et al. 1989). The strategy in this program is to use pyramiding to obtain a broad spectrum of resistance (Li et al. 2001; Rao et al. 2002).

Half-normal probability plot analysis method would be applicable more broadly beyond analysis of disease resistance data. Utility of the method would be determined by the type of data and gene action involved in regulation of the trait, rather than the type of trait. Half-normal probability plot analysis might be useful for data from populations that are not normally distributed, and particularly for data sets involving traits which are affected by epistactic gene action.

Table 4 Summary of regression and R^2 of P_k versus average diseased areas by genotypes

Isolate	Genotype							
	Susceptible		F_1		Low R		CLN-R	
	Regression	R^2	Regression	R^2	Regression	R^2	Regression	R^2
US-11	$y = 0.026\ x + 0.66$	0.99	$y = 0.430\ x + 0.30$	0.83	NA	NA	$y = 0.125\ x + 0.63$	0.40
US-17	$y = 0.022\ x + 0.62$	0.88	$y = 0.051\ x + 0.39$	0.93	$y = 0.085\ x + 0.29$	0.93	$y = 0.076\ x + 0.30$	0.96
DR4B	$y = 0.039\ x + 0.55$	0.97	$y = 0.237\ x + 0.29$	0.91	$y = 0.085\ x + 0.44$	0.65	$y = 0.136\ x + 0.38$	0.83
NC-1	$y = 0.045\ x + 0.36$	0.98	$y = 0.050\ x + 0.37$	0.94	$y = 0.076\ x + 0.27$	0.90	$y = 0.049\ x + 0.38$	0.95
US-7	$y = 0.052\ x + 0.18$	0.98	$y = 0.081\ x + 0.11$	0.98	$y = 0.043\ x + 0.16$	0.95	$y = 0.056\ x + 0.12$	0.99

Regression and R^2 for CU-R genotype could not be calculated.

In breeding, selection of extreme individuals involves selection from tails of the population distribution (for example the highest 5% percent). However, if more than one gene is involved in the trait, some of the plants in the tail could be extreme due to some other cause (for example gene and environment interaction) rather than line alone. Using Half-normal probability plot analysis to identify and select individuals which are outliers from the common distribution of the population could be a more effective method to select for true genetic differences. The ability to identify and eliminate true outliers would also be of great benefit to systems used for the identification of quantitative trait loci.

Acknowledgements We would like to thank Dr. Bill Fry (Department of Plant Pathology, Cornell University) for providing pathogen isolates and for advice on pathogen maintenance and use and Edward D. Cobb (Department of Plant Breeding, Cornell University) for assistance in plant growth. We would also like to thank Drs. Hanson and Black of the AVRDC, Tainan Taiwan, for generously supplying seeds of their tomato lines, and for cooperative discussions on the differences among late blight resistant lines. We also thank the person who reviewed this manuscript in July 2005 for the clarifying comments provided. This work was supported by Hatch Project 149484 and gifts from the California League of Tomato Processors.

References

AVRDC (1994) 1993 Progress Report. Asian Vegetable Research and Development Center, Shanhua, Tainan, Taiwan, pp 201–203

Birnbaum A (1959) On the analysis of factorial experiments without replication. Technometrics 1:343–357

Castano J, Mackenzie DR, Nelson RR (1989) Components analysis of race non-specific resistance to blast disease of rice caused by *Pyricularia-Oryzae*. Phytopathology 127:89–99

Chunwongse J, Chunwongse C, Black L, Hanson P (2002) Molecular mapping of the *Ph-3* gene for late blight resistance in tomato. Hort Sci Biotech 77:281–286

Daniel C (1959) Use of half-normal plots in interpreting factorial two-level experiments. Technometrics 1:311–342

El-Kharbotly A, Pereira A, Stiekema WJ, Jacobsen E (1996) Race specific resistance against *Phytophthora infestans* in potato is controlled by more genetic factors than only R-genes. Euphytica 90:331–336

Freialdenhoven A, Scherag B, Hollricher K, Collinge DB, Thordal CH, Schulze LP (1994) *Nar-1* and *Nar-2*, two loci required for *Mla-12*-specified race-specific resistance to powdery mildew in barley. Plant Cell 6:983–994

Kim MJ (2003) Characterization of the late blight resistance of *Lycopersicon pimpinellifolium* accession L3708, and its transfer to elite processing tomato lines. Cornell University, Ph.D Thesis

Kim MJ, Mutschler MA (2005) Transfer to Processing Tomato and Characterization of Late Blight Resistance Derived from *Solanum pimpinellifolium* L3708. J Amer Soc Hort Sci (in press)

Krane SA (1963) Half-normal plots for multi-level factorial experiments. Proceedings of the eighth conference on the design of experiments in army research development and Testing. ARO-D Report 63-2. U.S. Army Research Office, Durham, pp 261–285

Kruger J, Thomas CM, Golstein C, Dixon MS, Smoker M, Tang S, Mulder L, Jones JDG (2002) A tomato cysteine protease required for *Cf-2*-dependent disease resistance and suppression of autonecrosis. Science 296:744–747

Legard DE, Lee TY, Fry WE (1995) Pathogenic specialization in *Phytophthora infestans*: aggressiveness on tomato. Phytopathology 85:1355–1361

Li ZK, Sanchez A, Angeles E, Singh S, Domingo J, Huang N, Khush GS (2001) Are the dominant and recessive plant disease resistance genes similar?: a case study of rice R genes and *Xanthomonas oryzae* pv. *oryzae* races. Genetics 159:757–765

Pan Q, Wang L, Ikehashi H, Tanisaka T (1996) Identification of a new blast resistance gene in the indica rice cultivar Kasalath using Japanese differential cultivars and isozyme markers. Phytopathology 86:1071–1075

Rao KK, Lakshminarasu M, Jena KK (2002) DNA markers and marker-assisted breeding for durable resistance to bacterial blight disease in rice. Biotech Adv 20:33–47