Plant Pathology (2009) **58**, 1128–1133 Doi: 10.1111/j.1365-3059.2009.02140.x

BSPP

Plant Pathology

Comparison of three diagrammatic keys for the quantification of late blight in tomato leaves

F. M. Corrêa[a*], J. S. S. Bueno Filho[b] and M. G. F. Carmo[a]

[a]*Departamento de Fitotecnia, Universidade Federal Rural do Rio de Janeiro, BR-465 km 7, 23890-000, Rio de Janeiro, RJ; and* [b]*Departamento de Ciências Exatas, Universidade Federal de Lavras, 37200-000, Lavras, MG, Brazil*

Three diagrammatic grading keys were designed for the assessment of the severity of late blight (caused by *Phytophthora infestans*) in tomato leaves. Simplified and broad keys considered, respectively, six (3, 12, 22, 40, 60 and 77%) and eight (3, 6, 12, 22, 40, 60, 77 and 90%) levels of disease severity, whilst a modified key based on a previous proposal for potato late blight considered six levels (1, 5, 10, 16, 32 and 50%). The keys were validated by 24 evaluators who assessed digital images of tomato leaves exhibiting different areas with lesions. Evaluator errors were compared using a mixed model in which evaluators were considered as random effects and the keys and evaluations as fixed effects. The accuracy and precision of the evaluators were compared by simple linear regression between the estimated and actual values of disease severity. The repeatability of evaluators was assessed using Pearson's correlation coefficient. There was significant ($P \leq 0{\cdot}001$) variability amongst the errors made by evaluators, although the precision of each of the three keys was high with a coefficient of determination (R^2) of 0·96, 0·93 and 0·83 for the simplified, broad and modified key, respectively. Repeatability of estimations amongst the evaluators was adequate (correlation coefficients of 0·91, 0·91 and 0·90 for the three keys, respectively). The simplified and broad keys resulted in higher precision and accuracy for the estimation of severity than did the modified key. Since the simplified key considers a smaller number of disease severity levels, its use is recommended in the assessment of late blight in tomato leaves.

Keywords: disease assessment, disease severity, *Phytophthora infestans, Solanum lycopersicum*

Introduction

Late blight, caused by the oomycete *Phytophthora infestans*, is a serious disease that has resulted in significant losses of tomato (*Solanum lycopersicum*) crops in Brazil (Reis *et al*., 2003). The majority of epidemics of late blight occur in the autumn and winter seasons during cold and humid periods. The disease is recognized by the appearance of grey to green spots on the leaves of infected plants, and these may evolve into necrotic brown areas that ultimately cause leaf loss and, in some cases, plant death (Erwin & Ribeiro, 1996; Cerkauskas, 2005). Whilst symptoms of the disease are typically observed on leaves, the stems, fruits and petioles are also affected. Control of the disease traditionally relies on the frequent application of fungicides (Suassuna *et al.*, 2004).

The assessment of affected crops and the development of strategies for selecting resistant tomato phenotypes depend on the availability of accurate methods for the quantification of the disease that allow comparisons to be made between different researchers and locations (James, 1974; Zadoks & Schein, 1979). Various techniques have been employed to measure the severity and spread of fungal diseases in plants, including video imaging, infrared photography, spectral reflectance and diagrammatic grading keys (Nilsson, 1995). The last of these approaches is the most straightforward and cost effective, and involves the estimation of a the affected area, the leaf lesion area (LLA), on the basis of a series of illustrations of whole plants or parts of plants presenting different levels of disease severity (Nutter *et al.,* 2006). However, a number of aspects must be considered in the design of such keys, including the limits of the scale and the range of intensities detected in field crops, the correct representation of symptoms and the limitations of human visual acuity as defined by Weber-Fechner's stimulus-response law (Horsfall & Barratt, 1945; James, 1974; Nutter & Schultz, 1995).

The first grading key based on Weber-Fechner's law was designed by Horsfall & Barratt (1945) and considered 12 levels of disease severity. Later, various other keys were developed for different pathogens and hosts (James, 1971; Slopek, 1989; Leite & Amorim, 2002; Nita *et al.,* 2003; Gomes *et al.,* 2004; Martins *et al.,* 2004; Spósito

*E-mail: fmcron@gmail.com

Published online 11 August 2009

et al., 2004; Andrade *et al.,* 2005; Godoy *et al.,* 2006; Halfeld-Vieira & Nechet, 2006). However, the only keys available for late blight are specific to potato leaves. One of these uses four levels of disease severity (1, 10, 25 and 50% LLA) and is applied to leaves (James, 1971), whilst the other considers eight levels of severity (≤ 10, 11–25, 26–40, 41–60, 61–70, 71–80, 81–90 and >90% LLA) and is applied to whole plants (Cruickshank *et al.,* 1982). Although these two keys are often used for the assessment of late blight in tomato (Jaime-Garcia *et al.,* 2001; Duarte *et al.,* 2007), the structure and leaf morphology of tomato plants differ from those of potato, thus increasing the possibility of errors in grading.

Other sources of error in determining disease severity are related to precision (i.e. variations associated with an estimate) and accuracy (i.e. differences between estimated and actual intensity of disease) of the evaluator (Nutter *et al.,* 2006). Such errors need to be detected and analysed in order to allow systematic deviations to be corrected, either by training or by using calibration equations (Shaw & Royle, 1989; Nutter & Schultz, 1995). Several approaches may be used to assess the precision, accuracy and repeatability of the evaluations, including the application of linear correlation (Shokes *et al.,* 1987) and variation coefficients (Nutter *et al.,* 1993). Additionally, simple linear regression has been employed to quantify evaluator error (Nutter *et al.,* 2006), and this approach, together with analysis of variance, has been used to identify differences between evaluators and methods of evaluation (Shokes *et al.,* 1987).

The aim of the present study was to develop and validate diagrammatic grading keys suitable for the evaluation of the severity of late blight in tomato leaves.

Materials and methods

The leaves employed in the design of the keys were obtained from 66 different tomato genotypes planted in the experimental fields of the horticulture sector of the Departamento de Fitotecnia (Instituto de Agronomia, Universidade Rural do Rio de Janeiro – UFRRJ, Rio de Janeiro, Brazil), in an area historically associated with the natural occurrence of tomato late blight. The genotypes employed included a commercial susceptible cultivar (Super Sweet; Azevedo, 2006), two resistant cultivars (Andrea-1 and Andrea-2; Azevedo, 2006) and 63 cultivars that were under evaluation at UFRRJ.

To develop the grading keys, a sample of 198 leaves was used (one from each cultivar in each of the three rows). Samples were collected in June 2006 from the apical, median and basal parts of tomato plants. Digital images of the leaves were recorded at a resolution of 600 d.p.i. in a flatbed scanner and the diseased areas were measured with SIARCS software (Jorge & Crestana, 1996). The maximum and minimum levels of late blight severity were determined based on the infected leaves collected from the field.

Three separate grading keys were designed. Two were based on histograms of the levels of late blight severity, on which a logarithmic scale was manually set according to the principles of Weber-Fechner's law (Horsfall & Barratt, 1945), whilst a third was based on the levels of severity of late blight in potato leaves, as proposed by James (1971). For the validation of the diagrammatic keys, two evaluations were conducted 7 days apart by 24 evaluators who had no previous experience in disease quantification. Slides of images of 50 tomato leaves displaying different disease levels were presented to the evaluators, who were then instructed to estimate the percentage of diseased area according to the three proposed keys. For each key, 1200 evaluations were recorded.

Comparison of the errors made by the evaluators was performed by analysis of variance (ANOVA) of the absolute errors (estimated severity minus actual severity). Actual severity was based on the readings of digitalized images and estimated severity was given by evaluators. Evaluators were not trained prior to the first evaluation, but were trained prior to the second. Possible effects of training were measured in the ANOVA.

The statistical model used in the ANOVA is represented by:

$$E[Y_{ijk}|[\delta, \varepsilon, \tau] = \mu + \alpha_i + \beta_j + \delta_k + \varepsilon_{ik} + \tau_{jk}$$

in which Y_{ijk} is the absolute error in the ith key in the jth evaluation of the kth evaluator; μ is the effect associated with the general mean; α_i is the effect associated with key i (where i = 1, 2 and 3), considered as a fixed effect; βj is the effect associated with evaluation j (where j = 1 and 2), considered as a fixed effect; δ_k is the effect associated with evaluator k (where k = 1, 2...24), considered as a random effect with $N(0, \sigma^2)$; ε_{ik} is the effect associated with scale i within the evaluator k, considered as a random effect; and τ_{jk} is the effect associated with evaluation j within evaluator k, considered as a random effect.

The accuracy of the evaluators was determined by simple linear regression between the estimated (or predicted, in the case of random effects) and actual values of disease severity according to the model $Y = \alpha + \beta X$, in which Y is the vector of predicted response, X the vector of actual values, and α and β the parameters of the equation to be estimated. The precision was determined from the coefficient of determination (R^2) of the regression (Nutter & Schultz, 1995). Estimations derived from the regression equations *t*-test were used to test two hypothesis regarding the parameter estimates, the intercept, Ho: a = 0, and the slope coefficient, Ho: b = 1 (Parker *et al.,* 1995). The repeatability of the evaluators in applying the proposed keys was assessed through the determination of Pearson's coefficient of correlation between the two evaluations (Nutter & Schultz, 1995). All of the statistical analyses were performed using R.2.5.0 software (R Development Core Team, 2008).

Results

The severity values employed in the proposed diagrammatic grading keys were established from the distribution

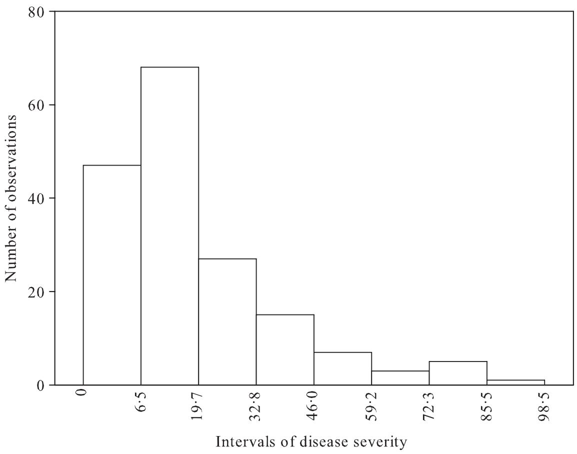

Figure 1 Distribution of the severity of late blight (*Phytophthora infestans*) observed in tomato leaves in the field.

of disease severity observed in tomato leaves (Fig. 1). The histogram indicated that the symptoms of late blight could affect nearly 100% of the leaf area, although leaves exhibiting an LLA encompassing 6·5–19·7% of the total surface were found most frequently. Based on the severity values presented in Fig. 1, two grading keys were proposed, one considering six levels of severity (the simplified key) and the other considering eight levels of severity (the broad key). The severity intervals of these two keys were based on the logarithmic increments postulated by Weber-Fechner's law (Nutter & Esker, 2006) and were 3, 12, 22, 40, 60 and 77% LLA for the simplified key, and 3, 6, 12, 22, 40, 60, 77 and 90% LLA for the broad key. A third key was based on a modified version of the proposals of James (1971) for potato late blight and considered six levels of severity, namely, 1, 5, 10, 16, 32 and 50% LLA.

The analysis of variance demonstrated significant differences in absolute error among keys and between evaluations, but not among evaluators. However, there was strong evidence of interactions between keys and evaluators and between evaluations and evaluators (Table 1). Variance components for interactions $\hat{\sigma}^2_{evaluator/keys}$ and $\hat{\sigma}^2_{evaluator/evaluation}$ were highly significant, indicating that evaluators' skills varied with the different keys and between evaluations, but there was no evidence that absolute error was reduced in the second evaluation (Table1).

Overall, the largest variation in absolute errors, and therefore the least precision, occurred with the James-modified key; the other two keys resulted in an improved level of precision, which was similar between them (Fig. 2). The James-modified key was also the least accurate, as evaluators using that key underestimated LLA by up to 40% in both evaluations. The problem of underestimating LLA increased with the James-modified key when the average LLA was more than 50%, because the scale did not distinguish these values (Fig. 3). Evaluators on

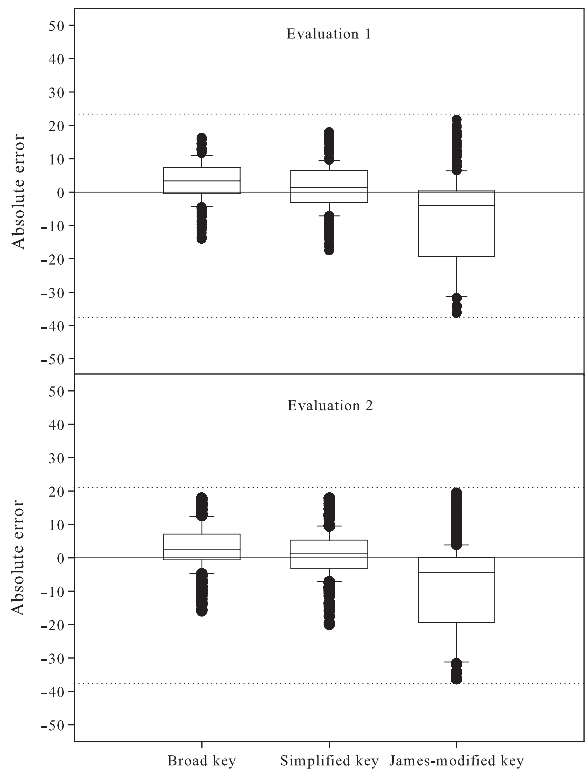

Figure 2 Box plot of the absolute errors in two separate evaluations of the severity of late blight (*Phytophthora infestans*) on tomato leaves using three different diagrammatic grading keys. Zero represents agreement between evaluator and reference values from scanned leaves.

Table 1 Analysis of variance of the absolute errors of 24 evaluators in the assessment of late blight (*Phytophthora infestans*) in tomato leaves using three different grading keys (simplified, broad and James-modified)

Source of variation	Degrees of freedom	Mean square	*F*-test
Key	2	47591·0000	<0·0001
Evaluator	23	1008·0000	<0·0965[a]
Evaluation	1	573·0000	<0·0134
Evaluator/key	46	224·0000	<0·0001[a]
Evaluator/evaluation	23	590·0000	<0·0001[a]
Residuals	7104	93·8454	

$\hat{\sigma}^2_{evaluator} = 0{\cdot}032 \pm 0{\cdot}566$ $\hat{\sigma}^2_{evaluator/scale} = 1{\cdot}30 \pm 1{\cdot}14$ $\hat{\sigma}^2_{evaluator/evaluation} = 3{\cdot}30 \pm 1{\cdot}81$

[a]*F*-test refers to the significance level against the hypothesis of null variance component for the random effect.

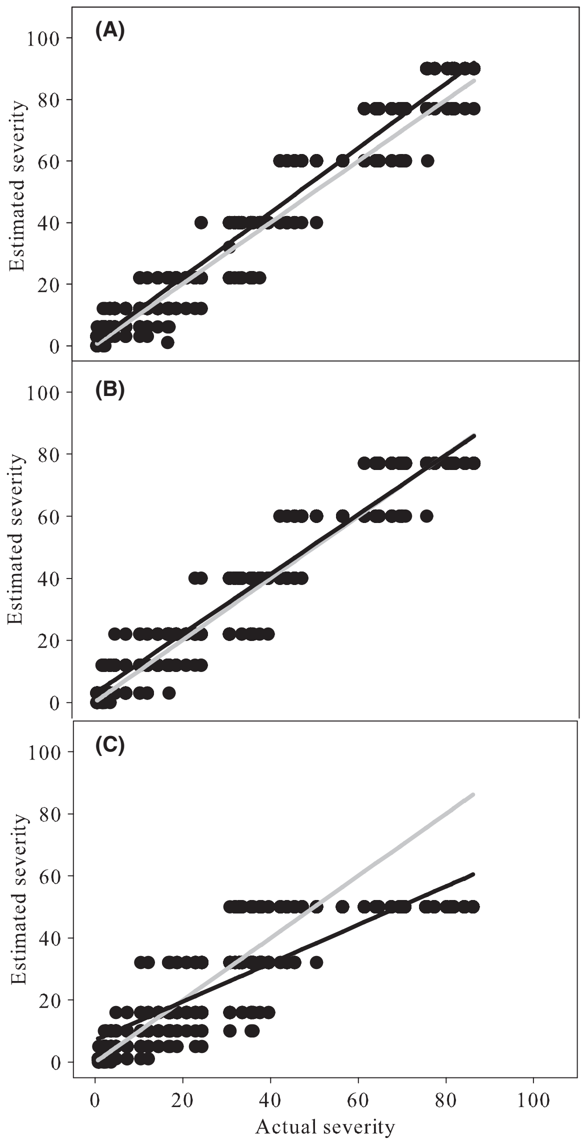

Figure 3 Linear regressions between actual severity of tomato late blight (*Phytophthora infestans*) and estimated severity evaluated using a broad (A), a simplified (B) and a James-modified (C) diagrammatic grading key. Dots represent the 100 estimations carried out by each of 24 evaluators. The continuous black line represents the regression between actual and estimated severity, whilst the continuous grey line represents the ideal relationship.

average tended to overestimate with the broad key, whilst accuracy was greatest with the simplified key (Fig. 2). Absolute error arising from the application of the simplified and broad keys ranged from 10 to 15%, but rarely exceeded the upper limit of the scales. In the first and second evaluations, the intercept values for the simplified and broad keys were near to 0, indicating the absence of constant deviations between estimated and actual severity. With the same keys, the slope coefficients were not different from 1, indicating the absence of systematic over- or underestimation associated with changes in disease severity (Table 2). In contrast, the intercept and slope for the James-modified key differed from 0 and 1, respectively (Table 2). Accuracy and precision were also evident by plotting predicted severity on actual severity (Fig. 3). Here also, the precision of evaluators was more stable across severity values for the simplified and broad keys. When using the James-modified key, evaluators were less precise at the mid-severity range. Overall, the repeatability of the estimations performed by the 24 evaluators was high, since the correlation between the two evaluations was 91% for the simplified and broad keys, and 90% for the James-modified key.

Discussion

The differences between evaluators in respect of their ability to measure dissimilar levels of intensity of late blight as determined in the present study were in agreement with the previous findings of Nutter & Schultz (1995). The maximum absolute error for the simplified and broad keys was about 20%. For the James-modified key the maximum was about 40%. In all cases maximum errors were underestimations of disease. These errors were similar to those previously reported (Nutter & Schultz, 1995; Leite & Amorim, 2002; Nita *et al.*, 2003; Gomes *et al.*, 2004; Martins *et al.*, 2004; Spósito *et al.*, 2004; Andrade *et al.*, 2005; Godoy *et al.*, 2006; Halfeld-Vieira & Nechet, 2006).

The absence of constant and systematic deviations, as well as the reduced R^2, arising from the application of the simplified and broad keys indicates that they were similar in both accuracy and precision and superior to the James-modified key. The inadequacy of the James-modified key in this case could be explained at least in part by the fact that it does not cover all of the severity levels observed in

Table 2 Intercepts (*a*), slope coefficients (*b*) and coefficients of determination (R^2) of simple linear regression equations relating the estimations of severity of late blight (*Phytophthora infestans*) in tomato leaves made by 24 evaluators (two evaluations per evaluator) using broad, simplified and James-modified diagrammatic grading keys, and the actual severity of the disease

	Grading keys								
	Broad			Simplified			James-modified[a]		
Evaluations	*a*	*b*	R^2	*a*	*b*	R^2	*a*	*b*	R^2
1	1·06	1·05	0·96	2·78	0·96	0·91	7·14[b]	0·61[b]	0·80
2	1·01	1·05	0·96	2·72	0·96	0·93	6·39[b]	0·62[b]	0·83

[a]James (1971).
[b]For these values, the null hypothesis (i.e. $a = 0$ and $b = 1$) was rejected according to the *t*-test ($P < 0·05$).

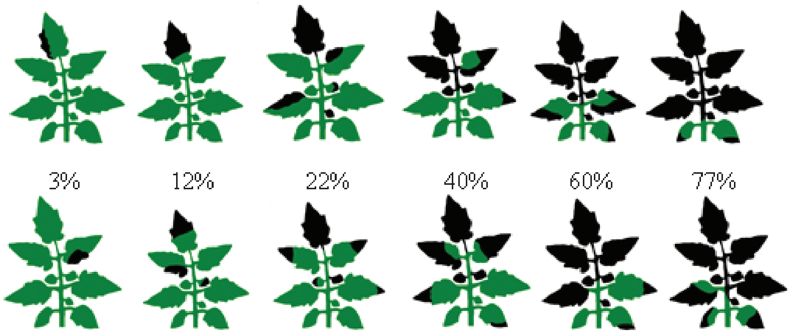

Figure 4 Illustration of the simplified diagrammatic grading key considering six levels of severity of late blight (*Phytophthora infestans*) in tomato leaves.

leaf samples collected in the field, especially those with more than 50% LLA (Nutter *et al.*, 1993).

Although use of grading keys for disease severity may introduce some absolute errors of measurement, any such shortcomings are compensated for by the speed and standardization they provide (Stonehouse, 1994). Thus, based on the criteria adopted herein, the simplified key is recommended over the broad key because it is less complex (considering only six levels of severity), but remains accurate and precise. The simplified key proposed represents a reliable tool for measuring late blight in the pathosystem *S. lycopersicum–P. infestans*. An illustration of the proposed simplified key for the evaluation of late blight in tomato leaves is presented in Fig. 4.

Acknowledgements

The authors wish to thank each of the evaluators who contributed to the study. We are also grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico for an MSc Fellowship to one of us (FMC).

References

Andrade GCG, Alfenas AC, Mafia RG, Maffia LA, Gonçalves RC, 2005. Escala diagramática para avaliação da severidade da mancha foliar do eucalipto causada por *Quambalaria eucalypti*. *Fitopatologia Brasileira* **30**, 504–9.

Azevedo VF, 2006 *Produção Orgânica de Tomate Tipo Cereja: Comparação entre Cultivares, Espaçamentos e Sistemas de Condução da Cultura*. Seropédica, RJ, Brazil: Universidade Federal Rural do Rio de Janeiro, MSc dissertation[http://www.ia.ufrrj.br/cpgf/DissertacaoVlamirFortesdeAzevedo.pdf].

Cerkauskas R, 2005. *Tomato Diseases: Late Blight*. Shanhua, Taiwan: AVDRC – The World Vegetable Center [http://www.avrdc.org/pdf/tomato/late_blight.pdf].

Cruickshank G, Stewart HE, Wastie RL, 1982. An illustrated assessment key for foliage blight of potatoes. *Potato Research* **25**, 213–4.

Duarte HSS, Zambolim L, Jesus Junior WC, 2007. Manejo da requeima do tomateiro industrial empregando sistema de previsão. *Summa Phytopathologica* **33**, 328–34.

Erwin DC, Ribeiro OK, 1996. *Phytophthora Diseases Worldwide*. St Paul, MN, USA: APS Press.

Godoy CV, Koga LJ, Canteri MG, 2006. Diagrammatic scale for assessment of soybean rust severity. *Fitopatologia Brasileira* **31**, 63–8.

Gomes AMA, Michereff SJ, Mariano RL, 2004. Elaboração e validação de escala diagramática para cercosporiose da alface. *Summa Phytopathologica* **30**, 38–42.

Halfeld-Vieira BA, Nechet KL, 2006. Elaboração e validação de escala diagramática para avaliação da mancha-de-cercospora em melancia. *Fitopatologia Brasileira* **31**, 46–50.

Horsfall JG, Barratt RW, 1945. An improved grading system for measuring plant disease. *Phytopathology* **35**, 655. Abstract).

Jaime-Garcia R, Orum TV, Felix-Gastelum R, Trinidad-Correa R, Vanetten HD, Nelson MR, 2001. Spatial analysis of *Phytophthora infestans* genotypes and late blight severity on tomato and potato in the Del Fuerte Valley using geostatistics and geographic information systems. *Phytopathology* **91**, 1156–65.

James WC, 1971. An illustrated series of assessment keys for plant diseases, their preparation and usage. *Canadian Plant Disease Survey* **51**, 39–65.

James WC, 1974. Assessment of plant disease and losses. *Annual Review of Phytopathology* **12**, 27–48.

Jorge LAC, Crestana S, 1996. SIARCS 3.0: novo aplicativo para análise de imagens. In: *XIII Congresso Latino Americano de Ciência do Solo, Águas de Lindóia, MG, Brazil, 1996*. Piracicaba, SP, Brazil: ESALQ/SBCS, 365–71.

Leite RMVBC, Amorim L, 2002. Elaboração e validação de escala diagramática para mancha de *Alternaria* em girassol. *Summa Phytopathologyca* **28**, 14–9.

Martins MC, Guerzoni RA, Câmara GMS, Mattiazi P, Lourenço AS, Amorim L, 2004. Escala diagramática para a quantificação do complexo de doenças foliares de final de ciclo em soja. *Fitopatologia Brasileira* **29**, 179–84.

Nilsson HE, 1995. Remote sensing and image analysis in plant pathology. *Canadian Journal of Plant Pathology* **17**, 154–66.

Nita M, Ellis MA, Madden LV, 2003. Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. *Phytopathology* **93**, 995–1005.

Nutter Jr FW, Esker PD, 2006. The role psychophysics in phytopathology: the Weber-Fechner law revisited. *European Journal of Plant Pathology* **114**, 199–213.

Nutter Jr FW, Schultz PM, 1995. Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. *Canadian Journal of Plant Pathology* **17**, 174–84.

Nutter Jr FW, Gleason ML, Jenco JH, Christians NC, 1993. Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology* **83**, 806–12.

Nutter Jr FW, Esker PD, Netto RAC, 2006. Disease assessment concepts and the advancements made in improving the accuracy and precision of plant disease data. *European Journal of Plant Pathology* **115**, 95–103.

Parker SR, Shaw MW, Royle DJ, 1995. The reliability of visual estimates of disease severity on cereal leaves. *Plant Pathology* **44**, 856–65.

R Development Core Team, 2008. *The R Project for Statistical Computing*. Vienna, Austria: The R Foundation for Statistical Computing. [http://www.R-project.org].

Reis A, Smart CD, Fry WE, Maffia LA, Mizubuti ESG, 2003. Characterization of isolates of *Phytophthora infestans* from southern and southeastern regions of Brazil, from 1998 to 2001. *Plant Disease* **87**, 896–900.

Shaw MW, Royle DJ, 1989. Estimation and validation of a function describing the rate at which *Mycosphaerella graminicola* causes yield loss in winter wheat. *Annals of Applied Biology* **115**, 425–42.

Shokes FM, Berger DH, Smith DH, Rasp JM, 1987. Reliability of disease assessment procedures: a case study with late leafspot of peanut. *Oleagineux* **42**, 245–51.

Slopek SW, 1989. An improved method of estimating percentage leaf area diseased using a 1 to 5 disease assessment scale. *Canadian Journal of Plant Pathology* **11**, 381–7.

Spósito MB, Amorim L, Junior JB, Bassanezi RB, Aquino R, 2004. Elaboração e validação de escala diagramática para avaliação da severidade da mancha preta em frutos cítricos. *Fitopatologia Brasileira* **29**, 81–5.

Stonehouse J, 1994. Assessment of Andean bean diseases using visual keys. *Plant Pathology* **43**, 519–27.

Suassuna ND, Maffia LA, Mizubuti ESG, 2004. Aggressiveness and host specificity of Brazilian isolates of *Phytophthora infestans*. *Plant Pathology* **53**, 405–13.

Zadoks JC, Schein RD, 1979. *Epidemiology and Plant Disease Management*. New York, USA: Oxford University Press.