Mexican Journal of Phytopathology
ISSN: 2007-8080

Crossref
Open access

Scientific Article

Aerial and terrestrial digital images for quantification of powdery mildew severity in Ayocote bean (*Phaseolus coccineus*)

Alfonso Muñoz-Alcalá, Gerardo Acevedo-Sánchez[1], Diana Gutiérrez-Esquivel, Oscar Bibiano-Nava, Ivonne García-González, Norma Ávila-Alistac[2], María José Armenta-Cárdenas, María del Carmen Zúñiga-Romano, Rene Gómez-Mercado, Juan José Coria-Contreras, Serafín Cruz-Izquierdo[3], Gustavo Mora-Aguilera[1]*, José Jesús Márquez-Diego, Programa de Fitosanidad-Fitopatología, [1]CP-LANREF, [3]Programa de Genética. Colegio de Postgraduados, Km 36.5 Carretera México-Texcoco, Montecillo, Texcoco, Estado México, México, C.P. 56230; [2]Departamento de Producción Agrícola y Animal, Universidad Autónoma Metropolitana, Unidad Xochimilco, Xochimilco, CDMX, México C.P. 04960.

***Corresponding Author**:
Gustavo Mora-Aguilera
morag@colpos.mx

Section:
Periodical Issue

Received:
02 December, 2023
Accepted:
09 March, 2024
Published:
21 March, 2024

Citation:
Muñoz-Alcalá A, Acevedo-Sánchez G, Gutiérrez-Esquivel D, Bibiano-Nava O, García-González I, Ávila-Alistac N, Armenta-Cárdenas MJ, Zúñiga-Romano MC, Gómez-Mercado R, Coria-Contreras JJ, Cruz-Izquierdo S, Mora-Aguilera G and Márquez-Diego JJ. 2024. Aerial and terrestrial digital images for quantification of powdery mildew severity in Ayocote bean (*Phaseolus coccineus*).Mexican Journal of Phytopathology 42(2): 17. https://doi.org/10.18781/R.MEX.FIT.2312-1

ABSTRACT

Objective/Background. Epidemiological research on *Phaseolus coccineus* is lacking. The aim was to develop and validate digital methods to quantify the severity associated with powdery mildew in ayocote bean.

Materials and Methods. An ayocote bean plot with 65.3 % incidence and 22.7 % average powdery mildew foliar severity was selected. Based on 250 leaves collected in field with varying severity degrees, eight 7- and 8-class logarithmic-diagrammatic scales (ELD) were designed and validated in a controlled environment (CEV) and field (FV). In Rstudio®, accuracy (β), precision (R^2), reproducibility (r), and agreement level were determined with Cohen's kappa index (κ_w) and Lin's concordance coefficient (LCC). Additionally, a *Hierarchical Cluster Analysis* (HCA) was performed by scale and assessment environment for clustering by similarity evaluation. In ArcMap® v10.3, in a 15-quadrant block, an *'image segmentation'* analysis was performed using supervised classification and maximum likelihood to estimate powdery mildew severity and an indicator of canopy coverage index (VCI).

Results. In VEC-1, *v1r2* (ELD-7c; β=1.07, R^2=0.93, r=0.87) and *v1r1* (ELD-8c; β=0.97, R^2=0.85, r=0.87) scales were best evaluated. In VEC-2, comparing clusters conformed in the HCA, the ELD-7c was the best scored with perfect accuracy (β>0.96), very high precision (R^2>0.94), very high reproducibility (r=0.97-0.99) and very high agreement (κw>0.96; LCC>0.97); and in ELD-8c reproducibility

and agreement decreased. In VCa, ELD-7c maintained optimal metrics, but ELD-8c reached ideal parameters for preventive ELD in early stages of powdery mildew (β>0.98, R^2>0.98, r=0.99, κw=0.99-0.999, LCC=0.98-0.999). Image analysis estimated severity = 8.4 % (*CI* = 5.3 - 12.6 %) and ICV = 0.88 (*CI* = 0.76 - 0.99), contrasting with field assessment 47 % (*CI* = 38.8 - 55.3 %) and 0.46 (*CI* = 0.76 - 0.99), respectively, mainly with ICV > 0.94 due to less symptomatic leaf exposure. Suggests applicability for canopy estimation with restrictions for severity based on pathogen expression.

Conclusion. A methodology for ELD development is proposed, comprising: image acquisition, processing and quantification; controlled validation and field validation. Validation statistics included precision (R^2); accuracy (β); reproducibility (Pearson's coefficient and *Hierarchical Cluster Analysis*); and agreement (Lin's Coefficient and Kappa Index), proposed in a comprehensive approach for first time. RGB-drone images are proposed to estimate a comprehensive vigor and severity coverage index.

Keywords: *Erysiphe vignae*, scales-logarithmic, RStudio.

Introduction

The Ayocote bean (*Phaseolus coccineus*) has a high productive potential in several regions of Mexico due to its tolerance to phytopathogens, environmental adaptability, and nutritional benefits (Ávila-Alistac *et al.*, 2023). However, scientific research on epidemiological and productive aspects is limited, including methodologies for disease quantification that allow proposing management strategies for the crop (Armenta-Cárdenas *et al.*, 2024). The parametrization of the damage subsystem, or pathometry, through severity assessment scales for several pathosystems has been the most widely used resource, emphasizing an etiological approach rather than a comprehensive epidemiological approach for crop management (Del Ponte *et al.*, 2017; Mora-Aguilera *et al.*, 2021). Over the past two decades (2007-2019), publication trends on development and implementation of severity scales suggest: 1) the field implementation has increased significantly, 2) the Horsfall & Barrat principles and Weber-Fechner law are being replaced by linear or arithmetic models, 3) the Weber-Fechner visual stimulus law is questioned, 4) it is used as tools strictly for assessment the damage subsystem, 5) there is no methodological consensus for determining severity intervals and optimal classes number (Del Ponte *et al.*, 2022; 2017; Franceschi *et al.*, 2020; Godoy *et al.*, 1996). In these new approaches, biological-epidemiological principles are compromised or

omitted by covering statistical parameters, e.g., the coefficient of determination (R^2) that justify their implementation. A review of 105 scientific articles published from 1991 – 2017 on development of Logarithmic Diagrammatic Scales (LDS) analyzed various approaches to generating scales and adjusting reliability parameters (Del Ponte *et al.*, 2017). In this meta-analysis, only five of 127 LDSs in three publications were developed for *Phaseolus sp.*, specifically for *Colletotrichum lindemuthianum*, *Uromyces appendiculatus*, *Phaeoisariopsis griseola*, *Pseudocercospora griseola*, and *Xanthomonas campestris* pv. *phaseoli* (Librelon *et al.*, 2015; Godoy *et al.*, 1997). Except for *Phaeoisariopsis griseola*, severity intervals between 0.1 – 60 % were divided into 6 – 9 classes based on the Weber-Fechner visual stimulus law. Additionally, the diagrammatic design (i.e., image of tissue on concern) was based on black-and-white drawn leaves due to digital technological restrictions (Del Ponte *et al.*, 2017; Godoy *et al.*, 1997). The most recent work using LDS was conducted for leaf blight caused by *Alternaria alternata* in *Phaseolus vulgaris* genotypes (Gonzalez-Cruces *et al.*, 2022). However, none for powdery mildew reported for the first time in *P. coccineus* from the central Altiplano in Mexico (Armenta-Cárdenas *et al.*, 2024). Therefore, an epidemiological approach was used in this study for LDS implementation in preventive disease management models in *Phaseolus coccineus*, which involved prioritizing the early classes of severity associated with pathogenic process of *Erisyphe vignae* for timely management control (Gonzalez-Cruces *et al.*, 2022; Mora-Aguilera *et al.*, 2021; Librelon *et al.*, 2015). In the experimental plot of Colegio de Postgraduados, Campus Montecillo, Estado de México, an area of 3,100 m^2 of *P. coccineus* established for seed production with genetic improvement purposes, and with high powdery mildew incidence was selected as a study area. The objective was to develop and validate digital methodologies for severity quantification using a logarithmic-diagrammatic scale and drone images, to establish preventive mechanisms potentially linked to official production and management programs.

Materials and Methods

Experimental plot. In July 2022, during the summer-autumn cycle, an experimental plot (50 x 62 m) of Ayocote bean in the flowering stage was identified with 65.3 % incidence and 22.7 % average foliar severity of powdery mildew (Armenta-Cárdenas *et al.*, 2022). Due to the irregular density and heterogeneity in coverage, a 13mpx image was taken from plot-centroid using a 50 m vertical flight with a Phantom 3 DJI® drone to plan the methodology for quantifying plant canopy and severity. The quadrant division used by Armenta-Cárdenas *et al.* (2022) with 80 quadrants and 720 sub-quadrants, was used as reference, and plant severity and vigor data were used to validate the estimates of this study.

Develop of a logarithmic-diagrammatic severity scale for powdery mildew. A total of 250 apical leaves from the middle stratum of *Phaseolus coccineus* plants were collected, photographed, and digitized, ensuring representativeness across the entire plot. The photographs were taken with mobile digital devices. A total of 50 images were discarded due to low resolution or poor damage discrimination. The collection criterion was aimed at severity range representative of disease progress from Y_o (healthy) to Y_{max} (maximum severity). The digitized images were processed in GIMP v2.10.32 for background removal and quantification of total area (*TA*), damaged area (*DA*), or signs of the fungus associated with powdery mildew. The severity percentage (*sev %*) per leaf was obtained using the equation: *sev %* = 100 * [*DA* / *TA*]. For develop of logarithmic-diagrammatic scale (LDS) based on severity, class midpoint (*CM*), lower limit (L_L), and upper limit (U_L) were determined by entering the number classes desired and maximum severity into *2Log*-Epidem v.2.0 using the modified Hosfall and Barrat method (1945) (Mora-Aguilera and Acevedo-Sánchez, 2022. CP-LANREF. Unpublished). Two scales of 7 (LDS-7) and 8 (LDS-8) classes were developed for validation purposes. The classes number selected considered the potential use in severity assessment studies, genetic improvement, or biological effectiveness tests.

Validation of severity scales in a controlled environment. Two validations were conducted in a controlled environment (*CEV*) for LDS-7 and LDS-8. In the first one, version 1 of each scale was validated, and in the second one, a version with fixes and improvements, e.g., change of photographic image. In total, four scales for LDS-7 and four for LDS-8 were validated. From the total collection of 200 digitized leaf images, 30 with different severity degrees were randomly selected. The images were uniformly sized and centered on a *PowerPoint*® 2016 slide. A randomization macro with 30 s of visual exposure per leaf was programmed using *Microsoft*® *Visual Basic*® 2016. After 30 s, the real severity value (*R-sev*) was automatically displayed bottom the respective image. The recording of estimated assessment per image, with support of the color-printed scale by rater, was carried out in Validar-PER v1.5 (CP-LANREF, 2022. Unpublished). This software was developed in *MS Excel*® to determine the accuracy (β, slope) and precision (R^2, coefficient of determination) associated with a linear regression model ($y = \beta_o + \beta_x + e$) between *R-sev* and estimated severity (Nutter and Schultz, 1995). In Validar-PER, each rater recorded the Class Number (#Class) and *CM* of respective scale. The file was projected in a classroom to nine raters for the assignment of severity class (*S-class*) per leaf. LDS-7 and LDS-8 were used in independent events. Each rater recorded in Validar-PER v1.5 the *S-class* and *R-sev* values for 30 leaves. Upon completing each assessment, β and R^2 coefficients were automatically generated associated with a proposed accuracy scale to measure the bias of real value relative

to estimated (β < 0.96, underestimated; β = 0.96 to 1.05, perfect; and β > 1.06, overestimated), and precision for the bias in set of real values relative to estimate in overall evaluation (R^2 < 0.69, unacceptable; R^2 = 0.7 to 0.8, low; R^2 = 0.81 to 0.9, medium; R^2 = 0.91 to 0.95, high; R^2 > 0.96, very high). Reproducibility, as indicator of consistency by LDS among raters (Nutter & Schultz 1995), was estimated in *RStudio*® 2023.06.2 through: 1) Pearson correlation coefficient (*r*) from a matrix with *S-class* values; 2) weighted Kappa index (κw); and 3) Lin's Concordance Correlation Coefficient (LCC), to determine the agreement: no concordance (< 0), insignificant (0.0 – 0.2), low (0.2 – 0.4), moderate (0.4 – 0.6), good (0.6 – 0.8), and very good (0.8 – 1.0).

Validation of severity scales in the field. The LDS-7 and LDS-8 with the best scores in *r*, β, and R^2 were selected for field validation (*FV*). Four out of nine initial raters were selected based on contrasts in plant pathologist profile, and the final β and R^2 scores from the controlled environment validation (*CEV*). In the plot, 240 leaves (30 per scale and rater) were randomly selected to determine *S-class* using the color-printed scale. Additionally, an image of assessment leaf was taken for digitization and to determine *R-sev* using the previously described method. Like the *CEV* validation, Validar-PER v1.5 was used for estimating accuracy and precision. Each rater recorded *S-class* and *R-sev* values for 30 leaves per scale. The β, R^2, *r*, κw, and LCC parameters were generated for each rater and scale based on the method described above.

Hierarchical cluster analysis for CEV and FV. An independent matrix was integrated for the *CEV* and *FV* processes for nine and four raters, respectively, including *S-class* and *R-sev* for 30 leaves. In RStudio® 2023.06.2, for each assessment event and severity scale, heat maps were performed using the *Heatmap* function to represent *S-class* and reproducibility among raters through the Pearson correlation coefficient (*r*). Positive *r* values equal to or close to 1 indicated high reproducibility. For each *Heatmap* and variable (rater and leaf), a *Hierarchical cluster analysis* (HCA) was performed using the '*complete*' method and '*Euclidean*' distance (*d*). A green-yellow-red gradient color scheme was used to represent the transition from healthy to damage through *S-class* and *r*. The *S-class* and *R-sev* values for each process, scale, and rater were plotted using the *ggplot* function, fitting a linear regression to determine β, R^2, and *p*-value parameters using the *stat_poly_eq* function.

Estimation of coverage vigor and severity using drone images. A block of nine quadrants (Column 8:10, Row 1:3) was selected based on the criterion of maximum powdery mildew inductivity through terrestrial-visual exploration. At the block

centroid, 10 RGB-drone images (13 mpx) were captured with the Phantom 3 DJI® at 27 m elevation. An image-control per quadrant at 5 m elevation was captured for higher-resolution symptom. In ArcMap® v10.3, an '*image segmentation*' analysis was conducted in two stages using supervised classification and maximum likelihood: 1) Training on an image-sample with a 5 m resolution, to create a '*RGB signature*' capable of discriminating between classification categories, i.e., powdery mildew severity (*sev*), soil (*s*), leaf tissue (*t*), and flowering (*f*); and 2) Image-segmentation by extrapolating the '*RGB signature*' from training to the 27 m resolution image corresponding to the whole block. The area for each classification category and total area of whole block were obtained ($TS = sev + s + t + f$). The powdery mildew severity estimation for the block was calculated as: $sev_b = 100 * (sev / TS)$; and the canopy coverage as $vc_b = 100 * ([\ sev + t + f\ ] / TS)$. Severity and canopy coverage were compared against field data for quadrants evaluated by Armenta-Cárdenas *et al.* (2024).

Results and Discussion

Design of logarithmic-diagrammatic severity scales (LDS). Eight scales were developed, two with 7 classes and two with 8 classes, with two repetitions per scale. The classes range used was like other studies with analogous objectives (5 – 12, *Mo* = 6; Del Ponte *et al.*, 2017). The representative image for each class was selected based on lower limit (L_L) and upper limit (U_L) obtained in *2Log*-Epidem v.2.0. The LDSs for 7 and 8 classes were composed with class image associated to L_L, *CM*, and U_L (Figure 1). Values of 0, 0.5, 2.1, 6.2, 16.6, 41.7, and 100 % were obtained as *CMs* for LDS-7c (Figure 1A); and 0, 0.2, 0.9, 2.7, 7.1, 17.8, 42.9, and 100 % for LDS-8c (Figure 1B).

First validation of eight scales in a controlled environment (CEV-1). The evaluation in LDS-7c had β, R^2, and r coefficients ranged in 0.99 – 1.02, 0.80 – 0.94, and 0.81 – 0.87, respectively, reported as acceptable-optimal in scale design and validation studies (Ortega-Acosta *et al.*, 2016; Librelon *et al.*, 2015; Godoy *et al.*, 1996; Nutter Jr. & Schultz, 1995). The scale *v1r2* was the best scored with β = 1.07 (overestimated by 7 %), R^2 = 0.93 (*high*) and r = 0.87 (*moderate*). Although *v2r1* had similar coefficients, it obtained the lowest reproducibility at 0.81 (Table 1). For the LDS-8c, the coefficients were notably lower: β = 0.81 – 0.97, R^2 = 0.80 – 0.85, and r = 0.79 – 0.94 (Table 1). The best scored scale was *v1r1* with β = 0.97 (perfect), R^2 = 0.85 (medium) and r = 0.87 (medium) (Table 1). Overall, it was observed that a higher classes number correlated with increased assessment error, mainly with a trend to underestimate, possibly due to difficulty in discriminating

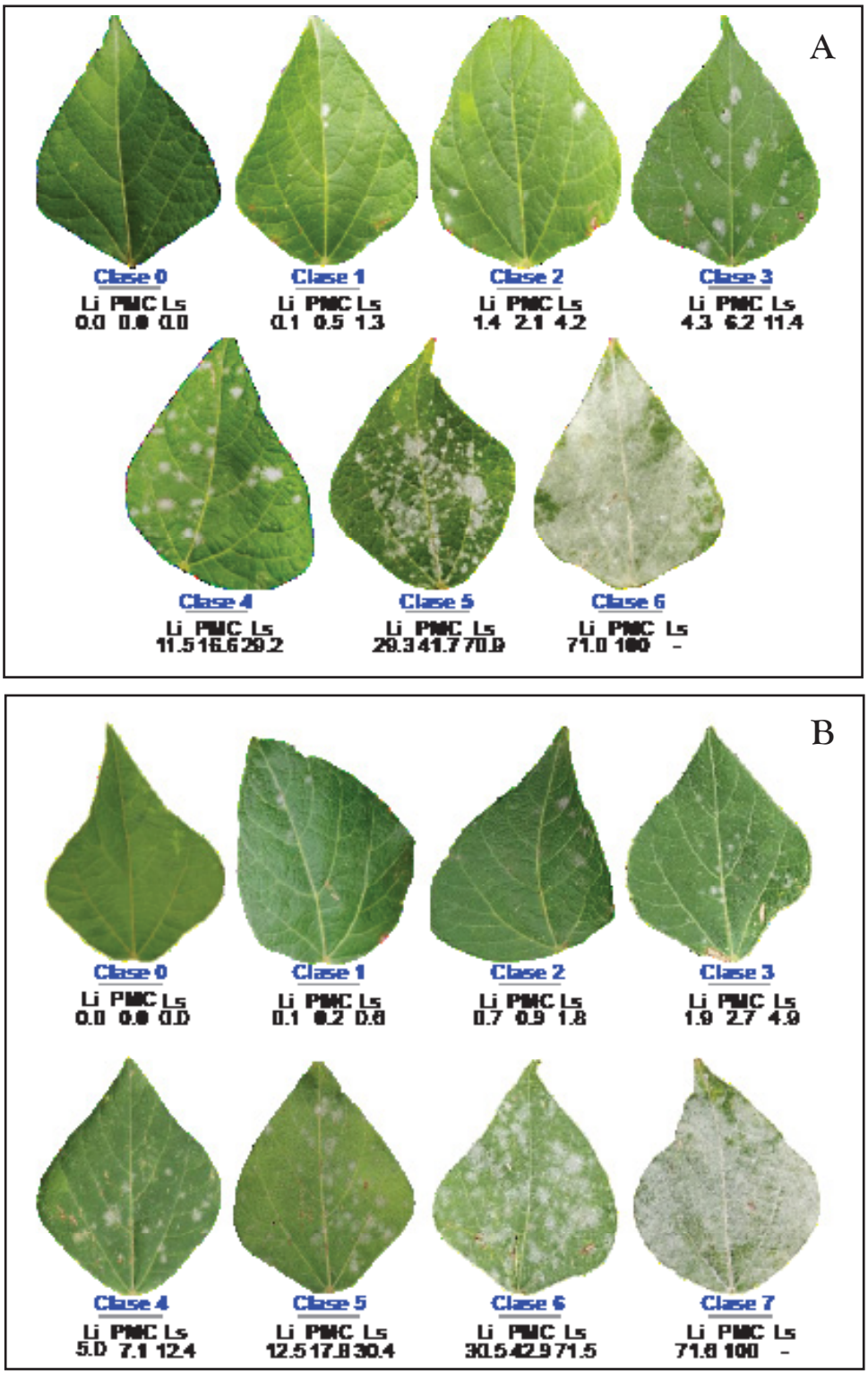


Figure 1. Final versions of logarithmic-diagrammatic severity scales for powdery mildew in Ayocote bean (*Phaseolus coccineus*), selected for field validation. **A)** 7-class logarithmic-diagrammatic scales; **B)** 8-class logarithmic-diagrammatic scales.

Table 1. Average accuracy (β), precision (R^2), and reproducibility (r) of eight logarithmic-diagrammatic severity scales for evaluating powdery mildew in *Phaseolus coccineus*.

ID[z]	Accuracy (β)	Qualitive class of β	Precision (R^2)	Qualitive class of R^2	Reproducibility (r, Range)
			ELD* 7 *classes		
v1r1	0.99	Perfect	0.81	medium	0.83 (0.60 – 0.98)
v1r2	**1.07**	**Overestimated**	**0.93**	**high**	**0.87 (0.67 – 0.99)**
v2r1	1.08	Overestimated	0.94	high	0.81 (0.61 – 0.98)
v2r2	1.02	Perfect	0.80	low	0.86 (0.72 – 0.94)
			ELD 8 classes		
v1r1	**0.97**	**Perfect**	**0.85**	**medium**	**0.87 (0.77 – 0.97)**
v1r2	0.81	Underestimated	0.84	medium	0.94 (0.81 – 1.00)
v2r1	0.96	Perfect	0.81	medium	0.79 (0.62 – 0.99)
v2r2	0.87	Underestimated	0.80	low	0.79 (0.62 – 0.99)

[z] The ID was formed based on the version number (v) and the repetition (r) of the scale. *Scales selected for best superior rating in the evaluated parameters are highlighted in **bold**.

damage between classes (Del Ponte *et al.*, 2022; Perina *et al.*, 2020). However, other factors that may influence outcomes in a controlled environment include image quality, exposure time, or rater experience (Librelon *et al.*, 2015; Godoy *et al.*, 1997). The scales *v1r2* and *v1r1*, of seven and eight classes respectively, best scored in this phase were selected for a second validation, analysis, and specific parameterization.

Second validation and parametric analysis of scales (CEV-2). In LDS-7c, the *Hierarchical cluster analysis* (HCA) formed a cluster for raters 4 – 9 ($p = 0.94$), characterized by parameters of *perfect* accuracy ($\beta > 0.96$), *very high* precision ($R^2 > 0.94$), *very high* reproducibility ($r = 0.97 – 0.99$), and *very good* concordance ($\kappa_w > 0.96$; LCC > 0.97) (Figure 2A1-A2, Figure 3A, and Table 2). Raters 1 – 3 formed independent clusters ($p > 0.75$), which underestimated severity by 10 – 18 % in classes 2 – 5 ($\beta = 0.82 – 0.97$), their precision was *moderate* ($R^2 = 0.82 – 0.89$), and although reproducibility was *high* ($r = 0.92 – 0.94$), the agreement was *moderate – good* ($\kappa_w = 0.9 – 0.95$; LCC = 0.89 – 0.93) (Figure 2A1-5A3, Figure 3A, and Table 2). In LDS-8c, two significant different clusters were formed ($p < 0.001$), with raters 4 and 8 as independent cases ($p > 0.56$). The main cluster was formed by raters 2, 3, 6, and 9 ($p = 0.96$), with *perfect* accuracy ($\beta > 0.96$), *very high* precision ($R^2 > 0.95$), *high – very high* reproducibility ($r > 0.93$), and *high – very high* concordance ($\kappa w = 0.80 – 0.97$; LCC = 0.97 – 0.98) (Figure 2B1-B3, Figure 3B, and Table 2). Raters 1, 5, and 7 formed the second cluster ($p = 0.56$), underestimating severity by 10 – 15 % in class 4 ($\beta = 0.68 – 0.80$).

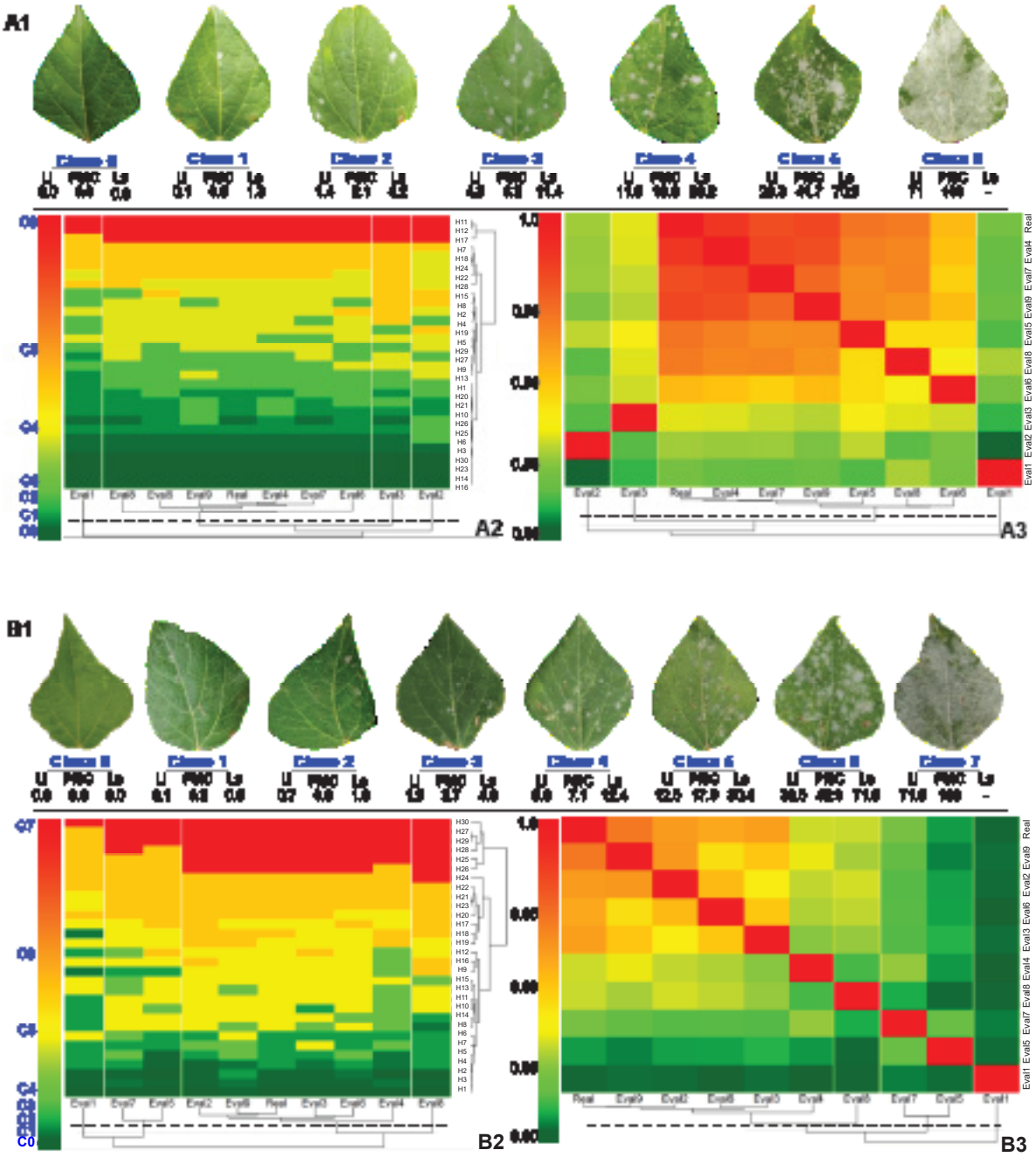


Figure 2. **A1 and B1**. Logarithmic-diagrammatic scales of 7 and 8 classes for assessment of powdery mildew severity on Ayocote bean (*P. coccineus*) leaves, during the Controlled Environment Validation (CEV) process of 30 leaves by nine raters. **A2 and B2**. *Heatmap* of Pearson correlation coefficient (*r*) among nine raters by severity scale. Values of $r = 0.8 – 1$ indicate the reproducibility of each scale among raters. **A3 and B3**. *Heatmap* of severity class in 30 leaves evaluated by scale and rater. The color represents the class value assigned by the rater for each leaf. For rater and leaves, a *Hierarchical cluster analysis* is plotted, grouped by the *complete* method and *Euclidean* distance.

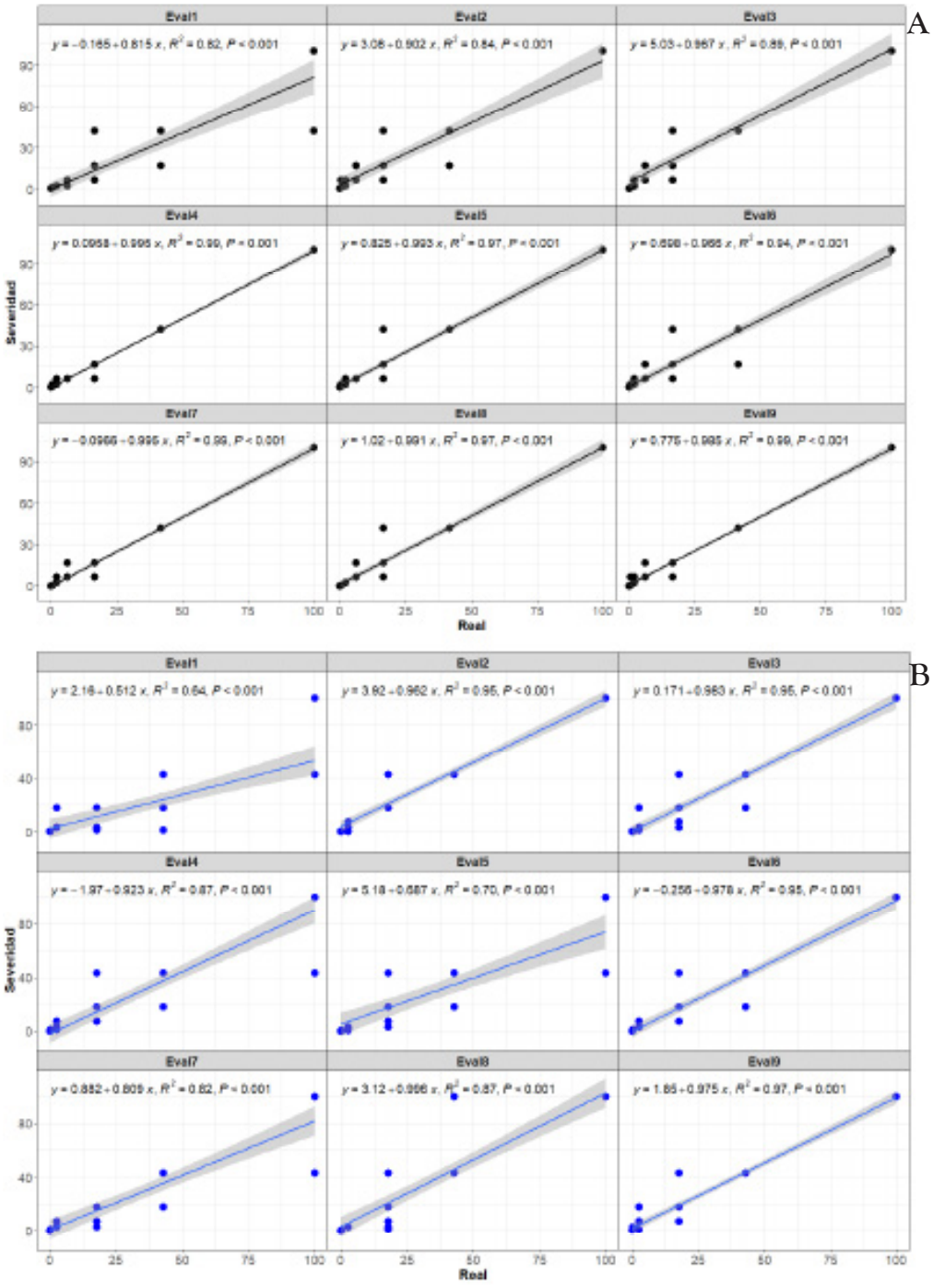


Figure 3. Correlation graphs between severity (y) assessed using the scale and real values (x) by nine raters during the Controlled Environment Validation (CEV) with 30 *Phaseolus coccineus* leaves. The linear regression equation ($y = \beta o + \beta x + e$) is fitted to determine β, R^2, and p-value parameters using the *stat_poly_eq* function. **A.** LDS-7 classes. **B.** LDS-8 classes.

Table 2. Parametric comparison of nine raters relative to the real value, to determine accuracy (βx), precision (R^2), and agreement (LCC, κ_w) by severity class assessed during the validation process in a controlled environment (CEV) and in the field (FV).

Stage[x]	Rater *vs* real	Hits *vs* real [y]	βo	β_x	*r*	$R^{2\,z}$	LCC	κ_w
					ELD 7 Classes			
	Eval1	0.63	-0.2	0.82	0.91	0.82	0.89	0.95
	Eval2	0.57	3.1	0.90	0.92	0.84	0.92	0.90
	Eval3	0.67	5.0	0.97	0.94	0.89	0.93	0.95
	Eval4	0.87	0.1	0.99	0.99	0.99	0.99	0.98
	Eval5	0.87	0.8	0.99	0.98	0.97	0.98	0.98
	Eval6	0.73	0.7	0.96	0.97	0.94	0.97	0.96
	Eval7	0.87	-0.1	0.99	0.99	0.99	0.99	0.98
	Eval8	0.90	1.0	0.99	0.98	0.97	0.98	0.99
	Eval9	0.83	0.8	0.96	0.99	0.99	0.99	0.96
	$\bar{x}$	**0.77**	**.**	**0.95**	**0.96**	**0.93**	**0.96**	**0.96**
CEV	***CI***	**0.69-0.85**	**.**	**0.91-0.99**	**0.94-0.98**	**0.89-0.98**	**0.94-0.98**	**0.94-0.98**
					ELD 8 Classes			
	Eval1	0.50	2.1	0.51	0.80	0.64	0.65	0.80
	Eval2	0.83	3.9	0.96	0.97	0.95	0.97	0.97
	Eval3	0.80	0.2	0.98	0.97	0.95	0.97	0.94
	Eval4	0.47	-1.9	0.92	0.93	0.87	0.93	0.93
	Eval5	0.63	5.2	0.68	0.84	0.70	0.81	0.92
	Eval6	0.80	-0.3	0.98	0.97	0.95	0.97	0.97
	Eval7	0.67	0.9	0.80	0.91	0.82	0.89	0.93
	Eval8	0.73	3.1	0.99	0.93	0.87	0.93	0.92
	Eval9	0.67	1.8	0.98	0.98	0.97	0.98	0.91
	$\bar{x}$	**0.68**	**.**	**0.87**	**0.92**	**0.86**	**0.90**	**0.92**
	CI	**0.59-0.76**	**.**	**0.76-0.98**	**0.88-0.96**	**0.78-0.94**	**0.83-0.97**	**0.89-0.95**
					ELD 7 Classes			
	Eval1	0.87	1.7	0.86	0.86	0.88	0.93	0.96
	Eval2	0.77	2.6	0.98	0.94	0.97	0.98	0.96
	Eval3	0.77	3.3	0.97	0.93	0.95	0.97	0.96
	Eval4	0.80	-1.2	0.99	0.95	0.98	0.99	0.97
	$\bar{x}$	**0.80**	**.**	**0.95**	**0.92**	**0.95**	**0.97**	**0.96**
FV	***CI***	**0.77-0.83**	**.**	**0.91-0.99**	**0.96-0.98**	**0.91-0.97**	**0.95-0.98**	**0.96-0.97**
					ELD 8 Classes			
	Eval1	0.87	-1.2	0.99	0.99	0.98	0.98	0.98
	Eval2	0.87	1.3	0.98	0.99	0.98	0.98	0.98
	Eval3	0.90	0.5	0.99	0.99	0.99	1.0	0.99
	Eval4	0.87	-0.3	1.01	0.99	0.99	0.99	0.98
	$\bar{x}$	**0.88**	**.**	**0.99**	**0.99**	**0.99**	**0.99**	**0.98**
	CI	**0.86-0.88**	**.**	**0.98-1.0**	**0.99**	**0.98-0.99**	**0.98-0.99**	**0.98-0.99**

[X] **CEV** = Controlled Environment Validation**. FV** = Field Validation. [Y] Proportion of the correct estimation number among the total sheets assessment. [Z] The fit of Coefficient of Determination (R^2) and correlations (*r*) were significant $p<0.001$.

Rater 1 was notoriously the most inaccurate ($\beta = 0.51$), with erratic assessment of underestimation and overestimation in classes 3 – 6. In reproducibility, this cluster was *medium* ($r = 0.80 – 0.91$) with *moderate* agreement ($\kappa_w = 0.80 – 0.92$; LCC = 0.65 – 0.86; Figure 2A2, Figure 3B, and Table 2). Although raters 4 and 6 were close to main cluster, they were significant different ($p = 0.78$ and 0.56, respectively), associated with strong underestimation or overestimation in classes 4 – 5 ($\beta = 0.92, 0.99$), which resulted in *medium* precision ($R^2 = 0.87$) and *moderate* agreement ($\kappa_w = 0.93$; LCC = 0.93; Figure 2B1-B2, Figure 3B, and Table 2). Comparatively, LDS-7c was 15 % more accurate, with underestimation in classes 4 – 6 and overestimation in classes 1 – 3 ($\beta = 0.91 – 0.99$; Table 2). Conversely, in LDS-8c, underestimation in classes 1 – 7 was more evident ($\beta = 0.76 – 0.98$). LDS-7c was 9 % more precise with *high* to *very high* metrics ($R^2 = 0.89 – 0.99$ vs. $R^2 = 0.78 – 0.94$; Table 2). Overall, precision, accuracy, and reproducibility were similar to analogous studies using scales (Ortega-Acosta *et al.*, 2016; Martelli *et al.*, 2017; Perina *et al.*, 2019; da Silva *et al.*, 2019; Franceschi *et al.*, 2019; Arias *et al.*, 2020). The second evaluation improved overestimation bias in the LDS-7c and the precision in 8 LDS-8c (Table 1 and 4). Furthermore, the best-clusters assessed by scale had similar coefficients (Table 2), suggesting that training increases efficiency regardless of classes number (Librelon *et al.*, 2015; Telíz-Ortíz *et al.*, 2003; Martelli *et al.*, 2017).

Lin's Concordance Correlation Coefficient (LCC), with an increase in the implementation over the last decade because it integrates precision and accuracy into the same statistic (Del Ponte *et al.*, 2017; 2022), showed better parameters for LDS-7c (LCC = 0.96, *CI* = 0.94 – 0.98; Table 2) than LDS-7c (LCC = 0.9, *CI* = 0.83 – 0.97; Table 2). As reported in previous studies, the LCC did not show explanatory differences with respect to *r* and R^2, which had similar trends (Table 2) (Perina *et al.*, 2019). A Pearson correlation among the three parameters showed a similarity of $R^2 > 0.98$ between β, R^2, and LCC. The proposal to include in this study the weighted Cohen's Kappa index (κw), in addition to the LCC, to determine the agreement degree with the real value, validated the results of linear coefficients and LCC, showing better agreement for LDS-7c (κw = 0.94 – 0.98) where 60 % of raters had a *very good* agreement, in contrast to LDS-8c (κw = 0.89 – 0.95) with 50 % of raters in *moderate* and 22 % *very low* (Table 2). The κw index overall corresponded with β, R^2, and LCC at $r = 0.70$; interestingly, it was higher in LDS-8c ($r = 0.86$) associated with a broad distribution of assessment errors between all severity classes (Figures 2-5 and Table 2). Comparing each indicator respect to the proportion of real hits, κw was higher than LCC with $r = 0.91$ for LDS-7c and $r = 0.68$ for LDS-8c, and similar respect to β and R^2, therefore it can be considered as a complementary indicator to linear statistics in the absence of other statistics like LCC (Arias *et al.*, 2020; Del Ponte *et al.*, 2017; Martelli *et al.*, 2017; Ortega-Acosta *et al.*, 2016; Librelon *et al.*, 2015).

Field validation (FV) of LDS-7c and LDS-8c scales. At field condition validation, LDS-7c maintained accuracy (β = 0.95, *CI* = 0.91 – 0.99), precision (R^2 = 0.95, *CI* = 0.92 – 0.97), and reproducibility (*r* = 0.92, *CI* = 0.89 – 0.95) compared to CEV (Figures 2-5 and Table 2). The parametric analysis of this scale showed a main cluster comprising the real values and raters 2 – 4 (p = 0.93) with closely *perfect* accuracy (β > 0.97) and *very high* precision (R^2 > 0.95) (Figure 4A2, Figure 5A, and Table 2). Regarding reproducibility in this cluster, *very high* values (*r* = 0.95 – 0.97) suggested an optimal error margin for field conditions in *Phaseolus coccineus* (Gonzalez-Cruces *et al.*, 2022; Librelon *et al.*, 2015). Rater 1 formed a significantly different clade (p > 0.93) with lowest assessment of accuracy, precision, and reproducibility (β = 0.86, R^2 = 0.88, *r* = 0.86), tending to underestimate or overestimate classes 5 – 6 (Figure 4A1-A2, Figure 5A, and Table 2). In agreement, the LCC (0.97, *CI* = 0.95 – 0.98) and κw (0.96, *CI* = 0.96 – 0.97) showed no differences compared to the controlled environment evaluation. Although accuracy, precision, and agreement appeared to show analogous trends under field conditions, in classes lower 16.6 % severity, particularly the class 4 which was underestimated in CEV, its discrimination improved notably in the field (Figures 2A2 and 4A2). This effect, reported by Del Ponte *et al.* (2022), where the benefit of using scales is marginal and greater at lower severity classes, for a study with epidemiological purposes it is appropriate in order to improve precision in detecting early disease phases where intervening in the epidemic process can be more effective (Gonzalez-Cruces *et al.*, 2022; Mora-Aguilera *et al.*, 2021).

On the contrary, the LDS-8c scale performed better metrics compared to the CEV process, and even better than LDS-7c in field. The analysis formed a main cluster with real values and raters 3 – 4 (p = 0.99). Raters 1 – 2 formed independent clades without significant differences (p > 0.85, Figures 4B2 and 5B). Accuracy improved considerably and was close to *perfect* (β = 0.98 – 1.01), with main underestimation errors in classes 4 – 6 (Figures 4B1-B2, 5B, and Table 2). Precision increased notably to 0.98 – 0.99 in four raters, showing a significant improvement with 8 classes at the field assessment (Figures 4B2, 5B, and Table 2). The reproducibility of LDS-8c among raters had the best assessment in comparison to the real (*r* = 0.99), even despite slight overestimation in classes 2 – 4 (Figures 4B3, 5B, Table 2). In agreement, both LCC (0.99, *CI* = 0.98 – 0.999) and κw (0.98, *CI* = 0.99 – 0.999) achieved optimal agreement levels and significantly improved compared to CEV process (Table 2). LCC maintained similarity with linear statistics (Perina *et al.*, 2019). Overall, the eight-class scale applied in a field assessment improved in statistical metrics to optimal levels desired for a logarithmic-diagrammatic scale (LDS). Qualitatively, initial damage attributable to classes 1 - 4 could be differentiated with greater precision and biological criteria than ELD-7c, which enables the strategies design aimed at interrupting chains of infection prior to 7.1 %

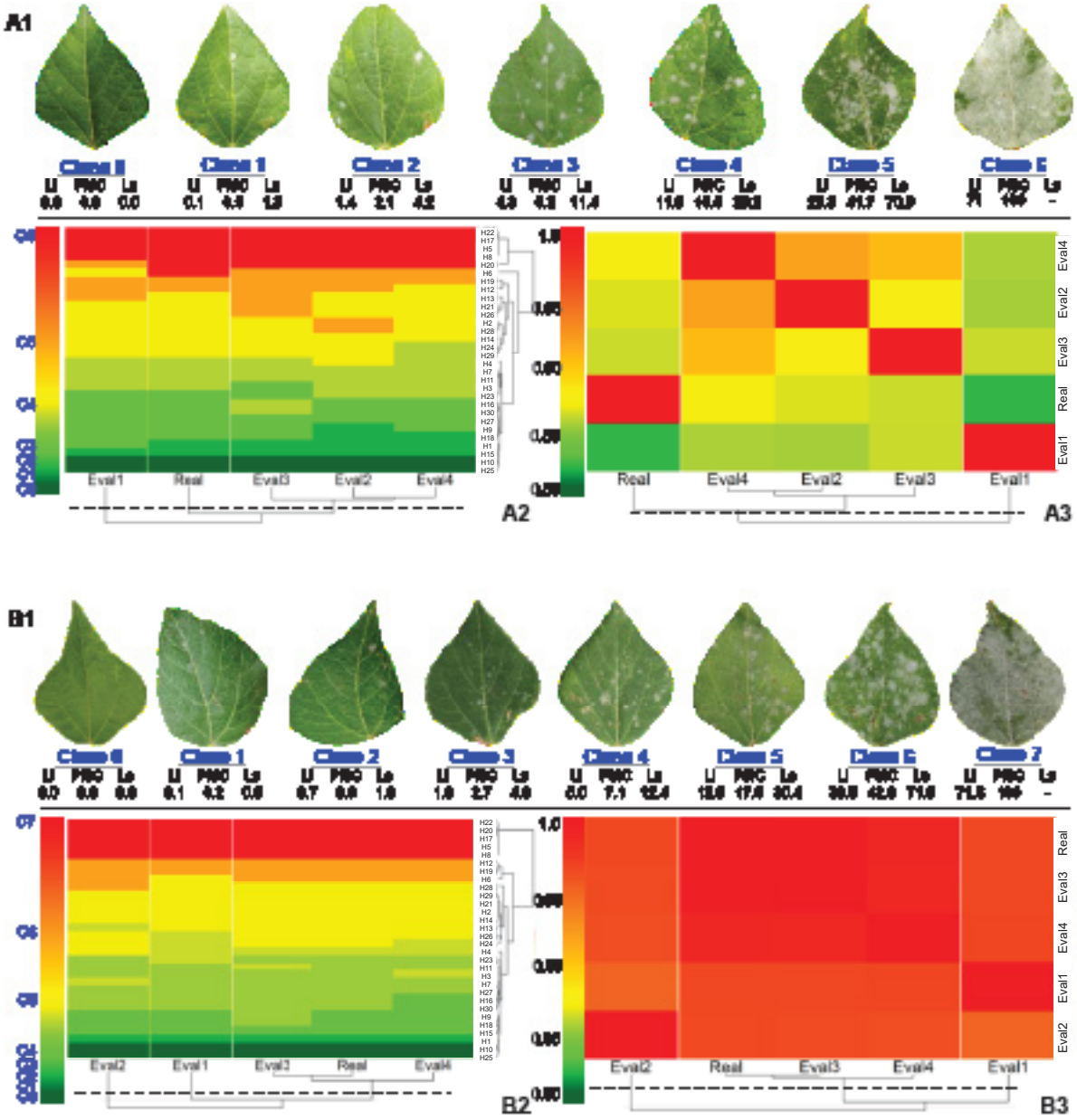


Figure 4. **A1 and B1**. Logarithmic-diagrammatic scale of 7 and 8 classes for assessing powdery mildew severity on Ayocote bean (*P. coccineus*) leaves during the validation process of 30 leaves in the field by four selected raters. **A2 and B2**. *Heatmap* of Pearson correlation coefficient (*r*) among nine raters based on severity scale. Values of $r = 0.8 – 1$ indicate the reproducibility level of each scale among raters. **A3 and B3.** *Heatmap* of severity class on 30 leaves assessed by scale and rater. The color represents the class value assigned by the rater to each leaf. *Hierarchical cluster analysis* is performed by grouping raters and leaves using the '*complete*' method and *Euclidean* distance.

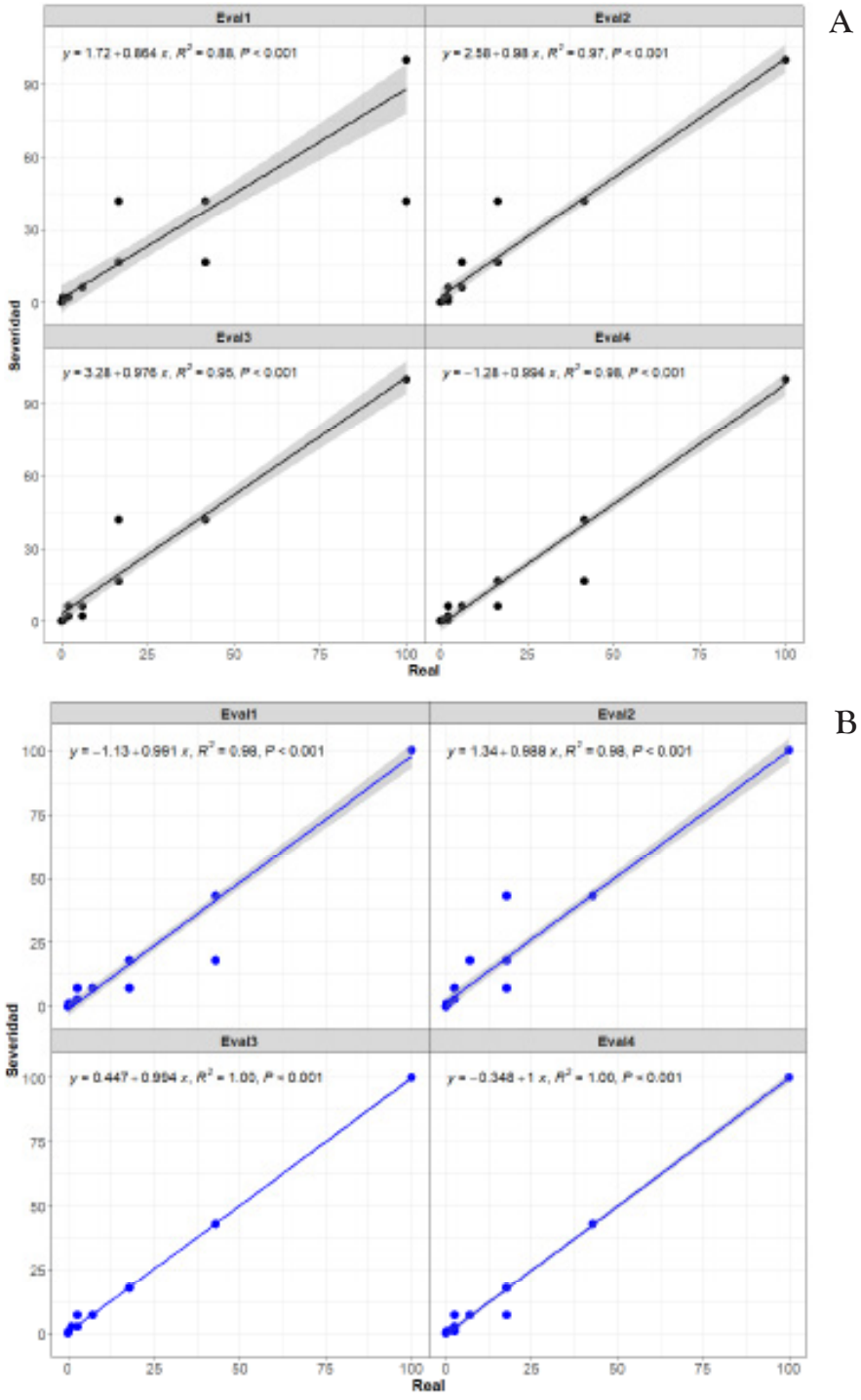


Figure 5. Correlation plots between severity (y) assessed by scale and actual values (x) from four raters during Field Validation (VCa) with 30 *Phaseolus coccineus* leaves. The linear regression equation ($y = \beta o + \beta x + e$) is fitted to determine parameters β, R^2, and p-value using the *stat_poly_eq* function. **A.** LDS-7classes. **B.** LDS-8classes.

severity (L_L = 5.3, U_L = 12.4; Figure 1), integrating into Surveillance Systems for disease monitoring, in breeding or control programs (Gonzalez-Cruces *et al.*, 2022; Mora-Aguilera *et al.*, 2021; Franceschi *et al.*, 2019; Ortega-Acosta *et al.*, 2016).

RGB image analysis (13mpx) taken with DJI® Phantom 3 Drone. The analysis at 27 m within the 15 selected quadrants of the study plot (Figure 6A1) allowed estimating an average powdery mildew severity of 8.4 % (*CI* = 5.3 – 12.6 %) and a canopy coverage index (VCI) of 0.88 (*CI* = 0.76 – 0.99), contrasting with the field assessment which obtained a severity of 47 % (*CI* = 38.8 – 55.3 %) and VCI = 0.46 (*CI* = 0.76 – 0.99) (Figure 6A2 and Table 3). The image analysis revealed an inverse relationship between severity and VCI (R^2 = 0.68), suggesting that higher foliage density corresponds to lower damage intensity (Table 3). However, in comparison to severity and plant canopy index (PVI) assessed with App-Monitor v1.0, showed a directly proportional trend, although not significant (R^2 = 0.2) due to severity varying within the experimental plot (Figure 6). The classification algorithm underestimated by 46.4 % to 63.7 % in quadrants with VCI > 0.94, associated with the occurrence of *Erysiphe vignae* primarily in the lower to middle canopy strata, less exposed for aerial imaging (Table 3, Figure 6A1-10A2).

These findings suggest that the image classification algorithm was less efficient in estimating foliar severity associated with non-systemic fungal organisms (i.e., *Erysiphe vignae*) when VCI was low, indicating potential overestimation of severity in quadrants with limited tissue availability (Table 3, Figure 6A2). However, field assessments also revealed significant bias in severity compared to PVI (R^2 = 0.19, Table 3), linked to prior discussions on successful pathogenesis processes in organisms thriving under favorable developmental microclimates within foliage, coupled with reduced exposure to sunlight for *P. coccineus* plants (Craig & Weyne, 2012; Hückelhoven & Panstruga, 2011). This suggests that estimation error is associated more with biological factors of the pathosystem than with digital analysis processes. It has been documented that such analyses may have favorable implications for other disease organisms such as viruses, wilt diseases, blights, etc., or for calculating indices and epidemiological indicators in crops (Gonzalez-Cruces *et al.*, 2022). Overall, exploratory results from image analysis suggest that these methodologies hold significant epidemiological potential; however, they should be complemented with field assessments linked to sampling systems (i.e., sampling pattern) designed *ad hoc* based on biological-epidemiological criteria of the pathosystem, enabling informed decision-making.

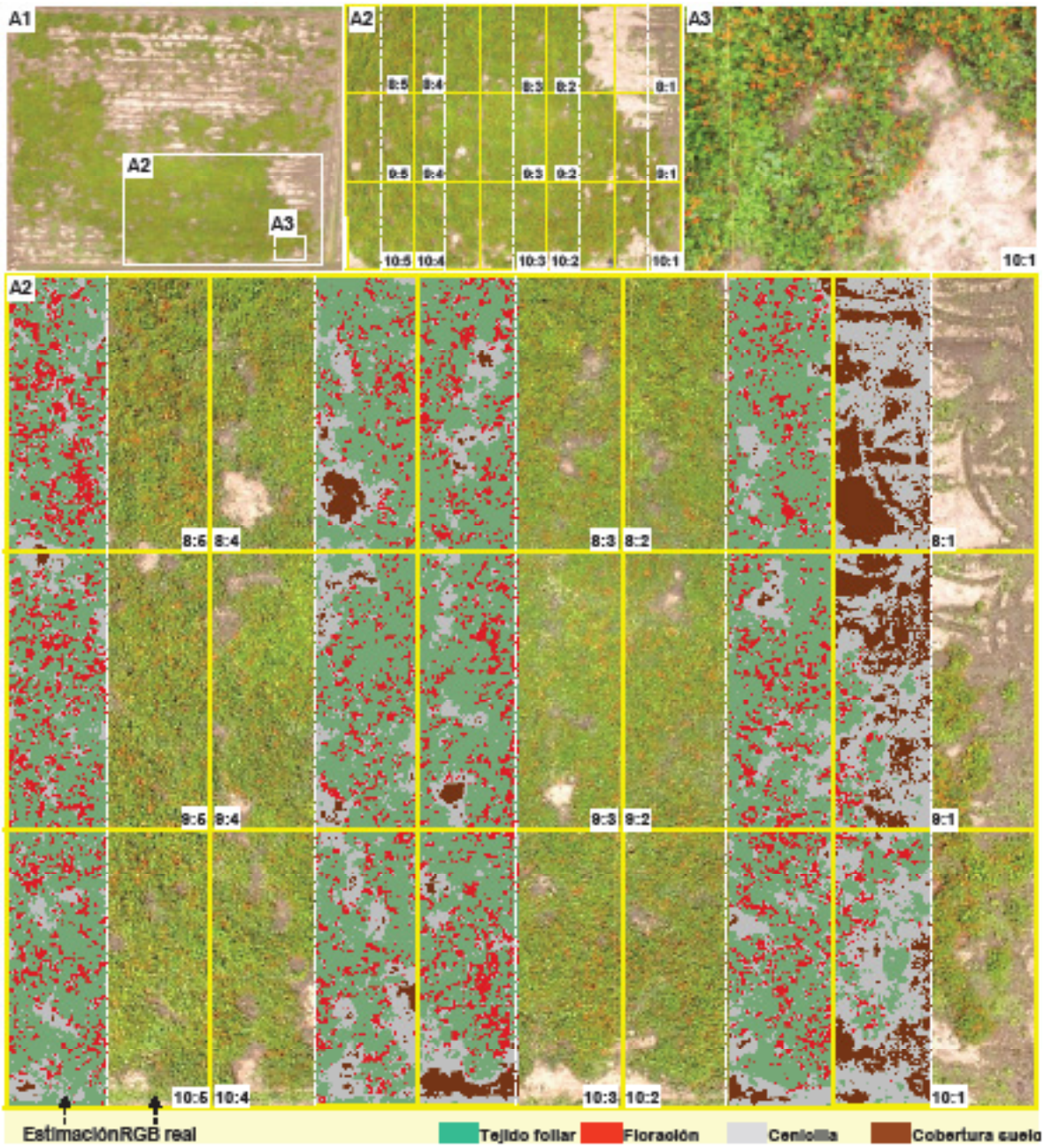
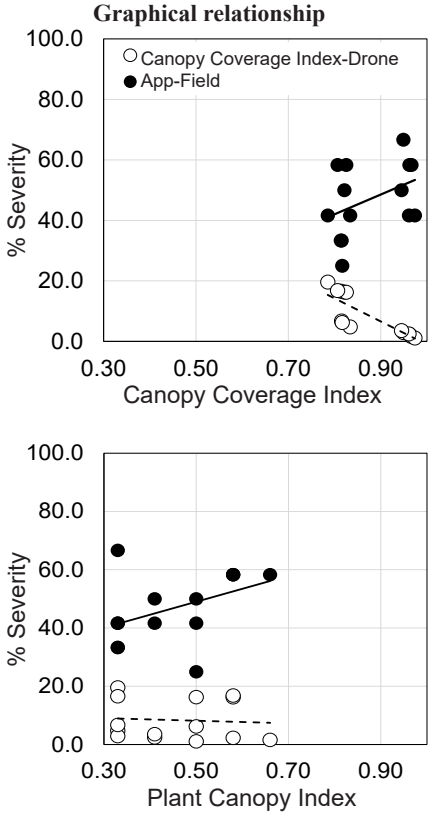


Figure 6. Estimation of canopy and severity indicators using RGB imagery (13 mpx) from Phantom 3 processed through supervised segmentation algorithm in ArcMap® v10.3. **A1.** Image of the total experimental area (40 x 52 m). Captured at 50 m altitude. **A2.** Block of 15 selected quadrants based on uniformity in host continuity, canopy, and maximum inductivity. Continuous yellow lines depict quadrant divisions. Dashed white lines represent selected blocks for algorithm versus real image estimation. Captured at 27 m. **A3.** Image at 5 m of a selected sector for designing *'RGB signature'* with crop categories (foliar tissue, flowering, powdery mildew, and soil coverage).

Table 3. Comparison of canopy coverage index (VCI), plant canopy (PVI), and powdery mildew severity percentage estimated using RGB-drone image and field assessments.

C:B [x]	Drone		Fiel [y]	
	Canopy coverage index	% Severity	Plant canopy index	% Severity
8:1	0.01	.	0.00	0.00
8:2	0.83	4.8	0.33	41.7
8:3	0.97	1.6	0.66	58.3
8:4	0.96	2.3	0.58	58.3
8:5	0.95	2.9	0.33	66.7
9:1	0.82	6.7	0.33	33.3
9:2	0.82	6.2	0.50	25.0
9:3	0.97	1.1	0.50	41.7
9:4	0.96	2.5	0.41	41.7
9:5	0.95	3.6	0.41	50.0
10:1	0.79	19.6	0.33	41.7
10:2	0.81	16.6	0.33	33.3
10:3	0.82	16.3	0.50	50.0
10:4	0.83	16.2	0.58	58.3
10:5	0.81	16.8	0.58	58.3
$\bar{x}$ (±sd)	**0.88 (±0.24)**	**8.4 (±7.1)**	**0.46 (±0.16)**	**47.0 (±16.8)**
CI	**0.76 – 0.99**	**5.3 – 12.6**	**0.38 – 0.53**	**38.8 – 55.3**

Graphical relationship



[x] Quadrant (C) and Block (B) IDs. Field estimates with App-Monitor v1.0 using 5 classes (0-100%) (Armenta-Cárdenas *et al.*, 2024). SD = Standard Deviation, *CI* = Confidence Interval (α = 0.05).

Conclusions

The validation process in controlled environments and field enabled the development of scales with optimal quantitative statistical parameters. In field, the selected and validated logarithmic-diagrammatic scales of seven and eight classes exhibited best accuracy, precision, and agreement values for assessment powdery mildew severity caused by *Erysiphe vignae* in Ayocote beans. However, for epidemiological purposes aimed at managing larger-scale crops, the use of eight classes is recommended as it provided greater precision in the initial phases of the *E. vignae* pathogenesis process. Regarding the use of images for damage estimation, the underestimation of severity estimated with RGB-drone images compared to field assessments with App-Monitor suggests the need to establish complementary digital and ground-based methods due to the occurrence of *E.*

vignae in the middle plant canopy and its heterogeneous foliar coverage habits. Additionally, other digital methodologies should be tested in addition to drone imagery. A methodology for developing logarithmic-diagrammatic scales is proposed, which includes image capture, processing, and quantification, as well as validation in controlled environments and fields. For scale validation, precision metrics (R^2); accuracy (β); reproducibility (Pearson coefficient and *Hierarchical cluster analysis*); and agreement (Lin's coefficient and Kappa Index) are proposed, for the first time, comprehensively. RGB-drone images are proposed to estimate an integrated canopy coverage index and severity.

Acknowledgments

The authors acknowledge CONAHCYT for the scholarship awarded to graduate students and the CP-LANREF team for their logistical and operational support in carrying out field activities.

Literature Cited

Arias MS, Guerrero ÁGE and González PPA. 2020. Diagrammatic scale for measuring severity of gray mould in thornless Castilla blackberry (*Rubus glaucus* Benth). Crop protection Ciencia Rural 50:11, e20190859. https://doi.org/10.1590/0103-8478cr20190859.

Armenta-Cárdenas MJ, Ávila-Alistac N, Zúñiga-Romano MC, Acevedo-Sánchez G, Muñoz-Alcalá A, Gómez-Mercado R, Coria-Contreras JJ, Gutiérrez-Esquivel D, Cruz-Izquierdo S, García-González I, Bibiano-Nava O and Mora-Aguilera G. 2024. Epidemiological etiology of *Erysiphe* sp. and putative viral and phytoplasma-like symptoms in Ayocote bean (*Phaseolus coccineus*). Mexican Journal of Phytopathology 42 (2): 11. https://doi.org/10.18781/R. MEX.FIT.2310-7.

Ávila-Alistac N, Rivas-Valencia P y Espinosa-Calderón A (Eds.). 2023. El Maíz y Frijol en México: Etiología, Epidemiología y Mejoramiento Genético. Ed. Sociedad Mexicana de Fitopatología, A.C. (SMF). Primera Edición. ISBN:978-607-69698. https://doi.org/10.18781/R.MEX.FIT:2024-1.

Craig A and Weyne WF. 2012. Effects of sunlight exposure on grapevine powdery mildew development. Phytopathology 102 (9): 857-866. https://doi.org/10.1094/PHYTO-07-11-0205.

da Silva GCBM, Pio R, Pereira RCM, Peche PM and Pozza EA. 2019. Development and validation of a severity scale for assessment of fig rust. Phytopathologia Mediterranea 58 (3): 597-605. https://doi.org/10.14601/Phyto-10967

Del Ponte EM, Cazón LI, Alves KS, Pethybridge SJ and Bock CH. 2022. How much do standard area diagrams improve accuracy of visual estimates of the percentage area diseased? A systematic review and meta-analysis. Tropical Plant Pathology 47: 43–57. https://doi.org/10.1007/s40858-021-00479-5

Del Ponte EM, Pethybridge SJ, Bock CH, Michereff SJ, Machado FJ and Spolti P. 2017. Standard area diagrams for aiding severity estimation: scientometrics, pathosystems, and methodological trends in the last 25 years. Phytopathology 107:1161–1174. https://doi.org/10.1094/PHYTO-02-17-0069-FI.

Franceschi VT, Alves KS, Mazaro SM, Godoy CV, Duarte HSS and Del Ponte EM. 2020. A new standard area diagram set for assessment of severity of soybean rust improves accuracy of estimates and optimizes resource use. Plant Pathology 69: 495–505. https://doi.org/10.1111/ppa.13148.

Godoy CV, Carneiro SMTPG, Iamauti MT, Pria MD, Amorim L, Berger RD and Filho AB. 1997. Diagrammatic scales for bean diseases: development and validation. Journal of Plant Diseases and Protection 104(4): 336–345. http://www.jstor.org/stable/43215167.

Gonzalez-Cruces A, Arista-Carmona E, Díaz-Arias KV, Ramírez-Razo K, Hernández-Livera A, Acevedo-Sánchez G, Mendoza-Ramos C and Mora-Aguilera G. 2022. Epidemiology of *Bean common mosaic virus* and *Alternaria alternata* in 12 *Phaseolus vulgaris* genotypes. Mexican Journal of Phytopathology 40 (2): 188-220. https://doi.org/10.18781/R.MEX.FIT.2202-8.

Hückelhoven R and Panstruga R. 2011. Cell biology of the plant–powdery mildew interaction. Current Opinion in Plant Biology 14 (6): 738-746. https://doi.org/10.1016/j.pbi.2011.08.002.

Librelon SS, Souza EA, Pereira R, Pozza EA and Abreu AFB. 2015. Diagrammatic scale to evaluate angular leaf spot severity in primary leaves of common bean. Australasian Plant Pathology 44: 385–395. https://doi.org/10.1007/s13313-015-0360-9.

Martelli IB, Pacheco CdeA, Bastianel M, Schinor EH, Conceição Pmda and Azevedo de FA. 2017. Diagrammatic scale for assessing foliar symptoms of *Alternaria* brown spot in citrus. Agronomy Science and Biotechnology 2 (2): 57. https://doi.org/10.33158/ASB.2016v2i2p57.

Mora-Aguilera G, Acevedo-Sánchez G, Guzmán-Hernández E, Flores-Colorado OE, Coria-Contreras JJ, Mendoza-Ramos C, Martínez-Bustamante VI, López-Buenfil A, González-Gómez R and Javier-López MÁ. 2021. Web-based epidemiological surveillance systems and applications to coffee rust disease. Mexican Journal of Phytopathology 39 (3): 452-492. https://doi.org/10.18781/r.mex.fit.2104-6.

Nutter JRFW and 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 (2): 174–184. https://doi.org/10.1080/07060669509500709.

Ortega-Acosta SA, Velasco-Cruz C, Hernández-Morales J, Ochoa-Martínez DL and Hernández-Ruiz J. 2016. Diagrammatic logarithmic scales for assess the severity of spotted leaves and calyces of roselle. Mexican Journal of Phytopathology 34: 270-285. https://doi.org/10.18781/R.MEX.FIT.1606-6.

Perina FJ, Belan LL, Moreira SI, Nery EM, Alves E and Pozza EA. 2019. Diagrammatic scale for assessment of *Alternaria* brown spot severity on tangerine leaves. Journal Plant Pathology 101: 981–990. https://doi.org/10.1007/s42161-019-00306-6.

Teliz-Ortiz D, Mora-Aguilera G and Ávila-Quezada GD. 2003. Logarithmic systems for measuring severity of anthracnose and scab in avocado fruits. Proceedings V World Avocado Congress, Mexico. 585-589 p. http://www.avocadosource.com/WAC5/Papers/wac5_p585.pdf.