Multivariate Comparison of Papaya Ringspot Epidemics

Gustavo Mora-Aguilera, Daniel Nieto-Angel, C. Lee Campbell, Daniel Téliz, and Eliseo García

First, second, and fourth authors: research assistant, research assistant, and professor, respectively, Instituto de Fitosanidad, Colegio de Postgraduados, 56230 Montecillo-Chapingo, México; third author: professor, Department of Plant Pathology, North Carolina State University, Raleigh 27695-7616; and fifth author: research assistant, Colegio de Postgraduados-Campus Veracruz, 56230 Montecillo-Chapingo, México.

We thank D. Riestra, L. Matheis, A. Mora, and E. Rosas for their technical assistance. We thank S. Webb, D. Chellemi, and T. Gottwald for their comments on previous drafts of this paper. We also thank the anonymous reviewers for their criticisms and suggestions.
We thank CONACYT/México and Colegio de Postgraduados for the financial support they provided.
Accepted for publication 13 September 1995.

ABSTRACT

Mora-Aguilera, G., Nieto-Angel, D., Campbell, C. L., Téliz, D., and García, E. 1996. Multivariate comparison of papaya ringspot epidemics. Phytopathology 86:70-78.

Nine epidemiological parameters, estimated to characterize each of 60 disease progress curves of papaya ringspot (papaya ringspot virus type P [PRSV-P]) incidence on papaya (*Carica papaya*), were analyzed using a combination of principal component and cluster analysis techniques to select the best subset of variables with the highest explanatory capacity and to classify epidemics according to their degree of similarity. The effects of cultural methods, transplanting date, planting density, and field location on epidemic classification and control of papaya ringspot also were examined. Standardized area under disease progress curve ($AUDPC_s$), shape parameter (c) of the Weibull distribution function, and time between transplant date and first symptoms (X_o) were selected, using principal component analysis, as the most important variables and represented 83.5% of the overall variance. The remaining variability was explained by the apparent infection rate (r_G-Gompertz) standardized by the Richard's method, the scale parameter (b) of the Weibull distribution function, initial and final disease incidence, time to reach 50% incidence, and time of epidemic duration. Five epidemic categories were defined through a direct linkage of the principal component scores associated with $AUDPC_s$, c, and X_o, via cluster analysis. Experimental site and transplanting date had more influence on the definition of epidemic categories than did planting density. No category indicated complete suppression of papaya ringspot. Epidemics associated with early transplanting dates (February, April, and June) were in the category representing the smallest $AUDPC_s$ ($\approx$ 25%-days day^{-1}), the longest time to first symptoms ($\geq$ 120 days), and low values of c (1 to 3). $AUDPC_s$ was more sensitive to experimental site than the other factors. Transplanting dates in February and June were suggested as additional alternatives to April and May transplanting dates practiced by growers, to delay the epidemic onset and to reduce the incidence of papaya ringspot of Central Veracruz, México.

Additional keywords: comparative epidemiology, virus epidemiology.

The study of experimental manipulations on the development of epidemics of a particular disease or the ecological behavior of epidemics of various diseases can be investigated using a number of distinct epidemiological approaches (7). When a relatively large series of plant disease epidemics are studied with different cultivars, cultural practices, or pathosystems and the overall goal is to characterize them using descriptive variables, it is practical to use a comparative epidemiological approach (1,22,25). Comparison of epidemics is often carried out by univariate analyses such as slope comparisons of linearizing transformations of disease progress curves (4,11) or parameter comparisons of epidemic curves described by nonlinear models (14). Multivariate analyses, however, appear to provide a more complete characterization of epidemics by taking into account several biological facets of disease progress curves or disease development (1,8,9,22,24).

The comparison of epidemics using a set of variables simultaneously, which underlies a multivariate analysis, encounters, however, at least one important problem: the presence of collinearity, i.e., the existence of variables highly correlated to others in the data set (8,22). For example, final disease incidence has been found to be correlated with initial level of inoculum and area under disease progress curve (AUDPC) in some pathosystems (8,22,25). Although such types of correlations may provide clues to understand the development of epidemics (8), collinearity should be eliminated or reduced in comparative studies. Pragmatic situations, in which collinearity is undesirable in the comparison of epidemics, apply when assessing resistance to various diseases (1,9,23) or when measuring the effect of cultural practices.

A corollary stage in multivariate comparison of epidemics is the classification of disease progress curves according to a set of variables (8,22). Often this stage is seen as being independent of the problem of collinearity and, thus, it is perceived as a standalone analytical stage (1,9).

This research was concerned with the multivariate comparison of 60 epidemics induced in papaya (*Carica papaya* L.) by papaya ringspot virus type P (PRSV-P) with three different agronomic sites, six transplant dates, and four planting densities. These cultural practices were established to assess the control of papaya ringspot under the tropical conditions of Central Veracruz, México. A set of variables associated with the disease progress curves was investigated for problems of collinearity through principal component analysis (PCA). The effects of cultural practices on the selected variables were then studied by classifying the epidemics via cluster analysis (CA) using principal components as input variables (19).

Papaya ringspot has been studied with respect to temporal and spatial development of epidemics and with regard to the implications of biological and climatic variables on disease spread in the tropical conditions of southeastern México (2,15,29,30,36). In

Corresponding author: C. Lee Campbell; E-mail address: lee_campbell@ncsu.edu

Publication no. P-1996-1110-01R

this research, the influence of location, planting date, and density on epidemic classification was determined and evaluated from the perspective of control of papaya ringspot.

MATERIALS AND METHODS

Establishment of field plantings. Seventy-two plantings of papaya type Cera were established in three fields at the Veracruz station of the Instituto de Recursos Naturales at Paso San Juan, Veracruz (19° 10′N, 98° 16′W). The field at site A was nearly level. The field at site B, approximately 2 km from site A, was located on a hill with a slope of 3 to 4%. The field at site C was flat and was located about 1 km from sites A and B. Each of the three sites had clay-loam soil, and maize was the previous crop.

The papaya plantations were established at each site on six transplanting dates (beginning December 1987 and every 2 months thereafter), with planting densities of 135, 84, 51, and 42 plants per plot (6.6 by 40 m) to give an equivalent of 4,444, 2,500, 1,600, and 1,111 plants/ha, respectively. Standard farming practices were used in all plantations. Twelve plantings, corresponding to the first three transplanting dates at site C, were damaged and lost because of the effects of heavy rain and strong winds. Therefore, data corresponding to a total of 60 papaya plantings were included in this study.

Disease assessment. From the time of transplanting in December 1987 until August 1989 each plant in each of the 60 plantings was observed every 2 weeks for symptoms typical of natural infection by papaya ringspot virus type P (PRSV-P). Typical symptoms included mosaic and distortion of leaves, watery spots on the stem and leaf petioles, and ringspots on the fruit. Disease progress curves were obtained from the cumulative proportion of plants with typical symptoms of PRSV-P over all dates of assessment.

Curve elements. The 60 disease progress curves were characterized with 10 associated variables (Fig. 1). One of these descriptive variables was the rate of apparent infection (r). This variable was estimated with the slope parameter of the monomolecular, Gompertz, and logistic models in subsequent linear regression analyses. These models were evaluated for goodness-of-fit to the data of the 60 epidemics by examination of r^2-values and standard deviations of the estimated slope parameter (7). The GLM procedure of SAS (35) was used for these analyses. The best model fitted to each epidemic was selected and the values of the slope parameter from these various models were transformed onto a standardized scale through the use of the weighted mean absolute rate of disease incidence as ρ-parameter of the Richard's model (7) for the overall linear model that was selected most frequently.

Epidemics also were characterized with the Weibull distribution function (32) modified as a two-parameter model. This model was selected because of its flexibility (32) and capacity to describe disease progress curves more naturally than some linear models (14,22). This model can be written as:

$$y = \exp\left\{-\left(\frac{t}{b}\right)^{c}\right\}, \quad t > 0$$

In which y = disease incidence, t = time of disease assessment (days) after transplanting, b = scale parameter, and c = shape parameter.

Because b and c can be interpreted biologically (32), these parameters were selected as descriptive variables. The b parameter is related inversely to apparent infection rate (r), and the c parameter is related to the shape and slope of the density function (dy/dt per unit) for disease progress curve (32). The initial estimates of b and c were made empirically to begin the iterations of nonlinear regression (DUD method) with a nonlinear regression procedure (35). Additional descriptive variables included: AUDPC estimated with the midpoint rule for area estimation (7) programmed in SAS; AUDPC standardized by duration time in days ($AUDPC_s = AUDPC/T_t$) (7); initial incidence of disease (Y_o); final disease incidence (Y_f); time in days to reach 50% disease incidence (T_{50}); total duration of the epidemic in days measured from the time of first observation of symptoms to fruit harvest (T_t); and time from transplant to appearance of the first symptoms (X_o) (Fig. 1).

Outliers detection and variables selection. PCA was performed to reduce the original set of descriptive variables. This smaller set of variables ideally should allow the characterization of the original data set without loss of explanatory capacity, because of the elimination of variables with a relatively high degree of correlation (i.e., redundant) (17,18,19). The data matrix used in this analysis was constructed with 60 epidemics and 10 descriptive variables. Although PCA automatically normalizes variables when a correlation matrix is used, the angular transformation (arc sine $[Y_i]^{1/2}$) was applied to Y_o and Y_f in order to normalize these specific variables. This transformation was selected because of the broad ranges of Y_o and Y_f.

Reduction of the original set of variables was achieved in several steps as described previously (27). In the first step, potential multivariate outliers (i.e., those that are detected only when several variables are studied simultaneously) were detected and tested as influential observations in the estimation of coefficients, variances, and scores of principal components. The methods of Hawkins and Fatti (17), Pack et al. (31), and an extension of the biplot technique suggested by Gabriel (13) were integrated and performed with PRINCOMP and FACTOR procedures of SAS (35). A detailed description of this approach is presented elsewhere (28). In the second step, the correlation matrix of the original variables, as calculated with the SAS PRINCOMP procedure (35), was examined to detect the degree of collinearity among variables. Variables were grouped into two sets of less correlated variables within-group to avoid the effect of collinearity in the estimation of variances and coefficients by PCA (17). Each group of variables was evaluated with PCA to identify principal compo-

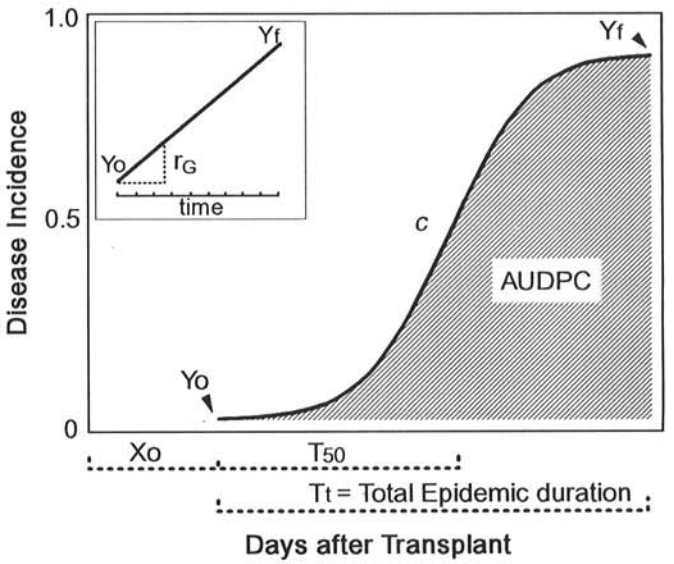


Fig. 1. Variables estimated for comparison of papaya ringspot epidemics on papaya at Veracruz, México. X_o is time in days from transplanting in the field until first symptoms were detected; T_{50} is time in days to reach 50% disease incidence; T_t is total duration of an epidemic in days; Y_o and Y_f are the initial and final disease incidence, respectively, measured in percentage (%); AUDPC is the area under disease progress curve (%-days), $AUDPC_s$ (not shown) is the AUDPC standardized by dividing AUDPC by T_t; c and b are respectively, the curve-shape and scale parameters estimated by the Weibull model (b is not shown in the figure); r_G is the apparent infection rate (per unit day^{-1}) standardized by Richard's method to the Gompertz model (estimated by the slope of the line fitted to each epidemic).

TABLE 1. Numbers of papaya ringspot epidemics in papaya (*Carica papaya* L.) best described by the Gompertz, logistic, monomolecular, and Weibull models and ranges of the coefficient of determination (r^2), growth rates, and mean square errors (MSE) of the estimate of the growth rate

Model	Number of epidemics[a]	r^2	Growth rate[b]	MSE[c]
Gompertz	28	0.73 – 0.99	0.043 – 0.009	0.0008 – 0.0097
Logistic	23	0.94 – 0.99	0.082 – 0.021	0.0002 – 0.0011
Monomolecular	9	0.95 – 0.98	0.027 – 0.002	0.0001 – 0.0007
Weibull	60	0.94 – 0.99	1/80.4 – 1/458.3	0.0012 – 0.0431

[a] Number of epidemics out of 60 best fitted with a particular model.
[b] Growth rate estimated with the slope of the Gompertz, logistic, and monomolecular linearized model forms and with the inverse of the Weibull scale parameter (b).
[c] Mean square error of the estimate of the growth rate.

TABLE 2. Mean, standard deviation, and range for 10 variables associated to disease progress curves of papaya ringspot in papaya (*Carica papaya* L.)

Variable (units)[a]	Mean	Standard deviation	Range
AUDPC (%-days)	7584.44	2479.30	3085.57 – 13071.43
$AUDPC_s$ (%-days day^{-1})	36.45	10.56	14.96 – 61.6
T_t (days)	209.63	35.86	111.00 – 263.0
T_{50} (days)	116.07	40.30	57.00 – 211.0
X_o (days)	87.54	32.89	52.00 – 199.0
b (per unit day^{-1})$^{-1}$	186.32	71.00	80.45 – 458.3
c (-)	3.54	2.28	0.620 – 9.900
r_G (per unit day^{-1})	0.02	0.01	0.004 – 0.050
Y_o (proportion incidence)	0.02	0.02	0.001 – 0.104
Y_f (proportion incidence)	0.82	0.17	0.357 – 1.000

[a] X_o is time in days from transplanting in the field until first symptoms were detected; T_{50} is time in days to reach 50% disease incidence; T_t is total duration of an epidemic in days; Y_o and Y_f are the initial and final disease incidence, respectively, measured in percentage (%); AUDPC is the area under disease progress curve (%-days), $AUDPC_s$ is the AUDPC standardized by dividing AUDPC by T_t; c and b are respectively, the curve-shape and scale parameters estimated by the Weibull model; r is the apparent infection rate (per unit day^{-1}) standardized by Richard's method to the Gompertz model (estimated by the slope of the line fitted to each epidemic).

nents that accounted for a small amount of variance ($\lambda < 0.7$, λ = eigenvalue) (18,19,31) and to select variables for elimination from those principal components by using the approach of minor principal components described by Hawkins and Fatti (17). In the third step, a final elimination of variables was done by combining the remaining variables, and the analysis was performed again with PCA. The third step was repeated until the number of variables remaining was similar to the number of major principal components. Biplot displays of major principal components (those with a large variance, $\lambda > 0.7$) (17,31) both with and without varimax rotation were used to assist in the elimination of variables (1,5,13,35).

Classification of epidemics. Principal components associated with the variables retained above were subject to multivariate analysis of variance (MANOVA) and CA (35) for the purpose of classification of epidemics. In this way, these procedures were applied to the orthogonal multivariate space produced by PCA and, thus, took advantage of its normality and noncollinearity properties. The aim of MANOVA was to compare the overall influence of site on the development of epidemics and to determine whether or not the CA should be applied to each experimental location. The classification of epidemics was achieved with the AVERAGE clustering method of the CLUSTER procedure followed by the TREE and PLOT procedures of SAS to diagram the structure of clusters (dendrogram) (35). Dendrograms were cleaved at the highest range of the similarity index to achieve the best definition of clusters (34).

RESULTS

Curve elements. Of the 60 epidemics, 28 were best described by the Gompertz disease progress model, nine by the monomolecular model, and 23 by the logistic model. With one exception, the best model fitted to each epidemic had $r^2 \geq 0.94$ (Table 1). For comparison purposes, the estimates of the rate of disease progress for epidemics described by the monomolecular and logistic model were transformed with the Richard's procedure (7) to provide values of parameters rate equivalent to those for the Gompertz model. Thus, only actual or adjusted r_G values were used for comparative purposes. The Weibull model adequately

TABLE 3. Correlation coefficients and levels of significance for 10 standardized epidemiological variables estimated from 60 epidemics of papaya ringspot in papaya (*Carica papaya* L.)

Variables[a]	Y_o	Y_f	T_t	T_{50}	c	b	$AUDPC_s$	AUDPC	X_o
r_G	0.125	0.857[b]	–0.733	–0.594	0.714	–0.649	0.328	–0.092	–0.227
	0.343	0.0001	0.0001	0.0001	0.0001	0.0001	0.004	0.732	0.040
Y_o		0.147	–0.144	–0.192	–0.062	–0.057	0.181	0.087	0.122
		0.380	0.183	0.018	0.532	0.665	0.133	0.521	0.345
Y_f			–0.453	–0.620	0.515	–0.831	0.597	0.391	–0.353
			0.011	0.0001	0.0001	0.0001	0.0001	0.0004	0.001
T_t				0.495	–0.501	–0.307	–0.177	0.377	–0.137
				0.0001	0.0001	0.041	0.246	0.002	0.245
T_{50}					–0.203	0.685	–0.669	–0.361	–0.107
					0.397	0.0001	0.0001	0.012	0.416
c						–0.222	–0.167	–0.429	0.003
						0.057	0.315	0.002	0.828
b							–0.683	–0.465	0.407
							0.0001	0.0001	0.0002
$AUDPC_s$								0.839	–0.140
								0.0001	0.126
AUDPC									–0.206
									0.046

[a] X_o is time in days from transplanting in the field until first symptoms were detected; T_{50} is time in days to reach 50% disease incidence; T_t is total duration of an epidemic in days; Y_o and Y_f are the initial and final disease incidence, respectively, measured in percentage (%); AUDPC is the area under disease progress curve (%-days), $AUDPC_s$ is the AUDPC standardized by dividing AUDPC by T_t; c and b are respectively, the curve-shape and scale parameters estimated by the Weibull model; r is the apparent infection rate (per unit day^{-1}) standardized by Richard's method to the Gompertz model (estimated by the slope of the line fitted to each epidemic).
[b] Underlined numbers represent relatively high ($r > 0.60$) correlation between the variables involved.

described all epidemics ($r^2 \geq 0.94$) (Table 1). Mean value for b was 186.3 and for c was 3.5 among all epidemics (Table 2).

Outliers detection and variables selection. Epidemics 4, 8, 10, and 59 were detected as outliers. These epidemics had at least one variable with a value outside the range expected for a normal distribution. However, the exclusion of these epidemics did not change the estimation of coefficients and variances by PCA substantively in comparison with the case when all epidemics were included in PCA. The same number of principal components was selected whether potential outliers were excluded from the data matrix or the full data set was used. Thus, data recorded in epidemics 4, 8, 10, and 59 were not considered as distorting observations, and all epidemics were considered in successive analyses.

When the correlation matrix of the 10 standardized descriptive variables was inspected, several relatively large correlations were found between some curve elements (Table 3). The highest correlations were Y_f with r_G, b, and T_{50} (r = 0.86, –0.83, and –0.62, respectively); AUDPC$_s$ with AUDPC, b, and T_{50} (r = 0.84, –0.68, and –0.67, respectively); and r_G with T_t, c, and b (r = 0.71, –0.73, and –0.65, respectively). The high correlation between AUDPC$_s$ and AUDPC was expected, and only the variable AUDPC$_s$ was kept in the rest of the analyses.

High correlations among several descriptive variables also were detected in biplot displays after varimax rotation (Fig. 2). For the purpose of the biplot analysis, the first four factors ($\lambda \geq 0.7$) were selected and accounted for 47.2, 17.9, 15.0, and 9.4% of the total variance in the data. In the biplot made with the first two factors, the plane represented 65.1% (47.2 and 17.9%) of the total variance (Fig. 2A). Most of this variance was explained by all variables except Y_o and X_o, which were represented by shorter vectors. Thus, X_o and Y_o were not taken into account in the inspection of correlations for this multivariate subspace. In this biplot (Fig. 2A), r_G was positively correlated to Y_f and c, and negatively correlated to T_t, as indicated by small (« 90°) and large (≅ 180°) vector angles. Similarly, Y_f was negatively correlated to b and T_{50}.

In the biplot made with the third and four factors (Fig. 2B), the plane represented 24.4% (15.0 and 9.4%) of the total variance, most of which was explained by Y_o and X_o which were represented by the longest and more parallel vectors to axes. The more parallel the vectors are to the axes, the better the overall represented variance is explained by the plane. Additionally, because the vectors associated with X_o and Y_o were perpendicular to each other, these variables were fairly independent (Table 3, Fig. 2B).

From the correlation matrix and the biplot display results, two groups of variables with low within-group collinearity were selected. The first group comprised of r_G, AUDPC$_s$, b, and Y_o; the second group comprised of T_t, T_{50}, X_o, c, and Y_f. After PCA was performed independently for each group, b and r_G from the first group and T_t and Y_f from the second group were eliminated (Table 4).

In three successive PCA with the pooled set of remaining variables, T_{50} and Y_o were eliminated. In the final of these successive PCA, the variables AUDPC$_s$, c, and X_o were retained and associated with three principal components that had $\lambda > 0.7$. The proportion of variance explained by these variables was 83.5% of the total for the nine epidemic parameters. Thus, the essential dimensionality of the data was three with one variable associated with each particular axis of a 3-dimensional plot (Figs. 3A to 5A). The first axis, defined by AUDPC$_s$, explained 31.5% of the overall variance; the second axis, defined by the c-parameter, contributed 28.6%; and the third axis, defined by X_o, contributed 23.3%.

Classification of epidemics. CA was performed with the principal components associated with the variables' AUDPC$_s$, c, and X_o. Because of differential influences of site in the development of epidemics as indicated with MANOVA (Wilks lambda, $p < 0.01$), CA was performed independently for each site. Clusters in the dendrograms and 3-dimensional plots were defined by a cutoff at 0.7 Euclidean distance units (Figs. 3 to 5). At site A, six general epidemic types were defined. Each general type was associated with individual clusters (Fig. 3). A total of 11 epidemics characterized by the smallest AUDPC$_s$, relatively late times to epidemic onset (X_o), and small curve-shape values (c) were associated with two clusters defined by April and June transplanting dates at the

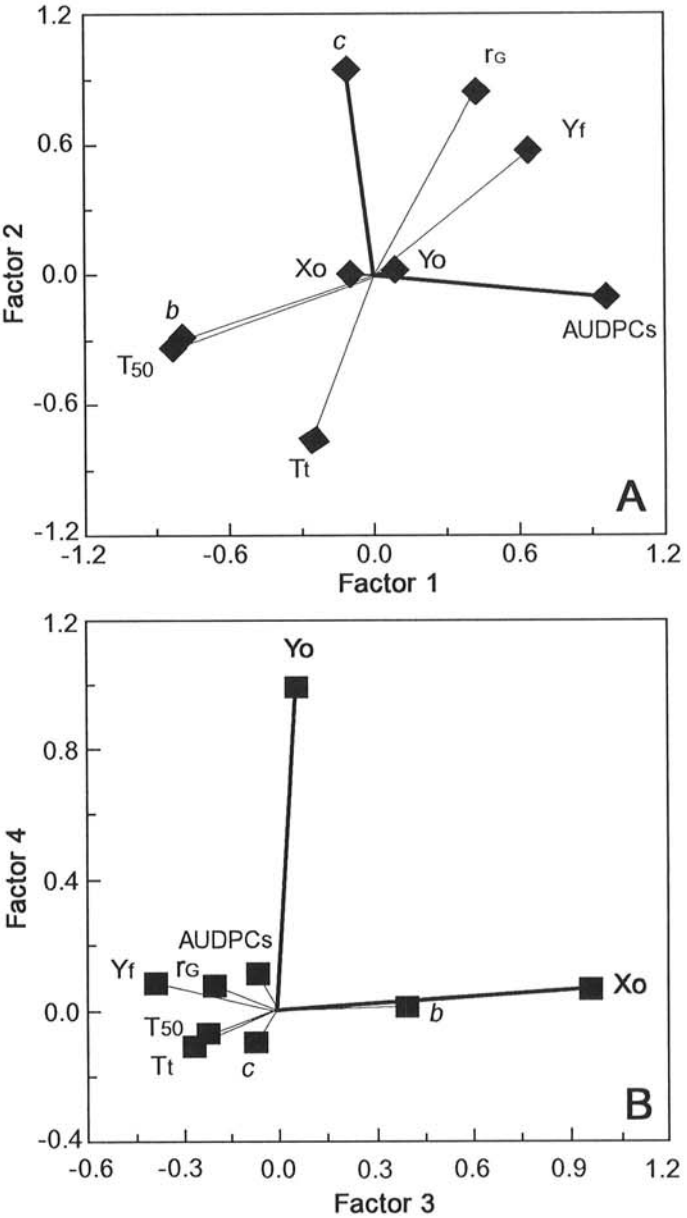


Fig. 2. Biplots of approximate correlation and relative importance of variables measured for 60 papaya ringspot epidemics, at Veracruz, México. **A,** Varimax rotated biplot of the first and second factors. **B,** Varimax rotated biplot of the third and fourth factors. Longer vectors fairly parallel to an axis (thicker lines) represent variables that are best represented by the two factors biplotted. Small vectors represent variables that are not well-represented by the plane and may be associated with any of the remaining factors not plotted. Variables with small angles between vectors and vectors in the opposite direction are positively and negatively correlated, respectively. X_o is time in days from transplanting in the field until first symptoms were detected; T_{50} is time in days to reach 50% disease incidence; T_t is total duration of an epidemic in days; Y_o and Y_f are the initial and final disease incidence, respectively, measured in percentage (%); AUDPC is the area under disease progress curve (%-days), AUDPC$_s$ is the AUDPC standardized by dividing AUDPC by T_t; c and b are respectively, the curve-shape and scale parameters estimated by the Weibull model; r_G is the apparent infection rate (per unit day^{-1}) standardized by Richard's method to the Gompertz model (estimated by the slope of the line fitted to each epidemic).

TABLE 4. Eigenvectors and eigenvalues of principal components derived from two groups of epidemiological variables associated to papaya ringspot epidemics in papaya (*Carica papaya* L.)

Variables[a]	Principal components				
	C1	C2	C3	C4	C5
			Group 1[b]		
r_G	0.528[c]	−0.092	0.737	0.410	...
$AUDPC_s$	0.551	0.026	−0.669	0.496	...
Y_o	0.155	0.977	0.077	−0.119	...
b	−0.626	0.188	0.051	0.754	...
Eigenvalues	2.194	0.984	0.640	0.180	...
			Group 2		
T_t	0.503	−0.289	−0.133	0.798	−0.083
T_{50}	0.482	−0.159	0.658	−0.197	0.518
X_o	0.051	0.880	−0.000	0.321	0.345
c	−0.451	0.025	0.727	0.379	−0.349
Y_f	−0.553	−0.340	−0.141	0.274	0.694
Eigenvalues	2.408	1.201	0.807	0.427	0.155

[a] X_o is time in days from transplanting in the field until first symptoms were detected; T_{50} is time in days to reach 50% disease incidence; T_t is total duration of an epidemic in days; Y_o and Y_f are the initial and final disease incidence, respectively, measured in percentage (%); AUDPC is the area under disease progress curve (%-days), $AUDPC_s$ is the AUDPC standardized by dividing AUDPC by T_t; c and b are respectively, the curve-shape and scale parameters estimated by the Weibull model; r is the apparent infection rate (per unit day^{-1}) standardized by Richard's method to the Gompertz model (estimated by the slope of the line fitted to each epidemic).

[b] Groups with low within-group collinearity selected through scrutiny of the correlation matrix and the biplots display. Group 1 has only four principal components due to the number of variables.

[c] Column of numbers in each group of variables contain the coefficients of the eigenvector associated with each principal component. Underlined numbers represent the highest coefficient values associated with a minor principal component; variables associated with such coefficient values were considered redundant and excluded from subsequent analyses. A minor principal component was characterized with an eigenvalue (λ) smaller than 0.70.

four planting densities. Epidemics in plantings established in February at 4,444 and 1,600 plants/ha, and September at 1,600 plants/ha were also included in one of these clusters. Seven epidemics with the largest $AUDPC_s$ and the smallest X_o and c-value were associated with a cluster that included three of the four planting densities of September and all planting densities in November transplants. Epidemics in plantings established in December were split into three clusters with the largest c-value, and in two of these clusters with large $AUDPC_s$ (Fig. 3A).

At site B, six general epidemic types were defined (Fig. 4). Six epidemics characterized by the smallest $AUDPC_s$, relatively large X_o, and small c-values were associated with one cluster defined by the February transplanting date at the four planting densities. Epidemics associated with June and December transplanting dates at 4,444 and 2,500 plants/ha, respectively, were also included in this cluster. Nine epidemics characterized by relatively large $AUDPC_s$, the longest X_o, and the smallest c-value were associated with two clusters defined by the four planting densities of April. Epidemics associated with June and November at 1,600 and 4,444 plants/ha, respectively, and December at 4,444, 1,600, and 1,111 plants/ha were also included in these clusters. The four epidemics of September were well-defined by a cluster characterized with large $AUDPC_s$, the shortest X_o, and the largest c-value. Three and two epidemics from low planting densities of November and June, respectively, were classified in two clusters defined with large $AUDPC_s$, short X_o, and small c-values (Fig. 4A).

At site C, five general epidemic types were determined (Fig. 5). A total of three epidemics characterized with the smallest $AUDPC_s$, the largest X_o, and small c-values were associated with one cluster defined by the June transplanting date. The higher planting density of this transplanting date defined by itself a single

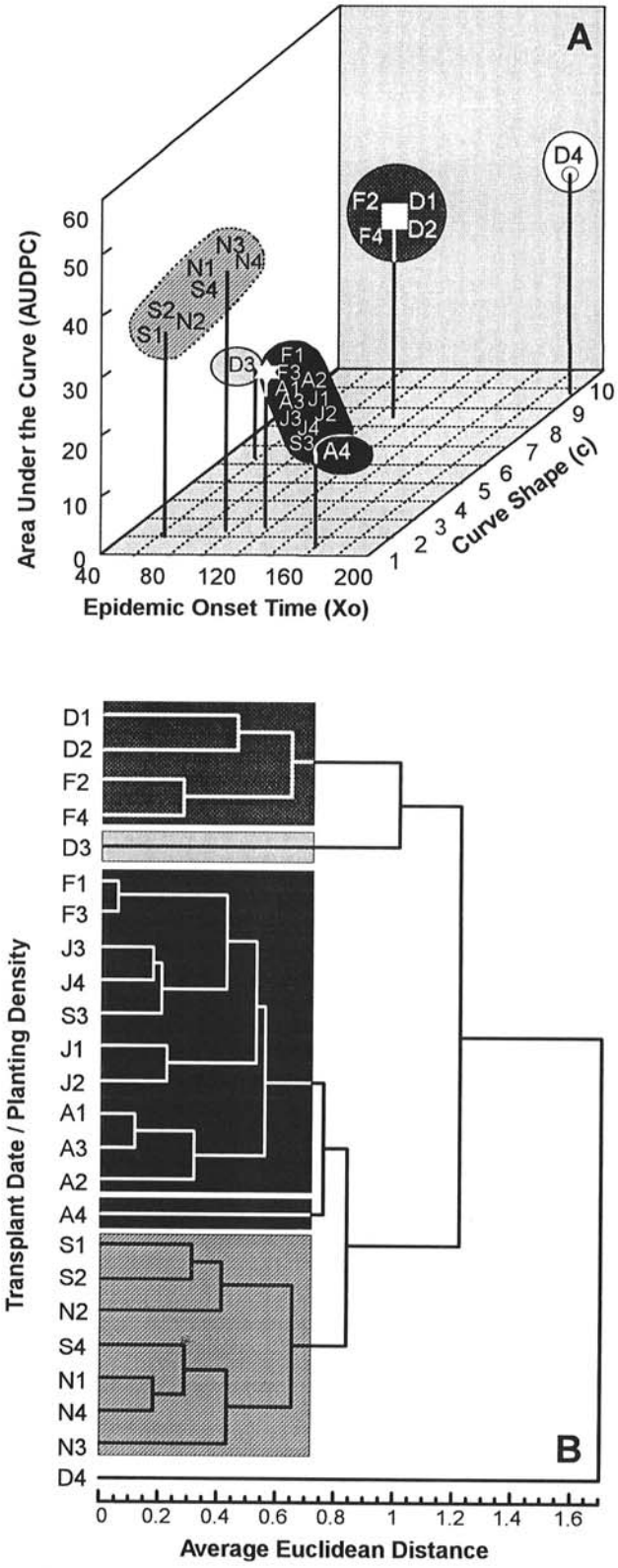
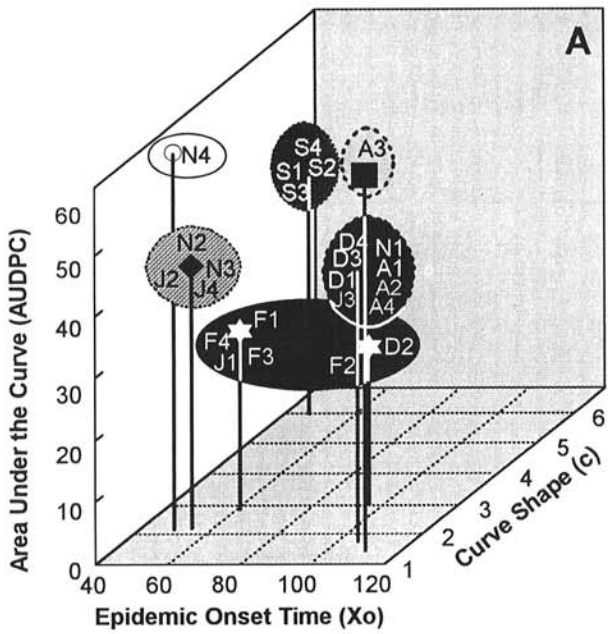


Fig. 3. Site A, Veracruz, México. Classification of papaya ringspot epidemics according to cluster analysis of scores of three principal component obtained after three successive principal component analyses. **A,** Epidemic categories represented by the original values of $AUDPC_s$, X_o, and c associated with the three principal components. Lines projected on to each cluster are average values of $AUDPC_s$, X_o, and c for a cluster or subcluster. Gray-scale shading correspond to the same cluster defined in the dendrogram. **B,** Dendrogram representing relative similarities among 24 epidemics. Gray-scale shading indicates clusters defined at 0.7 Euclidean distance units. Alphanumeric fields displayed within clusters in A and vertically in B, represent transplant date (F = February, A = April, J = June, S = September, N = November, D = December) and planting density (1 = 4,444, 2 = 2,500, 3 = 1,600, 4 = 1,111 plants/ha).



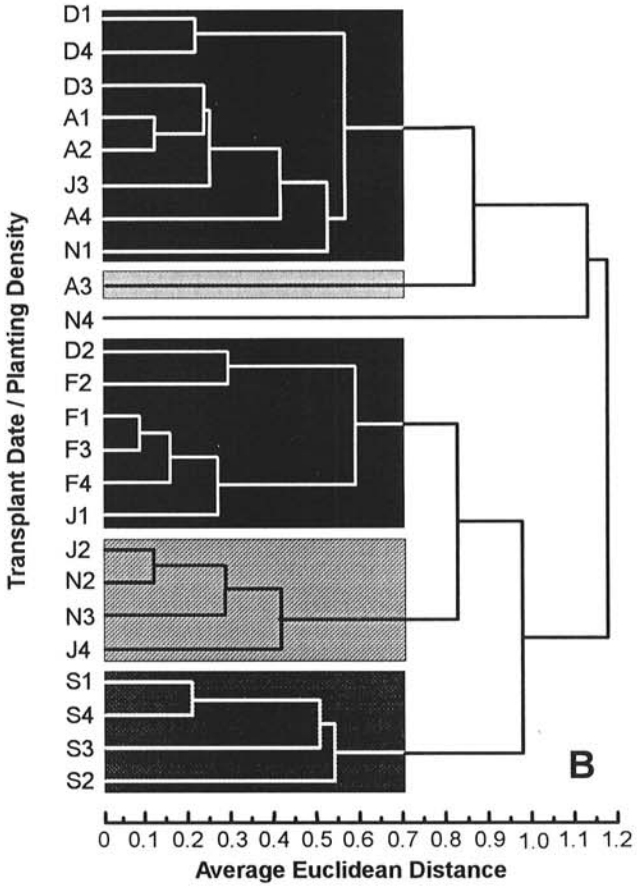
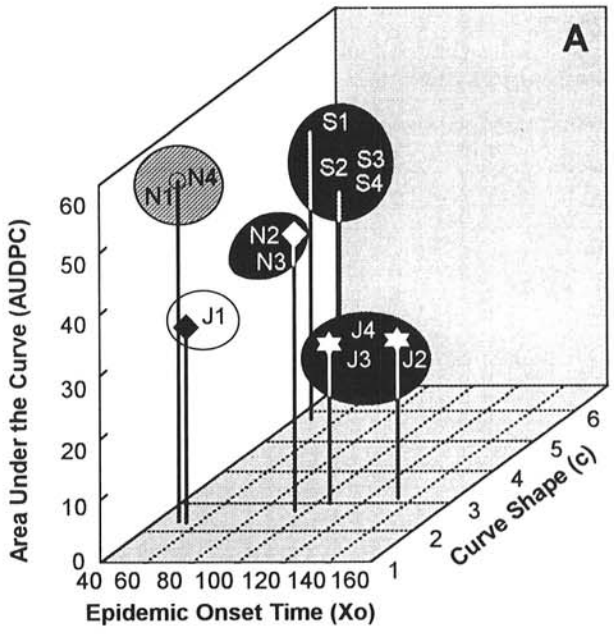


Fig. 4. Site B, Veracruz, México. Classification of papaya ringspot epidemics according to cluster analysis of scores of three principal components obtained after three successive principal component analyses. **A,** Epidemic categories represented by the original values of AUDPC$_s$, X_o, and c associated with the three principal components. Lines projected on to each cluster are average values of AUDPC$_s$, X_o, and c for a cluster or subcluster. Gray-scale shading correspond to the same cluster defined in the dendrogram. **B,** Dendrogram representing relative similarities among 24 epidemics. Gray-scale shading indicates clusters defined at 0.7 Euclidean distance units. Alphanumeric fields displayed within clusters in A and vertically in B, represent transplant date (F = February, A = April, J = June, S = September, N = November, D = December) and planting density (1 = 4,444, 2 = 2,500, 3 = 1,600, 4 = 1,111 plants/ha).



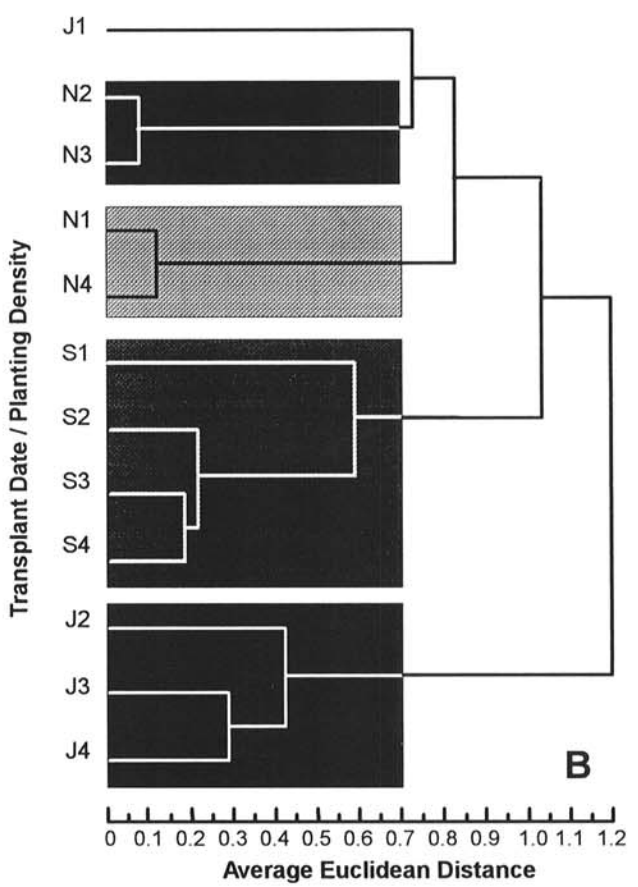


Fig. 5. Site C, Veracruz, México. Site A, Veracruz, México. Classification of papaya ringspot epidemics according to cluster analysis of scores of three principal components obtained after three successive principal component analyses. **A,** Epidemic categories represented by the original values of AUDPC$_s$, X_o, and c associated with the three principal components. Lines projected on to each cluster are average values of AUDPC$_s$, X_o, and c for a cluster or subcluster. Gray-scale shading correspond to the same cluster defined in the dendrogram. **B,** Dendrogram representing relative similarities among 12 epidemics. Gray-scale shading indicates clusters defined at 0.7 Euclidean distance units. Alphanumeric fields displayed within clusters in A and vertically in B, represent transplant date (F = February, A = April, J = June, S = September, N = November, D = December) and planting density (1 = 4,444, 2 = 2500, 3 = 1,600, 4 = 1,111 plants/ha).

cluster with similar characteristics except that X_o was shorter than the previous cluster. The four epidemics of September and November were classified in one and two clusters, respectively; these groups of epidemics were characterized by the largest $AUDPC_s$ and shortest X_o, and, in the case of November clusters, with a small c-value. The September epidemics cluster had the largest c-value (Fig. 5A).

DISCUSSION

The examination of descriptive elements of disease progress curves in studies of comparative epidemiology was originally suggested by Kranz (22). Several researchers have used this approach to compare epidemics which occurred with different cultivars (1,8,9), cultural practices (14), or pathosystems (22,24). These attempts were successful because the curve elements utilized represented certain biological facets that were expressed in the disease progress curve as a whole (22). In our research, some of the curve elements suggested by Kranz (22) were included and were complemented with the b and c parameters of the Weibull model (8,9) to introduce nonlinear elements into the analysis. Nonlinearity has been suggested to reveal important properties of disease progress such as asymptotic behavior and shape of the curve (14).

PCA has been applied to detect collinearity among variables and to identify those variables that could explain most of the variance of a data matrix used in comparison of epidemics (8,22,25). Additionally, CA has been applied, as the next step, to classify epidemics according to their similarity. Grouping has been done previously with the complete set of original variables without the removal of spurious variables (1,9,24). In our analysis, several variations to this general approach were taken by combining different statistical methods (5,13,17,18,31). A direct linking of a reduced set of transformed variables (principal components) with CA was suggested ideally to improve the clustering process and derive more stable and valid conclusions in multivariate comparison of epidemics.

The consideration of multivariate outliers (i.e., those aberrant observations that are usually expressed when variables as a whole are considered in a multivariate space [17,31]) is important because conclusions in comparative epidemiology usually rely upon a set of variables rather than on single ones. In this study, none of the four outliers were categorized as an influential observation; therefore, no correction or elimination of epidemics was made.

Collinearity is a common problem in the multivariate comparison of epidemics (8,22). Interrelations among the curve elements of disease progress should be determined prior to comparative attempts (1,8). In this study, Y_o and X_o were not well-correlated (r-values less than –0.19 and 0.41, respectively) with any variable. This result is similar to that of Campbell et al. (8), but differs from findings of Kranz (22) in 40 different host-pathogen combinations. In the study of Kranz (22), X_o and Y_o were highly correlated with the "consequences of the epidemics" and to Y_f and AUDPC, respectively, indicating that epidemic build up was correlated with the initial inoculum level. $AUDPC_s$ was not correlated highly with r_G, as was found by Kranz (22). $AUDPC_s$ was, however, correlated relatively highly with b and T_{50}, and r_G was correlated highly with Y_f, T_i, c, and b (Table 3). The correlation among b and r_G was previously indicated by Pennypacker et al. (32) as a general property of b of the Weibull model. These results suggest that interrelations among epidemic curve parameters may be pathosystem dependent. For example, in epidemics of papaya ringspot epidemics, the initial inoculum level (Y_o) and the epidemic onset (X_o) did not contribute to the build up of the epidemics as might be expected for epidemics involving an aerially dispersed, fungal, polycyclic pathogen. Rather, epidemic build up relied upon the arrival of transient, viruliferous aphids from sources outside the field (29,30).

In the initial screening for redundant variables, the minor principal components approach (17), applied to two groups of less intercorrelated variables, was complemented with the observation of biplot displays, particularly after varimax rotation (5,13,20). It can be argued that subsetting variables with less within-group correlation is somewhat subjective and that there is no guarantee that independent PCA with different groups of variables may lead to elimination of those variables that are highly correlated in the correlation matrix as a whole. However, this approach was fairly effective in our study, because the most highly correlated variables, r_G, b, Y_f, and T_i, were eliminated.

The approach of subsetting variables was suggested in our study to overcome problems associated in the estimation of eigenvalues and eigenvector coefficients when strong collinearity does exist among a relatively large number of variables. It is likely that the elimination of redundant variables under such a situation becomes a more complicated issue and acquires a high degree of subjectivity. For example, PCA with the whole set of variables may lead to the elimination of $AUDPC_s$ using the criteria of minor principal components alone (data not shown), when, in fact, that variable was proven to be an important descriptor in our data.

After three successive PCA, $AUDPC_s$, c of the Weibull model, and X_o were selected to characterize the papaya ringspot epidemics. Kranz suggested at least two variables to characterize an entire epidemic, one for measuring epidemic intensity (i.e., AUDPC) and one for determining the curve pattern (i.e., c) or for defining a position parameter value (i.e., X_o) (23). Our results showed that at least three variables, similar to those indicated by Kranz, were needed to avoid a detrimental effect on the overall variance explained. These variables contained similar percentages (31.5, 28.6, and 23.3%) of the overall variance (83.5%), which indicated their relative importance in the definition of clusters.

Successive PCA allowed a single association of axis-variables, thus determining a clear characterization of the different categories of epidemics defined by CA, i.e., each epidemic category can be described in terms of $AUDPC_s$, c, and X_o (Figs. 3A to 5A). Epidemiological interpretations are often more difficult to achieve when this property is not attained through data analysis (1,9,24,25). The completion of the intended analysis in the final PCA was reached following the criteria of Jolliffe (18) and Pack et al. (31) for selecting major principal components ($\lambda > 0.7$). A more conservative criteria for selecting principal components ($\lambda > 0.9$) (17) was not considered in our study because of the detrimental effect on the relative proportion of the explained variance. For example, using such criteria, $AUDPC_s$ would be eliminated and the remaining variables would explain 70.9% of the overall variance.

CA performed by using the principal components associated with $AUDPC_s$, c of the Weibull model, and X_o allowed us to identify the presence of at least five categories of epidemics for each sites at 0.7 Euclidean distance units (Figs. 3 to 5). Experimental site and transplanting date influenced the development of papaya ringspot epidemics more than planting density. The multivariate space defined by $AUDPC_s$, X_o, and c determined different number, properties, and composition of clusters at each site. Within each site, clusters were, in general, more determined by transplanting dates than by planting densities, i.e., the four densities of a particular transplanting date were allocated in the same or nearby clusters (Figs. 3 to 5). Only two epidemics from late transplanting dates (December and November) and the lowest density (1,111 plants/ha) were found forming two independent clusters at sites A and B (Figs. 3 and 4).

The influence of experimental site and transplanting date on papaya ringspot epidemic development may be explained by the population dynamics of the aphid vectors. Aphid populations exhibit a bimodal curve for population dynamics with catch numbers being highly variable among sites at Central Veracruz,

México, even at relatively short distances (about 2 km) (2,29,30). The highest population peak of the most important aphid vectors (*Aphis gossypii*, *A. nerii*, and *Myzus persicae*) is detected late in the spring and its impact on incidence of papaya ringspot can be detected early in the summer (29). However, aphid vectors are present during the whole year, which was demonstrated in that there was not any epidemic category which provided complete control of PRSV-P through cultural practices. Avoidance of the highest vector population peak through alteration of planting time is well documented in the literature (26,33).

Insensitivity of epidemic development to planting density may be due to the behavior of aphid vectors. Vectors transmit the pathogen nonpersistently and do not colonize papaya plants. Disease spread apparently exhibits uniform and/or random spatial pattern in the field (30), suggesting that an increase in disease incidence does not depend upon plant spacing. However, high planting density is advantageous in that it results in higher papaya production as reported with other virus diseases (6,10,33). The positive or negative influence of plant density on vector landing and, thus, on virus transmission found in the literature cannot be generalized to different virus diseases. Rather, results seem dependent upon the specific virus-vector relationship and crop system (6,10,16,21,26,33).

Two epidemic categories were of interest from the perspective of papaya ringspot control. The first epidemic category was characterized by the smallest $AUDPC_s$ ($\approx$ 25%-days day^{-1}), the latest X_o ($\geq$ 120 days), and small *c*-value (1 to 3) (Figs. 3A to 5A). This epidemic type was defined by transplant dates in February, April, or June for the three sites. The second epidemic category was characterized by properties similar to that in the first category, except for having a large $AUDPC_s$ ($\approx$ 50%-days day^{-1}). However, the relatively high $AUDPC_s$ value was accumulated late in the season, leaving the plant free of disease through most of the vegetative stage. This epidemic type was observed at site B, and was defined mostly by December and April transplanting dates. This result cannot be generalized to site C because plantings of the December and April transplanting dates, as well as that of February, were destroyed by heavy rains and strong winds.

Transplanting dates associated with the first epidemic category could be useful for controlling papaya ringspot in Central Veracruz, México. This epidemic type was consistently present among sites, included 20 epidemics overall, and was associated with early transplanting dates. The second category was restricted to the site where disease was more severe (site B), comprised eight epidemics, and was associated with early transplanting dates except February, which belonged to the first epidemic category. The second category seemed to characterize the effect of early transplanting dates under high inoculum pressure and did not invalidate the assumption of PRSV-P control by utilizing an early transplanting date.

Both epidemic categories had relatively small $AUDPC_s$ and X_o occurred at least 14 weeks after papaya transplanting, which means that primary infections in the field were detected when plants had 8 to 12 fully expanded mature leaves and had started flowering. Thus, disease had less effect on fruit production compared to the earlier epidemic onsets observed in other epidemic categories. Prevalence of the papaya ringspot during the last 3 to 6 months before harvesting can cause an average loss of 500 g per marketable fruit, and fruit appearance depends upon period of time in which plants remain infected (15). Therefore, transplanting dates such as February, April, or June, which allow the latest epidemic onset in addition to a small $AUDPC_s$, could be useful as part of an integrated program to control PRSV-P.

Growers commonly establish low density plantations (1,111 plants/ha) during late April and May to take advantage of the rainy season. Although other factors should be studied with regard to papaya production in the region, such as commercialization, labor, and water availability, the growers production scheme could be improved. February and June transplanting dates may be considered as an alternative to reduce the incidence of papaya ringspot in Central Veracruz, México. Higher planting density (4,444 plants/ha) also could be considered to increase papaya production. Results of this research and preliminary control of papaya ringspot attempts with the application of mineral and vegetable oil, different reflective material barriers, trap crops, and cross protection with a mild strain of PRSV-P from Hawaii (3,12,15,36,37) indicate that application of a single control method is not completely effective. Therefore, an interdisciplinary approach conducted in cooperation with growers at Central Veracruz to manage papaya ringspot is currently under investigation to validate and integrate these results with other control strategies for the papaya crop (15).

LITERATURE CITED

1. Anderson, W. F., Beute, M. K., Wynne, J. C., and Wongkaew, S. 1990. Statistical procedures for assessment of resistance in a multiple foliar disease complex of peanut. Phytopathology 80:1451-1459.
2. Becerra, N. L. 1987. Dinámica poblacional y hospederas de áfidos vectores del virus de la mancha anular del papayo en Veracruz. (Resumen). Congr. Nac. Fitopatol., 14th. Sociedad Mexicana de Fitopatologia, Morelia, Michoacan, Mexico.
3. Becerra, N. L. 1989. Preferencia al color de papayo (*Carica papaya* L.) y barreras de jamaica (*Hibiscus sabdariffa* L.) como medios para reducir la transmision por áfidos de virosis en papayo. Rev. Mex. Fitopatol. 7:218-222.
4. Berger, R. D. 1988. The analysis of effects of control measures on the development of epidemics. Pages 137-151 in: Experimental Techniques in Plant Disease Epidemiology. J. Kranz and J. Rotem, eds. Springer-Verlag KG, Berlin.
5. Bradu, D., and Gabriel, K. R. 1978. The biplot as a diagnostic tool for models of two-ways tables. Technometrics 20:47-68.
6. Brink, G. E., and McLaughlin, M. R. 1990. Influence of seeding rate and interplanting with tall fescue on virus infection of white clover. Plant Dis. 74:51-53.
7. Campbell, C. L., and Madden, L. V. 1990. Introduction to Plant Disease Epidemiology. John Wiley & Sons, New York.
8. Campbell, C. L., Madden, L. V., and Pennypacker, S. P. 1980. Structural characterization of bean root rot epidemics. Phytopathology 70:152-155.
9. Campbell, C. L., Pennypacker, S. P., and Madden L. V. 1980. Progression dynamics of hypocotyl rot of snapbean. Phytopathology 70:487-494.
10. Davies, J. C. 1976. The incidence of rosette disease in groundnut in relation to plant density and its effect on yield. Ann. Appl. Biol. 82:489-501.
11. Fulton, W. C. 1979. On comparing values of Vanderplank's *r*. Phytopathology 69:1162-1164.
12. Galindo, A. J., Olivas, E., and Rodríguez, R. 1978. Experimento para el control del virus de la mancha anular del papayo. (Resumen). Congr. Nac. Fitopatol., 8th. Sociedad Mexicana de Fitopatologia, Chapingo, Mexico.
13. Gabriel, K. R. 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58:453-467.
14. Gilligan, C. G. 1990. Comparison of disease progress curves. New Phytol. 115:223-242.
15. Grupo Interdisciplinario del Papayo. 1992. La virosis del papayo en Veracruz: Etiología y control. Pages 62-71 in: Memoria, Quinta Reunión Científica del Sector Agropecuario y Forestal del Estado de Veracruz. SARH, INIFAP, and UAV, Veracruz, Mexico.
16. Halbert, S. E., and Irwin, M. E. 1981. Effect of soybean canopy closure on landing rates of aphids with implications for restricting spread of soybean mosaic virus. Ann. Appl. Biol. 98:15-19.
17. Hawkins, D. M., and Fatti, L. P. 1984. Exploring multivariate data using the minor principal components. Statistician 33:325-338.
18. Jolliffe, I. T. 1973. Discarding variables in a principal component analysis. II: Real data. Appl. Stat. 22:21-31.
19. Jolliffe, I. T. 1986. Principal Component Analysis. Springer-Verlag KG, Berlin.
20. Jolliffe, I. T. 1989. Rotation of III-defined principal components. Appl. Stat. 38:139-147.
21. Kousalya, G., Ayyavoo, R., Krishnamurthy, C. S., Kandaswamy, T. K., and Bhaskaran, S. 1971. Effect of spacing, roguing and weeding on the incidence of rosette disease of groundnut with observations on the aphid vector *Aphis craccivora*. Madras Agric. J. 58:495-505.
22. Kranz, J. 1974. Comparison of epidemics. Annu. Rev. Phytopathol. 12:355-374.

23. Kranz, J. 1983. Epidemiological parameters of plant resistance. Pages 141-161 in: Durable Resistance in Crops. F. Lamberti, J. M. Waller, and N. A. Van der Graaff, eds. Plenum Press, New York.
24. Kranz, J., and Lörincz, D. 1970. Methoden zum automatischen vergleich epidemischer ablaufe bei pflanzenkrankheiten. Phytopathol. Z. 67:225-233.
25. Madden, L. V., and Pennypacker, S. P. 1979. Principal components analysis of tomato early blight epidemics. J. Phytopathol. 95:364-369.
26. Maelzer, D. A. 1986. Integrated control of insect vectors of plants virus diseases. Pages 483-512 in: Plant Virus Epidemics. G. D. McLean, R. G. Garrett, and W. G. Ruesink, eds. Academic Press, Sydney, Australia.
27. Mora-Aguilera, G., and Campbell, C. L. Multivariate techniques for selection of epidemiological variables. In: Exercises in Plant Disease Epidemiology. L. Francl and D. A. Neher, eds. American Phytopathological Society, St. Paul, MN. In press
28. Mora-Aguilera, G., Kiltie, R., Nieto-Angel, D., and Téliz, D. 1993. Outliers and redundant variables in principal components and factor analysis. Pages 185-194 in: Proc. Annu. Reg. Conf., 1st. SAS Institute, Cary, NC.
29. Mora-Aguilera, G., Nieto-Angel, D., Téliz, D., and Campbell, C. L. 1993. Development of a prediction model for papaya ringspot in Veracruz, México. Plant Dis. 77:1205-1211.
30. Mora-Aguilera, G., Téliz, D., Campbell, C. L., and Avila, C. 1992. Temporal and spatial development of papaya ringspot in Veracruz, México. J. Phytopathol. 136:27-36.
31. Pack, P., Jolliffe, I. T., and Morgan, J. T. 1988. Influential observation in principal component analysis: A case study. J. Appl. Stat. 15:39-52.
32. Pennypacker, S. P., Knoble, H. D., Antle, C. E., and Madden, L. V. 1980. A flexible model for studying plant disease progression. Phytopathology 70:232-235.
33. Reddy, D. V. R., Amin, P. W., McDonald, D., and Ghanekar, A. M. 1983. Epidemiology and control of groundnut bud necrosis and other diseases of legume crops in India caused by tomato spotted wilt virus. Pages 93-102 in: Plant Virus Epidemiology. R. T. Plumb and J. M. Thresh, eds. Blackwell Scientific Publications Ltd., Oxford.
34. Romesburg, H. C. 1990. Cluster Analysis for Researchers. Robert E. Krieger Publishing Co., Inc., Melbourne, FL.
35. SAS Institute, Inc. 1988. SAS User's Guide: Statistics. Release 6.03 ed. SAS Institute, Inc. Cary, NC.
36. Téliz, D., Mora, A. G., Nieto, A. D., Gonsalves, D., García, E., Matheis, L., and Avila. C. 1991. La mancha anular del papayo en México. Rev. Mex. Fitopatol. 9:64-68.
37. Trujillo, P. G., Vegas, A., and Monteverde, E. 1989. Control del virus de la mancha anillada y distorsionante de la lechosa (DRSV) mediante aspersiones con aceite blanco. Rev. Fac. Agron. (Venezuela) 15:141-155.