

A Numerical Taxonomic Study of the Avocado (*Persea americana* Mill.)¹

A. M. Rhodes, Simon E. Malo, Carl W. Campbell², and S. G. Carmer³
University of Illinois, Urbana

ABSTRACT. Principal components and cluster analyses, based on 67 characters, were applied to 38 cultivars, which collectively exemplified the 3 races of avocado and their racial hybrids. Diagrams constructed from principal component analysis clearly showed the phenetic diversity of the 3 races and their racial hybrids. Correlation and distance phenograms from cluster analyses did not show overall phenetic diversity as well as principal component diagrams. The phenograms were most useful, however, in showing phenetic similarities among closely related cultivars, which were obscure in principal component analysis. The 2 methods are, thus, complementary, and both methods are recommended in studying patterns of variation with species such as avocado.

Three general horticultural groups or races of avocados have been recognized. These races are called Mexican, Guatemalan, and West Indian. They have been described and classified by Popenoe (4). Additional descriptions of the races are given by Hodgson (2) and Ruehle (7). In general, the Mexican race is the most resistant to cold injury of the 3 groups. The leaves are usually anise-scented. The fruit has a smooth, thin skin, and is smaller in size, but higher in oil content than the other 2 races. Cultivars showing characteristics intermediate between the Mexican and Guatemalan races dominate the California avocado industry.

The fruit of the Guatemalan race has a thick, brittle rind, usually with a pebbled surface. It is later in maturity than the other 2 races. Cultivars combining characteristics of the Guatemalan and West Indian races are commercially important in Florida.

The West Indian race is the least cold-hardy of the 3 groups. The shoots characteristically have shorter internodes and lighter colored leaves than the other 2 races. The fruit has a smooth, leathery rind, which is generally glossy. The oil content is the lowest of the 3 groups. Since the West Indian race is native to the lowland American tropics, it is well adapted to south Florida, but it is not adapted to California conditions.

Because of hybridization among the races, intermediate forms exist that cannot be classified according to Popenoe's key (4). Anderson (1) classified a hybrid population derived from the Mexican and Guatemalan races using a pictorialized scatter diagram technique based on a few leaf characters. His study helps show a part of the variation in the avocado germplasm complex, but a general classification of the complex would

¹ Received for publication January 8, 1971.

² University of Florida, Homestead.

³ The authors express their appreciation to R. A. Conover

be desirable. For this reason a study using methods of numerical taxonomy was made to quantify the phenetic relationships among various genotypes that are growing in Florida. This paper reports the results of that study.

Materials and Methods

Thirty-two of the 38 cultivars used in the study are located at the Sub-Tropical Experiment Station, Institute of Food and Agricultural Sciences, University of Florida, Homestead. The other 6 cvs., Black Prince, Nadir, Peterson, Ruehle, Trapp, and Winter Mexican, are located in groves around Homestead, Florida.

Table 1. List of code numbers, names and presumed racial types of OTU's used in this study. The predominant racial type is listed first.

Code no.	Name	Racial type ^a
01	Ajax	Guat. W.I.
02	Areu	W.I.
03	Areu Seedling	W.I.
04	Avila	W.I.
05	Black Prince	W.I. Guat.
06	Blake Seedling	Mex.
07	Booth 5	Guat. W.I.
08	Booth 7	Guat. W.I.
09	Booth 8	Guat. W.I.
10	Brogdon	Mex. W.I.
11	Catalina	W.I.
12	Chandler	W.I. Guat.
13	Courtright	Mex. W.I.
14	Ettinger	Mex. Guat.
15	Harris	Guat.
16	Hass	Guat.
17	Hickson	Guat. W.I.
18	Lawhon	W.I.
19	Lounsbury	Guat.
20	Lula	Guat. W.I.
21	Major	Guat. W.I.
22	Mexican Seedling	Mex.
23	Monroe	Guat. W.I.
24	MT-4	Mex. Guat.
25	Nadir	W.I. Guat.
26	Nezahualcoyotl	Mex. Guat.
27	Peterson	W.I.
28	Pollock	W.I.
29	Reinecke 1	Guat. Mex.
30	Reinecke 12	Guat. Mex.
31	Ruehle	W.I.
32	Schaff	W.I. Guat.
33	Tappen	W.I. Guat.
34	Taylor	Guat.
35	Tonnage	Guat.
36	Trapp	W.I.
37	Waldin	W.I.
38	Winter Mexican	Guat. Mex.

^aGuat. = Guatemalan, Mex. = Mexican, W.I. = West Indian

Data were recorded during 1969 and 1970. Table 1 lists the code number, name, and presumed racial type of each cultivar. The presumed racial type is based on the opinion of one of the authors (S.E.M.) before the data were analyzed. Table 2 lists the characters used in the study. Where appropriate the character states for each character and integer code for each character state are given.

The data were analyzed by 2 different methods: (a) principal components and (b) unweighted pair-group clustering. The 2 techniques have been shown by Rohlf (6) to be complementary. He noted that scatter diagrams resulting from principal component analysis show the general overall pattern of phenetic diversity at the expense of detail, while phenograms constructed from cluster analysis show detailed patterns of phenetic similarity in the terminal branches or tips, but the general pattern

of diversity is less clear.

For cluster analysis conventional numerical taxonomic procedures were used (9). These involved standardization of characters, computation of product moment correlation and Sokal's distance similarity matrices of OTU's (Operational Taxonomic Units), cluster analysis of the similarity matrices by the unweighted pair group method with arithmetic averages, drawing of phenograms to show clusters of OTU's (which are cultivars in this study), and measurement of reliability of results by cophenetic and correlation coefficient techniques.

For principal component analysis the R - technique, described in detail by Orloci (3), was used to obtain normalized vectors of the characters and projections of the OTU's along the principal component axes. The projections were calculated from normalized vectors and standardized characters.

Table 2. List of characters and character states used in present study.

01. Relative cold tolerance. (1) low (2) intermediate (3) high.
02. Relative annual yield. (1) low (2) moderate to high.
03. Relative blooming date. (1) early (2) mid-season (3) late.
04. Dioecious tendency when flowering. (1) none (2) slight (3) moderate.
05. Months from fruit-set to maturity.
06. Branching habit. (1) creeping (2) spreading (3) intermediate (4) upright.
07. Average internode length of shoot (mm).
08. Leaf color. (1) light (2) medium (3) dark green.
09. Color of new vegetative growth. (1) green (2) bronze (3) red.
10. Leaf habit. (1) flat (2) slightly folded.
11. Anise-scent of crushed leaf. (1) none (2) weak (3) strong.
12. Median leaf shape. (1) ovate (2) elliptic (3) obovate.
13. Apex of leaf. (1) acuminate (2) subacuminate (3) acute.
14. Median leaf blade length (mm).
15. Ratio of leaf blade length to width.
16. Ratio of leaf blade length to petiole length.
17. Petiole sulcus. (1) incomplete (2) complete.
18. Color of midrib hairs. (1) white (2) brown tint.
19. Relative density of midrib hairs. (1) sparse (2) few (3) many
20. Length of midrib hairs (mm).
21. Panicle color. (1) pale yellow (2) light green (3) medium green.
22. Average number of panicles per shoot.
23. Average number of branches per terminal panicle.
24. Average length of longer panicle branches (mm).
25. Average diameter of panicle branches (mm).
26. Lenticel color of panicles. (1) white or green (2) purple.
27. Flower type. (1) A (2) B.
28. Flower diameter (mm).
29. Tip of outer perianth lobes. (1) acuminate (2) acute (3) obtuse.
30. Width of outer perianth lobes (mm).
31. Fruit stem length (mm).
32. Fruit stem diameter (mm).
33. Length of cupule-enlarged section of fruit stem which originates as part of the calyx tube (mm).
34. Mid-diameter of cupule (mm).
35. Obliqueness of stem to longitudinal axis of fruit (radian).
36. Fruit surface (1) smooth (2) slightly pebbled or rough (3) pebbled.
37. Skin appearance of fruit. (1) glossy (2) intermediate (3) dull.
38. Skin color of fruit. (1) yellowish to light green (2) medium green (3) dark green.
39. Skin of ripe fruit sometimes purple. (1) no (2) yes.
40. Relative tendency of fruit to sunburn. (1) none (2) some.
41. Relative tendency of fruit to crack. (1) none (2) some (3) much.
42. Skin texture of fruit. (1) papery (2) leathery and pliable (3) leathery but brittle (4) woody and brittle.
43. Skin thickness of larger fruit (mm).
44. Lenticel of fruit prominent. (1) no (2) yes.
45. Relative size of fruit lenticels. (1) small (2) intermediate (3) large.
46. Flesh color of fruit. (1) light or greenish-yellow (2) medium yellow.
47. Relative flesh moisture. (1) watery (2) intermediate (3) dry.
48. Flesh texture of fruit. (1) soft (2) firm.
49. Relative flesh sweetness. (1) none to slight (2) moderate.
50. Relative nut-like flavor of fruit (1) none (2) slight (3) moderate to high.
51. Relative oil content of flesh. (1) low (2) intermediate (3) high.
52. Relative fiber content of fruit. (1) none (2) slight (3) moderate.
53. General fruit quality. (1) fair (2) good (3) excellent.
54. Fruit shape. (1) oval to round (2) pear.
55. Fruit flattened near stylar end. (1) no (2) yes.
56. Average fruit weight (oz.).
57. Fruit length (mm).
58. Ratio of fruit length to width.
59. Ratio of fruit length to seed cavity length.
60. Index of flesh thickness: Ratio of fruit diameter to seed cavity diameter.
61. Ratio of fruit volume to seed volume.
62. Seed weight (oz.)
63. Seed shape. (1) globose (2) variable (3) pointed.
64. Cotyledon surface. (1) smooth (2) rough (3) rough and ridged.
65. Seed integuments. (1) adhere (2) loose.
66. Texture of seed integuments. (1) smooth (2) rough.
67. Relative thickness of seed integuments. (1) thin (2) intermediate (3) thick.

The computations were performed using University of Illinois Agronomy Statistical Laboratory programs on an IBM 360/75 computer.

Results and Discussion

Principal components 1 and II. The general overall pattern of phenetic diversity is shown in Fig. 1 as a 2-dimensional model representing the projection of the OTU's (cultivars) onto the first 2 principal component axes. These 2 components account for about 32 percent of the total variance. Component I accounts for 19.5 percent and component II 12.3 percent. Axis I best separates cultivars of the West Indian race from cultivars of the Mexican race while Axis II separates cultivars of the Guatemalan race from the other two.

The 11 cultivars in the lower left-hand corner of Fig. 1 are typical of the West Indian race. This group includes OTU 25 which was designated as a W. I. - Guat. hybrid in Table 1.

The 4 cultivars (OTU's 15, 19, 34 and 35) in the upper center of the diagram are representative of the Guatemalan race. The 2 cultivars (OTU's 6 and 22) in the lower right-hand corner are members of the Mexican race.

The remaining cultivars located between racial groups represent racial hybrids. Their respective positions agree well with the racial types given in Table 1, except for OTU 16 which is considered to be pure Guatemalan.

Cluster analysis. The results of the cluster analysis are shown in Figs. 2 and 3 as phenograms based on correlation and distance coefficients,

respectively. Cophenetic correlations, which measure phenogram distortion in summarizing the results of a similarity matrix, were the same (0.73) for both correlation and distance. Generally, a cophenetic value below 0.80 is thought to indicate a

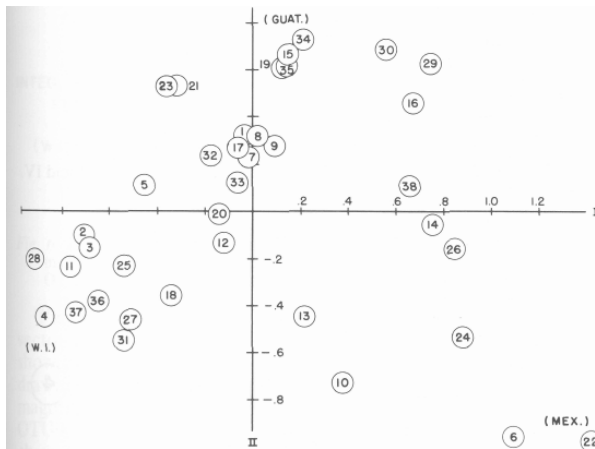


Fig. 1. Scatter diagram. Axes represent principal components I and II. Numbers identify cultivars in Table 1.

principal component results than does the phenogram based on distance coefficients. In the phenogram based on correlation coefficients (Fig. 2), 3 large clusters can be seen which represent the 3 races and most of their respective predominant racial hybrids of Table 1. Five cultivars (OTU's 21, 23, 29, 33, and 38) did not join their predominant racial type but instead joined their lesser racial type. OTU's 15 and 16 appear to be out of place because they do not cluster with other cultivars (OTU's 19, 34 and 35) of the Guatemalan race.

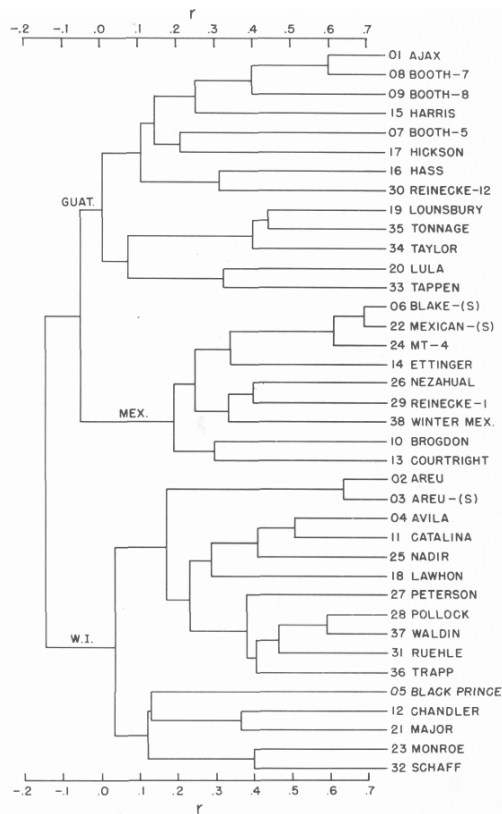


Fig. 2. Correlation phenogram resulting from unweighted pair-group method of clustering using simple averages.

relatively poor representation of a similarity matrix (8). The correlation between the correlation and distance similarity matrices was only 0.71, which implies that different relationships are depicted in the 2 matrices and that the resulting phenograms are also different. The magnitude of these differences is indicated by the low correlation (0.33) between our phenograms. Although neither phenogram shows overall phenetic diversity as well as the *principal component* analysis (Fig. 1), the phenogram based on correlation coefficients agrees much more with the

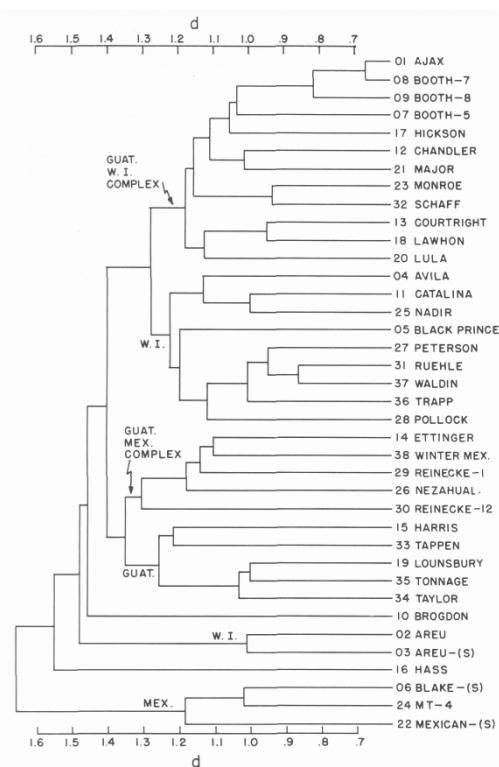


Fig. 3. Distance phenogram resulting from unweighted pair-group method of clustering using simple averages.

In the phenogram based on distance coefficients (Fig. 3), most of the Guatemalan-West Indian hybrids cluster together, and then join with the largest of 2 West Indian clusters.

Most Guatemalan-Mexican hybrids join the Guatemalan race. The cluster representing the Mexican race contains the 2 Mexican cultivars (OTU's 6 and 22) and a Guatemalan-Mexican hybrid (OTU 24). Cultivars that show poor placement are OTU's 13 and 18. OTU's 10 and 16 simply did not cluster with any specific group.

Although the overall phenetic diversity is not well shown by the distance phenogram, certain clusters of OTU's tend to agree with results of the principal component analysis when OTU's of the Guatemalan-West Indian hybrids and the West Indian race are plotted separately with references to Axes III and IV.

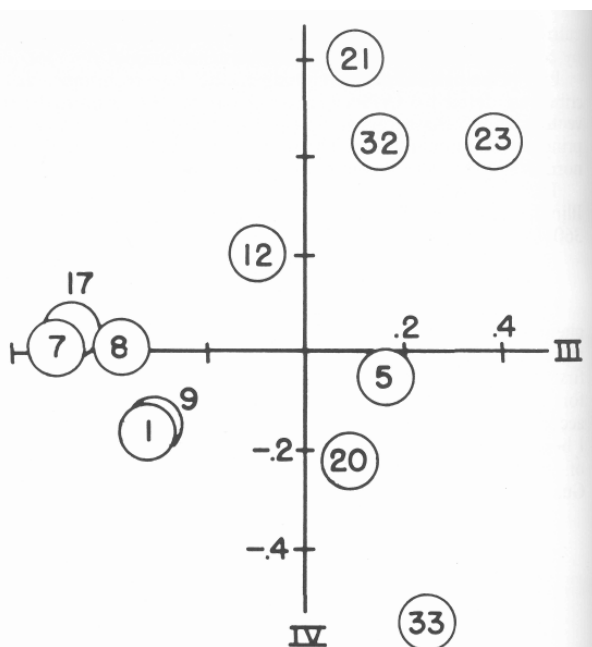


Fig. 4. Scatter diagram. Axes represent principal components III and IV. Numbers identify Guat. - W. I. hybrids in Table 1.

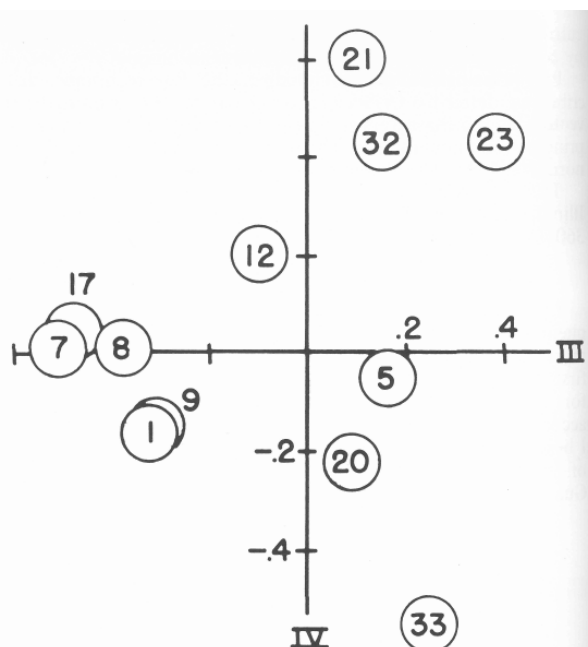


Fig. 4. Scatter diagram. Axes represent principal components III and IV. Numbers identify Guat. - W. I. hybrids in Table 1.

Principal components III and IV. Component III accounts for 7.4 percent of the total variance and component IV 6.5 percent. The scatter diagram in Fig. 4 shows the Guatemalan-West Indian hybrids with relation to Axes III and IV. The Ajax-Booth-Hickson complex (OTU's 1, 7, 8, 9 and 17) remains intact, which agrees well with the results of both phenograms. The other hybrid cultivars (OTU's 5, 12, 20, 21, 23, 32 and 33) are scattered and separated from the Ajax-Booth-Hickson complex. In Fig. 1, OTU's 32 and 33 appeared to be part of the complex, but both phenograms indicate that they are not part of this complex.

The scatter diagram in Fig. 5 represents the OTU's of the West Indian race with respect to Axes III and IV. OTU's 2 and 3 are located in the lower part of the diagram away from the rest of the group, thus agreeing somewhat with the results of both phenograms. In the experience of one of the authors (S.E.M.), OTU 2 is typical of West Indian cultivars grown in tropical western South America while the other cultivars, except OTU 3, are typical of those grown in Florida. The South American type has large conspicuous lenticels on the fruit while the other type has small inconspicuous lenticels. The South

American type also has a characteristic yellowish cast to the leaves which is absent in other avocados.

A remote but conceivably possible connection to our evidence for 2 distinct sub-types in the West Indian race is found in a report by Popenoe (5) concerning prototypes of the 3 races. He gives 2 possible locations for the West Indian prototype, one near Santa Marta, Colombia and the other in the Central American lowlands of Honduras and Costa Rica.

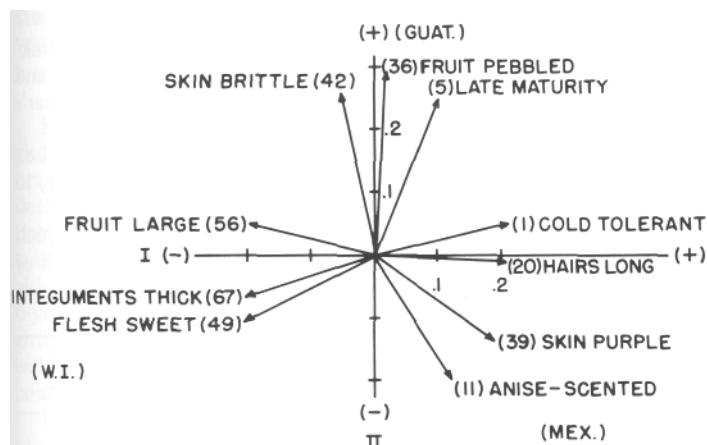


Fig. 6. Example of how vectors obtained from the correlation coefficient matrix of characters can be used to characterize certain groups of OTU's found in Fig. 1. See text for details.

CHARACTERS AND VECTORS. In the form of principal component analysis employed herein, normalized vectors are computed from the correlation coefficient matrix of characters and are then used to obtain the coordinates of each OTU for each axis as in Figs. 1, 4 and 5. The vectors may also be used to help characterize the germ plasm of various OTU clusters in terms of the covariation of characters. For example, with reference to Axes I and II, the end points of vectors 5, 36 and 42

(related to characters 5, 36 and 42, respectively, in Table 2) are found in the upper center of Fig. 6. If Fig. 6 is superimposed on Fig. 1, it can be seen that late maturity, pebbled fruit and brittle skin are characteristic of Guatemalan germplasm. Conversely, in the opposite direction of the arrows and at equal magnitude, these vectors indicate germplasm of early maturity, smooth fruit and papery skin.

Furthermore, a line drawn at right angles to a particular vector indicates that OTU'S along this line are not expected to show any differences for the character in question. Thus, a line drawn at right angles to vector 42 and at a slightly negative magnitude will pass through OTU's 4, 12, 36 and 38. These 4 OTU's have leathery and pliable skin, an intermediate state for character 42.

From Fig. 6, it can also be seen that Mexican germplasm has anise-scented leaves (vector 11) and sometimes purple skin (vector 39), while West Indian germplasm has moderately sweet flesh (vector 49) and thick seed integuments (vector 67). Likewise, Guatemalan-West Indian hybrids have large fruits (vector 56), and Guatemalan-Mexican hybrids are relatively cold tolerant (vector 1) and have long midrib hairs (vector 20).

Since the vectors shown in Fig. 6 agree well with our subjective knowledge of the character patterns of the cultivars used in this study, additional studies are in progress to quantify the germplasm patterns of the important cultivars of avocados grown in Florida. Such studies should be especially useful to breeders who have the continuous problem of deciding what genotypes should be saved in order to conserve as much potential variability of a crop as practical for future use.

Conclusions

The results of the principal component analysis best agreed with our pre-study concepts of the phenetic diversity of the germplasm. Cluster analysis using distance coefficients gave poorer overall results than cluster analysis using correlation coefficients. The results of cluster analyses were most useful to clarify indistinct groups of the principal component analysis. Thus the 2 types of analyses are complementary, and it is recommended that both methods should be used in numerical taxonomic studies to obtain as much information as possible.

Literature Cited

1. Anderson, E. 1969. Plants, man, and life. Univ. of Calif. Press. 251 p. (Sec. print.)
2. Hodgson, R. W. 1950. The avocado - a gift from the middle Americas. *Econ. Bot.* 4:253-293.
3. Orloci, L. 1967. Data centering: a review and evaluation with reference to component analysis. *Syst. Zool.* 16:208-212.
4. Popenoe, W. 1920. Manual of tropical and subtropical fruits. MacMillan, New York. 474 p.5.
5. _____. 1936. Origin of the cultivated races of avocado. *Cal. Avo. Assoc. Yrbk.* 1935:184-194.
6. Rohlf, F. J. 1968. Stereograms in numerical taxonomy. *Syst. Zool.* 17:246-255.
7. Ruehle, G. D. 1963. The Florida avocado industry. *Fla. Agr. Expt. Sta. Bui* 602. 102 p.
8. Schnell, G. D. 1970. A phenetic study of the suborder Lari (Aves) II. Phonograms discussion, and conclusions. *Syst. Zool.* 19:264-302.
9. Sokal, R. R., and P. H. A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco and London. 359 p.