A COMPUTER-AIDED MORPHOLOGICAL CLASSIFICATION OF *MANIHOT ESCULENTA CRANTZ*

— *by* —

David J. Rogers
Colorado State University

The following classification is based on the collections made by the author in the more important growing areas of *Manihot esculenta* in the Western Hemisphere. It is a classification of the cultivars, and does not indicate the relationship of the species as a whole to other members of the genus *Manihot*. The classification does not employ herbarium specimens made by other collectors because the typical collections are not adequate for the purposes of differentiation of the cultivars of this species.

The classification is based on 231 population samples, carefully selected from over 500 samples made by the author. These 231 samples represent the most complete of the collections, and contain amongst them the major morphological variations to be found over the geographical range of the species.

Samples were made in Jamaica, both from private holdings and from the extensive collections of the Department of Agriculture's experiment station at Bodles, near Old Harbor. In Costa Rica, collections are from private holdings and from the collection established by Dr. Jorge Leon at the Inter-American Institute of Agricultural Sciences, Turrialba. In order to provide a check on the stability of characters employed, these two areas were sampled at least twice in separate years. Other collections were made at the Federal Agronomic Experiment Station of Brazil at Belem, Para, the collection of which was assembled from the Amazon Basin by Dr. Milton de Albuquerque; from the state experiment station of Tambe and Ararapina in the state of Pernambuco, both stations under the over-all direction of Dr. Jaime Coelho; from the Federal Agronomic Experiment Station at Sete Lagoas, state of Minas Geraes; at the state experiment station of Sao Paulo, in Campinas under the direction of Dr. Edgard Normanha; at several private locations in Bolivia and Peru, and along the Amazon River near Manaus. Collections from nearly every major growing area in Central and South America have been examined, and if possible, specimens included in this study. It is felt that most of the types of variation existing in this complex has been sampled.

The classification is based on morphological criteria alone. While it is regrettable that biochemical and cytological information are not included, this first survey of the species in its various areas of distribution could not, because of the practical problems, include these types of information. Clearly, it is important to include this type of data as our knowledge of *M. esculenta* expands. It is hoped that this morphological classification will be accepted for its intended purpose—a method by which various morphological types within the complex can be identified, and by which workers from various places can relate the plants of their collections to those of other areas.

It will be noted that no collections from other important regions of the Eastern Hemisphere tropics have been included. Since this plant is of Western

Hemisphere origins, it was felt more critical to study the plant in its native areas of distribution first, and to perhaps later expand the work to those areas where the crop is an evident introduction (cf. Rogers, 1963, 1965).

CHARACTERS EMPLOYED IN THIS STUDY

Tables 1 and 2 illustrate the information upon which this classification is based. From an examination of these characters, it becomes evident why the typical herbarium specimen of the botanist is not satisfactory for this type of classification. The samples I made consistently included materials which would make it possible to make comparisons of those parts of the plant most important to the cultivator, the roots, plus other obvious parts of the stem and leaf. Note that no information is gathered for the flowering parts of the plant. This is done because many cultivars do not normally come to flower by the time the plant is harvested, and the information about the flowers, while critical to a classification within the genus *Manihot,* is not critical to this classification among the cultivars alone. From my examination of the flowering material of many cultivars, I do not feel that the morphological characters of the flower vary significantly. It has been found by Moh (1966) in Costa Rica, that there is considerable variability in pollen viability, and this no doubt is an important feature to be kept in mind when breeding considerations are made.

COMPUTER PROGRAM

The methods employed to provide an objective classification of *Manihot esculenta* are described in a series of papers (Wirth, et al. 1966; Estabrook and Rogers, 1966; etc.) and the program itself is available to anyone interested in using the methods. A more detailed description of the computer methods will be presented in the demonstration on Tuesday afternoon. At this point, only a cursory description is given of the methodology.

For each sample, a recording is made of each of the characters given in Tables I and II. All samples are treated in the same way. From these character records, a measure of similarity between every sample is computed by taking the fraction of the characters for which both samples possessed the same attribute. Thus, relations between all objects are produced, in a measure of over-all similarity. The method used has been designated by some workers as the simple matching coefficient.

The similarity table generated in this step is used as input to a procedure for clustering the objects. The model used as the mathematical base of this particular clustering method is derived from graph theory. A series of "subgraphs" of the samples under study can be made, each subgraph connecting samples with a certain value of similarity, and partitions between the different connected subgraphs. By lowering the similarity values, a greater number of samples are joined by subgraphs and we eventually place all objects into a single subgraph but along the way, pick out groups of organisms which satisfy certain requirements. This procedure is best illustrated by the following, over simplified diagrams (Figs. 1-5).

DESIGNATION OF THE GROUPS

We have chosen not to erect any formal nomenclature for the clusters of interest to us. Since the purpose has been to show the "constellations" of morphological expressions, and to give some means to differentiate amongst an essentially reticulate relationship, we have not felt that any formalized names could be adequately applied. Therefore, we designate our clusters as "groups," and number them with Roman numbers for convenience in discussion. This convention is flexible.

One may discover many interesting inter-relations by using the computer program, and many more than I have time to indicate in this discussion. There are several ways in which the results of the graph theory model give useful information. In addition to the delimiting of clusters or groups, the method indicates how widely divergent one group is from another—how much morphological isolation one group has from another—and also which samples of one cluster act as linking agents to another cluster. In this latter case, the most likely hybrids between groups are indicated—that is, which samples in one cluster have the most properties in common with samples in another cluster. This gives the worker some idea of the nature of the variation types within any cluster, a useful indication if he wishes to develop some breeding program.

But perhaps the most useful attribute of the computer program is the fact that many different ideas about relationship may be tested. If one wishes to discover the influence of a certain new piece of information on an old classification, the computer is sufficiently rapid that the test can be made in a short enough time to allow several tests of ideas. We must not assume under any circumstances that the computer "tells you" what you must do, rather it is used as a tool to aid our idea-testing.

THE GROUP OF M. ESCULENTA

The graph-clustering procedure for 231 samples of *M. esculenta* was completed in 101 levels, and all samples were placed together in a single cluster at the similarity value of 0.62500. In examination of the intermediate levels, we discover that the population was essentially divided into clusters at the 59th level of clustering, similarity value of 0.81875 with two major sets of characters. The clusters we can form at this level are indicated in Fig. 8, showing the primary division of *M. esculenta* into subequal groups which we designate as "rough" and "smooth" rooted. Note character number thirteen. It can also be demonstrated that these groups are differentiated as well on the external root color and external stem color—those with "smooth" surfaced roots are light tan or pink, and have silver-colored stems; those with "rough" roots are brown to dark brown or brown-yellow, and have brown, yellow, or infrequently silver-brown stems. We feel that these characters are sufficiently constant to provide a major division within the species complex. No geographic differentiation accompanies this division, and to the author's knowledge the plants with these characters occur in about the same proportions in all areas sampled. This being the case, no subspecific definition in the sense of wild-plant taxonomy, is possible, for this taxon is usually designated as allopatric. In the past, I have proposed that the term "convariants", used in regularly recommended nomenclature for cultivated plants be used in designation of these two, but do not feel that this is justified.

The smaller number of plants (or cultivars) belongs to the group designated smooth. But the same type of variation in HCN content occurs in plants of the smooth-rooted group as occurs in the larger, rough-rooted, group of cultivars. It is also true that characters other than the HCN content of the root vary equally in both the smooth and rough-rooted groups. For these reasons again, I do not feel justified in making a formal taxonomic category. We are more interested in the morphological variations, and ways to identify them, than a classification for the sake of some finalized botanical nomenclature.

FURTHER GENERAL SUBDIVISIONS

It will be noted in the breakdown given in Fig. 6 that both the rough and smooth-rooted varieties are subdivided into two groups, each with the mnemonic, descriptive terms, "linear-lobed" or "obovate-lobed" groups. These groups remain separate until about a similarity value of 0.75, which would indicate that about 3 to 4 characters out of 15 differ between them. When there is this much difference, we have more confidence that the groups have some validity, and are not based on a single character difference, or, in other words, are not based on purely subjective judgment, as would be the case without the computer analysis. These groups have more internal connections among themselves than with the groups outside,a fact that is best demonstrated on the computer print-out.

Below this level of subdivision, and within each of the above named breakdowns (linear-lobed or obovate-lobed) we reach the lowest categories that we wish to circumscribe for the purposes of classification. Each of these groups, designated by Roman numerals and lower case letters, constitutes a number of cultivars which for convenience' sake have also, as for the larger groups, been given a morphological, informal, epithet. These divisions are still capable of being recognized as groups, but their relationships among themselves are more complex than the larger divisions due to their reticulate nature. The lowest level of similarity that can be demonstrated for these is about 0.80. In other words, any organisms showing less than this amount of similarity to the other, included, organisms, should be considered as a separate entity. This might be the case if we were dealing with separate species, but since these are cultivars within a single complex species, we have not felt bound to continue separations, although a relatively small number of cultivars are much less similar than the 0.80 figure. It is not the purpose of our classification to split down to the individual cultivar. Let it be said, however, that the computer method allows each organism to be related to all other cultivars with accuracy, as will be shown in the following discussions.

By far the largest number of cultivars fall into the category of rough roots, with brown, yellow or reddish brown stems. Within this group, the preponderance are obovate-lobed (Fig. 6). Our subgraphing technique indicates that there are eleven subdivisions within the obovate-lobed, rough-rooted group. It is a matter of choice whether we will accept still other categories at this level, but we feel that if we follow a set of rules to delimit them, that we must not designate more than this. My meaning on this point will be explained in some detail below.

THE SMOOTH -ROOTED, SILVER-STEMMED GROUPS

Since the time limitations on this paper do not permit a full exposition of the classification of *M. esculenta,* we will concentrate our attention on the smaller of the large subdivisions, in order to demonstrate the rational for the classification (Fig. 7).

As noted above, the silver-stemmed, smooth-rooted group can be divided the same way the rough-rooted group is divided, namely: linear- or obovate-lobed. There are no sharp or deep "moats" within the cluster of linear-lobed cultivars. They are not particularly closely related, inter-alia, but their relationships clearly differentiate them from the obovate group. The differences between the clusters of linear-lobed and the obovate-lobed are noted in the accompanying Fig. 7. The obovate-lobed, smooth-rooted group obviously has the larger number of subdivision, as was the case in the rough-rooted group.

We will take the groups in the order given. First, the linear-lobed smooth group may be characterized as follows: the roots are externally light brown, tan, or light tan, and a few are light pink; the root cortex is white to cream and rarely yellow, but there are no pink pigments in the cortex. The leaf-scars on the stem are usually slightly to moderately raised, seldom large. The lobes of the leaves are moderate to long, and most frequently are sinuate-margined. The petioles are usually green, but may be red-green, or rarely red. The linear-lobed, smooth-rooted group tends toward little or no branching, with few cultivars fully branched.

Figure 8 illustrates the inter-relationships among the 17 cultivars included in the linear-lobed group. The boxes enclose cultivars related above 0.85 similarity, and nearest-neighbors are immediately related within smaller boxes. The "core" of linears include those listed from 114 to 225 (numbers of cultivars are my collection numbers). Two closely-related, but separate, "strings" are attached to specimens numbered 303 and 300. Outlying individuals are connected to the center core. All connections indicate (by a bracketed number over the lines) the similarity measure which join the specimens. Arrows indicate the specimens within this group which are closest to members outside the group. Interestingly, both specimens (114 and 131) join to the same group in the obovate-lobed cluster, and at the same similarity value. The secondary connections betwen objects have not been indicated on this diagram, but a very useful summary of the 10 closest objects is given by a part of the computer print-out called the "nodal distance array." We have used the diagramming technique of connecting only closest neighbors because of the nature of the biological relationships with the species *M. esculenta.* Were we to indicate all the relationships that each specimen has with all others of the complex, we would lose sight of the major objective of a classification, namely, the placing of an individual in a grouping which the worker can visualize. Other groups of organisms tested with our method indicate that other application techniques with the computer print-out are valuable. This is, the nature of the variation to be found in different genera and families, obviously, is not identical with that found in this species, this genus, or perhaps in this family. For example, computer studies of a section of the genus *Cassia* done in collaboration with Dr. Howard Irwin of the New York Botanical Garden indicate that the most important part of the print-out was that dealing with the internal connections within the clusters. In *Cassia,* the connectedness within a cluster was a measure of the "goodness" of the species. Certainly the amount of connectedness within the species *M. esculenta* is no measure of, nor indicator of, the goodness of this or that cluster.

Looking back at Fig. 7, we see the structure for the larger group of smooth-rooted cultivars, here designated II, obovate-leaves. This designation actually refers to the lobes of the simple leaf, and all characters—length, width, margin, etc.,—are of the central lobe. Group II has five subdivisions, whereas the equivalent group (IV) in the rough-rooted category was divided into eleven. Obviously, there is no real equivalence in these subdivisions of obovate-lobed groups between rough-and smooth-rooted cultivars.

The same rules apply—as with the cluster of linear-lobed cultivars—all connections between pairs shown is above 0.80. As with other groups, this does not indicate that all specimens included in a cluster share the same or higher similarity values. Rather, as indicated earlier, most of the relationships found in these clusters are clinal, with the ends of the clines having little in common. But the clusters are not particularly well described as clinal, since the implications of a cline are frequently some geographic distribution, and we cannot assign any of our larger groupings to be specified geography. Part of the explanation is the fact that many of my specimens were made from experiment station collections, from diverse areas, and did not necessarily represent any one local area, at the point where I collected it.

Figures 9 through 12 represent the four groupings of obovate-lobed, smooth-rooted plants. We have given each of these a morphological name for the most obvious common trait. IIa, unbranched, (Fig. 9) has a specimen (No. 120) from which most of the variations are derived. In this case, 120 acts as an articulator, holding together a group of smaller clusters which would otherwise not connect.

Figure 10 illustrates that the group IIb can be subdivided. In this case, we have indicated the points of the other clusters to which this cluster joins. Also indicated in IIb_2 are the points where cluster I, the linear-lobed group joins, and the similarity values at which they join.

Group IIc (Fig. 11) contains the largest, and perhaps the most common set of morphological conditions to be found amongst the smooth-rooted group. If one were working in a "typological" manner, one would select group IIc as the type for the smooth-rooted clusters. However, we are aware of the hazards of such a designation, and only point out that more specimens fall in this category than in any of the others.

The last of the smooth-rooted group, IId, (Fig. 12) contains two specimens 229 and 301, whose character-combinations indicate one of the major difficulties with assigment of a morphological name to designate the group. These two specimens are clearly 7-lobed, but according to all other characters, should be associated with group IId. While the assignment is good, according to the over-all similarity measure, the naming of the group is poor. But to give another morphological epithet to the group would require a polynomial designation. This is as unsatisfactory as the name we chose, and much more cumbersome.

This group also illustrates the point made earlier, that the similarity measure can drop below 0.80 and still include specimens adequately assignable to the group. The two specimens are 329, paired with 365 at 0.69 similarity, and 316, paired with 119 at similarity of 0.67.

SUMMARY

We have summarized the variation in *Manihot esculenta,* and discovered a satisfactory method for dividing the cultivars into related constellations. The relationships amongst the cultivars is reticulate, but by employing the graph theory model, we have been able to discover the major categories, and found that the recognizable groups are "strings" of clinal relationships. There will probably be some variations to the groups we have established, particularly when new biochemical information is found, but without the addition of further information, we are certain that an investigator can reliably relate his materials to the categories we have provided.

The classification is based on the collections made by the author, and no other herbarium material has been employed to structure the classification. The specimens used in this classification will be housed in the herbarium of the United States National Arboretum, central locality from which other interested workers may borrow these materials.

Table 1

Manihot esculenta characters — January, 1967

STEM CHARACTERS, Nos. 1, 2, 14, 15

K_1 Color of Stem

1. Silver N = 6
2. Silver-brown K = 1
3. 0
4. 0
5. Brown
6. Yellow

K_2 Branching of Plant

1. One branch at top or no branches
2. One or two branches but not 1 branch if at top
3. More than 2 branches

LEAF CHARACTERS, Nos. 3-10

K_3 Leaf Lobe Shape

1. Obovate
2. Linear

K_4 Number of Lobes of Leaf

1. 3 or 4 lobes
2. 3, 4 or 5 lobes
3. 5 or 6 lobes
4. 4, 5 or 6 lobes
5. 7 or 8 lobes
6. 9 or 10 lobes

K_5 Length of Median Lobe

1. Less than 14 cm
2. 14-17 cm
3. Greater than 17 cm

K_6 Width of Median Lobe (widest point)

1. Narrow (1.5 cm—2.4 cm)
2. Medium (2.6 cm—4.8 cm)
3. Broad (5.0————)

K_7 Sinuosity of Lobes of Linear Leaves

1. Pandurate
2. Some sinuosity
3. Simple (not sinuous)
4. Logical (obovate)

K_8 Sinuosity of Lobes of Obovate Leaves

1. Pandurate
2. Some sinuosity
3. Simple (not sinuous)
4. Logical (linear)

K_9 Color of Young Foliage

1. Reddish-blue
2. Bluish-green
3. Green

Table II.

K_{10} Petiole Color

1. Red
2. Green-red
3. Red-green
4. Green

	R	GR	RG	G
R	1			
GR	75	1		
RG	25	75	1	
G	0	25	75	1

ROOT CHARACTERS, Nos. 11, 12, 13

K_{11} External Color of Root

1. Light brown-yellow
2. Brown, dark brown, reddish brown
3. Light brown, tan, light tan
4. Pinkish brown, pinkish tan
5. Pinkish white, light pink, pink

K_{12} Cortex (root) color

1. White to cream
2. White to cream with pink
3. Cream-yellow to yellow
4. Cream-yellow to yellow with pink

	1	2	3	4
1	1			
2	75	1		
3	0	0	1	
4	0	0	75	1

K_{13} Surface of Root

1. Smooth
2. Rough

STEM CHARACTERS, Nos. 1, 2, 14, 15

K_{14} Nature of Scars on Stem

1. Smooth
2. Slightly raised
3. Moderately raised
4. Very large

N = 4
K = 1

K_{15} Storey Length

1. 4—8 cm
2. 9—20 cm
3. 21—28 cm

Fig. 1.

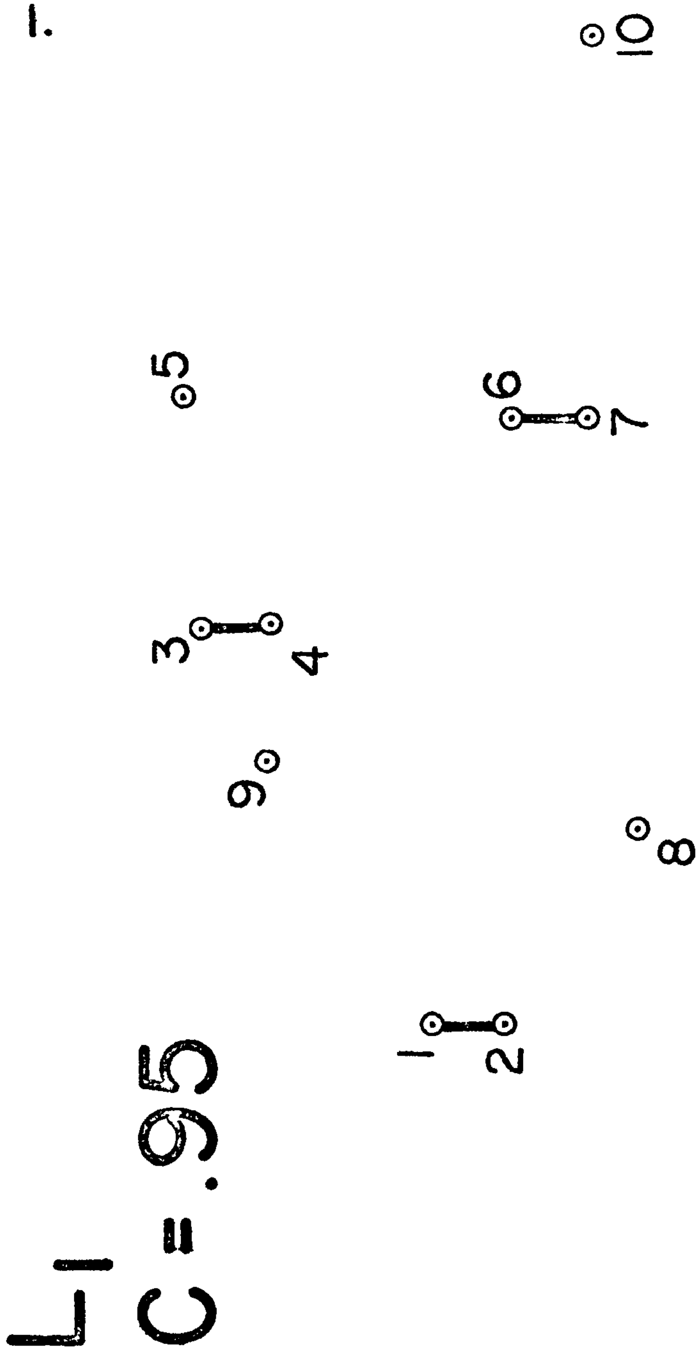


Fig. 2

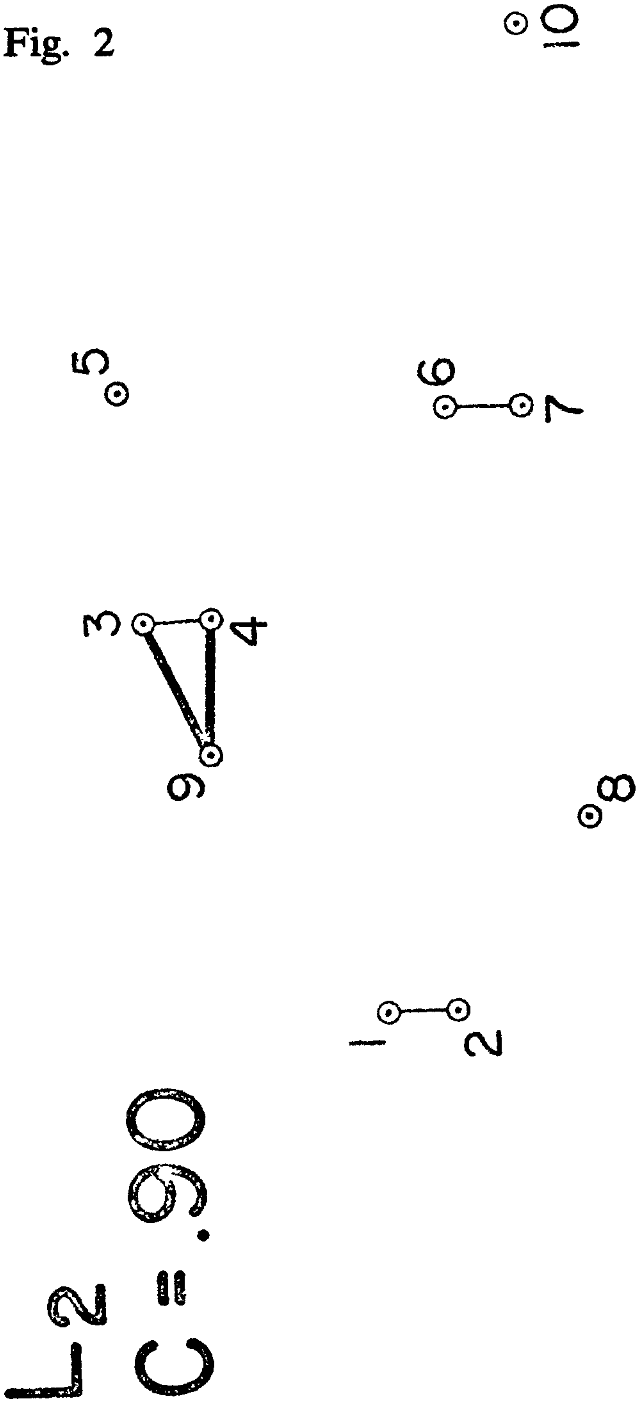


Fig. 3

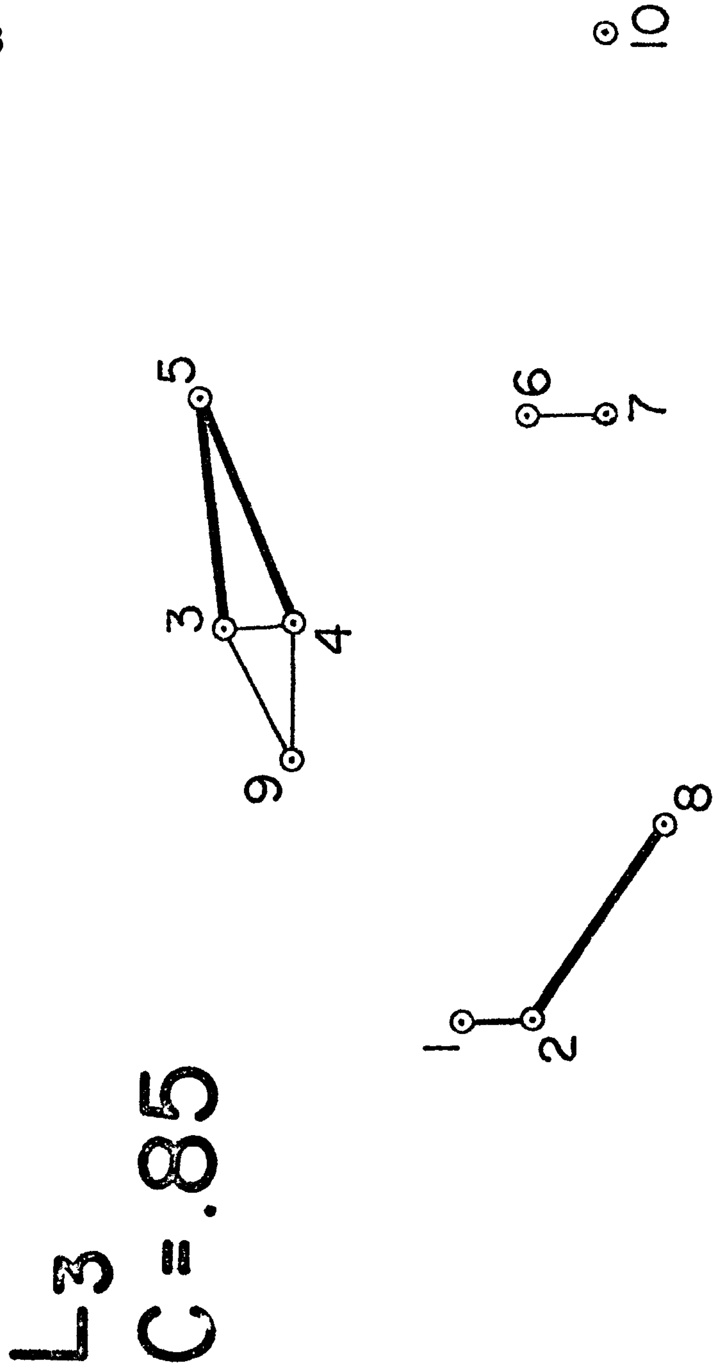


Fig. 4

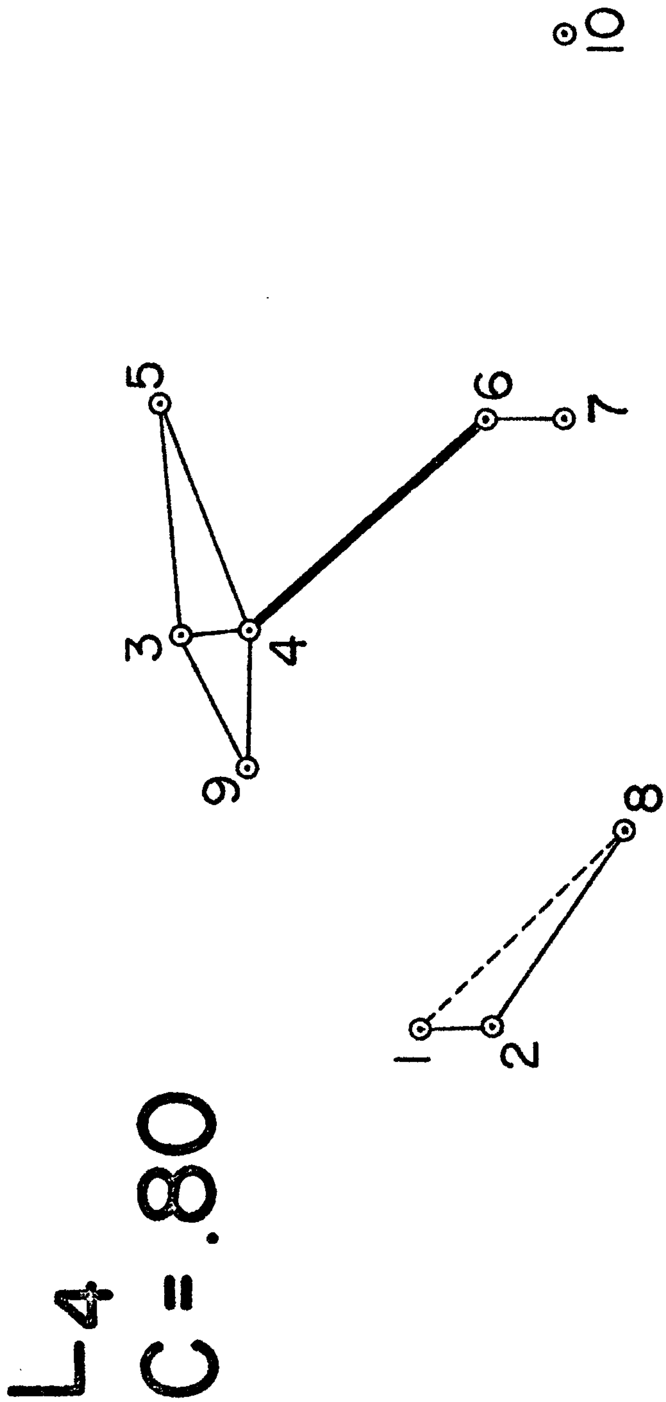


Fig. 5

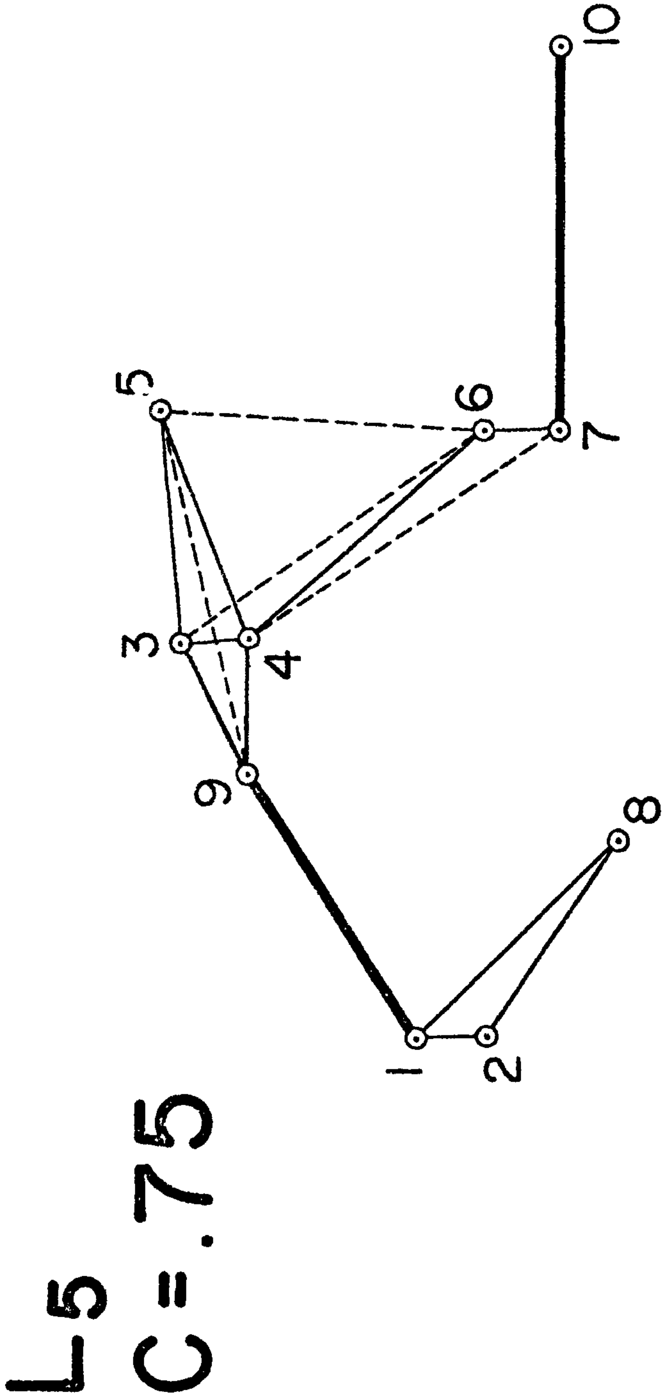
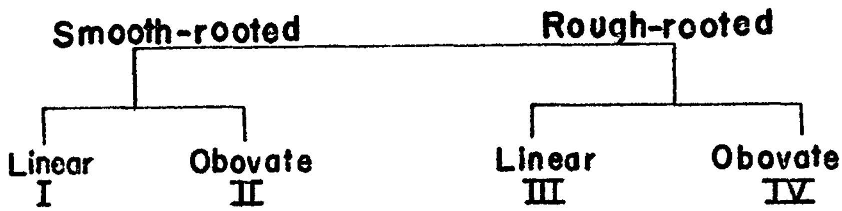


Fig. 6 Manihot esculenta, Crantz



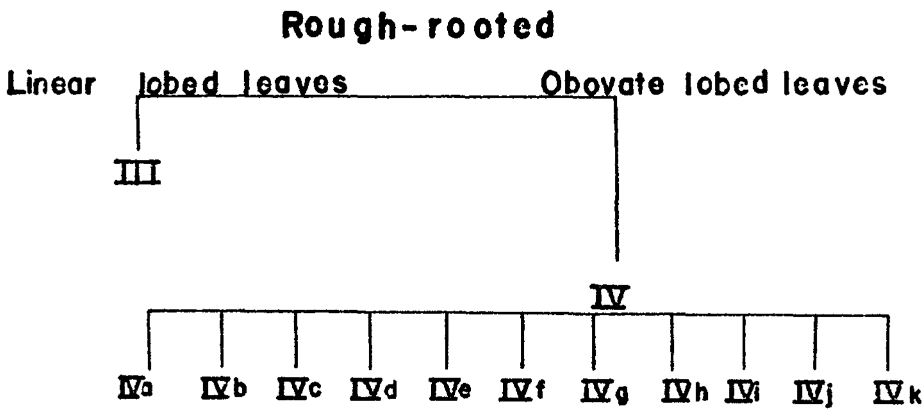


Fig. 7

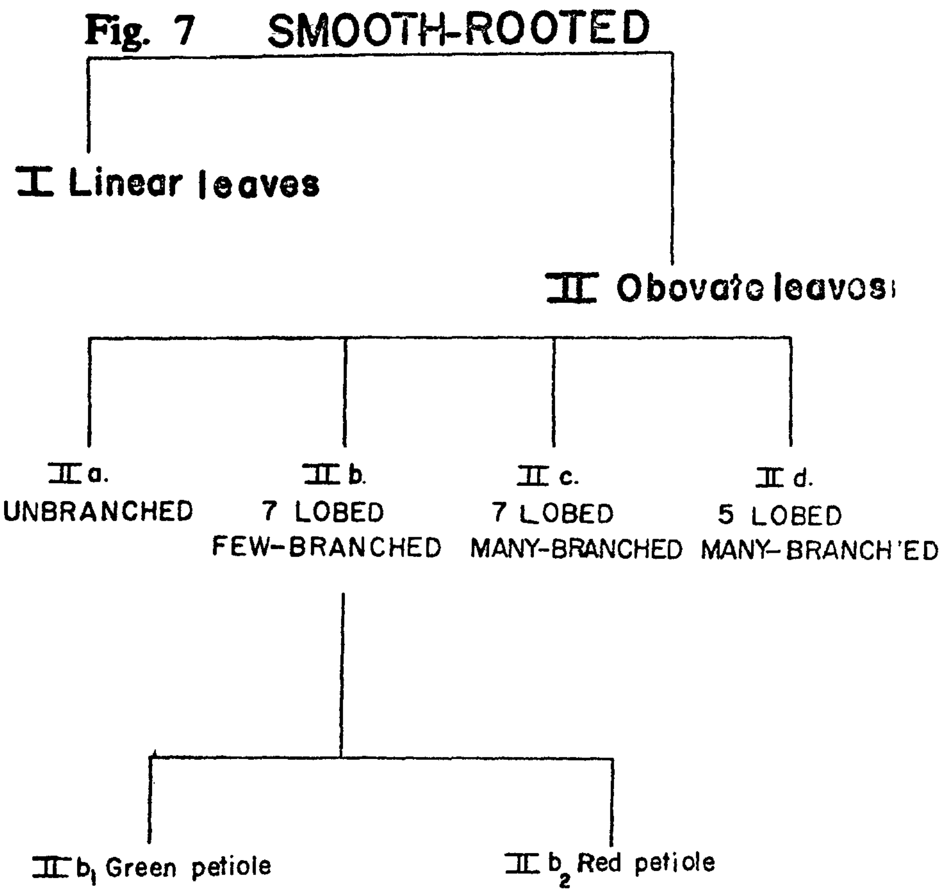
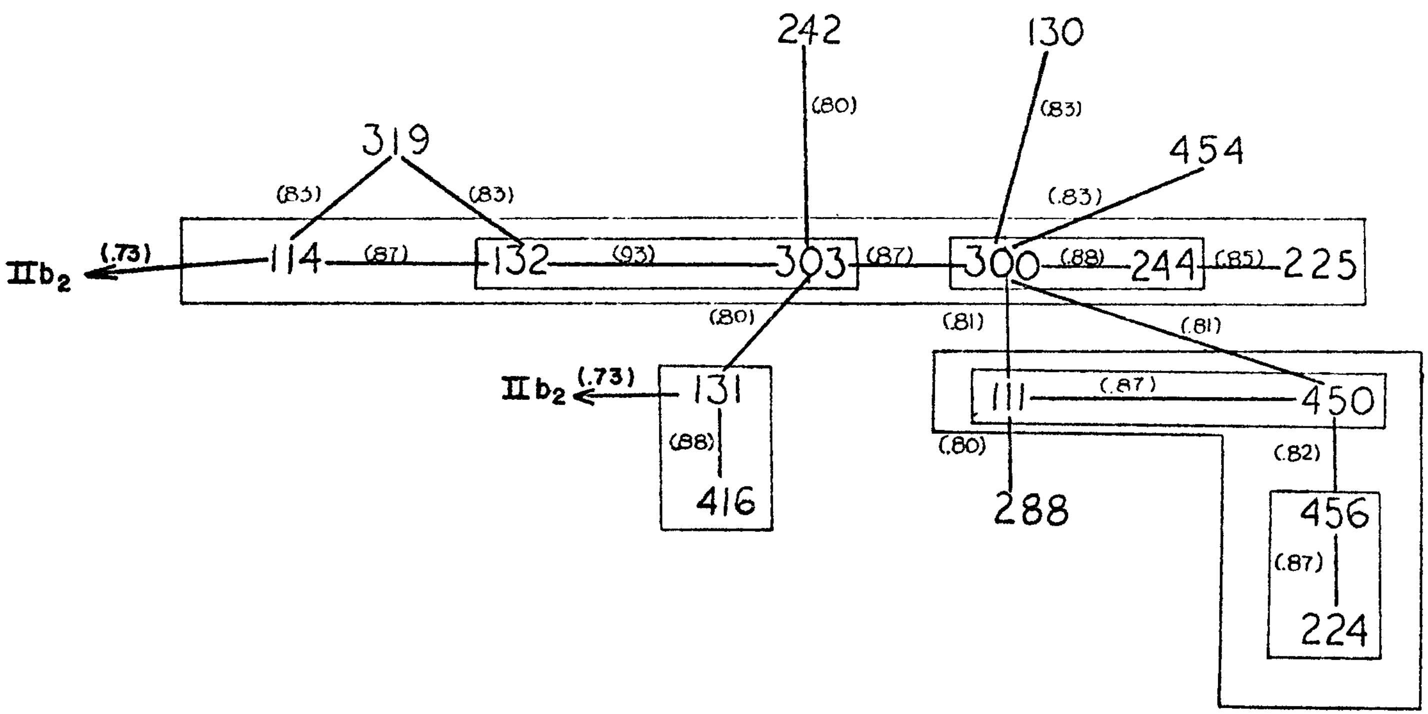

SMOOTH-ROOTED
I Linear leaves
II Obovate leaves
II a.
UNBRANCHED
II b.
7 LOBED
FEW-BRANCHED
II c.
7 LOBED
MANY-BRANCHED
II d.
5 LOBED
MANY-BRANCH'ED
II b1 Green petiole
II b2 Red petiole

Fig. 8 **Manihot esculenta**

Group I, Linear-lobed foliage



Group IIa, Unbranched

Fig. 9

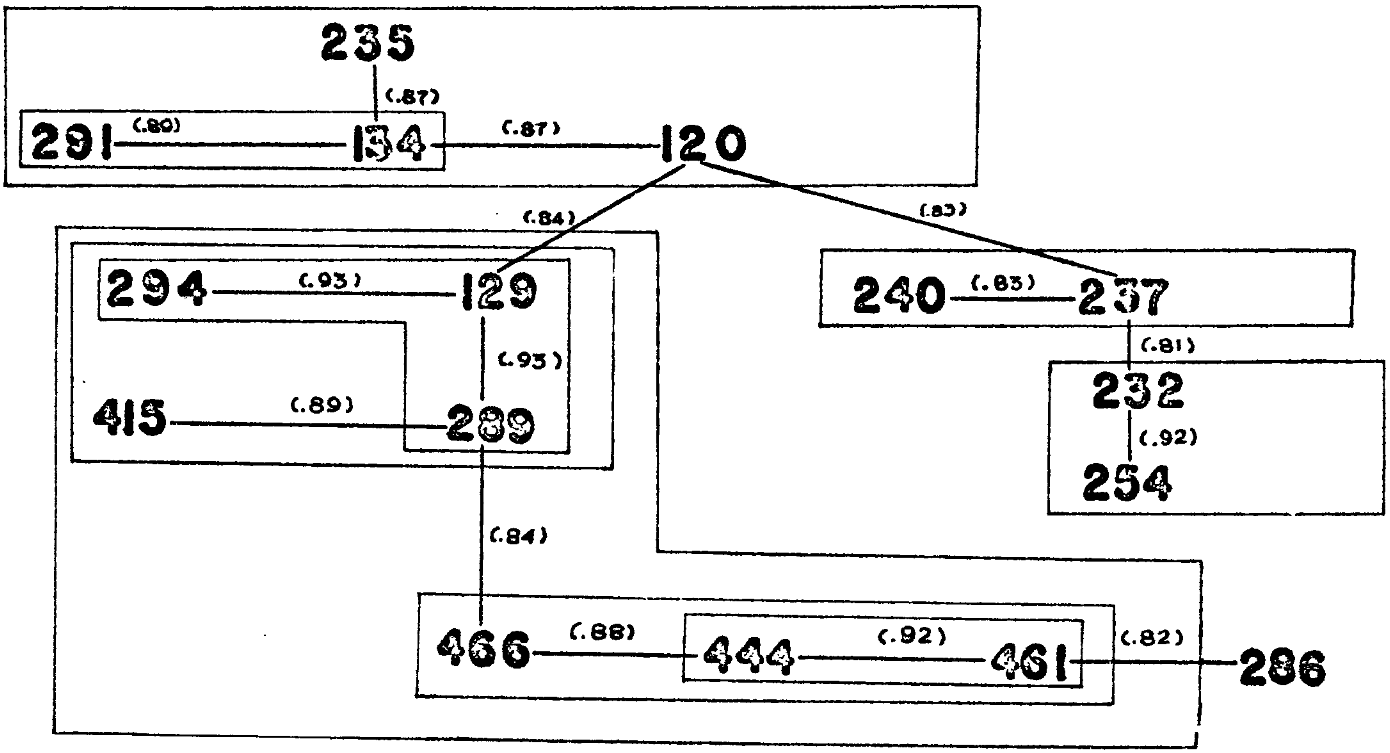


Fig. 10 **Group II b, 7 Lobed, few-branched**

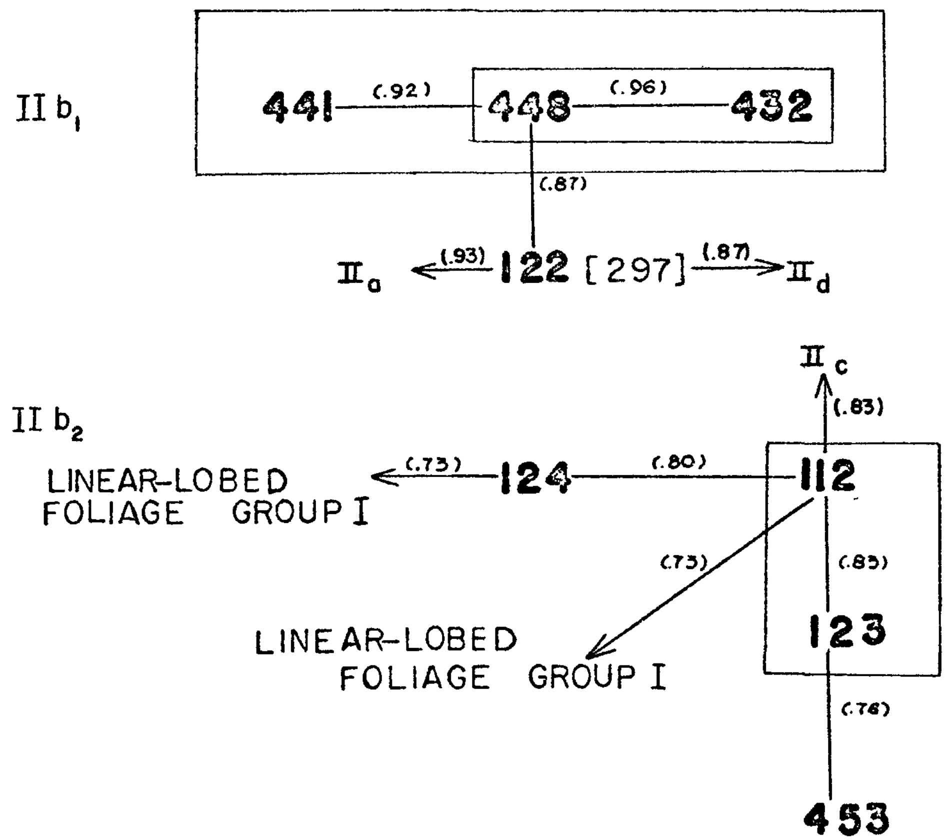
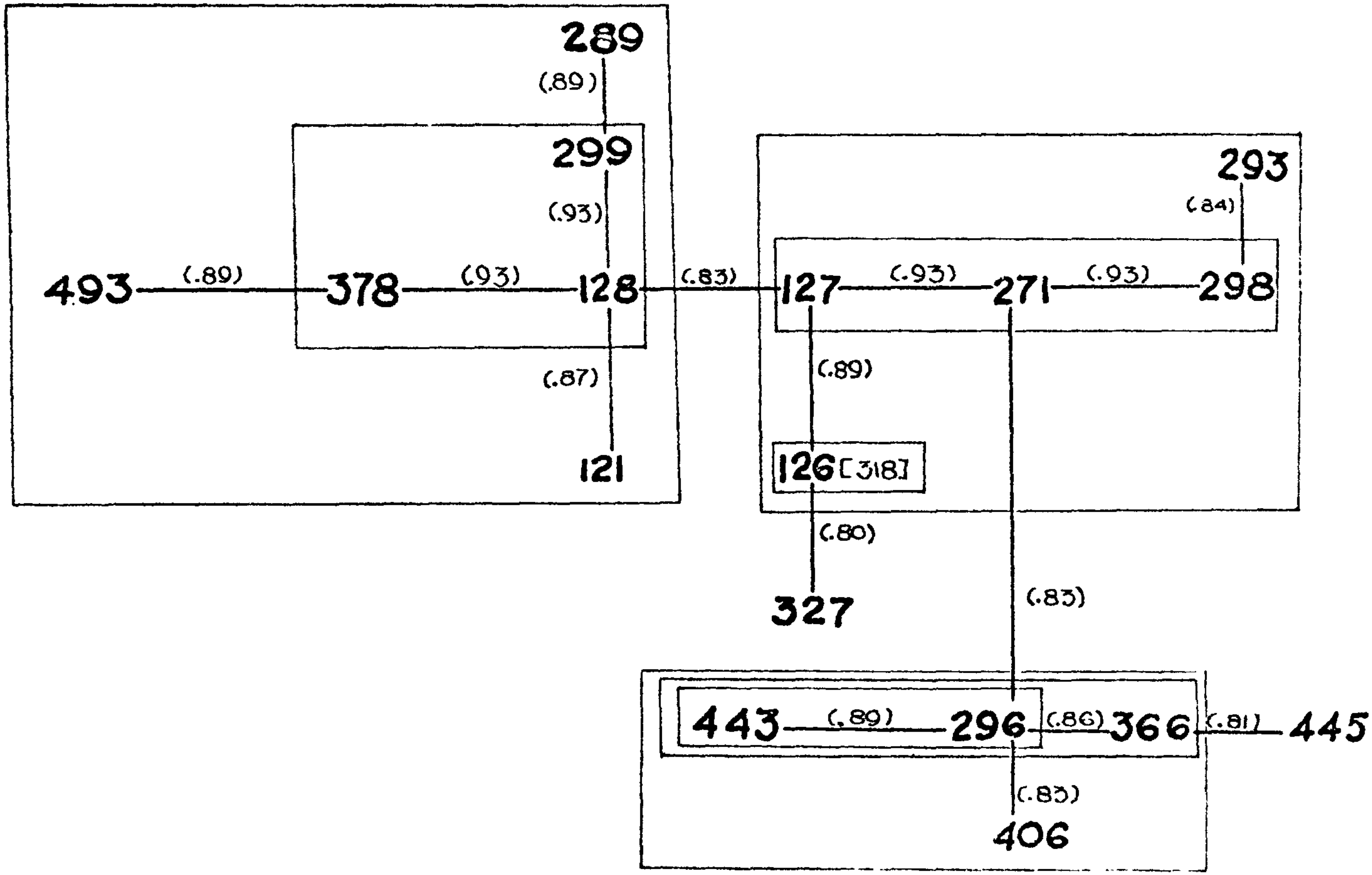


Fig. 11 **Group IIc, 7 Lobed obovate, many-branched**



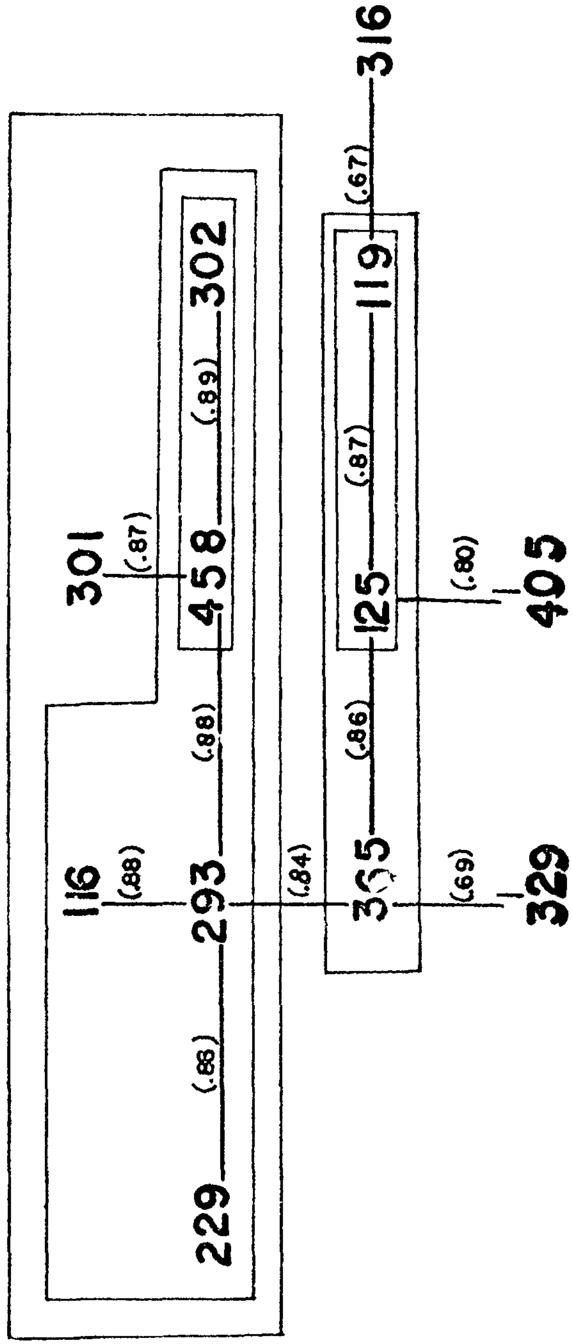


Fig. 12 Group II d, 5 Lobed, many-branched

REFERENCES

1. Estabrook, G.F. 1966. A mathematical model in graph theory for biological classification, **J. Theoret. Biol.** 12 : 297-310.

2. Estabrook, G.F. and D.J. Rogers, 1966. A general method of taxonomic description for a computed similarity measure. **Bio Science** 16 : 789-793.

3. Rogers, D.J. 1963. Studies of **Manihot esculenta** Crantz and related species. **Bul. Torr. Bot. Club** 90 : 43-54.

4. ——— 1965. Some botanical and ethnological considerations of **Manihot esculenta. Econ. Bot.** 19: 369-377.

5. Moh, C.C. 1966. Preliminray observation on meiotic chromosome pairing in **Manihot esculenta.** The Application of Nuclear Energy to Agriculture. Inter-American Institute of Agricultural Science of the OAS Training and Research Center Turrialba, **Costa Rica. Annual Report,** July 1, 1966.

6. Wirth, M., G. F., Estabrook, and D. J. Rogers, 1966. A graph theory model for systematic biology with an example for the Oncidiinae (Orchidaceae). **Syst. Zool.** 15: 59-69.