

# COMPUTER TAXONOMY IN THE FUNGI IMPERFECTI<sup>1</sup>

W. BRYCE KENDRICK

*Plant Research Institute, Canada Department of Agriculture, Ottawa*

AND

JEAN R. PROCTOR

*Statistical Research Service, Canada Department of Agriculture, Ottawa*

Received July 3, 1963

## Abstract

In Part I of this paper, the principles of numerical taxonomy are discussed and the practical processes outlined, with particular reference to their applicability to the Fungi Imperfecti.

In Part II, one of the authors (J.R.P.) discusses several scoring methods, including one introduced by the authors in which a 'primary' character is weighted according to the number of 'secondary' characters used to describe it. The merits of the various methods are considered.

In Part III, one of the authors (W.B.K.) discusses the results obtained when 30 strains of Hyphomycetes belonging to the *Leptographium* complex were compared using a computer, the taxonomic data being scored in six ways. The classification agreeing most closely with the established one was obtained when 'primary' characters were weighted and 'negative matches' were included. Of the methods applied, this is considered in Part II to be the most logically satisfying.

## Part I. Introduction

In all taxonomy, organisms are placed into groups according to their similarities and differences. The processes of comparison and sorting are often carried out intuitively and subjectively. Sometimes an integrated pattern of similarity involving many characters may be discerned almost instantaneously; in other cases where numerous small and independently varying characters exist, the systematist may be unable to arrive at satisfactory conclusions. In some groups this has meant that no acceptable classification has yet been devised, while in other fields examples are known of repeated transfer and retransfer of taxa between closely related groups.

In an attempt to increase the objectivity of the comparative process, numerical methods have been introduced in recent years. The availability of electronic computers for carrying out the otherwise laborious comparisons has made these methods practicable. The aim of numerical taxonomy as outlined by Sneath (1957, 1962) is to group organisms according to the number of features in common. This, to some extent, has been the process used by orthodox taxonomists in approaching groups of previously unclassified organisms (although much conscious weighting is often applied to characters believed to be important in the evolutionary sense). We therefore hoped that there would be a reasonable degree of correspondence between a classification obtained when we applied numerical methods to the taxonomy of the Fungi Imperfecti and the existing scheme. When there were important differences, we reconsidered both approaches. In doing this, we discovered some anomalies in the numerical method. These were removed by modifications based on theoretical considerations. In its modified form, the numerical method gave a classification agreeing more closely with the orthodox one.

<sup>1</sup>Contribution No. 328, Plant Research Institute and Statistical Research Service.

The Fungi Imperfecti are suitable subjects for numerical methods because no system of classification yet devised for this group can claim any support from genetic or palaeontologic studies, and taxa are based only on morphological (Saccardo 1886), and more recently also on developmental (Vuillemin 1910*a*, *b*, 1911, 1912; Hughes 1953), studies. Thus taxonomic decisions in the Fungi Imperfecti are by nature arbitrary. Recently the use of more advanced observational tools—phase contrast illumination and electron microscopy—has resulted in new micromorphological characters being made available to the taxonomist, and it is on these that the genera employed in the present study are differentiated.

#### *Methods of Computer Taxonomy*

Basically, numerical taxonomy as envisaged by Sneath (1957, 1962) involves four processes:

- (1) Compilation of detailed taxonomic data on the group of organisms under consideration. (See Table III.)
- (2) Coding the taxonomic information in a form suitable for presentation to the computer.
- (3) Computation of a coefficient for each pair of *operational taxonomic units* (OTUs) (individuals, species, etc.) using all available characters.
- (4) Cluster analysis of the coefficients to produce a diagrammatic representation of the results.

#### *(1) Compilation of Data*

In the studies we have carried out, the number of defined characters is not very large (see Table III), and all were included. As the Hyphomycetes display a great range of morphology, most classifications have relied almost exclusively on such features. In the present study, in addition to the morphological characters, a few others such as pigmentation, growth rate in culture, host, and geographic origin were available. In dealing with morphological data it is desirable to guard against the inclusion of numbers of highly correlated characters. However, such associations do not always occur where expected. For instance, in the group of Hyphomycetes studied, the thickness and height of a conidiophore stipe are not necessarily closely correlated, and so it is reasonable to include both in the analysis.

The characters of an organism can be divided into two main types, quantitative and qualitative. The latter can be either simple 'presence or absence' characters, or multistate characters with two or more mutually exclusive alternatives; for example color and shape can be multistate qualitative characters. Quantitative characters are arbitrarily divided into a number of states, the number ranging from two to six in the present study.

The word character has been used rather loosely in numerical taxonomy, and this has led to some confusion. In this work we have, as far as possible, defined the term 'character' as a number of mutually exclusive states. If, in a group of character states, it is possible to score more than one similarity or difference, then more than one character is represented.

#### *(2) Information Coding*

This procedure is discussed in Part II of this paper.

*(3) Similarity and Matching Coefficients*

If the number of characters present and absent in two OTUs, X and Y, are represented by the entries in Table I, then the similarity coefficient used by Sneath (1957) and many others is defined as  $S = a/(a+b+c)$ . Characters absent from both X and Y are excluded in forming this coefficient.

Sokal and Michener (1958), followed by Rogers and Tanimoto (1960) and others, considered that common absences or 'negative matches' should be taken into account, and used what has been called a matching coefficient, defined as  $M = (a+d)/(a+b+c+d)$ .

The relative merits of these coefficients and the modifications necessitated by the inclusion of multistate characters are discussed in Part II of this paper.

TABLE I

	Present in Y	Absent from Y
Present in X	<i>a</i>	<i>b</i>
Absent from X	<i>c</i>	<i>d</i>

*(4) Cluster Analysis*

The simplest method of presenting the results of an analysis in a form in which they can be readily understood is by a *rearranged matrix of coefficients* (Sneath and Sokal 1962). The calculated coefficients are tabulated in matrix form with the OTUs in the same order horizontally and vertically. Each coefficient on the principal diagonal of the matrix is then unity (this being the coefficient obtained when an OTU is compared with itself) and the other coefficients are symmetrical about this diagonal. The order of the OTUs is then altered so that the higher coefficients lie close to the principal diagonal, thus making the most similar OTUs adjacent. This can be shown diagrammatically by using higher densities of shading as the magnitude of the coefficients increases (see examples in Part III, Figs. 11, 12). If a good arrangement of the OTUs has been obtained, the darker areas of the diagram should be in blocks adjacent to the principal diagonal. In our studies it has been possible by this method to pick out the main groups into which the OTUs fall.

The results can also be displayed by means of a *dendrogram* (Sneath and Sokal 1962). The first step in forming a dendrogram is to group the pair of OTUs with the highest coefficient. Once a group has been formed, it is necessary to decide on a measure of similarity between that group and other groups or individuals. In preliminary studies we examined the use of each of the maximum, mean, and median of the coefficients between the members of the groups under comparison. Although the maximum is the simplest to employ, classifications obtained using this measure are likely to be unreliable, since an extreme value may be unrepresentative of the group of similarity values. In this work we have employed the median coefficient to form the dendrograms (Figs. 7-10) since it is both easier to obtain than the mean, and is unaffected by outlying values.

Although we have experienced little difficulty in cluster analysis, other workers have not been so fortunate (Sneath 1962), and further study of this subject is needed.

## Part II. Scoring Methods in Numerical Taxonomy

The scoring of characters is one of the more controversial aspects of numerical taxonomy. In an attempt to arrive at a consistent and logically satisfying system, we have considered several approaches.

### *The Unweighted Method*

Our initial scoring method was based on that of Sneath (1957), although it appears to be closer to that recently advocated by Beers and Lockhart (1962). For characters which could only be present or absent, a presence was scored + and an absence -. For multistate characters, both qualitative and quantitative, the following system was used:

State	1	2	3	4	...
OTU in state 1	+	-	-	-	
" " " 2	NC	+	-	-	
" " " 3	NC	NC	+	-	
" " " 4	NC	NC	NC	+	

where NC stands for 'no comparison'. The number of characters in common ( $a$  in Table I) between any pair of OTUs is the number of character states in which both OTUs score +. The number of characters on which a comparison can be made ( $a+b+c$  in Table I) is the number of character states in which at least one OTU scores +, but excluding those in which the other scores NC.

With the system of scoring just described, it is impossible to score more than one similarity or one difference for any one character.

### *Primary and Secondary Characters*

Some of the characters employed in our studies were what we later termed 'primary' characters—qualitative characters which, if present, were described by further 'secondary' characters, either qualitative or quantitative. For instance, the primary character 'presence of setae' in the Hyphomycete genera *Lacellinopsis* and *Lacellina* (Ellis 1957) was described by eight 'secondary' or 'descriptive' characters: in *Lacellinopsis sacchari*, for example, these were as follows: (1) simple, (2) straight, (3) subulate, (4) thick-walled, (5) pigmented, (6) septate, (7) 400–700  $\mu$  long (maximum), (8) 7–9  $\mu$  thick (maximum) (Proctor and Kendrick 1963). When the primary character was present in an OTU, the secondary characters were scored in the usual way. When the primary character was absent, the secondary characters were scored 'NC'.

Using the unweighted scoring system, it was found that the presence of secondary characters led, in some cases, to anomalies in the relationships between OTUs measured by the similarity coefficients. This occurred because it was possible for differences in secondary characters to outweigh similarities in primary characters.

To demonstrate this with a hypothetical example, consider three OTUs, X, Y, and Z. Two of these, Y and Z, are similar in every respect except that Z lacks a primary character  $\alpha$  possessed by Y, and also by X. Suppose, further, that  $\alpha$  is described by two secondary characters,  $\beta_1$  and  $\beta_2$ , in both of which X and Y differ. From a commonsense standpoint, X and Y are more similar than X and Z, because X and Y have the primary character  $\alpha$  in common.

If  $\alpha$ ,  $\beta_1$ , and  $\beta_2$  are ignored, the coefficients of similarity  $S_{XY}$  and  $S_{XZ}$  are the same:  $S_{XY} = S_{XZ} = k/n$ , where  $k$  is the number of characters in common between X and Y or X and Z, and  $n$  is the number of characters on which comparisons can be made.

If  $\alpha$ ,  $\beta_1$ , and  $\beta_2$  are then included, the coefficients become:

$$S'_{XY} = (k+1)/(n+3)$$

since X and Y have  $\alpha$  in common and differ on  $\beta_1$  and  $\beta_2$ ,

$$S'_{XZ} = k/(n+1)$$

since X and Z differ on  $\alpha$  and no comparisons can be made on  $\beta_1$  and  $\beta_2$ . Hence, if  $k > (n+1)/2$ ,  $S'_{XZ} > S'_{XY}$ .

Thus, using the coefficient and scoring system described above, it is possible to arrive at results which conflict with intuitive ideas. In the example just given, the primary character was described by only two secondary characters, but primary characters in our studies have been described by as many as 11 secondary characters, making anomalous results even more likely.

*Alternatives to the Original Unweighted Method: Introduction of Weighting based on Primary and Secondary Characters*

To avoid these difficulties, three approaches seemed possible.

(i) Omit all secondary characters. This is proposed by Beers and Lockhart (1962) and may be practicable in groups of organisms where there are few secondary characters. In the Fungi Imperfecti, however, secondary characters represent a high proportion of the total characters available and may be entirely responsible for 'species' differentiation.

(ii) Give secondary characters less but non-zero weight. One possible method would be to weight the secondary characters of a primary character so that the effect on a similarity index of all these secondary characters could not be greater than that obtained by the addition of another primary character. Thus, if a primary character had eight associated secondary characters, each of the latter might be given the weight 1/8. A practical objection to this approach is that the scoring system would be complicated by the introduction of fractional values. A second objection is that a simple primary character with only one secondary character could affect the classification as much as a complex primary character with many secondary characters.

(iii) Give additional weight to the primary characters according to the number of associated secondary characters. The method adopted was to give a primary character with  $m$  secondary characters a weighting of  $(m+1)$ . Thus in the above example,  $\alpha$  is given a weight of three and the coefficients become

$$S''_{XY} = (k+3)/(n+5),$$

$$S''_{XZ} = k/(n+3).$$

$S''_{XZ}$  cannot then be greater than  $S''_{XY}$ , whatever the values of  $k$  and  $n$ . This is the 'weighted' method of scoring which we have applied in our studies on the fungi.

*Additive Scoring for Quantitative Characters*

A disadvantage of the scoring systems discussed so far, when applied to quantitative characters, is that the magnitude of differences between character states is ignored. A single difference is scored both when the pair of OTUs

compared are in adjacent character states and when they lie at opposite ends of the total range of states.

A system of scoring which has been proposed to overcome this difficulty is the 'additive' method (Sneath 1962). A quantitative character divided into four states is scored as follows:

State	1	2	3	4
OTU in state 1	+	-	-	-
" " " 2	+	+	-	-
" " " 3	+	+	+	-
" " " 4	+	+	+	+

Although these divisions are referred to as states, under our definition they would be regarded as separate characters since it is possible to score more than one similarity or difference in a comparison of two OTUs. For example, if an OTU in state 1 is compared with one in state 4, the number of characters in common will increase by 1, and the number on which the comparison is made by 4.

Although this scheme has the advantage that it makes some allowance for the magnitude of quantitative differences, it can produce other anomalies. One objection to it is that a quantitative character assumes more importance (i.e. has more weight) than a qualitative one. A second is that the similarity coefficient between two large individuals is likely to be higher than that between two small, but otherwise equally similar, individuals. (This objection does not apply when matching coefficients are used.)

It might be better to give fractional values to differences in quantitative characters. A difference of unity between a pair of OTUs would only be scored if they fell at opposite ends of the range of character states. Although this system seems theoretically more desirable, we have not yet applied it because of practical difficulties associated with the introduction of fractional scoring.

#### *The Inclusion of Common Absences: Matching Coefficients*

In Sneath's original scheme, only the common presence of a character was counted as a similarity, characters absent from both OTUs being ignored. It has since been argued (Sokal and Michener 1958; Hill *et al.* 1961; Beers and Lockhart 1962) that there are good reasons for considering the common absence of a character as a similarity.

The principal argument against the inclusion of such 'negative matches' is that there would then be no limit to the number of characters which could be considered. However, the main effect of including redundant characters, that is, those which are either always present or always absent in the organisms under comparison, is to reduce the range of possible values of the matching coefficients. For the purposes of classification, a widely ranging coefficient is likely to be preferred, and so an investigator is not likely to include many redundant characters. In the study described in Part III, all redundant characters are excluded, but, provided matching coefficients are used, the inclusion of redundant primary characters makes little difference to the classification obtained.

An advantage of matching coefficients is that the number of characters on which a comparison is made ( $a+b+c+d$  in Table I) is less variable than with similarity coefficients ( $a+b+c$  in Table I). With the latter, because of differences between denominators (i.e. in values of  $a+b+c$ ), the inclusion of redundant characters can change the relative values and hence the classification obtained from them. To demonstrate this, consider two hypothetical similarity coefficients,  $S_1$  and  $S_2$ . For  $S_1$ ,  $a=10$ ,  $a+b+c=20$ ; for  $S_2$ ,  $a=20$ ,  $a+b+c=40$ . Then:

$$S_1 = \frac{10}{20} = .5, S_2 = \frac{20}{40} = .5.$$

Suppose that 20 'presence or absence' characters, present in all OTUs under consideration are now added. The coefficients become:

$$S'_1 = \frac{10+20}{20+20} = .75,$$

$$S'_2 = \frac{20+20}{40+20} = .67,$$

i.e. the coefficients are no longer equal and the dendrogram obtained from all the coefficients may be altered. With matching coefficients, classifications are more stable because of the comparative constancy of the denominator, and should not be altered by the inclusion of redundant primary characters.

A difficulty which is removed by the inclusion of negative matches is that of deciding which is the positive and which the negative aspect of a character. With some characters, such as the presence or absence of spore curvature in fungi, it is not possible to know which aspect to score positively.

Multistate characters are treated in the same way for matching as for similarity coefficients, except when quantitative characters are scored by the 'additive' method.

To enable a computer program originally designed for the calculation of similarity coefficients to be used for the calculation of matching coefficients, it was merely necessary to divide the single state 'presence or absence' characters into two states and score them as follows:

State	Present	Absent
OTU with character present	+	-
OTU with character absent	NC	+

A 'presence or absence' character is now a two-state qualitative character and, provided the 'additive' method is not used for quantitative characters, the scoring system for all characters is uniform.

#### *Discussion of Methods*

Of the methods used so far, that with (1) weighting of primary characters, (2) scoring of negative as well as positive matches, (3) non-additive scoring of quantitative characters, seems to us to be the most satisfactory both from the logical and practical points of view.

(1) The term 'weighting' is unfortunately associated with the assignment of different levels of importance to characters on a subjective basis. Because of

this, it has been argued many times (Sneath 1957; Sneath and Sokal 1962) that all characters should be given equal weight. However, once it has been decided which characters to include, the weighting scheme described in this paper is objective. Admittedly, it would be possible to give excessive weight to a particular primary character by including redundant secondary characters. To avoid this danger, in our work we have excluded all secondary characters which do not vary in the group of fungi under consideration.

It may be of interest to note that a similar result to that obtained by our weighted method could be obtained by an extension of Sneath's method B (Sneath 1957).

Although in our application of numerical methods in taxonomy, we have encountered only primary and secondary characters, in other studies several levels of character dependence could occur. Thus secondary characters might be described by tertiary characters, etc. There is no real difficulty in extending the weighting method to this situation.

(2) The inclusion of negative matches appears to be a definite improvement, and no objections to this have arisen in the course of our experimental work.

(3) Ignoring the magnitude of differences between character states when scoring quantitative characters is not entirely satisfactory, but it seems preferable to assigning greater importance to quantitative than qualitative characters, which is the implication of the 'additive' system.

### Part III. Application of Numerical Taxonomy to Hyphomycetes

The members of the group of Hyphomycetes chosen for this study are characterized by a darkly pigmented, mononematous conidiophore bearing at its apex a complex sporogenous apparatus consisting of one to six multiplicative series of metulae, the ultimate series of which bears numerous sporogenous cells. The conidia are always slimy amerospores which accumulate in a mucilaginous drop around the sporogenous apparatus. All usually occur on or are isolated from wood, sometimes causing a diseased condition in living trees and frequently producing blue stain.

The numerous similarities of morphology and habitat led earlier workers to assign the members of this group to a single genus. In 1953, however, Hughes proposed a system of classification for Hyphomycetes in which the different methods of spore production were used to delineate the major groups. On closer examination, the Hyphomycetes under consideration in our study were seen to fall into two of these groups. Some species produced their spores in basipetal succession from open-ended cells called phialides which did not increase in length with spore production (Hughes 1953, Section IV). Others produced spores singly as blown-out ends, at the apex of the sporogenous cell and of the successive new growing points which developed just behind and to one side of the previously terminal conidium. At maturity, a sporogenous cell of this type possesses a number of scars, each of which was, in turn, terminal before being pushed aside by the development of a new growing point. These sporogenous cells show a perceptible increase in length as spore production proceeds, and have recently been termed sympodulae (Kendrick 1962; Hughes 1953, Section II).

The phialide-bearing species are placed in the genus *Phialocephala* Kendrick (1961, 1963a, b) and the sympodula-bearing species belong to the genus *Verticicladiella* Hughes (Hughes 1953; Kendrick 1962). It appears that this group is an example of parallel evolution to produce the spore dispersal method which Ingold (1961) calls the 'stalked spore drop' (although he does not include among his examples the group of fungi being considered here).

As this group is still the subject of intensive taxonomic study, it has been chosen as the object of the present numerical study in the hope that the affinities of 'unknowns' may in future be revealed by the computer, and that meaningful lines may be drawn between apparently intergrading species, or, where this is not possible, that the existence of the biological continuum may be recognized and accepted.

The taxonomic protocol used here is derived from data sheets compiled prior to the publication of papers on the two genera concerned (Kendrick 1961, 1962, 1963a). A total of 30 isolates were included, representing seven species of *Verticicladiella* and three species of *Phialocephala*. They are listed in Table II.

#### Methods

In this study, three of the scoring methods described in Part II of this paper were employed, using both similarity and matching coefficients: (1) the

TABLE II  
List of isolates studied

Isolate Code No.	Herb. DAOM No.	Name
1	29111	<i>Verticicladiella penicillata</i> (Grosm.) Kendrick
2	84338	<i>V. antibiotica</i> Kendrick
3	71293	"
4	33961(b)	<i>V. brachiata</i> Kendrick
5	45373(a)	"
6	33974	"
7	34360	"
8	33941	<i>V. abietina</i> (Peck) Hughes
9	37980	"
10	60338	"
11	60807	"
12	62102	"
13	64328(a)	"
14	63700	<i>V. procera</i> Kendrick
15	63686	"
16	62093	"
17	62094	"
18	62095	"
19	62096	"
20	33940	"
21	87324	<i>V. wagnerii</i> Kendrick
22	34896	<i>V. serpens</i> (Goid.) Kendrick
23	71381	<i>Phialocephala dimorphospora</i> Kendrick
24	71465(c)	"
25	87232	"
26	87233	"
27	28736	<i>P. bactrospora</i> Kendrick
28	63900	"
29	71357	<i>P. repens</i> (Cooke & Ellis) Kendrick
30	89659	"

unweighted method, (2) the additive method, (3) the weighted method of Proctor and Kendrick.

The list of characters originally compiled is given in Table III, primary and secondary characters being indicated.

TABLE III  
*Verticicladiella* and *Phialocephala*: list of characters and states

No.	Type (see Key)	Description	States
1	P(3)	Geographical source	(1) Western North America (2) Eastern North America (3) Europe
2	P(4)	Host	(1) <i>Pinus</i> (2) <i>Picea</i> (3) <i>Pseudotsuga</i> (4) Broadleaved sp.
3	P*	Associated w. bark beetles	
4	P(3)	Associated w.	(1) Brown rot (2) Wedge blue stain (3) Annular blue stain
5	P*	Associated w. root disease	
6	P*	Found fruiting on the host	
7	P*	Perfect state known	
8	PR	COLONY	
9	S(5) of 8	Diam. at 15 days at 25° C on malt agar	(1) < 2 cm (2) 2-4 cm (3) 4-6 cm (4) 6-8 cm (5) > 8 cm
10	S* of 8	Conidiophores in concentric zones	
11	S* of 8	Optimum growth temp. > 20° C	
12	S* of 8	Producing diffusible antibiotic	
13	PR	MYCELIUM	
14	S* of 13	Pigmented hyphae present	
15	S* of 13	Hyaline hyphae present	
16	S* of 13	Fine contorted hyphae present	
17	S(3) of 13	Hyphal width	(1) < 4 μ (2) 4-8 μ (3) > 8 μ
18	PR	CONIDIOPHORE STIPE	
19	S(6) of 18	Length of stipe	(1) < 200 μ (2) 200-400 μ (3) 400-600 μ (4) 600-800 μ (5) 800-1000 μ (6) > 1000 μ
20	S(3) of 18	Width of stipe at base	(1) < 10 μ (2) 10-15 μ (3) > 15 μ
21	S(4) of 18	Width of stipe at apex	(1) < 5 μ (2) 5-7 μ (3) 7-9 μ (4) > 9 μ
22	S(3) of 18	Number of septa in stipe	(1) < 5 (2) 5-10 (3) > 10

TABLE III (Continued)

No.	Type (see Key)	Description	States
23	S(3) of 18	Thickness of stipe wall	(1) <0.5 $\mu$ (2) 0.5-1.5 $\mu$ (3) >1.5 $\mu$
24	S(3) of 18	Basal pigmentation	(1) Pale (2) Deep (3) Very dark
25	S* of 18	Stipe pigment strongly reddish	
26	S* of 18	Stipe wall rough	
27	S* of 18	Stipe wall w. amorphous brown deposits	
28	S* of 18	Stipe w. characteristic lateral outgrowths	
29	S* of 18	Stipe w. blunt fingerlike projections growing downward from base	
30	S* of 18	Stipe w. pigmented rhizoidal hyphae at base	
31	PR	SPOROGENOUS APPARATUS	
32	S(4) of 31	Total length of sporogenous apparatus	(1) <60 $\mu$ (2) 60-100 $\mu$ (3) 100-140 $\mu$ (4) >140 $\mu$
33	S(3) of 31	Number of primary metulae	(1) <4 (2) 4-5 (3) >5
34	S(3) of 31	Length of primary metulae	(1) <18 $\mu$ (2) 18-35 $\mu$ (3) >35 $\mu$
35	S(3) of 31	Width of primary metulae	(1) <5 $\mu$ (2) 5-7 $\mu$ (3) >7 $\mu$
36	S* of 31	Primary metulae rather divergent	
37	S* of 31	Primary metula sometimes growing on and producing secondary sporogenous apparatus	
38	S(3) of 31	Number of series of metulae	(1) 3 or fewer (2) Up to 4 (3) Up to 5
39	S(4) of 31	Pigment extending into	(1) Primary metulae only (2) Secondary metulae (3) Tertiary metulae (4) Sporogenous cells
40	S(4) of 31	Number of sporogenous cells	(1) <50 (2) 50-150 (3) 150-250 (4) >250
41†	P(2)	Type of sporogenous cell	(1) Sympodula (2) Phialide
42	S(3) of 41(1)	Length of sympodulae	(1) <15 $\mu$ (2) 15-30 $\mu$ (3) >30 $\mu$
43	S(3) of 41(1)	Width of sympodulae at base	(1) <1.5 $\mu$ (2) 1.5-2.5 $\mu$ (3) >2-5 $\mu$

TABLE III (*Concluded*)

No.	Type (see Key)	Description	States
44	S(2) of 41(2)	Width of phialides	(1) <3 $\mu$ (2) >3 $\mu$
45	S(2) of 41(2)	Length of collarette	(1) <6 $\mu$ (2) >6 $\mu$
46	S(2) of 41(2)	Width of collarette	(1) <2.3 $\mu$ (2) >2.3 $\mu$
47†	P(2)	Type of spore	(1) Sympodioconidium (2) Phialoconidium
48	S(4) of 47(1)	Length of sympodioconidia	(1) <5 $\mu$ (2) 5-7 $\mu$ (3) 7-9 $\mu$ (4) >9 $\mu$
49	S(3) of 47 (1)	Width of sympodioconidia	(1) <1.5 $\mu$ (2) 1.5-4.0 $\mu$ (3) >4.0 $\mu$
50	S* of 47(1)	Sympodioconidia mostly curved	
51	S* of 47(1)	Abstriction scar well marked	
52	S(2) of 47(2)	Length of phialoconidia	(1) <3 $\mu$ (2) >3 $\mu$
53	S(2) of 47(2)	Width of phialoconidia	(1) <2 $\mu$ (2) >2 $\mu$
54	S* of 47(2)	Phialoconidia mostly spherical	
55	S* of 47(2)	First phialoconidium differing from others	

## KEY TO SYMBOLS:

P(*n*) = Primary character with *n* states.

P\* = Primary 'presence or absence' character.

PR = Redundant primary character (not used in forming our coefficients).

S(*n*) of *x* = Secondary character with *n* states associated with primary character *x*.S\* of *x* = Secondary 'presence or absence' character associated with primary character *x*.

†In this study, a primary character with more than one positive state was weighted according to the sum of the secondary characters of the different states. This procedure is debatable, but was adopted because of the difficulties presented by unequal numbers of secondary characters.

The calculations of similarity and matching coefficients were carried out on an IBM 1620 computer using a program originally compiled for the IBM 650 computer (Colwell and Liston 1961) and simulated by Colwell for the IBM 1620. It is designated CAB-1 (see *Taxometrics* No. 1, June 1962, feature 3, page 1).

The results are presented by means of rearranged matrices (Figs. 11, 12) and dendrograms (Figs. 7-10).

All 'redundant' characters, that is, those possessed by all organisms in the group under consideration, have been removed, and the dendrograms thus emphasize the differences between the organisms, rather than their overall similarity. If this analysis were expanded to include fungi not possessing certain of the characters found in all members of the present sample, those characters would have to be reinstated.

*Discussion*

The justification for these techniques has been discussed in Part II of this paper. It is obvious that using the weighted method of scoring, and similarity

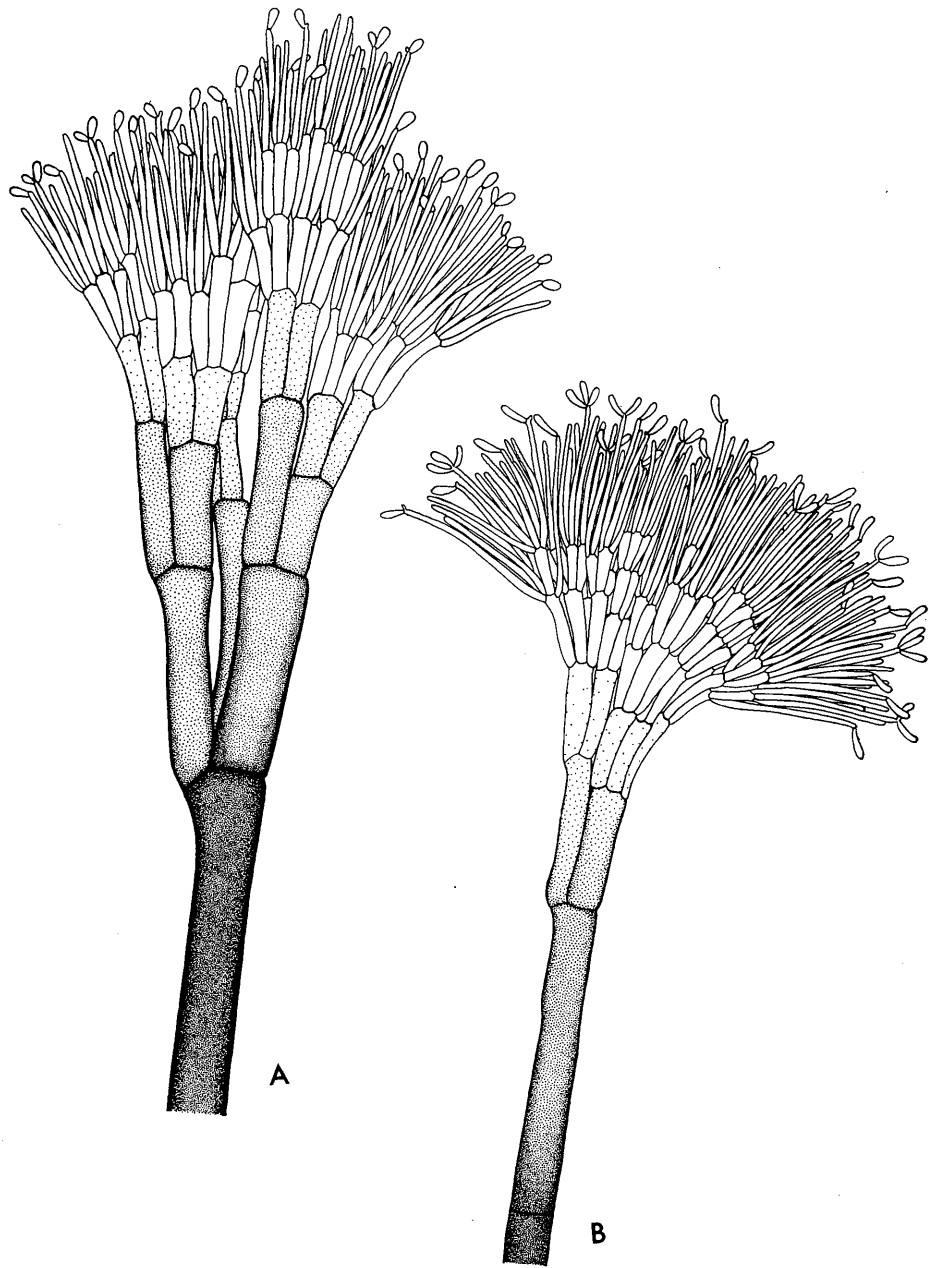


FIG. 1. A, apical portion of conidiophore of *Verticicladiella procera*. B, apical portion of conidiophore of *Verticicladiella abietina*. ( $\times 1000$ )

or matching coefficients, the two genera dealt with in this study are very adequately separated (Figs. 9, 10, 12).

This cannot be said when the unweighted or additive scoring methods are applied. When the additive method is used (Fig. 8), the generic line is crossed twice. It has already been established in Part II of this paper that the additive method suffers from a disadvantage in that each *state* of a quantitative multi-state character is, in fact, treated as a *character* and such quantitative multi-state characters accordingly assume disproportionate importance in the analysis. Here such characters as the long conidiophore stipe of *P. bactrospora* have tended to link it more closely with a group of tall *Verticicladiella* species than with other species of *Phialocephala*. In addition various dimensional and numerical similarities have apparently linked *Verticicladiella brachiata* with *Phialocephala dimorphospora* and *P. repens* rather than with the other species of *Verticicladiella*. With the original unweighted method (Figs. 7, 11) using either similarity or matching coefficients, anomalous results are obtained because of the presence of large numbers of secondary characters, and the generic separation is lost, as it was in our earlier study (Proctor and Kendrick 1963).

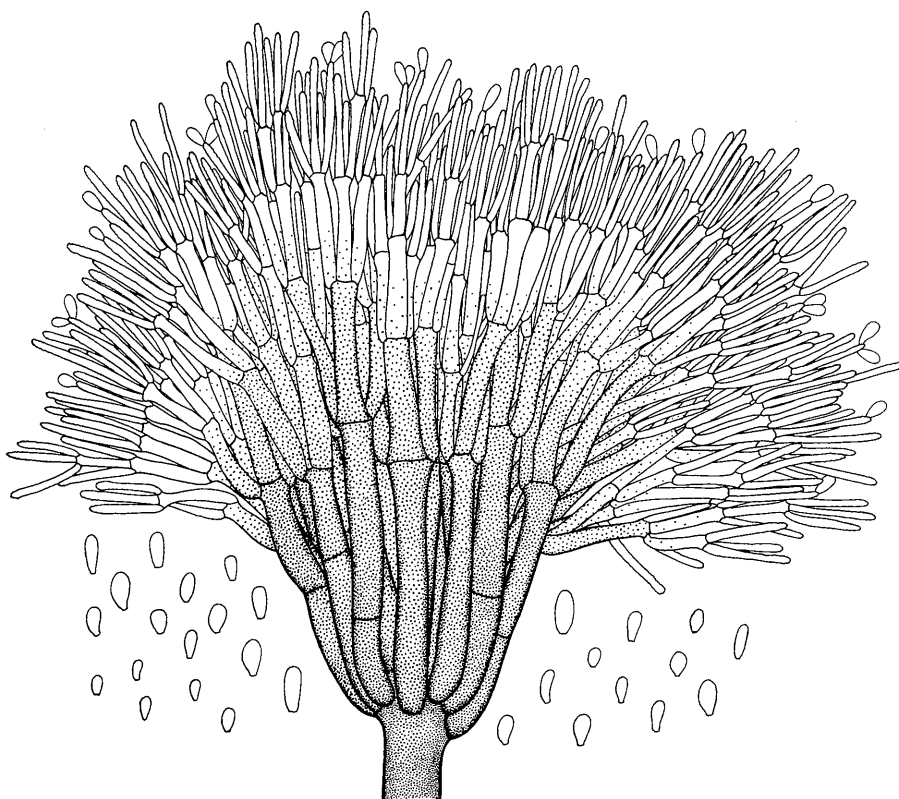


FIG. 2. Apical portion of conidiophore of *Verticicladiella wagnerii*. ( $\times 800$ )

While the use of a matching coefficient has already been justified on logical grounds in Part II of this paper, it is interesting to note how this logic is also borne out in the dendrograms derived from the weighted method. That derived from matching coefficients (Fig. 10) maintains the arrangement of isolates predicted from the rearranged matrix, while, on that derived from similarity coefficients (Fig. 9), *V. antibiotica* (isolates 2, 3) is linked with *V. brachiata* (isolates 4-7) at a higher level than either is linked with the other species of the genus. From examination of the matrices and of the isolates it would appear best that these two species should be placed, for the present at

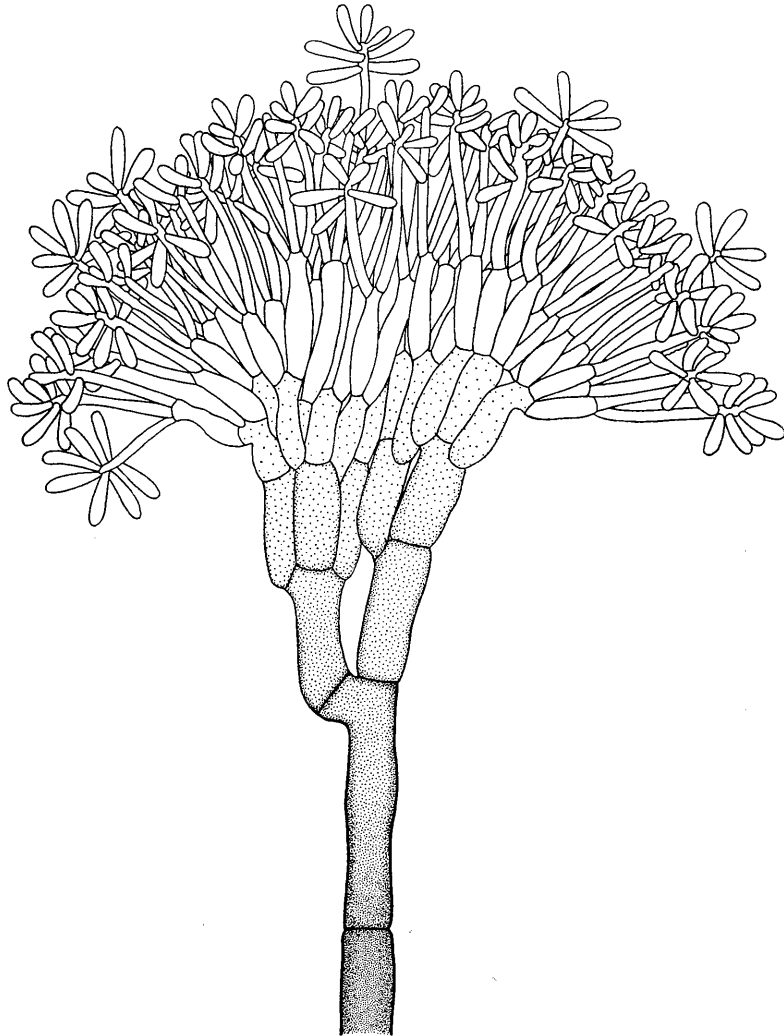


FIG. 3. Apical portion of conidiophore of *Verticicladiella penicillata*. ( $\times 1000$ )

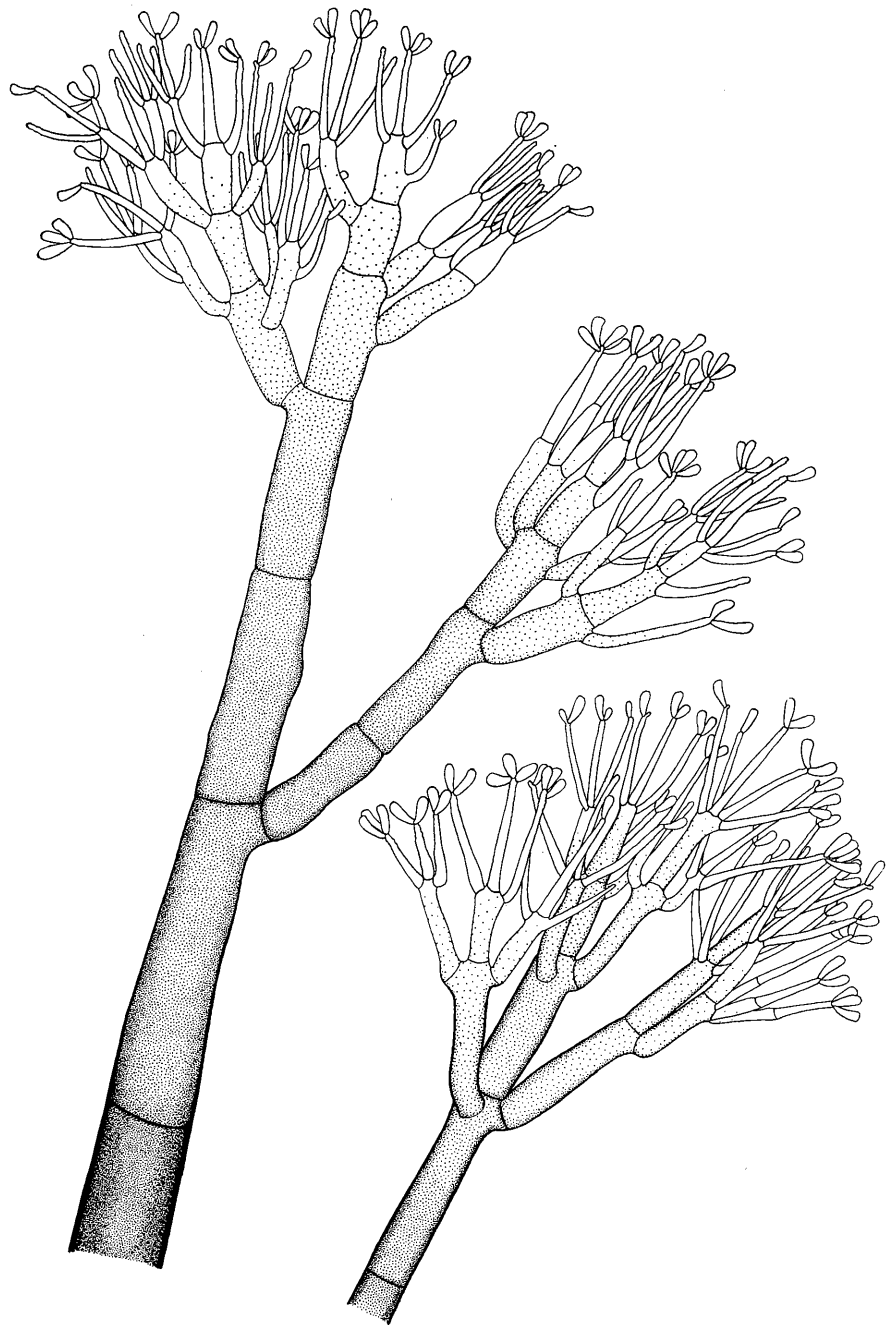


FIG. 4. Apical portions of conidiophores of *Verticicladiella serpens*. ( $\times 1000$ )

least, at opposite ends of the spectrum within the genus *Verticicladiella*, as in Figs. 10 and 12. During preliminary work in which some redundant characters were included, the dendrogram derived from the weighted method using matching coefficients was of the same form as that derived after the removal of redundancies. Where the weighted method using similarity coefficients was employed, the corresponding dendrograms differed in the arrangement of some isolates.

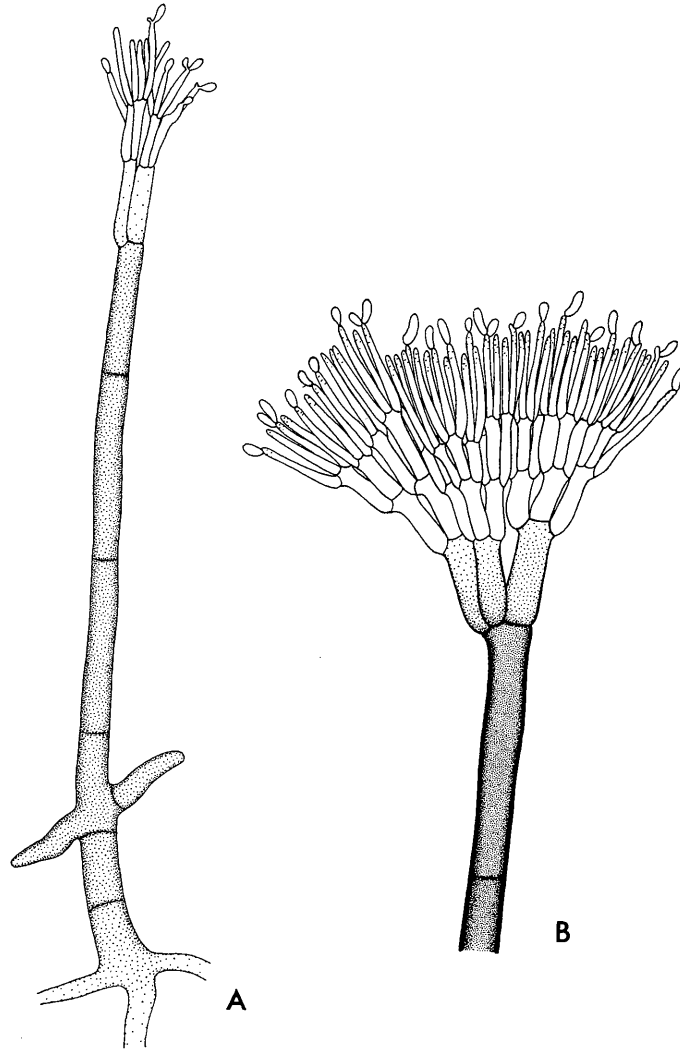


FIG. 5. A, conidiophore of *Verticicladiella brachiata*. B, apical portion of conidiophore of *Verticicladiella antibiotica*. ( $\times 1000$ )

The dendrogram which most accurately reflects contemporary taxonomic thought on the group under consideration is that obtained by using the weighted method of scoring and matching coefficients (Fig. 10). In addition these approaches have been shown to be logically the most satisfying of those considered in this paper. Certain features of this dendrogram, however, require some explanation, and camera lucida drawings of several species of *Verticicladiella* and *Phialocephala* (Figs. 1-6) are given for comparison and to aid the reader in the discussion which follows.

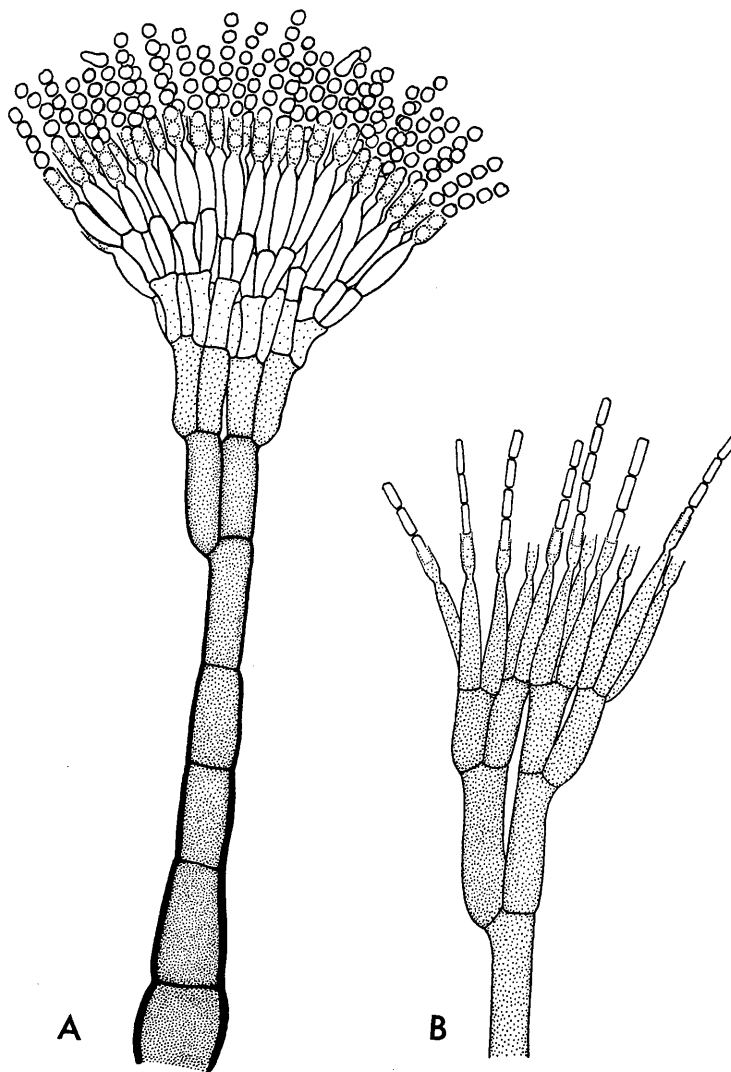


FIG. 6. A, conidiophore of *Phialocephala dimorphospora*. B, apical portion of conidiophore of *Phialocephala bactrospora*. ( $\times 1000$ )

*Verticicladiella abietina* (isolates 8–13; Fig. 1B), the type species of the genus, resembles *V. procera* (isolates 14–20; Fig. 1A) in many ways, differing chiefly in the size and number of the various components of the conidiophore, and in spore proportions and curvature. However, it is noticeable that because of our present reliance on morphological characters, and often on differences in size of such characters, *V. wagnerii* (isolate 21; Fig. 2), a species with large and complex conidiophores, appears in the dendrogram more closely linked to *V. procera* than *V. abietina*. Again largely because of dimensional similarities, both *V. penicillata* (isolate 1; Fig. 3) and *V. serpens* (isolate 22; Fig. 4) are linked with *V. abietina* before *V. abietina* is linked with *V. procera*.

As a taxonomist applying the intuitive approach to this group, W.B.K. considers that if a wider range of data, including physiological and genetic characters, were available, the dendrogram would show closer affinity between *V. abietina* and *V. procera* than between either of these species and any other *Verticicladiella* yet described. Some improvement in the method of dealing with quantitative characters, such as the fractional scoring mentioned in Part II, may also help. Some hint of this may be seen in the rearranged matrix (Fig. 12) in which isolates 8 and 9 of *V. abietina* show a high similarity to most of the *V. procera* isolates.

*V. procera* appears to be a tightly knit species, but *V. abietina* may be segregated into two groups, each of three isolates. Detailed examination of data shows the differences to be due mainly to larger dimensions of the various parts of the conidiophores in the 11–13 group. Two possible explanations arise at once.

The 8–10 group were originally cultured in 1952, 1953, and 1951, respectively, while the 11–13 group were isolated more recently, in 1958, 1957, and 1958, respectively. In this group of fungi, aging of a culture sometimes has the effect of causing a degeneration of the conidiophores formed, and often a complete loss of ability to produce recognizable stipitate conidiophores; even the ability to produce the normal brown pigment may be lost. In the present case, while the conidiophores of the 8–10 group are smaller than those of the 11–13 group, they are nevertheless well-defined and sufficiently constant to be easily identifiable to species.

A second possible reason for the differences between the two groups might be that the 8–10 group are all derived from species of *Picea* (though from such diverse locations as Ontario, British Columbia, and Colorado), and might constitute a 'race' peculiar to that host genus. Whatever the causes of this apparent differentiation, they seem to me quite insufficient to warrant the erection of two form species rather than the single taxon which at present encompasses these six isolates.

It may be argued that the level at which 8–10 and 11–13 groups are linked is very close to that at which *V. penicillata* (isolate 1) and *V. serpens* (isolate 22) are maintained as separate species. However, rather than taking this to imply any weakness in the concept of *V. abietina*, I incline to the view that *V. penicillata* and *V. serpens* are closely 'related' in the genetic sense of that word. The numerical scheme is here merely corroborating the opinion of Hunt (1956), who suggested that '*C. [eratocystis] penicillata*, and . . . *C. serpens* [the perfect states of *V. penicillata* and *V. serpens*] have perithecia, ascospores,

and imperfect states which are very similar. *C. penicillata* . . differ[s] from *C. serpens* in the longer ascospores'. I have observed that the conidia of *V. penicillata* are longer than those of *V. serpens*. If the metulae of *V. serpens* (Fig. 4) were not so divergent, this species would resemble *V. penicillata* quite closely, and it is probable that the original isolate was, in fact, more similar to that species than it appears today. It seems possible that these two species, both known thus far only from Europe, have diverged comparatively recently from a common ancestral stock, and remain 'difficult' or 'critical' (Burgess 1955).

The rearranged matrices convey more detailed information than the dendrograms, which give only a generalized picture of the groupings, and

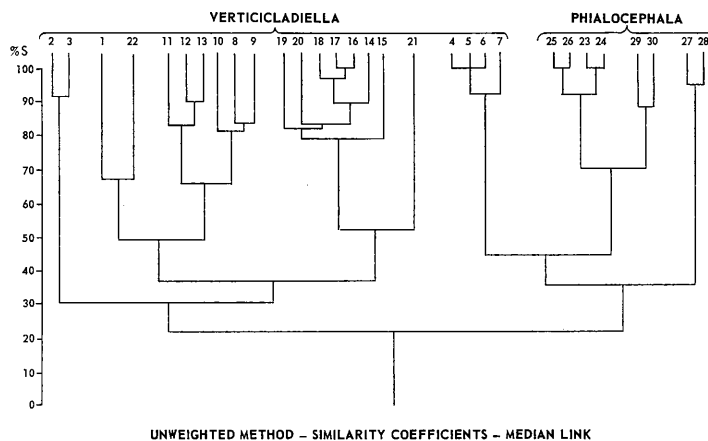


FIG. 7. Note that *Verticicladiella brachiata* (isolates 4-7) is linked with *Phialocephala* rather than with the other species of *Verticicladiella*.

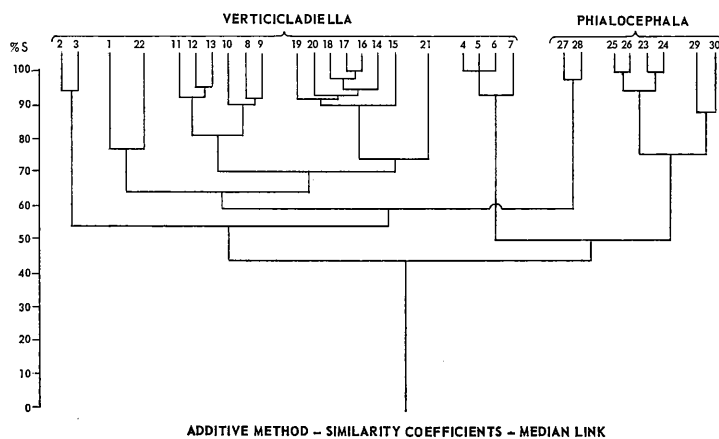


FIG. 8. Note that the generic line is crossed twice; *Phialocephala bactrospora* (isolates 27, 28) apparently linking with *Verticicladiella*, and *Verticicladiella brachiata* (isolates 4-7) with *Phialocephala*.

it is from a consideration of the matrices (of which examples are given in Figs. 11 and 12) that the arrangement of isolates shown in the dendrograms has been derived. An examination of the matrix in Fig. 12 will show why isolates 8-13 (*V. abietina*) and 14-20 (*V. procera*) have been placed at the center of the genus *Verticicladiella*, and why isolates 2 and 3 (*V. antibiotica*) (Fig. 5B) and 4-7 (*V. brachiata*) (Fig. 5A) have been placed peripherally.

In *Verticicladiella* it is perhaps fortunate that *V. abietina*, the type species, does at present seem very 'typical' of the genus; that is, it represents in many ways a midpoint between the extremes so far found in the genus. *Phialocephala dimorphospora* (Fig. 6A), the type species of *Phialocephala*, is unlikely to represent such a median form, as it possesses several important characters not

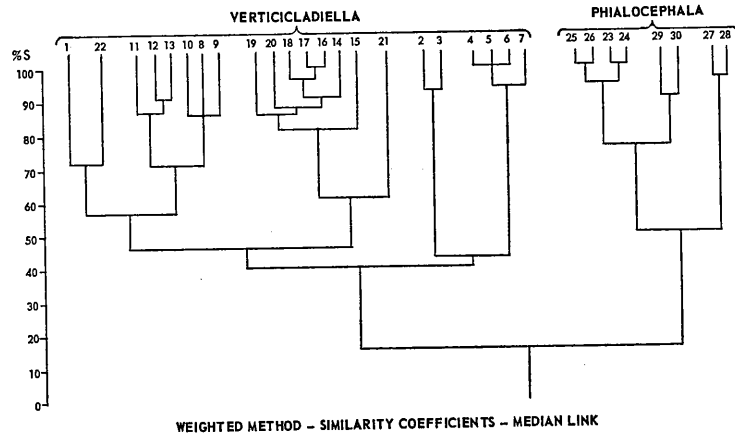


FIG. 9. Note the clear separation of the two genera, but the slightly anomalous association between *Verticicladiella antibiotica* (isolates 2, 3) and *V. brachiata* (isolates 4-7).

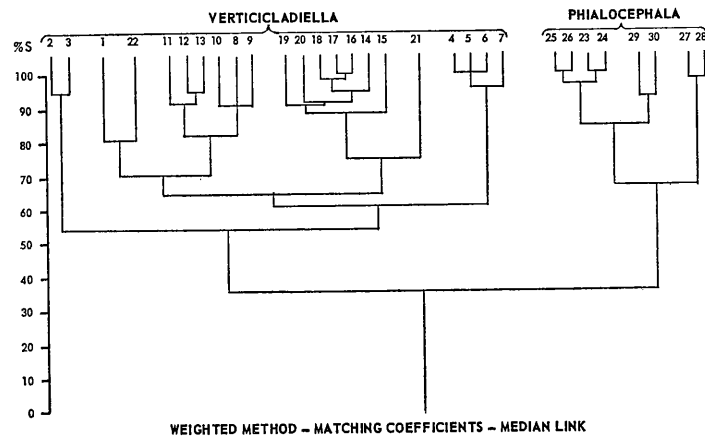


FIG. 10. What seems to us the most logical method has produced the most satisfactory dendrogram.

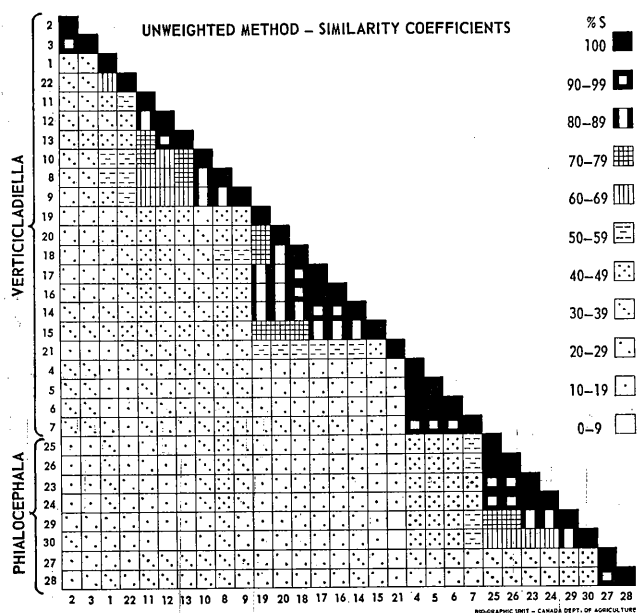


FIG. 11. Rearranged matrix of coefficients. Note the poor definition of some species and the high similarities between isolates 4-7 and isolates 23-26, 29, and 30. More detailed comparisons, especially between individual isolates, can be made from this type of diagram than from the dendrograms.

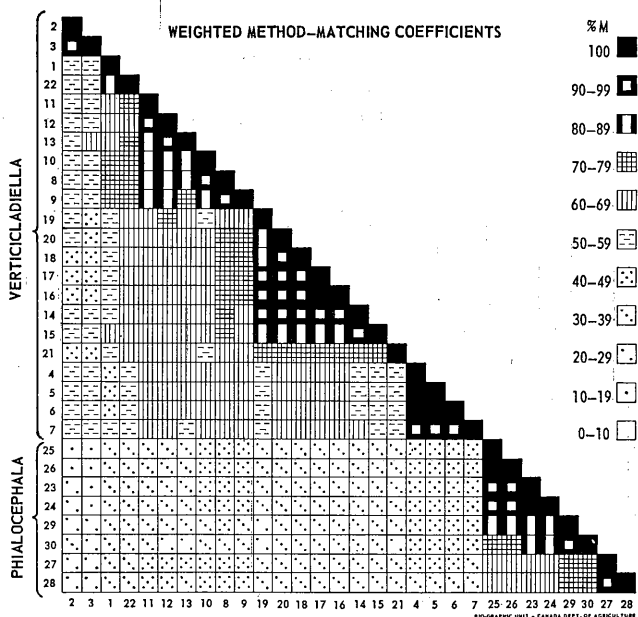


FIG. 12. Rearranged matrix of coefficients. Note the clear generic separation. In no case does an intergeneric coefficient exceed an intrgeneric coefficient.

found in all other species of the genus. While only three species are dealt with here, two further species have been described recently (Kendrick 1963*b*); among the five species so far described and several further species awaiting description, *P. dimorphospora* is the only one to produce two morphologically distinct types of spore. In addition it shares its possession of a long cylindrical collarette only with *P. bactrospora* (Fig. 6B). *P. dimorphospora* was, in fact, chosen as type species of *Phialocephala* because it exhibited the phialide so clearly. With only three species under consideration here, it is difficult to draw any conclusions as to what a 'median' species might look like, but further numerical studies with a larger number of species may ultimately reveal one.

The weighted methods discussed above have now been successfully applied to two quite different groups of the Fungi Imperfecti for which an acceptable system of classification already exists. An analogy may be drawn here with the electron microscope, which, since its relatively recent introduction, has made great contributions to knowledge. However, during the early days of its application it was often found necessary to check many borderline observations by referring back to the light microscope, and in this way many artifacts were discovered and eliminated. It seems to us that a similar approach can legitimately be made to numerical taxonomy, especially in groups where a relatively well-established classification already exists. The established bacterial classification, on the other hand, was apparently considered very unsatisfactory by some bacteriologists who felt that a completely fresh approach was necessary. This, in fact, is probably the reason for the development of numerical taxonomy by bacteriologists, and its wider acceptance to date among bacteriologists than by specialists in any other group.

It has been our intention in this paper to draw the attention of mycologists and workers in other related disciplines to some of the potentialities of numerical taxonomy. We intend in future to employ and explore further these methods in the study of groups of organisms whose taxonomy is confused or poorly understood—for example the genus *Leptographium* itself—and it is hoped that these applications will be the subject of future communications.

#### Acknowledgments

We are indebted to Dr. J. W. Groves, Mrs. P. M. Morse, Dr. P. Robinson, and Dr. L. K. Weresub for many helpful suggestions during the preparation of the manuscript.

#### References

- BEERS, R. J. and LOCKHART, W. R. 1962. Experimental methods in computer taxonomy. *J. Gen. Microbiol.* **28**, 633–640.
- BURGES, N. A. 1955. Problems associated with the species concept in mycology. *In* Species studies in the British flora. *Edited by* J. E. Lousley. Arbroath. pp. 65–82.
- COLWELL, R. R. and LISTON, J. 1961. Taxonomic relationships among the pseudomonads. *J. Bacteriol.* **82**, 1–14.
- ELLIS, M. B. 1957. *Haplobasidium*, *Lacellinopsis* and *Lacellina*. Commonwealth Mycol. Inst. Mycol. Paper No. 67, 1–15.
- HILL, L. R., TURRI, M., GLARDI, E., and SILVESTRI, L. G. 1961. Quantitative methods in the systematics of Actinomycetales—II. *Giorn. Microbiol.* **9**, 56–72.
- HUGHES, S. J. 1953. Conidiophores, conidia, and classification. *Can. J. Botany*, **31**, 577–659.
- HUNT, J. 1956. Taxonomy of the genus *Ceratocystis*. *Lloydia*, **19**, 1–59.
- INGOLD, C. T. 1961. The stalked spore-drop. *New Phytologist*, **60**, 181–183.

- KENDRICK, W. B. 1961. The *Leptographium* complex. *Phialocephala* gen. nov. Can. J. Botany, **39**, 1079-1085.
- 1962. The *Leptographium* complex. *Verticicladiella* Hughes. Can. J. Botany, **40**, 771-797.
- 1963a. The *Leptographium* complex. *Penicillium repens* C. & E. Can. J. Botany, **41**, 573-577.
- 1963b. The *Leptographium* complex. Two new species of *Phialocephala*. Can. J. Botany, **41**, 1015-1023.
- PROCTOR, J. R. and KENDRICK, W. B. 1963. Unequal weighting in numerical taxonomy. Nature, **197**, 716-717.
- ROGERS, D. J. and TANIMOTO, T. T. 1960. A computer program for classifying plants. Science, **132**, 1115-1118.
- SACCARDO, P. A. 1886. Sylloge fungorum. **4**, 1-807. Pavia.
- SNEATH, P. H. A. 1957. The application of computers to taxonomy. J. Gen. Microbiol. **17**, 201-226.
- 1962. The construction of taxonomic groups. Microbiol. Classification (12th Symp. Soc. Gen. Microbiol.), 289-332.
- SNEATH, P. H. A. and SOKAL, R. R. 1962. Numerical taxonomy. Nature, **193**, 855-860.
- SOKAL, R. R. and MICHENER, C. D. 1958. A statistical method for evaluating systematic relationships. Univ. Kansas Sci. Bull. **38**, 1409-1438.
- VUILLEMIN, P. 1910a. Matériaux pour une classification rationnelle des Fungi Imperfecti. Compt. Rend. Acad. Sci. **150**, 882-884.
- 1910b. Les conidiosporés. Bull. Sci. Soc. Nancy, Ser. III, **11** (2), 129-172.
- 1911. Différence fondamentale entre le genre *Monilia* et les genres *Scopulariopsis*, *Acmosporium*, et *Catenularia*. Bull. Soc. Mycol. France, **27**, 137-152.
- 1912. Sur une nouvelle espèce de *Tilachlidium* et les affinités de ce genre. Bull. Soc. Mycol. France, **28**, 113-130.