

**ON QUANTITATIVE STATISTICAL METHODS IN TAXONOMY;
SUBDIVISION OF A POLYMORPHOUS SPECIES:
PLANCHONELLA SANDWICENSIS (GRAY) PIERRE**

by

J. E. B O E K E

(Leiden).

"It is an attempt to carry on consciously and mathematically the same sort of process which with a good naturalist is subconscious and unmathematical".
Anderson and Abbe.

Introduction.

In his paper on Pacific Sapotaceae, Lam (4) dealt at some length with the extreme polymorphy in *Planchonella sandwicensis* (Gray) Pierre from the Hawaiian Islands. For systematical and morphological data concerning this species I may refer to that publication. In the following lines I will restrict myself to a discussion of some attempts towards a subdivision of herbarium material of the said species, using mainly quantitative, statistical methods.

I am very much indebted to Dr H. J. Lam for his stimulating support as well as for his continuous interest in the progress of my work, which have largely contributed to its eventual results. I am equally indebted to Dr J. B. D. Derksen for his critical reading of the mathematic part of the manuscript. However, full responsibility for any errors is, of course, my own.

In applying quantitative statistical methods I followed two different ways, which will be discussed in § 1 and § 2 respectively; § 3 contains an attempt to check my own results mathematically.

1. In order to subdivide the polymorphous material, I led my investigations, first of all, along the lines, suggested by Anderson and Whitaker (3, cf. also 1 and 2). The publication mentioned first is, so far as I know, the first one which treats taxonomical problems in the following way. By counting, measuring and calculating, each specimen of the material to be subdivided (into genera, species, subspecies, varieties or formae or, in general, into taxa of some kind) is given an "index". The indices, thus obtained, of course will not be all the same but will be varying within a certain range. Gaps occurring in this range, may point to taxonomic discontinuities, i. e. to natural delimitations of separate groups.

However, this interesting working hypothesis has been worked out in a way, I cannot consider unassailable. Anderson and Whitaker worked with dried material of *Uvularia grandiflora* Sm. and of *Uvularia perfoliata* L., two related, yet distinct North-American Liliaceae. Fifteen specimens (entire plants) were taken from either species and in each specimen certain measurements were made at a particular leaf, inserted at the main axis and bearing the lowest (sterile) shoot. After this, in each specimen the absolute values (a), some of their ratios (b) and the sum of the squares of these ratios (c) were compared with a, b and c of all other specimens respectively. It then appeared that in the two species the variation-ranges of the values of either a, b or c were overlapping each other, any discontinuity being, therefore, absent. However, the discontinuity demanded by the apparent specific differences showed itself — at any rate in the thirty specimens studied — if to value c was added: 1. the squared two-tenths part of the number of leaves on the main axis below the lowest branch, and 2. the squared tenth part of the difference between 10 and the number of leaves on this sterile branch.

Only a small part of the figures, leading to this rather astonishing result, have been published; rather astonishing, for it would be an exception to the rule that the sum of the squares of continuously varying numbers shows continuous variability itself. Most of the facts mentioned, however, are shown by graphic representation only. Therefore, although this cannot be proved, this "discontinuity between two species" is most probably a collective character of these thirty specimens only and may be called accidental; that is to say, as soon as the material would be enlarged with a sufficiently great number of specimens, both in what is regarded as *Uvularia grandiflora* and as *Uvularia perfoliata* a discontinuity of this kind would, most probably, be no longer apparent. Moreover, the method described seems just a little too artificial to allow trustworthy results regarding a subdivision of a living material.

In view of these considerations I would rather not apply to our material the method suggested by Anderson and Whitaker unmodified. Moreover, as our material consisted of twigs of the most different size, appearance and leafedness, it did not allow to gather the data in exactly the same way as was described above, i. e. in each specimen from any one particular leaf or any other special feature. For the sake of objectivity I had to start from average values in each specimen.

Every specimen in this material is represented by one or more twigs, bearing the most different numbers of leaves, flowers and fruits. Besides, no character of any part of a specimen, be it the shape of the leaves or the length of the pedicels etc., can be regarded to be of constant value in the twig it belongs to.

There is, on the contrary, a certain variability in all characters in the same specimen. In a given twig x, a special character p, therefore, is not sufficiently characterized by its mean value m_{px} ; its variability v_{px} should be mentioned as well. In a twig y this same feature is characterized by m_{py} and v_{py} . If we want to subdivide a given material on account of a character p, this character is useful only when the following conditions are fulfilled:

1. If two twigs, x and y , are parts of one and the same specimen, any difference between m_{px} and m_{py} should not surpass a certain value dependent on v_{px} and v_{py} ; if this upper limit is surpassed, the character chosen is of no use for a subdivision.

2. if x and y are hailing from two different specimens, belonging to one and the same taxonomic group (taxon), the value of m_{px} minus m_{py} may be somewhat larger than in the previous case, but still it has a distinct upper limit, depending on the variability of p within the taxon.

3. if x and y represent two different taxa, the difference mentioned must surpass this limit. A significant difference of this kind is a counterpart of the discontinuity mentioned above, but statistically it is better justified.

Moreover, a useful character p has to comply with the general condition that it must be possible to count or measure it in all specimens, since the specimens it is lacking in cannot be put in their proper places in the taxonomic system to be designed. For this reason, only characters of the stem and of the leaves are suitable, and those of flowers and fruits are to be left out of account, as in many specimens these are not extant. Therefore, the comparative research work, discussed in this paragraph, was based upon the leaves, the more so, since the astonishing diversity (cf. Lam, l. c., figs. 4—5) was most attractive in view of our purpose.

The first condition mentioned, to be fulfilled by a useful character p was that the difference $m_{px} - m_{py}$ should not surpass a certain amount, dependent on the values of v_{px} and v_{py} . As every textbook of mathematical statistics will show the mathematical basis of this thesis, I will not but recall here, that m_{px} and m_{py} are, with some reserve, considered to be representing material of an essentially different nature only, if their mutual difference is greater than its "standard error", $3\sigma (m_{px} - m_{py})$, a mathematical function of v_{px} and v_{py} and of the number of leaves, measured regarding the character p in the specimens x and y respectively.

Among the characters of the material at hand, I have been looking for such, as might be suitable for basing a subdivision upon. Possible items were the length of the petioles (a), the length of the lamina (b) (the length of the apex in acuminate leaves was not included), the distance from the base of the leaf to the point on the costa where the leaf is broadest (c), and the largest width of the lamina perpendicular to the costa (d).

None of the values of a, b, c or d complied with the condition mentioned above sub 1, for in some specimens not only fertile twigs with normal leaves were extant, but also sterile sucker shoots with much larger leaves, which, for the sake of objectivity, could not be left out of account. As, however, those larger leaves showed generally the same shape as the normal leaves, it seemed, a priori, not impossible that, for instance, the ratios between length and breadth should be similar or at least comparable. In view of this possibility, I choose six different ratios and calculated the average values for each twig of a set of six from the same specimen (*H. St. John 11584*). The results are given in the following table (I):

TABLE I.

specimen nr. \ Means of:	$\frac{b}{a}$	$\frac{c}{a}$	$\frac{d}{a}$
116a	3.0 ± 0.10	1.78 ± 0.040	2.07 ± 0.029
116b I	3.2 ± 0.08	1.70 ± 0.062	2.11 ± 0.066
116b II	2.6 ± 0.06	1.51 ± 0.033	1.71 ± 0.042
117a	2.9 ± 0.07	1.73 ± 0.039	1.87 ± 0.042
117b	3.2 ± 0.08	1.76 ± 0.066	1.85 ± 0.046
117c	3.0 ± 0.07	1.68 ± 0.053	1.93 ± 0.086
$m_{\max} - m_{\min}$	0.60	0.27	0.40
$3 \varepsilon (m_{\max} - m_{\min})$	0.31	0.15	0.24

specimen nr. \ Means of:	$\frac{b}{c}$	$\frac{b}{d}$	$\frac{c}{d}$
116a	1.72 ± 0.019	1.47 ± 0.062	0.86 ± 0.015
116b I	1.85 ± 0.052	1.47 ± 0.036	0.80 ± 0.014
116b II	1.70 ± 0.023	1.48 ± 0.019	0.87 ± 0.040
117a	1.66 ± 0.024	1.55 ± 0.027	0.93 ± 0.012
117b	1.79 ± 0.046	1.71 ± 0.030	0.96 ± 0.030
117c	1.74 ± 0.028	1.52 ± 0.025	0.89 ± 0.023
$m_{\max} - m_{\min}$	0.19	0.24	0.16
$3 \varepsilon (m_{\max} - m_{\min})$	0.17	0.14	0.10

Here again a, b, c and d are dimensions of leaf-parts as mentioned above. In each category of ratios the largest (m_{\max}) and the smallest (m_{\min}) average value have been printed in heavy type and the difference ($m_{\max} - m_{\min}$) has been recorded underneath. This shows that each of these differences is greater than the amount of $3 \varepsilon (m_{\max} - m_{\min})$. Thus, none of the ratios used is in compliance with the first condition mentioned above and is, therefore, of any use as a criterion for subdividing this polymorphous material.

Beside its shape there is yet another character of the leaf showing a considerable variability, viz. the angle between the midrib and the lateral nerves, which is varying between 40° and 90°. As the angles at the base

of the leaves are always greater than those near the apex, I choose to measure two particular angles, α and β , situated at $\frac{1}{3}$ and $\frac{2}{3}$ of the length of the costa from the leaf-base. In this way I measured α and β in the leaves of six twigs from one single specimen (*J. F. Rock 8684*) and calculated the average values of α , β and α/β for each of these twigs. The results may be seen from the following table (II):

TABLE II.

specimen nr.	Means of:	α	β	$\frac{\alpha}{\beta}$
129a		6°09 ± 1°31	4°225 ± 1°13	1.45 ± 0.033
129b		6°455 ± 0°54	5°10 ± 0°91	1.27 ± 0.027
129c		6°33 ± 1°19	4°41 ± 1°446	1.45 ± 0.025
129d		6°585 ± 1°07	4°79 ± 0°82	1.38 ± 0.030
129e		6°475 ± 0°73	4°585 ± 3°05	1.42 ± 0.039
131		6°185 ± 1°23	4°47 ± 1°45	1.40 ± 0.043
$m_{\max} - m_{\min}$		0°495	0°875	0.18
$3 \varepsilon (m_{\max} - m_{\min})$		0°507	0°435	0.13

In each category the largest (m_{\max}) and the smallest (m_{\min}) value has been printed in heavy type, the difference ($m_{\max} - m_{\min}$) being recorded underneath. Each of these differences is greater than the value of $3 \varepsilon (m_{\max} - m_{\min})$, except in α , but this case is not very striking, since the two values are almost the same. Thus, the values of the various angles measured, provides no suitable criterion either for a subdivision of the material.

After this second failure I stopped looking for a criterion of this kind and I made another attempt along a quite different line in order to solve, if possible, the problem of the subdivision of a polymorphous material on the basis of statistical methods.

2. My second attempt was based upon the following working hypothesis. If, in a polymorphous material, there are specimens with certain morphological characters in an extreme condition (in the following to be called *extremes*) and others, having these characters in an intermediate condition (to be called *intermediates*), the procentual number of the latter is showing to which degree the former, although being morphologically opposite, may be considered to belong together, in other words: in how far the phases of the characters in question may be considered to represent the various materializations of one and the same genotypical variation.

If the number of intermediates between two groups of extremes is small, the two groups may be considered distinct. If, on the other hand, the intermediates are numerous, there is no ground to base separate groups on the extremes. Once the groups being determined, each specimen may be inserted in its proper place or taxon on the basis of the characters used.

Our starting point for working out this scheme consisted of five pairs of opposite characters, viz.

The surface of the leaf-blade, being large (*a*) or small (*b*);

The shape of the leaf, being obovate-oblong (*c*) or acute (acuminate)-rhomboid (*d*);

the pedicels, being long (*e*) or short (*f*);

the sepals, being rounded (*g*) or acute (*h*);

the fruits, being rounded (*i*) or acute (*j*).

From each pair of characters, in the sequence mentioned, one is applicable to a specimen. In case of doubt both figures are to be used. Such a doubt may arise, both when the part concerned is missing and when it is in an intermediate condition¹).

In this way every specimen may be represented by a formula, e. g. a twig with large (*a*) round-shaped (*c*) leaves, long (*e*) pedicels and rounded (*g*) sepals, but acute (*j*) fruits by the formula *acegj*. Another specimen, showing the first four characters in the same condition but possessing no fruits, by *acegⁱ_j*. A third example, for the greater part equal to the preceding one, but leaving doubt as to the interpretation of the length of the pedicels, by *ac^e_fgⁱ_j*.

Every combination of characters without intermediate factors, i. e. all extremes, the formulae of which may be composed out of the five pairs of characters mentioned, have been inserted in the following table (III) and have been designated by a capital:

TABLE III.

A <i>acegi</i>	I <i>adegi</i>	Q <i>bdegi</i>	Y <i>bcegi</i>
B <i>acegj</i>	J <i>adegj</i>	R <i>bdegj</i>	Z <i>bcegj</i>
C <i>acehj</i>	K <i>adehj</i>	S <i>bdehj</i>	AA <i>bcehj</i>
D <i>acehi</i>	L <i>adehi</i>	T <i>bdehi</i>	BB <i>bcehi</i>
E <i>acfhi</i>	M <i>adfhi</i>	U <i>bdfhi</i>	CC <i>bcfhi</i>
F <i>acfhj</i>	N <i>adfhj</i>	V <i>bdfhj</i>	DD <i>bcfhj</i>
G <i>acfgj</i>	O <i>adfgj</i>	W <i>bdfgj</i>	EE <i>bcfgj</i>
H <i>acfgi</i>	P <i>adfgi</i>	X <i>bdfgi</i>	FF <i>bcfgi</i>

¹) Which condition may be called intermediate? After some time's survey of the entire material this question answered itself. Yet, I am afraid that some absurdities, showing themselves in the results of this investigation and mentioned later on were partly caused by an erroneous interpretation of this intermediate condition. If this investigation could have been repeated, slightly altered as to this point, I suppose these absurdities would disappear to a large extent.

According to this table, the examples given above agree with the combinations B, (A or B) and (A, B, G or H) respectively.

All combinations (formulae) actually represented in our material are mentioned in the following table (IV). In this table all of the 143¹⁾ specimens (some of the 177 numbered, afterwards proved to be duplicates) are divided into two categories, one comprising the extremes, the other containing the intermediates; the latter has been divided into groups of combinations with one, two, three or four intermediate factors respectively. Of each combination in this table the first item is its formula. Next follow one or more capitals designating this formula and the numbers of the specimens represented by this formula.

TABLE IV.

Extremes:

$acegi$ — A — 9, 11, 13, 14, 16, 20, 23, 25C, 27, 28, 29, 39, 41, 42, 43, 45, 49, 51, 52, 57, 59, 70, 75, 78, 79, 81, 83, 84, 85, 89, 93, 96, 97.	
$acphi$ — E — 104, 118, 119	$bdfhj$ — V — 155, 158, 167, 171
$acfhj$ — F — 129, 162	$adfgj$ — O — 102
$acfgj$ — G — 125	$bcegi$ — Y — 21
$acfgi$ — H — 99, 106, 127	$bcfgi$ — FF — 132, 134, 154
$adfhj$ — N — 144	

Intermediates (with one, two, three and four characters in an intermediate condition):

$aceg^i_j$ — (A or B) — 3*, 4*, 7*, 19*, 24*, 25A*, 37*, 60*, 61*, 69*, 74*, 88*, 92*, 152*, 10?, 30?, 46?, 62?, 63?, 64?, 65?, 73?, 112?	
ace^g_hi — (A or D) — 72?	a_bdfhi — (M or U) — 128?
ac^e_fgi — (A or H) — 31?, 34?, 66?, 122?	adf^i_hj — (M or N) — 145*, 146*, 161?
a^c_degi — (A or I) — 86?, 90?	a_bdfhj — (N or V) — 170?
a_bcegi — (A or Y) — 12?, 22?	b^c_dfhj — (V or DD) — 169?
$acfh^i_j$ — (E or F) — 115*, 116*, 124*	$b^c_d^e_hj$ — (S or AA) — 101?

! Doubtful regarding condition of one or more pairs of characters.

* Fruits wanting.

A, E, F, etc. Component of a combination also occurring as an "extreme".

M or **U** Neither component of this combination is found among the "extremes".

For further explanation see text.

¹⁾ This number is somewhat smaller than that mentioned in Lam's paper (160) as some specimens were not received on loan but after the present paper was secluded.

TABLE IV (continued).

$acfg_h^i$	— (E or H) — 130?	bc^e_fgi	— (Y or FF) — 150?
a^c_dfgi	— (H or P) — 105?, 120?	$bdfg_j^i$	— (W or X) — 107*, 139*
a_bcfgi	— (H or FF) — 147?	$bcfg_j^i$	— (EE or FF) — 137*
ad^e_fgj	— (J or O) — 98?	$bcfh_j^i$	— (CC or DD) — 166*
$aceg^i_{hj}$	— (A, B, C or D) — 1?, 176? 5*, 47*	bc^e_fgj	— (B, G, Z or EE) — 113?
ac^e_fgj	— (A, B, G or H) — 2*	$b^c_dfh_j^i$	— (U, V, CC or DD) — 164*, 165* 172*, 173*
$a^c_deg^i_j$	— (A, B, I or J) — 8*, 44*, 54*	$bc^e_fgh_j^i$	— (Z, AA, DD or EE) — 149?
$ac^e_fgh_i$	— (A, D, E or H) — 17?	$bc^e_fgj^i$	— (Y, Z, EE or FF) — 153?
$a^c_dfh_j^i$	— (E, F, M or N) — 163B*	$ad^e_fgj^i$	— (I, J, O or P) — 177?
$adf^g^i_{hj}$	— (M, N, O or P) — 126*?	$b^c_dfg_j^i$	— (W, X, EE or FF) — 123A?, 138? 141?, 100*
$acfg^i_{hj}$	— (E, F, G or H) — 135?		
$ac^e_fgh_j^i$	— (A, B, C, D, E, F, G or H) — 18†, 25B†, 26†, 58†, 68†, 77†		
$a^c_dfg^i_{hj}$	— (E, F, G, H, M, N, O or P) — 123B?		
$a^ce_dfg^i_j$	— (A, B, G, H, I, J, O or P) — 114*?		
$ad^e_fgh_j^i$	— (I, J, K, L, M, N, O or P) — 6†, 103†? 136†		
$ac^e_dfg^i_{hj}$	— (A, B, C, D, E, F, G, H, I, J, K, L, M, N, O or P) — 15†?		
$b^ce_dfg^i_{hj}$	— (Q, R, S, T, U, V, W, X, Y, Z, AA, BB, CC, DD, EE or FF) — 157†?		
$a^ce_dfg^i_{bcfh_j}$	— (A, B, C, D, E, F, G, H, Y, Z, AA, BB, CC, DD, EE or FF) — 143†?		
$ac^e_dfg^i_{bdfhj}$	— (E, F, G, H, M, N, O, P, U, V, W, X, CC, DD, EE or FF) — 140?		

‡ Doubtful regarding condition of one or more pairs of characters.

* Fruits wanting.

† Sterile specimens (flowers and fruits wanting).

For further explanation see p. 53.

This table shows that the combinations A and (A or B) are numerically best represented and this concentration of specimens is still corroborated by combining both groups of specimens, since

1° combination (A or B) means, properly speaking, a transition between A and B. However, as A is representing a great number of specimens and B none at all, this condition can hardly be called an intermediate one, but can better be considered lying within the variation-range of combination A.

2° in many specimens, possessing both older and younger fruits, the latter show some degree of acuteness (in the formula: $\begin{smallmatrix} i \\ j \end{smallmatrix}$), whilst the former, as a matter of course, whether rounded (*i*) or acute (*j*), are perfectly decisive. As the combination (A or B) meaning doubt as to the condition of the fruits (in the nrs. 10?, 30?, 46? etc.), often showed this $\begin{smallmatrix} i \\ j \end{smallmatrix}$ character in small fruits, it is not impossible, that these fruits would have got a rounded shape (character *i*), had they been allowed to ripen (cf. Lam, l. c., pp. 28 and 30).

This correction, bringing together all specimens characterized by $acegi$ and $aceg\begin{smallmatrix} i \\ j \end{smallmatrix}$, gives rise to a sort of nucleus, consisting of $33 + 23 = 56$ specimens, i. e. 39% of the entire material. All other combinations of characters are more or less distinctly connected with this nucleus. On the other hand, combination V ($bdfhj$), representing 4 specimens, has nothing in common with it, being directly opposite to it. Less different of it are, of course, those combinations, that have in common all characters but one.

Of this latter kind five different combinations, B, D, H, I and Y, may be indicated. In the same way five other combinations can be found, each of them different from B only as to one single character, viz. the combinations A, C, G, J and Z. The combinations that have most characters in common with combination D are C, A, E, L and BB. In general, each combination out of those, included in table III, has five others differing from it as to one of its characters only. Symbolizing each of these "single differences" by a distance of constant length all 32 combinations could be arranged in a three-dimensional diagram. However, if we consider only those combinations occurring as extremes, a two-dimensional diagram will suffice. In the scheme, shown in *Fig. 1* (p. 56), O and G, G and F, F and E, etc., each differing from the other as to one single character, show the same mutual distance, both horizontally and vertically.

In *Fig. 1* an application is demonstrated of our working hypothesis, already mentioned on page 51. Considering two groups of extremes differing only as to one single pair of opposite characters and, besides, a group of intermediates, only being in an intermediate condition as to the opposite characters in question, the morphological difference between the two groups of extremes mentioned may be called insignificant, if the number of intermediates mentioned is relatively large; it may, on the other hand, be called significant, if the said number is relatively small

or if it is nought. In the former case the two groups belong together from a morphological point of view.

Let me take as an example the groups of extremes characterized as A and as H, differing as to the length of the pedicels only. The first group, consisting of 33 specimens, has long, the second group, comprising 3 specimens, possesses short pedicels. In addition, a group of

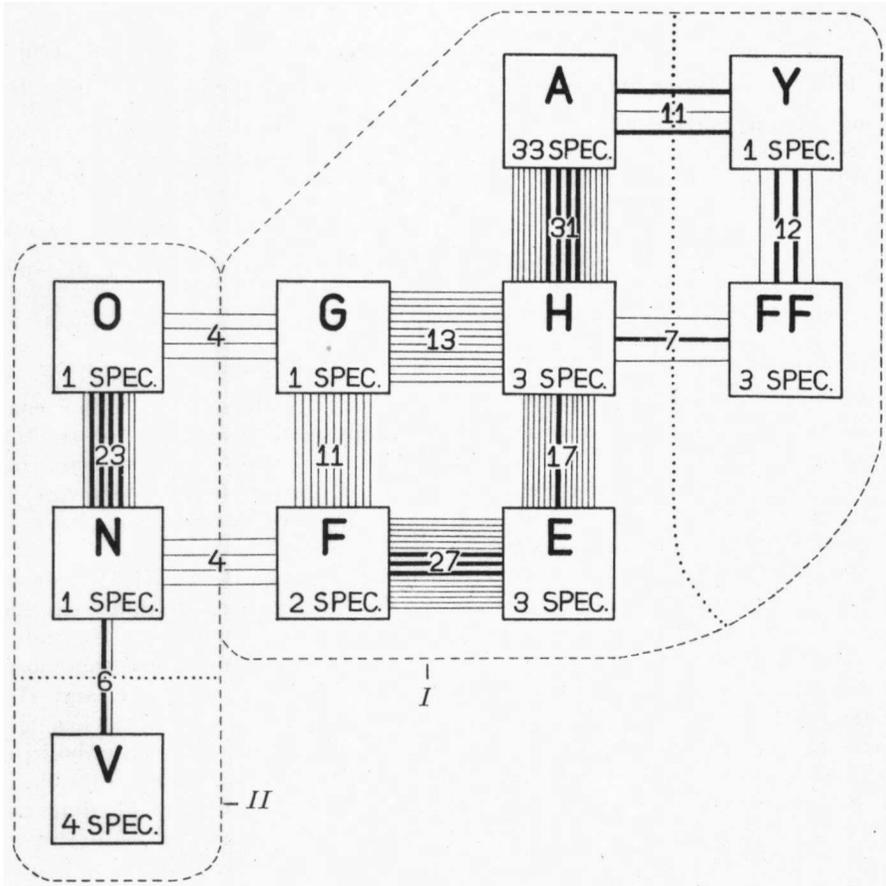


Fig. 1 — Arrangement of 10 combinations (extremes) according to their natural relations as shown by the number of their direct (heavy lines) and indirect (thin lines) intermediates, which connect them.

4 specimens is characterized by (A or H), the length of their pedicels being doubtful. As this number of four specimens is relatively large (group H including only three), the different length of the pedicels in A and H is not to be called a significant point of distinction.

These four specimens, establishing, in a way, a bridge between A and H, are indicated by four heavy lines in the scheme mentioned. However, as a matter of fact, there are many other specimens, forming some

sort of connection between A and H, viz. those, characterized by (A, B, G or H), by (A, D, E or H), by (A, B, C, D, E, F or H) etc., all including A and H as components in their combinations of characters. Of course, each of these does not indicate as strong a connection between A and H as the specimens of (A or H). Therefore, in the scheme the graphic representation of each specimen of this kind is by a thin line between A and H. But not as a thin line between A and H only. The other components of the said combinations of characters had also to be connected by thin lines, if possible, i. e. if occurring in our scheme. A specimen, for instance, characterized by (A, B, G or H) is indicated by one thin line between A and H and another between G and H, not, however, by a thin line between A and G, two combinations which are not to be connected directly, since they are different in more than one category of characters.

Out of this procedure all heavy and thin lines resulted, which are shown in *Fig. 1*. A simple calculation, based upon the principle of the "weighted mean", but not specified in this notice, shows, that one heavy line symbolizes as strong a "connection" between two combinations (the capitals in the scheme) as five thin lines. On this calculation the totals between the combinations are based. Considering this, it seems acceptable to divide the scheme into at least two parts, the first consisting of O, N and V, the second comprising E, F, G, H, A, Y and FF. Though it is evident, that a strict disjunction is not in the nature of this material, the mutual connection (shown in the number of connecting lines between the capitals) within either part is undoubtedly more evident than the connection between the two parts. In other words, the "intrapartial" connection is stronger than the "interpartial" one. Other, though weaker delimitations lie between A—H and Y—FF, and also between N and V.

Between the two main groups, to be called *II* (including the capitals O, N and V) and *I* (including the other seven combinations of characters) the most striking difference concerns the shape of the leaf, that is the pair of characters $\begin{smallmatrix} c \\ d \end{smallmatrix}$, since the character *c* (leaves obovate-oblong) is extant in group *I* without any exception, and is entirely lacking in group *II*. Just as in the arrangement, composed by Lam (4, fig. 4), it is shown here, that the shape of the leaf is a valuable criterion for a subdivision in *Planchonella sandwicensis*¹⁾, the other characters mentioned giving rise to homologous variations within either group.

This being stated, a number of questions arise. Are we right in dividing all 143 specimens (not to mention other specimens, not represented in this material) either into group *I* or into group *II*, on account of the shape of the leaves only? And if so, is this sufficiently justified by the mutual relations of only 40 % of the material at hand, the number of extremes comprised in our scheme being not more than 52 (out of 143)? Are the specimens, which are intermediate as to the shape of their leaves, i. e. characterized by $\begin{smallmatrix} c \\ d \end{smallmatrix}$, to be considered also actually intermediate between

¹⁾ This statistical result needs, of course, not contradict Lam's choice of the length of the pedicels for a practical subdivision.

the groups? Questions like these I will reply to with the aid of a statistical check, to be described in the next paragraph.

3. The 143 specimens of the material may be grouped, regarding the size and the shape of the leaves, the length of the pedicels and the shape of the sepals and fruits, in 50 different combinations of characters, and each specimen is, concerning the characters mentioned, completely determined by one of these combinations.

As yet, we restricted ourselves to a distribution of these combinations to the groups *I* and *II*, only regarding the condition of the second pair of characters, i. e. the shape of the leaf (*c* and *d*). Each *c* has been put into group *I*, each *d* into group *II*, and each $\frac{c}{d}$ is considered an intermediate. In this way all other characters are considered homologous variations and, therefore, of no importance for a subdivision. If this procedure is correct, these other characters will be equally distributed to the groups *I* and *II*, i. e. group *I* will, proportionally speaking, comprise as many specimens with one of these characters (for instance *b*) as group *II*.

This means, that, if in group *I* the numbers of specimens, showing character $\frac{a}{b}$ or *b* (*a* and *b* are representing the size of the leaf) are, for instance, 50, 10 and 30 respectively and in group *II* these numbers are 20, 4 and 12 (that is in the same proportion), these characters are to be considered homologous variations. If, however, the proportions in the two groups are not the same, for instance 50, 10, 30 and 6, 4, 26 respectively, the distribution of these characters cannot be considered a homologous variation, but is significantly different within the two groups and is, therefore, of real importance for a subdivision.

Though this conclusion is admissible only in case many specimens are available, I checked its consequences in a purely quantitative way in the 143 specimens at hand. It must be admitted that the results of this investigation will be unfavourably biased by this relatively small number; I hope, therefore, that this method may some time be applied to a more extensive material.

Group *I* includes 4 times the combination V (*b d f h j*), once the combination N (*a d f h j*) and once the combination O (*a d f g j*). This means that character *a*, being present in N and O only, is represented in two specimens of this group, i. e. (the total number of specimens in group *II* being six) in 33.3%; in the same way character *b* shows a percentage of 66.7. Character *c* is, naturally, extant in 0% and character *d* in 100% of the specimens of this group, and the characters *e* and *f* show the same figures. For *g*, extant in combination O (one single specimen) and for *h* in V and N (including 5 specimens together), they are 16.7% and 83.3% respectively. Finally, the characters *i* and *j* are extant in 0% and 100% of the specimens in group *II* respectively.

Summarizing the relative significance of each of the characters, after having been calculated in this way, it is found that:

$$II = \begin{matrix} 33.3 \% a & 0 \% c & 0 \% e & 16.7 \% g & 0 \% i \\ 66.7 \% b & 100 \% d & 100 \% f & 83.3 \% h & 100 \% j \end{matrix}$$

The figures concerning group *I* may be calculated in the same way, leading to the following results:

$$I = \begin{matrix} 91.3 \% a & 100 \% c & 73.9 \% e & 89.1 \% g & 93.5 \% i \\ 8.7 \% b & 0 \% d & 26.1 \% f & 10.9 \% h & 6.5 \% j \end{matrix}$$

I think, that a further accumulation of figures may be interrupted here and that the introduction of a simple cipher-key will open the door to the well-known ground of systematics. The formulae given may then be interpreted in the following way:

Group *I*: leaves chiefly (91.3 %) large, shape of the leaves always (100 %) obovate-oblong,
pedicels mostly (73.9 %) long,
sepals mostly (89.1 %) rounded,
fruits almost always (93.5 %) rounded.

Group *II*: leaves chiefly (66.7 %) small, always (100 %) of acute (acuminate)-rhomboid shape,
pedicels always (100 %) short,
shape of the sepals mostly (83.3 %) acute,
shape of the fruits always (100 %) acute.

In these terms the same is expressed as in the formulae given above, namely a "synthetic description" of the groups *I* and *II*. As the distribution of characters not concerning the shape of the leaf is by no means proportionally the same within these groups, it is obvious that the four characters in question are not to be considered homologous variations; they are of some more taxonomic importance.

Now every specimen, represented by five characters as shown above, may be compared with either of these "synthetic" groups *I* and *II*. With a purely quantitative method any given specimen of *Planchonella sandwicensis* may be stated to have more or less in common either with group *I* or with group *II*.

This method of allocation may be demonstrated with a specimen, characterized by $acf \begin{smallmatrix} g \\ h \\ j \end{smallmatrix}$, i. e. with the combination (E, F, G or H). If we could dispose of 100 specimens of the same type, the characters *a*, *e* and *f* would be encountered 100 times, and the characters *b*, *d* and *e* never. It may, therefore, be said, that the leaf is $\begin{matrix} 100 \% a \\ 0 \% b \end{matrix}$, the leafshape is $\begin{matrix} 100 \% c \\ 0 \% d \end{matrix}$ and the pedicels are $\begin{matrix} 0 \% e \\ 100 \% f \end{matrix}$. But as both the shape of the sepals and of the fruits are subject to doubt, in one half of cases we should be inclined to call them rounded (*g* and *i* respectively) and in the other half acute (*h* and *j* respectively). Thus the formula $\begin{matrix} g \\ h \\ j \end{matrix}$ may also be written as $\begin{matrix} 50 \% g & 50 \% i \\ 50 \% h & 50 \% j \end{matrix}$. The entire formula may then be written thus: $\begin{matrix} 100 \% a & 100 \% c & 0 \% e & 50 \% g & 50 \% i \\ 0 \% b & 0 \% d & 100 \% f & 50 \% h & 50 \% j \end{matrix}$ in which way

even one single specimen is as perfectly characterized, as 100 specimens would be.

This implies, that if one specimen shows a character in an intermediate condition, for instance $\begin{matrix} i \\ j \end{matrix}$, 50 out of 100 of the same type are to be considered i and 50 j . In the case of the "synthetic groups", however, some characters are not in an intermediate condition, but they are not entirely pure either. As the leaf in group I , for instance, is mostly large (91.3 %), this means that 913 out of 1000 specimens of group I may be expected to be large-leaved and the other 87 small-leaved. In 1000 specimens of the combination (E, F, G or H), however, all thousand have large leaves. Of either group of 1000 specimens, therefore, 913 possess the same kind of leaves, viz. large ones, which means that there is a conformity of 91.3 % of the two groups of 1000 specimens.

Considering the shape of the leaf in specimens of the combination (E, F, G or H) and of group I the conformity is total (100 %), for in 1000 specimens of either type all thousand have rounded leaves.

As to the pedicels, out of 1000 specimens of group I , 739 are to be considered long-pedicellate and the other 261 short-pedicellate. In 1000 specimens of the type $a c f \begin{matrix} g \\ h \\ j \end{matrix} i$, however, all have short pedicels. The conformity of the two groups of 1000 specimens is, therefore, only 26.1 %, since 261 specimens of either group show the same character b .

Considering the shape of the sepals in 1000 specimens of both types mentioned, the specimens with rounded and with acute sepals would be 891 and 109 in number respectively in group I , and 500 and 500 respectively in the type of the combination (E, F, G or H). The conformity of both groups, therefore, concerns firstly 109 specimens showing the character g and secondly 500 specimens showing the character h , or in total 609 specimens or 60.9 %.

As to the fruits, in a material of 1000 specimens of group I , 935 specimens would possess more or less rounded fruits and 65 specimens would have more or less acute ones. Of 1000 specimens of the combination (E, F, G or H), however, fruits with rounded and with acute tips would be found in equal numbers (500 of each category); and 500 specimens with rounded fruits are also equally extant in the first-named group; but since only 65 specimens in this group have acute fruits, the conformity of the two groups is 565 specimens or 56.5 %.

The conformity with reference to five different pairs of characters, however, may be stated not only in groups of 1000 specimens of either kind, but is also existing in their single representatives. In fact, the groups of 1000 specimens were only mentioned in order to elucidate our explanation.

The foregoing considerations show that the conformity of a specimen, characterized by $a c f \begin{matrix} g \\ h \\ j \end{matrix} i$ and a specimen of group I was found to be

- 91.3 % concerning the leaf,
- 100 % concerning the shape of the leaf,
- 26.1 % concerning the length of the pedicels,

60.9 % concerning the shape of the sepals, and

56.5 % concerning the shape of the fruit.

The "average procentual conformity" (A.P.C.) is the average of these five amounts, viz. 67 %. This means that a specimen characterized by $acfg_{hj}^g$ may be considered to belong to group *I*, although the A.P.C. is not greater than $\frac{2}{3}$ of the total similarity.

In the same way its A.P.C. may be calculated with regard to group *II*. It then proves to be 50 %. This means that a specimen characterized by $acfg_{hj}^g$ cannot be considered to belong to group *II*, for although it has exactly one half of its characters in common with that group (its A.P.C. being 50 %), the A.P.C. with group *I* is greater (67 %).

Now the A.P.C. with group *I* and group *II* has been calculated for all specimens in the material at hand. In Table V these figures have been given in the fourth and fifth columns respectively. The sixth column indicates whether the specimens have ultimately been inserted in group *I* or in group *II*. The second column contains the formulae of the specimens (combinations of characters), the third one the number of specimens representing these formulae in the material.

It appears from this table that in the majority of the cases, in which the character *c* is occurring, the A.P.C. with group *I* is greater than that with group *II*, and that, whenever the character *d* is extant the *II* percentage is greater. Predomination of the A.P.C. with *II*, in case the

TABLE V.

Conformity of 50 combinations of characters with regard to group *I* and group *II*.

Nr.	Formula	Number of specimens	Average Procentual Conformity (A. P. C.)		Group
			with <i>I</i>	with <i>II</i>	
	Extremes				
1	<i>acegi</i>	33	89.6	10.0	<i>I</i>
2	<i>acfgi</i>	3	80.0	30.0	<i>I</i>
3	<i>acfhi</i>	3	64.4	43.3	<i>I</i>
4	<i>acfgj</i>	1	62.6	50.0	<i>I</i>
5	<i>acfhj</i>	2	47.0	63.3	<i>II</i>
6	<i>adfhj</i>	1	27.0	83.3	<i>II</i>
7	<i>adfgj</i>	1	42.6	70.0	<i>II</i>
8	<i>bdfhj</i>	4	10.4	90.0	<i>II</i>
9	<i>bcfgi</i>	3	63.5	36.7	<i>I</i>
10	<i>bcegi</i>	1	73.1	16.7	<i>I</i>

TABLE V (continued).

Nr.	Formula	Number of specimens	Average Procentual Conformity (A. P. C.)		Group
			with I	with II	
	Single Intermediates				
11	$aceg_j^i$	23	82.2	20.0	I
12	$ace_h^g i$	1	83.9	20.0	I
13	$ac_f^e g i$	4	90.0	20.0	I
14	$a_d^c e g i$	2	79.6	20.0	I
15	$a_b^c e g i$	2	83.0	20.0	I
16	$acfh_j^i$	3	57.0	53.3	I?
17	$acf_h^g i$	1	74.4	40.0	I
18	$a_d^c f g i$	2	70.0	40.0	I
19	$ad_f^e g j$	1	52.6	60.0	II?
20	$adf_h^i j$	3	37.0	73.3	II
21	$a_b^a dfhj$	1	20.4	93.3	II
22	$a_b^a dfhi$	1	38.1	73.3	II
23	$a_b^a c f g i$	1	73.5	40.0	I
24	$b_d^c e h j$	1	30.0	60.0	II
25	$bc_f^e g i$	1	73.5	26.7	I
26	$b_d^c f h j$	1	20.4	80.0	II
27	$bdfg_j^i$	2	36.1	66.7	II
28	$bcf_g_j^i$	1	56.1	46.7	I
29	$bcf_h_j^i$	1	40.4	60.0	II
	Complex Intermediates				
30	$ace_h^g i_j$	4	76.5	30.0	I
31	$ac_f^e g_j^i$	1	82.6	30.0	I

TABLE V (continued).

Nr.	Formula	Number of specimens	Average Percentual Conformity (A. P. C.)		Group
			with I	with II	
32	$a_d^c e g_j^i$	3	72.2	30.0	I
33	$a c_f^e g h_i$	1	84.4	30.0	I
34	$a_d^c f h_j^i$	1	47.0	63.3	II
35	$a c f_g^e h_j^i$	1	67.0	50.0	I
36	$b_d^c f h_j^i$	4	30.4	70.0	II
37	$a^c e f g_j^i$	1	50.4	50.0	?
38	$b c_f^e g h_j^i$	1	50.4	56.7	II?
39	$b c_f^e g_j^i$	1	66.1	36.7	I
40	$a d_f^e g_j^i$	1	62.6	50.0	I
41	$b_d^c f g_j^i$	4	46.1	56.7	II
42	$a c_f^e g h_j^i$	6	77.0	40.0	I
43	$a_d^c f g_j^i h_j^i$	1	57.0	60.0	II?
44	$a^c e d f g_j^i$	1	72.6	40.0	I
45	$a d_f^e g_j^i h_j^i$	3	57.0	60.0	II?
46	$a c e g_d^i f h_j^i$	1	67.0	50.0	I
47	$b_d^c e g_i f h_j^i$	1	50.4	56.7	II?
48	$a c e g_i f h_j^i$	1	70.4	50.0	I
49	$a c f_g^e h_j^i$	1	60.4	70.0	II
50	$a d f_g^e h_j^i$	1	47.0	70.0	II

character *c* is extant is found only in the nrs. 5 (2 specimens), 29 (1 specimen), and 38 (1 specimen). Predomination of the A. P. C., in case the character *d* is extant, is found only in nr. 40 (1 specimen). As out of 101 specimens with the character *c* only 4 specimens (4%) have less conformity with group I than with group II and as out of

19 specimens with the character d only 1 specimen (this cannot be said to be 5 %, as the number concerned is too small) show less conformity with group *II* than with group *I*, it may be concluded that, on the whole, a subdivision according to the shape of the leaf is justified by a quantitative check. Only in cases of doubt as to the shape of the leaf, this quantitative check does not justify the allotment of an intermediate position between the groups, for, firstly, the 23 cases of $\overset{c}{d}$ include only 9 % (two specimens) showing nearly equivalent *I* and *II* conformity and, secondly, these two specimens (nrs. 43 and 47) form only 19 % of the total number of specimens with equivalent *I* and *II* conformity. As to the subdividing criterion used by Lam (4, fig. 5) in his demonstrative arrangement of specimens of *Planchonella sandwicensis*, on account of the length of the pedicels, their condition $\overset{e}{f}$ is found only seven times among the eleven specimens with equivalent *I* and *II* conformity and out of (in total) 25 cases of $\overset{e}{f}$ only 28 % (7 specimens), show equivalent *I* and *II* conformity.

In conclusion, it may be stated that, on the whole, this quantitative check proves fairly well that the shape of the leaves is a useful criterion for subdividing *Planchonella sandwicensis*, though it is, evidently, not suitable for locating the cases in the demarcation zone, i. e. the specimens to be placed between the two groups (subspecies or varieties). But this would undoubtedly be to high a demand for a mathematical method applied to living material.

Summary.

An attempt has been made to subdivide a very polymorphous plant species by means of a quantitative statistical method. This method has been based upon the following working hypothesis: 1° as some morphological characters of the material, concerning e. g. the shape of the leaves, the length of the pedicels etc., show an extremely great variation, each of these characters in every specimen at hand may be stated to be in one of three (one intermediate and two extreme) conditions; 2° if a character happens to be in an intermediate condition in a relatively great number of specimens the difference between the extreme conditions of that character may be considered insignificant from a taxonomical point of view; 3° the fewer characters of two or more specimens are differing significantly (in the way mentioned), the more reason there is to consider those specimens to belong together; and, on the contrary, the more numerous the significant differences are, the more reason there is to distribute the specimens to two (or more) different groups.

On the basis of this working hypothesis the material at hand, consisting of 143 specimens (all considered to belong to the Sapotaceous *Planchonella sandwicensis*, which was discussed in a paper by Lam); could be subdivided into two different groups. Five characters were chosen, each allowing to state one intermediate and two opposite extreme conditions. Of

these the shape of the leaf proved to be a most important criterion for a subdivision. After this had been stated, a purely quantitative check was made which largely endorsed the result.

Literature.

1. ANDERSON, E. and ABBE, E. C., A quantitative comparison of specific and generic differences in the Betulaceae — *Journ. Arn. Arb.* 15, 1934, 43—49.
2. ANDERSON, E. and HUBRIGT, L., Phylogenetic relationships, as deduced from a study of leaf-variation — *Bot. Gaz.* 100, 1938, 312—323.
3. ANDERSON, E. and WHITAKER, TH. W., Speciation in *Uvularia* — *Journ. Arn. Arb.* 15, 1934, 28—42.
4. LAM, H. J., A tentative list of wild Pacific Sapotaceae, except those from New Caledonia — *Blumea* 5, 1942, 14—30.