

CHARACTER ANALYSIS IN THE *BANISTERIOPSIS CAMPESTRIS* COMPLEX (MALPIGHIACEAE), USING SPATIAL AUTO-CORRELATION

G. F. Estabrook and B. Gates¹

Summary

The choice, construction, and evaluation of characters is central to systematic botany. Yet, how a systematist does this is rarely revealed or even discussed in his published works. Although the method suggested here is quantitative in nature, it does not depend either conceptually or computationally on numerical clustering. Indeed, it is our intent to provide a tool that can be used by a non-numerical taxonomist, prior to reaching any taxonomic decisions, to choose, construct, and evaluate potentially useful characters that will work for taxa recognition, without the need for arbitrary or subjective numerical clustering techniques. Here we will exemplify an idea for evaluating characters, by presenting first the basic idea, then the specimens and characters that constitute our example, next a discussion of the choice, construction and evaluation of these characters in a traditional context without consideration of spatial auto-correlation character analysis, and finally we present the results of spatial auto-correlation character analysis and discuss them in the light of the traditional assessment of characters. Our purpose in offering this method of character analysis is *not* to enable the worker to stop thinking about characters, but to enable him to think about them even more thoroughly by revealing their pattern of variation in phenetic space.

Basic Idea

Most workers would agree that taxa are based on groups of characters that tend to go together, i.e. are correlated. Methods, such as the product-moment correlation coefficient (r), for quantifying the amount of correlation of two numerically measured characters, have been extant for many years (Pearson, 1936). Many ordination techniques such as principal component analysis (Anderson, 1958) are based on this coefficient. A method for quantifying the dependency of two *qualitative* taxonomic characters was suggested by Estabrook (1967) and has been exemplified by Hawksworth et al. (1968), Bisby (1970), and Baum (1974). The concept of preferring those characters that are best correlated with all the characters under consideration is discussed by these authors and found to be consistent with traditional results.

The concept of spatial auto-correlation, discussed by geographers, Cliff and Ord (1973), and recently applied to biogeography by Sokal and Oden (1978), can serve us as a means of evaluating quantitative characters. If the entities being measured (collections) are arranged in space, then a basis for measurement (character) is positively spatially auto-correlated if there is a tendency for entities close together to have similar measurements, and for entities far apart to have dissimilar measurements. Actually all that is needed to calculate a measure of spatial auto-correlation is to know how far apart are any two entities. There are many procedures for quantifying distance or dissimilarity of two collections based on quantitative characters (Sneath and Sokal, 1973). In order to apply spatial auto-correlation in the evaluation of a quantitative taxonomic character: first calculate a measure of dissimilarity or distance for every pair of collections, based on all the remaining char-

¹ Division of Biological Sciences and the Herbarium, the University of Michigan, Ann Arbor, MI 48109, U.S.A.

acters; then use this distance to calculate the spatial auto-correlation of this character. A character whose values differ slightly for similar collections, but differ greatly for dissimilar collections will have a high value for spatial auto-correlation. It is suggested that these characters might have a high taxonomic value.

We believe that spatial auto-correlation character analysis may be most useful near the species level in the context of a species complex (such as that presented in example here), or in a subspecific study. Because this method is purely phenetic in its approach and makes no consideration of historical evolutionary processes, it seems likely that a character analysis method such as that reviewed by Estabrook (1983a, b) based explicitly on such considerations would be more appropriate for studying relationships above the species level.

The Formulae

We will use the capital letters K and L to represent characters; small letters *a*, *b*, *c*, to represent collections; and the composite symbol K(*a*) to represent the measurement of *a* determined by the measuring criterion (character) K. It is helpful to remember that K(*a*), K(*b*), L(*a*), etc., all stand for numbers, and thus they can be used in arithmetic expressions. For example, if K were average leaf length (Table 2), and *a* were A369 (Table 1), then K(*a*) = 6.0 cm (Table 3).

The measures of difference that we used in the example to follow are chosen from the Minkowski family of difference measures. We use the composite symbol d(*a*,*b*) to represent the difference (distance) between collection *a* and collection *b*. For use in evaluating character L, it is calculated as follows:

$$d(a,b) = \left(\sum_{K \neq L} |K(a) - K(b)|^z \right)^{1/z} \quad (1)$$

This is the Zth root of the sum over all characters K that are not character L, of the Zth power of the difference, according to character K, between collection *a* and collection *b*.

This formulation may seem complicated, but by choosing different values for Z, many familiar distance measures result. When Z = 1, the difference between collection *a* and collection *b* is simply the sum of their difference for each character. When Z = 2, the difference between collections is the distance that would actually be measured with a ruler if they were placed in space at positions with coordinates equal to their character values. This is often called ordinary, or Euclidean, distance. When Z is near 0, the Zth power of the difference between *a* and *b*, according to character K, is near 1, except when *a* and *b* are the same (or extremely close) in which case even the Zth power of this difference will still be near zero. Thus, when Z is near 0, the distance between collection *a* and collection *b* resembles a simple count of the number of characters for which *a* and *b* are different. High values of Z result in a distance measure that reflects the difference according to that one character for which *a* and *b* are most different.

We wish to calculate a measure of distance between pairs of collections, that is independent (computationally) from the character, L, that we wish to evaluate using spatial auto-correlation. Thus, in computing distance, the sum is always taken over all characters different from character L.

Now a measure of spatial auto-correlation can be calculated for L as follows:

$$G(L) = \sum_{a \neq b} d(a,b)^X |L(a) - L(b)|^Y \quad (2)$$

G(L) is the goodness of character L. When it assumes a high value, L may be considered a good character for taxonomic purposes. To express the desire for good characters to give very different values to very distant collections, choose a value of X larger than 1 (X = 2 or 3); but to express the desire for good characters to give very similar values to very close

Table 1. Collections used.

Code used	Collector	Number	Locality	State in Brazil
<i>Banisteriopsis campestris</i> (Adr. Jussieu) Little				
I 270	Irwin et al.	27034	Serra do Cabral,	MG
I 275	Irwin et al.	27506	vicinity Diamantina,	MG
H 315	Hatschbach	31519	Jaboticatubas,	MG
I 203	Irwin et al.	20354	Serra do Cipo,	MG
I 146	Irwin et al.	14628	vicinity Barreiras,	BA
A 369	Anderson et al.	36906	Espigao Mestre,	BA
A 673	Anderson	6738	Chapada dos Veadeiros,	GO
I 265	Irwin et al.	26562	Lagoa Paranoa,	DF
A 790	Anderson	7905	Serra Geral do Parana,	GO
A 798	Anderson	7980	Serra dos Cristais,	GO
D 149	Dusen	14991	Jaguariaiva,	PR
D 256	Dusen	256a	Villa Velha,	PR
M 234	Mimura	234	vicinity Sao Jose dos Campos,	SP
A 125	Anderson	11255	Rio Verde,	MT
A 120	Anderson	11204	Bataguacu,	MT
G 344	Gates	344	vicinity Diamantina,	MG
<i>Banisteriopsis angustifolia</i> (Adr. Jussieu) Gates				
G 341	Gates	341	vicinity Minas Novas,	MG
G 346	Gates	346	vicinity Diamantina,	MG
G 348	Gates	348	vicinity Diamantina,	MG
G 367	Gates	367	vicinity Diamantina,	MG
A 844	Anderson	8445	vicinity Diamantina,	MG
A 875	Anderson	8759	Rio Manso,	MG
I 310	Irwin et al.	31081	vicinity Seabra,	BA
<i>Banisteriopsis cipoensis</i> Gates				
G 386	Gates	386	Serra do Cipo,	MG
E 110	Eiten & Eiten	11086	Serra do Cipo,	MG
<i>Banisteriopsis arborea</i> Gates				
I 227	Irwin et al.	22747	vicinity Diamantina,	MG
I 279	Irwin et al.	27954	vicinity Diamantina,	MG
A 897	Anderson	8971	near Jequití,	MG
<i>Banisteriopsis andersonii</i> Gates				
A 838	Anderson	8383	vicinity Diamantina,	MG
A 850	Anderson	8500	vicinity Diamantina,	MG
G 351	Gates	351	vicinity Diamantina,	MG

collections, choose a value of X smaller than 1 ($X = 0.5$ or 0.2). Similarly, if you feel that it is most important for collections that are very different, according to character L, to be very distant, choose a value for Y larger than 1 ($Y = 2$ or 3); but if you feel that it is most important for collections that are very similar, according to character L, to be very close, choose a value for Y less than 1 ($Y = 0.5$ or 0.2).

Considerations for large X may seem the same as considerations for small X. Similarly, considerations for large Y may seem the same as considerations for small Y. Further, considerations for large or small X may seem the same as considerations for large or small Y. But they are not. Our example will show, that these parameter values result in different

Table 2. Characters.

-
- I. Habit: height and degree of branching of the stem
1.0 = shrublet to 4 dm, unbranched
2.0 = subshrub or shrub 0.5 to 1.5 m, unbranched
3.0 = shrub to 3 m, much branched, with main stem branching near base
4.0 = small tree to 4 m, main stem unbranched for about 1 m from base, much branched distally
- II. Petiole length in mm
Range of petiole length of leaves not associated with the inflorescence
1.0 = 1–3 mm
2.0 = 2–5 mm
3.0 = 3–6 mm
4.0 = 4–8 mm
- III. Position of the leaf glands
1.0 = at the base of the lamina or on the petiole
2.0 = just above the base of the lamina beside the midrib, sometimes with a second pair of glands further from the base
3.0 = 1 or 2 pairs of glands more than 5 mm from the base of the lamina, on the lamina beside the midrib, and less than $\frac{1}{4}$ of the lamina length from the base
4.0 = near the leaf margin, $\frac{1}{4}$ to $\frac{1}{2}$ of the lamina length from the base
- IV. Pubescence of the abaxial leaf surface
1.0 = almost glabrous except for the midrib and some of the lateral veins
2.0 = loosely tomentose, the hairs with more or less flexuous arms and the leaf surface visible to some extent between the hairs
3.0 = densely tomentose, at least the outermost hairs with straight arms arranged more or less parallel and with the leaf surface not visible between the hairs
- V. Pubescence of the adaxial leaf surface
1.0 = short sessile appressed hairs
2.0 = short-stalked hairs with short trabeculae, diverging from the leaf surface
3.0 = short-stalked hairs with long trabeculae, more or less diverging from the leaf surface
4.0 = long-stalked hairs with long trabeculae
- VI. Leaf angle
1.0 = leaves diverging widely from stem, more or less at 90° to incident light
2.0 = leaves more or less erect, parallel to incident light
- VII. Average leaf length in cm
Average length of midvein from base of lamina to tip for 10 leaves not associated with the inflorescence
- VIII. Average leaf width in cm
Average width of 10 leaves not associated with the inflorescence, measured at their widest point
- IX. Leaf length/width ratio
1.0 = 1.0–1.5
2.0 = 1.0–2.0
3.0 = 1.5–2.0
4.0 = 1.5–3.0
5.0 = 2.0–3.0
6.0 = 2.0–5.0
7.0 = 4.0–15.0
8.0 = 7.0–30.0
9.0 = more than 15.0
- X. Position of widest place on leaf
1.0 = nearer to base of leaf
2.0 = at, or distal to, half-way
- XI. Leaf margin
1.0 = flat
2.0 = slightly revolute
3.0 = strongly rolled
-

Table 2. Continued.

XII. Leaf apex	1.0 = acute
	2.0 = acute to apiculate
	3.0 = apiculate to emarginate
XIII. Pedicel length in mm	Average length of 10 pedicels
XIV. Surface of the fruit nut	1.0 = smooth to rugulose
	2.0 = rugose to tuberculate
	3.0 = rugose to tuberculate with lateral winglet
XV. Length of fruit wing in mm	Average length from fruit nut to tip of dorsal wing for 10 fruits
XVI. Width of fruit wing in mm	Average width of 10 fruits measured at widest point
XVII. Bracteoles length in mm	Average length of 10 bracteoles

analysis; by altering the principles for choosing good characters, expressed through values for X and Y, different choices for good characters result (Table 5).

The Banisteriopsis campestris Complex

Banisteriopsis is a large neotropical genus in the family Malpighiaceae. Members of this genus are characterized by small interpetiolar stipules, the peduncle subtending the flower greatly reduced or absent, flowers with ten fertile stamens and three styles with terminal stigmas, and a fruit of three samaras, each with an enlarged dorsal wing thickened along its upper margin.

Within this genus, the *Banisteriopsis campestris* complex, as recognized by one of us (Gates, 1977, 1982), consists of five species, one, *B. campestris*, very widespread and common in the cerrados throughout the Planalto of Brazil, and four other species with more or less restricted distributions in the Serra do Espinhaco. The specimens used in this study are representative of the five taxa recognized in this group (Table 1).

The flowers of the species in this complex are very similar with only minor differences in size and proportions between them; it is not possible to assign a plant to a species within the complex using flowers alone. For this reason, flower characters have not been used in this study. The fruits in the complex are also similar, but there is sufficient variation in the size of the fruit wing for this to be taxonomically useful. Most of the differences that distinguish the members of this complex are vegetative; none are vines, but there is variation in the habit, and in the size, shape and pubescence of the leaves. For these reasons, the characters used in this study are fruit or vegetative characters (Table 2). The values that are associated with each collection used in this analysis are given in Table 3, the basic data matrix.

A Ranking of the Characters by the Specialist

In order to assess the taxonomic usefulness of a character, each character was evaluated with regard to its intrinsic measurability. Measurements of a character made in several places on the same specimen typically show some variation around an average value; when this variation is small when compared with the variation of this average value among specimens, the character is considered a sounder basis for assessing similarities and differences among specimens. In considering the measurements for a character for all specimens at once, if there were some gaps or discontinuities in the distribution of measurements over the range of possible values, the character was considered potentially more useful. If

Table 3. Basic data matrix.

Collec- tion	Character																
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
I 270	2	2	2	2	2	1	7.9	4.0	4	1	2	2	10	1	15	13	1.4
I 275	2	1	2	2	2	1	5.7	4.1	2	1	1	2	10	3	17	13	1.6
H 315	2	2	3	2	4	1	6.9	5.6	1	2	1	3	11	2	15	11	1.5
I 203	2	2	2	2	4	1	5.0	3.7	1	2	1	3	11	1	11	10	1.9
I 146	2	2	3	2	2	1	7.6	4.6	3	2	2	2	12	1	21	15	1.6
A 369	2	1	2	2	2	1	6.0	4.1	2	1	1	2	12	2	21	15	1.4
A 673	2	1	3	1	2	1	7.6	3.1	5	1	2	2	11	1	20	13	1.6
I 265	2	1	2	2	2	1	6.4	4.2	2	2	1	3	16	1	22	17	1.4
A 790	2	1	3	2	2	1	7.4	3.0	4	1	1	2	11	1	20	12	1.4
A 798	2	1	2	2	2	1	5.6	3.5	2	1	1	2	10	1	17	12	1.3
D 149	2	1	2	2	2	1	7.2	3.7	3	1	1	2	13	2	20	11	1.6
D 256	2	1	3	2	3	1	6.0	3.4	3	1	1	2	14	1	22	13	1.2
M 234	2	2	3	2	2	1	5.7	3.9	2	1	1	2	14	3	19	14	1.5
A 125	2	1	2	2	2	1	7.9	5.8	1	1	1	2	10	1	19	13	1.3
A 120	2	2	2	2	3	1	6.6	3.7	3	1	1	2	13	1	19	10	1.5
G 344	2	2	2	3	1	1	4.6	2.7	2	2	2	2	9	1	14	9	1.3
G 341	3	2	1	3	1	1	8.9	0.2	9	1	3	1	11	2	11	7	1.2
G 346	3	2	1	3	1	1	7.8	0.6	8	1	3	1	14	1	15	10	1.2
G 348	3	2	1	3	1	1	9.5	1.7	7	1	3	1	15	1	16	12	1.7
G 367	3	2	1	3	1	1	6.4	0.8	7	1	3	1	11	2	19	13	1.4
A 844	3	2	1	3	3	1	8.0	0.8	7	1	3	1	16	1	16	13	1.5
A 875	3	2	1	3	1	1	7.6	0.8	7	1	3	1	16	1	17	13	1.9
I 310	2	3	2	3	1	1	7.2	2.5	6	1	3	1	11	1	14	11	2.0
G 386	1	2	2	2	4	1	5.1	2.3	6	1	1	3	10	1	6	5	1.8
E 110	1	2	2	2	4	1	4.7	2.1	5	1	1	3	10	1	6	5	1.9
I 227	4	4	4	2	1	1	4.5	1.9	5	1	1	2	21	3	28	17	2.0
I 279	4	4	4	2	1	1	3.7	2.0	3	1	1	2	14	3	23	12	1.6
A 897	4	4	4	2	1	1	4.9	2.0	5	1	1	2	13	3	23	13	1.5
A 838	3	2	2	1	1	2	5.5	0.4	8	1	3	1	17	1	18	10	1.5
A 850	3	1	2	1	1	2	6.2	0.3	9	1	3	1	16	2	13	9	1.4
G 351	3	2	2	1	1	2	5.3	0.7	7	1	3	1	10	1	15	10	1.3

Table 4. Normalized data matrix.

Collection	Character																
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
I 270	-0.56	0.11	-0.23	-0.21	0.09	-0.33	1.07	0.87	-0.20	-0.41	0.28	0.24	-0.97	-0.68	-0.46	0.49	-0.56
I 275	-0.56	-1.06	-0.23	-0.21	0.09	-0.33	-0.53	0.93	-1.01	-0.41	-0.81	0.24	-0.97	1.96	-0.03	0.49	0.31
H 315	-0.56	0.11	0.94	-0.21	2.04	-0.33	0.34	1.90	-1.42	3.22	-0.81	1.72	-0.61	0.64	-0.46	-0.23	-0.13
I 203	-0.56	0.11	-0.23	-0.21	2.04	-0.33	-1.04	0.67	-1.42	3.22	-0.81	1.72	-0.61	-0.68	-1.31	-0.59	1.62
I 146	-0.56	0.11	0.94	-0.21	0.09	-0.33	0.85	1.25	-0.60	1.40	0.28	0.24	-0.24	-0.68	0.82	1.20	0.31
A 369	-0.56	-1.06	-0.23	-0.21	0.09	-0.33	-0.31	0.93	-1.01	-0.41	-0.81	0.24	-0.24	0.64	0.82	1.20	-0.56
A 673	-0.56	-1.06	0.94	-1.86	0.09	-0.33	0.85	0.29	0.21	-0.41	0.28	0.24	-0.61	-0.68	0.60	0.49	0.31
I 265	-0.56	-1.06	-0.23	-0.21	0.09	-0.33	-0.02	1.00	-1.01	1.40	-0.81	1.72	1.23	-0.68	1.03	1.92	-0.56
A 790	-0.56	-1.06	0.94	-0.21	0.09	-0.33	0.70	0.22	-0.20	-0.41	-0.81	0.24	-0.61	-0.68	0.60	0.12	-0.56
A 798	-0.56	-1.06	-0.23	-0.21	0.09	-0.33	-0.61	0.55	-1.01	-0.41	-0.81	0.24	-0.97	-0.68	-0.03	0.12	-1.00
D 149	-0.56	-1.06	-0.23	-0.21	1.07	-0.33	0.56	0.67	-0.60	-0.41	-0.81	0.24	0.13	0.64	0.60	-0.23	0.31
D 256	-0.56	-1.06	0.94	-0.21	1.07	-0.33	-0.31	0.48	-0.60	-0.41	-0.81	0.24	0.50	-0.68	1.03	0.49	-1.44
M 234	-0.56	0.11	0.94	-0.21	0.09	-0.33	-0.53	0.80	-1.01	-0.41	-0.81	0.24	0.50	1.96	0.39	0.84	-0.13
A 125	-0.56	-1.06	-0.23	-0.21	0.09	-0.33	1.07	2.03	-1.42	-0.41	-0.81	0.24	-0.97	-0.68	0.39	0.49	-1.00
A 120	-0.56	0.11	-0.23	-0.21	1.07	-0.33	0.12	0.67	-0.60	-0.41	-0.81	0.24	0.13	-0.68	0.39	-0.59	-1.00
G 344	-0.56	0.11	-0.23	1.43	-0.88	-0.33	-1.33	0.03	-1.01	1.40	-0.81	0.24	-1.34	-0.68	-0.67	-0.95	-1.00
G 341	0.77	0.11	-1.40	1.43	-0.88	-0.33	1.79	-1.58	1.84	-0.41	1.38	-1.24	-0.61	0.64	-1.31	-1.67	-1.44
G 346	0.77	0.11	-1.40	1.43	-0.88	-0.33	0.99	-1.32	1.43	-0.41	1.38	-1.24	0.50	-0.68	-0.46	-0.59	-1.44
G 348	0.77	0.11	-1.40	1.43	-0.88	-0.33	2.23	-0.61	1.02	-0.41	1.38	-1.24	0.87	-0.68	-0.25	0.13	0.75
G 367	0.77	0.11	-1.40	1.43	-0.88	-0.33	-0.02	-1.19	1.02	-0.41	1.38	-1.24	-0.61	0.64	0.39	0.49	-0.56
A 844	0.77	0.11	-1.40	1.43	-0.88	-0.33	1.14	-1.19	1.02	-0.41	1.38	-1.24	1.23	-0.68	-0.25	0.49	-0.13
A 845	0.77	0.11	-1.40	1.43	-0.88	-0.33	0.85	-1.19	1.02	-0.41	1.38	-1.24	1.23	-0.68	-0.03	0.49	1.62
I 310	-0.56	1.28	-0.23	1.43	-0.88	-0.33	0.56	-0.10	-0.62	-0.41	1.38	-1.24	-0.61	-0.68	-0.67	-0.23	2.06
G 386	-1.89	0.11	-0.23	-0.21	2.04	-0.33	-0.97	-0.23	0.62	-0.41	-0.81	1.72	-0.97	-0.68	-2.38	-2.38	1.18
E 110	-1.89	0.11	-0.23	-0.21	2.04	-0.33	-1.26	-0.36	0.21	-0.41	-0.81	1.72	-0.97	-0.68	-2.38	-2.38	1.62
I 227	2.10	2.45	2.11	-0.21	-0.88	-0.33	-1.41	-0.48	0.21	-0.41	-0.81	0.24	3.07	1.96	2.31	1.92	2.06
I 279	2.10	2.45	2.11	-0.21	-0.88	-0.33	-1.99	-0.42	-0.60	-0.41	-0.81	0.24	0.50	1.96	1.24	0.13	0.31
A 897	2.10	2.45	2.11	-0.21	-0.88	-0.33	-1.11	-0.45	0.21	-0.41	-0.81	0.24	0.13	1.96	1.24	0.49	-0.13
A 838	0.77	0.11	-0.23	-1.86	-0.88	3.06	-0.68	-1.45	1.43	-0.41	1.38	-1.24	1.60	-0.68	0.18	-0.59	-0.13
A 850	0.77	-1.06	-0.23	-1.86	-0.88	3.06	-0.17	-1.51	1.84	-0.41	1.38	-1.24	1.23	0.64	-0.89	-0.95	-0.56
G 351	0.77	0.11	-0.23	-1.86	-0.88	3.06	-0.82	-1.26	1.02	-0.41	1.38	-1.24	-0.97	-0.68	-0.46	-0.59	-1.00

Table 5a. Selected character rankings of 7 best characters for Z = 0.2.

X = 0.2					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-3}$	Character	Score $\times 10^{-3}$	Character	Score $\times 10^{-4}$
8	5.32	9	6.95	6	4.19
7	5.22	8	6.89	10	3.63
9	5.12	12	6.59	2	2.91
15	5.11	7	6.47	13	2.86
17	4.84	15	6.42	15	2.86
16	4.80	11	6.41	3	2.84
13	4.80	1	6.36	5	2.81

X = 1.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-8}$	Character	Score $\times 10^{-8}$	Character	Score $\times 10^{-9}$
8	1.54	9	2.36	6	1.56
9	1.53	11	2.30	10	1.09
7	1.47	12	2.30	9	1.08
15	1.46	8	2.28	12	1.06
12	1.41	1	2.09	8	1.05
1	1.41	3	1.99	11	1.01
13	1.39	4	1.97	3	0.98

X = 3.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-11}$	Character	Score $\times 10^{-11}$	Character	Score $\times 10^{-5}$
8	6.22	9	1.09	6	9.40
9	6.20	12	1.04	9	6.08
7	6.16	8	1.03	8	5.68
1	6.16	11	1.02	12	5.43
12	6.09	4	0.87	10	5.29
15	6.07	1	0.85	11	4.53
16	6.04	16	0.83	16	4.07

Table 5b. Selected character rankings of 7 best characters for Z = 0.2.

X = 0.2					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-2}$	Character	Score $\times 10^{-2}$	Character	Score $\times 10^{-3}$
8	7.85	9	9.75	6	5.79
7	7.82	8	9.73	10	5.23
15	7.61	7	9.57	2	4.23
9	7.47	15	9.37	13	4.14
17	7.25	17	9.27	15	4.11
16	7.15	16	9.14	16	4.06
13	7.13	12	9.10	1	3.97

Table 5b. Continued.

X = 1.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-3}$	Character	Score $\times 10^{-4}$	Character	Score $\times 10^{-4}$
8	7.76	9	1.03	6	6.03
7	7.60	8	1.02	10	5.25
15	7.51	12	1.00	13	4.67
9	7.48	1	0.99	15	4.66
17	7.04	15	0.98	4	4.50
16	7.00	7	0.95	1	4.47
13	6.97	11	0.95	12	4.42

X = 3.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-11}$	Character	Score $\times 10^{-11}$	Character	Score $\times 10^{-7}$
8	3.16	12	4.60	6	2.53
7	3.08	1	4.58	13	2.43
9	3.08	9	4.53	15	2.34
15	3.06	8	4.36	10	2.29
1	2.94	15	4.33	12	2.25
17	2.88	11	4.27	5	2.23
16	2.88	13	4.12	1	2.12

Table 5c. Selected character rankings of 7 best characters for Z = 0.2.

X = 0.2					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-2}$	Character	Score $\times 10^{-2}$	Character	Score $\times 10^{-3}$
7	5.93	9	7.28	6	4.32
8	5.92	8	7.27	10	3.91
15	5.75	7	7.23	2	3.19
9	5.63	15	7.05	13	3.11
17	5.49	17	7.01	15	3.08
16	5.40	16	6.89	16	3.04
13	5.39	13	6.85	1	2.98

X = 1.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-3}$	Character	Score $\times 10^{-3}$	Character	Score $\times 10^{-4}$
8	1.84	9	2.35	6	1.37
7	1.82	8	2.32	10	1.20
15	1.79	15	2.30	15	1.09
9	1.76	1	2.27	13	1.08
17	1.68	7	2.27	1	1.06
16	1.67	12	2.26	2	1.05
13	1.66	16	2.19	5	1.03

Table 5c. Continued.

X = 3.0					
Y = 0.2		Y = 1.0		Y = 3.0	
Character	Score $\times 10^{-4}$	Character	Score $\times 10^{-4}$	Character	Score $\times 10^{-5}$
8	3.67	1	5.27	15	2.87
7	3.62	15	5.07	13	2.80
15	3.61	12	4.99	1	2.71
9	3.56	9	4.93	6	2.61
17	3.37	5	4.78	5	2.56
16	3.34	8	4.77	16	2.48
13	3.12	13	4.73	10	2.39

a character were easy to measure, and concerns a part that is likely to be present on most specimens, it was considered potentially more useful. In addition, a character must also "work" taxonomically. Two specimens with similar values for a good character should belong to the same taxon. This is a somewhat circular criterion, because at the time the character is being evaluated, there are no taxa. Ideally similar specimens should belong to the same taxon, and dissimilar specimens should belong to different taxa. Thus, a character that seems to assign similar values to similar specimens, and different values to dissimilar specimens will be considered taxonomically useful.

The above considerations were used to generate the following grouping of characters, ranked in order of their taxonomic usefulness.

- most useful: 1 habit
 4 pubescence of lower leaf surface
 5 pubescence of upper leaf surface
 6 position of leaves
 7 leaf length
 8 leaf width
 9 length/width ratio
 15 length of fruit wing
 16 width of fruit wing
- less useful: 2 petiole length
 3 position of leaf glands
 11 leaf margin
 12 leaf apex
- least useful: 10 position of leaf's widest point
 13 pedicel length
 14 surface of fruit nut
 17 bracteole length

The first 9 characters are all considered good characters for taxonomic purposes. Characters 1 and 6 exhibit no within specimen variation, but character 6 has only 2 states, one of which isolates only a few specimens with erect leaves. However, specimens with erect leaves are all fairly similar to each other, and this character state is easily observed and unambiguous; character 6 is a good character in these respects. Character 1 is a very good character where field observations are available, as in addition to its lack of within specimen variation, it is easily observed and the subdivision it effects places similar specimens together. Unfortunately, plants with the habit of tree or shrub cannot be distinguished on an herbarium sheet without field notes, and young and regenerating individuals whose mature form is a tree or shrub often have the habit of an unbranched shrub. Characters 8

and 9 show considerable within specimen variation, but are useful because the variation among specimens is very high and corresponds to the overall similarity of the specimens. These characters are also easily measured and always obtainable. Characters 8 and 9 are highly correlated ($r = 0.91$) which suggests much shared information between them. Character 7 shows high variation among specimens, is easy to measure, and is usually available for measurement. But it is only within the group of collections with very short leaves, or within the group with very long leaves, that collections show high overall similarity. Character seven is completely uncorrelated with character eight ($r = 0.01$). Character 15 shows little within specimen variation and considerable variation between specimens, is easily recorded when fruits are present, and the variation corresponds well to the overall similarity. Character 16 shows more within specimen variation, and the range of measurements represented by this character is less, compared with character 15 so that although characters 15 and 16 are highly correlated ($r = 0.85$), character 16 is somewhat less useful in practice. Characters 4 and 5 show little within specimen variation, and correspond well to overall similarity. There is some difficulty in recording these characters because of a lack of discontinuity in the variation.

Characters 2, 3, 11 and 12 are considered less useful for taxonomic purposes. Character 2 is easily measured, but the variation among specimens in this character is relatively low, and there are no discontinuities in the measurements. Thus although the pattern of variation within this character more or less corresponds to the overall similarity, individual measurements are not very useful for taxonomic purposes. Character 3 has fairly high within specimen variation, and its variation does not correspond to the overall similarity. In character 11, slightly rolled leaf margins are difficult to distinguish from flat margins, and sometimes both states are represented in the same collection. In character 12, both acute and apiculate (or both apiculate and emarginate) apices are sometimes present in the same collection. For these reasons, although characters 11 and 12 seem to correspond fairly well to the overall similarity they are not included in the group of most useful characters. These two characters are strongly correlated with each other ($r = 0.85$), and with characters 5, 8 and 9.

Characters 10, 13, 14 and 17 are considered least useful for taxonomic purposes. Character 14 has high within specimen variation; it is difficult to record this character because on a single specimen fruits with rugose nuts and nuts with lateral projections will sometimes occur. Character 10 also shows considerable within specimen variation, such that a single herbarium sheet may not sample all of the variation on one individual. Field observations suggest that some changes in leaf shape occur during ontogeny, with young and regenerating shoots often bearing leaves of a different shape. Character 13 serves principally to isolate one distinct species. Character 17 exhibits little within specimen variation, and the variation among specimens more or less corresponds to overall similarity; however, it was not used to establish and recognize species because it did not correlate well with other characters considered better indications of specific status, and its total range of variation is small.

Characters 8 and 9 are highly correlated ($r = -0.91$) because width is the denominator in length-width ratio. Character pairs: 11 and 12 ($r = -0.85$); 5 and 12 ($r = 0.81$); 8 and 11 ($r = -0.72$); and 9 and 11 ($r = 0.84$) also show high correlation, but these characters measure different parts of the plant. In this case, such a group of correlated characters may reflect adaptive syndromes. This would make them very desirable for taxonomic purposes.

Results

Character analysis using phenetic spatial auto-correlation with formulae (1) and (2) was applied to the Normalized Data Matrix of Table 4. Just as exposure time, aperture, and focus affect the picture made by a camera, so do the parameters X, Y, Z affect the "picture" made of a character's pattern of variation in phenetic space. It is not clear to us what values will be appropriate in what situations, so we present 27 analyses to represent three values

for each of X, Y, and Z. We anticipate that the effect of these values, and their relative appropriateness will become more clear as others gain experience with this method. The most striking aspect of the results of character analysis by spatial auto-correlation (Table 5a, b, and c) is the similarity of the rankings to each other for low values of Y (=0.2) and all values of X and Z. In these cases, characters 7, 8, 9, and 15 were ranked high. These were considered "good" in the opinion of the specialist, but also included in the top seven for these cases were 12, 13, and 17 which were judged less useful. One collection in the study group shows unusually long pedicels (over 3 standard deviations). This collection is generally large in all parts. Character 13 (pedicel length) isolates it most clearly. Thus, in the context of this study collection, character 13 may be overrated. Bracteole length, character 17, shows virtually no within collection variation and seems to correlate well with overall similarity, but because its range of variation is only 0.8 mm, and because in large groups of fairly similar collections much of this range is represented, it was considered somewhat impractical. Thus, in the assessment of the specialist, character 17 has been somewhat overrated as well. Remember that within specimen variability, and the overlap of ranges, are not factors in spatial auto-correlation.

Rankings for Y = 1.0 and X = 1.0 or 3.0 and all values for Z include a slightly different selection of characters with 1 ranked high and 7 absent or ranked lower. All rankings for X = 1.0 and all values of Y and Z include one or more of the less useful characters 3, 11, 12, 13 or 17.

Rankings for X = 0.2 or 1.0 and Y = 3.0 and all values of Z are very different from those at low and medium Y. In these cases, characters 6 and 10 are ranked first and second and there is a large drop in the score for the third ranked characters. Characters 6 and 10 are each 2-state characters with only a few individuals in one of the states. Very high Y tends to favor this kind of character. Character 10 gives a very unsatisfactory taxonomy of the *B. campestris* complex. Also in these cases, the leaf size and shape characters that were favored for other parameter values are missing from the top 7 or are ranked lower, and characters 2 and 13 (considered less useful) have been included. Of the three remaining cases with high Y, those for which X = 3.0 and Y = 3.0 and Z = 0.2 or 1.0 rank character 6 first but give lower rank to character 10; still three of the top seven characters are considered less useful. When X, Y, and Z are all 3.0, character 10 falls to seventh and only the overrated 13 is still in the top six. High values of Y seem to give undesirable results, that can be counteracted somewhat by high values of X and Z.

In assessing the overall usefulness of the different rankings for taxonomic purposes, a ranking is considered good if the top seven characters are all or mostly from the list of "good" characters, but remember that the criteria are different, so do not overrate our rankings. By this criterion, the best rankings are given by: X = 3.0, Y = 0.2, and Z = 0.2; X = 3.0, Y = 0.2, and Z = 1.0; and X = 1.0, Y = 1.0, and Z = 3.0. The six characters 1, 7, 8, 9, 15 and 16 are common to all three best rankings, and are all considered "good." They describe leaf size and shape, fruit wing size, and habit. Use of these characters would classify the *B. campestris* complex into groups similar to those of Gates (1977, 1982).

The rankings for X = 1.0, Y = 0.2, and Z = 0.2; X = 0.2, Y = 1.0, and Z = 0.2; and X = 1.0, Y = 1.0, and Z = 1.0 each include the same five "good" characters. These five are the same as the six "good" characters discussed in the previous paragraph except that fruit wing width (16) is replaced by various less useful characters. Use of this slightly changed group of characters would produce nearly the same taxonomy as before. The following rankings also contain five of these six "good" characters, but habit has been replaced by a less useful character: X = 0.2, Y = 0.2, and all values of Z; X = 1.0, Y = 0.2, and Z = 1.0 or 3.0; X = 1.0, Y = 1.0, and Z = 1.0; X = 0.2, Y = 1.0, and Z = 3.0, and X = 3.0, Y = 0.2, and Z = 3.0. The less useful characters that appear in this context are 12, 13 and 17, which have been earlier discussed.

The best ranking with five "good" characters is given by: X = 3.0, Y = 1.0, and Z =

3.0. This list includes characters 1, 5, 8, 9, and 15. These characters provide information about many different aspects of the plant including habit, leaf pubescence, leaf size, and fruit wing size.

In summary, character analysis by phenetic spatial auto-correlation has fairly consistently identified good characters for use in a difficult complex of closely related taxa, by revealing for each character, the pattern of variation in the phenetic space of the other characters in the study. This information can be considered in conjunction with other information that may be available, such as: within specimen variability; ontogenetic variability; ease with which data can be observed; environmental plasticity, to choose more effectively what characters might best be used as a basis for defining and recognizing taxa. The most appropriate values of X, Y and Z in the context of other taxonomic studies remains to be determined by experience. A Fortran Program is available from the authors.

Literature Cited

- Anderson, T. W. 1958. *An introduction to multivariate statistical analysis*. John Wiley and Sons, Inc., New York.
- Baum, B. R. 1974. Classification of the oat species (*Avena*, Poaceae) using various taximetric methods and an information theoretic model. *Canadian J. Botany* 52: 2241-2262.
- Bisby, F. A. 1970. The evaluation and selection of characters in angiosperm taxonomy: An example from *Crotalaria*. *New Phytol.* 69: 1149-1160.
- Cliff, A. D. and J. K. Ord. 1973. *Spatial autocorrelation*. Pion, London.
- Estabrook, G. F. 1967. An information theory model for character analysis. *Taxon* 16: 86-97.
- . 1983a. Phylogenetic trees and character state trees. In: T. O. Duncan and T. F. Stuessy (eds.), *Cladistics: Perspectives on the reconstruction of evolutionary history*. Columbia University Press, N.Y.
- . 1983b. The causes of character incompatibility. In: J. Felsenstein (ed.), *Proceedings of the NATO Advanced Study Institute on Numerical Taxonomy, Bad Windsheim FRG 1982*. Springer-Verlag, Berlin.
- Gates, B. 1977. *A monograph of the Central Brazilian species of Banisteriopsis (Malpighiaceae)*. Ph.D. Thesis, The University of Michigan, Ann Arbor.
- . 1982. *Banisteriopsis, Diplopterys (Malpighiaceae)*. *Flora Neotropica Monograph* Number 30, pp. 1-237.
- Hawksworth, F. G., G. F. Estabrook and D. J. Rogers. 1968. An application of an information theory model for character analysis in the genus *Arceuthobium* (Viscaceae). *Taxon* 17: 605-619.
- Sokal, R. R. and N. L. Oden. 1978. Spatial autocorrelation in biology 1. Methodology. *Biol. J. Linn. Soc.* 10(2): 199-228.
- Pearson, K. 1926. On the coefficient of racial likeness. *Biometrika* 18: 105-117.
- Sneath, P. H. A. and R. R. Sokal. 1973. *Numerical taxonomy*. W. H. Freeman and Co., San Francisco.