Complex Data Produce Better Characters

By: B. K. KIRCHOFF[1], S. J. RICHTER[2], D. L. REMINGTON[1], AND E. WISNIEWSKI[3]

Kirchoff, B. K., S. J. Richter, D. L. Remington, and E. Wisniewski. 2004. Complex Data Produce Better Characters. Systematic Biology 53: 1-17.

Abstract:

Two studies were conducted to explore the use of complex data in character description and hybrid identification. In order to determine if complex data allow the production of better characters, eight groups of plant systematists were given two classes of drawings of plant parts, and asked to divide them into character states (clusters) in two separate experiments. The first class of drawings consisted only of cotyledons. The second class consisted of triplets of drawings: a cotyledon, seedling leaf, and inflorescence bract. The triplets were used to simulate complex data such as might be garnered by looking at a plant. Each experiment resulted in four characters (groups of clusters), one for each group of systematists. Visual and statistical analysis of the data showed that the systematists were able to produce smaller, more precisely defined character states using the more complex drawings. The character states created with the complex drawings also were more consistent across systematists, and agreed more closely with an independent assessment of phylogeny. To investigate the utility of complex data in an applied task, four observers rated 250 hybrids of *Dubautia ciliolata X arborea* based on the overall form (Gestalt) of the plants, and took measurements of a number of features of the same plants. A composite score of the measurements was created using principal components analysis. The correlation between the scores on the first principal component and the Gestalt ratings was computed. The Gestalt ratings and PC scores were significantly correlated, demonstrating that assessments of overall similarity can be as useful as more conventional approaches in determining the hybrid status of plants. [Analytic processing; Banskia; character; character state; complexity; configural processing; *Dubautia*; holistic processing; phylogenetic analysis.]

Article:

In current usage, morphological characters and character states are abstractions that are conceptually separated from the whole organism. Stevens (personal comm.) has referred to these types of characters as “atomistic.” They have little relationship to the context in which they occur. This lack of attention to the context may be one reason why it is possible to produce alternative character states based on the same data (Gift and Stevens, 1997). Separating character

---

[1] Department of Biology, P. O. Box 26170, University of North Carolina, Greensboro, North Carolina 27402, USA; E-mail: kirchoff@uncg.edu (B.K.K.), dlreming@uncg.edu (D.L.R.)
[2] Statistics Division, Department of Mathematical Sciences, University of North Carolina, Greensboro, North Carolina 27402, USA; E-mail: sjricht2@uncg.edu
[3] Department of Psychology, University of North Carolina, Greensboro, North Carolina 27402, USA; E-mail: edw@uncg.edu

states from their context simplifies them to the extent that alternative groupings are equally plausible.

In a survey of 512 phylogenetic studies spanning the years 1986–1997, Poe and Wiens (2000) found that only 20% of the papers mentioned any type of criterion by which morphological characters were excluded from or included in the study. Likewise, in her survey of the botanical literature, Hawkins (2000) found little discussion of why a particular approach to character construction was used.

This lack of detail may be due to the fact that systematists believe that character state description is an objective and repeatable process (Theile,1993; Poe and Wiens, 2000). However, several lines of evidence suggest that it is neither objective nor repeatable. First, Poe and Wiens (2000) point to studies of the same organism by different researchers to show that people select characters in different ways (e.g., Good, 1988; Campbell and Frost, 1993). Second, Stevens (1991) cites examples where scientists who reject the use of quantitative characters nevertheless use characters that either have a quantitative component (size less than or greater than a given value), or that can be interpreted to have such (large versus small). The implication is that these scientists were not fully aware of how they were creating their characters. Third, Gift and Stevens (1997) demonstrate that different systematists create different character states based on the same graphical presentation of data. The work presented here extends Gift and Stevens's (1997) result to simple qualitative data, such as are frequently used in phylogenetic analyses. Our research is directed at these problems.

This article is about character state construction, the process by which variation is partitioned into character states. We approach characters as clusters of traits that are subdivided into smaller clusters called character states. This operational definition allows us to look at the process of character state construction as a grouping activity, and to apply insights from cognitive psychology to help improve performance in this task. According to this perspective, character states arise when traits that have been judged homologous are placed into subcategories. In some cases, transformation relationships are specified between these categories, but these transformations are not necessary for the construction of undirected characters. The characters described in this article are undirected, and consist of only two hierarchical levels. We refer to the highest level as the character itself, and to the set of partitions as the character states. No transformation relationships between these states are specified. We deal only with multistate, undirected characters.

Before proceeding with a description of our research, we take a brief look at more standard methods of describing characters. According to standard definitions, a morphological character is any feature that is shared among organisms (Davis and Heywood, 1973; Colless, 1985; Stuessy, 1990; Fristrup, 1992). Implicit in this definition is the idea that characters are groups of features. Yet this idea is seldom operationalized. Characters are commonly described by studying a number of individuals belonging to distinct taxa, determining similarities of the attributes of these organisms to create a "trait" for each taxon, and finally determining the homologies of these traits to create characters (Jardine, 1969; Stevens, 1984; de Pinna, 1991; Brower and Schawaroch, 1996). In a phylogenetic sense, characters are hypotheses of primary homology (de Pinna,1991). They are clusters of traits that have been assessed as homologous. Since the criteria by which homologies are determined have been well studied (Remane, 1952; Patterson, 1982; Hall, 1994, 1999), the criteria by which characters are created should be relatively easy to describe. They are those by which the traits were homologized.

Character states are defined as one or more alternative manifestations of a character (Kitching et al., 1998:201). If we allow as many unique character states as there are taxa, then the specification of character states is unproblematic. Each taxon has its own character state, which corresponds to what we have called a "trait." The problem with these types of highly articulated characters is that they are useless for phylogenetic studies. They do not suggest relationships among taxa. For this, it must be possible to group taxa based on the possession of shared character states (Platnick,1979). At least some character states must describe more than one taxon. To achieve this, we must make similarity judgments that unite certain traits into character states. In conventional practice, this clustering is often done implicitly through the use of verbal labels (Patterson, 1982; Stevens, 1991). The labeling process begins with the reduction of the initial variability in attributes to produce a trait of a taxon through the use of a label. For instance, a complex labellum may be labeled "obdeltoid," thereby reducing all aspects of its shape to something that is inverted and broadly triangular (Kirchoff, 2001). Once developed, this label is then applied to the scoring of new attributes. A mental image of a "typical" obdeltoid labellum is created, compared to the attribute at hand, and an initial assessment of that taxon's trait is made. If further study of the attributes of the taxon is undertaken, it is usually only to confirm the initial assessment of the shape, as specified by the shape term. In this way, character states are transformed from frequency distributions (Thiele, 1993) to discrete values—the traits of the taxa. Once these traits are established, they must be coded for inclusion in a data matrix. Hawkins et al. (1997) and Hawkins (2000) have explored some of the pitfalls of this process.

The use of labels makes it unnecessary to group attributes explicitly into character states. The labels themselves provide the means by which the clusters are created. In labeling an attribute, we place it into a class that is implicitly identical with all other attributes that receive the same label. A flabellate cotyledon of Taxon A belongs to the same character state as one from Taxon B. Both are flabellate. Using this process there is no need to physically compare the two shapes. The comparison is done via our mental image.

In this article we explore an alternative to the use of verbal labels for the construction of character states. This process involves the direct comparison of traits and the placement of these traits into clusters (character states). We do not deal with the transformation of attribute frequency distributions into traits, though this is an important issue in its own right. The process of grouping traits into character states is perhaps the least studied of all aspects of character construction because it is transparent in normal practice. It is made transparent by the use of verbal labels. The process can be made explicit by dispensing with the labels and directly comparing sets of images representing the traits of the taxa under study. Placing these images into clusters creates character states. This creation is explicit and lends itself to analysis. Using this method we are able to investigate the quality of the character states created with simple versus complex traits. We can also investigate the reliability with which different investigators construct character states from the same data.

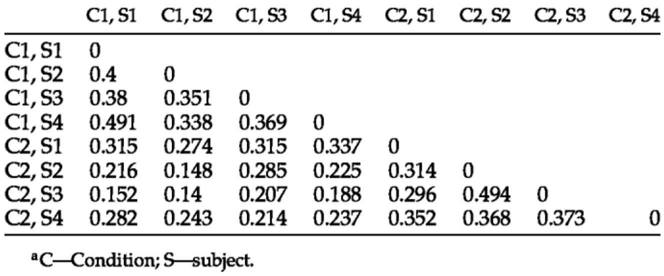
The studies presented here address the question of trait complexity in two ways. First, we report the results of two grouping tasks: one involving relatively simple drawings of plant parts and the other relatively complex drawings of plant parts. With these tasks we test the hypothesis that more complex traits allow systematists to produce smaller, more precisely defined character states with greater inter-investigator reliability, and with greater agreement with an independently derived phylogeny. Our second study tests the utility of global assessments of plant form as a means of determining the hybrid status of plants in the genus *Dubautia* (Asteraceae, Madiinae). This study is included to show that assessments of complex data are as reliable as more conventional approaches in at least one practical application, the identification of hybrids. We discuss the results of both studies in light of research on visual processing systems.

CHARACTER STATE EXPERIMENTS

Methods

To investigate the use of simple versus complex data in the production of character states, four groups of plant systematists at the Royal Botanic Gardens, Sydney, Australia (the Subjects), were given 87 drawings (the Traits) of one of two types (the Conditions), and asked to divide them into clusters corresponding to the states of each character. The subjects were all trained systematists with many years experience. A few were Masters- level herbarium assistants, but most had completed their Ph.D. Groups were self-selecting, ranged in size from two to four systematists, and differed in composition between the two trials. For each Condition, the systematists were given approximately forty-five minutes to create clusters of similar traits. Extra time was available for those who felt the need for it. The clustering tasks were carried out by placing all of the drawings on a table, and asking the subjects to arrange them into as many character states (sub-clusters) as they desired. Hierarchical clusters were not permitted. The drawings were taken from a published morphological phylogeny of the Australian plant genus *Banksia* (Thiele and Ladiges,1996). Although these are well known plants in Australia, none of the systematists were specialists in this genus, and only a few species of the genus grow in the area around Sydney. In addition, all of the drawings used in the study were of plant parts that are not visible to casual observation. The plants must be grown from seed to observe cotyledon and seedling leaf shape, and bract shape is only visible upon dissection of the inflorescence. Thus, it is unlikely that any of the systematists could identify the species from the drawings.

TABLE 2. Matrix of intersubject correlations.[a]



	C1, S1	C1, S2	C1, S3	C1, S4	C2, S1	C2, S2	C2, S3	C2, S4
C1, S1	0							
C1, S2	0.4	0						
C1, S3	0.38	0.351	0					
C1, S4	0.491	0.338	0.369	0				
C2, S1	0.315	0.274	0.315	0.337	0			
C2, S2	0.216	0.148	0.285	0.225	0.314	0		
C2, S3	0.152	0.14	0.207	0.188	0.296	0.494	0	
C2, S4	0.282	0.243	0.214	0.237	0.352	0.368	0.373	0

[a] C—Condition; S—subject.

FIGURE 1 IS OMITTED FROM THIS FORMATTED DOCUMENT

The first class of traits consisted of outlines of the cotyledons (Fig. 1). The grouping task using them is called Condition 1. The second class of traits was created by pairing the first drawings with two others taken from the same publication (Thiele and Ladiges,1996). The additional drawings show the first seedling leaf, and one of two paired floral bracts (Fig. 2). These three plant parts were not shown to scale in these drawings, and the systematists were informed of this fact. Two of the traits in this second class lacked floral bracts, and consisted only of a cotyledon and seedling leaf (Fig. 2, row four last trait; Fig. 6Q). The grouping task with the composite drawings is called Condition 2. Both classes of traits were identified with the numbers used by Thiele and Ladiges (1996) to identify the cotyledons (see their fig. 16). Composite drawings were used to simulate more realistic complex traits, which were not available at the time of this study. The opportunities and limitations imposed by the use of these composites are addressed in the “General Discussion” section.

Data analysis was carried out with PAUP* (Swofford, 2002), Component (Page, 1993), SAS (SAS Institute, 1999), and PopTools (Hood, 2002). PAUP* and Component were used to create consensus trees for the two Conditions. These trees were imported into Tree View (Page, 1996) for printing and eventual incorporation into this article. SAS and PopTools were used for the statistical analyses. A detailed treatment of our analytic procedures is given in the next section.

Results and Discussion

Each Condition resulted in four clusterings of traits, one for each group of subjects who participated in the experiments. These clusterings can be represented as trees with a central unresolved polytomy connected to terminal clusters of various sizes (Fig. 3). In these trees, the leaves are the traits, and the branches represent clusters of traits. A cluster is defined as a group of two or more traits, and corresponds to a character state. Saying that a branch is smaller is the same as saying that it contains fewer traits.

Visual inspection of the trees suggests that the subjects were able to produce smaller, more precisely defined clusters when presented with more complex traits (Fig. 3B; Condition 2). This impression is supported by the calculation of the average number of clusters in each Condition (Table 1). The subjects in Condition 2 produced almost twice the number of clusters as those in Condition 1. A measure of how well the subjects were able to discriminate the traits can be calculated by dividing the number of clusters by the number of traits (resolution, Table 1). A tree that lacks clusters will have

FIGURE 2 IS OMITTED FROM THIS FORMATTED DOCUMENT

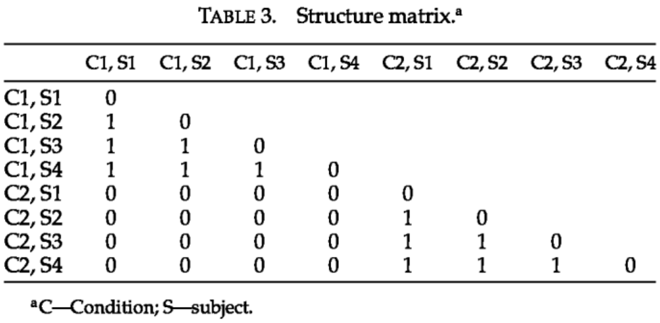


TABLE 3. Structure matrix.[a]

	C1, S1	C1, S2	C1, S3	C1, S4	C2, S1	C2, S2	C2, S3	C2, S4
C1, S1	0							
C1, S2	1	0						
C1, S3	1	1	0					
C1, S4	1	1	1	0				
C2, S1	0	0	0	0	0			
C2, S2	0	0	0	0	1	0		
C2, S3	0	0	0	0	1	1	0	
C2, S4	0	0	0	0	1	1	1	0

[a] C—Condition; S—subject.

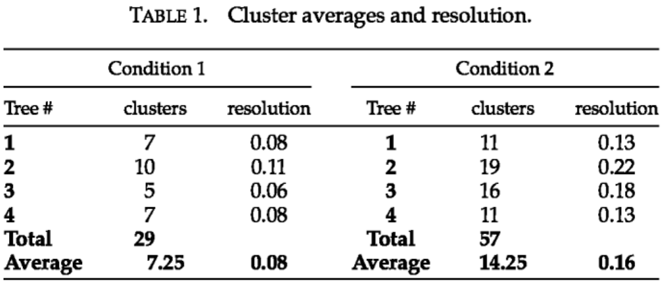


TABLE 1. Cluster averages and resolution.

Condition 1			Condition 2		
Tree #	clusters	resolution	Tree #	clusters	resolution
1	7	0.08	**1**	11	0.13
2	10	0.11	**2**	19	0.22
3	5	0.06	**3**	16	0.18
4	7	0.08	**4**	11	0.13
Total	**29**		**Total**	**57**	
Average	**7.25**	**0.08**	**Average**	**14.25**	**0.16**

a resolution of 0 [0/87], while a tree consisting solely of clusters of two will have a resolution of 0.5 [(87/2)/87]. Resolutions above 0.5 are only possible in hierarchical clusterings, which were not permitted in this experiment. By this measure, the clusters from Condition 2 are twice as resolved as those from Condition 1 (Table 1).

Matrix correlations.—To investigate if the subjects were making use of the additional information provided in the triplets of Condition 2, we compared a matrix of pair- wise intersubject correlations (Table 2) with a structure matrix that represents greater agreement within than between Conditions (Table 3). A significant correlation between these matrices implies greater agreement within than between Conditions. This result is only expected if the subjects were making use of the additional information in the triplets.

To calculate the intersubject correlations (Table 2) we first derived the distance between each pair of traits for

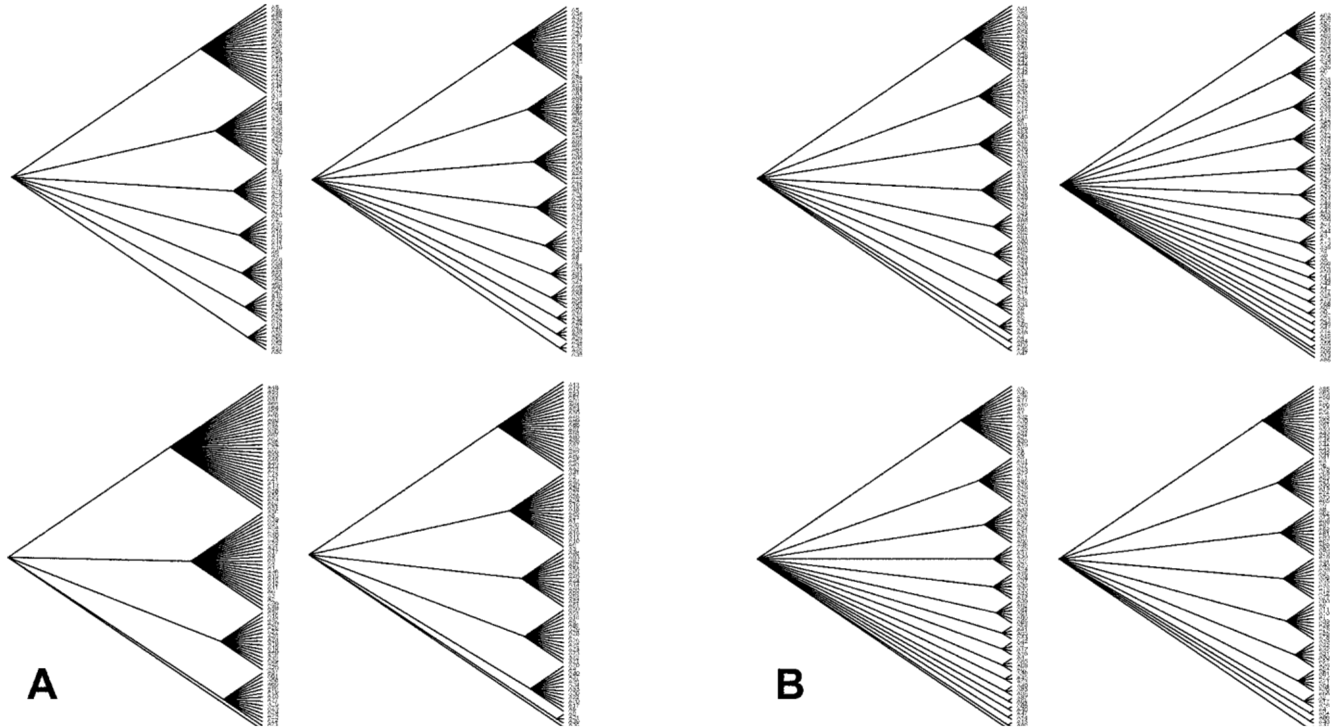


FIGURE 3. Tree representations of groupings produced in the two Conditions. (A) Condition 1. (B) Condition 2. Note that there is a greater number of smaller clusters in Condition 2 than Condition 1.

each grouping. We defined this distance as the number of traits in the cluster that contains both traits. Thus, the maximum distance is the total number of traits, and is assigned to all pairs of traits that were placed into different clusters. The minimum distance is two, and is assigned to any pair of traits that is separated from all other traits by being placed in their own cluster. For each subject, a matrix of pairwise distances between traits was created, and the correlation between these distances for each pair of subjects was computed. The result was the matrix of pairwise intersubject correlations (Table 2). This correlation matrix was compared to the structure matrix (Table 3) using the Mantel Test routine (Mantel, 1967) in PopTools. The two matrices had a correlation of $r = 0.763$ ($p = 0.023$). This implies that agreement is higher and more consistent within Conditions than between Conditions.

To interpret this test, recall that the cotyledon outlines used in Condition 1 were also part of the triplets used in Condition 2. It is possible that the subjects in Condition 2 were ignoring the other two drawings, and making use only of the drawings of cotyledons. If this were true, we would expect to find significant correlations between subjects, across Conditions. However, if the subjects were making use of the additional information present in Condition 2, we would expect to find significant correlations only within Conditions, not between Conditions. The significant correlation between the intersubject correlation matrix (Table 2) and the structure matrix (Table 3) supports the latter supposition. We interpret this result, along with the result that the clusters from Condition 2 were about twice as resolved as those from Condition 1, to mean that the subjects were making use of the additional drawings in Condition 2.

Consensus trees.—Similarities among the clusters within a Condition were assessed by computing consensus trees (Adams, 1972; Swofford, 1991). We computed four commonly used consensus trees: strict, majority rule, semi-strict, and Adams. Each type of tree tells something slightly different about the data. Strict consensus trees contain only those clusters that are common to all of the trees being compared. The strict consensus tree for Condition 1 has no clusters. The tree is completely unresolved, with all traits connected to a central polytomy. The strict consensus tree for Condition 2 contains a single cluster of two traits. That is, every group of systematists placed these two traits, and only these two traits, in the same cluster. Majority rule consensus trees contain those clusters that occur in more than half of the comparison trees. Since four trees were compared in each Condition, a cluster must occur on three of the four trees to be included in the majority rule tree. The Condition 1 tree is completely unresolved. The Condition 2 tree has three clusters. One of these is the same one found on the strict consensus tree. The other two are new to the majority rule tree. Semi-strict consensus trees contain clusters that are found in a subset of trees, and are not contradicted by the other trees. The semi- strict tree for Condition 1 contains four clusters, two of which are nested within a third. The average cluster size is 6.3. The Condition 2 tree contains seven clusters, two of which are nested in a much larger third. Average cluster size is 4.4. All of the clusters from the majority rule tree are found on the semi-strict tree. We interpret these results to mean that use of the more complex traits (Condition 2) permitted greater intersubject agreement (more clusters on the consensus trees), and greater precision in the placement of the traits (smaller clusters).

TABLE 5. Clusters Shared by pairs of trees.

Condition 1				Condition 2			
Tree #	1	2	3	Tree #	1	2	3
2	0			**2**	2		
3	0	0		**3**	4	4	
4	0	0	0	**4**	4	4	8
Avg		**0**		**Avg**		**4.3**	

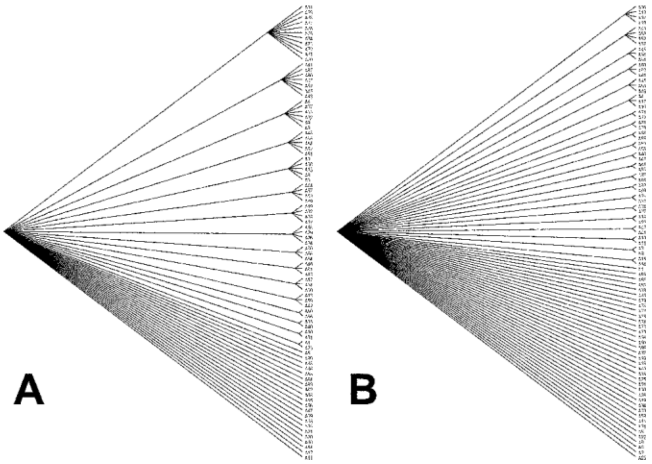


FIGURE 4. Adams trees for the two Conditions. (A) Condition 1. (B) Condition 2. There is a greater number of smaller clusters in Condition 2.

Quite a bit more structure of the data is revealed when we count nested sub-clusters that are common to all of the trees in a Condition. Adams consensus trees contain these clusters. A group of traits appears on the Adams tree if these traits occur together on all comparison trees. In our analysis, the Adams trees contain all of the clusters from the semi-strict trees, but lack the hierarchical resolution of these trees. In addition, the Adams trees contain other clusters not found in any other consensus tree. The main differences between the Adams trees for the two Conditions are the number and sizes of the clusters (Fig. 4, Table 4). The Condition 2 tree has a greater number of smaller clusters than the Condition 1 tree. The number of binary clusters is four in Condition 1, and 15 in Condition 2. This result is consistent with our conclusion that the more complex traits used in Condition 2 allowed the subjects to make more precisely defined clusters (character states) with higher intersubject agreement.

Pairwise comparisons between the subject trees reveal structure that is not apparent in any of the consensus trees (Table 5). None of the pairs of trees in Condition 1 contain any clusters in common. In contrast, pairs of trees from Condition 2 share between two and eight clusters, with an average of 4.3 clusters per pair. This result attests to the greater intersubject agreement in Condition 2 than Condition 1.

TABLE 4. Adams tree clusters.

	Condition 1	Condition 2
Clusters	16	21
Avg cluster size	4.1	2.4
No. traits in clusters	66	50
No. of clusters of 2	4	15

Consistency indices.—Another method of evaluating the quality of the groups of clusters produced in the two Conditions is to treat them as characters, and to compare them to an independently derived phylogenetic tree. The characters that are more consistent with the tree will have a higher consistency index. For purposes of this comparison, we used the chloroplast DNA tree for *Banksia* and *Dryandra* constructed by Mast and Givnish (2002). To determine if the clusters produced in Condition 2 were more consistent with this tree, we coded our eight trees (Fig. 3) as multistate, undirected characters (Appendix 1; available at http://systematicbiology.org) and calculated consistency (CI), and rescaled consistency indexes (RCI) (Kluge and Farris, 1969; Farris, 1989) of these characters on their tree. We translated our clusters into characters by assigning the same character state code (a letter) to all of the traits that occurred in the same cluster (Appendix 1). Means of both indices were calculated for each Condition, and the statistical significance of the difference in means was evaluated based on both parametric (independent-samples t-tests with no assumption of equal variances) and nonparametric (Mann-Whitney U Statistic) tests. We found that both consistency indices were higher in Condition 2 than in Condition 1 (Table 6). The means were almost double for both indices. These differences were found to be statistically significant using both the parametric ($p = 0.008$) and nonparametric ($p = 0.0145$) tests (Table 6). We interpret this result to mean that our subjects produced better estimates of the chloroplast DNA phylogeny when they used more complex traits (Condition 2).

Assessment of the clusters.—In order to gain insight into why the use of more complex traits resulted in better character states, we assessed which aspects of the traits the subjects were using by inspecting the clusters that appear on the Adams trees (Figs. 5, 6). The traits that were not part of these clusters are also reproduced here for purposes of comparison (Figs. 1, 2). To evaluate the criteria by which the clusters were created we compared both the features of the traits in a cluster and the Gestalts of the traits (Figs. 5, 6) to these same elements in those traits that were not categorized (Figs. 1, 2).

In the case of the simple traits, one of the clusters must have been created based on the possession of a single analytic feature (Fig. 5B). The only feature that unites these cotyledons is the apical notch, and all the cotyledons with a notch occur in this cluster. A second cluster is based on multiple of analytic features (Fig. 5C). This cluster contains the cotyledons that are asymmetric, convex, and crenulate. The Gestalts of these cotyledons are relatively dissimilar. Two other clusters could have been created based on either a combination of analytic features or similar Gestalts. One cluster contains all of the cotyledons that are symmetric, convex, and crenulate (Fig. 5A). The other contains all of the cotyledons that are broadly obdeltoid, symmetric, and crenulate (Fig. 5D). The remaining clusters cannot be explained based on the possession of unique features or Gestalts (Figs 5E–5P). Any combination of features that could be used to define these clusters would result in the inclusion of other cotyledons that were excluded by at least one group of subjects (compare Figs. 1, 5).

TABLE 6. Consistency indices.

	Condition	G1	G2	G3	G4	Mean	t-test	Mann-Whitney U
CI	1	0.13	0.2	0.15	0.16	0.16	p = 0.008	p = 0.0145
	2	0.24	0.34	0.34	0.23	0.288		
RCI	1	0.03	0.07	0.08	0.07	0.059	p = 0.008	p = 0.0145
	2	0.1	0.13	0.13	0.09	0.112		

In the majority of the clusters in Condition 2 all three drawings were similar among the triplets (Fig. 6A–6O), while six of the triplets were similar in at most two of their constituent drawings (Fig. 6P–6U). Because of their composite nature, shape terms are more difficult to apply to these traits. Seventeen of the 21 clusters on the Adams tree lacked any single feature by which they might have been created (Fig. 6A–6L, 6O–6S). In these cases, it seems likely that the clusters were constructed based on an assessment of the forms of at least two of the drawings. We discuss these possibilities more fully below, in the context of the findings of cognitive psychology on featural and configural processing (see “General Discussion”).

The results of this study show that the use of more complex data yields character states that are more precisely defined, have greater intersubject agreement, and are more consistent with an independently derived phylogeny. The next study demonstrates that the use of complex characters, in the form of global ratings of plant form, has practical utility in the study of hybrids.

DUBAUTIA HYBRID RATING

Background

Dubautia ciliolata (DC) Keck and *D. arborea* (A. Gray) Keck are closely related species within the Hawaiian silversword alliance (Asteraceae, subtribe Madiinae), and are the only *Dubautia* spp. endemic to the island of Hawaii (Carr, 1985; Baldwin and Sanderson, 1998). In spite of their close relationship, the two species differ markedly in appearance. *D. ciliolata* is a low shrub

FIGURE 5 IS OMITTED FROM THIS FORMATTED DOCUMENT

FIGURE 6 IS OMITTED FROM THIS FORMATTED DOCUMENT

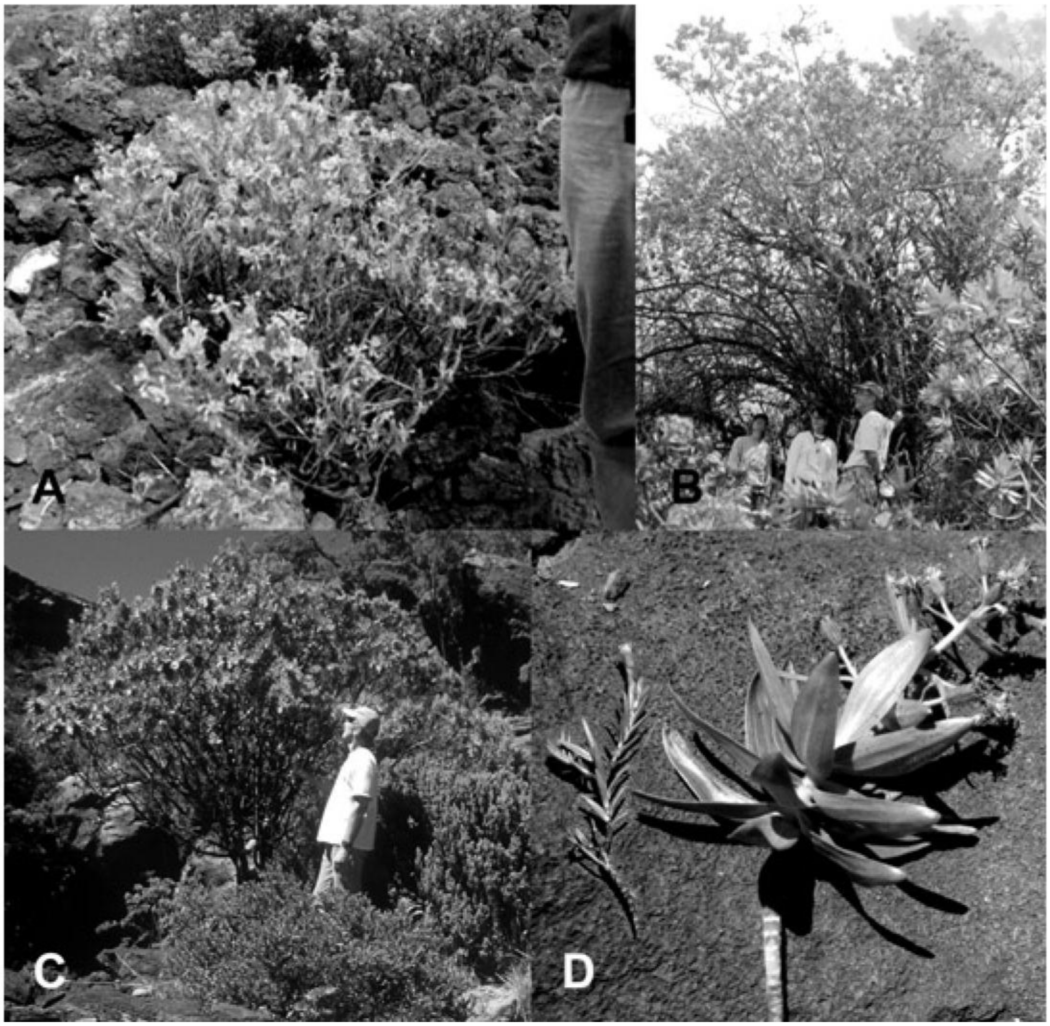


FIGURE 7. (A) *Dubautia ciliolata*, lower slopes of Mauna Loa, Hawaii. (B) *Dubautia arborea*, Pu'u La'au, Hawaii. (C) Natural *Dubautia arborea* x *ciliolata* hybrid, Waipahoehoe Gulch, Hawaii. (D) Typical *Dubautia ciliolata* (left) and *D. arborea* (right) flowering shoots, showing differences in leaf size, leaf shape, and capitulescence architecture.

seldom much more than 1 m tall, with small lanceolate leaves (<20 mm long), ca. 5 flowers per head, and usually only 1–5 heads per capitulescence (Fig. 7A). By contrast, *D. arborea* is a large shrub or tree up to 6 m tall, leaves >40 mm long, heads with 15–25 flowers, and 7–20 heads per capitulescence (Fig. 7B) (Carr, 1985).

D. ciliolata and *D. arborea* typically occur in distinct but often contiguous high-elevation woodland and shrubland habitats, respectively, on Mauna Kea (Big Island of Hawaii), but occasionally overlap and hybridize in narrow gulches. A spectacular hybrid zone occurs in Waipahoehoe Gulch, with hundreds of hybrid plants that span the entire morphological range between the two parental species with various combinations of intermediate characteristics (Carr, 1985). The overall objective of the larger research project from which this study is extracted is to evaluate associations between plant morphology and genetic polymorphisms in order to study the genetic basis of trait evolution. This requires morphological descriptions of 250 reproductively mature natural hybrid plants from Waipahoehoe Gulch, probably representing a combination of F1, F2, backcross, and advanced- generation hybrids, for which tissue has been collected for molecular analysis. This requirement provided a convenient opportunity to test the utility of complex data in an applied context.

Methods

Measurements of the 250 Waipahoehoe Gulch plants were taken on several traits that tend to distinguish the vegetative morphology of *D. ciliolata* and *D. arborea*, including length and width measurements of three randomly selected leaves, plant height, crown width, length of the terminal 20 internodes on a vigorous upper crown shoot, and branching order in the apical 30 cm of the selected shoot (Appendix 2; available at http://systematicbiology.org). Branching order refers to the number of orders of branches on a shoot, and serves as a measure of branch compactness. For example, a shoot with no lateral branches in the apical 30 cm has a branching order of 1. A shoot with second order lateral branches that are themselves unbranched has a branching order of 2, and so on.

Size variation, which can be substantial between young and old *D. arborea*-like individuals, was reduced by selecting only reproductively mature plants. Nevertheless, there was still considerable apparent age-related variation in leaf size, branching habit, and internode length. In many cases, leaf size, internode length, and compactness of branching also varied greatly within individual plants. On some plants, all of the vigorous mature shoots in the upper and outer crown were flowering, so vegetative shoots had to be selected from some of the taller “sucker” shoots emerging from within the crown. These shoots tended to have longer internodes, larger leaves, and less lateral branching than more mature shoots.

In addition to the measurements, each plant was rated on a 1-to-5 scale based on its overall appearance (or Gestalt) with respect to the form of the parental species. This rating system was developed largely out of a concern that morphological measurements alone could not adequately characterize the morphology of the hybrid plants, due to differences in age, microsite conditions, and vigor. A rating of 1 signified a plant with a completely *D. ciliolata*-like appearance, and a rating of 5 corresponded to a plant with a completely *D. arborea*-like appearance. Ratings of 2–4 were used for plants with varying degrees of intermediate characteristics, with a 3 rating denoting plants that showed about equal degrees of *D. arborea*-like and *D. ciliolata*-like appearance (Fig. 7C). Because this rating system is based on the Gestalt of the plant and not on their isolated features, intermediate plants with somewhat different appearances could have similar ratings. For example, plants that are relatively compact like *D. ciliolata*, but have stout *D. arborea*-like leaves and stems, might receive a 3 rating, but so would plants that were completely intermediate in all traits.

Four observers rated each hybrid plant individually. Each of the four individual ratings was recorded and averaged to obtain a composite rating. On a prompt from the recorder, the observers gave their ratings more or less simultaneously using hand signals. The four individual ratings were not always independent. Individual observers sometimes hesitated if they were uncertain about what rating to give a particular plant, and could be influenced by the other observers’ ratings. In practice, this was sometimes used to deliberately “shade” the composite rating up or down to reflect the observer’s uncertainty. For example, if an observer was undecided between a 3 and a 4 rating, he or she might give the plant a rating of 3 if the previous observers scored the plant a 4, or vice versa. Occasionally, there was discussion of why observers chose particular ratings. In a few of these cases, observers decided to modify their ratings to reflect either agreement or disagreement with others. However, in spite of the complexity of trait combinations in the hybrids, the individual ratings never differed by more than 1 point on any of the 250 plants that were scored. This level of agreement occurred in spite of the fact that different observers had contrasting research interests related to different aspects of plant morphology (see Stevens, 2001, and Kirchoff, 2001 for a discussion of how this might have affected the ratings).

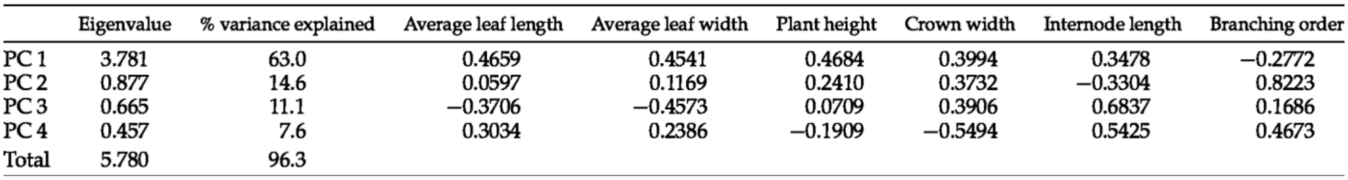
In order to summarize the morphological measurements, a principal components analysis (PCA) on the morphological measurements was conducted using PROC PRINCOMP of SAS (SAS Institute, 1999). The first four principal components were used. Correlations between the traits and the Gestalt rating, and between the first principal component and the Gestalt rating, were evaluated using PROC CORR of SAS.

Results and Discussion

The first principal component (PC 1) explained 63% of the total variation in the six traits (eigenvalue = 3.78, Table 7), and was strongly correlated with the Gestalt ratings (r = 0.902). This correlation demonstrates that the Gestalt ratings are as useful as more standard methods for estimating the hybrid status of individuals.

Investigating the correlations between the variables that comprise this principal component and the Gestalt ratings can provide insight into how these variables contribute to the overall correlation. Leaf length and width are almost as highly correlated with the Gestalt rating as is PC 1 (r = 0.897, r = 0.882), while plant height is only a bit less so (r = 0.809). Crown width and length of the terminal 20 internodes are less strongly correlated with the Gestalt rating (r = 0.608, r = 0.511), while branching order has a relatively weak correlation (r = − = −0.495). The strength of the correlations between the leaf measurements and the Gestalt ratings may reflect the fact that the difference in leaf size between *D. arborea* and *D. ciliolata* is large relative to intraspecific variation (Fig. 7D). Consequently, leaf length and width should be good indicators of species identity, and can be expected to show high correlations with other measures of identity such as the Gestalt ratings. Plant height is also indicative of species identity, and can be expected to show high correlation with the Gestalt ratings when measurements are taken from individuals that have grown to near their maximum size. On the other hand, including measurements of crown width, internode length, and branching order may have added confounding effects not related to species differences. These confounding effects

TABLE 7. Principal component eigenvalues and weights for the first four (of six) principal components.



	Eigenvalue	% variance explained	Average leaf length	Average leaf width	Plant height	Crown width	Internode length	Branching order
PC 1	3.781	63.0	0.4659	0.4541	0.4684	0.3994	0.3478	−0.2772
PC 2	0.877	14.6	0.0597	0.1169	0.2410	0.3732	−0.3304	0.8223
PC 3	0.665	11.1	−0.3706	−0.4573	0.0709	0.3906	0.6837	0.1686
PC 4	0.457	7.6	0.3034	0.2386	−0.1909	−0.5494	0.5425	0.4673
Total	5.780	96.3						

may help explain why the correlation between PC 1 and the Gestalt ratings is not much higher than the correlations between the leaf size and the Gestalt ratings.

It is also possible that the range of phenotypes in intermediate individuals was greater for aspects of crown shape than for leaf size. Different combinations of features (mosaics) were visually apparent in these individuals, which may have been due to the presence of advanced generation hybrids with variable proportions and assortments of *D. arborea* and *D. ciliolata*-derived alleles. The single numerical scale we used to rate the hybrid phenotypes ignored these multiple combinations of features.

GENERAL DISCUSSION

The results of our study on characters states show that the use of more complex traits (triplets) results in character states with smaller clusters, better intersubject reliability, and better agreement with an independently derived phylogeny. With simple traits, our results parallel those of Gift and Stevens (1997) with qualitative data. They found little intersubject agreement when they asked subjects to divide graphical data into distinct clusters. Using simple traits, we found that our subjects likewise did a poor job of creating similar clusters. One of the strongest expressions of this fact is that there are no clusters in common between any pair of subjects when they grouped simple traits (Table 5).

The correlation between the matrix of pairwise inter- subject correlations and the structure matrix (Tables 2, 3) shows that the improvement we found with complex traits can be directly related to the presence of the additional drawings. This improvement can thus be related to the complexity of the traits. Insight into the nature of this complexity can be gained from McShea's (1996) classification of types of complexity. He bases this classification on three dichotomies: hierarchical versus nonhierarchical, object versus process, and differentiation versus configuration. Combinations of these terms give eight categories of complexity. Our triplets exhibit differentiated, hierarchical, object complexity. To understand this category, we must first understand the nature of a part.

A part can be defined as an integrated pattern of objects or processes that is informed by both higher-level boundary conditions and lower-level "initiating conditions" or substructure (Salthe,1985:125). A part is thus the middle member of a hierarchical triadic system. McShea (2001), and McShea and Venit (2001) gave a similar definition when they defined a part as a system that is internally integrated and isolated from its surround. Here, the "surround" is the next higher level in the hierarchy. The lower hierarchical level is implicit in this definition.

In our triplets, the individual drawings are parts because they are integrated objects nested within a triplet, and are composed of various features that serve as their substructure (representations of teeth, veins, etc.). The features that compose the substructure are not parts because they themselves have no lower-level substructure (no initiating conditions). This is an artifact of using drawings. Real organs would of course have substructure (veins, cells, etc.). The triplets themselves are also not parts because they have no "boundary conditions." No higher level of organization exists above the triplet. The triplet itself is a whole, of which the individual drawings are parts. This is one reason that we feel justified in calling the triplet a trait of a taxon, even though they do not occur in this form on any plant.

Returning to McShea's (1996) classification, our triplets are composed of objects (drawings). They are arranged in rows with little or no configuration, and therefore are "differentiated." They also have some hierarchical structure, though this structure is constructed, not inherent in the organism from which the parts are taken. Still, from the subjects' standpoint, hierarchical structure is present. These features tell us that the triples have differentiated, hierarchical, object complexity. We will use the understanding we have gained from this analysis in our discussion of the improvement we found with complex traits.

We begin an exploration of this subject by dispensing with one possible interpretation of our results. It is possible that the subjects produced the complex clusters based on gaps, or partial gaps, in the pattern of variation of the individual drawings (Fig. 8). For instance, a cluster of complex traits may have been created because there was a clear distinction in bract shape between the triplets included in and excluded from the cluster (Fig. 8). This production method would explain the smaller, more precisely defined clusters that we found in Condition 2. However, for this to be the production method there would have to be gaps in the variation of at least one of the members of the triplet at the boundaries of the clusters. We were unable to find any such gaps in our clusters (Fig. 6). In the majority of cases, all three members of the triplets were similar (Fig. 6A–6O), and there are no gaps in the variation of any of the drawings at cluster boundaries. For instance, obdeltoid bracts are relatively uncommon in the traits. One cluster (Fig. 6H) could have been created based on this bract shape. However, very similar bracts also occur in other clusters (Fig. 6L), and in some of the uncategorized triplets (Fig. 2, first row). Continuous variation can also be documented for most of the other clusters (compare Figs. 2, 6). In only two cases does one of the component drawings show the type of gap in variation necessary for this production method to work (Fig. 6M, 6N; bracts). Yet even in these cases, all three drawings are similar across the cluster, so they could have been created based on similarity in all drawings.

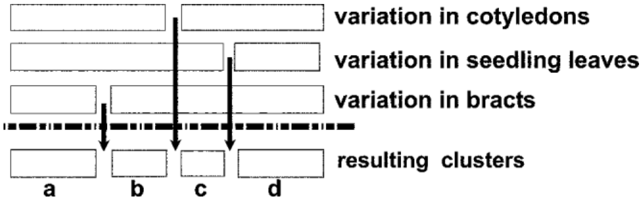


FIGURE 8. Partitioning of the variation of the complex traits based on variation in the individual drawings. The pattern of variation shown in Fig. 6 demonstrates that this type of partitioning does not occur. a, b, c, d, = arbitrary labels of clusters.

Because most of the complex clusters lack any obvious distinctive features, it is natural to speculate that the subjects were attending to some aspect of the overall form of the triplet. There are two ways that this could happen, corresponding to two cognitive processing modes for visual stimuli. Featural or analytic processing involves selectively attending to distinct, separate features of a stimulus (e.g., the nose, eyes, and mouth of a face; Brooks, 1978). Configural or holistic processing involves attending to aspects other than those of the primary features. These aspects may be global features such as the Gestalt (Goldstone, 2000; Jacoby and Brooks, 1984) or second- order relational features such as the distance between primary features (Diamond and Carey, 1986; Tanaka and Sengco,1997; Wells and Hryciw,1984).

Previous work by biologists on the intuitive classification of shapes has been carried out under the implicit assumption of featural processing. Sokal and Rohlf (1980) studied classifications of Caminalcules by 22 subjects ranging from professional systematists to children. Subjects created classifications and indicated whether their classification was based on phenetic, cladistic, or phyletic principles of classification. They found that both those intending cladistic and phyletic schemes actually came considerably closer to a phenetic standard than a cladistic standard. From this they concluded that human taxonomic judgment reflects a subjective, multivariate analysis of visually striking features. This is the definition of featural processing (Brooks, 1978). To our knowledge, the use of configural processing in making similarity judgments has not been explored in the biological literature. However, Moss (1971) compared intuitive and numerical similarity judgments of Caminacules, and found that close-relative similarities tended to be less variable when obtained intuitively than numerically. This finding is consistent with the use of a different (configural) processing mode to assess close versus distant relative similarities. Also related to this topic is Anderson's (1951) work on introgression, which we discussed in the context of our hybrid rating study, below.

At present, the factors that determine whether a stimulus is processed analytically or configurally are not completely known. However, research suggests that expertise with the stimuli affects the way that they are processed (Gauthier et al., 1998; Gauthier and Tarr, 2002; Goldstone, 2000; Takane and Sergent, 1983). In particular, configural processing of stimuli develops as a result of experience. For example, most people have consid erable experience in identifying faces and process these stimuli configurally. Dog experts have considerable experience judging dog forms and process photographs of dogs configurally (Diamond and Carey, 1986).

Because configural processing is typically seen in experts, it is possible that our subjects (all trained systematists) were able to process the triplets configurally. This would account for the fact that all three drawings are similar in most clusters of triplets (Fig. 6). Direct evidence for this supposition would come from experiments in which the subjects are shown inverted or distorted triplets, and their recognition ability assessed by measuring their fidelity and response time to questions such as "Have you seen this stimulus before?". Inversion and distortion of stimuli have been shown to disrupt configural processing, but have no significant effects on analytic processing (Rhodes et al., 1993; Young et al., 1987; Farah et al., 1995; Tanaka and Sengco,1997; Diamond and Carey, 1986).

A second feature that may stimulate configural processing concerns the nature of the stimulus. If the stimulus possesses analytic features that allow it to be easily recognized and categorized, it is likely to be processed analytically (see Spalding and Murphy, 1996, for a review of the psychological literature). Configural processing is stimulated when configural features are distinctive and analytic features are not (Diamond and Carey, 1986).

The use of analytic processing is consistent with two of the clusters produced in Condition 1 (Fig. 5B, 5C). The cotyledons in these clusters all have one or more unique analytic features and have Gestalts that are relatively dissimilar. Two additional clusters could have been created with either analytic or configural processing (Fig. 5A, 5D). Although the cotyledons in these clusters have a number of unique features, they also have similar Gestalts. The remaining clusters lack unique distinguishing features or Gestalts. We are unable to determine if they were produced with analytic or configural processing.

In most Condition 2 clusters, the corresponding individual drawings of each triplet are similar. Thus, the processing method used to produce them must have taken all of the drawings into account (Fig. 6). This could have happened with either analytic or configural processing.

In order for analytic processing to be used, the complex traits (triplets) must possess analytic features. These features only exist to the extent that there are parts, or part-like elements, within a trait. Considered as a whole, the complex traits have Gestalts, but do not have features. The same is true for the individual drawings that compose a trait. Considered as a unit, the individual drawings have Gestalts, but not features. The features only appear when we make the conceptual shift from seeing a drawing as a whole to seeing it composed of elements (Bortoft, 1996:61–68). Although these features are not parts (see above), they form the substructure of the drawings, and are therefore part-like. To compare the triplets analytically, we must look at these features, which only exist at the lowest hierarchical level of the triplet. When we do this, we can compare the drawings on a feature-by-feature basis. If the drawings were compared in this way, similarity across all three drawings of a triplet would result from sequential comparisons of the features, one drawing at a time.

If the subjects used configural processing to create the clusters, then they would focus either on the trait as a whole or on the individual drawings of the trait, each considered as a whole. In this perceptual mode they would have access to the configural features that exist at the two upper hierarchical levels. This information is not accessible to analytic perception. Attention to the Gestalts of the triplets produces initial similarity assessments. Attention to the individual drawings allows the subjects to check their assessments of the whole triplet, one drawing at a time. If they shifted hierarchical levels to the features of each drawing, they would have to shift perceptual modes, for only analytic features exist at the lowest level. Shifting to analytic perception would allow them to process the features of the triplets. If they did not shift to analytic processing, we would expect to see variation in the features coupled with a general uniformity of Gestalts within a cluster. The subject would have compared the Gestalts of the triplets and the individual drawings, but not the features of the drawings. Inspection of the complex clusters shows that many of their traits do have similar Gestalts, but variable features (Fig. 6). For instance, the spatulate cotyledons in some clusters have similar shapes but quite distinct projections at their bases. The projections can point downward, to the side, or horizontal, all in one cluster (Fig. 6E). This suggests that the subjects did not use these features in creating their groups.

The hypothesis that the complex traits were grouped with the help of configural processing suggests that use of this processing mode will result in the production of characters that have higher intersubject agreement, and better concordance with an independent assessment of phylogeny. This is possible because the use of complex traits provides configural information (triplet Gestalts) not available in simple traits. Configural processing gives the subjects the option of processing the Gestalts at two hierarchical levels. Analytic processing only allows access to one level.

If configural processing allows greater intersubject agreement in grouping tasks, it may be significant that many of the techniques used in creating character states (groups) encourage analytic processing. For instance, it is standard practice to extract small parts from organisms in order to create characters (Kirchoff, 2001; Fristrup, 2001; Hennig, 1950/1966). Removing these parts from the organism removes much of the context in which they appear—especially the Gestalt and configural aspects of features, whose presence provides the raw material for configural processing. The use of more complex features to create characters and character states allows the possibility of configural processing. The use of stimuli with little configural information (graphs) may be why Gift and Stevens's (1997) results showed so little intersubject agreement.

An additional characteristic that distinguishes featural and configural processing is the verbalizability or reportability of the qualities of the stimulus. When people process a stimulus analytically they are typically able to report quite clearly and accurately the features to which they are attending. The features are identifiable and verbalizable. In contrast, when people use configural processing, they typically cannot easily or accurately report the features to which they are attending. In fact, forcing subjects to verbally describe a stimulus that they normally process configurally impairs subsequent recognition of that stimulus (Schooler and Engstler-Schooler, 1990). This non-verbalizability does not mean that configural processing lacks precision. Numerous studies show that stimuli that are normally processed configurally (e.g., faces) are more accurately recognized in a context that encourages configural rather than analytic processing (McDonald, 2001). This suggests that configural processing is precise, even though it is not verbalizable.

The non-verbalizability of configural processing is relevant to character state description because, in most cases, the process of going from uncategorized features to character states is mediated by the development of a verbal description, often a single word. Verbal descriptions of this nature are inherently analytic. They restrict the features of the stimulus that receive attention (Kirchoff, 2001). They also exclude configural features of the part in relationship to the whole organism, and even exclude some features of the part itself. For instance, some of the cotyledons used in our experiments were termed spatulate by Thiele and Ladiges (1996), but this term does not capture the shapes of the small projections at the bottom of each cotyledon (Figs. 1, 5). Drawings or photographs provide much more accurate representations of morphological characters (Kirchoff, 2001).

Turning to the literature, Donoghue and Sanderson (1994) carried out research similar to that reported here. They postulate that the power of similarity assessments to reveal true evolutionary similarity (secondary homology) varies with the complexity of the character being investigated. Thus, more complex characters should be amenable to more stringent similarity tests, and should show greater agreement with phylogenies supported by other characters. They tested this hypothesis by reanalyzing data that has traditionally been viewed as simple (characteristics of hairs) and complex (flower characters) from 10 published papers. Their results were consistent with the prediction, but were not statistically significant. They speculate that the lack of statistical significance may be due to a failure of their categories (hairs versus floral) to correspond to differences in complexity (see Donoghue and Sanderson, 1994, for a fuller discussion).

The difficulty in creating good character states is in striking contrast to the reliability with which homologies (i.e., the groups we call "characters") are established. Part of the reason for this may be that many of the criteria for homology assessment simulate configural processing. Remane (1952) developed what is perhaps the most influential set of criteria for the assessment of homology.

Of his three main criteria, two depend on configural information. His first criterion, and the one that has been applied most extensively, is similarity in position. This criterion is configural because position can only be established based on the relationship of the part to the rest of the organism. For instance, Rieppel (1994) used the example of the zeugopodium (the paired radius/ulna of the arm, or tibia/fibula of the leg) to illustrate this point. The zeugopodium is defined as the paired elements located distal to the unpaired proximal element of the stylopodium (humerus or femur). This is a con- figural definition. To locate the bones comparable to our radius and ulna in a previously unstudied organism, we look for paired bones that lie distal to a single bone in the limb.

Remane's (1952) second criterion is also dependent on configural information, but slightly less so than his first. This is the criterion of linkage with intermediate forms. Two structures are similar if they can be placed into a transformation series with other elements such that there is a gradual transformation in form between the elements of the series. The identification of structures that belong to the series depends on comparing each new element with those already in the series. These comparisons may be made based on some special (analytic) feature of the elements (the third criterion), or they may be based on their Gestalts. For instance, most of the triplets used in Condition 2 lack analytic distinguishing features. If they were to be placed in a transformation series, it would have to be on the basis of their Gestalts.

Remane's third criterion is the possession of some special feature that makes the part distinctive. This is an analytic criterion. The possession of the long acuminate tips of some bracts is an example of this type of special feature (Fig. 6N). This tip makes these bracts distinctive, and suggests that they have a common evolutionary origin.

The character concept developed here has important consequences for character description. Conceptualizing character states as clusters of complex traits allows us to work with characters that include a portion of their context. Naturally complex traits are those that contain at least three hierarchical levels, and which, unlike the traits used in our experiments, include the relationship between these levels as they exist in the organism. The middle level is what is normally called a character (e.g., upper lemma muticous or awned, caryopsis compressed dorsiventrally or laterally, etc.; Zuloaga et al., 2000). Note that this hierarchical level can be described, but it cannot be pictured. Any picture would show more than the presence or absence of an awn, or the compression of the caryopsis. It would also show the precise shape of the lemma apex, or the shape of the caryopsis. Pictures represent all three hierarchical levels, not just the middle level.

The upper hierarchical level includes the context of the feature. The inclusion of this level allows the complex trait to be represented pictorially. At this level, the armature of the lemma would be represented as a picture of the lemma, the compression of the caryopsis as a picture of the fruit. At the lower level, the trait has part- like substructures. These substructures would include specific aspects of the shape of the lemma or caryopsis: whether the tip of the lemma is acute or obtuse, the ornamentation of the caryopsis, etc. These aspects would be at least partially represented in the picture. For instance, a photograph of a caryopsis might show its ornamentation. The cellular composition of these parts-like features is not an aspect of the complex trait. Cells occupy an even lower hierarchical level, and would be included only in complex traits whose middle levels focused on features such as ornamentation, or on anatomical traits such as presence or absence of Krantz anatomy (Zuloaga et al., 2000).

Once complex traits are defined, placing them into clusters based on similarity criteria creates character states. These states are hypotheses of primary homology. This method allows the creation of hierarchical clusters that, if fully resolved, would have very high information contents. Hierarchical clusters of this nature can be represented as step matrices for phylogenetic analysis (Swofford and Begle,1993:15–18). Coding the characters in this way should remove the coding distortions demonstrated by Hawkins et al. (1997) and Hawkins (2000).

In the hybrid rating study, the observers originally decided to collect Gestalt ratings because of the lack of any consistently reliable diagnostic features that could be used to determine the status of the hybrids. The primary existing classification of the Hawaiian Madiinae (Carr, 1985) makes extensive use of multiple morphological features at the generic, specific, and subspecific levels, contributing in part to the set of vegetative characters measured and reported here. However, diagnostic features that would potentially be useful for evaluating the degree of hybrid introgression between *D. arborea* and *D. ciliolata* were confounded by non-genetic sources of variation (see “Methods”). Such variation is perhaps unavoidable when measuring plants in natural populations, but it introduced discrepancies in measurements from plants that were judged similar by their Gestalt ratings.

Several of the measured traits show variation related to microsite (e.g., some plants were growing out of cliff faces) or damage such as partial uprooting or breakage, as well as variation due to the age of the plant. Discrepancies due to these factors were more likely to influence the trait measurements than the composite ratings. For example, observers had little difficulty distinguishing a very young, and thus small, *D. arborea*-like plant from a mature *D. ciliolata*-like plant. Thus, the ratings were consistent across observers even though various combinations of individual features were present on different intermediate plants. The observers noted the occurrence of these different features, but were able to consistently interpret the whole plants’ phenotypic distance from the two parental types. The individual observers’ ratings never differed by more than one rating point on any of the 250 hybrid plants. Consequently, Gestalt ratings may be more robust to the effects of plant age and environmental variation than are phenotypic measurements, resulting in a character with reduced non-genetic sources of variance. This in turn would result in higher heritibilities for genetic studies and more reliable discrimination of *D. arborea*-like and *D. ciliolata*-like phenotypes. On the other hand, individual morphological measurements and their principal components are likely to be more useful for studies that estimate the magnitude and trait specificity of individual loci, and for characterizing the variability of intermediate phenotypes.

Different aspects of the phenotype that vary somewhat independently of each other in the hybrid plants, resulting in a mosaic of recombinant phenotypes, are not well addressed by a simple linear Gestalt rating scale such as the one used here. Alternative Gestalt approaches that group the various intermediate phenotypes based on perceived similarity of complex features may be able to overcome this deficiency, and would benefit from further study. A more rigorous evaluation of the efficacy of alternative Gestalt vs. metric classifications would be possible with artificial hybrids of known genetic configurations growing in controlled environments.

The results of this study demonstrate that Gestalt assessments of plant form can be as useful as more conventional assessments (i.e., PCA) as a means of determining a plant's hybrid status. In combination with the character state study, this result suggests that observers can reliably assess hybrid status in the field. This conclusion is similar to that of Anderson (1951), Anderson and Gage (1952), and Anderson and Rudolf (1956). Anderson (1951) found that he was able to detect the results of hybridization in the field, even in organisms with which he was previously unfamiliar. He substantiated these assessments through morphometric measurements and ordinations (Anderson and Gage, 1952; Anderson and Rudolf, 1956).

Even more important for our purposes is Anderson's (1951) analysis of how he is able to make these assessments. In brief, he found that hybrid populations show "discordant" variation, while the variation in true species is "concordant" (Anderson, 1951). The difference between these two types of variation has to do with what Anderson (1951) calls the "trend in proportion" of corresponding parts. For instance, the leaves of both *Prunus virginiana* and *P. serotina* vary in size. In *P. virginiana* the longer leaves are also correspondingly broader, while in *P. serotina* the longer leaves are only slightly wider. These two species have different trends in their length/width ratios. In concordant variation, variation in these trends between individuals is no greater than the variation within the same individual. In discordant variation, there is more variation between than within individuals. Because of this, one receives a "harmonious" impression when viewing a population with concordant variation. Hybrid populations do not give this impression. "Variation is no longer concordant, it is DISCORDANT" (Anderson, 1951:139; emphasis in original).

Both concordant and discordant variation result from variations in the relationship between features. Subjects using configural processing have been found to attend to these types of details (Diamond and Carey, 1986; Tanaka and Sengco, 1997; Wells and Hryciw, 1984). Additional evidence that Anderson (1951) was using configural processing comes from the fact that he was initially unable to articulate the means by which he recognized the hybrid populations. He dismissed his ability to recognize hybrid populations until he was able to explain it. Configural processing is non-verbalizable (Schooler and EngstlerSchooler,1990).

We do not have any direct evidence as to the processing mode used by the observers in our study. The mere fact that the ratings were "intuitive" does not mean that they were made with configural processing. The fact that there are high correlations between leaf length, leaf width, plant height, and the Gestalt ratings suggests that the observers may have been attending to a combination of these features. If attended to separately, they would have used analytic processing. If attended to as part of the overall configuration of the plants, configural processing would be indicated. Lower correlations between the Gestalt ratings and crown width, length of the terminal 20 internodes, and branching order suggest that the ratings did not take all aspects of plant form into account. These facts favor an analytic processing mode. However, it is also possible that these lower correlations merely reflect the greater amount of non-genetic variation in these three measurements, and that the observers were able to see beyond this in making their ratings.

Regardless of the processing mode, our studies indicate that complex data are more useful than simple data in the construction of character states, and as useful as analytic data in determining hybrid status. More complex data provide more information for the creation of character states. States created with this type of data are more precisely defined, show greater similarity across investigators, and have greater agreement with an independent assessment of phylogeny. In assessing hybrid status, Gestalt ratings of plant form can be as reliable as principle component analysis. Complex data should be used more often in these types of studies.

ACKNOWLEDGMENTS

The authors thank Robert Robichaux for suggesting the hybrid rating study, and Susan McDonald for help in interpreting the literature on analytic and configural processing. The observers in the hybrid rating study were Robert Robichaux, Amy Lawton-Rauh, Sola Halldorsdottir, and David L. Remington. Portions of this research were supported by a grant from the Rudolf Steiner Foundation, and a Research Fellowship from the Royal Botanical Gardens, Sydney (both to the first author), and by NSF grant BCS-9975198 (to the fourth author). Special thanks are due Peter Weston for hosting Dr. Kirchoff's stay at the Royal Botanical Gardens, and to the members of the Character Discussion Group, who served as subjects in the character state experiments. The hybrid rating study was supported, in part, by NIH Individual Postdoctoral Fellowship 5F32GM20554 to DLR, and NSF grant DEB9708540 to Michael Purugganan and Robert Robichaux.

REFERENCES

Adams, E. N. III. 1972. Consensus techniques and the comparison of taxonomic trees. Syst. Zool. 21:390–397.

Anderson, E.1951. Concordant versus discordant variation in relation to introgression. Evolution 5:133–141.

Anderson, E., and A. Gage. 1952. Introgressive hybridization in Phlox bifida. Am. J. Bot. 39:399–404.

Anderson, E., and E. D. Rudolph. 1956. An analysis of variation in a variable population of Cladonia. Evolution 10:147–156.

Baldwin, B.G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. 95:9402–9406.

Bortoft, H.1996. The Wholeness of nature. Lindisfarne Press, Hudson, NY.

Brooks, L. R. 1978. Nonanalytic concept formation and memory for instances. Pages 169–211 in Cognition and categorization (E. Rosch and B. B. Lloyd, eds.). Erlbaum, Hillsdale, NJ.

Brower, A. V. Z., and V. Schawaroch. 1996. Three steps of homology assessment. Cladistics 12:265–272.

Campbell, J. A., and D. R. Frost. 1993. Anguid lizards of the genus Abronia: Revisionary notes, descriptions of four new species, phylogenetic analysis, and key. Bull. Am. Mus. Nat. Hist. 216:1– 121.

Carr, G. D. 1985. Monograph of the Hawaiian Madiinae (Asteraceae): Argyroxiphium, *Dubautia*, and Wilkesia. Allertonia 4:1–123. Colless, D. 1985. On "character" and related terms. Syst. Zool. 34:229–233.

Davis, P. H., and V. Heywood. 1973. Principles of angiosperm taxonomy. Robert Krieger, Huntington, New York.

de Pinna, M. C. C.1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7:367–394.

Diamond, R., and S. Carey. 1986. Why faces are and are not special: An effect of expertise. J. Exp. Psychol. Gen. 115:107–117.

Donoghue, M. J., and M. J. Sanderson. 1994. Complexity and homology in plants. Pages 394–421 in Homology: The hierarchical basis of comparative biology (B. K. Hall, ed.). Academic Press, San Diego.

Farah, M. J., J. N. Tanaka, and M. Drain. 1995. What causes the face inversion effect? J. Exp. Psychol.: Hum. Percept. Perform. 21:628– 634.

Farris, J. S. 1989. The retention index and rescaled consistency index. Cladistics 5:417–419.

Fristrup, K.1992. Character: Current usages. Pages 45–51 in Keywords in evolutionary biology (E. F. Keller and E. A. Lloyd, eds.). Harvard University Press, Cambridge, MA.

Fristrup, K. 2001. A history of character concept in evolutionary biology. Pages 13–35 in The character concept in evolutionary biology (G. Wagner, ed). Academic Press, San Diego.

Gauthier, I., and M. J. Tarr. 2002. Unraveling mechanisms for expert object recognition: Bridging brain activity and behavior. J. Exp. Psychol. Hum. Percept. Perform. 28:431–436.

Gauthier, I., P. Williams, M. J. Tarr, and J. Tanaka. 1998. Training "greeble" experts: A framework for studying expert object recognition processes. Vision Res. 38:2401–2428.

Gift, N., and P. F. Stevens, 1997. Vagaries in the delimitation of character states in quantitative variation—an experimental study. Syst. Biol. 46:112–125.

Goldstone. R. L. 2000. Unitization during category learning. J. Exp. Psychol. Hum. Percept. Perform. 26:86–112.

Good, D. A. 1988. Phylogenetic relationships among gerrhonotine lizards. Univ. Calif. Publ. Zool. 121:1–138.

Hall, B. K. 1994. Homology: The hierarchical basis of comparative biology. Academic Press, San Diego.

Hall, B. K.1999. Homology. John Wiley, Chichester.

Hawkins, J. A. 2000. A survey of primary homology assessment: Different botanists perceive and define characters in different ways. Pages 22–53 in Homology and systematics. Systematics Association Special Volume 58 (R. Scotland and R. T. Pennington, eds.). Taylor and Francis, New York.

Hawkins, J. A., C. E. Hughes, and R. W. Scotland. 1997. Primary homology assessment, characters and character states. Cladistics 13:275– 293.

Hennig, W. 1950/1966. Phylogenetic systematics. University of Illinois Press, Urbana (Translation by D. E. Rosen, G. Nelson, and C. Patterson).

Hood, G. 2002. PopTools, ver. 2.4.9. Pest Animal Control Co-operative Research Centre, CSIRO, Canberra, Australia. http://www.cse. csiro.au/CDG/poptools/index.htm

Jacoby, L. L., and L. R. Brooks. 1984. Nonanalytic cognition: Memory, perception, and concept learning. Psychol. Learn. Motiv. 18:1–47.

Jardine, N. 1969. The observational and theoretical components of homology: A study on the morphology of dermal skull-roofs of riphidistrian fishees. Biol. J. Linn. Soc. 1:327–361.

Kirchoff, B. K. 2001. Character description in phylogenetic analysis: Insights from Agnes Arber's concept of the plant. Ann. Bot. 88:1203– 1214.

Kitching, I. J., P. L. Forey, C. J. Humphries, and D. M. Williams. 1998. Cladistics. 2nd ed. Oxford University Press, Oxford.

Kluge, A. G., and J. S. Farris. 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1–32.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Mast, A., and T. J. Givnish. 2002. Historical biogeography and the origin of stomatal distributions in *Banksia* and *Dryandra* (Proteaceae) based on their cpDNA phylogeny. Am. J. Bot. 89:1311–1323.

McDonald, S. L. 2001. A competitive evaluation of the relative judgment versus context-specific processing hypotheses for the sequential superiority effect. Dissertation. The University of North Carolina at Greensboro.

McShea, D. W.1996. Complexity and homoplasy. Pages 207–225 in Homoplasy: The recurrence of similarity in evolution (M. J. Sanderson and L. Hufford, eds.). Academic Press, San Diego.

McShea, D. W. 2001. Parts and integration: Consequences of hierarchy. Pages 27–60 in Evolutionary patterns: Growth, form and tempo in the fossil record (J. B. C. Jackson, S. Lidgard, and F. K. McKinney, eds.). University of Chicago Press, Chicago.

McShea, D. W., and E. P. Venit. 2001. What is a part? Pages 259–284 in The character concept in evolutionary biology (G. Wagner, ed). Academic Press, San Diego.

Moss, W. W.1971. Taxonomic repeatability: An experimental approach. Syst. Zool. 20:309–330.

Page, R. D. M. 1993. COMPONENT: Tree comparison software for Microsoft Windows, version 2.0. The Natural History Museum, London.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci.12:357–358.

Patterson, C.1982. Morphological characters and homology. Pages 21– 74 in Problems in phylogenetic reconstruction (K. A. Joysey and A. E. Friday, eds.). Academic Press, London.

Platnick, N. I. 1979. Philosophy and the transformation of Cladistics. Syst. Zool. 28:537–546.

Poe, S., and J. J. Wiens. 2000. Character selection and the methodology of morphological phylogenetics. Pages 20–36 in Phylogenetic analysis of morphological data (J. J. Wiens, ed.). Smithsonian Institution Press, Washington, DC.

Remane, A.1952. Die Grundlagen des naturlichen Systems der vergle-ichenden Anatomie und der Phylogenetik. Gesst und Portig, Leipzig. Rhodes, G., S. Brake, and A. P. Atkinson. 1993. What's lost in inverted faces? Cognition 47:25–57.

Rieppel, O. 1994. Homology, topology, and typology: The history of modern debates. Pages 63–100 in Homology: The hierarchical basis of comparative biology (B. K. Hall, ed.). Academic Press, San Diego.

Salthe, S. N.1985. Evolving hierarchical systems. Columbia University Press, New York.

SAS Institute. 1999. The SAS System for Windows, Version 8.00, SAS Institute, Cary, NC.

Schooler, J. W., and T. Y. Engstler-Schooler.1990. Verbal overshadowing of visual memories: Some things arebetter left unsaid. Cogn. Psychol. 22:36–71.

Sokal, R. R., and F. J. Rohlf. 1980. An experiment in taxonomic judgment. Syst. Bot. 5:341–365.
Spalding, T. L., and G. L. Murphy. 1996. Effects of background knowledge on category construction. J. Exp. Psychol.: Learn. Mem. Cogn. 22:525–538.
Stevens, P. F. 1984. Homology and phylogeny: Morphology and systematics. Syst. Bot. 9:395–409.
Stevens, P. F.1991. Character states, morphological variation, and phylogenetic analysis: A review. Syst. Bot. 16:553–583.
Stevens, P. F. 2000. On characters and character states: Do overlapping and non-overlapping variation, morphology and molecules, all yield data of the same value? Pages 81–105 in Coding characters of phylogenic analysis (R. Scotland and R. T. Pennington, eds.). Systematics Association Special Volume Series 58. Cambridge, UK.
Stuessy, T. F. 1990. Plant taxonomy: The systematic evaluation of comparative data. Columbia University Press, New York.
Swofford, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 in Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford University Press, New York.
Swofford, D. L., and D. P. Begle.1993. PAUP: Phylogenetic Analysis Using Parsimony, Ver. 3.1. Illinois Natural History Survey. Champaign, IL.
Swofford, D. L. 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta for Macintosh. Sinauer Associates, Sunderland, MA.
Takane, Y., and J. Sergent. 1983. Multidimensional scaling models for reaction times and same-different judgments. Psychometrika 48:393– 423.
Tanaka, J. W., and J. A. Sengco. 1997. Features and their configuration in face recognition. Mem. Cogn. 25:583–592.
Thiele, K. 1993. The Holy Grail of the perfect character: The cladistic treatment of morphometric data. Cladistics 9:275–304.
Thiele, K., and P. Y. Ladiges. 1996. A cladistic analysis of *Banksia* (Proteaceae). Aust. Syst. Bot. 9:661–733.
Wells, G. L., and B. Hryciw. 1984. Memory for faces: Encoding and retrieval operations. Mem. Cogn. 12:338–344.
Young, A. W., D. Hellawell, and D. C. Hay. 1987. Configural information in face recognition. J. Cogn. Neurosci. 3:59–70.
Zuloaga, F. O., O. Morrone, and L. M. Giussani. 2000. A cladistic analysis of the Paniceae: A preliminary approach. Pages 123–135 in Grasses: Systematics and evolution (S. W. L. Jacobs and J. Everett, eds.). CSIRO, Melbourne.