HAWAIIAN BIOGEOGRAPHY

EVOLUTION ON A HOT SPOT ARCHIPELAGO

EDITED BY
WARREN L. WAGNER
AND V. A. FUNK

SMITHSONIAN
INSTITUTION
PRESS

WASHINGTON
AND LONDON

Technical preparatory editing by Robynn K. Shannon
Copyediting and typesetting by Peter Strupp/Princeton Editorial Associates
Cover art by Alice Tangerini and Robynn K. Shannon
Book design by Linda McKnight
Proofreading by Eileen D'Araujo

Project management by Deborah L. Sanders

Library of Congress Cataloging-in-Publication Data

Hawaiian biogeography : evolution on a hot spot archipelago / edited by Warren L. Wagner and V. A. Funk
p. cm.—(Smithsonian series in comparative evolutionary biology)
Papers grew out of a symposium cosponsored by the American Society of Plant Taxonomists and the Association for Tropical Biology in 1992.
Includes bibliographical references (p.) and index.
ISBN 1-56098-462-7 (clothbound).—ISBN 1-56098-463-5 (paperbound)
1. Biogeography—Hawaii. 2. Island ecology—Hawaii. I. Wagner, Warren Lambert. II. Funk, V. A. (Vicki A.), 1947– . III. Series.
QH198.H3H38 1995
574.9969—dc20 94-5353

British Library Cataloguing-in-Publication Data available

Throughout the text, places of deposit for plant voucher specimens are indicated by herbarium abbreviations as given in Holmgren et al. (1990).

This material is based upon work supported by the National Science Foundation under award no. DEB-9214264 to the American Society of Plant Taxonomists. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Manufactured in the United States of America

01 00 99 98 97 96 95 5 4 3 2 1

♾ The paper used in this publication meets the minimum requirements of the American National Standard for Permanence of Paper for Printed Library Materials Z39.48-1984.

3 Cladistic Methods

V. A. FUNK

The authors in this volume have used the methods of phylogenetic systematics, also called cladistics, to develop phylogenies and examine monophyletic groups (referred to as *clades*)[1], in a rigorous way. Thorough explanations of cladistics can be extremely complex. This discussion is not intended to be comprehensive; rather it is an introduction to the concepts and terminology necessary for the reader inexperienced in phylogenetic theory to understand the analytic aspects of the chapters in this volume. Additional discussions can be found in Hennig (1966), Nelson and Platnick (1981), Wiley (1981), Swofford and Olsen (1990), Wiley et al. (1991), Forey et al. (1992), Maddison and Maddison (1992), Swofford (1993), and references cited therein.

Cladistics seeks to answer the following question: Given any group of more than three taxa, which taxa are more closely related to one another than to any other taxa? Relatedness is identified by the sharing of one or more uniquely derived characters that other taxa outside the group do not possess. For example, within vertebrates the unique derived character "feathers" identifies all birds as being most closely related to each other. The branching pattern of the tree that illustrates this relatedness is formed by the distribution of the unique characters in the way that

[1]Most of the clades identified in this book do not have formal taxonomic names and, for emphasis, are given in italics and without capitalization (unless derived from a proper name).

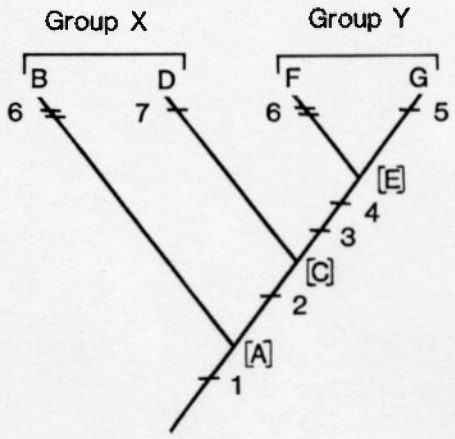

FIGURE 3.1. Cladogram. Letters represent taxa, whereas letters in brackets are hypothetical ancestral taxa. Numbers represent apomorphic characters of the transformation series; those with *single bars* are apomorphic, and those with *double bars* are independently derived. Group X is paraphyletic, a grade. Group Y is monophyletic, a clade.

requires the least amount of convergent or parallel evolution and character loss. A tree formed solely by these unique characters can be called a *cladogram* but is also called a *phylogenetic tree* or *tree* (Figure 3.1). Cladograms are characterized by the fact that their information is contained in the branching sequence and not in the physical proximity of the terminal branches. For instance, Figure 3.2 shows the same branching sequence as Figure 3.1, and as far as information content is concerned, it is identical. In Figure 3.1, B is next to D, but in Figure 3.2 B is next to F. Neither of these physical locations gives the correct relationship because the branching sequence of both figures shows that the actual relationship is one of B being most closely related to the group of taxa DGF (see discussion on Venn diagrams below). A cladogram in which the branch

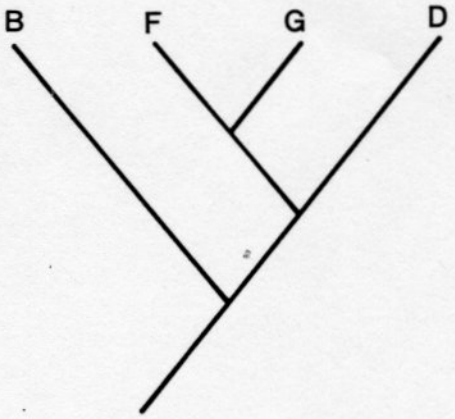

FIGURE 3.2. Cladogram with the same branching sequence and the same information content as Figure 3.1.

TABLE 3.1. Character Matrix for Figures 3.1 to 3.3

	Transformation series						
Taxon[a]	1	2	3	4	5	6	7
OG	0	0	0	0	0	0	0
B	1	0	0	0	0	1	0
D	1	1	0	0	0	0	1
F	1	1	1	1	0	1	0
G	1	1	1	1	1	0	0
[A]	1	0	0	0	0	0	0
[C]	1	1	0	0	0	0	0
[E]	1	1	1	1	0	0	0

[a]B, D, F, G, and the OG (outgroup) are actual taxa, whereas A, C, and E are hypothetical taxa whose character data are inferred from the most-parsimonious tree.

and internode lengths reflect the number of characters on that branch or internode is called a *phylogram*.

Cladistics has as its basis three concepts: apomorphy, monophyly, and parsimony. An *apomorphy* is a uniquely derived evolutionary character. Hennig (1966) called these *apomorphous characters,* but various other permutations of the term now include *apomorphic character* and *apotipic.* There are related terms; for instance, every apomorphy either is found in one taxon, an *autapomorphy* (Figure 3.1, apomorphic characters 5 and 7; Table 3.1), or is shared by more than one taxon, a *synapomorphy* (Figure 3.1, apomorphic characters 1 to 4). A synapomorphic character, in the true sense, is one that has evolved once in the ancestor of a group of taxa marking a common evolutionary history for that group. Every apomorphous character is paired with the character from which it is derived, the *plesiomorphous character* (or *plesiomorphic character* or *plesiomorphy*). In the bird example, "feathers" is the apomorphic character, and because feathers are believed to be derived from scales, then "scales" becomes the plesiomorphic character.

The apomorphic and plesiomorphic characters together form an *evolutionary transformation series* (often abbreviated TS) (Hennig, 1966; Wiley et al., 1991). The transformation series can contain more than one apomorphic character, provided they are evolutionarily homologous. Some authors refer to individual characters as *character states* and transformation series as *characters,* and both systems are used in this book. However, this alternative terminology necessitates placing apomorphic

and plesiomorphic character states into characters rather than transformation series. Unfortunately, users of the term *character state* sometimes incorrectly shift to the term *character* in the discussion section. To be unambiguous, the transformation series concept is preferred.

The terms *apomorphy* and *plesiomorphy* are dependent on their relative position on a cladogram. A character that is synapomorphous at a node when one is discussing group Y (Figure 3.1, apomorphic character 3) will be plesiomorphous if one is discussing the characters that delimit taxon G. When characters are found in more than one taxon, they are considered to be evolutionarily or phylogenetically *homologous* (Patterson, 1982, 1988). If what appears to be the same apomorphic character is found in two unrelated groups, it is considered to be nonhomologous and therefore not a single apomorphy (Figure 3.1, apomorphic character 6) and is referred to as a *homoplasious character.* If a character occurs as a synapomorphy on a cladogram and is subsequently lost in one or more taxa, then it is a *character loss* (also referred to as a *reversal,* but this term can be confused with genetic terminology). Homoplasious characters and character losses may obscure the phylogenetic pattern. These seemingly contradictory characters are referred to as *character conflict.* Such conflicts are resolved by parsimony analysis, and once they are recognized and understood, become apomorphic characters themselves.

The *parsimony* criterion governs how cladograms are constructed. It is nearly identical to Hennig's Auxiliary Principle: "Never assume convergence or parallel evolution, always assume homology in the absence of contrary evidence" (Hennig, 1966, according to Wiley et al., 1991; Farris, 1983). This principle does not preclude the possibility of convergent or parallel evolution; it simply states that when there is no reason to think otherwise, two characters that appear to be the same are treated as homologous. This means that the character has the potential for grouping taxa if it is apomorphous. When characters support conflicting groups (Figure 3.1, apomorphic character 6), the explanation that is the simplest is chosen (i.e., the one that requires the smallest number of homoplasious characters and character loss). Therefore, the user of parsimony is not making any statement about the process of evolution.

A *monophyletic group* is a group of taxa that share a common ancestor and includes all descendants of that ancestor, also referred to as a *clade.* On a cladogram, this translates into any group that includes all taxa that share at least one synapomorphy (Figure 3.1, group Y). Figure 3.3 is a Venn diagram for Figures 3.1 and 3.2; each ellipse represents

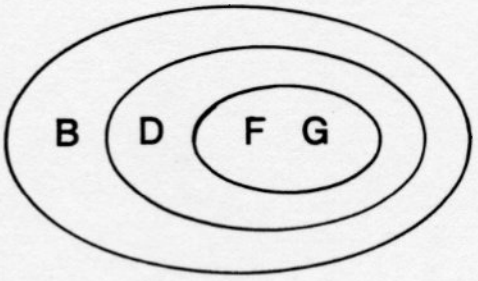

FIGURE 3.3. Venn diagram of Figure 3.1.

a monophyletic group so that one can easily see three such groups, FG, DFG, and the whole clade, BDFG. However, the concept of monophyly is far more than a definition of a group of taxa. Concomitant with it is the notion that the only groups that are evolutionarily meaningful (natural) are monophyletic ones. Therefore, in this view, the only groups that can be recognized in formal classifications are monophyletic ones. The justification for this position lies in the nature of the groups. If groups include an ancestor and all its descendants (monophyletic), then the groups reflect a common evolutionary history and can be used to study speciation, biogeography, pollination biology, and other evolutionary concepts. *Non-monophyletic* groups are of two types (Farris, 1974). In Figure 3.1, A is the ancestor of taxa B, D, F, and G. Group X contains the common ancestor A, but only two of the descendants, B and D, and so it is not monophyletic. Such a group, one that includes some but not all of the descendants (Figure 3.1, group X), is called *paraphyletic*, which is also referred to as a *grade*. *Polyphyletic* groups have been defined several ways, but, in general, they consist of taxa taken from more than one monophyletic group. Under certain circumstances, it is difficult to separate paraphyletic and polyphyletic groups, so often authors simply refer to any group of taxa that does not satisfy the criterion of monophyly as non-monophyletic.

Both monophyly and parsimony depend on apomorphous characters; therefore, apomorphies are the central concept of cladistics. The process of assigning the status of apomorphy to a character is called determining *polarity*. Using an *outgroup* (or outgroups) is the most common way of determining which characters are apomorphic (Watrous and Wheeler, 1981; Farris, 1982; Maddison et al., 1984). Characters found in the outgroup as well as in some of the taxa of the group being studied (the *ingroup*) are considered to be plesiomorphous. Those characters found only in some of the taxa of the ingroup but that are absent in the rest of

the ingroup and in the outgroup are considered to be apomorphous. Many exceptions and extenuating circumstances to be considered when using the outgroup criterion cannot be covered in this brief discussion. Additional information can be found in the general references listed in the first paragraph and in Watrous and Wheeler (1981), Farris (1982), and Maddison et al. (1984). An outgroup can be, but is not necessarily, the taxon most closely related to the ingroup, the *sister group*. In Figure 3.1, DFG is the sister group of B. The outgroup(s) should be a closely related taxon that does not contain large numbers of autapomorphous characters. Sometimes a specific outgroup cannot be identified, and a composite outgroup is constructed by evaluating each transformation series separately to determine which character(s) was apomorphic. Authors in this volume who use this approach have discussed how the composite outgroups were formed. Another method that is occasionally used to assign polarity is ontogeny (Patterson, 1982).

The process of tree construction has changed greatly in the past two decades. Instead of the manual constructing of small character trees for each transformation series, which necessitates examining each character to decide if it is apomorphous and then looking for groups of taxa that can be nested, computer programs are now used. These programs construct networks based on the distribution of shared characters without assigning polarity or evolutionary direction, then root the tree based on the characters present in the outgroup(s), either by using the outgroup(s) as part of the analysis or by attaching it to the network after the analysis is completed. The two most commonly used programs are PAUP (Swofford, 1993) and HENNIG86 (Farris, 1988). These computer programs have increased the speed and accuracy of cladogram production. Moreover these programs have introduced many options that give the user a powerful resource for investigating the phylogeny of the taxa in question. Another program available for analyzing characters and cladograms is MacClade (Maddison and Maddison, 1992), which also has a broad array of options. On occasion, different programs will give different answers to the same questions. It is the user's responsibility to make sure she or he understands and endorses the assumptions that underlie the options in all the programs; otherwise, the results will be misleading (at best) or erroneous.

For years, many phylogeneticists have tried to measure the robustness of data used to construct cladograms, to find a way to assign a value that would indicate how "robust" the cladogram was. The simplest measure is the *tree length*, or total number of *steps*. The tree length is

equal to the total number of characters actually on the tree, including all conflicting characters. The first index, and still the most popular, is the *consistency index* (CI) (Kluge and Farris, 1969). Currently, the index is calculated using only synapomorphies and taking the minimum number of steps necessary if all the data agreed and dividing it by the actual number of steps. The other commonly used index is the *rescaled consistency index* (RC) (Farris, 1989), which multiplies the CI by the *retention index* (RI; ratio of apparent synapomorphy to actual synapomorphy). The RC excludes characters that do not contribute to the "fit" of the tree by excluding autapomorphic characters as well as totally homoplasious ones. The CI and the RC can be used for each individual transformation series (character) as well as the cladogram as a whole. Several other indices have been proposed (e.g., F-ratios, d-measures) (Wiley et al., 1991) but are not used in this volume. Each index has certain strengths and weaknesses, and no one index has been found that really gives us the information we seek, the answer to the question "How good is this cladogram?"

Whereas the indices give information on the tree as a whole (or on the individual transformation series), there is another approach to estimating the value of a particular cladogram with respect to the data and that is by placing confidence limits on the individual branches. Some authors provide such values based on *bootstrapping*. This technique involves randomly sampling with replacement the character information from a data set to build many "bootstrap" data sets of the same size as the original data set, which are then analyzed to give one or more trees. The percentage of occurrences (usually out of 100) that a particular monophyletic group appears among the trees of the sample data sets can be considered an index of support for that monophyletic group. This technique does not result in true confidence limits in a statistical sense. One of the biggest problems is that the values can be related to the size of the data set. Also, it takes three synapomorphies at an internode for a confidence level of 95% to be reached, and these could all be homoplasious characters that occur many times on the tree. There are additional problems with the assumptions required by bootstrapping that can cause either over- or underestimates of confidence (for further discussion, see Sanderson, 1989).

As data sets grow, there is an ever greater chance of the analysis resulting in more than one equally parsimonious tree. A method of working with multiple trees is the implementation of consensus trees (Wiley et al., 1991; Swofford, 1993). Two types of consensus trees are

common in the literature, strict and majority rule. *Strict consensus trees* reflect only the groups that are found in all the equally parsimonious trees. *Majority rule consensus trees* show the branching sequences that are found in most of the trees. Both consensus tree methods have the potential of producing unresolved areas or branching patterns on the consensus tree that are not found in any of the equally parsimonious trees. Although consensus trees are useful in identifying the areas of agreement and conflict among the competing trees, unless a consensus tree is identical to one of the equally parsimonious trees, it cannot be used as a phylogeny beyond the point of agreement found in all trees. For instance, *polytomies* (nodes with more than two branches) that are the result of conflicting branching sequences in competing trees and are not found in any of the competing equally parsimonious trees should not be used as part of the phylogeny. One should consider selecting one of the equally parsimonious trees for use as a phylogenetic tree. Another option for dealing with multiple trees that was used in this book is *successive weighting,* an a posteriori weighting based on the fit of the characters to the trees (Farris, 1989; Swofford, 1993). There are several types of a priori weighting as well, but none were used by the authors in this volume.

When many equally parsimonious trees are produced, especially with molecular data sets, the methods of bootstrapping and majority rule consensus trees are often combined to produce a tree. Extreme caution must be used with such a tree, for there is no way to gauge what relationship it holds with any of the equally most-parsimonious trees.

Once a phylogenetic tree has been produced, one of the most interesting things to do with it is to use it to study evolution. Indeed, the ability to ask questions about evolution is why many researchers are interested in producing phylogenies in the first place. One technique used in this book to facilitate such evolutionary studies is *optimization* or *mapping.* The method is examined in detail in Funk and Brooks (1990), Brooks and McLennan (1991), and Maddison and Maddison (1992); a simplified explanation is offered here. Once a cladogram has been constructed, any feature or condition is selected to be examined in the light of the phylogeny of the group. Examples include habitat, habit, chromosome number, and home range. The condition of each terminal taxon is identified on the cladogram, and hypothetical conditions are assigned to the nodes that reflect the most-parsimonious arrangement of those conditions at each node. This allows one to determine the potential ancestral conditions. In this volume, the method is primarily used to examine

biogeography, but other features examined include speciation and habitat evolution as well as adaptive radiation and coevolution. Of all these features, only biogeography has its own special term: A cladogram in which the terminal taxa have been replaced by their respective distributions is called an *area cladogram*.

Phylogenetic systematics is an interesting, growing, and constantly changing field of study. This brief discussion is an introduction in the hopes that the reader will be able to better understand the chapters in this volume.

ACKNOWLEDGMENTS

I thank Paul Manos, François Lutzoni, Peter Cannell, and Warren L. Wagner for reading and commenting on drafts of the manuscript. Their willingness to offer criticism should not be taken to infer their acceptance of its contents.