Evolution, 52(2), 1998, pp. 486–502

BUTTERFLIES AND PLANTS: A PHYLOGENETIC STUDY

NIKLAS JANZ[1] AND SÖREN NYLIN
Department of Zoology, University of Stockholm, 106 91 Stockholm, Sweden
[1]*E-mail: niklas.janz@zoologi.su.se*

Abstract.—A database on host plant records from 437 ingroup taxa has been used to test a number of hypotheses on the interaction between butterflies and their host plants using phylogenetic methods (simple character optimization, concentrated changes test, and independent contrasts test). The butterfly phylogeny was assembled from various sources and host plant clades were identified according to Chase et al.'s rbcL-based phylogeny. The ancestral host plant appears to be associated within a highly derived rosid clade, including the family Fabaceae. As fossil data suggest that this clade is older than the butterflies, they must have colonized already diversified plants. Previous studies also suggest that the patterns of association in most insect-plant interactions are more shaped by host shifts, through colonization and specialization, than by cospeciation. Consequently, we have focused explicitly on the mechanisms behind host shifts. Our results confirm, in the light of new phylogenetic evidence, the pattern reported by Ehrlich and Raven that related butterflies feed on related plants. We show that host shifts have generally been more common between closely related plants than between more distantly related plants. This finding, together with the possibility of a higher tendency of recolonizing ancestral hosts, helps to explain the apparent large-scale conservation in the patterns of association between insects and their host plants, patterns which at the same time are more flexible on a more detailed level. Plant growth form was an even more conservative aspect of the interaction between butterflies and their host plants than plant phylogeny. However, this is largely explained by a higher probability of colonizations and host shifts while feeding on trees than on other growth forms.

Key words.—Coevolution, host shifts, insect-host plant interactions, Lepidoptera, Papilionoidea, specialization.

Received March 22, 1996. Accepted November 26, 1997.

Few systems have played such an important role in our understanding of how species interactions evolve as butterflies and their host plants. To a large extent this is the result of a single influential paper by Ehrlich and Raven (1964). Their essay inspired a flood of papers on different aspects of this association, and a number of related hypotheses on the evolution of insect-plant interactions have emerged. However, there have been few attempts to exploit the large database on butterfly–host plant affiliations to test such hypotheses using phylogenetic methods (Mitter and Brooks 1983; Miller 1987a). A major reason for this is that well-supported phylogenies, for both butterflies and seed plants, have been unavailable. However, this is slowly changing, and today it is possible to put together reasonably robust phylogenies for both groups. Chase et al. (1993) have recently published a molecular phylogeny for all seed plants, which is probably the best estimate of large-scale angiosperm phylogeny to date. Butterfly phylogenies are also emerging and we have constructed a plausible phylogeny across the butterflies by combining these published estimates.

Ehrlich and Raven (1964) argued that the patterns of host plant association that we observe today have been shaped by a stepwise coevolutionary process in which plants evolve defenses against natural enemies, and these enemies in turn evolve new capacities to cope with these defenses. Plants that escape from herbivores can diversify in the absence of enemies. Insects that eventually manage to colonize one of these plants will enter a new adaptive zone and can in turn diversify onto the relatives of this plant, because they will be chemically similar. Ehrlich and Raven argued that these processes have led to the main pattern they had observed, namely that related butterflies tend to feed on related groups of plants.

Most or all plant diversification up to the level of resolution used in our analysis had probably already taken place at the time the butterflies started to diversify. The oldest known butterfly fossil dates back to 48 M.Y.B.P. and the diversification of the butterfly families probably took place at the end of the Cretaceous, about 66 M.Y.B.P. (Emmel et al. 1992). At least some families even in the most recently derived of the plant clades used in this analysis date to this time, such as, Urticaceae (a member of Rosid 1 in Chase et al., 1993): 90 M.Y.B.P., Rutaceae (Rosid 2): 52 M.Y.B.P., Apiaceae (Asterid 2): 52 M.Y.B.P., Apocynaceae (Asterid 1): 60 M.Y.B.P. (dates from Eriksson and Bremer 1992). Therefore, the clades themselves must be even older. It is reasonable to regard the evolution of current associations as arising generally through butterfly colonization of already-diversified hosts, and that is the approach we shall take. This is not to say that coevolution is an unimportant process in the interaction between butterflies and their host plants, only that evidence for it should be sought at other levels of resolution.

There are two fundamentally different approaches to comparative analyses using phylogenetic data. One approach seeks to find and explain general ecological or evolutionary correlations (Felsenstein 1985; Grafen 1989; Harvey and Pagel 1991; Pagel 1992), while the other seeks to reconstruct and explain particular historical events or sequences of events along branches in a phylogeny (Mitter and Brooks 1983; Coddington 1988; Sillén-Tullberg 1988; Maddison 1990; Brooks and McLennan 1991). These approaches are complementary (Coddington 1994; Nylin and Wedell 1994; Pagel 1994) and we have in the present paper used both, depending on the problem.

The question of ancestral host associations is a clearly historical problem. On what plant did the first butterfly feed? This question is interesting in its own right, but answering it also provides necessary information for any phylogenetic tests regarding direction of evolution of host plant associations. Ehrlich and Raven (1964) regarded it most likely that the ancestral host plant family was Aristolochiaceae. In con-

trast, Scott (1986) noted that Fabaceae were eaten by the most basal branches of several butterfly families and suggested that the ancestral host probably was a legume. More recently, Ackery (1991) suggested that Malvales may instead be the ancestral host association.

Ehrlich and Raven (1964) noted several factors influencing the association between butterflies and their host plants, but particularly stressed the importance of the plant's secondary metabolic substances. They noted that many higher taxa of plants are characterized by distinctive secondary chemistry, and cited a number of examples in which related butterflies feed on related plants. They also cited examples where related butterflies are feeding on unrelated plants with chemical similarities. These observations are consistent with, though not sufficient to demonstrate, the importance of plant chemistry. However, nobody has tried to determine whether and at what taxonomic scales the tendency to feed on related plants is statistically demonstrable for butterflies as a whole, as opposed to selected examples.

Although plant chemistry has been viewed as the prime factor governing the evolution of butterfly–host plant associations (e.g. Feeny 1975, 1976, 1991; Jermy 1976, 1984; Scriber and Slansky 1981; Berenbaum 1983; Zangerl and Berenbaum 1993; Fiedler 1995b), other aspects of the host, not necessarily well correlated with phylogeny, might also have a large effect (Benson et al. 1975; Smiley 1978; Price et al. 1980; Courtney 1984; Bernays and Graham 1988; Anderson 1993). One example is host growth form. Different growth forms can dominate in different habitat types, which have distinct combinations of microclimate, enemies, etc. These require specialized adaptations, making it more difficult for a butterfly to colonize a new growth form and habitat than to colonize a new host plant with a different chemical composition in the same habitat. Herbaceous host plants may also require a different search behavior by ovipositing females than do arboreal hosts. Thus, host growth form might be more conserved than host clade membership, when both are traced on the butterfly phylogeny.

We might also expect that the propensity to shift between host taxa would differ between butterflies feeding on different growth forms. For example, Feeny (1976) hypothesized that herbaceous plants are defended by diverse "qualitative" toxins that require corresponding diverse physiological and behavioral adaptations in the attacking insect, while trees are characterized by "quantitative" defense consisting of a limited number of digestion-reducing agents such as tannins, which do not require specialized detoxification tactics. Under this hypothesis, trees make up a chemically more homogenous group. Host shifts should thus be easier and more common between trees than between herbs. This would influence the relationship between phylogenetic and growth form conservatism, especially if tree feeding is common among butterflies. Similar arguments have been advanced to explain an apparent association between tree feeding and polyphagy (Futuyma 1976; Fiedler 1995a), but these hypotheses has never been tested using phylogenetic methods.

In this paper, we have performed phylogenetic analyses of the interaction between butterflies and their host plants to address the following questions regarding the patterns and causes of host shifts, colonizations and specialization: (1) Are patterns of butterfly host plant utilization nonrandom so that related butterflies feed on related plants, as suggested by Ehrlich and Raven? (2) What was the ancestral host plant association, and has this association constrained host plant utilization in butterflies? (3) Are host shifts involving closely related plant species more common than shifts to more distantly related plants? (4) Are there identifiable groups of unrelated plants that often occur together as hosts? (5) Is plant phylogeny a more conservative aspect of butterfly-plant associations than plant growth form, or vice versa? (6) Are major host shifts more common in woody-plant-feeding than in herb-feeding lineages? (7) Are tree-feeding butterfly taxa associated with a larger number of host plant clades than are herb-feeding taxa?

Methods

Unless otherwise stated, all analyses have been carried out using the computer program MacClade (vers. 3.05, Maddison and Maddison 1992).

Phylogenies

The plant phylogeny used in this study follows the rbcL-based analysis of seed plant relationships by Chase et al. (1993). They performed two different searches using slightly different taxon sampling and weighting procedures. These searches produced very similar trees. We have for the present analysis used the tree produced by their search 2 (or tree B), which they judged to be the most reliable. Chase et al. summarized their findings in a simplified cladogram (their fig. 2) in which most terminal taxa were given informal names, reflecting their approximate correspondence to groupings in previous classifications. This summary phylogeny is presented in a modified form in figure 2 (Chase et al. 1993). With minor modifications to be noted, we have used the Chase et al. terminal clades and nomenclature as the character states in our analysis and discussion of butterfly host associations. For better resolution within their "Rosid 1" clade, exceptionally important as butterfly–host plants, we have recognized six subclades, following the branching pattern of their "tree B," which we labeled "Rosid 1A . . . F." Unless otherwise specified, the term "plant clade" in this paper refers to these terminal clades and subclades in the Chase et al. phylogeny. There are problems with this analysis, mainly arising from the computational difficulties of analyzing a dataset of this size. In a critique of the analysis, Baum (1994) notes that although it is very likely that Chase et al. have not found the most parsimonious tree, the final phylogeny include many higher level groupings suggested by traditional systematists, and that even the unconventional placements of some taxa often fit surprisingly well with morphological data. In any case, it is the most comprehensive attempt so far to reconstruct a phylogeny for the seed plants as a whole, and should be a better estimate of the true phylogeny than one inferred from previous taxonomy. For most of our analyses we have only used the terminal clades in this phylogeny, not the deeper branchings, which may be less reliable. In a few cases where host plant families were not included in the analysis by Chase et al., their positions were inferred from the classification of Cronquist (1981). These plant families

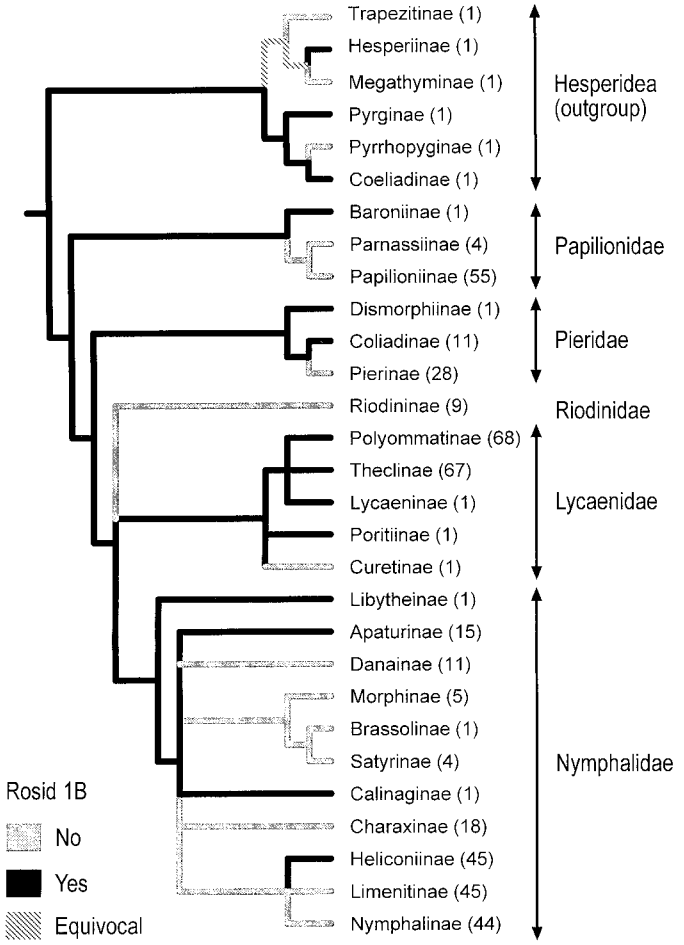

FIG. 1. Simplified version of the phylogeny of Papilionoidea used in the analyses, showing the major relationships within butterflies. Total number of taxa in the phylogeny actually used in the analyses is given within parentheses behind each taxon name. Utilization of the plant clade Rosid 1B (Cannabidaceae, Fabaceae, Krameriaceae, Moraceae, Polygalaceae, Rhamnaceae, Rosaceae, Ulmaceae, Urticaceae, and Zygophyllaceae) is traced on the phylogeny. The assignment of states to branches follows optimization on the complete phylogeny.

were Salicaceae (placed in Rosid 1A), Plantaginaceae (Asterid 1), and Cactaceae (Hamamelid 1).

There has not yet been any comparable attempt to perform a combined phylogenetic analysis of the butterflies as a whole. For this reason we have synthesized results from several smaller studies into a single phylogeny for all butterflies (Fig. 1, Appendix 2). Sources for the different parts of the phylogeny, and the type of evidence they presented, are listed in Table 1.

The phylogeny has been resolved to generic level in most groups, the exceptions being groups with little or no variation in host use (e.g., Satyrinae, which has been resolved to tribal level) and groups with large variation in host use and for which a detailed phylogeny is available (e.g., *Papilio*, which has been resolved to species level). The level of resolution is likely to have some effect on the reported patterns (Sillén-Tullberg 1993, see Results). The complete phylogeny consists of 437 ingroup taxa, and is given in parenthetical format in Appendix 2, (the subfamily level relationships are shown in Fig. 1).

There are perhaps even more uncertainties in the butterfly phylogeny than in the plant phylogeny. Parts of the phylogeny are poorly resolved. This is particularly true for the large family Lycaenidae, but also for Pieridae, where we have chosen to collapse branches about which the phenetic analyses by Geiger (1980) and Ehrlich and Ehrlich (1967) were in conflict. The basal structure of Nymphalidae is left unresolved where Harvey (1991) conflicts with Scott and Wright (1990), but the nymphalid taxonomic groupings in our phylogeny follow the taxonomy of Harvey (1991). Where Miller (1987b) conflicts with Hancock (1983) on the resolution of species groups in Papilionini, we have followed Miller's more recent study. We have also, when possible, tried to estimate the effect of the phylogenetic reconstruction on our results.

The choice of outgroup to the butterflies is somewhat problematic. When tracing character evolution on a phylogenetic tree, one should ideally use several outgroups, including the sister group, to correctly reconstruct ancestral character states (Maddison et al. 1984). In our case this criterion is difficult to fulfill as the phylogeny of higher Lepidoptera is almost completely unknown (Nielsen 1989). The skippers (Hesperiidae) are strong candidates as the sister taxon to the true butterflies, but it has been suggested that the wholly South American group Hedyloidea should be positioned between the true butterflies and the skippers (Scoble 1986). Hedyloidea is poorly known (including its host plant affiliations) and its position is controversial. The most recent analyses actually favor Hesperiidae as sister group to the Papilionoidea (Weller and Pashley 1995; de Jong et al. 1996; Scoble 1996). For this reason we have chosen to exclude Hedyloidea from the study, but we test the effect that this may have on the reconstruction of the ancestral host plant association.

Host Plant Data

Data on host plant utilization were collected from several sources (Ehrlich and Raven 1964; Common and Waterhouse 1972; Larsen 1974, 1991; Smart 1975; Johnston and Johnston 1980; Higgins and Hargreaves 1983; Scott 1986; de la Maza Ramíres 1987; DeVries 1987; Migdoll 1987; Miller 1987a; Ackery 1988; Parsons 1991; Corbet and Pendlebury 1992; Ebert 1993). Because it is difficult to evaluate the validity of literature data on host plant affiliations, false records certainly exist. We have therefore tried to be conservative, by excluding anecdotal host plant records. We included a host plant record only if it was corroborated by two independent sources, if the plant was recorded for more than one species in the butterfly genus, if there were records of more than one host plant genus from the same plant family, or if it was the only plant recorded for this butterfly group. Atypical hosts with little support were excluded. The risk we thereby run of erroneously excluding host plant data that are correct is, we believe, outweighed by the advantage of excluding a greater number of records that are incorrect. The host plant database is provided in Appendix 3.

Character Coding

Optimizing a complex character such as host plant utilization is problematic. One major issue is to treat multiple

TABLE 1. Sources for phylogenetic information regarding the butterflies. Morph, morphological data; Behav, behavioral data; Tax, taxonomic grouping, no phylogeny.

Taxon/taxa	Level of resolution used by us	Type of data	References
Papilionoidea	Butterfly families and choice of outgroups	Morph	Kristensen 1976
		rRNA	Martin and Pashley 1992
		Morph + mtDNA	Weller and Pashley 1995
		Morph	de Jong et al. 1996
Hesperioidea	Subfamilies	Morph + Behav	Scott and Wright 1990
Papilionidae	Subfamilies, tribes	Morph	Miller 1987b
Parnassiini	Genera	Morph	Hancock 1983
Graphini, Troidini	Genera, species	Morph	Miller 1987b
	Species	Morph	Munroe 1961
Graphium	Species	Morph	Saigusa et al. 1982
Papilionini	Genera and species groups	Morph	Munroe 1961
		Morph	Hancock 1983
		Morph	Miller 1987b
Papilio	Species	mtDNA	Sperling 1992, 1993
Pieridae	Subfamilies	Morph	Ehrlich and Ehrlich 1967
		Allozymes	Geiger 1980
	Tribes, genera	Morph (Tax)	Smart 1975
Riodinidae	Genera	Morph (Tax)	Smart 1975
Lycaenidae	Subfamilies	Morph	Ackery 1984
	Tribes, genera	Morph	Eliot 1973
		Morph (Tax)	Smart 1975
Nymphalidae	Subfamilies, tribes	Morph	Scott and Wright 1990
	Tribes, genera	Morph (Tax)	Harvey 1991
	Genera	Morph (Tax)	Smart 1975
Danainae	Genera	Morph	Ackery 1984
	Genera	Morph	Ackery and Vane-Wright 1984
Acreini	Species groups	Morph	Pierre 1987

associations. Although most butterfly taxa in our analysis are restricted to one plant clade, 36% use plants belonging to more than one clade. There is no method for character coding capable of handling this problem in a completely satisfactory way. However, several approaches can be taken to work around the problem, all with different problems and advantages.

One method is to simply optimize plant clade use as an unordered multistate character with multiple associations treated as ambiguity. This makes it easy to interpret patterns of multiple host use, but leads to loss of information. It does not help to code multiple associations as polymorphisms, as MacClade 3 does not allow polymorphic states to be assigned to internal nodes in the phylogeny. An alternative is to optimize each plant clade as an independent binary character. This allows easy investigation of specific host associations, but makes it difficult to interpret multiple host use and can lead to false reconstructions of nodes as having no association at all. Currently the only way to correctly handle multiple associations, while allowing ancestral polymorphisms, is to code host use as a multistate character, using a separate state for each host plant association and, in addition, a separate state for each possible *combination* of plant clade associations, and then assign transformation weights using a step matrix (see Appendix 4 and Maddison and Maddison 1992, p. 83). A difficulty with this approach is that the number of states grows exponentially with the number of plant groups. Limitation on how many states the current version (3.05) of MacClade can handle restricts analysis to a maximum of four plant groups simultaneously. The method can therefore only be used on either a broad scale or on subsets of the butterfly hosts. In the following tests we have chosen different approaches to this dilemma, depending on the problem.

Ancestral Host Plant Association

To estimate the phylogenetic sequence of butterfly–host plant associations we created a multistate character with multiple host use coded as polymorphisms. To control for the possibility of erroneous ancestral state assignment due to the fact that polymorphisms are only allowed at terminals in MacClade 3, we also optimized the same data as a matrix of binary characters.

There is some controversy over the degree of certainty with which a character state can be ascribed to an internal node of a phylogeny (e.g., Frumhoff and Reeve 1994). Optimization of a character that changes too quickly relative to the branching pattern of the phylogeny will not result in a plausible historical reconstruction. To address this problem, we performed several randomizations to test to what extent our reconstruction of the ancestral state is dependent on the frequency and distribution of character states among extant taxa and on the structure of the butterfly phylogeny. These tests follow the same logic and use the same null hypothesis as the "permutation tail probability" test (Archie 1989; Faith and Cranston 1991).

First, if host plant utilization changes too fast relative to speciation there will be no phylogenetic signal in this character and ancestral hosts cannot be reconstructed by optimization on the butterfly phylogeny (Frumhoff and Reeve 1994). Reconstruction at the ancestral node will in that case be a function of the relative frequencies of current host as-

sociations. If our reconstruction is an artifact of a host plant association being common because it is for some reason easy to colonize, then this association should be more or less randomly distributed among butterflies today. It follows that if the reconstruction of the ancestral node using the actual distribution of character states among living taxa differs significantly from that under a random distribution, given the same frequencies of character states, then the ancestral host assignment is more likely to reflect the actual evolutionary history. To address this question, a randomization test was carried out by reassigning the observed states 1000 times at random and reconstructing the ancestral node each time. The states of all extant taxa were randomized, including the outgroup.

Second, as the butterfly phylogeny is uncertain on several points, we tested how sensitive our reconstruction of the ancestral host was to the tree topology by partially randomizing tree structure. First, we completely randomized the structure within each family, only retaining the structure between families. We then repeated the randomizations of branches within families, while retaining the most basal branch in each family. In each case, the ancestral character states were reconstructed for 1000 randomizations.

After the above analysis had indicated the ancestral host plant clade, we further separated this clade into smaller units in an attempt to determine the ancestral plant association more closely.

Colonization and Phylogenetic Distance

To test the prediction that host shifts and colonizations are more common between closely related host taxa we performed two different analyses. First we analyzed shifts from the presumably ancestral clade (Rosid 1B) to three major plant groups. This test was carried out on very broad plant categories, namely colonizations or shifts from Rosid 1B to the three major plant groups "rosids" (other than Rosid 1B), "asterids" and "other plants" (Chase et al. 1993), making the step matrix coding described above in Character Coding appropriate. Each of these groups and each possible combination of them were treated as separate states, with 15 in all. Gains and losses were given equal weight, that is, a gain and a loss of one of these plant groups both carried a cost of one (Appendix 4). Because step matrices cannot be used with unresolved trees, the test was carried out on 100 phylogenies with randomly resolved polytomies.

The test hypothesis predicts that shifts from the ancestral host plant should most commonly be to plants in the same major clade (rosids) and least commonly to plants most distantly related to the original host (other seed plants). The number of unambiguous changes from feeding on Rosid 1B to feeding on the three groups were counted using the step matrix and the "chart state changes" option in MacClade. We tested these numbers against the null hypothesis that colonizations should be equally distributed among the three groups (one-third to each), assuming they are of approximately equal diversity and thus provide equally sized "targets" for random colonization. Accurate and meaningful measures of diversity for these groups are very hard to obtain, as they do not easily translate to the traditional taxonomical groups. However, using the number of plant families in our Appendix 1 as a crude measure of diversity, this assumption seems reasonable; there are 54 rosid families, 47 asterid families, and 51 families of other seed plants among the butterfly hosts.

As there is always variation between resolutions we first needed to know whether the found differences were consistent over the resolutions. In this test and in similar tests that follow, we have used an approach to the problem of irresolution similar to Losos (1994) and Martins (1996). For each random resolution we calculated the pairwise differences between number of colonizations of rosids, asterids, and other seed plants from the ancestral clade Rosid 1B. For instance, if there were 35 colonizations of other rosids, 23 of asterids, and 17 of other seed plants in a given resolution, the pairwise difference between rosids and asterids would be +12, between rosids and other seed plants +18, and between asterids and other seed plants +6. The distribution of these differences were examined to get an estimation of the number of resolutions where the hypothesis was corroborated or refuted. In our example all differences were in accordance with the hypothesis, as they were all positive. Note that this procedure only gives an estimate of the consistency of the found difference over the examined phylogenies, the magnitude of the difference may still be small enough to be a result of chance. We therefore performed chi-square tests of fit to the null hypothesis mentioned above, both on the average numbers of colonizations over the 100 resolutions and on each individual resolution.

In the other analysis we investigated colonizations within and between the two large sister groups "rosids" and "asterids." We compared the number of colonizations to and from plants belonging to the same group (rosids or asterids) with the number of colonizations between these groups, making no assumption about the ancestral host plant. To make the test as conservative as possible, we excluded all changes within the terminal clades of the Chase et al. (1993) phylogeny (Rosid 1–3 and Asterid 1–5), counting only changes between these rather large clades as changes "within rosids" and "within asterids." The numbers of unambiguous colonizations were tested by goodness-of-fit, against the null hypothesis that the number of changes within asterids or rosids and between them should be equal. The rationale is that since we assume that the two groups are approximately equally diverse, if colonizations are random, half should be to the other major clade. As the total number of colonizations could be summed using binary characters, there was no need for step matrix coding and hence no need to resolve polytomies.

To identify groups of unrelated plants that often occur together as butterfly host plants, we constructed a seed plant phylogeny using only presence or absence of butterfly taxa as characters, using a method similar to what is often used in cospeciation studies (e.g., Paterson et al. 1993). These were analyzed using PAUP 3.1.1 (Swofford 1991) with default "factory" settings (simple addition sequence, one tree held at each step during stepwise addition, tree bisection-reconnection [TBR] swapping algorithm, MULPARS option in effect, no topological constraints). We compared the plant groups suggested by this analysis with the groups in the Chase et al. (1993) phylogeny. The butterfly characters and

plant taxa follow the matrix shown in Appendix 3. We used a hypothetical ancestor with zero (no butterfly associations) as outgroup, as the origin of angiosperms predates the butterflies, and we did not want to put a constraint on what plant groups could be united by the analysis. Well-defined groups in this analysis that are not supported by the Chase et al. phylogeny were interpreted as host shift tendencies not corresponding to plant phylogeny.

Plant Phylogeny versus Growth Form

To assess whether plant phylogeny or growth form is the more evolutionarily conservative aspect of butterfly host use, we compared the number of steps needed to trace each on the butterfly phylogeny, using a coding that assigns an equal number of states to both features. The categories we used were for plant groups, "rosids," "asterids," and "other plants," and for growth forms, "herbs," "vines," and "trees and shrubs." Plant group and growth form were coded as multistate characters with one state for each category and a separate state for each combination of categories (seven states in total). Transformation weights were assigned using step matrices (see above and Appendix 3). The numbers of steps needed to trace the two characters were averaged over 100 phylogenies with randomly resolved polytomies, and tested by goodness-of-fit against the null hypothesis of equal numbers, that is, that plant growth form and phylogeny are equally conservative aspects of the association. To investigate the consistency of the differences in the number of states needed to trace the two characters, their distributions were examined over the random resolutions (see above). This analysis could potentially be influenced not only by the diversity of the plant clades (see above), but also by the "availability" of different growth forms for colonization. However, at least on the family level, the distribution of growth forms does not seem to be alarmingly unequal, among the families listed in the Cronquist system on the Flowering Plant Gateway website, 282 contain trees or shrubs, 91 vines, and 208 herbs (Watson and Dallawitz 1992).

The character states are very unequal in frequency, with strong bias toward feeding on the plant group "rosids" and the growth form "trees." The question therefore arises of whether there is a different tendency to shift between plant groups while feeding on trees or to shift between growth forms while feeding on rosids.

We used two procedures to test this. First, the possibility that changes in plant clade use are more likely to occur on a certain growth form (and vice versa) was tested with the concentrated-changes test (Maddison 1990; Maddison and Maddison 1992). Specifically we tested if major host shifts were concentrated to branches reconstructed as woody-plant feeders, and if shifts between growth forms are concentrated to branches reconstructed as feeding on rosids. Thus the test was repeated twice, using "plant group" and "growth form" in turn as the dependent variable. As the test cannot handle multistate characters the characters were recoded as binary. All taxa that use rosids were coded as rosid feeders (regardless of what else they feed on). The other state thus represent butterfly taxa that do not feed on rosids. Growth forms were treated the same way. To ensure that only complete shifts were included in the analysis (i.e., when a colonization is followed by specialization on the novel plant), we only counted a shift if no species in the butterfly taxon has retained the old association. Shifts were counted using step matrix coding of plant use and the "chart all changes" option in MacClade. The concentrated-changes test can only be used on completely resolved phylogenies, so the test was iterated over 100 butterfly phylogenies with randomly resolved polytomies. MacClade (Maddison and Maddison 1992) uses a simulation algorithm to calculate the statistics on large phylogenies and our calculations were based on 1000 such simulations for each randomly resolved tree, using the default settings of the program (allowing either state to be ancestral and using actual changes).

As an alternative test of the association between tree-feeding and tendency to host shift, we counted the number of terminal plant clades from the Chase et al. (1993) phylogeny that were used as hosts by the terminal taxa in our butterfly phylogeny. If this measure of "host use diversity" is treated as a continuous character, its correlation with tree feeding can be assessed by the approach of independent contrasts (e.g. Felsenstein 1985; Pagel 1992; Purvis and Rambaut 1995). As the butterfly terminal taxa are most often genera, taxa with multiple host plant associations may represent collections of species specialized on different plants, polyphagy of individual species, or both. In any case, "host use diversity" should approximate the number of host colonizations that have taken place within that taxon, regardless of whether these have led to increases in host range or to divergence in host plant use among species.

These data were analyzed with Comparative Analysis using Independent Contrasts (CAIC; Purvis and Rambaut 1995), using the "brunch" algorithm, which is designed to test if changes in a discrete character (like tree feeding) are associated with changes in a continuous character (like host range). This test differs from the concentrated changes test in simply testing whether changes in two characters are correlated, without identifying one character as logically independent or causative. One advantage of CAIC is that it has an algorithm for handling polytomies. Details of the tests can be found in Pagel (1992) and Purvis and Rambaut (1995). The contrasts generated by CAIC were tested both qualitatively with a sign test or quantitatively with a one-sample *t*-test (see Höglund and Sillén-Tullberg 1994).

CAIC uses explicit assumptions about branch lengths and offers two default alternatives: all branch lengths assumed equal (corresponding to a punctuational model of evolution) or ages of taxa assumed proportional to the number of included species (corresponding to a gradual model of evolution), using an algorithm by Grafen (1989). As we had no information on branch lengths, we used both.

Results

Patterns of Host Plant Utilization

Butterflies use almost all major seed plant families, and even a few nonseed plants, some species do not even feed on plants (Fig. 2). Nevertheless, as was pointed out by Ehrlich and Raven (1964), the use of plant clades appears nonrandom. Some families are heavily utilized by many butterfly groups

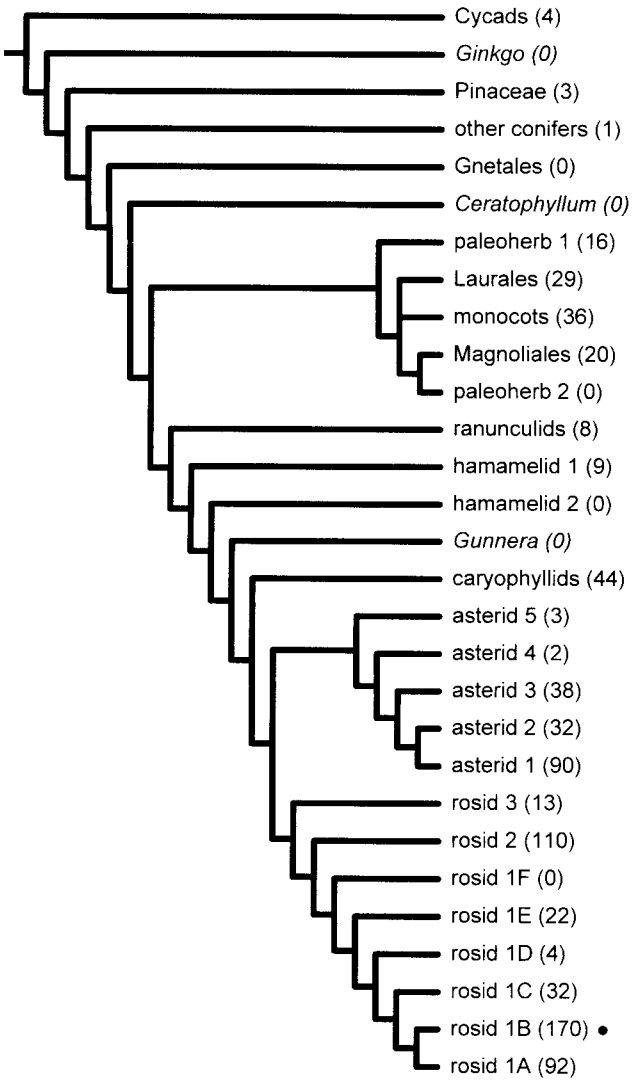

FIG. 2. Phylogenetic relationships among seed plants from Chase et al. (1993). The names Ros 1A to E are assigned by us to subclades identified by Chase et al. (1993). Numbers of associated taxa in the butterfly phylogeny are indicated after plant clade names. The dot denotes presumed ancestral host plant clade.

(e.g., Fabaceae), while others are dominant hosts for particular subsets of butterflies. In Papilionidae, Aristolochiaceae (Paleoherb 1) and Rutaceae (Rosid 2) are clearly dominating themes, while Pieridae are typically associated with plants in Brassicaceae and related families, such as Capparidaceae (Rosid 2). Fabaceae (Rosid 1B) are common hosts in Riodinidae and Lycaenidae, and perhaps Urticales (Rosid 1B) together with Passifloraceae and related families (Rosid 1A) could be said to be predominant hosts in Nymphalidae, although this is a very large generalization. Other, equally conspicuous plant groups, however, are only rarely used as hosts by butterflies. The very large family Orchidaceae is only used by a few genera in Lycaenidae (*Hypolycaena* and *Chliaria* in Theclinae) and Nymphalidae (*Faunis* in Amathusiinae). Likewise, gymnosperms are used by only a handful of species in Pieridae (*Neophasia* in Pierinae) and Lycaenidae (*Callophrys, Eumaeus*, and *Strymon* in Theclinae, and *Theclinesthes* and *Luthrodes* in Polyommatinae). Other examples include Asteraceae, which, considering its size, is also relatively rarely used by butterflies.

Further supporting Ehrlich and Raven's (1964) contention that related butterflies tend to feed on related groups of plants, the Chase et al. (1993) phylogeny suggests that some ostensibly disparate host plant assemblages are more phylogenetically homogenous than previous taxonomy suggested. For example, under the classification of Cronquist (1981), the diverse host list of the nymphalid tribe Nymphalini includes as dominant themes three families in the order Urticales (Hamamelidae), two in Rosales (Rosidae), one in Salicales (Dillenidae), and two in Fagales (Hamamelidae). In addition there are species feeding on Ericales (Dillenidae), Asterales, (Asteridae) Rhamnales (Rosidae), Malvales (Dillenidae), and Liliales (Liliidae). In the Chase et al. (1993) analysis, however, the most frequently used groups; Rosaceae (Rosales), Urticales, Salicales, Fagales and Rhamnales, formerly in three subclasses, all fell within the Rosid 1 subclade, while Malvales belonged to the sister group Rosid 2 and Grossulariaceae (Rosales) to Rosid 3. Thus, the nine plant orders of main nymphalid hosts, previously scattered over five subclasses probably have relatively close affinities.

Ancestral Host Plant Association

The optimization of host use as a multistate character and as binary characters both suggest that the ancestral host plant was located within the clade we have called "Rosid 1B" (Fig. 1, see Appendix 1 for a list of included families). This was the only clade that even came close to being drawn back to the root of the butterfly phylogeny. Plant taxa such as Rosid 2 and Asterid 1 (see Appendix 1), also used by many butterflies (Fig. 2), are apically distributed on the butterfly phylogeny.

Hesperioidea, the most likely sister group, was used as outgroup for the optimizations. This group is mainly associated with Fabaceae (Rosid 1B) and monocotyledons. If the sister group to Papilionoidea instead turns out to be Hedyloidea, the reconstruction will be somewhat more uncertain. Host plant data on Hedyloidea are scarce but indicate that Sterculiaceae (Rosid 2) is the most important host plant group. Sterculiaceae is a member of Malvales, suggested by Ackery (1991) to be the ancestral host group for butterflies. However, even in this case our optimizations suggest that the colonizations of Rosid 2 by Hedylidae, some Hesperiidae, and some Papilionidae are independent evolutionary events and that Rosid 1B is the most probable ancestral host plant group. The same is true if Hedyloidea is used as the only outgroup or if no external outgroup is used at all.

The reconstructed ancestral node appears to be significantly different from reconstructions under random character state assignment. Rosid 1B feeding was ascribed to the ancestral node in only 49 of 1000 randomizations ($P = 0.049$). The relatively high proportion of Rosid 1B feeding on the phylogeny (39%) is therefore not a sufficient explanation for its reconstruction as the ancestral state. Moreover, as the distribution of Rosid 1B feeders in the phylogeny is nonrandom, it is unlikely that the widespread use of this group (and the reconstruction of the ancestral state) should simply follow

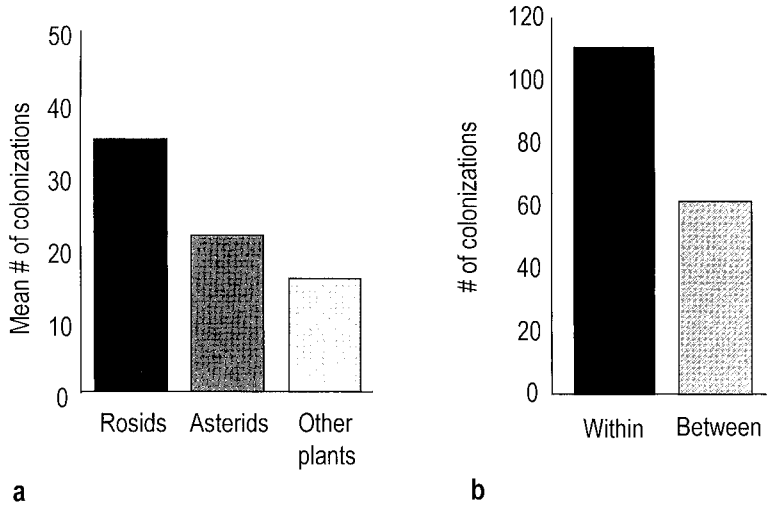

FIG. 3. Association between likelihood of colonizations and phylogenetic distance. (a) Mean number of colonizations from the ancestral clade Rosid 1B to three large plant groups of approximately equal diversity, averaged over 100 phylogenies with randomly resolved polytomies (see text). (b) Number of colonizations of plants belonging to the same major clade, that is, from a rosid to a rosid or an asterid to an asterid, compared with the number of colonizations from an asterid to a rosid or vice versa.

from the fact that Rosid 1B might for some reason be easy to colonize. A historical explanation of the utilization of at least this plant group is therefore more probable.

It follows from this result alone that the phylogenetic structure must be more important for the reconstruction of the ancestral association than simply frequency of use of each plant taxon. A thousand random resolutions of all branches within the butterfly families (keeping only the structure between families) generated only 30 cases where utilization of Rosid 1B was ascribed to the butterfly ancestor ($P = 0.03$), indicating that this reconstruction is a very improbable outcome if we regard the phylogenetic relationships within butterfly families as completely unknown. However, it is enough to retain the most basal branch in each family and randomly resolve all branches above to make Rosid 1B the most likely ancestral association. This means that the reconstruction of the basal branches within each family is critical for the reconstruction of the ancestral butterfly–host plant association.

After optimizing a multistate character where Rosid 1B had been further divided into two subclades (1B1: Fabaceae and Polygalaceae; 1B2: remaining families, see Appendix 1) indicated that the ancestral host plant family most likely was Fabaceae (as Polygalaceae is a very minor host). The result of this more detailed analysis was not used further, and for this reason it was not put through the randomization tests described above.

Colonization and Phylogenetic Distance

A comparison of the number of unambiguous changes from what we suggest to be the original host plant clade (Rosid 1B) to the three large groups of roughly equal size, "other rosids," "asterids," and "other seed plants," gives some support to the hypothesis that the probability of a host shift is related to the phylogenetic distance between the plant groups involved (Fig. 3a). There was no single resolution where the number of colonizations of rosids was fewer than the number of colonizations of asterids and "other plants." In only four resolutions out of 100 there were more colonizations of "other plants" than of asterids. Thus, we conclude that the differences were consistent over the phylogenies with randomly resolved phylogenies.

The number of colonizations of other rosids averaged 35.0 (range: 25–49) over 100 phylogenies with randomly resolved polytomies. Colonizations of asterids, sister group to the rosids, averaged 23.2 (16–33), while colonizations of any plant outside the rosids and asterids only averaged 17.4 (8–23). These average numbers depart significantly from the null hypothesis of equal proportions ($\chi^2 = 6.38$, df = 2, $P = 0.041$). However, among the 100 actual resolutions, there were 39 where the differences in colonizations proved nonsignificant. Thus we cannot safely conclude that colonizations are not evenly distributed among groups in the "true" resolution, even though they are consistent in direction.

Comparisons of shifts just between rosids and asterids gave stronger support to the hypothesis (Fig. 3b). The total number of colonizations between these groups accounted for 62 of 173 unambiguous changes on the phylogeny, or 35.8%, while 111 shifts occurred among plant clades within rosids or asterids ($\chi^2 = 13.88$, df = 1, $P < 0.001$). This was a very conservative test, as shifts within the terminal plant clades of Chase et al. (1993) were not counted.

Ehrlich and Raven (1964) noted that some groups of genealogically unrelated plants often occurred together as hosts, suggesting an underlying chemical convergence. Our analysis using the butterfly clades as characters suggests additional such convergences extending Ehrlich and Raven's observation. The most notable grouping of unrelated plants by butterflies was utilization of plants in Asterid 1 together with plants in Rosid 1 and 2. Twenty-seven butterfly taxa in our analysis use plants in all these clades as hosts, representing between 18 and 23 independent evolutionary events. The plant families used were most often Rosaceae and Ulmaceae in Rosid 1, Rutaceae, Tiliaceae, and Sapindaceae/Sterculiaceae in Rosid 2, and Oleaceae, Rubiaceae, and Verbenaceae

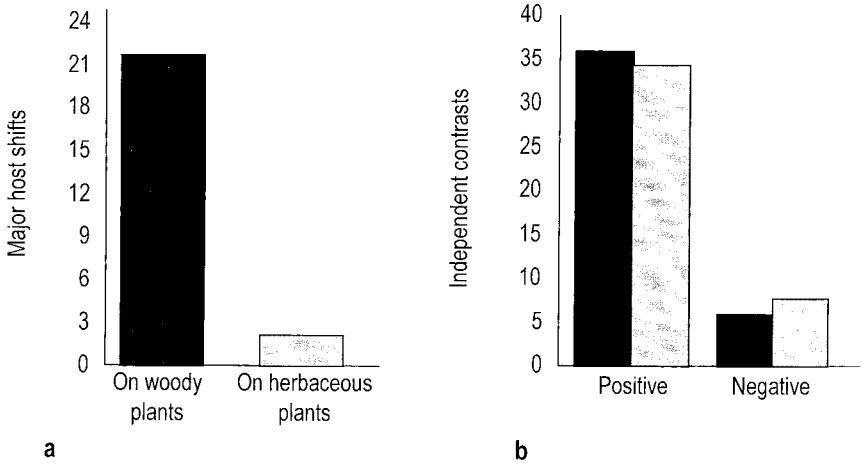

FIG. 4. Association between feeding on woody plants and host shifts or colonizations. (a) Mean number of shifts between rosids and other plants when feeding on woody versus herbaceous plants, averaged over 100 phylogenies with randomly resolved polytomies. (b) Independent contrasts between tree- and nontree-feeding lineages. Positive contrasts are those in which the tree-feeding lineage uses more hosts clades per butterfly taxon. Black bars represent results using scaled branch lengths, and shaded bars represent results using equal branch lengths (see text).

in Asterid 1. Other, less strongly supported groupings in our analysis that differed from those of Chase et al. (1993) include Rosid 3 united with Asterid 3 (5–6 independent events), and Asterid 2 united with ranunculids (3 independent events).

Plant Phylogeny versus Growth Form

The mean number of steps needed to trace the character "major plant groups" (rosids, asterids and other plants) on 100 phylogenies with randomly resolved polytomies was 217.1 (range 210–226), while it was 168.6 (range 163–174) for the character "growth form." This result departs significantly from the null hypothesis that changes in growth form and plant clade are equally common ($\chi^2 = 6.10$, df = 1, $P = 0.014$). The difference is consistent as the distributions of numbers of steps for the two characters across randomizations do not overlap. This indicates that, even at this low level of plant group resolution, growth form may in fact be a more evolutionarily conservative aspect of butterfly–host plant associations than plant phylogeny.

There is, however, one plant clade and one growth form that are much more widely utilized than the others. Rosids are used as host plants by 69% of the butterfly taxa, while asterids and other plants are used by only 30% and 35%, respectively. Likewise, 73% of the butterfly taxa feed on trees and/or shrubs, while 21% feed on vines and 31% feed on herbs. Thus, a higher tendency to shift between plant clades when feeding on trees would increase the apparent average rate of host plant shift. Likewise, a lower tendency to shift between growth forms when feeding on rosids could decrease the apparent rate of shifts among growth forms.

Interpretation of the concentrated changes test is complicated, as the results from the 100 random resolutions polytomies vary substantially (Fig. 4a). For a majority of resolutions (61), the number of major host shifts taking place on branches characterized by feeding on trees and shrubs is higher than expected by chance ($P < 0.05$). However, the P-values ranged from 0.002 to 0.349.

Not surprisingly, there was no evidence at all for the complementary hypothesis, that feeding on rosid plants should discourage shifts between different growth forms, as in no randomization was the P-value below 0.1.

The independent contrasts test showed a strong association between tree feeding and use of an increased number of host plant clades by butterfly taxa, giving further support to the hypothesis that colonizations of new host plants are facilitated by feeding on trees (Fig. 4b, exemplified in Fig. 5). This relationship was highly significant under both the sign test and the one-sample t-test (Table 2). The trend was consistent over all butterfly families, although not significant in all families, probably due to small numbers of contrasts. The two algorithms for estimating branch lengths gave very similar results.

A possible problem with these tests is that diversity of host use and taxon size are confounded. One can not entirely rule out the possibility that tree-feeding taxa in general contain more species that herb-feeding taxa.

DISCUSSION

Ehrlich and Raven's (1964) main observations, that related butterflies often feed on related host plants, and that some plant groups are commonly used by butterflies while other large plant groups are not, are upheld by our reanalysis (Fig. 2). As others have noted (e.g., Jermy 1976, 1984), these patterns can be explained by sequential colonization of related plant groups without the coevolutionary twist that the plants have escaped and radiated in the absence of butterfly feeding. Under a coevolutionary interpretation, underutilization of some plant groups by butterflies would be ascribed to chemical defences evolved to exclude butterflies from feeding. In the light of our results, however, it is not likely that such groups, which are also much older than the butterflies, have ever been used by butterflies. Some may have evolved chemical defenses against other herbivores that are equally effective against butterflies. Others, especially those

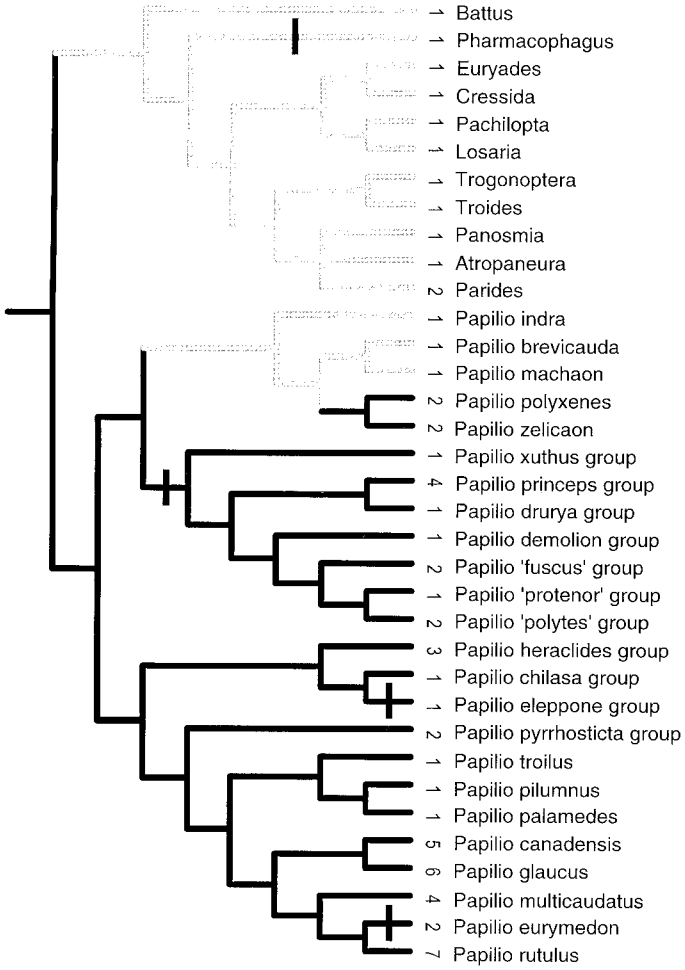

FIG. 5. Association between tree feeding and propensity for host shifts and colonizations exemplified by Troidini and *Papilio* in Papilionidae. White branches indicate herb-feeding, black branches woody-plant-feeding lineages. Number of plant clades from Chase et al. (1993) used as hosts by each butterfly taxon are indicated below butterfly names. One of the two to three shifts between rosids and other plants, taking place on herb-feeding branches, and three of the 13–32 such shifts taking place on woody plant-feeding branches, are shown in this figure (bars on branches). The figure also shows three of the 42 independent contrasts between herb- and tree-feeding taxa. In all 3 contrasts shown the tree-feeding taxa are associated with higher numbers of plant clades than are the herb-feeding taxa.

TABLE 2. Relationship between tree feeding and number of plant clades used by butterfly taxa, calculated using independent contrasts. Two different algorithms have been used to estimate branch lengths: (1) all branch lengths equal; or (2) branch lengths scaled so that the ages of taxa are proportional to the number of species they contain.

		t-test			Sign test	
Taxon	Branch lengths	*t*	df	Probability	pos/neg contrasts	Probability
Papilionoidea	Equal	4.23	62	< 0.001	34/8	< 0.001
	Scaled	3.77	62	< 0.001	36/6	< 0.001
Papilionidae	Equal	1.27	5	0.259	2/1	1.000
	Scaled	1.62	5	0.166	3/0	0.248
Pieridae	Equal	1.41	7	0.201	2/0	0.480
	Scaled	1.24	7	0.255	3/0	0.248
Riodinidae	Equal	*	0	*	0/0	*
	Scaled	*	0	*	1/0	*
Lycaenidae	Equal	2.91	19	0.009	11/3	0.061
	Scaled	2.29	19	0.034	11/3	0.061
Nymphalidae	Equal	2.93	26	0.007	17/4	0.009
	Scaled	2.88	26	0.008	18/3	0.002

* Could not be calculated.

very distantly related to the original butterfly host, may simply not have been "discovered" yet by butterflies, as such distant colonizations appear to have been rare. Given the current state of knowledge, it is impossible to give any definite explanation to these patterns.

Our analysis also extends Ehrlich and Raven's observation that some groups of unrelated plants repeatedly co-occur in the host lists of butterfly taxa, which they interpreted as reflecting chemical or other underlying similarity. The clearest prediction from our analysis is that plants in Rosid 1 (most importantly Rosaceae and Ulmaceae), Rosid 2 (most importantly Rutaceae, Tiliaceae and Sapindaceae/Sterculiaceae), and Asterid 1 (most importantly Oleaceae, Rubiaceae, and Verbenaceae) share some chemical or other feature that affects butterfly host selection. Another prediction, that may be easier to test, is that if these plants do share such a trait, they should also be linked in the diets of other groups of phytophagous insects.

Considering the variation in dominating host taxa among butterfly families, it is somewhat remarkable that the most basal branches in each family feed on plants in Rosid 1B, such as Fabaceae, Urticaceae, Ulmaceae, or Rosaceae. The Rosid 1B clade is also by far the most likely to have included the ancestral butterfly host. The character state randomizations strongly suggest that this pattern does reflect evolutionary history, not this plant group being for some reason unusually easy to colonize. Even if Rosid 1B were easier to colonize, there could still be a historical explanation: all butterflies may be literally preadapted to the chemical and other traits of these plants, simply because their ancestors fed upon plants containing them. Further subdivision of Rosid 1B corroborates Scott's (1986) suggestion that Fabaceae was the most likely ancestral host plant family.

The strong conservatism of butterfly association with major plant clades does not preclude frequent shifts among related host species. Indeed, many butterflies feed on several species or genera within the same plant family, suggesting that there have been many colonization events between plants too closely related to be distinguished by our analysis.

Restriction of most colonizations to related plants could reflect constraints on genetic variation in the capacity to feed on novel host plants, making a shift to an ancestral host plant more likely than to a completely novel plant (Futuyma 1991). There is some evidence for this in chrysomelid leaf beetles (Futuyma et al. 1993, 1994, 1995) and in butterflies (N. Janz and S. Nylin, unpubl.). In fact, one of the examples Ehrlich and Raven (1964) give in their paper on colonization of related plant groups by related butterflies could, in the light of the new plant phylogeny, be better understood as a recolonization of the ancestral host plant clade: the switch of one genus in Parnassiini (*Hypermnestra*) from the dominant Ar-

istolochiaceae (Pal 1) and Rutaceae (Rosid 2) theme to feed on Zygophyllaceae (Rosid 1B), which Ehrlich and Raven claimed to be closely related to Rutaceae. If such recolonizations are common, it means that a high number of host shifts could go unnoticed in a phylogenetic study, because they tend to shift back to the original host, thus making the overall pattern look more conservative than it is on a finer level. Or, put in another way, an opportunistic pattern of host plant utilization on a microevolutionary scale may well result in a conservative pattern on a macroevolutionary scale.

The fact that plant growth form was the more conservative aspect of host associations in our analysis suggests that other factors than plant chemistry, such as habitat or community structure, play an important role in shaping the large-scale patterns of butterfly-host plant association. A similar conclusion was reached in a recent phylogenetic study of weevils and their host plants (Anderson 1993). The influence of growth form is accentuated by the elevated rate of host shifts in lineages feeding on trees, the most frequently used plant growth form among butterflies, while changes among growths form appear to have occurred independently of the host plant clade. This, in turn, can perhaps be explained by Feeny's (1976) distinction between the different kinds of defenses utilized by "apparent" trees and "unapparent" herbs. As the mature foliage of trees with different taxonomic origins will have a convergent chemical defence, evolving a capacity to feed on mature leaves of a particular tree will preadapt the insect to feed on mature leaves from other trees (Feeny 1976, 1991). It follows that these aspects of plant chemistry should in fact be better correlated with plant growth form than with phylogeny.

In this study we have explicitly focused on the patterns and determinants of host shifts, through colonization (when a new plant is added to the host plant range) and specialization (narrowing of the host range, in the case of a host shift to include only the novel plant), as these seem to be the most important processes shaping the association between butterflies and their host plants. Even if the patterns that emerge on this taxonomic level cannot themselves have been caused by coevolution, the general mechanisms behind host shifts are of great importance for understanding the dynamics of the coevolutionary process.

Acknowledgments

We wish to thank O. Eriksson, C. Mitter, B. Tullberg, C. Wiklund, and three anonymous referees for valuable comments on the manuscript. This work was supported by a grant from the Swedish National Science Research Council to SN.

Literature Cited

Ackery, P. R. 1984. Systematic and faunistic studies on butterflies. Pp. 9–21 *in* R. I. Vane-Wright and P. R. Ackery, eds. The biology of butterflies: their cladistics and biology. British Museum (Natural History), London.

———. 1988. Host plants and classification: a review of nymphalid butterflies. Biol. J. Linn. Soc. 33:95–203.

———. 1991. Hostplant utilization by African and Australian butterflies. Biol. J. Linn. Soc. 44:335–351.

Ackery, P. R., and R. I. Vane-Wright. 1984. Milkweed butterflies: their cladistics and biology. British Museum (Natural History), London.

Anderson, R. S. 1993. Weevils and plants—phylogenetic versus ecological mediation of evolution of host plant associations in Curculioninae (Coleoptera, Curculionidae). Mem. Entomol. Soc. Can. 165:197–232.

Archie, J. W. 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38:239–252.

Baum, D. 1994. rbcL and seed-plant phylogeny. Trends Ecol. Evol. 9:39–41.

Benson, W. W., K. S. Brown, and L. E. Gilbert. 1975. Coevolution of plants and herbivores: passion flower butterflies. Evolution 29:659–680.

Berenbaum, M. R. 1983. Coumarins and caterpillars: a case for coevolution. Evolution 37:163–179.

Bernays, E. A., and M. Graham. 1988. On the evolution of host specificity in phytophagous arthropods. Ecology 69:886–892.

Brooks, D. R., and D. H. McLennan. 1991. Phylogeny, ecology, and behavior. Univ. of Chicago Press, Chicago.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, F. H. Strauss, Q. Xiang, G. M. Plunkett, P. S. Soltis, S. M. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr, S. W. Graham, S. C. H. Barrett, S. Dayanandan, and V. A. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. Ann. Mo. Bot. Gard. 80:528–580.

Coddington, J. A. 1988. Cladistic tests of adaptational hypotheses. Cladistics 4:3–22.

———. 1994. The roles of homology and convergence in studies of adaptation. Pp. 53–78 *in* P. Eggleton and R. Vane-Wright, eds. Phylogenetics and ecology. Academic Press, London.

Common, I. F. B., and D. F. Waterhouse. 1972. Butterflies of Australia. Angus and Robertson, Sydney, New South Wales, Australia.

Corbet, A. S., and H. M. Pendlebury. 1992. The butterflies of the Malay Peninsula. Malayan Nature Society, Kuala Lumpur.

Courtney, S. P. 1984. Habitat versus foodplant selection. Pp. 89–90 *in* R. I. Vane-Wright and P. R. Ackery, eds. The biology of butterflies. Academic Press, Orlando, FL.

Cronquist, A. 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York.

de Jong, R., R. I. Vane-Wright, and P. R. Ackery. 1996. The higher classification of butterflies (Lepidoptera): problems and prospects. Entomol. Scand. 27:65–101.

de la Maza Ramíres, R. 1987. Mariposas mexicanas. Fondo de Cultura Económica, Mexico City.

DeVries, P. J. 1987. The butterflies of Costa Rica and their natural history, Papilionidae, Pieridae, Nymphalidae. Princeton Univ. Press, Princeton, NJ.

Ebert, G. 1993. Die Schmetterlinge Baden-Württembergs. Verlag Eugen Ulmer, Stuttgart, Germany.

Ehrlich, P. R., and A. H. Ehrlich. 1967. The phenetic relationships of the butterflies. I. Adult taxonomy and the non-specificity hypothesis. Syst. Zool. 16:301–317.

Ehrlich, P. R., and P. H. Raven. 1964. Butterflies and plants: a study in coevolution. Evolution 18:586–608.

Eliot, J. N. 1973. The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. Bull. Br. Mus. (Nat. Hist.) Entomol. 28:373–506.

Emmel, T. C., M. C. Minno, and B. A. Drummond. 1992. Florissant butterflies: a guide to the fossil and present-day species of central Colorado. Stanford Univ. Press, Stanford, CA.

Eriksson, O., and B. Bremer. 1992. Pollination systems, dispersal modes, life forms, and diversification rates in angiosperm families. Evolution 46:258–266.

Faith, D. P., and P. Cranston. 1991. Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. Cladistics 7:1–28.

Feeny, P. 1975. Biochemical coevolution between plants and their insect herbivores. Pp. 5–19 *in* L. E. Gilbert and P. H. Raven,

eds. Coevolution of animals and plants. Univ. of Texas Press, Austin.

———. 1976. Plant apparency and chemical defence. Rec. Adv. Phytochem. 10:1–40.

———. 1991. Chemical constraints on the evolution of swallowtail butterflies. Pp. 315–340 *in* P. W. Price, T. M. Lewinsohn, G. W. Fernandes, and W. W. Benson, eds. Plant-animal interactions: evolutionary ecology in tropical and temperate regions. Wiley, New York.

FELSENSTEIN, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

FIEDLER, K. 1995a. Lycaenid butterflies and plants: hostplant relationships, tropical versus temperate. Ecotropica 1:51–58.

———. 1995b. Lycaenid butterflies and plants: is myrmecophily associated with particular hostplant preferences? Ethol. Ecol. Evol. 7:107–132.

FRUMHOFF, P. C., AND H. K. REEVE. 1994. Using phylogenies to test hypotheses of adaptation: a critique of some current proposals. Evolution 48:172–180.

FUTUYMA, D. J. 1976. Food plant specialization and environmental predictability in Lepidoptera. Am. Nat. 110:285–292.

———. 1991. Evolution of host specificity in herbivorous insects: genetic, ecological, and phylogenetic aspects. Pp. 431–454 *in* P. W. Price, T. M. Lewinsohn, G. W. Fernandes, and W. W. Benson, eds. Plant-animal interactions: evolutionary ecology in tropical and temperate regions. Wiley, New York.

FUTUYMA, D. J., M. C. KEESE, AND S. J. SCHEFFER. 1993. Genetic constraints and the phylogeny of insect-plant associations—responses of *Ophraella communa* (Coleoptera, Chrysomelidae) to host plants of its congeners. Evolution 47:888–905.

FUTUYMA, D. J., J. S. WALSH, T. MORTON, D. J. FUNK, AND M. C. KEESE. 1994. Genetic variation in a phylogenetic context—responses of two specialized leaf beetles (Coleoptera, Chrysomelidae) to host plants of their congeners. J. Evol. Biol. 7:127–146.

FUTUYMA, D. J., M. C. KEESE, AND D. J. FUNK 1995. Genetic constraints on macroevolution: the evolution of host affiliation in the leaf beetle genus *Ophraella*. Evolution 49:797–809.

GEIGER, H. J. 1980. Enzyme electrophoretic studies on the genetic relationships of Pierid butterflies (Lepidoptera: Pieridae). I. European taxa. J. Res. Lepid. 19:181–195.

GRAFEN, A. 1989. The phylogenetic regression. Philos. Trans. R. Soc. Lond. B Biol. Sci. 326:119–156.

HANCOCK, D. L. 1983. Classification of the Papilionidae (Lepidoptera): a phylogenetic approach. Smithersia 2:1–48.

HARVEY, D. J. 1991. Higher classification of the Nymphalidae, Appendix B. Pp. 255–273 *in* H. F. Nijhout. The development and evolution of butterfly wing patterns. Smithsonian Institution Press, Washington, DC.

HARVEY, P. H., AND M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford Univ. Press, Oxford, U.K.

HIGGINS, L. G., AND B. HARGREAVES. 1983. The butterflies of Britain and Europe. William Collins and Sons, London.

HÖGLUND, J., AND B. SILLÉN-TULLBERG. 1994. Does lekking promote the evolution of male-biased size dimorphism in birds? On the use of comparative approaches. Am. Nat. 144:881–889.

JERMY, T. 1976. Insect-host plant relationship-coevolution or sequential evolution? Symp. Biol. Hung. 16:109–133.

———. 1984. Evolution of insect/host plant relationships. Am. Nat. 124:609–630.

JOHNSTON, G., AND B. JOHNSTON. 1980. This is Hong Kong: butterflies. Hong Kong Government Publications, Hong Kong.

KRISTENSEN, N. P. 1976. Remarks on the family-level phylogeny of butterflies (Insecta, Lepidoptera, Rhopalocera). Zeitschrift für Zoologische Systematik und Evolutionsforschung 14:25–33.

LARSEN, T. B. 1974. Butterflies of Lebanon. National Council for Scientific Research, Beirut.

———. 1991. The butterflies of Kenya and their natural history. Oxford Univ. Press, Oxford, U.K.

LOSOS, J. B. 1994. An approach to the analysis of comparative data when a phylogeny is unavailable or incomplete. Syst. Biol. 43:117–123.

MADDISON, W. P. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution 44:539–557.

MADDISON, W. P., AND D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Vers. 3.05. Sinauer, Sunderland, MA.

MADDISON, W. P., M. J. DONOGHUE, AND D. R. MADDISON. 1984. Outgroup analysis and parsimony. Syst. Zool. 33:83–103.

MARTIN, J. A., AND D. P. PASHLEY. 1992. Molecular systematic analysis of butterfly family and some subfamily relationships (Lepidoptera: Papilionoidea). Ann. Entomol. Soc. Am. 85:127–139.

MARTINS, E. P. 1996. Conducting phylogenetic comparative studies when the phylogeny is not known. Evolution 50:12–22.

MIGDOLL, I. 1987. Field guide to the butterflies of southern Africa. C. Struik Publishers, Cape Town, South Africa.

MILLER, J. S. 1987a. Host-plant relationships in the Papilionidae (Lepidoptera): parallel cladogenesis or colonization? Cladistics 3:105–120.

———. 1987b. Phylogenetic studies in the Papilioninae (Lepidoptera: Papilionidae). Ph.D. diss. Am. Mus. Nat. Hist., New York.

MITTER, C., AND D. R. BROOKS. 1983. Phylogenetic aspects of coevolution. Pp. 65–98 *in* D. J. Futuyma and M. Slatkin, eds. Coevolution. Sinauer, Sunderland, MA.

MUNROE, E. 1961. The classification of the Papilionidae (Lepidoptera). Can. Entomol. Suppl. 17:1–51.

NIELSEN, E. S. 1989. Phylogeny of major lepidopteran groups. Pp. 281–294 *in* B. Fernholm, K. Bremer, and H. Jörnvall, eds. The hierarchy of life. Elsevier, Amsterdam, The Netherlands.

NYLIN, S., AND N. WEDELL. 1994. Sexual size dimorphism and comparative methods. Pp. 253–280 *in* P. Eggleton and R. Vane-Wright, eds. Phylogenetics and ecology. Academic Press, London.

PAGEL, M. D. 1992. A method for the analysis of comparative data. J. Theor. Biol. 156:431–442.

———. 1994. The adaptationist wager. Pp. 29–52 *in* P. Eggleton and R. Vane-Wright, eds. Phylogenetics and ecology. Academic Press, London.

PARSONS, M. 1991. Butterflies of the Bulolo-Wau Valley. Bishop Museum Press, Honolulu, HI.

PATERSON, A. M., R. D. GRAY, AND G. P. WALLIS. 1993. Parasites, petrels and penguins: does louse presence reflect seabird phylogeny? Int. J. Parasitol. 23:515–526.

PIERRE, J. 1987. Systématique cladistique chez les *Acraea* (Lepidoptera, Nymphalidae). Ann. Soc. Entomol. Fr. 23:11–27.

PRICE, P. W., C. E. BOUTON, P. GROSS, B. A. MCPHEARSON, J. N. THOMPSON, AND A. E. WEIS. 1980. Interactions among three trophic levels: influence of plants on interactions between insect herbivores and natural enemies. Annu. Rev. Ecol. Syst. 11:41–65.

PURVIS, A., AND A. RAMBAUT. 1995. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. Comput. Appl. Biosci. 11:247–251.

SAIGUSA, T., A. NAKAMICHI, H. SHIMA, AND O. YATA. 1982. Phylogeny and geographical distribution of the swallow-tail subgenus *Graphium* (Lepidoptera: Papilionidae). Entomol. Gen. 8: 59–69.

SCOBLE, M. J. 1986. The structure and affinities of the Hedyloidea: a new concept of the butterflies. Bull. Br. Mus. Nat. Hist., Entomol. Ser. 53:251–286.

———. 1996. In search of butterfly origins: morphology and molecules. Trends Ecol. Evol. 11:274–275.

SCOTT, J. A. 1986. The butterflies of North America. Stanford Univ. Press, Stanford, CA.

SCOTT, J. A., AND D. M. WRIGHT. 1990. Butterfly phylogeny and fossils. Pp. 153–208 *in* O. Kudrna, ed. Butterflies of Europe. AULA-Verlag, Wiesbaden, Germany.

SCRIBER, J. M., AND F. SLANSKY. 1981. The nutritional ecology of immature insects. Annu. Rev. Entomol. 26:183–211.

SILLÉN-TULLBERG, B. 1988. Evolution of gregariousness in apo-

sematic butterfly larvae: a phylogenetic analysis. Evolution 42: 293–305.

———. 1993. The effect of biased inclusion of taxa on the correlation between discrete characters in phylogenetic trees. Evolution 47:1182–1191.

SMART, P. 1975. The illustrated encyclopedia of the butterfly world. Salamander Books, London.

SMILEY, J. 1978. Plant chemistry and the evolution of host specificity: new evidence from *Heliconius* and *Passiflora*. Science 201:745–747.

SPERLING, F. A. H. 1992. Mitochondrial DNA variation and Haldane's rule in the *Papilio glaucus* and *P. troilus* species groups. Heredity 71:227–233.

———. 1993. Mitochondrial DNA phylogeny of the *Papilio machaon* species group (Lepidoptera: Papilionidae). Mem. Entomol. Soc. Can. 165:233–242.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony. Vers. 3.1.1. Illinois Natural History Survey, Champaign, IL.

WATSON, L., AND M. J. DALLAWITZ. 1997. The families of flowering plants: descriptions, illustrations, identification, and information retrieval. Vers. 16. March 1997. URL http://www.keli.ukans.edu/delta. *In:* The flowering plant gateway, URL http://www.isc.tamu.edu/FLORA/cronang.htm.

WELLER, S. J., AND D. P. PASHLEY. 1995. In search of butterfly origins. Mol. Phylogenet. Evol. 4:235–246.

WELLER, S. J., D. P. PASHLEY, AND J. A. MARTIN. 1996. Reassessment of butterfly family relationships using independent genes and morphology. Ann. Entomol. Soc. Am. 89:184–192.

ZANGERL, A. R., AND M. R. BERENBAUM. 1993. Plant chemistry, insect adaptations to plant chemistry, and host plant utilization patterns. Ecology 74:47–54.

Corresponding Editor: C. Mitter

APPENDIX 1

List of important host families for butterflies included in the clades recognized by Chase et al. (1993). Question marks denote families that were not included in the analysis by Chase et al., and for which positions have been inferred from the classification of Cronquist (1981).

Rosid 1: (A): Celastraceae, Erythroxylaceae, Euphorbiaceae, Linaceae, Malphigiaceae, Ochnaceae, Passifloraceae, Salicaceae?, Violaceae; **(B)**: Cannabidaceae, Fabaceae, Krameriaceae, Moraceae, Polygalaceae, Rhamnaceae, Rosaceae, Ulmaceae, Urticaceae, Zygophyllaceae; **(C)**: Begoniaceae, Betulaceae, Casuarinaceae, Cucurbitaceae, Fagaceae, Juglandaceae, Myricaceae; **(D)**: Oxalidaceae (*Oxalis*); **(E)**: Combretaceae, Melastomaceae, Myrtaceae, Punicaceae.

Rosid 2: Aceraceae, Anacardiaceae, Bataceae, Bombacaceae, Brassicaceae, Burseraceae, Capparaceae, Caricaceae, Geraniaceae, Hippocastanaceae, Malvaceae, Oxalidaceae (*Hypserocharis*), Resedaceae, Rutaceae, Sapindaceae, Simaroubaceae, Sterculiaceae, Tiliaceae, Tropaoleaceae.

Rosid 3: Crassulaceae, Grossulariaceae (*Ribes* . . .), Hamamelidaceae, Saxifragaceae (*Saxifraga*).

Asterid 1: Acanthaceae, Apocynaceae, Asclepiadaceae, Bignoniaceae, Boraginaeae, Convolvulaceae, Cornaceae (*Aucuba*), Gentianaceae, Gesneriaceae, Hydrophyllaceae, Lamiaceae, Loganiaceae, Oleaceae, Plantaginaceae?, Rubiaceae, Scrophulariaceae, Solanaceae, Verbenaceae.

Asterid 2: Apiaceae, Aquifoliaceae, Araliaceae, Asteraceae, Campanulaceae, Caprifoliaceae, Cornaceae (*Corokia*), Cornaceae (*Griselinia*), Cornaceae (*Helwingia*), Dipsacaceae, Menyanthaceae, Pittosporaceae, Valerianaceae.

Asterid 3: Diapensiaceae, Ebenaceae, Epacridaceae, Ericaceae, Myrsinaceae, Primulaceae, Sapotaceae, Symplocaceae, Theaceae.

Asterid 4: Alangiaceae, Araliaceae, Cornaceae (*Cornus*), Nyssaceae, Hydrangeaceae.

Asterid 5: Dilleniaceae, Vitidaceae.

Ranunculids: Berberidaceae, Fumariaceae, Menispermaceae, Papaveraceae, Ranunculaceae.

Paleoherbs 1: Aristolochiaceae, Piperaceae.

Monocots: Arecaceae, Bromeliaceae, Commelinaceae, Cyperaceae, Dioscoreaceae, Heliconiaceae, Liliaceae, Musaceae, Orchidaceae, Poaceae, Smilacaceae, Zingiberaceae.

Laurales: Hernandiaceae, Lauraceae, Monimiaceae.

Magnoliales: Annonaceae, Cannelaceae, Magnoliaceae, Winteraceae.

Paleoherbs 2: Chloranthaceae, Illiciaceae, Nympheaceae.

Hamamelid 1: Platanaceae, Proteaceae, Sabiaceae, Aizoaceae, Amaranthaceae, Cactaceae?

Caryophyllids: Caryophyllaceae, Chenopodiaceae, Nyctaginaceae, Olacaceae, Plumbaginaceae, Polygonaceae, Portulaceae, Santalaceae, Viscaceae/Loranthaceae.

Pinaceae.

(other) conifers: Cupressaceae, Podocarpaceae, Taxaceae, Taxodiaceae.

Cycads: Cycadaceae, Zamiaceae.

APPENDIX 2

Description of the butterfly phylogeny used in this study. Clades are identified by parantheses, numbers refer to taxon # in Appendix 3 The phylogeny can be obtained from the authors in electronic form upon request.

(((440,(438,439)),(441,(442,443))),((1,((5,(3,(2,4))),(((18,19),(20,(21,(22,((35,36),((26,(23,(25,24))),((32,(33,34)),(27,((31,30),(28,29)))))))))),(17,((6,(7,(((9,8),(11,10)),((12,13),(16,14,15))))),(((50,((49,46),(48,47))),(39,((45,44),(40,(43,(41,42))))))),((60,(37,38)),(59,((56,(58,57)),((52,51),(55,(54,53))))))))))),((61,(((90,91),(93,92),94,95,(96,97),98,(100,99)),((88,86,87,89),(80,(72,75,(84,85)),82,83,((81,71,70),(78,73,79,77,76,74)),(67,69,68),65,66,(63,64),62)))),((109,102,(101,103),(108,107,106,105,104)),(((180,((181,182),(((221,222,223),(219,220,225,224,226)),213,(229,(227,228)),202,((195,196),(198,197)),183,(244,247,231,230,232,233,234,235,236,237,238,239,240,241,242,243,246,245),((216,215),218,214,217),(212,(209,210),211),(206,205,(207,208),(204,203)),(201,(199,200)),(190,(191,192)),(189,(186,184,185,187,188)),193,194))),(((116,115),(112,(114,113))),((158,(156,153,154,155,157)),(176,170,172,163,164,175,161,173,171,174,168,165,169,166,120,162,167,121,(160,178,177))),(159,(128,127)),((129,131,130),(132,133,135,134),(((136,138,137,139,140),141),(((142,143),(151,152,150)),(144,((149,148),(147,146,145))))))),((125,(124,(122,123))),126,(118,(117,119)))),179),111,110),(437,((436,435,(((433,434),(431,432)),((426,427),(428,(430,429))))),425,(((418,419),(415,(417,416))),(420,(422,421,423,424))),((393,(399,395,394,396,397,398)),(382,(383,384,385)),(391,390,392),(387,386),388,389),(414,(404,403),(401,402,400),405,410,409,(408,407,406),(411,412),413),(((267,(266,259,260,264,263,262,261,265)),((257,256,255),258,254,253,252,251,250),248),(249,((268,((269,270),((271,(272,273)),((274,(275,276)),(277,(278,279))))))),(280,(281,((282,(284,283)),(292,((285,(286,287)),(288,(289,(290,291))))))))))),(337,(338,(351,(350,349),348),(340,(344,343),(341,342)),339,345,346,347),(381,(380,379,378)),(((358,359),(357,(352,355,356,353,354)),364,((360,361),(362,363))),369,(368,367,366,365),(372,370,377,376,375,374,373,371))),((((295,(296,298)),294,297),293,300,299,301,302,(303,(304,305))),((316,313,317,315,314,312,(309,308),307,306,(311,310)),(((323,322,321,320),327,(328,326,325,324),319),318,(332,331,333,335,334,330,336,329)))))))))));

APPENDIX 3

Data matrix used in the analyses. 0, not used as host; 1, used as host; ?, uncertain. H, herbs; V, vines; S, shrubs; and T, trees. Plant clade names follow Chase et al. (1993), see Appendix 1 for a list of plant families included in the clades.

#	Family	Subfamily	Tribe	Taxon	RRRRRRRAAAAARPMLPMHCPCCHVST 11111231234baaoaaaaaioy ABCDE 1 2 1
1	Papilionidae	Baroniinae	Baroniini	Baronia	0100000000000000000000000??
2		Parnassiinae	Parnassiini	Parnassius	0000001000001000000000001000
3				Archon	0000000000000010000000000100
4				Hypermnestra	0100000000000000000000000010
5			Zerynthiini	Zerynthiini	0000000000000010000000000100
6		Papilioninae	Troidini	Battus	0000000000000010000000001100
7				Pharmacophagus	0000100000000000000000000100
8				Cressida	0000000000000010000000000100
9				Euryades	0000000000000010000000000100
10				Losaria	0000000000000010000000000100
11				Pachliopta	0000000000000010000000000100
12				Trogonoptera	0000000000000010000000000100
13				Troides	0000000000000010000000000100
14				Atrophaneura	0000000000000010000000000100
15				Parides	0000000000000110000000000100
16				Panosmia	0000000000000010000000000100
17			Teinopalpini	Meandrusa	0000000000000000100000000001
18			Graphiini	Eurytides	0000000000000000101000000011
19				Protographium	0000010000000000001000000100
20				Iphiclides	0100000000000000000000000011
21				Lamproptera	0000000000000000100000000011
22				Pazala	0000000000000000100000000011
23				Graphium agamemnon	0000000000000000101000000101
24				" " macfarlanei	0000000000000000001000000001
25				" " mendana	0000000000000010000000000010
26				" " wallacei	0000000000000000001000000001
27				" " macleyanum	0000000000000000101000000001
28				" " empedovana	0000000000000000100000000001
29				" " codrus	0000000000000000100000000001
30				" " cloanthus	0000000000000000100000000011
31				" " sarpedon	0000100001000001010000000011
32				" " doson	0000000000000000101000000101
33				" " evemon	0000000000000000001000000???
34				" " euryplus	0000010000000000001000000171
35				Nomius/Pathysa	0000000000000000001000000171
36				Arisbe	1000010000000000001000000101
37			Papilionini	Papilio chilasa group	0000000000000000100000000011
38				" " eleppone group	0000010000000000000000000011
39				" " xuthus group	0000010000000000000000000001
40				" " demolion group	0000010000000000000000000001
41				" " 'protenor' group	0000010000000000000000000011
42				" " 'polytes' group	0000010100000000000000000011
43				" " 'fuscus' group	0000010100000000000000000011
44				" " drurya group	0000010000000000000000000011
45				" " princeps group	0100010000000001010000000011
46				" " machaon	0000000010000000000000001000
47				" " zelicaon	0000010010000000000000001001
48				" " polyxenes	0000010010000000000000001010
49				" " brevicauda	0000000010000000000000001000
50				" " indra	0000000010000000000000001000
51				" " glaucus glaucus	0110010100000001010000000011
52				" " glaucus canadensis	1110010100000000000000000011
53				" " rutulus	1110010100000001001000000011
54				" " eurymedon	0110000000000000000000000011
55				" " multicaudatus	0100010100000000001000000011
56				" " troilus	0000000000000001000000000011
57				" " palamedes	0000000000000001000000000001
58				" " pilumnus	0000000000000001000000000001
59				" " pyrrhosticta group	0000000000000001010000000101
60				" " heraclides group	0000010000000110000000001111
61	Pieridae	Dismorphiinae		Dismorphiinae	0100000000000000000000001011
62		Pierinae		Aporia	0100000000000000000000000011
63				Neophasia	0000000000000000000001000001
64				Eucheira	0000000001000000000000000010
65				Pieris/Pontia	0000010000000000000000001000
66				Ascia	0000010000000000000000001001
67				Leptophobia	0000010000000000000000001000
68				Itaballia/Pieriballia	0000010000000000000000000011
69				Perrhybris	0000010000000000000000000011
70				Anapheis/Belenois	0000010000000000000000000011
71				Dixeia	0000010000000000000000000011
72				Pinacopteryx	0000010000000000000000000011
73				Mylothris	0000000000000000000010001010
74				Ixias	0000010000000000000000000011
75				Hesperocharis	0000000000000000000010000010
76				Catasticta	0000000000000000000010000010
77				Pereute	0000000000000000000010000010
78				Delias	0000000000000000000010000010
79				Cepora	0000010000000000000000000010
80				Prioneris	0000010000000000000000000001?
81				Appias	1000010000000000000000001011
82				Elodina	0000010000000000000000000010
83				Leptosia	0000010000000000000000000011
84				Anthocharis/Zegris	0000010000000000000000001000
85				Euchloe/Elphinstonia	0000010000000000000000001000
86				Colotis	1000010000000000000000000011
87				Hebemoia	0000010000000000000000000??
88				Eronia/Nepheronia	1000010000000000000000000??
89				Pareronia/Valeria	0000010000000000000000000??
90			Coliadinae	Phoebis/Aphrissa	0100000100000000000000011??
91				Catopsilia	0100000000000000000000000011
92				Anteos	0100000000000000000000000011
93				Gonepteryx	0100000000000000000000000010
94				Kricogonia	0100000000000000000000000011
95				Dercas	0100000000000000000000000???
96				Eurema	0100010000000000000000001011
97				Terias	1100000001000000000000001011
98				Nathalis	0000000010000000000000001000
99				Colias	1100000001000000000000001011
100				Zerene	0100000000000000000000001000
101	Riodinidae	Riodininae		Abisara	0000000001000000000000000???
102				Dodona	0000000001000000000000000???
103				Zemeros	0000000001000000000000000???
104				Apodemia	0100000000000000001000101?
105				Calephelis	0000000001000100000000001110
106				Caria	0100000000000001000000000?1
107				Emesis	0100000000000000000000001000
108				Melanis	0100000000000000000000000001
109				Hamearis	0000000001000000000000001000
110	Lycaenidae	Curetinae		Curetinae	0100000000000000000000000?1
111		Poritinae		Poritia	0010000000000000000000000001
112		Theclinae	Luciini	Pseudodipsas	0000000101000010000000001?1
113				Paralucia	0000000010000000000000000010
114				Lucia	0001000000000000000000001000
115				Hypochrysops	1110110101000010000100001111
116				Philiris	1100000000000001000000000001
117			Theclini	Thecla	0110000000000000000000000011
118				Quercusia	0010000000000000000000000001
119				Laeosopis	0000000100000000000000000?1
120				Habrodais	0010000000000000000000000001
121				Hypaurotis	0010000000000000000000000001
122			Arhopalini	Arhopala	1010100000000000000000000001
123				Narathura	0000011010000001000000000001
124				Flos	0010100000000000000000000011
125				Surendra	0100000000000000000000000010
126			Ogyrini	Ogyris	0000000000000000000010000010
127			Zesiini	Jalmenus	0100010000000000000000000011
128				Pseudalmenus	0100000000000000000000000011
129			Amblypoidini	Myrina	0100000000000000000000000001
130				Amblypodia	0000000000000000000010000?10

APPENDIX 3. Continued.

#	Family	Subfamily	Tribe	Taxon	RRRRRRRAAAAARPMLPMICPCCHVST 1111123123456aaaaaaaaioy ABCDE 1 2 3	#	Family	Subfamily	Tribe	Taxon	RRRRRRRAAAAARPMLPMICPCCHVST 1111123123456aaaaaaaaioy ABCDE 1 2 3
131				Iraota	0100000000000000000000000??	196				Phlyaria	0100000000000000000000000??
132			Aphnaeini	Aphnaeus/Paraphnaeus	1100000000000000000000000??	197				Cacyreus	0000010100000000000000001000
133				Desmolycaena/Chloroselas	0100000000000000000000000??	198				Harpendyreus	0100000100000000000000001000
134				Axiocerses	0100000000000000000010000??1	199				Leptotes	1100000000000000001000101?
135				Spindasis	0100100100000010000000001?11	200				Syntarucus	0100000000000000000100020?0
136			Iolaini	Iolaus/Stugeta	0000000000000000000100000?1	201				Cyclyrius	0100000000000000000000001010
137				Pratapa	0000000000000000000100000?10	202				Castalius/Zintha	0100000000000000000000000010
138				Tajuria	0000000000000000000100000?10	203				Zizeeria	1101000000000000000010001020
139				Creon	0000000000000000000100000010	204				Zizina	0100000000000000000000001010
140				Jacoona	0000000000000000000100000010	205				Actizera	0101000000000000000000001020
141				Hemiolaus	0000000000000000000100000??	206				Zizula	0101000100000000000000002??0
142			Remelanini	Remelana	0000000001000000000000000??	207				Brephidium	0000010000000000000010001010
143				Ancema	0000000000000000000100000?10	208				Oraidium	0000000000000000000100020?0
144			Catapaecilmatini	Catapaecilma	0000100000000000000000000001	209				Everes	0100000000000000000000001010
145			Loxurini	Loxura	0000000000000010000000001000	210				Cupido	0100000000000000000000001010
146				Yasoda	0000000000000010000000000100	211				Pithecops	0100000000000000000000002??0
147				Eooxylides	0000000000000010000000000100	212				Azanus	0100000000000000000000001?11
148			Cheritrini	Cheritra	0100000000000010000000000001	213				Eicochrysops	0100000000000000000010001012
149				Drupadia	0100000100000000000000000???	214				Neopithecops	0000001000000000000000000010
150			Hypolycaenini	Hypolycaena	0110110100000011000100011111	215				Megisba	1000010000000000000000000002
151				Chliaria	0000000000000010000000001000	216				Celastrina	0110010111101000000000011?1
152				Leptomyrina/Gonatomyrina	0000001000000000000000001010	217				Acytolepis	1100010000000000000000000021
153			Deudorigini	Bindahara	1000000000000000000000000100	218				Vaga	0100001010000000000000000011
154				Sinthusa	0100000000000000000000000010	219				Glaucopsyche	0100000000000000000000001010
155				Rapala	0100110011000000000000002?21	220				Scolitandides	0000001000000000000000000000
156				Deudorix/Virachola	0100110100000001000110001111	221				Philotes	0000001000000000000000002020
157				Artipe	0000000100000000000000000??	222				Philiotiella	0000000000000000000010001000
158				Capys	0000000000000000001000000??	223				Euphilotes	0000000000000000000010001010
159			Tomarini	Tomares	0100000000000000000000001000	224				Pseudophilotes	0000000100000000000000001000
160			Eumaeini	Eumaeus	0000000000000000000000010010	225				Iolana	0100000000000000000000000010
161				Callophrys	0100011111100010000111011111	226				Maculinea	0100000100000000000000001000
162				Erora	0010000000000000000000000001	227				Euchrysops	0100000100000000000000001100
163				Fixsenia	0110000000000000000000000001	228				Lepidochrysops	0100000000000000000000002020
164				Satyrium	1110010101000000000000000?1	229				Oboronia	0000000000000001000000000000
165				Harkenclenus	0100000000000000000000000001	230				Polyommatus	0100000000000000000000001010
166				Chlorostrymon	0000001001000000000000000110	231				Plebejus	0100001100100000000010001010
167				Electrostrymon	0000010000000000000000000010	232				Lycaeides	0100000000000000000000001010
168				Ministrymon	0100000000000000000000000001	233				Cyaniris	0100000000000000000000001000
169				Phaeostrymon	0000010000000000000000000001	234				Aricia	0000010000000000000000001000
170				Strymon	1110011101000010000110010?1	235				Pseudaricia	0000010000000000000000001000
171				Calycopis	1010010000000000000000021010	236				Agriades	0100000001000000000000001000
172				Parrhasius	0010010000000000000000000001	237				Vacciniina	0000000001000000000000000010
173				Tmolus	1100010100000000000000011?1	238				Albulina	0100000000000000000000001000
174				Oenomaus	0000000000000000100000000010	239				Melaegeria	0100000000000000000000000000
175				Arawacus	0000000100000000000000001000	240				Agrodiaetus	0100000000000000000000001000
176				Allosmaitia	1000000000000000000000000100	241				Lysandra	0100000000000000000000001000
177				Atlides	0000000000000000000010000010	242				Plebicula	0100000000000000000000001000
178				Thereus	0100000000000000000000000010	243				Eumedonia	0000001000000000000000001000
179		Lycaeninae		Lycaeninae	0100001001000000000100001110	244				Chilades	0100010000000000000000000??
180		Polyommatinae	Lycaenesthini	Anthene/Trichlema	0110111100000010000000021111	245				Luthrodes	0000000000000000000000010001
181			Candalidini	Candalides	0100010100000011001100011111	246				Freyeria	0100000100000000000000001000
182				Nesolycaena	0000010000000000000000002020	247				Hemiargus/Cyclargus	0100010000000000000000001011
183			Polyommatini	Cupidopsis	0100000000000000000000001000	248	Nymphalidae	Heliconiinae	Pardopsini	Pardopsis	1000000000000000000000002020
184				Nacaduba	0100010001000000001000000011	249			Cethosiini	Cethosia	1000000000000000000000001000
185				Prosotas	0100010000000000000000000021	250			Heliconiini	Heliconiiti	1000000000000000000000000110
186				Catopyrops	0100000000000000000000000011	251				Vindula	1000000000000000000000001000
187				Erysichton	0000010100000000001100000011	252				Cirrochroa	1000000000000000000000000??
188				Neolucia	0100010101000000000100000010	253				Terinos	1000000000000000000000000???
189				Theclinesthes	0100010000000000000000001011	254				Lachnoptera	1000000000000000000000000021
190				Danis	0100000000000000000000000010	255				Cupha	1000010000000000000000000001
191				Jamides	0100010000000000000000001???	256				Vagrans	1000000000000000000000000?1
192				Pepliphorus	0100000000000000000000002??	257				Phalanta	1000000000000000000000000011
193				Catochrysops	0100000000000000000000002010	258				Euptoieta	1000001100001000000100011120
194				Lampides	0100000000000000000000001010	259				Argynnis	1000000000000000000000001000
195				Uranothauma	0100000000000000000000000?1	260				Argyreus	1000000000000000000000001000

APPENDIX 3. Continued.

#	Family	Subfamily	Tribe	Taxon	ABCDE 1 2 1	#	Family	Subfamily	Tribe	Taxon	ABCDE 1 2 1
261				Pandoriana	10000000000000000000000001000	325				Thessalia	00000001000000000000000001000
262				Mesoacidalia	10000000000000000000000001000	326				Texola	00000000100000000000000001000
263				Fabriciana	10000000000000000000000001000	327				Antillea	00000001000000000000000020?0
264				Issoria	10000000000000000000000001000	328				Dymasia	00000001000000000000000010?0
265				Speyeria	10000000000000000000000001000	329				Phyciodes	00000001100000000000000010?0
266				Brenthis	11000000000000000000000001000	330				Anthanassa	10000001000000000000000010?0
267				Boloriiii	11100010010000000001000010?0	331				Eresia	01000001000000000000000010?0
268			Acraeini	Actinote 1	10000000?0001000000000000???	332				Castilia	00000001000000000000000020?0
269				Actinote 2a	11000000000000000000000011??	333				Tegosa	00000000?0000000000000010?1
270				Actinote 2b	01000000000000100000000101?	334				Telenassa	00000001000000000000000020?0
271				Actinote 4a	01000000000000000000000022?	335				Ortilia	00000001?0000000000000011?0
272				Actinote 4b	01000000000000100000000101?	336				Phystis	00000001?0000000000000020?0
273				Actinote 4c	01000000000000100000000101?	337		Limenitinae	Coloburini	Coloburini	01000000000000000000000011
274				Actinote 5a	01001100000000000000000102?	338			Biblidini	Biblis	10000000000000000000000010?
275				Actinote 5b	00000?01000000000000000202?	339				Eurytelii	10000000000000000000000110
276				Actinote 5c	00000100000000000000000012?	340				Myscelia	10000000000000000000000020?
277				Actinote 6a	01000000000000000000000?021	341				Eunica	10000?00000000000000000002
278				Actinote 6b	11000000100000000001000111?	342				Sallya	10000000000000000000000021
279				Actinote 6c	00000000100000000000000010?	343				Catonephele	10000000000000000000000020?
280				Acraea 1	10000000000000000000000011	344				Nessaea	10000000000000000000000020?
281				Acraea 2	10000000000000000000000100	345				Ageroniiti	10000000000000000000000110?
282				Acraea 3a	10000000000000000000000011?	346				Epiphiliti	00000100000000000000000020?
283				Acraea 3b:1	10000000000000000000000100	347				Dynamine	10000000000000000000000020?
284				Acraea 3b:2	10000100100100000000000111	348				Diaethria	01000100000000000000000021
285				Acraea 4a:1	10000000000000000000000022?	349				Cyclogramma	00000100000000000000000100
286				Acraea 4a:2	10000000000010000000000022?	350				Haematera	00000100000000000000000100
287				Acraea 4a:3	10000000000000000000000011?	351				Callicore	00000100000000000000000110
288				Acraea 4b:1	10000000000000000000000100	352			Limenitidini	Limenitis	11000001100000000000000111
289				Acraea 4b:2	10000000000000000000000?11?	353				Basilarchia	11100101000000000000000011
290				Acraea 4b:3	10000000000000000000000???1	354				Ladoga	00000000100000000000000110
291				Acraea 4b:4	10000000000000000000000011?	355				Moduza	00000001000000000000000111
292				Acraea 5 'Bematistes'	10000000000100000000000011?	356				Parasarpa	00000000100000000000000020?
293		Nymphalinae	Nymphalini	Mynes	01000000000000000000000011	357				Adelpha	01001101010000000000000011
294				Hypanartia	01000000000000000000000101?	358				Athyma/Parathyma	10000001000000000000000012
295				Cynthia	01100101?100000000010001100	359				Pandita	00000001000000000000000100
296				Vanessa	01000000000000000000001000	360				Cymothoe	10000001000000000000000011
297				Antanartia	01000000?000000000000001022	361				Harma	10000000000000000000000011
298				Bassaris	01000000000000000000001000	362				Euptera	00000000?000000000000000??
299				Symbrenthia	01000000000000000000001012?	363				Pseudathyma	00000000?000000000000000??
300				Araschnia	01000000000000000000001000	364				Pseudacraea	00000000100000000000000021
301				Inachis	01000000000000000000001000	365				Pantoporia	01000000000000000000000?1?
302				Aglais	01000000000000000000001000	366				Lasippa	01001000010000000000000?1??
303				Nymphalis	11100101000000000001001111	367				Neptis	11001101000000000000000111?
304				Kaniska	00000000000000100000000100	368				Phaedyma	01000101010000000000000111
305				Polygonia	11100010010000000000001111	369				Partheniti	10000000000000000000000?0?0
306			Kallimini	Anartia	00000001000000000000000001000	370				Taenecia	00000000100000000000000???
307				Siproeta	00000001000000000000000001000	371				Euthalia	01100103000000000?000001?
308				Precis	00000001000000000000000101?	372				Lexias	00000000010000000000000???
309				Junonia	01000001000000000001000111?	373				Catuna/Evena	00000000?000000000000000???
310				Hypolimnas	01000111000000000001000?010	374				Hamanumida	00001000000000000000000???
311				Salamis	01000001000000000000000?01?	375				Euryphura	00000?00010000000000000021
312				Yoma	00000101000000000000000?020	376				Euphaedra	00000100000000000000000011
313				Rhinopalpa	01000000000000000000000???	377				Bebearia	00000001010000100000000??1
314				Kallima	00000001000000000000000010	378				Marpesia	01000100000000000000000011
315				Doleschallia	00000001000000010000000?010	379				Cyrestis	01000000000000000000000?001
316				Amnosia	01000000000000000000000010	380				Chersonesia	01000000000000000000000???
317				Catacroptera	00000001000000000000000?010	381				Dichorragia	00000100000000000000000???
318			Melitaeini	Euphydryiti	11100001110000000000000011	382		Charaxinae	Charaxini	Polyura	01000100000000000000000021
319				Mellicta	00000001100000000000000001000	383				Charaxes	11001100010000?10?1000021?
320				Melitaea	00000001100000000000000001000	384				Zingha	10000000000000000000000100
321				Poladryas	00000001000000000000000?020	385				Stonehamia	00000100000000000000000?10
322				Didymaeformia	00000001100000000000000000	386			Euxanthini	Euxanthe	00000100000000000000000001?
323				Cinclidia	00000001100000000000000000	387				Hypomeleana	00000100000000000000000???
324				Chlosyne	00000001100000000000001010	388			Pallini	Pallini	00000010000000000000000120

APPENDIX 3. Continued.

#	Family	Subfamily	Tribe	Taxon	#	Family	Subfamily	Tribe	Taxon
389			Prothoini	Prothoini	417				Caerois
390			Preponini	Archaeoprepona	418			Amathusiini	Amathusiini 1
391				Prepona	419				Amathusiini 2
392				Agrias	420		Brassolinae		Brassolinae
393			Anaeini	Zaretiditi	421		Satyrinae	Haeterini	Haeterini
394				Hypna	422			Biini	Biini
395				Anaea	423			Elymniini	Elymniini
396				Polygrapha	424			Satyrini	Satyrini
397				Consul	425		Calinaginae		Calinaga
398				Fountainea	426		Danainae	Danaini	Danaus
399				Memphis	427				Tirumala
400		Apaturinae		Rohana	428				Amauris
401				Bremeria	429				Parantica
402				Apatura	430				Ideopsis
403				Asterocampa	431			Eupoeini	Euploea
404				Chitoria	432				Idea
405				Doxocopa	433				Lycorea/Ituna
406				Euripus	434				Aneeta
407				Hestina	435		Ithomiinae		Ithomiinae
408				Sasakia	436		Tellervinae		Tellervinae
409				Eulaceura	437		Libytheinae		Libytheinae
410				Helcyra	438	Hesperiidae	Hesperiinae		Hesperiinae
411				Dilipa	439		Megathyminae		Megathyminae
412				Thaleropsis	440		Trapezitiinae		Trapezitiinae
413				Sephisa	441		Pyrginae		Pyrginae
414				Timelaea	442		Pyrrhopyginae		Pyrrhopyginae
415		Morphinae	Morphini	Morpho	443		Coeliadinae		Coeliadinae
416				Antirrhea	444	Hedylidae			Hedylidae

APPENDIX 4

Descriptions of step matrices used in this paper.

Step matrix "a" was used to describe transformation costs between the 15 possible combinations of "Rosid 1B," "other rosids" (all rosids except Rosid 1B), "asterids," and "other seed plants." State numbers translate as follows: **0**, Rosid 1B; **1**, other rosids; **2**, asterids; **3**, other seed plants; **4**, Rosid 1B + other rosids; **5**, Rosid 1B + asterids; **6**, Rosid 1B + other seed plants; **7**, other rosids + asterids; **8**, other rosids + other seed plants; **9**, asterids + other seed plants; **A**, Rosid 1B + other rosids + asterids; **B**, Rosid 1B + other rosids + other seed plants; **C**, Rosid 1B + asterids + other seed plants; **D**, other rosids + asterids + other seed plants; **E**, all groups.

Step matrix "b" was used to describe transformation costs between the seven possible combinations of the plant groups "rosids," "asterids," and "other seed plants." The same step matrix was also used for the seven possible combinations of the growth forms "herbs," "vines," and "trees and shrubs." State numbers translate as follows: **0**, rosids (herbs); **1**, asterids (vines); **2**, other seed plants (trees and shrubs); **3**, rosids + asterids (herbs + vines); **4**, rosids + other seed plants (herbs + trees and shrubs); **5**, asterids + other seed plants (vines + trees and shrubs); **6**, all groups.

a

From: \ To:	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E
0	0	1	1	1	1	2	2	2	2	2	2	3	3	3	3
1	1	0	2	2	2	1	1	1	3	3	3	2	2	2	4
2	1	2	0	2	2	1	3	3	1	1	3	2	2	4	2
3	1	2	2	0	2	3	1	3	1	3	1	2	4	2	2
4	1	2	2	2	0	3	3	1	3	1	1	4	2	2	2
5	2	1	1	3	3	0	2	2	2	2	4	1	1	3	3
6	2	1	3	1	3	2	0	2	2	4	2	1	3	1	3
7	2	1	3	3	1	2	2	0	4	2	2	3	1	1	3
8	2	3	1	1	3	2	2	4	0	2	2	1	3	3	1
9	2	3	1	3	1	2	4	2	2	0	2	3	1	3	1
A	2	3	3	1	1	4	2	2	2	2	0	3	3	1	1
B	3	2	2	2	4	1	1	3	1	3	3	0	2	2	2
C	3	2	2	4	2	1	3	1	3	1	3	2	0	2	2
D	3	2	4	2	2	3	1	1	3	3	1	2	2	0	2
E	3	4	2	2	2	3	3	3	1	1	1	2	2	2	0

b

From: \ To:	0	1	2	3	4	5	6
0	0	2	2	1	1	3	1
1	2	0	2	1	3	1	2
2	2	2	0	3	1	1	2
3	1	1	3	0	2	2	1
4	1	3	1	2	0	2	1
5	3	1	1	2	2	0	1
6	2	2	2	1	1	1	0