Syst. Biol. 49(2):202–224, 2000

More Taxa or More Characters Revisited: Combining Data from Nuclear Protein-Encoding Genes for Phylogenetic Analyses of Noctuoidea (Insecta: Lepidoptera)

ANDREW MITCHELL,[1,3] CHARLES MITTER,[1] AND JEROME C. REGIER[2]

[1]*Department of Entomology, University of Maryland, College Park, Maryland 20742, USA*
[2]*Center for Agricultural Biotechnology, University of Maryland Biotechnology Institute, College Park, Maryland 20742, USA*

Abstract.—A central question concerning data collection strategy for molecular phylogenies has been, is it better to increase the number of characters or the number of taxa sampled to improve the robustness of a phylogeny estimate? A recent simulation study concluded that increasing the number of taxa sampled is preferable to increasing the number of nucleotide characters, if taxa are chosen specifically to break up long branches. We explore this hypothesis by using empirical data from noctuoid moths, one of the largest superfamilies of insects. Separate studies of two nuclear genes, elongation factor-1α (EF-1α) and dopa decarboxylase (DDC), have yielded similar gene trees and high concordance with morphological groupings for 49 exemplar species. However, support levels were quite low for nodes deeper than the subfamily level. We tested the effects on phylogenetic signal of (1) increasing the taxon sampling by nearly 60%, to 77 species, and (2) combining data from the two genes in a single analysis. Surprisingly, the increased taxon sampling, although designed to break up long branches, generated greater disagreement between the two gene data sets and decreased support levels for deeper nodes. We appear to have inadvertently introduced new long branches, and breaking these up may require a yet larger taxon sample. Sampling additional characters (combining data) greatly increased the phylogenetic signal. To contrast the potential effect of combining data from independent genes with collection of the same total number of characters from a single gene, we simulated the latter by bootstrap augmentation of the single-gene data sets. Support levels for combined data were at least as high as those for the bootstrap-augmented data set for DDC and were much higher than those for the augmented EF-1α data set. This supports the view that in obtaining additional sequence data to solve a refractory systematic problem, it is prudent to take them from an independent gene. [Combining data; independent genes; Noctuidae; Noctuoidea; taxon sampling.]

Quantifying empirical support for a phylogeny is now the norm in systematics, and weak support is typically regarded as a problem in need of solving. The solution most often proposed, particularly for molecular studies, is to collect more data. However, given financial and other constraints, one may have to choose between either collecting additional sequence data for the taxa already sampled or sampling additional taxa. The benefits of collecting more characters are evident: By definition, consistent methods of phylogeny estimation will converge on the correct answer or true tree as the number of characters increases. Increasing taxon sampling can help phylogeny estimation by reducing long branch effects, but the benefits are less obvious because the number of potential trees, and thus the size of the estimation problem, increases geometrically with the number of taxa (Felsenstein, 1978a).

Taxon-sampling density has been the subject of much debate in the recent systematics literature, e.g., *Systematic Biology*, March 1998. Many authors have noted that increased taxon sampling might generally increase the accuracy of estimates of phylogeny (e.g., Hillis, 1996; but see Kim, 1996). Graybeal (1998) explicitly addressed the issue in a simulation study, concluding that if taxa are chosen specifically to break up long branches, increasing the number of taxa sampled is preferable to increasing the number of nucleotide characters. However, there have been few empirical studies on these questions.

[3]Current address (and address for correspondence): Department of Biological Sciences, CW-405 Biological Sciences Building, University of Alberta, Edmonton, Canada T6G 2E9. E-mail: am16@ualberta.ca

We present a case study of noctuoid moths (Insecta: Lepidoptera) in the context of this debate. Given previously low levels of support for the hypotheses of interest, in this case the deeper nodes in the tree, we ask whether it is better to increase the number of characters or the number of taxa sampled to improve the overall robustness of the phylogeny? We also expand the question to contrast the possibility of collecting additional characters from the same source, as in the simulations of Graybeal (1998), to that of obtaining such characters from a second, independent gene. With the recent development of several nuclear protein–encoding genes for use in systematics (Slade et al., 1994; Cho et al., 1995; Gupta, 1995; Waters, 1995; Friedlander et al., 1996; Ortí and Meyer, 1996; Fang et al., 1997; Regier and Shultz, 1997; Galloway et al., 1998), and the promise of more to come (e.g., Friedlander et al., 1992; Graybeal, 1994), there is now greater potential for utilizing independent sources of nucleotide characters in a single analysis.

We investigated these questions as part of an ongoing systematic study of the Noctuoidea, one of the largest superfamilies of insects (>45,000 species; Scoble, 1992). Separate studies of two nuclear genes, elongation factor-1α (EF-1α) and dopa decarboxylase (DDC), have shown each to carry much information about noctuoid relationships (Mitchell et al., 1997; Fang et al., 1999). The two genes yield very similar trees for overlapping sets of 49 exemplar species and show almost complete concordance with groupings that have been strongly supported by earlier morphological evidence. However, in each gene tree, support was weak for the deeper nodes, which represent relationships above the subfamily level.

Seeking a more robust phylogeny estimate, in this study we investigate the effect on phylogenetic signal of (1) increasing the taxon sampling by nearly 60%, to 77 species, and (2) combining data from the two genes in a single analysis. To contrast the effect of combining data from different sources to collection of the same total number of characters from a single gene, we simulate the latter by bootstrap augmentation of the single-gene data sets. The results confirm that adding data from a second gene can yield greater benefit for phylogeny reconstruction than obtaining more characters from a single gene. Somewhat surprisingly, however, we find that our increased taxon sampling, even though designed to break up long branches, produced markedly greater incongruence between genes and, if anything, *decreased* support for deeper nodes.

Phylogenetic Framework: Current Understanding of Noctuoid Relationships

Phylogenetic relationships within Noctuoidea have been problematic (Kitching, 1984), but recent work has identified several probable monophyletic groups, for which Poole (1995) and Kitching and Rawlins (1999) cite at least one synapomorphy. We will refer to these as "concordance groups" (Mitchell et al., 1997). Although these groupings are not beyond all doubt, they are surely close approximations; we therefore used recovery of concordance groups as one gauge of phylogenetic utility for EF-1α, DDC, and their combination. We sampled multiple representatives from 24 such groups, 21 of which are indicated in Table 1 and in the figures. The remaining three concordance groups are (Euteliinae + Stictopterinae), *Spodoptera*, and (*S. frugiperda* + *S. ornithogalli*); the monophyly of the latter two groups was confirmed by a morphological study in progress (M. Pogue, pers. comm.).

Of the four traditional noctuoid families, Notodontidae s.l. is widely agreed to be sister group to the others, i.e., Noctuidae, Arctiidae, and Lymantriidae (Miller, 1991; Kitching and Rawlins, 1999). Arctiidae and Lymantriidae are very likely monophyletic groups (Kitching and Rawlins, 1999). In contrast, unambiguous synapomorphies for Noctuidae are lacking (see Speidel et al., 1996, vs. Kitching and Rawlins, 1999). Within Noctuidae, recent reviews support two groups, trifines and quadrifines. Trifines appear to be monophyletic (Poole, 1995; Speidel et al., 1996; Kitching and Rawlins, 1999), but quadrifines probably

TABLE 1. Species of noctuids and outgroups sampled.

Higher taxa[a]	Exemplars[b]	Abbrev.	GenBank accession no.[c] EF-1α	DDC
Notodontidae[d]				
Notodontinae	*Furcula cinerea*	Fci[e]	U85665	AF151539[f]
	Gluphisia septentrionis	Gsep[g]	AF151603[f]	AF151542
Phalerinae	*Datana perspicua*	Dpe[e,g]	U85666	AF151540
Nystaleinae	*Symmerista albifrons*	Sal[e,g]	U85667	AF151541
Heterocampinae	*Nerice bidentata*	Nbid[e,g]	AF151604	AF151543
Arctiidae[d]				
Lithosiinae	*Hypoprepia miniata*	Hmi[e,g]	U85669	AF151547
Arctiinae[d]				
Arctiini	*Estigmene acrea*	Eac[e,g]	U85670	AF151549
	Hyphantria cunea	Hcun[e]	U85671	AF151550[f]
Ctenuchini	*Cisseps fulvicollis*	Cfu	AF151606[f]	AF151548[f]
Lymantriidae[d]				
Lymantriini	*Lymantria dispar*	Ldi[e,g]	U85672	AF151544
Orgyiini	*Dasychira obliquata*	Dob[e,g]	U85673	AF151545
	Orgyia leucostigma	Oleu	AF151605[f]	AF151546[f]
"Quadrifine" Noctuidae				
Aganainae	*Asota caricae*	Asot	AF151607[f]	AF151551[f]
Herminiinae[d]	*Palthis angulalis*	Pan[e]	U85678	AF151552
	Idia aemula	Iae[g]	AF151608[f]	AF151553[f]
Catocalinae	*Catocala ultronia*	Cul[e,g]	U85677	AF151561
	Caenurgina crassiuscula	Ccr[g]	AF151613[f]	AF151562
Calpinae	*Hypsoropha monilis*	Hyp[g]	AF151612[f]	AF151560
Hypeninae	*Hypena scabra*	Psc[g]	AF151609[f]	AF151554
Euteliinae	*Paectes pygmaea*	Ppy[e,g]	U85674	AF151555
Stictopterinae[d]	*Odontodes aleuca*	Oale	AF151611[f]	AF151557[f]
	Lophoptera sp.	Loph	AF151610[f]	AF151556[f]
Nolinae[d]				
Nolini	*Meganola* sp.	Mmi[e,g]	U85675	AF151558
Sarrothripini	*Baileya levitans*	Blev[e]	U85676	AF151559[f]
"Trifine" Noctuidae[d]				
Acontiinae[d]	*Tarachidia candefacta*	Tca[e,g]	U85681	AF151565
	Sprageueia leo	Sleo[e]	U85680	AF151566[f]
Eustrotiinae	*Thioptera nigrofimbria*	Thn[e,g]	U85679	AF151564
	Lithacodia musta	Lms[g]	AF151614[f]	AF151563
Plusiinae[d]	*Anagrapha falcifera*	Afa[e,g]	U85686	AF151567
	Chrysanympha formosa	Cfor	AF151615[f]	AF151568[f]
	Trichoplusia ni	Tni[e,g]	U20140	U71401

are not. Kitching and Rawlins, for example, excluded the quadrifine subfamily Nolinae from Noctuidae, regarding it as related to Lymantriidae and Arctiidae. Three recent molecular studies, using 28S ribosomal RNA and the mitochondrial gene ND1 (Weller et al., 1994), EF-1α (Mitchell et al., 1997), and DDC (Fang et al., 2000), similarly suggest that some quadrifines are related to lymantriids and arctiids, but in none of the studies were these groupings strongly supported. Regarding trifines, Poole (1995)

TABLE 1. Continued

Higher taxa[a]	Exemplars[b]	Abbrev.	GenBank accession no.[c] EF-1α	DDC
Heliothinae[d]	*Pyrrhia exprimens*	Pex[e,g]	U20137	U71430
	Eutricopis nexilis	Ene[e,g]	U20126	U71410
	Schinia arcigera	Sar[e,g]	U20138	U71431
	Adisura bella	Abe[e,g]	U20123	U71407
	Heliocheilus albipunctella	Hal[e,g]	U20127	U71413
	Heliothis (Masalia) terracottoides	Mte[g]	AF151631[f]	U71427
	Heliothis virescens	Hvi[g]	U20135	U71428
	Helicoverpa zea	Hze[e,g]	U20136	U71429
Stiriinae[d]	*Basilodes chrysopis*	Bch[e,g]	U20125	U71405
	Plagiomimicus olvello	Poli[e,g]	AF151620[f]	U71406
Oncocnemidinae[d]	*Oncocnemis obscurata*	Oob[e]	U85685	AF151585[f]
	Catabena lineolata	Clin	AF151621[f]	AF151586[f]
Cuculliinae	*Cucullia convexipennis*	Ccon[e]	U85694	AF151584[f]
Psaphidinae[d]	*Psaphida resumens*	Pres[e,g]	U85695	AF151581
	Feralia major	Fma[e,g]	U85696	AF151579
	Triocnemis saporis	Tsap	AF151619[f]	AF151580[f]
Amphipyrinae	*Amphipyra pyramidoides*	Apyr[e,g]	U85693	AF151578
Pantheinae	*Panthea furcilla*	Pfur[e]	U85684	AF151572[f]
	Charadra deridens	Cder[e]	U85683	AF151573[f]
Raphiinae	*Raphia abrupta*	Rabr[e]	U85689	AF151574[f]
Acronictinae	*Acronicta* sp. 1	Asp1[g]	AF151618[f]	AF151575
	Acronicta sp. 2	Asp2[e]	U85687	AF151576[f]
	Polygrammate hebraeicum	Phe[e,g]	U85688	AF151577
Condicinae[d]				
Condicini	*Condica videns*	Cvi[e,g]	U85682	AF151569
	Homophoberia sp.	Homo	AF151616[f]	AF151570[f]
Leuconyctini	*Leuconycta diptheroides*	Ldip	AF151617[f]	AF151571[f]
Agaristinae[d]	*Eudryas grata*	Egr[e,g]	U85690	AF151582
	Psychomorpha epimenis	Pepi[e,g]	U85691	AF151583
Eriopinae	*Callopistria mollissima*	Cmol	AF151622[f]	AF151587[f]
Noctuinae s.l.[dh]				
"Caradrinini"				
Prodeniini	*Spodoptera frugiperda*	Sfr[e,g]	U20139	U71403
	Spodoptera exigua	Sex[g]	AF151624[f]	U71404
	Spodoptera ornithogalli	Sor	AF151623[f]	AF151588[f]
	Elaphria grata	Elg[e,g]	U85697	AF151589
	Galgula partita	Gpa	AF151626[f]	AF151591[f]
	Anorthodes tarda	Atar	AF151625[f]	AF151590[f]
Dypterygiini	*Nedra ramulosa*	Nra[e,g]	U85698	AF151592

used a broad definition of the group, and combined the true cutworms of Lafontaine (1993) into an expanded subfamily Noctuinae s.l., containing most of the trifine species. Kitching and Rawlins (1999) used a more restricted definition of trifines, excluding Acronictinae and associated taxa as well as Pantheinae from this group. One goal of the present study was to further test these hypotheses of noctuid nonmonophyly and of the boundaries of the trifine clade.

TABLE 1. Continued

Higher taxa[a]	Exemplars[b]	Abbrev.	GenBank accession no.[c] EF-1α	DDC
Hadenini				
Hadenina	*Lacinipolia renigera*	Lre[e,g]	U85700	AF151594
	Pseudaletia unipuncta	Pun[g]	AF151627[f]	AF151595
Eriopygina	*Orthodes crenulata*	Ocr[e,g]	U85699	AF151593
Apameini[d]	*Apamea amputatrix*	Aamp	AF151629[f]	AF151597[f]
	Papaipema sp.	Papp	AF151628[f]	AF151596[f]
Xylenini[d]	*Anathix ralla*	Aral[e,g]	U85702	AF151598
	Lithophane hemina	Lhem[e,g]	U85701	AF151599
Noctuini s.l.[d]				
Noctuina	*Abagrotis alternata*	Aalt	AF151630[f]	AF151601[f]
Agrotina	*Agrotis ipsilon*	Aip[e,g]	U85704	AF151600
Aniclina	*Anicla infecta*	Ainf[e,g]	U85703	AF151602

[a]Nomenclature follows Kitching and Rawlins (1999) and Poole (1995); where these classifications were in conflict, we used the more finely split alternative.

[b]All species were collected in the USA, except as follows: *Asota caricae, Odontodes aleuca,* and *Lophoptera* sp. from Thailand; *Heliocheilus albipunctella* and *Heliothis (Masalia) terracottoides* from Mali.

[c]The entire alignment is available from the Society of Systematic Biologists web site (http://www.utexas.edu/admin/systbiol/).

[d]Concordance groups, supported by morphological synapomorphies. The following are also concordance groups: Euteliinae + Stictopterinae, Cuculliinae + Oncocnemidinae, *Spodoptera,* and *S. frugiperda* + *S. ornithogalli.*

[e]Species included in reduced data set of 49 taxa, closely matching the taxon sampling of Mitchell et al. (1997).

[f]Gene sequences new to this study.

[g]Species included in reduced data set of 49 taxa, closely matching the taxon sampling of Fang et al. (2000).

[h]Although there is strong evidence for the monophyly of Noctuinae s.l., its limits are poorly defined; "Caradrinini" is placed here tentatively (Poole, 1995).

MATERIALS AND METHODS

Species Sampled

Table 1 lists the species sampled and GenBank accession numbers of their EF-1α and DDC partial sequences used in this study, arranged according to recent proposals on noctuoid classification. The new data comprise 56 gene sequences. The entire alignment is also available from the Society of Systematic Biologists web site (http://www.utexas.edu/admin/systbiol/).

Our earlier studies of Noctuoidea, using EF-1α (Mitchell et al., 1997) and DDC (Fang et al., 2000), sampled 49 exemplar species each, only 35 of which were shared between data sets. Because each data set was sensitive to taxon-sampling effects, we decided not to attempt combining data for analysis until we had determined both gene sequences for most species. This goal was achieved for 60 species, after which sequence data for both genes were obtained from an additional 17 species, chosen to break up long branches present in trees from our earlier analyses. Seven of these new species were from previously unsampled clades within Noctuoidea (Ctenuchini, Apameini, Eriopinae, and the Old World subfamilies Aganainae and Stictopterinae), seven were from previously underrepresented clades (Lymantriidae, Plusiinae, Oncocnemidinae, Psaphidinae, Condicinae, and "Caradrinini") and the last three were species of Heliothinae, as reported in Cho et al. (1995) and Fang et al. (1997). All told, our sample includes all four traditional families; 23 of the 31 subfamilies of Noctuidae recognized by Kitching and Rawlins (1999), including the two noctuid subfamilies which they raise to family rank; and 27 of the 45 tribes recognized (both explicitly and implicitly) within Noctuidae.

Laboratory Protocols

Specimens were either live-frozen at −80°C or collected directly into 100% ethanol at ambient temperature in the field and subsequently stored at temperatures as cold as −20°C for as long as 6 months before being cooled to −80°C.

TABLE 2. Sequences of DDC primers used in this study. International Union of Biochemistry (1986) codes are used to indicate degeneracy. All primers included an M13 sequence at the 5′ end to facilitate automated sequencing. Primer names ending in F identify forward primers, which bind to the antisense strand of DNA; primer names ending in RC identify reverse complement primers.

Primer	Sequences (5′-3′)	Position[a]
1.7sF	GCH TGY ATY GGN TTY WCN TGG AT	471-493
1.9sF	ATG HTN GAY TGG YTV GGY CAR ATG	528-551
3.2sF	TGG YTN CAY GTN GAY GCN GCN TAY GC	975-1000
3.3sRC	CCA YTT RTG NGG RTT RAA RTT RAA	1065-1088
4sRC	GG DAT YTG CCA RTG HCK RTA RTC	1203-1225

[a]Nucleotide position relative to the DDC sequence of *Manduca sexta* (GenBank accession no. U03909).

Wings and abdomens were removed from specimens and vouchered. Genomic nucleic acids were extracted from the thorax (head + thorax for smaller specimens) by using a commercially available DNA/RNA isolation kit (Amersham Corp., Arlington Heights, IL). Polymerase chain reaction (PCR) primer sequences for EF-1α were those from Cho et al. (1995). DDC primers incorporated slightly more degeneracy than those of Fang et al. (1997) and are listed in Table 2. An additional DDC primer (3.3sRC) was also made. The 3.3sRC primer binds downstream from 3.2sF and allows amplification of two overlapping PCR fragments (1.7sF/3.3sRC or 1.9sF/3.3sRC and 3.2sF/4sRC). PCR amplification protocols for EF-1α and DDC were unchanged from Mitchell et al. (1997) and Fang et al. (1997), respectively. Note that reverse transcription (RT)-PCR was used to avoid introns in DDC. To obtain purer templates for sequencing, we always used a nested PCR approach; that is, a first amplification used external primers, and reamplifications utilized nested internal primers. Sequences were generated with an Applied Biosystems 373A Stretch DNA sequencer using Taq polymerase and dye-primer. Templates were sequenced in both directions.

Data Analysis

We used the XDAP program within the software package STADEN (Staden, 1992) to check chromatograms for accuracy of base-calling and to assemble contiguous DNA fragments. Sequences were aligned manually by using the Genetic Data Environment software package (GDE 2.2; Smith et al., 1994). No gaps were needed for alignment.

Nucleotide sequences for each gene were first analyzed separately for the 77-taxa data sets, and then sequences for the two genes were concatenated for the combined data analyses. To facilitate direct comparisons between our current data set with 77 taxa and those of our earlier studies with 49 taxa each (Mitchell et al., 1997; Fang et al., 1999), we also analyzed reduced data sets with 49 taxa each. Taxa included in these smaller data sets are indicated in Table 1 and were selected to match the taxon sampling of the earlier studies as closely as possible. Exact matches were not possible because DDC sequences could not be obtained for some of the species used in the earlier EF-1α study, and vice versa.

To test the hypothesis that the increased support levels seen in the combined data set were a direct consequence of combining data from independent genes, as opposed to simply increasing the total number of informative characters, we made additional data sets as follows. Seqboot, part of the Phylip package (Felsenstein, 1995), was used to generate bootstrap pseudodata sets from the original data for each gene, separately. For EF-1α, the first 709 bp from a bootstrap pseudodata set were added to the 1,240 bp of the original data set to get 1,949 bp. For DDC, 1,240 bp from three separate bootstrap pseudodata sets were added to the 709 bp of the original data set to get 1,949 bp. To account for the considerable random variation expected among bootstrap pseudodata sets (Hillis and Bull, 1993), we repeated this procedure twice. Thus there were three bootstrap-aug-

mented data sets for either gene, each equal in size to the combined data set.

Parsimony analysis.—Calculation of uncorrected pairwise divergences and all maximum parsimony (MP) analyses were performed by using test versions of PAUP*4.0 (D. Swofford, pers. comm.). MP analyses consisted of heuristic searches with 1,000 random addition sequences and tree bisection-reconnection (TBR) branch swapping, unless stated otherwise. Initial MP analyses used equal weights for all taxa. Subsequent analyses used six-parameter parsimony step matrices, calculated by using a spreadsheet described by Cunningham (1997a) and supplied to us by the author. In these analyses, transformations were weighted by the negative of the natural logarithm of their frequencies (our "ln-weighting").

Bootstrap percentages (BPs) were based on 1,000 bootstrap replicates with three random addition sequences per replicate, for each gene separately and for combined data, using equal weighting for all characters. Each bootstrap analysis performed under ln-weighting was restricted to 200 replicates, because the use of step matrices increased computer run time approximately fivefold.

Partitions.—Before combining the single-gene data sets in an effort to increase phylogenetic resolution, we considered how, and indeed whether, a combined analysis should be performed. Proponents of conditional combination argue that the degree of incongruence among data partitions should determine whether independent data sets can be combined (e.g., Swofford, 1991; Bull et al., 1993; Huelsenbeck et al., 1996). The incongruence length difference (ILD) test (Farris et al., 1994) appears to be the most appropriate method currently available for this purpose, but it may be too conservative a test (Sullivan, 1996; Cunningham, 1997b). Thus, adherence to conventional significance levels under this test could mean not combining some independent data sets, even when doing so would yield a more robust and accurate phylogeny. To investigate this possibility, we carried out a combined analysis for the 77-taxa data set despite a significant incongruence test result.

The ILD test was performed in PAUP*4.0b2 (D. Swofford, 1999). As suggested by Cunningham (1997a), invariant characters were excluded before the tests were performed. Details of parameters used for each test are given under Results. All ILD tests used 200 replicates with random addition sequences for taxa and TBR branch swapping. Tests performed on the reduced 49-taxa data sets used 10 random addition sequences, whereas tests performed on the 77-taxa data sets used three random addition sequences per replicate. To determine the contributions of individual taxa to overall significance of ILD test results, we excluded taxa for which placement varied markedly between genes or within a gene among methods and repeated the tests.

Distance analyses.—Some data partitions showed significant heterogeneity of base frequencies among taxa. Nonstationarity of base frequencies can cause inconsistency of some phylogenetic methods, but LogDet distances (Lockart et al., 1994) are robust to nonstationarity (Swofford et al., 1996). An initial minimum evolution (ME; Rzhetsky and Nei, 1992) analysis using LogDet distances with invariable sites was performed for comparison with ME trees derived under the preferred maximum likelihood (ML) model (see below). A ME bootstrap analysis also was performed, with the ML distances estimated by using the same model and parameters as the ML analyses (below). Five hundred bootstrap replicates were performed, each of which used five random-addition sequences for taxa and TBR branch swapping. Complete-and-partial bootstrap analyses were carried out by using the software "njbootli" (Zharkikh and Li, 1995), which uses a synonymous versus nonsynonymous rate model, based on Li (1993). This technique corrects bootstrap values based on the number of taxa in a data set, facilitating comparison of support levels among data sets with different taxon samples.

Comparison of support levels.—Support levels were compared in two ways. First, the overall frequency of strong support was used as a measure of resolution strength per se, regardless of which groups these were. Second, clade-by-clade comparisons of support levels were made, in which we also kept track of concordance groups. This provides a combined measure of support

and accuracy. For comparisons between 49-taxa and 77-taxa data sets, we considered only those groups represented by multiple taxa in both data sets. We used sign tests, as described by Zar (1984), to evaluate the statistical significance of differences in BPs between data sets; only the direction (sign) of differences between members of a pair is used in this test, not the actual difference. Nodes that broke up concordance groups and therefore were expected to have low support levels were treated differently from other groups; that is, the sign was reversed before performing the test. Differences of zero are not considered in sign tests, so n is the number of differences having a sign.

Maximum likelihood analyses.—ML analyses were performed on the 77-taxa combined data set by using PAUP*4.0b2 (Swofford, 1999). As suggested by Frati et al. (1997), 16 models (four substitution models permuted with four rate-distribution models) were evaluated on the most-parsimonious (mp) trees derived under ln-weighting. Likelihood ratio tests are used to determine which models have the significantly higher likelihood scores. The best model is then chosen from the set of models with the best score, preference being given to the model with the fewest free parameters. This model was selected for further analyses, with all parameters estimated once from the data on the mp trees, and fixed subsequently. The ML heuristic search strategy entailed (1) TBR branch swapping to completion on user-input trees, including the mp trees obtained under equal and ln-weighting, and ME trees obtained under the preferred ML model and the LogDet model; (2) 15 random addition sequence replicates for taxa with nearest neighbor interchange (NNI) branch swapping to completion; (3) 10 random addition sequence replicates for taxa with TBR branch swapping, with only a single replicate swapping to completion because of time constraints. The Kishino–Hasegawa test (Kishino and Hasegawa, 1989) was used to choose among ML trees.

Results

Third-codon position base composition was significantly heterogeneous among taxa for both EF-1α and DDC (χ^2-tests, P = 0.013 and 0.000, respectively). For EF-1α, exclusion of Arctiidae and Lymantriidae from the data set resulted in homogeneity of base composition (P = 1). For DDC we had to exclude all or part of Notodontidae, Arctiidae, Lymantriidae, and quadrifine Noctuidae before nonsignificance was achieved. However, the LogDet-model ME trees recovered the same concordance groups as the ln-weighted mp tree and differed from the ML-model ME trees only in minor rearrangements among adjacent nodes.

ILD tests performed on the two combined-gene 49-taxa data sets yielded nonsignificant results under both equal weighting (P = 0.450 and 0.500) and ln-weighting (P = 0.210 and 0.570).

For the 77-taxa data set results of ILD tests are summarized in Table 3. Under equal weighting, an initial ILD test suggested statistically significant incongruence between the two genes (P = 0.005, the minimum value possible for 200 test replicates). Incongruence became less significant on exclusion of certain relatively long-branch taxa (P = 0.035). Under ln-weighting, ILD test P-values were usually higher, and a marginally significant test result was obtained by excluding eight taxa (P = 0.040).

Maximum Parsimony Analyses, 77 Taxa

Table 4 summarizes the more notable differences among the mp trees recovered by each gene and by the combined data set under both equal weighting and six-parameter parsimony ln-weighting. The number of concordance groups recovered by data partitions under different weighting schemes is shown in Table 5. Summary statistics for various data partitions plotted onto one of the mp trees are shown in Table 6.

EF-1α.—Uncorrected pairwise divergence values (total nucleotides) for EF-1α ranged from 1.0% (within *Spodoptera*) to 10.7% (Aganainae + Herminiinae vs. Arctiidae + Lymantriidae). EF-1α divergence values were greater within the Arctiidae + Lymantriidae clade because of differences in third-codon position base composition (Mitchell et al., 1997). Equally weighted MP analysis of EF-1α alone resulted in 25 mp trees of 2,736 steps each, with the retention index (RI) = 0.443. The strict consensus tree

TABLE 3. Incongruence length difference tests on 77-taxa data set.

Taxa excluded[a]	No. of taxa excluded	*P*-values	
		Equal weight	ln-weight
None	0	0.005	0.010
Hmi	1	0.005	0.035
Blev	1	0.015	0.005
Hmi, Blev	2	0.035	0.020
Hmi, Blev, Gsep	3	0.010	0.020
Nolinae	2	0.035	0.020
Nolinae, Hmi, Gsep	4	0.005	0.010
Acontiinae	2	0.010	0.005
Plusiinae	3	0.005	0.010
Plusiinae, Hmi	4	0.035	0.040
Plusiinae, Hmi, Gsep	5	0.025	0.040
Plusiinae, Acontiinae, Hmi	6	0.015	0.020
Nolinae, Acontiinae	4	0.005	0.005
Nolinae, Plusiinae	5	0.035	0.015
Acontiinae, Plusiinae	5	0.010	0.005
Notodontidae	5	0.005	0.005
Acontiinae, Plusiinae, Gsep, Hmi, Blev	8	0.015	0.040

[a]Abbreviations of species names are given in Table 1.

had only 41 resolved nodes out of a possible 75. Under ln-weighting a single mp tree was found, with a length of 3,869 steps and RI = 0.458 (Fig. 1, left).

DDC.—Pairwise divergence values for DDC ranged from 2.8% (within Xylenini) to 26.7% (within Nolinae). Equally weighted MP analysis of DDC alone resulted in four mp trees of 4,661 steps each, with RI = 0.456. The strict consensus tree had 66 resolved nodes, many more than EF-1α (equally weighted), the conflict being primarily among the clades of higher trifine noctuids. The ln-weighting scheme gave two mp trees with lengths of 7,255 steps and RI = 0.474. The strict consensus of these trees had a 12-branch polytomy at the base of the trifine clade. This polytomy was primarily the result of alternative placements of Acontiinae, either at the base of the trifine noctuid clade or as a highly derived trifine group. The similarity between the two trees is thus better illustrated by an Adams consensus tree (Fig. 1, right).

Combined data.—Equally weighted MP analysis of the combined data set gave four

TABLE 4. Summary of most notable differences among trees inferred from equally weighted versus ln-weighted parsimony analyses.

Relationships recovered in mp trees	EF-1α only		DDC only		Combined data	
	Eq.	ln	Eq.	ln	Eq.	ln
Ingroup monophyletic?	N	Y	Y/N	Y	Y	Y
Arctiidae monophyletic?	N	N	Y	Y	Y/N	Y
(Arctiidae + Lymantriidae) monophyletic?	Y/N	Y	N	N	Y/N	Y
Nolinae placed with other quadrifines?	Y/N	Y	Y	Y	N	Y
Plusiinae placed within trifine Noctuidae?	N	Y	N	N	N	N
Acontiinae placed within trifine Noctuidae?	Y/N	Y	N	Y	Y	Y
Acontiinae basal within trifine Noctuidae?	N	Y	N/A	Y/N	N	Y

Eq., equal weighting; ln, six-parameter parsimony, logarithmic weighting; Y/N, yes for some trees and no for others; N/A, not applicable.

TABLE 5. Recovery of concordance groups by 77-taxa data sets under different methods.

Data sets	MP eq. wt.[a]	MP ln wt.[a]	ML[a]
Combined	21	21	21
EF-1α	13	17	–
DDC	19	20	–
EF-1α BA1[b]	13	–	–
EF-1α BA2[b]	12	–	–
EF-1α BA3[b]	15	–	–
DDC BA1[b]	20	–	–
DDC BA2[b]	19	–	–
DDC BA3[b]	18	–	–

[a]Maximum number possible = 24; dashes indicate analysis not performed.
[b]Bootstrap-augmented data sets with 1,949 bp from single genes.

mp trees, the strict consensus of which had 71 resolved nodes. Three mp trees were recovered under ln-weighting, with lengths of 11,259 steps and RI = 0.461 (Fig. 2).

Maximum Likelihood Analyses, 77 Taxa

The best-fitting likelihood model for the ln-weighted mp tree was the most general and parameter-rich model, the general time reversible model (GTR; Lanave et al., 1984; Rodriguez et al., 1990), with invariant sites and gamma-distributed rates, that is, GTR + I + Γ. This model had a score of −ln L = 33,226.45 (where L refers to likelihood), which proved significantly better than the second-best model, HKY85 (Hasegawa et al., 1985) + I + Γ, based on the likelihood ratio test with 4 df (Δln L = 54.07, $P < 0.001$).

Table 7 shows the estimated parameters of the GTR substitution model with rate heterogeneity, for different partitions of the data, based on the mp tree (Fig. 2). EF-1α had ~1.5 times as many invariable sites as DDC and showed somewhat greater rate heterogeneity, even when the proportion of invariable sites in each gene was accounted for. Site-specific rate models (not tabulated) yielded the following rate estimates: the partition EF-1α:DDC gave relative rates of ~1:3.3, and the partition first:second:third codon position gave relative rates of 4:1:81 and 2.7:1:22 for EF-1α and DDC, respectively.

The best tree recovered for the combined data set under the ML criterion (Fig. 3) had

TABLE 6. Summary statistics for data partitions mapped onto one of four mp trees for the combined data set, under equal weighting.

	No. characters	No. (%) variable	No. (%) inform[a]	No. steps on tree	CI	RI
Combined	1,949	803 (41)	666 (34)	7,485	0.168	0.443
EF-1α						
All sites	1,240	399 (32)	329 (27)	2,792	0.196	0.428
nt1	413	35 (8)	15 (4)	145	0.174	0.450
nt2	413	18 (4)	6 (1)	37	0.259	0.500
nt3	414	346 (84)	303 (73)	2,610	0.197	0.426
DDC						
All sites	709	404 (57)	337 (48)	4,693	0.152	0.452
nt1	236	110 (47)	79 (33)	609	0.206	0.544
nt2	236	65 (28)	32 (14)	237	0.262	0.570
nt3	237	229 (97)	226 (95)	3,847	0.137	0.428

CI = consistency index, excluding uninformative characters; RI = retention index.
[a]Parsimony informative characters.

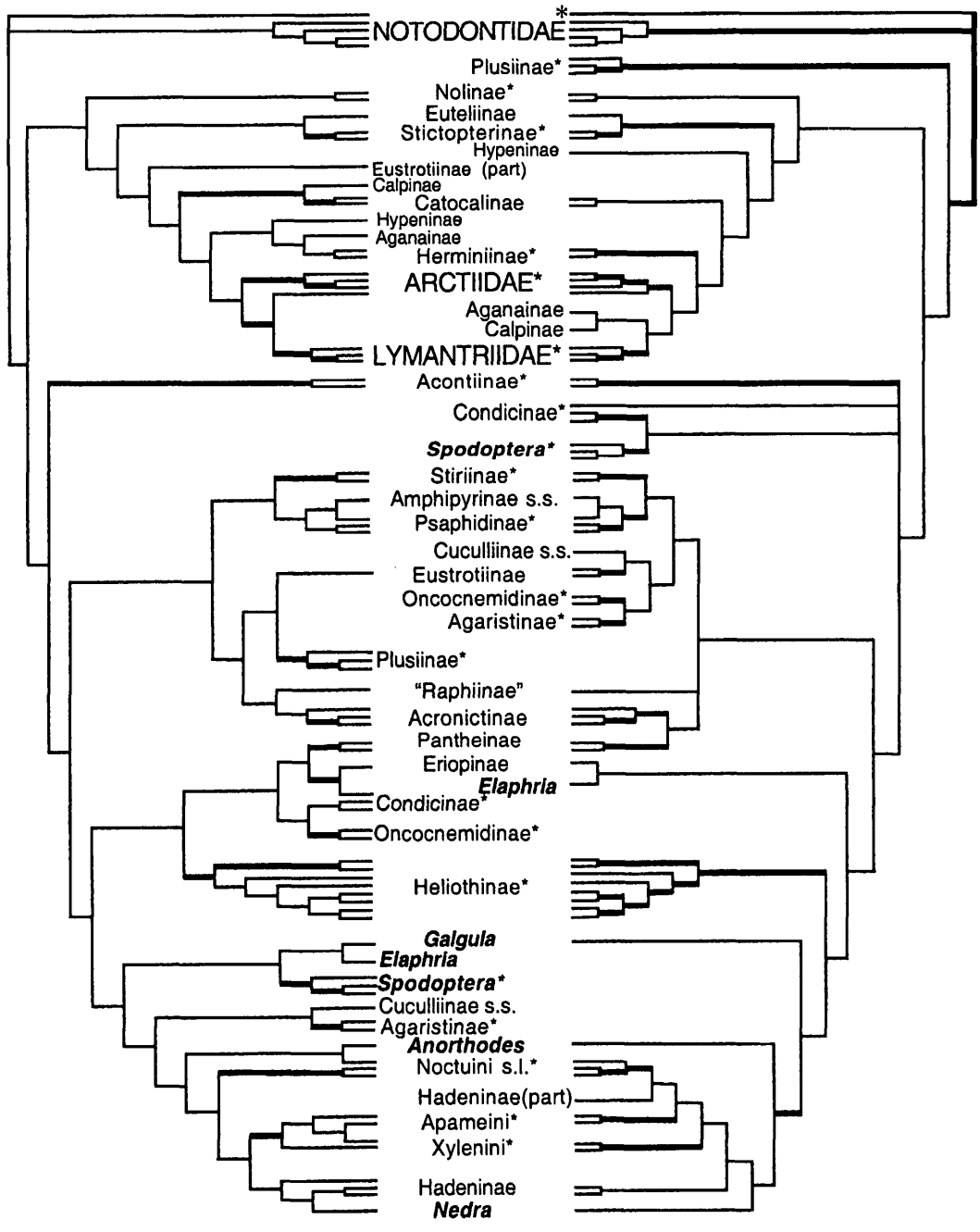

FIGURE 1. Comparison of mp trees derived under ln-weighting for EF-1α and DDC gene sequences. For DDC, tree shown is an Adams consensus of 2 mp trees. Thick branches have ≥70% bootstrap support. The genera of "Caradrinini" are bold. Asterisks indicate concordance groups. Four additional concordance groups are Euteliinae + Stictopterinae, Cuculliinae + Oncocnemidinae, *Spodoptera* (Sex + Sfr + Sor), and *S. frugiperda* + *S. ornithogalli* (Sfr + Sor).

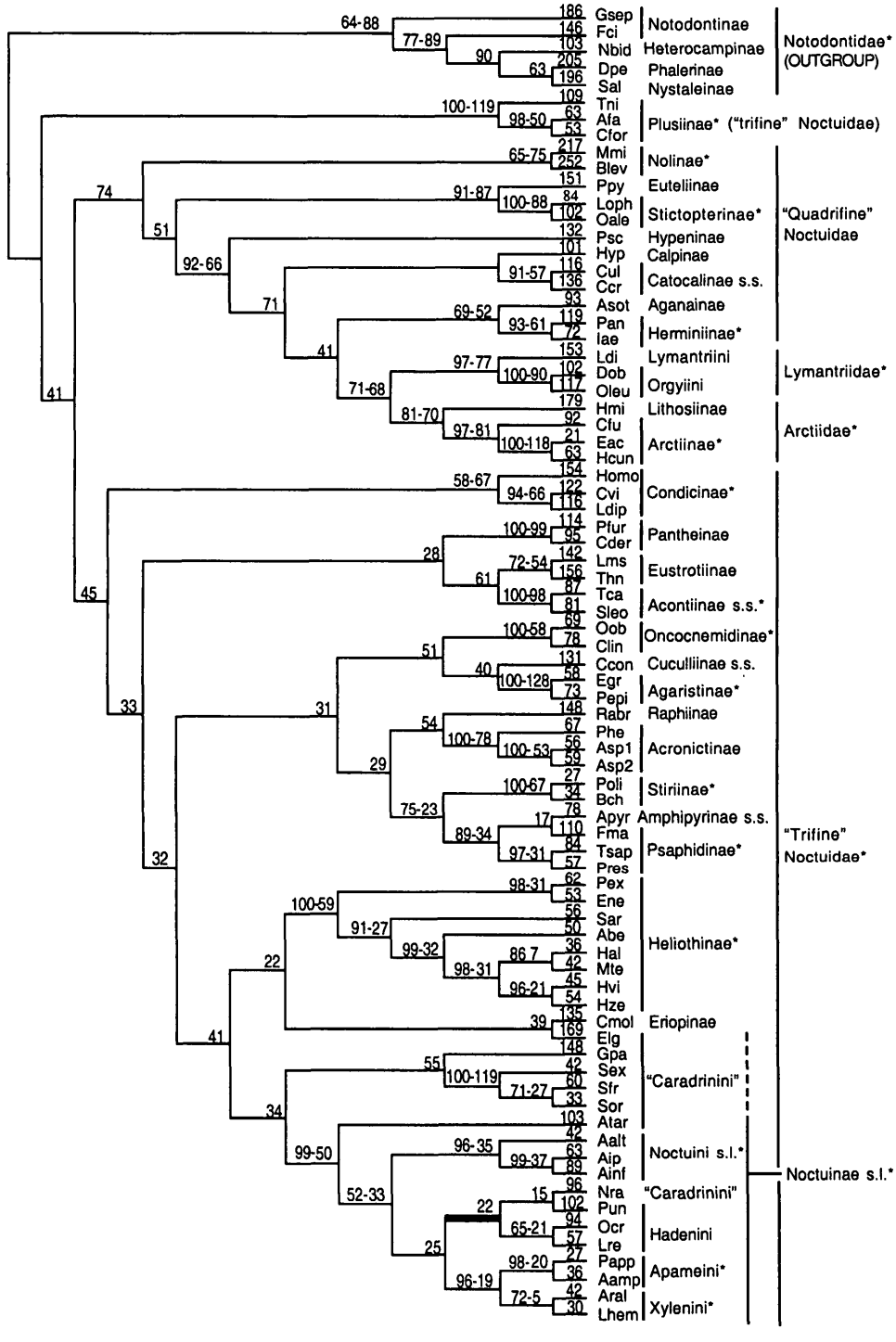

FIGURE 2. One of three mp trees for the combined data set, under ln-weighting. The single branch that collapses in the strict consensus tree is indicated by a thick line. Numbers above branches are bootstrap percentages (shown only if ≥50%), followed by branch lengths under ACCTRAN optimization. Asterisks indicate concordance groups. Four additional concordance groups are as detailed in Figure 1. Taxa included in the Noctuinae s.l. by Poole (1995) but not strongly allied with that group by our data are indicated by a dotted line.

TABLE 7. Parameters for the general time reversible (GTR) substitution model with rate heterogeneity, estimated by maximum likelihood on the mp tree derived under ln-weighting.

Parameters	EF-1α	DDC	Combined
Γ-distr. shape parameter (α)[a]	0.82	0.94	0.86
Proportion of invariable sites (ϕ)[a]	0.64	0.45	0.58
α estimated with ϕ = 0	0.20	0.31	0.23
Relative substitution rate parameters[b]			
A-C	2.163	1.939	2.044
A-G	11.32	7.741	9.085
A-T	4.402	2.160	3.027
C-G	1.825	1.140	1.302
C-T	20.01	7.327	10.44
G-T	1	1	1

[a]α and ϕ estimated simultaneously.
[b]R-matrix values.

a score of −ln L = 33,181.26. Nine of the 12 random addition sequence replicates using TBR branch swapping, including the single replicate that was allowed to swap to completion, found this tree. The latter replicate used approximately the same time as all 15 random addition sequence replicates that used NNI branch swapping, combined. None of the NNI-swapped replicates found trees as good as the TBR-swapped trees, but 13 replicates found trees that were not significantly worse (Kishino–Hasegawa tests, $P > 0.05$).

The relationships recovered by the best ML tree (Fig. 3) were similar to those of the mp tree under ln-weighting (Fig. 2). The most notable difference was in the placement of Condicinae, which was the basal trifine lineage under MP but a much more derived group under ML. Although none of the intervening nodes were strongly supported under either MP or ME, all the mp trees had significantly lower likelihoods than the ML tree (Kishino–Hasegawa tests, $P \leq 0.040$). The two methods recovered the same number of concordance groups, although Psaphidinae was recovered by ML only, and the expected relationships within *Spodoptera* were recovered by MP only.

The following rearragements of taxa could be performed under the ML criterion without significantly altering tree scores, based on Kishino–Hasegawa tests: Plusiinae could be moved to the base of the trifine clade; *Cucullia convexipennis* (Ccon) could be placed as sister group to Agaristinae, as in the mp trees; *Raphia abrupta* (Rabr) could be placed as sister group to Pantheinae; and the four basal-most clades of trifines in the ML tree could be united as a monophylextic group.

Taxon Sampling: Effects on Support Levels

EF-1α analyses yielded 18 groups with multiple representatives in both the 49-taxa and 77-taxa data sets (Table 8). Under MP, half of these groups showed <10% difference in BP between data sets and were not considered further. Of the remaining nine clades, seven had greater BPs with 49 taxa than with 77 taxa (clades 1, 7, 9, 19, 27, 29, and 33); only two had greater BPs with 77 taxa (clades 4 and 8). However, clade 8 breaks up a concordance group (Arctiidae), which means that the lower BP obtained with the 49-taxa data set is more congruent with existing phylogenetic evidence. Thus, for EF-1α, eight of nine clades have "better" BPs with 49 taxa than with 77 taxa. This trend is statistically significant (sign test, $0.05 < P < 0.02$).

For DDC, 19 BP comparisons were available (note that the DDC and EF-1α 49-taxa data sets sampled different taxa). Under MP, 12 clades showed <10% difference between data sets and were not considered further. Of the remaining seven clades, three had greater BPs with 49 taxa (clades 1, 4, and 33) and four had greater BPs with 77

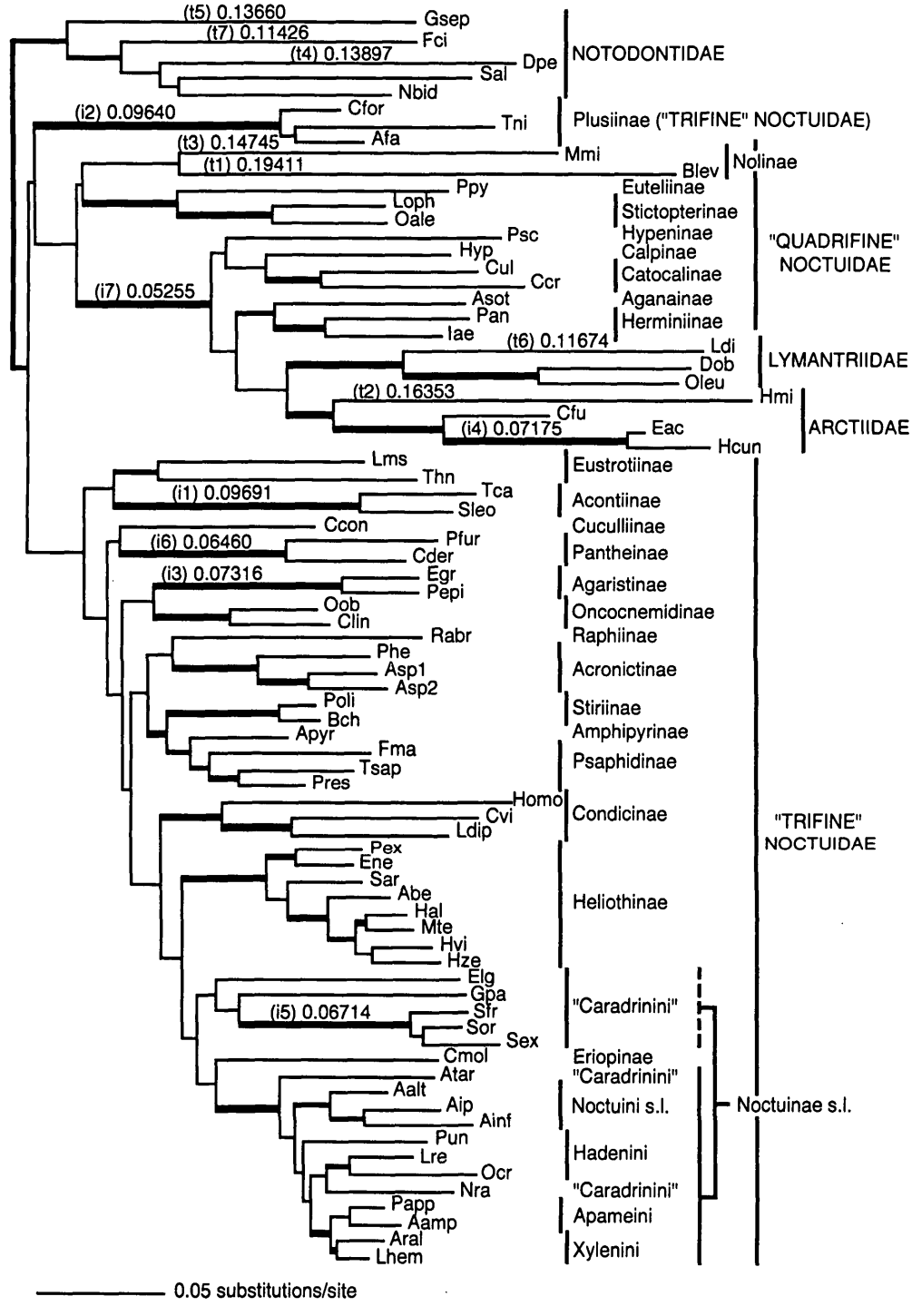

FIGURE 3. Maximum likelihood tree obtained under GTR + I + Γ model, −ln L = 33,181.26. Branches with ≥70% and ≥90% bootstrap support under ME analysis using the same ML model are drawn 2× and 3× thicker, respectively. The seven longest terminal branches and the seven longest internal branches are indicated by "t" or "i," respectively, and their rank order in parentheses above the branch; following the parentheses is the branch length in number of substitutions per site. The dotted line indicates taxa included in the Noctuinae s.l. by Poole (1995) but not strongly allied with that group by our data.

taxa (clades 2, 9, 19, and 29). The sign test for BP comparisons was nonsignificant ($P > 0.5$).

The complete-and-partial bootstrap, designed to correct BPs for the number of taxa used and implemented for neighbor-joining (Zharkikh and Li, 1995) was performed next. For EF-1α, all five of the clades that showed ≥10% difference in BP between data sets (clades 6, 7, 19, 25, and 27) had larger BPs for the 49-taxa data set. This trend was not statistically significant because of the small number of comparisons available ($n = 5$; $0.5 < P < 0.2$). For DDC, eight of nine comparisons (clades 1, 2, 15, 17, 25, 29, 30, and 33) showed the same trend ($0.05 < P < 0.02$).

Combining Data: Effect on Support Levels

Overall support, measured by the percentage of nodes with parsimony BP ≥ 50%, was greater for the combined-data mp trees than for either gene alone: Under ln-weighting, 48% and 55% of nodes in the EF-1α and DDC trees, respectively, had BP ≥50%, whereas 65% of the nodes in the combined-data mp tree are supported at that level.

The effect of combining data on levels of support for particular groups, under three different methods of analysis, is shown in Table 8 (ignoring the values in parentheses). The five clades that break up concordance groups (i.e., clades 3, 8, 35, 36, and 37) almost certainly represent incorrect relationships. It is therefore better that these clades receive little support. Indeed, under equally weighted parsimony, these clades have substantially lower BPs in the combined analysis than the highest value for either gene analyzed separately. Of the remaining 32 clades, all but seven have higher BPs in the combined analysis than for each gene alone. For five of the remaining seven clades (11, 13, 20, 33, and 34), the combined-data BP is much nearer the higher than the lower of the two BPs from each gene alone. Of the remaining two clades, one (clade 9) reflects strong conflict between the two genes over the placement of the arctiid *Hypoprepia miniata*, and the other (clade 21) reflects conflict, albeit weaker, over the placement of *Raphia abrupta*. Thus, 30 of 37 clades have better BPs in the combined analysis and only two clades have substantially better BP in the analysis of a single gene. This difference is highly significant (sign test, $P < 0.001$). If one excludes BP differences of ≤10% from consideration, then 16 of 20 clades show the same trend ($0.02 > P > 0.01$). The same trends were seen under ln-weighted parsimony and distance methods.

The BPs for each of the above 37 clades were also tabulated separately for the combined data set and each of the six bootstrap-augmented, single-gene data sets (data not shown). Results of these BP comparisons are summarized in Table 9. For DDC, the differences in BPs between the bootstrap-augmented data sets and the combined data set were nonsignificant. For EF-1α, the differences were significant, with the combined data set having higher BPs.

The number of concordance groups recovered by the bootstrap-augmented data sets was also noted (Table 5). None of the six bootstrap-augmented data sets recovered as many concordance groups as the combined data set did.

DISCUSSION

The source and effects of third-position base composition bias in EF-1α were identified by Mitchell et al. (1997) and are mentioned here only for completeness. For EF-1α the source of heterogeneity was the Arctiidae and Lymantriidae. These families shared a distinctive third-position base composition, but the similarity of LogDet ME trees and ML ME trees, among other things, suggested that this bias had little if any effect on phylogeny reconstruction. A similar conclusion was reached by Fang et al. (2000) for DDC third positions, although the source of heterogeneity in the 77-taxa data set is not as easily definable, including Notodontidae and at least some quadrifine noctuids, arctiids, and lymantriids.

Combining disparate phylogenetic data sets in a single analysis can be problematic if there is substantial disagreement among them (e.g., Bull et al., 1993). However, defining "substantial disagreement" is difficult. The ILD test (Farris et al., 1994) has been proposed as an objective test of the degree of disagreement among data partitions, but recent studies suggest that the

TABLE 8. Bootstrap percentages for clades recovered by different data sets, under equal and ln-weighting schemes for parsimony analysis, and using the complete-and-partial bootstrap program Njbootli (Zharkikh and Li, 1995) for distance analysis. Only clades showing substantially different bootstrap values (i.e., ≥10% difference) among data set partitions were tabulated.

		MP, equal weighting			MP, ln-weighting			Distance, Njbootli[b]		
Clade no.	Taxa included in clade[a]	EF-1α[c]	DDC[c]	Comb.	EF-1α	DDC	Comb.	EF-1α[c]	DDC[c]	Comb.
1	Ingroup (all except Notodontidae)	12 (24)	42 (64)	50	13	69	64	4 (12)	19 (70)	40
2	Notodontidae, excl. Gsep	<5	45 (<5)	53	9	75	77	–	44 (76)	50
3[d]	Gsep, Blev	34	12	<5	22	9	15	–	–	–
4	Arctiidae, Lymantriidae, some quadrifine noctuids[e]	53 (19)	31 (49)	84	68	57	92	23 (–)	– (–)	65
5	Arctiidae, Lymantriidae, Herminiinae, Aganainae	<5 (11)	35 (39)	52	<5	37	43	– (–)	– (18)	–
6	Arctiidae, Lymantriidae	38 (36)	24 (28)	52	56	31	71	36 (61)	36 (28)	82
7	Lymantriidae	54 (71)	75 (69)	85	72	88	97	64 (94)	– (81)	89
8[d]	Lymantriidae, Hmi	77 (58)	<5 (8)	43	74	<5	9	74 (76)	– (–)	54
9	Arctiidae	<5 (35)	91 (77)	52	<5	92	81	– (–)	66 (71)	–
10	Herminiinae, Aganainae	14	20	29	36	33	69	–	–	79
11	Herminiinae	50	95	93	59	95	92	72	95	89
12	Euteliinae, Stictopterinae	34	74	81	33	85	91	–	47	80
13	Catocalinae, Calpinae	70	7 (10)	58	75	<5	44	–	– (17)	–
14	Nolinae	15 (18)	16	27	19	47	65	– (–)	–	–
15	Eustrotiinae	6	56 (52)	72	<5	80	72	–	14 (63)	–
16	some Plusiinae (Afa, Cfor)	67	89	95	70	87	98	98	100	99
17	Stiriinae, Psaphidinae, Apyr	16 (9)	31 (30)	53	18	63	75	11 (10)	45 (64)	65
18	Psaphidinae, Apyr	13 (21)	68 (60)	72	16	80	89	– (–)	44 (49)	–
19	Psaphidinae	54 (64)	78 (68)	86	32	36	33	23 (66)	51 (–)	79
20	some Psaphidinae (Pres, Tsap)	25	89	74	44	90	97	40	89	71

TABLE 8. Continued

Clade no.	Taxa included in clade[a]	MP, equal weighting			MP, ln-weighting			Distance, Njbootli[d]		
		EF-1α[c]	DDC[c]	Comb.	EF-1α	DDC	Comb.	EF-1α[c]	DDC[c]	Comb.
21	Rabr, Phe	44 (43)	<5	<5	33	<5	<5	28 (–)	–	–
22	Oncocnemidinae, Agaristinae	8 (<5)	43	51	7	36	45	– (–)	–	30
23	Condicinae	7	42	69	8	35	58	13	38	–
24	some Condicinae (Cvi, Ldip)	34	91	93	76	95	97	29	71	86
25	some Heliothinae (Sar, Abe, Hal, Mte, Hvi, Hze)	35 (27)	84 (82)	90	51	84	91	69 (84)	62 (75)	96
26	some Heliothinae (Hvi, Hze)	29	67 (64)	77	67	94	96	72	48 (–)	76
27	Noctuinae s.l., excl. all "Caradrinini" but Atar	60 (76)	94 (93)	100	54	86	99	40 (73)	80 (88)	97
28	Noctuinae s.l., excl. all "Caradrinini"	8	50	57	<5	62	52	–	24	24
29	Hadeninae, Apameini, Xylenini, Nra	12 (60)	27 (16)	44	13	22	13	– (52)	17 (76)	34
30	Hadeninae, Apameini, Xylenini	14 (20)	21 (16)	43	15	12	22	– (–)	19 (53)	–
31	Apameini, Xylenini	90	47	97	95	18	96	81	79	95
32	Apameini	15	81	90	20	97	98	40	59	96
33	Xylenini	19 (95)	75 (95)	68	25	75	72	– (90)	77 (100)	90
34	Cmol, Elg	<5	47	33	<5	42	42	–	30	17
35[d]	Cmol, Homo	67	<5	<5	84	5	11	38	–	–
36[d]	Sex, Sor	76	20	28	74	32	22	61	94	58
37[d]	Aamp, Aral	57	<5	<5	48	<5	<5	–	–	–

[a]Species abbreviations are from Table 1.
[b]Dashes denote no value given (the "njbootli" software reports bootstrap values for those clades recovered in the neighbor-joining tree only).
[c]Bootstrap percentages in parentheses are values derived for comparable groups (if present) in the 49-taxa data sets.
[d]Clades that break up concordance groups (clades for which monophyly has been established by morphological evidence); it is expected that these groups receive little support.
[e]Quadrifine noctuid subfamilies included are Herminiinae, Aganainae, Catocalinae, Calpinae, and Hypeninae.

TABLE 9. Sign tests performed on bootstrap percentage (BP) comparisons between each of the six bootstrap-augmented data sets and the combined data set for the 37 clades listed in Table 8.

Data sets	+[a]	−[b]	0[c]	*n*	*P*[d]
EF-1α 1	34	3	0	37	<0.001
EF-1α 2	33	4	0	37	<0.001
EF-1α 3	32	5	0	37	<0.001
DDC 1	15	17	5	32	>0.5
DDC 2	14	20	3	34	>0.5
DDC 3	13	20	4	33	<0.5, >0.2

[a]Number of clades for which the combined data set had the larger BP.
[b]Number of clades for which the combined data set had the smaller BP.
[c]Number of clades showing no difference in BP.
[d]Two-tailed test.

ILD test is not entirely appropriate as a test for combinability of separate data partitions. For example, Cunningham (1997b) suggested that "a significance threshold of 0.05 may be too conservative for the ILD test," after finding that ILD test *P*-values had to be much less than 0.01 ($P \cong 0.001$) before combined analysis of data partitions led to reduced recovery of a known phylogeny. By this criterion, many of our ILD tests performed under ln-weighting are effectively nonsignificant ($0.05 < P < 0.01$), justifying our analysis of the combined data set. Nonetheless, much information may be gleaned from comparison of *P*-values among different tests, as discussed below.

Because there can be only one true species phylogeny, incongruence must indicate either that phylogeny reconstruction for at least one of the data partitions was inaccurate, attributable to poor data quality or inappropriate phylogenetic analysis, or that the partitions have different evolutionary histories, attributable to lineage sorting, for example. The latter explanation is more plausible when internodes are relatively short, as is seen for the deeper branches of the ML tree. Thus lineage sorting cannot be ruled out for this data set. In our analyses, however, much of the incongruence can be eliminated by excluding highly diverged taxa, which involve relatively deep nodes in the tree. This observation seems to favor error in the gene tree(s), possibly from long branch attraction (Felsenstein, 1978b), as the explanation for the incongruence. It is therefore surprising that ILD tests performed on the reduced 49-taxa data sets gave consistently higher *P*-values than equivalent tests performed on the complete 77-taxa data sets; that is, incongruence between data partitions *increased* with increased taxon sampling. Long branch and taxon sampling effects could reasonably be expected to decrease with the addition of taxa, making it easier to reconstruct the true phylogeny (e.g., Hillis, 1996). A possible reason for this apparent contradiction is that we could be exacerbating the long branch problem when adding more taxa, because we are inadvertently selecting highly divergent exemplar species from a poorly known higher-level phylogeny.

Robustness of Results to Variation in Phylogenetic Methods

Almost all nodes strongly supported by MP analysis were also strongly supported by ME analysis, and vice versa, with the following exceptions: Three groups with moderate support under MP (≥65%; Nolinae, Arctiidae + Lymantriidae, and *Spodoptera frugiperda* + *S. ornithogalli*) had much less support under ME (≤55%). In contrast, support for Psaphidinae, Condicinae, and Stiriinae + Amphipyrinae + Psaphidinae increased from 33%, 58%, and 75% under MP to 83%, 90%, and 96%, respectively, under ME. Relationships within Plusiinae and within *Spodoptera* were recovered "correctly," that is, in agreement with morphology, by the mp tree (Fig. 2) but not by the ML tree (Fig. 3). Perhaps not coincidentally, both groups are subtended by long internal branches. The placement of plusiines within noctuids also differed: The plusiines were sister group to the rest of the ingroup

in the mp tree but were sister group to the quadrifine noctuids, arctiids, and lymantriids in the ML tree. However, under the ML criterion, Plusiinae can be placed in either position, or as sister group to all other trifine noctuids, without significantly changing the likelihood score (Kishino–Hasegawa tests, $P > 0.5$).

More Taxa or More Characters?

Taxon sampling effects.—Under the MP criterion, higher support values were seen in the 49-taxa data set than in the 77-taxa data set, in comparisons made for EF-1α. For DDC, slightly more than half the clades compared had higher support with 77 taxa, but the trend was not statistically significant. Note that the extra taxa added in the 77-taxa data set join below the root node of the same group in the smaller data set for three clades (1, 4, and 27). This might be predicted to result in lower BPs for the 77-taxa data set because the subtending branch is split. However, for EF-1α clade 4 actually increased in BP with increased taxon sampling. None of the clades affected in this way in the DDC analysis showed a change ≥10%.

When a distance method with a correction for the number of taxa in each data set was used (complete-and-partial bootstrap; Zharkikh and Li, 1995), the 49-taxa data sets clearly had higher support values for both genes. All available comparisons for EF-1α supported this conclusion. For DDC, almost all clades compared showed the same trend, and the difference was statistically significant. The nonsignificance of the EF-1α result obtained by this method was obviously the result of the small number of comparisons available ($n = 5$).

Thus, in cases where increasing taxon sampling changed the support levels for a clade, the effect was a negative one. This result was surprising, in light of the large (57%) increase in taxon sampling, especially because the taxa had been selected in an attempt to break up long branches (Swofford et al., 1996). Furthermore, our results are contrary to the findings of Poe (1998) that when adding taxa decreased the accuracy of a phylogeny estimate, the decrease usually did not involve preexisting relationships.

The reason for the decrease in support levels with addition of taxa appears to be that the additional groups sampled were sampled too sparsely relative to the local saturation level for these genes. The effect is that of adding long branches to the tree rather than breaking up long branches (Graybeal, 1998; Hillis, 1998). Thus, "lineages that previously were not spuriously attracted to each other could become 'long' in a relative sense by virtue of the shortening of another branch on which the added taxa connect" (Poe, 1998).

An alternative explanation for the decrease in support levels, that the original 49 species samples were just fortuitous combinations of taxa, and that different combinations might yield lower support values, was discounted by randomly deleting 28 species from subfamilies for which more than a single exemplar was sampled in the 77 species data set (this was repeated five times). Bootstrap analyses of these new 49-taxa data sets produced support values very similar to those of the original 49-taxa data sets.

We have thus provided empirical data in support of simulation studies (e.g., Poe and Swofford, 1999) that suggest when trying to solve a difficult phylogenetic problem, adding taxa is not a panacea because it can create additional problems under certain circumstances.

Addition of characters.—The combined data sets gave both higher overall levels of support, measured as percentage of nodes with strong support, and higher support levels for the specific groups recovered in the trees, than did any of the single-gene data sets alone. That trend was apparent under all phylogenetic methods used. Those results are no surprise given that consistent methods of phylogeny estimation, by definition, will converge on the true tree with increasing numbers of characters sampled. However, the absolute number of characters sampled is not the only consideration. Our data suggest that equal attention should be paid to potential nonindependence among characters.

Combining data from independent genes.—One caveat to the observation that consistent methods of phylogeny estimation will converge on the true tree as increasing numbers of characters are sampled is that

none of the assumptions of the method can be violated. However, violations of assumptions are not necessarily easily detected (Swofford et al., 1996). This fact alone may be grounds for preferring that additional characters be obtained from an independent source, all else being equal.

Other reasons also support the idea that obtaining characters from independent genes would hold greater benefit for phylogeny inference than obtaining the same total number of characters from a single gene. First, the endemic biases of any one gene (in base composition, evolutionary rate, and so forth) might be diluted in the combined data set. Cummings et al. (1995), for example, showed that data sets consisting of blocks of contiguous sites drawn from complete mitochondrial genomes were less likely to recover the genome phylogeny than were data sets comprising an equal number of single sites drawn randomly from the entire genome. They proposed that the effects of "location-dependent processes in sequence evolution" could be reduced, and thus the power of phylogenetic analysis improved, by sampling "sites from throughout the genome or from other genomes in the organisms." Second, on the positive side, each gene might carry a signal for groupings on which the other genes are silent. These effects could result in decreased support for spurious groupings, in addition to increased support levels for correct groupings.

Our comparisons of the combined-gene data set and the single-gene data sets of the same size obtained by bootstrap augmentation provide some support for the above postulates. For EF-1α, support levels for the bootstrap-augmented data set were consistently lower than for the combined data set. For DDC, there was little apparent difference in support levels between the bootstrap-augmented data set and the combined-gene data set, although the combined data set recovered more concordance groups. One might argue that as long as one used an especially informative gene, it would make little difference whether one sampled more of the same gene or an additional gene. However, the difficulty is knowing in advance that a given gene is especially informative for the group of interest. Besides, this misses the point that the resolution and support levels obtained by using the combined-gene data set were no less than those obtained from the bootstrap-augmented DDC data set, despite the fact that EF-1α proved to offer less information per nucleotide. Thus, while recognizing that the dynamics of this single case may not be generalizable, we argue that the prudent approach to collecting additional sequence data is to prefer that such data comes from additional genes.

Systematics of the Noctuoidea

The tentative hypothesis that Noctuidae is paraphyletic with respect to Arctiidae and Lymantriidae (Weller et al., 1994; Mitchell et al., 1997; Fang et al., 2000) is strongly supported by our new data set. There is 96% bootstrap support under MP, and 90% under ME, for a clade comprising Arctiidae, Lymantriidae, and the quadrifine noctuid subfamilies Aganainae, Herminiinae, Catocalinae, Calpinae, and Hypeninae, to the exclusion of other quadrifine and all trifine Noctuidae. It is thus likely that Noctuidae is not a natural group. However, in the interests of taxonomic stability, we decline to revise noctuoid classification to reflect the paraphyly of Noctuidae until the deeper divergences within the family are resolved with reasonable confidence.

Supporting a growing consensus among noctuid systematists (Beck, 1960, 1992; Holloway, 1989; Lafontaine and Poole, 1991; Lafontaine, 1993; Poole, 1995; Speidel et al., 1996; Kitching and Rawlins, 1999), our data also provide the strongest evidence to date (99% bootstrap value under MP and 100% under ME) for the monophyly of the "true cutworms" (Noctuinae s.l.; Poole, 1995). As circumscribed in our trees, Noctuinae s.l. includes Noctuini s.l., Hadenini, Apameini, Xylenini, and some "Caradrinini," (*Nedra* and *Anorthodes*). "Caradrinini" is a heterogeneous assemblage of genera, most of whose placements are still problematic; accordingly, further work is needed to clarify the limits of Noctuinae s.l.

Several other higher-level relationships within trifine noctuids are supported by our data. Support for the placement of *Amphipyra* with the Psaphidinae, and the Stiriinae as sister group to this clade, is relatively strong. Moreover, the broader trifine

relationships in Figure 2 and Figure 3 correspond well to recent morphological hypotheses. For example, groups previously thought to be relatively primitive, such as Acontiinae and Eustrotiinae (Poole, 1995), do indeed take basal positions, whereas subfamilies generally regarded as derived, such as Agaristinae, Cuculliinae s.s., Oncocnemidinae, and Amphipyrinae s.s., are allied with the highly derived Heliothinae and Noctuinae s.l. The placement of other groups, such as Condicinae and Plusiinae, remains difficult. Plusiinae are thought to be allied with the trifine clade (Poole, 1995; Speidel et al., 1996; Kitching and Rawlins, 1999), but our data set lacks the resolving power to distinguish among various basal placements for this group, plusiines being basal within trifines or within quadrifines, or even being basal to all other noctuids. Thus the deeper nodes in Noctuidae remain weakly supported, and clearly, much more evidence will be needed to fully sort out the higher-level relationships of this large paraphyletic group.

Conclusions

By combining data from independent nuclear genes, we obtained the first strong evidence that Noctuidae are paraphyletic with respect to Arctiidae and Lymantriidae, and we have increased confidence in many higher-level relationships within the family.

Concordance among multiple independent data sets is probably the most powerful evidence systematists can provide for phylogenetic relationships (Miyamoto and Fitch, 1995) and this may be reason enough to turn to an independent source when more data are needed to answer a difficult phylogenetic problem. We have provided some evidence for the importance of character-set independence. Our study also suggests that the gene-specific biases endemic to DNA sequence data will be diluted in a combined data set, thus reducing or even eliminating support for erroneous relationships.

The increased support levels in the combined data set provided a strong a posteriori argument for combining data in a single analysis. While we agree in principle with the desirability of a priori congruence testing, our results support proposals by other authors that the ILD test is too conservative for the purpose of deciding when to combine data sets if one applies the conventional significance level of $P = 0.05$, or even $P = 0.01$. An additional dilemma is that with large data sets, the time needed to perform a rigorous ILD test may be prohibitive. Quicker methods for testing the significance of incongruence would be welcomed.

Increased taxon sampling is often recommended as a fix for problematic molecular phylogenetic data sets, especially if taxa are chosen specifically to break up long branches. However, we present an empirical example of deleterious effects: Support levels for most clades decreased when we introduced previously unsampled, divergent groups to the data set. We suggest that this might be a more general phenomenon: If "new" taxa are not sampled densely enough, one can inadvertently increase apparent long branch effects, thereby reducing support for clades on the tree, even when trying to do the opposite! This situation is most likely to occur in taxa for which the phylogenetic relationships are poorly known. At present we are attempting to determine whether a much larger increase in the taxon sample will permit more confident resolution of deep nodes on the noctuid tree.

ACKNOWLEDGMENTS

This paper was greatly strengthened by the thoughtful suggestions made by Dick Olmstead, Chris Simon, Tim Collins, and Cliff Cunningham. We thank Jeff Shultz for providing useful comments on a draft of the manuscript, Bob Poole and Eric Metzler for help with species identifications, Suwei Zhao and Kongyi Jiang for technical assistance, Cliff Cunningham for supplying a spreadsheet for calculation of six-parameter parsimony step matrices, Dave Swofford for use of a test version of PAUP*4.0, Andrey Zharkikh for help with his complete-and-partial bootstrap software, and Bob Poole, Don Lafontaine, Mike Pogue, Tim Friedlander, Quentin Fang, and Soowon Cho for helpful discussions. We are deeply indebted to James Adams, Bob Denno, Ian Kitching, Ed Knudsen, Marcus Matthews, Ric Piegler, Bob Poole, Ron Robertson, and Andy Warren for providing fresh specimens for this study. Financial support from the USDA-NRICGP, USDA-ARS Systematic Entomology Laboratory, and the University of Maryland Graduate School is gratefully acknowledged.

REFERENCES

BECK, H. 1960. Die Larvalsystematik der Eulen (Noctuidae). Akademie Verlag, Berlin.

BECK, H. 1992. New view of the higher classification of the Noctuidae (Lepidoptera). Nota Lepid. 15:3–28.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

CHO, S., A. MITCHELL, J. C. REGIER, C. MITTER, R. W. POOLE, T. P. FRIEDLANDER, AND S. ZHAO. 1995. A highly conserved nuclear gene for low-level phylogenetics: Elongation factor-1αrecovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12:650–656.

CUMMINGS, M. P., S. P. OTTO, AND J. WAKELY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12:814–822.

CUNNINGHAM, C. W. 1997a. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. 46:464–478.

CUNNINGHAM, C. W. 1997b. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. 14:733–740.

FANG, Q. Q., S. CHO, J. C. REGIER, C. MITTER, M. MATTHEWS, R. W. POOLE, T. P. FRIEDLANDER, AND S. ZHAO. 1997. A new nuclear gene for insect phylogenetics: Dopa decarboxylase is informative of relationships within the Heliothinae (Lepidoptera: Noctuidae). Syst. Biol. 46:269–283.

FANG, Q. Q., A. MITCHELL, J. C. REGIER, C. MITTER, T. P. FRIEDLANDER, AND R. W. POOLE. 2000. Phylogenetic utility of the nuclear gene dopa decarboxylase (DDC) in noctuoid moths (Insecta: Lepidoptera: Noctuoidea). Mol. Phylogenet. Evol. (in press).

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, AND C. BULT. 1994. Testing the significance of incongruence. Cladistics 10:315–319.

FELSENSTEIN, J. 1978a. The number of evolutionary trees. Syst. Zool. 27:27–33.

FELSENSTEIN, J. 1978b. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.

FELSENSTEIN, J. 1995. PHYLIP (phylogeny inference package), version 3.572. Department of Genetics, University of Washington, Seattle.

FRATI, F., C. SIMON, J. SULLIVAN, AND D. L. SWOFFORD. 1997. Evolution of the mitochondrial cytochrome oxidase II gene in Collembola. J. Mol. Evol. 44:145–158.

FRIEDLANDER, T. P., J. C. REGIER, AND C. MITTER. 1992. Nuclear gene sequences for higher-level phylogenetic analysis: 14 promising candidates. Syst. Biol. 41:483–489.

FRIEDLANDER, T. P., J. C. REGIER, C. MITTER, AND D. L. WAGNER. 1996. A nuclear gene for higher-level phylogenetics: Phosophoenolpyruvate carboxykinase tracks Mesozoic-age divergences within Lepidoptera (Insecta). Mol. Biol. Evol. 13:594–604.

GALLOWAY, G. L., R. L. MALMBERG, AND R. A. PRICE. 1998. Phylogenetic utility of the nuclear gene arginine decarboxylase: An example from the Brassicaceae. Mol. Biol. Evol. 15:1312–1320.

GRAYBEAL, A. 1994. Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. Syst. Biol. 43:174–193.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47:9–17.

GUPTA, R. S. 1995. Phylogenetic utility of the 90kD heat shock family of protein sequences and an examination of the relationship among animals, plants and fungi species. Mol. Biol. Evol. 12:1063–1073.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 21:160–174.

HILLIS, D. M. 1996. Inferring complex phylogenies. Nature 383:130–131.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. Syst. Biol. 47:3–8.

HILLIS, D. M., AND J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42: 182–192.

HOLLOWAY, J. D. 1989. The moths of Borneo. Part 12. Noctuidae: Noctuinae, Heliothinae, Hadeninae, Acronictinae, Amphipyrinae, Agaristinae. Southdene, Kuala Lampur.

HUELSENBECK, J. P., J. J. BULL., AND C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11:152–158.

INTERNATIONAL UNION OF BIOCHEMISTRY. 1986. Nomenclature for incompletely specified bases in nucleic acid sequences. Mol. Biol. Evol. 3:99–108.

KIM, J. 1996. General inconsistency conditions for maximum parsimony: Effects of branch lengths and increasing numbers of taxa. Syst. Biol. 45:363–374.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29:170–179.

KITCHING, I. J. 1984. An historical review of the higher classification of the Noctuidae (Lepidoptera). Bull. Br. Mus. Nat. Hist. (Entomol.) 49:153–234.

KITCHING, I. J., AND J. E. RAWLINS. 1999. Noctuoidea. Pages 355–401 *in* Handbook of zoology. Lepidoptera, volume 1: Systematics and evolution (N. P. Kristensen, ed.). W. de Gruyter, Berlin.

LAFONTAINE, J. D. 1993. Cutworm systematics: Confusions and solutions. Mem. Entomol. Soc. Can. 165:189–196.

LAFONTAINE, J. D., AND R. W. POOLE. 1991. Noctuoidea, Noctuidae (Part), Plusiinae. Moths of America north of Mexico, Fascicle 25.1. The Wedge Entomological Research Foundation, Washington, DC.

LANAVE, C., J. PREPARATA, C. SACCONE, AND G. SERIO. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20:86–93.

LI, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. 36:96–99.

LOCKART, P. J., M. A. STEEL, M. D. HENDY, AND D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.

MILLER, J. S. 1991. Cladistics and classification of the Notodontidae (Lepidoptera: Noctuoidea) based on

larval and adult morphology. Bull. Am. Mus. Nat. Hist. 204:1–230.

MITCHELL, A., S. CHO, J. C. REGIER, C. MITTER, R. W. POOLE, AND M. MATTHEWS. 1997. Phylogenetic utility of elongation factor-1α in Noctuoidea (Insecta: Lepidoptera): The limits of synonymous substitution. Mol. Biol. Evol. 14:381–390.

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

ORTÍ, G., AND A. MEYER. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleost fishes. Mol. Biol. Evol. 13:556–574.

POE, S. 1998. The effect of taxonomic sampling on accuracy of phylogeny estimation: Test case of a known phylogeny. Mol. Biol. Evol. 15:1086–1090.

POE, S., AND D. L. SWOFFORD. 1999. Taxon sampling revisited. Nature 398:299–300.

POOLE, R. W. 1995. Noctuoidea. Noctuidae (part). Cuculliinae, Stiriinae, Psaphidinae (part). Moths of America north of Mexico, Fascicle 26.1. The Wedge Entomological Research Foundation. Washington, DC.

REGIER, J. C., AND J. W. SHULTZ. 1997. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. 14:902–913.

RODRIGUEZ, F., J. L. OLIVER, A. MARIN, AND J. R. MEDINA. 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142:485–501.

RZHETSKY, A., AND M. NEI. 1992. A simple method for estimating and testing minimum evolution trees. Mol. Biol. Evol. 9:945–967.

SCOBLE, M. J. 1992. The Lepidoptera. Form, function and diversity. Oxford Univ. Press, Oxford, U.K.

SLADE, R. W., C. MORITZ, AND A. HEIDEMAN. 1994. Multiple nuclear-gene phylogenies: Application to pinnipeds and comparison with a mitochondrial DNA gene phylogeny. Mol. Biol. Evol. 11:341–356.

SMITH, S. W., R. OVERBEEK, C. R. WOESE, W. GILBERT, AND P. M. GILLEVET. 1994. The genetic data environment and expandable GUI for multiple sequence analysis. Comput. Appl. Biosci. 10:671–675.

SPEIDEL, W., H. FANGER, AND C. M. NAUMANN. 1996. The phylogeny of the Noctuidae (Lepidoptera). Syst. Entomol. 21:219–251.

STADEN, R. 1992. STADEN package. MRC Laboratory of Molecular Biology, Cambridge, U.K.

SULLIVAN, J. 1996. Combining data with different distributions of among-site variation. Syst. Biol. 45:375–380.

SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 *in* Phylogenetic analysis of DNA sequences (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, New York.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

WATERS, E. R. 1995. An evaluation of the usefulness of the small heat shock genes for phylogenetic analysis in plants. Ann. Mo. Bot. Gard. 82:278–295.

WELLER, S. J., D. P. PASHLEY, J. A. MARTIN, AND J. L. CONSTABLE. 1994. Phylogeny of noctuoid moths and the utility of combining independent nuclear and mitochondrial genes. Syst. Biol. 43:194–211.

ZAR, J. H. 1984. Biostatistical analysis, 2nd edition. Prentice Hall, Englewood Cliffs, New Jersey.

ZHARKIKH, A., AND W.-H. LI. 1995. Estimation of confidence in phylogeny: The complete-and-partial bootstrap technique. Mol. Phylogenet. Evol. 4:44–63.

Received 10 November 1998; accepted 6 April 1999
Associate Editor: C. Simon