Phylogenetic Utility of *Tektin*, a Novel Region for Inferring Systematic Relationships Among Lepidoptera

Alaine Whinnett,[1,2] Andrew V. Z. Brower,[3] Ming-Min Lee,[3] Keith R. Willmott,[2] and James Mallet[1]

ABSTRACT   Rapidly evolving nuclear coding sequences are highly desirable for phylogenetic studies of closely related species. Here, we investigated an 807-bp region, homologous to the testis-specific *Tektin* gene from *Bombyx mori* (L.), in 34 nymphalid butterfly taxa in the subfamilies Ithomiinae, Danainae, and Heliconiinae. Within Ithomiinae, relationships inferred from *Tektin* sequence data were remarkably similar to those in trees based on combined morphological and ecological data. Partitioned Bremer analysis, with mitochondrial *cytochrome oxidase I* and *II*, and nuclear *wingless* and *elongation factor 1-α* sequences, revealed *Tektin* to have the greatest utility for inferring relationships at the genus, tribe, and subfamily levels across the studied taxa. We think *Tektin* will provide a useful source of molecular characters for inference of relationships among other butterflies, and perhaps among other insect taxa.

KEY WORDS   Ithomiinae, Lepidoptera, molecular phylogenetics, *Tektin*

Since the late 1980s, molecular sequence data have played a growing role in phylogenetic analyses. In insects, relatively few loci have contributed to this effort, among them 18S and 28S rDNA, *elongation factor 1-α* (*Ef1α*), *wingless* (*wg*), and the mitochondrial genes *cytochrome oxidase I* and *II* (COI and COII), 12S, and 16S. Because gene trees may deviate from the organismal topology due to ancestral polymorphism (Takahata and Nei 1985, Neigel and Avise 1986, Nei 1987, Edwards and Beerli 2000), horizontal gene transfer (Nei 1987, Kidwell 1993, Cummings 1994, Syvanen 1994, Philippe and Douady 2003), and introgressive hybridization (Nei 1987, Doyle 1992, Machado et al. 2002, Mallet 2005), as well as stochastic processes, the sequencing of multiple, unlinked loci is considered highly desirable for obtaining robust phylogenetic inferences that approximate the organismal phylogeny (Pamilo and Nei 1988, Brower et al. 1996, Edwards and Beerli 2000, Rokas et al. 2003). Yet, despite general recognition of these principles, many phylogenetic hypotheses are still based on a single (often mitochondrial DNA), or a few separate and putatively independent gene regions.

Mitochondrial sequences are widely used in studies of closely related species, or even subspecies or populations within species, largely because they have a more rapid substitution rate, lower or absent recombination, and smaller effective population size in comparison with nuclear regions (Harrison 1989). However, evidence is mounting that phylogenetic inferences from mitochondrial sequences alone may be confounded by selection (Hudson and Turelli 2003), nuclear copies (Bensasson et al. 2001), and cytoplasmic factors, such as the intracellular bacterium *Wolbachia* (Hurst and Jiggins 2005). Therefore, there exists an urgent need to develop additional nuclear gene sequences that evolve rapidly enough to be used at subspecies, species, and higher levels.

We have been investigating phylogenetic relationships among a group of nymphalid butterflies (Ithomiinae) that have long been the subject of research in evolution and biogeography. We sought to develop primers for novel phylogenetically informative nuclear gene regions in these Lepidoptera, amenable to direct sequencing, but that showed rapid evolutionary rates. We deliberately targeted a ≈1-kb intronless, easily alignable coding sequence, which could be directly sequenced from genomic DNA using external primers. We repeatedly identified sequences from Ithomiinae inferred to be homologous to a testis-specific *Tektin* from the silkworm, *Bombyx mori* (L.) (GenBank accession AB056651) (Ota et al. 2002), while attempting to clone a fragment of *triose-phosphate isomerase* (Beltrán et al. 2002). Initial comparisons with GenBank sequences suggested the potential phylogenetic utility of *Tektin*, so we further investigated the region in 34 species of the subfamilies Ithomiinae, Danainae, and Heliconiinae (Nymphalidae), sampled at a range of taxonomic ranks (Table 1). We refer here to Ithomiinae as a subfamily distinct from Danainae, following the most recent checklist for this group by Lamas (2004), although the two are

[1] Department of Biology, 4 Stephenson Way, University College London, London, NW1 2HE United Kingdom.
[2] The Natural History Museum, Cromwell Rd., London, SW7 5BD United Kingdom.
[3] Department of Zoology, Oregon State University, Corvallis, OR 97331.

Table 1. Specimen information and Genbank Accession numbers

Taxon	Voucher no.	Collection locality	Tektin	wg	Eflα	Mitochondrial
Bombycidae						
B. mori			AB056651	D14169	D13338	AY048187
Nymphalidae: Ithomiinae						
Athyrtis mechanitis salvini (Smka 1884)	RB359	BRAZIL: Rondônia, Cacaulândia	AY848713	DQ071873	—	DQ069230
Callithomia lenea epidero (Bates 1862)	02-3207	PERU: Cuzco, Palma Real	—	—	DQ085444	—
Callithomia lenea zelie (Guérin-Méneville 1844)	RB380	BRAZIL: Rondônia, Cacaulândia	AY848714	—	—	DQ069232
Dircenna dero (Hübner 1823) ssp	PE18-2	PERU: Madre de Dios, Tambopata Preserve	AY848715	DQ071874	DQ073024	DQ069233
Elzunia pavonii (Butler 1873)	E44-3	ECUADOR: Sucumbíos, El Recodo	AY848716	DQ073010	DQ073025	—
	E28-2	ECUADOR: Loja, San Pedro de Vilcabamba	AY848717	—	DQ073026	—
	E12-1	ECUADOR: Loja, Macará	AY848718	AF246562	DQ073027	—
Godyris zavaleta matronalis (Weymer 1883)	E44-1	ECUADOR: Sucumbíos, El Recodo	AY848719	DQ073011	DQ073028	DQ069234
Greta hermana joiceyi (Kaye 1918)	E39-46	ECUADOR: Sucumbíos, La Bonita	AY848720	DQ073012	DQ073029	DQ069235
Hyalyris antea amarilla (Vitale & Bollino 2000)	E30-4	ECUADOR: Zamora-Chinchipe, Quebrada Chorillos	AY848721	DQ073013	DQ085445	DQ069236
Hyposcada anchiala c. interrupta (Tessmann 1928)	02-2105	PERU: San Martín, km 7.2 Pongo-Barranquita	AY848711/AY848712	DQ085433	DQ085446	DQ069237
Hyposcada anchiala fallax (Staudinger 1884)	02-3519	PERU: Cuzco, Mazuko	AY848722	DQ085434	DQ085447	DQ078356
Hyposcada anchiala mendax (Fox 1941)	02-1645	PERU: San Martín, Puente Serranayacu	AY848723	—	DQ085448	DQ078477
Hyposcada zarepha flexibilis (Haensch 1909)	02-1644	PERU: San Martín, Puente Serranayacu	AY848724	DQ085435	DQ073030	DQ078357
Ithomia drymo (Hübner 1816)	02-198	PERU: San Martín, Chumía	AY848725	DQ085436	DQ073031	DQ078363
Methona sp	B16-5	BRAZIL: São Paulo, São Paulo	AY848726	DQ073014	—	DQ069238
Melinaea menophilus (Hewitson 1856)	RB296	BRAZIL: Rondônia, Cacaulândia	—	DQ073015	—	DQ069239
	RB239	BRAZIL: Rondônia, Cacaulândia		—	DQ085449	AF014156/DQ069240
	RB288	BRAZIL: Rondônia, Cacaulândia		DQ085437	—	—
	02-1541	PERU: San Martín, Shapaja		—	—	DQ069241
	02-613	PERU: San Martín		DQ073016	DQ073033	DQ085456
Napeogenes larilla (Hewitson 1877) ssp	E39-47	ECUADOR: Sucumbíos, La Bonita	AY848727	DQ085438	DQ085450	DQ085457
Oleria assimilis (Haensch 1903) ssp nov	02-3609	PERU: Cuzco, Quincemil	AY848728	DQ085439	DQ085451	DQ085458
Oleria gunilla lota (Hewitson 1872)	Ec467	ECUADOR: Napo, Jatun Sacha	AY848729	DQ085440	DQ085452	DQ078390
Oleria onega janarilla (Hewitson 1863)	Ec277	ECUADOR: Pastaza, Communidad Shuar Mirador, 70 km E of Macas-Puyo	AY848730	DQ085441	DQ085453	DQ085459
Oleria onega (Hewitson 1852) ssp nov	02-515	PERU: San Martín, km 7.2 Pongo-Barranquita	AY848731	DQ085442	DQ085454	DQ085460
Oleria rubescens (Butler & Druce 1872)	02-835	PERU: San Martín, Puente Serranayacu	AY848732	DQ085443	DQ085455	DQ157528
Pagyris cymothoe sylvella (Hewitson 1868)	8369	PANAMA: Chiriquí, Quebrada Hornito	AY848733	—	DQ073035	DQ073038
Patitia neglecta (Lamas 1979)	E16-2	ECUADOR: Pichincha, San Antonio	AY848734	DQ073017	DQ073034	DQ069242
Pteronymia veia linzera (Herrich-Schäffer 1864)	02-1244	PERU: San Martín, km 8 Tarapoto-Yurimaguas	AY848735	DQ073018	DQ073036	—
Tithorea harmonia c. martina (Fox 1956)	E43-16	ECUADOR: Sucumbíos, km 19 La Bonita-Tulcán	AY848736	—	DQ073037	—
	PE19-19	PERU: Huánuco, Tingo María	AY848737	—	—	—
	PE12-3	PERU: Cuzco, Rosalina	—	—	—	—
Tithorea harmonia furia (Staudinger 1884)	V20	VENEZUELA: Monagas, nr Barrancas	—	AF246561	—	DQ157546
Velamysta pupilla greeneyi (Vitale & Bollino 2003)	E43-3	ECUADOR: Sucumbíos, km 19 La Bonita-Tulcán	AY848738	DQ073020	DQ073021	DQ071864
Nymphalidae: Danainae						
Danaus gilippus (Cramer 1776)	AZ1-3	USA: Arizona, Portal and vicinity	AY848739	DQ175476	DQ071871	—
Danaus plexippus (L. 1758)	C3-8	COLOMBIA: Meta, Villavicencio, Carretera El Amor	AY848740	—	—	AY569150
Parantica melusine (Grose-Smith 1894)	PNG2-3	PAPUA NEW GUINEA: Miamafu	AY848741	—	DQ071870	DQ175477
Anetia briarea numidia (Hübner 1819–1826)	CU1	CUBA: Santiago de Cuba	—	AF246579	DQ071869	DQ071866
Amauris tartarea (Mabille 1876)	GH-058	GHANA: Bobiri Forest Preserve	AY848742	DQ071872	DQ071868	DQ071867

Table 1. Continued

Taxon	Voucher no.	Collection locality	Tektin	wg	Ef1α	Mitochondrial
Nymphalidae: Heliconiinae						
Philaethria dido (L. 1763)	P7-4	PANAMA: Colón, Gamboa	AY848743	—	—	U08554
	RB283	BRAZIL: Rondônia, Cacaulândia	—	AF014137	—	—
	690	PANAMA: Gamboa, El Renacer	—	—	AY747979	—
Heliconius sara (F. 1793) ssp	G42-2	FRENCH GUIANA: Laurent du Maroni, St. Laurent du Maroni	AY848744	—	—	—
	P1-7	PANAMA: Colón, Gamboa	—	AF014130	—	—
	308	FRENCH GUIANA: St Maurice on road to Apatou	—	—	AY747924	AF413695
Heliconius erato petiverana (Doubleday 1847)	STRI-B-850	PANAMA: Canal Zone, Pipeline Road	AY848745	—	—	—
	P29-2	PANAMA: Panamá, El Llano	—	AF014127	—	—
	CR13	COSTA RICA: Puntarenas, Sirena	—	—	AY748017	AF413684
	2981	PANAMA: Canal Zone, Pipeline Road	—	—	—	—
	STRI-B-2980	PANAMA: Canal Zone, Pipeline Road	AF848746	—	—	—
Heliconius erato (L. 1758) ssp	P32-7	PANAMA: Darién, Cañazas	—	AF014126	—	—
Heliconius erato hydara (Hewitson 1867)	G16-4	FRENCH GUIANA: Cayenne, Piste Coralie	—	—	AY747987	—
	440	FRENCH GUIANA: Sablance (La Victoire) Route N1 km 19	—	—	—	AF413687
	STRI-B-442	FRENCH GUIANA: Sablance (La Victoire) Route N1 km 20	—	—	—	—

clearly sister taxa and have been considered a single subfamily (Ackery et al. 1999, Brower 2000). Taxa for this study were selected from among those for which comparative data had already been obtained for several other frequently used Lepidoptera phylogenetic markers: the mitochondrial loci COI and COII, and the nuclear loci *wg* and *Ef1α*.

Tektins are a family of largely ($\approx$70% of their sequence) α-helical proteins, which form filaments 2 to 3 nm in diameter (Linck et al. 1985). The filaments are composed of a core heterodimer of Tektins A and B (Pirner and Linck 1994), with Tektin C homodimers either separate, or on the periphery of the Tektin A and B core (Norrander et al. 1996). The filaments are the primary constituent of at least one of 13 protofilaments in the A microtubule wall of doublet microtubules (Linck et al. 1985, Nojima et al. 1995) and are located near to the junction where the A and B microtubules bind (Linck 1990, Nojima et al. 1995). Tektins were first characterized in the purple sea urchin, *Strongylocentrotus purpuratus* (Linck 1982, Linck et al. 1982, Linck and Langevin 1982), which now boasts a well studied Tektin repertoire, having sequence data reported for *Tektin* A1, B1, and C1 (Norrander et al. 1992, 1996; Chen et al. 1993). *Tektin* GenBank entries are currently available for a number of species: sea urchin, human, mouse, rat, dog, zebra fish, sea squirt, green alga, nematode, fruit fly, and silkworm.

B. mori Tektin (GenBank AB056651) had 30% identity to the deduced amino acid (aa) sequence of purple sea urchin Tektin A1 and 28% identities to B1 and C1 over a 338-aa section (aa 159–496) (Ota et al. 2002). Ota et al. (2002) further reported a 38% aa identity of this 338-aa section of the silkworm Tektin to a then hypothetical protein sequence from the fruit fly *Drosophila melanogaster* (Meigen) (AAF44971). This fruit fly sequence remains the closest, non-Lepidoptera sequence to the silkworm Tektin sequence known to date. It is identical to gene product of the more recently submitted *Tektin-A* (NM_078853), suggesting that the only silkworm sequence reported to date is orthologous to *Tektin-A*. The 338-aa silkworm region also has 24% aa identity to the fruit fly Tektin-C (NM_079216) and 26% aa identity to a cloned fruit fly gene product (BT010079). The fruit fly Tektin-A, Tektin-C, and cloned product (BT010079) have $\approx$25–30% pairwise polypeptide identity over the region corresponding to silkworm aa 159–496, suggesting that the cloned product (BT010079) also belongs to the *Tektin* gene family and may represent the as yet unassigned third *D. melanogaster Tektin* gene family member (Tektin-B).

Our major goal was to develop region-specific primers for nuclear genes to extend the currently limited repertoire of gene regions useful for Lepidoptera phylogenetics. Here, we investigate the pattern of molecular evolution of a *Tektin* gene and investigate its utility for inferring relationships among Lepidoptera at different taxonomic levels, compared with expectations based on traditional classification and empirical data from other gene regions.

Table 2. Primers used to amplify *Tektin*

Primer	Sense or antisense	5′to 3′ Sequence	Nucleotide position
Tek2	AS	TGTCRCTCCAATCRWATTC	930–948
Tek3	S	CAAGACGTACAAGCTAGCAAGA	198–219
Tek4	S	ACTGGAGAATGGCGAAAGAAC	459–479
Tek5	AS	GGCCWMRTCKCKGCAGTT	1401–1418
TektinA	S	ACCAGTGGRGAYATYCTWGG	330–349
TektinB	S	CAGGMCAYATMGAYTGGA	376–393
Tektin3	AS	CGCAGTTTYTGATRCTYT	1084–1101
Tektin4	AS	TCATRTCTTGASWGCCTTTG	1157–1176

Positions are given relative to silkworm *Tektin* (AB056651).

Materials and Methods

DNA Extraction. DNA was extracted from one third of a thorax (*Hyposcada* and *Oleria*) or two meso- and metathoracic legs (*Paititia*) by using the DNAeasy kit (QIAGEN, Valencia, CA), according to the manufacturer's instructions, with an initial 3-h incubation at 55°C and a final elution volume of 300 μl. Dried wings were retained as vouchers at University College London. All other samples had previously been isolated from thorax, thorax and head, or thorax, head, and abdomen tissue by using the SDS phenol/chloroform method (Brower 1994). Dried wings, antennae, legs, and available abdomens were retained as vouchers by A.V.Z.B. at Oregon State University.

Primer Development, Polymerase Chain Reaction (PCR), and Sequencing. An individual of *Hyposcada anchiala mendax* (02–1645) (Ithomiinae) was used for initial primer development. Two pairs of primers (Table 2) were designed to amplify overlapping regions. Primers Tek2 and Tek5 were designed from the alignment of silkworm (accession code AB056651) and fruit fly (accession code NM_078853) *Tektin*. Primers Tek3 and Tek4 were designed from a 413-bp *Hyposcada* fragment (accession code AY848712) with high resemblance [99 of 121 identities (81%)] to silkworm *Tektin*. The 413-bp putative *Tektin* fragment had been repeatedly amplified in *Hyposcada* samples when attempting to amplify *triose-phosphate isomerase* by using TPI-1 and TPI-2 primers (Beltrán et al. 2002).

A PCR was performed in a 50-μl volume, by using 4 μl of *H. anchiala mendax* (02–1645) template DNA and the following conditions: 1× PCR buffer (0.1 M Tris-HCl, 0.5 M KCl, 0.01 volume of Triton X-100), 0.2 μM dNTPs, 5 μM MgCl₂, 0.2 μM each primer, 0.3 U/μl *Taq*, and an amplification profile of 94°C for 2 min, followed by 32 cycles of (94°C for 60 s, 50°C for 60 s, and 72°C for 90 s) and a final 10-min extension at 72°C. PCR products were excised from an agarose gel and purified using the QIAquick gel extraction kit (QIAGEN), according to the manufacturer's protocol and sent to a commercial facility for cycle sequencing, precipitation and sequencing. This resulted in a 1142-bp edited product (after data loss at three prime end) (accession code AY848711), with a single G→C difference from the original 413-bp amplification. Final pairwise primers; TektinA, TektinB, Tektin3, and Tektin4, were designed from an alignment of the new 1142 bp *Hyposcada* sequence and the silkworm *Tektin* sequence (Table 2). Using these four primers, the *Tektin* region was amplified and sequenced in the other taxa as described above, with the following modifications: 2.5 μM MgCl₂, 0.5 μM of forward and reverse primer, a 35 cycle program, and a 55°C annealing temperature.

Taxonomic Sampling. For the phylogenetic study, *Tektin* sequences were analyzed from 26 Ithomiinae species, spanning a range of taxonomic levels and representing all major Ithomiinae tribes. Four Danainae and four Heliconiinae species also were selected (Table 1). These sequences have been submitted to GenBank (Accession codes AY848711–AY848746).

Outgroup sequences were available from GenBank. In addition, all the *Ef1α, wg,* and mitochondrial comparative data were available from previous publications and/or other research projects; individuals 02–3207 and 02–613 for *Ef1α,* and 02–1541 for *wg* (M. Zimmermann, unpublished data); 440, 690, 308, and 2981 for *Ef1α* (Beltrán et al., personal communication); individuals of the genera *Oleria* and *Hyposcada* (02–2105, 02–3519, 02–1644, 02–198, 02–3609, Ec467, Ec277, 02–515, 02–835 and 8369) for mitochondrial DNA, *wg* and *Ef1α* (A.W., unpublished data); STRI-B-850, STRI-B-442, and STRI-B-2980 for the mitochondrial region (Beltrán et al. 2002); RB283, P1–07, CR13, G16-04, E12-01, CU1, and V20 for *wg* (Brower and DeSalle 1998, Brower 2000); PE12-03 for *Ef1α,* AZ1-3 for *Ef1α* and *wg,* PNG2-3 and CU1 for *Ef1α* and mitochondrial DNA, and GH-058 for *Ef1α, wg,* and
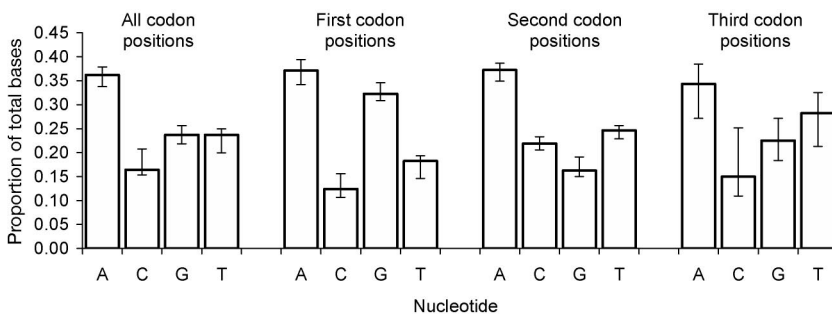
Fig. 1. Base compositions of the 807-bp aligned *Tektin* region, averaged over all specimens. Error bars depict minimum to maximum range.
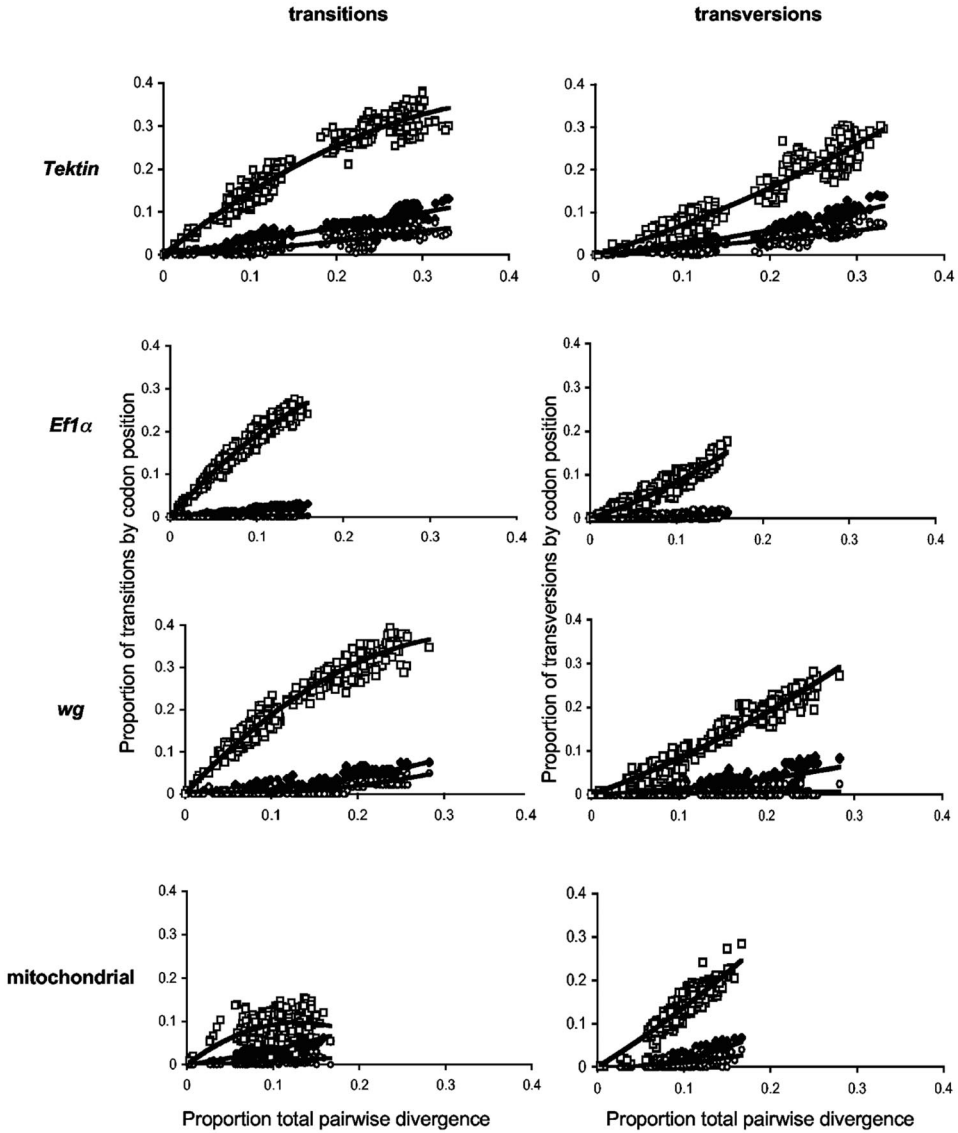
Fig. 2. Proportion of transitions or transversions against proportion total pairwise divergence for; *Tektin*, *Ef1α*, *wg*, and mitochondrial data. ◆ represents the first, ○ represents the second, and □ represents the third codon positions.

mitochondrial DNA (A.V.Z.B., personal correspondence); C3-08 for mitochondrial DNA (Brower and Jeansonne 2004); and P7-04 for mitochondrial DNA (Brower 1994). All other comparative sequences were obtained from Brower et al. (2005).

Protocols for *wg* closely followed those listed in Brower and DeSalle (1998), using primers LepWG1 and LepWG2 for amplification and sequencing. Parameters for amplification of the *Ef1α* region are described in Beltrán et al. (unpublished data), Mallarino et al. (2005), and Brower et al. (2005), and use primers listed therein (modified from primers in Cho et al. 1995). Protocols for the mitochondrial region largely follow those listed in Beltrán et al. (2002), Brower and Jeansonne (2004), Brower and colleagues (2005), and

Mallarino et al. (2005), using various combinations of the following primers for PCR and/or sequencing: Barbara I, Dick, Eva, George I/III, Ike, Imelda, Jane, Jerry, Jesse, Nancy, Liddy, Mamie, Pat, Phyllis, Romeo, Ron, Rudy, Rush, Strom, and Wyman.

Data Analyses. Sequences were edited using SeqEd v1.0.3 software (Applied Biosystems, Foster City, CA), and peptide sequences generated in MacClade 4.0 (Maddison and Maddison 1997). Sequences were deposited in GenBank (Accession numbers listed in Table 1). PAUP version 4.0b 10 (Swofford 2000) was used to calculate: numbers of variable and parsimony informative sites, uncorrected pairwise divergences, nucleotide composition, transition, and transversion statistics, and to perform a χ^2 test of base frequency
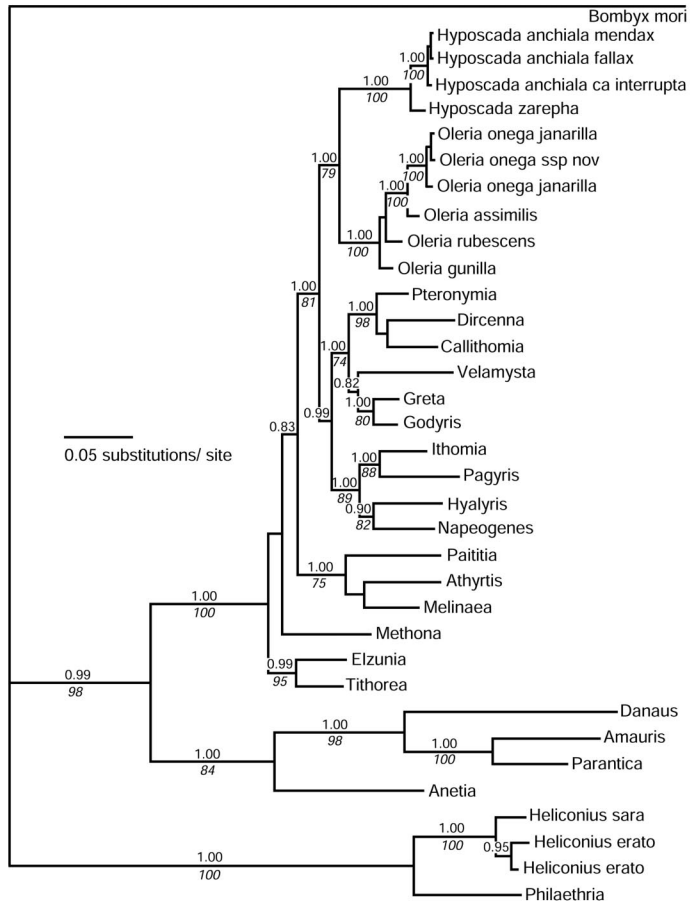
Fig. 3. Phylogenetic hypothesis based on *Tektin* nucleotide data. Tree topology is consensus of the last 9,000 trees inferred using Bayesian methods. Bayesian probabilities above 0.70 (above branch) and parsimony bootstrap support values higher than 70% (below branch) are given.

homogeneity. Data partition homogeneity tests (Farris et al. 1994) also were implemented in PAUP between all pairwise combinations of gene partitions in all individuals for which complete or nearly-full length data were available for both gene partitions. Phylogenetic analyses were performed using MrBayes 3.0 (Huelsenbeck and Ronquist 2001), with nst = 6 and a gamma rates heterogeneity model, with four simultaneous chains run for 1,000,000 generations, sampling a tree every 100 generations. A consensus tree with branch support in the form of posterior probabilities was derived from the final 9,000 trees (representing the final 900,000 generations), after confirmation that likelihood values had stabilized. We also conducted a heuristic search with TBR branch swapping in PAUP to find the most parsimonious trees, and summarized these with a majority rule consensus tree. Parsimony node support was assessed by Bremer support and bootstrapping with 1000 replicates. To dissect the contributions of the individual genes in recovering relationships at different taxonomic levels, partitioned Bremer support values (Baker and DeSalle 1997) were

calculated for the total combined data, by using Tree-Rot (Sorenson 1999). Finally, a phylogenetic hypothesis for the translated sequences with 1,000 bootstrap replicates was estimated using neighbor joining in PAUP.

Results

Of the pairwise final *Tektin* primer combinations tested, primers TektinA and Tektin3 consistently generated the strongest PCR products in all specimens except for Heliconiinae, for which primers TektinB and Tektin4 were optimal. Sequence data generated using TektinA with Tektin3 and TektinB with Tektin4 were edited to give 729 bp and 752–763-bp intronless regions, respectively (removal of ambiguous chromatogram peaks resulted in some data loss at the three prime end). Within the 807-bp final alignment, 685 bp of overlap was amplified using both primer pairs, of which 387 nucleotides were variable among all specimens (317 of which were parsimony informative [PI]), 364 variable (310 PI) among all Nymphalidae,
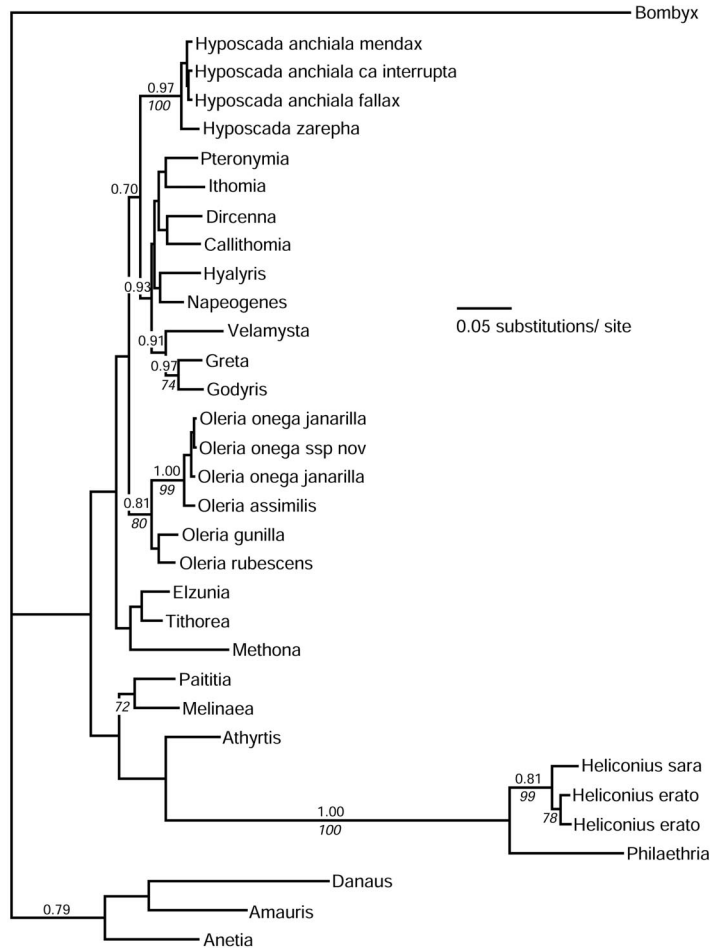
Fig. 4. Phylogenetic hypothesis based on *wg* nucleotide data. Tree topology is consensus of the last 9,000 trees inferred using Bayesian methods. Bayesian probabilities above 0.70 (above branch) and parsimony bootstrap support values higher than 70% (below branch) are given.

328 variable (257 PI) among Danainae and Ithomiinae, 278 variable (189 PI) within Ithomiinae, and 84 variable (67 PI) among *Oleria* and *Hyposcada* (Ithomiinae: Oleriini). These variable sites are approximately randomly distributed along the entire length of the amplified region.

Wingless sequences produced a 417-bp final alignment. Partially overlapping mitochondrial and *Ef1α* regions had been amplified, due to the different primer choices of collaborating laboratories. These sequences were edited to give final alignments of the 1618 bp (mitochondrial) and 1072 bp (*Ef1α*), which were complete for most taxa. The mitochondrial region has the greatest number of variable sites, for example, 577 (436 PI) within Nymphalidae and 497 (347 PI) among Ithomiinae. However, when accounting for length differences, *Tektin* is more variable, having 45.1 and 34.4% variable sites within Nymphalidae and Ithomiinae respectively, compared with the corresponding values of 35.7 and 30.7% for the mitochondrial data. Both absolutely and relative to length,

Tektin has higher variability than the other nuclear regions. The *Ef1α* fragment has 292 (27.2%) variable sites (235 PI) among the Nymphalidae and 238 (22.2%) (158 PI) among the Ithomiinae. And the shortest region, *wg*, has 163 (39.1%) (130 PI) and 121 (29.0%) (83 PI) variable sites within the Nymphalidae and Ithomiinae, respectively.

The amplified *Tektin* region is AT rich (Fig. 1). The sum of the mean contributions of A and T nucleotides to the 807-bp alignment is 59.9%, with first, second, and third codons having, respectively, increasing AT biases of 55.3, 61.8, and 62.5%. The overall AT bias is more pronounced than in *Ef1α* (49.5%) and *wg* (46.2%), but it was considerably less than that in mitochondrial DNA (75.1%) in corresponding taxa. Heterogeneity of nucleotide composition of synonymous sites was observed. For example, in *Tithorea*, there were strong departures at four-fold degenerate sites (21A, 9C, 8G, and 31T), from the overall nucleotide composition of the *Tektin* region (χ^2 test, $P < 0.001$). χ^2 tests for base frequency homogeneity, for the total
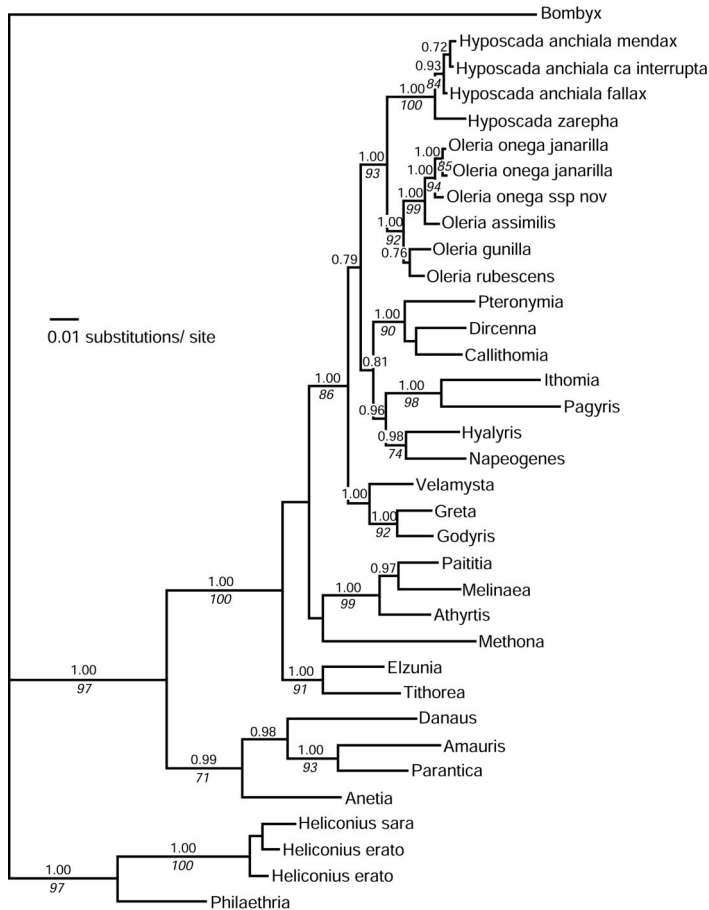
Fig. 5. Phylogenetic hypothesis based on *Ef1α* nucleotide data. Tree topology is consensus of the last 9,000 trees inferred using Bayesian methods. Bayesian probabilities above 0.70 (above branch) and parsimony bootstrap support values higher than 70% (below branch) are given.

Tektin data set, first, second, and third codon positions, revealed no significant differences across taxa ($P > 0.95$).

To assess the extent of saturation, the proportion of sites with transitional and transversional pairwise changes were plotted against uncorrected pairwise divergences (Fig. 2). The near linear trends of all transversions, and first and second codon position transitions, imply that saturation has not been reached for any of the pairwise comparisons, in any gene. The third position transition trend lines begin to plateau as the more phylogenetically distant taxa are considered, thus deeper evolutionary hypotheses from each data set might be confounded by multiple substitutions. This plateau effect was strongest for the mitochondrial data set.

The different genes gave strong heterogeneous support for phylogenetic hypotheses: six *wg* and mitochondrial, 20 *Ef1α*, and 22 *Tektin* branches were supported by both a Bayesian posterior probability >0.70, and a bootstrap value $>70\%$ in the parsimony analysis (Figs. 3–6; Table 3). There are cases of agreement between the phylogenetic hypotheses inferred by

Bayesian methods across all genes, including monophyly of both Heliconiinae and Danainae, the Danainae internal topology, and the basal position of *Hyposcada zarepha* in the *Hyposcada* clade. There are also a number of examples where a particular topology is unique to a specific gene, for example, *Ef1α* fails to place the two *Heliconius erato* as sister taxa, and the mitochondrial data do not recover *Elzunia* as sister to *Tithorea*, or *Velamysta* as basal to *Greta* + *Godyris*. Additionally, there are cases of repeated discordance, such as the weakly supported position of *Methona*. Discordances among the individual gene trees are found at all phylogenetic levels, and within clades several different patterns of topological agreement can be observed. For example, within the *Oleria* clade, *Tektin* and *wg* place *O. onega* ssp. nov. as sister to one of the two *O. onega janarilla* individuals, whereas mitochondrial DNA and *Ef1α* group the two *O. onega janarilla* individuals; all genes placed *O. assimilis* as outside *O. onega*, but *Ef1α* and *wg* place *O. gunilla* and *O. rubescens* as a sister clade to the remaining *Oleria* taxa, whereas mitochondrial DNA and *Tektin*, respec-
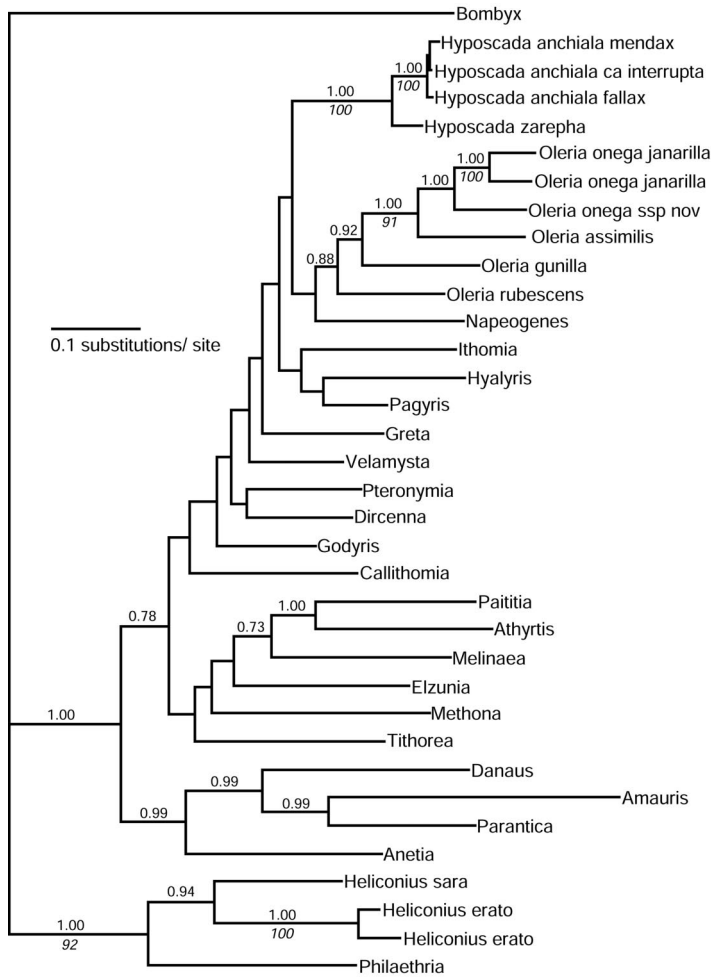
Fig. 6. Phylogenetic hypothesis based on mitochondrial nucleotide data. Tree topology is consensus of the last 9,000 trees inferred using Bayesian methods. Bayesian probabilities above 0.70 (above branch) and parsimony bootstrap support values higher than 70% (below branch) are given.

tively, support *O. rubescens* or *O. gunilla* as sister to all other *Oleria*.

Given these topological conflicts, we investigated which relationships emerged when the data were

Table 3. Number of Bayesian inferred branches with strong Bayesian, Parsimony, or Bayesian and Parsimony support

	Bayesian	Parsimony	Bayesian + Parsimony
Tektin	4 (0.90)		22 (0.99, 92%)
wg	4 (0.83)	2 (75%)	6 (0.93, 92%)
Ef1α	8 (0.87)		20 (1.00, 92%)
Mitochondrial	11 (0.93)		6 (1.00, 97%)

Branches are regarded as strongly supported if they have a Bayesian posterior probability >0.70, or bootstrap value >70% in the parsimony analysis. Mean Bayesian posterior probability and bootstrap values are given in parentheses.

combined in a single analysis. Partitioned homogeneity tests indicated little evidence for heterogeneity (*Ef1α*:*Tektin* $P = 0.91$, *Tektin*:*wg* $P = 0.94$, *Tektin*:mitochondrial DNA $P = 0.13$, *Ef1α*:*wg* $P = 0.98$, *Ef1α*:mitochondrial DNA $P = 0.45$, *wg*:mitochondrial DNA $P = 0.43$), so all genes were concatenated for the overall analysis. The Bayesian total evidence analysis (Fig. 7) and maximum parsimony heuristic search (Fig. 8) inferred slightly different topologies; maximum parsimony placed *Paititia*, *Athyrtis*, and *Melinaea* as the most basal clade, with *Elzunia*, *Tithorea*, and *Methona* forming a clade that was sister to remaining Ithomiinae (the clade containing these last two clades had a bootstrap value of 72%). In contrast, Bayesian analyses placed *Elzunia* and *Tithorea* as a clade sister to remaining Ithomiinae, then *Methona* (posterior probability 0.56), then the *Athyrtis*, *Paititia*, *Melinaea* clade as sister to remaining Ithomiinae (posterior probability 0.56).
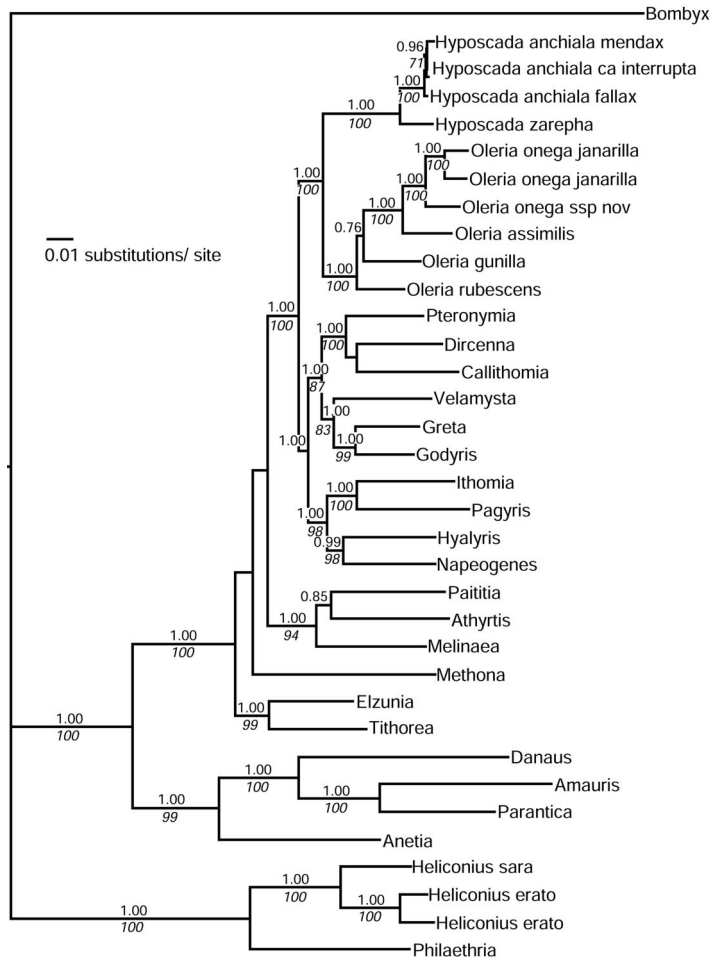
Fig. 7. Total evidence phylogenetic hypothesis. Tree topology is consensus of last 9,000 trees inferred using Bayesian methods. Bayesian probabilities above 0.70 (above branch) and parsimony bootstrap support values higher than 70% (below branch) are given.

Overall, the total evidence Bayesian topology was most similar to the individual topologies inferred using *Tektin* and *Ef1α*. The *Tektin* and *Ef1α* topologies both differed from the total evidence topology in arrangements of the *Paititia, Athyrtis,* and *Melinaea* clade, and placements of *Oleria rubescens* and *Oleria gunilla.* In addition, the *Tektin* topology differed from the total evidence topology at the intraspecific level (i.e., *Hyposcada anchiala* and *Oleria onega*), whereas *Ef1α* differed in the arrangement of the *Velamysta, Greta, Godyris* clade, position of *Methona,* and relationships between *Heliconius* species.

Mitochondrial DNA provides the strongest support for relationships between closely related taxa, as measured by Partitioned Bremer support values within the *Hyposcada* and *Oleria* clades. The Partitioned Bremer support analysis also revealed that *Tektin* most strongly supported the deeper nodes, indicating it contributed significantly to resolving positions of the most divergent taxa in the combined analysis.

Topologies inferred from Tektin protein (Fig. 9) and *Tektin* nucleotide data are in broad agreement, for example, by recovering the monophyly of the three nymphalid subfamilies, the basal positions of *Elzunia, Tithorea,* and *Methona* within the Ithomiinae clade, and the derived position of the monophyletic Oleriini. *Velamysta* assumes the most discordant position, despite having lowest protein character differences to the taxa with which it clusters based on nucleotide data (14 with *Greta* and 12 with *Godyris,* compared with, for example, 21 with *Napeogenes*).

Discussion

Tektin proved easy to amplify from genomic DNA in all specimens tested, was amenable to direct sequencing, was easily aligned and exhibited a level of variation particularly appropriate for resolving phylogenetic relationships at the genus, tribe, and subfamily levels.
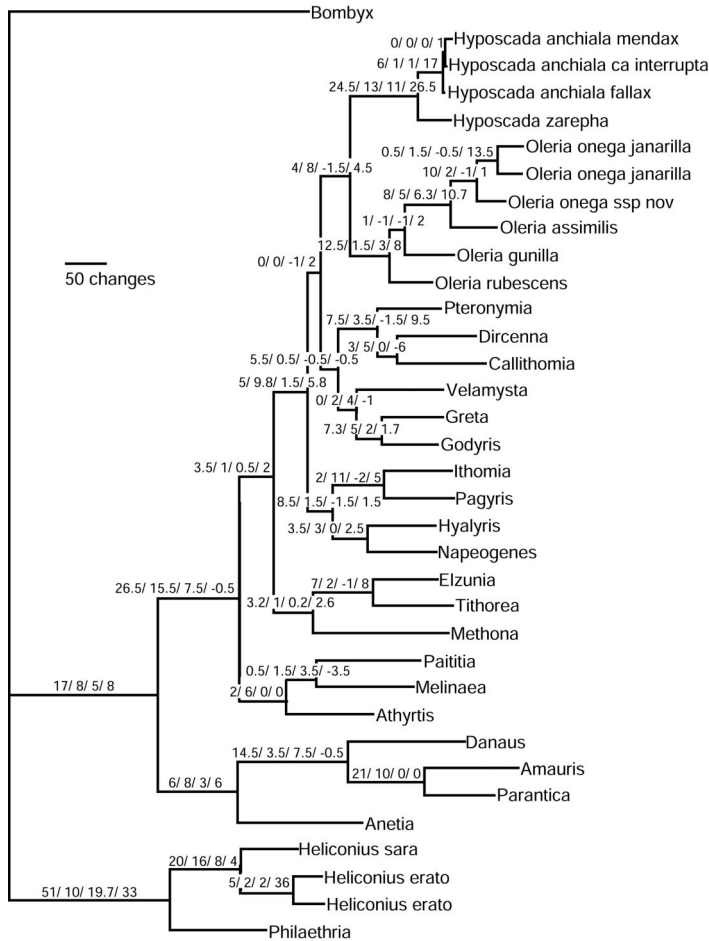
Fig. 8. Total evidence phylogenetic hypothesis. Tree topology was inferred with maximum parsimony. Partitioned Bremer support values are given above branches (*Tektin*/ *Ef1α*/ *wg*/ mitochondrial).

Tektin products belong to a protein family, thus the potential for amplifying paralogs was of concern. However, the high nucleotide differentiation of *Tektin* paralogs makes them easily distinguishable, for example, uncorrected pairwise distances of A:B 52%, A:C 48%, and B:C 52% were recovered for a 951-bp region of purple sea urchin *Tektin,* which aligns without insertions and deletions (GenBank Accession numbers M97188, L21838, and U38523). As the pairwise distance values recovered in this study are all significantly lower than the distances observed between *Tektin* paralogs in other species, for example, within ithomiines (0–15.2%, mean 10.3%) and even between the silkworm outgroup and all other individuals (26.5–32.1%, mean 29.1%), it is highly likely that all our sequences are actually orthologous. In addition, phylogenetic analysis of Genbank *Tektin* sequences in human, mouse, sea urchin, fruit fly, silkworm, and our ithomiine sequences (data not shown) shows that *Tektin* sequences cluster together by paralog rather than by the species to which they belong. This type of clustering indicates that individual *Tektin* paralogs really are distinct enough to behave as single-copy genes. Easily distinguishable paralogs have previously proven effective for phylogenetic analyses, including *Ef1α* and *wg* (Cho et al. 1995, Brower and DeSalle 1998).

The mitochondrial and *Tektin* regions had the highest number of variable sites. However, when accounting for length differences, *Tektin* had the highest per-site variability. Despite its faster evolutionary rate, and the resultant concern that deep level comparisons using highly variable regions are often more confounded by the effects of saturation, *Tektin* proved to be more informative than mitochondrial DNA for the more distant comparisons. This relationship is probably due to the biased nucleotide composition of the mitochondrial region (75.1%, compared with 59.9% in *Tektin*) hindering its deeper resolving power.

That the *Tektin* topology was in broad agreement to traditional classification and a recent cladogram inferred from morphology (Fig. 10; K.R.W. and A.V.L.
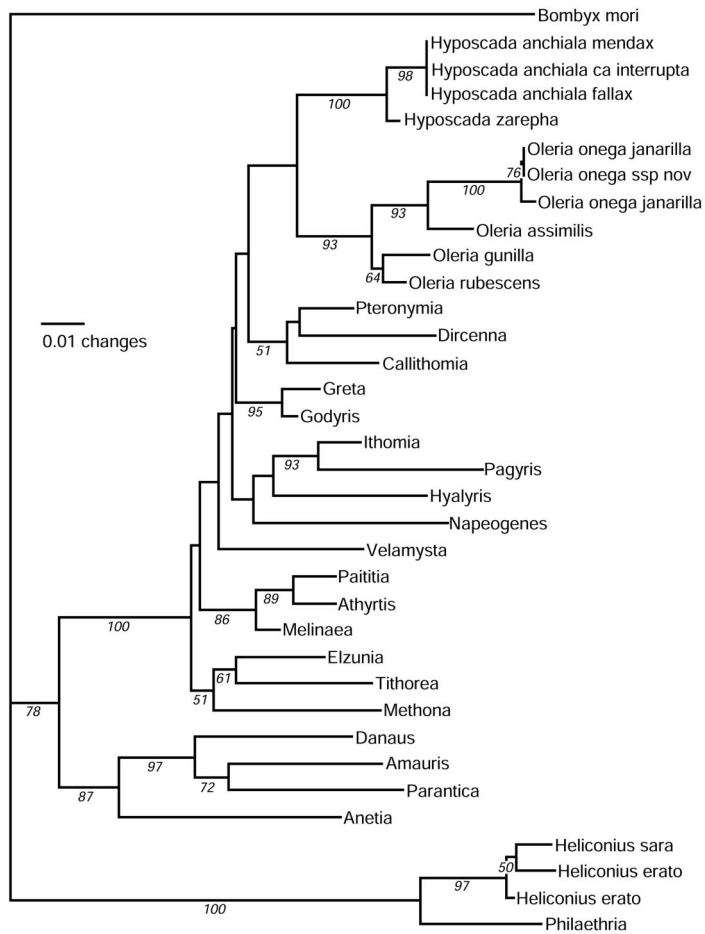
Fig. 9. Phylogenetic hypothesis based on Tektin amino acid data, inferred with neighbor joining. Neighbor joining bootstrap support values are given below branches.
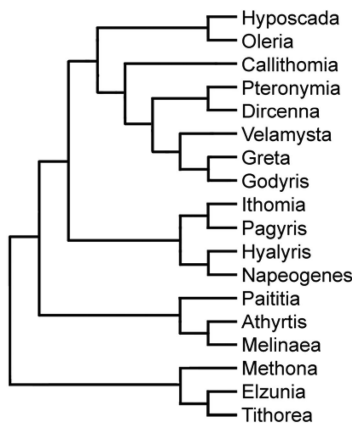

Fig. 10. Most parsimonious tree inferred from combined morphological and ecological data after successive approximations character weighting (K.R.W. and A.V.L. Freitas, personal communication).

Freitas, personal communication) strongly supports the utility of *Tektin* as a tool for successfully inferring relationships. The *Tektin* gene tree was most similar to that obtained from the *Ef1α* data set, despite the differing placements of *Velamysta, Greta,* and *Godyris.* Both *Tektin* and *Ef1α* topologies are well resolved, with most branches well supported. Mitochondrial DNA best differentiated closely related taxa within the Oleriini, but it failed to provide congruent support for intermediate depth internal branches and its placement of *Napeogenes* within Oleriini was clearly erroneous; both effects are probably due to the confounding effects of saturation, a by-product of the limited taxonomic sampling employed in this analysis. The topology generated from the *wg* data has more internal structure than the mitochondrial tree but is quite poorly resolved, in part because of short sequence length. These findings are based on our taxa of choice and may not have done justice to each gene, which have previously proven to be useful (Simon et al. 1994, Cho et al. 1995, Brower and DeSalle 1998, Caterino et al. 2000). Our purpose was not to advocate or discour-

age general use of these "paradigm" genes, but to provide comparative evidence to enable future studies to select the combination of genes most appropriate for their level of phylogeny estimate.

After combining *Tektin* with COI and COII, *wg* and *Ef1α* regions, partitioned Bremer analyses demonstrated *Tektin* to be particularly effective at recovering the genus, tribe, and subfamily relationships. The total evidence tree represents our preferred phylogenetic hypothesis, as is to be expected on theoretical grounds (DeSalle and Brower 1997), and also shows good agreement with the morphological tree (Fig. 10). Given the close correspondence between trees derived from *Tektin* alone and both combined evidence molecular and morphological trees, *Tektin* is arguably the most accurate single marker for inferring relationships among all taxa studied. We recommend future studies use *Tektin* in conjunction with other genomic regions, to obtain robust and reliable phylogenetic hypotheses that best imply the organismal phylogeny. We suggest *Tektin* to be a very useful addition to the molecular phylogenetic arsenal for studies of Lepidoptera. Our findings strongly endorse *Tektin* as a new candidate gene for phylogenetic and perhaps phylogeographical studies of other insect groups.

Acknowledgments

References Cited

Ackery, P. R., R. De Jong, and R. I Vane-Wright. 1999. The butterflies: Hedyloidea, Hesperioidea and Papilionoidea, pp. 263–300. *In* N. P. Kristensen [ed.], Lepidoptera, moths and butterflies. 1. Evolution, systematics and biogeography. Handbook of Zoology 4 (35), Lepidoptera. de Gruyter, Berlin, Germany.

Baker, R. H., and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46: 654–673.

Beltrán, M. S., C. D. Jiggins, V. Bull, M. Linares, J. Mallet, W. O. McMillan, and E. Bermingham. 2002. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. Mol. Biol. Evol. 19: 2176–2190.

Bensasson, D., D.-X. Zhang, D. L. Hartl, and G. M. Hewitt. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16: 314–321.

Brower, A.V.Z. 1994. Phylogeny of *Heliconius* butterflies inferred from mitochondrial DNA sequences (Lepidoptera: Nymphalidae). Mol. Phylogenet. Evol. 3: 159–174.

Brower, A.V.Z. 2000. Phylogenetic relationships among the Nymphalidae (Lepidoptera) inferred from partial sequences of the *wingless* gene. Proc. R. Soc. Lond. B Biol. Sci. 267: 1201–1211.

Brower, A.V.Z., and R. DeSalle. 1998. Patterns of mitochondrial vs. nuclear DNA sequence divergence in nymphalid butterflies: the utility of *wingless* as a source of characters for phylogenetic inference. Insect Mol. Biol. 7: 73–82.

Brower, A.V.Z., and M. M. Jeansonne. 2004. Geographical populations and 'subspecies' of New World monarch butterflies (Nymphalidae) share a recent origin and are not phylogenetically distinct. Ann. Entomol. Soc. Am. 97: 519–523.

Brower, A.V.Z., R. DeSalle, and A. Vogler. 1996. Gene trees, species trees, and systematics. A cladistic perspective. Annu. Rev. Ecol. Syst. 27: 423–450.

Brower, A.V.Z., A.V.L. Freitas, M.-M. Lee, K. L. Silva Brandão, and A. Whinnett. 2005. Phylogenetic relationships among the Ithomiini (Lepidoptera: Nymphalidae) inferred from one mitochondrial and two nuclear gene regions. Syst. Entomol. (in press).

Caterino, M. S., S. Cho, and F.A.H. Sperling. 2000. The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. 45: 1–54.

Chen, R., C. A. Perrone, L. A. Amos, and R. W. Linck. 1993. Tektin B1 from ciliary microtubules: primary structure as deduced from the cDNA sequence and comparison with Tektin A1. J. Cell Sci. 106: 909–918.

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 1-alpha* recovers morphology-based tree for heliothine moths. Mol. Biol. Evol. 12: 650–656.

Cummings, M. P. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol. Evol. 9: 141–145.

DeSalle, R., and A.V.Z. Brower. 1997. Process partitions, congruence and the independence of characters: inferring relationships among closely-related Hawaiian *Drosophila* from multiple gene regions. Syst. Biol. 46: 751–764.

Doyle, J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. Syst. Bot. 17: 144–163.

Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54: 1839–1854.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics 10: 315–319.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4: 6–11.

Hudson, R. R., and M. Turelli. 2003. Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. Evolution 57: 182–190.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.

Hurst, G.D.D., and F. M. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic, and phylogenetic studies: the effects of inherited symbionts. Proc. R. Soc. B. 272: 1525–1534.

Kidwell, M. G. 1993. Lateral transfer in natural populations of eukaryotes. Annu. Rev. Genet. 27: 235–256.

Lamas, G. 2004. Nymphalidae. Ithomiinae, pp. 172–191. *In* G. Lamas [ed.], Checklist: Part 4A. Hesperioidea - Papilionoidea. *In* J. B. Heppner [ed.], Atlas of Neotropical

Lepidoptera. Volume 5A. Association for Tropical Lepidoptera, Scientific Publishers, Gainesville, FL.

Linck, R. W. 1982. The structure of microtubules. Ann. N.Y. Acad. Sci. 383: 98–121.

Linck, R. W. 1990. Tektins and microtubules. Adv. Cell Biol. 3: 35–65.

Linck, R. W., and G. L. Langevin. 1982. Structure and chemical composition of insoluble filamentous components of sperm flagella microtubules. J. Cell Sci. 58: 1–22.

Linck, R. W., L. A. Amos, and B. Amos. 1985. Localisation of Tektin filaments in microtubules of sea urchin sperm flagella by immunoelectron microscopy. J. Cell Biol. 100: 126–135.

Linck, R. W., D. F. Albertini, D. M. Kenney, and G. L. Langevin. 1982. Tektin filaments: chemically unique filaments of sperm flagellar microtubules. Prog. Clin. Biol. Res. 80: 127–132.

Machado, C. A., R. M. Kliman, J. A. Markert, and J. Hey. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and its close relatives. Mol. Biol. Evol. 19: 472–488.

Maddison, W. P., and D. R. Maddison. 1997. MacClade: analysis of phylogeny and character evolution. Sinauer, Sunderland, MA.

Mallarino, R., E. Bermingham, K. R. Willmott, A. Whinnett, C. D. Jiggins. 2005. Molecular systematics of the butterfly genus *Ithomia* (Lepidoptera: Ithomiinae): a composite phylogenetic hypothesis based on seven genes. Mol. Phylogenet. Evol. 34: 625–644.

Mallet, J. 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20: 229–237.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation, pp. 515–534. *In* S. Karlin and E. Nevo [eds.], Evolutionary processes and theory. Academic, New York.

Nojima, D., R. W. Linck, and E. H. Egelman. 1995. At least one of the protofilaments in flagella microtubules is not composed of tubulin. Curr. Biol. 5: 158–167.

Norrander, J. M., L. A. Amos, and R. W. Linck. 1992. Primary structure of tekin A1: comparison with intermediate filament proteins and a model for its association with tubulin. Proc. Natl. Acad. Sci. U.S.A. 89: 8567–8571.

Norrander, J. M., C. A. Perrone, L. A. Amos, and R. W. Linck. 1996. Structural comparison of Tektins and evidence for their determination of complex spacings in flagellar microtubules. J. Mol. Biol. 257: 385–397.

Ota, A., T. Kusakabe, Y. Sugimoto, M. Takahashi, Y. Nakajima, Y. Kawaguchi, and K. Koga. 2002. Cloning and characterisation of testis-specific Tektin in *Bombyx mori.* Comp. Biochem. Physiol. B Biochem. Mol. Biol. 133: 371–382.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5: 568–583.

Philippe, H., and C. J. Douady. 2003. Horizontal gene transfer and phylogenetics. Curr. Opin. Microbiol. 6: 498–505.

Pirner, M. A., and R. W. Linck. 1994. Tektins are heterodimeric polymers in flagellas microtubules with axial periodicities matching the tubulin lattice. J. Biol. Chem. 269: 31800–31806.

Rokas, A., B. L. Williams, N. King, and S. B. Carroll. 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature (Lond.) 425: 798–804.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic usefulness of mitochondrial genes with a compilation of conserved PCR primers. Ann. Entomol. Soc. Am. 87: 651–701.

Sorenson, M. D. 1999. TreeRot, version 2. Boston University, Boston, MA.

Swofford, D. L. 2000. *PAUP**: phylogenetic analysis using parsimony (* and other methods). Sinauer, Sunderland, MA.

Syvanen, M. 1994. Horizontal gene transfer: evidence and possible consequences. Annu. Rev. Genet. 28: 237–261.

Takahata, N., and M. Nei. 1985. Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110: 325–344.