Syst. Biol. 55(1):57–72, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150500431213

Optimal Intron Analyses in the *Trimeresurus* Radiation of Asian Pitvipers

SIMON CREER, CATHARINE E. POOK, ANITA MALHOTRA, AND ROGER S. THORPE

School of Biological Sciences, University of Wales, Bangor, Gwynedd, LL57 2UW, UK; E-mail: s.creer@bangor.ac.uk (S.C.)

Abstract.—Nuclear introns are commonly used as phylogenetic markers, but a number of issues related to alignment strategies, indel treatments, and the incorporation of length-variant heterozygotes (LVHs) are not routinely addressed when generating phylogenetic hypotheses. Topological congruence in relation to an extensive mitochondrial DNA multigene phylogeny (derived from 2423 bp of 12S, 16S, ND4, and CYTB genes) of the Asian pitviper *Trimeresurus* radiation was used to compare combinations of "by eye" and edited and unedited ClustalX 1.8 alignments of two nuclear introns. Indels were treated as missing data, fifth character states, and assigned simple and multistate codes. Upon recovery of the optimal alignment and indel treatment strategy, a total evidence approach was used to investigate the phylogenetic utility of the indels and test new generic arrangements within *Trimeresurus.* Approximately one third of the intron data partitions exhibited LVHs, suggesting that they are common in introns. Furthermore, a simple concatenation approach can facilitate the incorporation of LVHs into phylogenetic analyses to make use of all available data and investigate mechanisms of molecular evolution. Analyses of ClustalX 1.8–assisted alignments were generally more congruent than the "by eye" alignment and the analysis of a simple coded, edited ClustalX 1.8 (gap opening cost 5, gap extension cost 1) alignment revealed the most congruent tree. The total evidence approach supported the new arrangements within *Trimeresurus,* suggesting that the phylogeny should be considered as a working benchmark in Asian pitviper systematics. Finally, a critical appraisal of the diverse array of indels (56 to 57 per intron, ranging from 1 to 151 bp in length) suggested that they are a combination of Hennigian and homoplasious events unrelated to indel size or location within the intron. [Alignment; indels; intron analysis; length-variant heterozygotes; *Trimeresurus*.]

The concept of using nuclear introns as phylogenetic markers was first introduced approximately 10 years ago (Lessa, 1992; Slade et al., 1993, 1994). Since then, intron data partitions have been used as stand-alone independent markers and as nuclear corroborative support for mitochondrial DNA (mtDNA) partitions in numerous studies (for examples, see Prychitko and Moore, 1997; Pitra et al., 2000; Johnson and Clayton, 2000; Rockman et al., 2001; Oakley and Phillips, 1999; Rowe and Honeycutt, 2002; Birks and Edwards, 2002; Creer et al., 2003). Although introns have been shown to be involved in alternative splicing (Leicht et al., 1995) and gene regulation mechanisms (Gasch et al., 1989; Alder et al., 1992; Kirby et al., 1995; Prychitko and Moore, 2003), empirical approaches have shown that they can be considered as neutral markers that possess a number of traits that are desirable for molecular phylogenetics (Friesen et al., 1997; Friesen, 2000). The overt lack of functional constraints and the fact that most bases have the potential to yield phylogenetically informative sites mean that introns generally exhibit lower levels of homoplasy than mtDNA (Slade et al., 1994; Prychitko and Moore, 2000, 2003). Furthermore, the convenience of utilizing an exon-primed intron crossing (EPIC) primer design strategy (Lessa, 1992; Slade et al., 1993) facilitates the identification and amplification of target regions.

There are, however, a number of issues that are central to the appropriate application of intron data in phylogenetic reconstruction that are not always comprehensively and explicitly addressed. These include the application of suitable alignment methodology, appropriate treatment of indels (or gaps—sequence positions that have experienced an insertion or deletion event) (Giribet and Wheeler, 1999), and the identification and incorporation of length-variant heterozygotes (LVHs) (Sota and Vogler, 2003). Inadequate treatment of one or more of the above aspects may at best lead to a poorer understanding of the molecular markers involved and at worst provide an incorrect phylogenetic hypothesis.

Alignment

Alignments can either be achieved manually "by eye" or assisted by computer program implementing the dynamic programming algorithm of Needleman and Wunsch (1970). The subjectivity of "by eye" alignments has been criticized (e.g., Giribet and Wheeler, 1999), but empirical studies have shown that manual alignments are not significantly worse than computer assisted alignments (Sanchis et al., 2001; Belshaw and Quicke, 2002). If, however, computer assisted alignments are used, a range of costs are available to simulate the relative frequency of the occurrence of indels relative to substitution events. Low gap costs will favor alignments with larger numbers of indels, whereas high gap costs will produce alignments with fewer indels (Page and Holmes, 1998; Phillips et al., 2000). Thus, due to the data partition specific combination of the size and frequency of the indel events, different alignment parameters will be more suitable for indel rich and poor datasets.

Indel Treatment

The treatment of indels in phylogenetic analyses has received considerable attention over the past few years, resulting in varying points of view (Giribet and Wheeler, 1999; Simmons and Ochoterena, 2000; Lutzoni et al., 2000). Consequently, a number of strategies have been suggested regarding the analysis of indel data with substitutional data in order to generate robust phylogenetic hypotheses. The classical approach is to consider indels as missing data (Kumar et al., 1993; Swofford et al., 1996; Swofford, 1998), but such a treatment effectively disregards a potentially valuable source of information that represents actual biological events. Alternatively,

indels can be incorporated into phylogenetic analyses by a number of other strategies. Indel positions may be regarded simply as a fifth character state (Swofford, 1998). Indels with different start and/or end positions may be replaced by a "simple coded" binary matrix (coded as presence/absence), which is analyzed alongside the DNA base characters (Simmons and Ochoterena, 2000—hereafter referred to here as simple coding). If sequences cannot be aligned equivocally, areas of ambiguous alignments can be replaced with multistate coded characters (step matrices) and analyzed alongside the DNA base characters (hereafter referred to here as multistate coding). Unlike simple coding, multistate coding identifies different character states within homologous regions and assigns costs relative to transformations between the identified states. Thus, the method aims to include areas of ambiguous alignment without violating positional homology (Lutzoni et al., 2000).

Finally, direct optimization (Wheeler, 1996; Wheeler and Gladstein, 2000) can be used, which incorporates indel events as additional transformations during the optimization step in tree evaluation instead of trying to reconcile sequence lengths by adding gaps as additional states. Unlike the above methods, which dissociate the process of aligning sequences from the phylogenetic analysis, in direct optimization, gap assignment is an intrinsic and inseparable part of parsimony inference (Cognato and Vogler, 2001; Hormiga et al., 2003).

LVHs

Introns of diploid organisms will either be heterozygotic or homozygotic. Individuals possessing heterozygotic introns of identical length are immediately visible in sequencing chromatograms (resulting from direct sequencing of PCR products) as dual peaks of approximately equal intensity occupy the same base position. However, if the heterozygotic alleles are different lengths, the result will almost certainly be the apparent termination of the sequencing reaction due to the superimposition of two separate sequence chromatograms occupying the same reading frame. If the latter is encountered, cloning the two alleles is the simple solution to obtain both copies of the locus. Although this issue was highlighted in the mid-1990s (Palumbi and Baker, 1994), a number of intron studies do not mention LVHs (Johnson and Clayton, 2000; Pitra et al., 2000; Jenkins et al., 2001; Rockman et al., 2001; Braband et al., 2002; Ericson et al., 2002; Rowe and Honeycutt, 2002), while others have tried but have not detected intraindividual length variation (Oppen et al., 2000; Prychitko and Moore, 2000; Birks and Edwards, 2002). Recently, however, studies have detected and incorporated LVHs into phylogenetic frameworks (Beltrán et al., 2002; Sota and Vogler, 2003; Pons et al., 2004), suggesting that the phenomenon is common at certain loci and should be considered as a matter of course in studies using introns as phylogenetic markers.

Once LVHs have been incorporated into data sets and the application of different alignment parameters and indel treatment methods have been explored, a decision should be made to favor one of the competing hypotheses (Giribet and Wheeler, 1999; Sanchis et al., 2001). If measures of taxonomic congruence are not available (e.g., morphological data, other genes less sensitive to alignment/indel homology statements, or extremely large data sets), all phylogenetic hypotheses should be presented and common topologies discussed regarding phylogenetic consensus (Giribet and Wheeler, 1999; Belshaw and Quicke, 2002; Arnedo et al., 2004). If, however, an independent robust phylogeny fulfilling the above criteria is available, taxonomic congruence (based on the assumption that synapomorphies are homologous) can be used as a measure favoring alignment and indel treatments that maximize phylogenetic consensus (Giribet and Wheeler, 1999; Cognato and Vogler, 2001; Giribet, 2001; Belshaw and Quicke, 2002; but see Simmons, 2004, for counterpoint).

Therefore, the ideal system to study the optimal analysis of intron data is a taxonomic group for which an independent robust phylogenetic hypothesis already exists. The Asian pitvipers, particularly those belonging to the *Trimeresurus* radiation, which have recently been split into seven genera by Malhotra and Thorpe (2004), offer this opportunity. Malhotra and Thorpe's (2004) phylogenetic hypothesis was derived from 2423 bp of 12S ribosomal RNA (12S), 16S ribosomal RNA (16S), NADH subunit 4 (ND4), and cytochrome *b* (CYTB) mtDNA gene partitions for New and Old World pitvipers, which includes extensive sampling of members of the *Trimeresurus* radiation.

Here, therefore, we aim to identify appropriate methodologies concerning alignment, indel treatment, and the incorporation of LVHs of two nuclear introns in the Asian pitvipers of the *Trimeresurus* radiation. Topological congruence with the mtDNA multigene phylogeny of Malhotra and Thorpe (2004) will be used to evaluate the optimal intron analysis method. Assuming that any mtDNA phylogenetic hypothesis represents an organismal phylogeny has its limitations. However, the Malhotra and Thorpe (2004) study offers the most comprehensive molecular appraisal of Asian pitvipers to date and the new generic arrangements are based on a combination of strongly supported monophyletic groupings and select diagnostic morphological characters. Finally, a total evidence (Kluge, 1989; Nixon and Carpenter, 1996) approach using six partitions of mitochondrial and nuclear data will be used to test the validity of the new generic arrangements within the *Trimeresurus* radiation and also critically appraise the utility of the indels (size, location, and mode of molecular evolution) in phylogenetic reconstruction.

MATERIALS AND METHODS

Taxon Sampling

Taxa were sampled to optimize representation of members of five out of seven recently designated genera within the *Trimeresurus* radiation included in the mtDNA analyses of Malhotra and Thorpe (2004).

Explicitly, these included individuals representing the genera *Cryptelytrops* (formerly the *Trimeresurus albolabris* species group—*C. albolabris*, *C. erythrurus*, *C. purpureomaculatus*, *C. septentrionalis*, *C. insularis*, *C. fasciatus*, *C. macrops,* and *C. venustus*), *Popeia* (formerly the *Trimeresurus popeiorum* species group—*P. popeiorum* and *P. sabahi*), *Viridovipera* (formerly the *Trimeresurus stejnegeri* species group— *V. vogeli* and *V. gumprechti*), *Parias* (formerly unassigned—*P. schultzei*, *P. flavomaculatus*, *P. malcolmi,* and *P. hageni,*) and *Trimeresurus* (formerly the Indian subcontinent species group—*T. trigonocephalus*, *T. gramineus*, *T. puniceus,* and *T. borneensis*). All former species groups can be cross-referenced in Malhotra and Thorpe (2000) and Creer et al. (2003). For the majority of taxa, intron gene regions were amplified from the same samples used in the Malhotra and Thorpe (2004) study. When this was not possible, introns were amplified from conspecifics from the same geographic locality/region.

Primer Design and Molecular Protocols

Whole genomic DNA extractions from liver and blood samples were as in Creer et al. (2003). The seventh intron from the β fibrinogen gene (7IβFIB) and the third intron from the TATA box-binding protein (3ITBP) were amplified using the following primer design (aided by primer 3 [Rosen and Skaletsky, 1998]) and PCR strategies.

The exon-primed intron crossing (EPIC) *Trimeresurus*-specific 7IβFIB primers TRIMFIB-BI7U—5′-AGAGAC-AATGATGGATG*GTAAG-3′ (* designates the 7th exon–7th intron boundary) and TRIMFIB-B17L—5′-CCTTTTGGGATCTGGGTGTA-3′ (22 bp downstream from the 7th intron–8th exon boundary) were designed from sequences from the same species/groups using the 7IβFIB primers listed in Prychitko and Moore (1997) using the PCR thermal cycling parameters in Creer et al. (2003). Thermal cycling parameters for the TRIMFIB-BI7 amplifications were an initial denaturation of 94°C for 5 min, followed by 35 cycles of denaturation at 93°C (1 min), annealing at 50°C ± 5°C (30 s), extension at 72°C (2 min), and a final extension of 72°C (15 min). The EPIC *Trimeresurus* specific 3ITBP primers TRIMTBPI3F—5′-CCTTTACCAGGAACCACACC-3′ (107 bp upstream from the 3rd exon–3rd intron boundary) and TRIMTBP-I3R—5′-CGAAGGGCAATGGTTTTTAG-3′ (60 bp downstream from the 3rd intron–4th exon boundary) were designed from alignments of cDNA mRNA sequences and genomic sequences of the TATA box-binding protein from *Viridovipera stejnegeri* and *Protobothrops flavoviridis* (referred to as *Trimeresurus gramineus* and *Trimeresurus flavoviridis,* respectively, in Nakashima et al. [1995]). Thermal cycling parameters for the TRIMTBPI3 amplifications were an initial denaturation of 94°C for 5 mins, followed by 35 cycles of denaturation at 94°C (1 min), annealing at 50°C to 55°C (1 min), extension at 72°C (1 min), and a final extension of 72°C (15 min). All 50 μl PCR reactions (1× PCR buffer, 0.5 to 3.5 mM $MgCl_2$, 0.8 mM total dNTPs, 0.4 mM primers, and 0.4 U Sigma JumpStart Taq) were performed on an Applied Biosystems GeneAmp PCR System 3700 alongside a negative control (upH_2O) to exclude the possibility of contamination. Double-stranded PCR products were electrophoresed on 1% agarose gels and visualized by ethidium bromide staining (Dowling et al., 1996). Target PCR products were then excised, pooled, and cleaned using Qiaquick columns (QIAGEN). PCR products were sequenced from both ends using dye-labeled terminators (ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit), and subsequently run on an ABI Prism 377 DNA sequencer. If premature stops, indicative of either size heterozygotes (Palumbi and Baker, 1994), or DNA secondary structure, were experienced in the sequencing reactions of any operational taxonomic unit (OTU), the PCR product was cloned using an Invitrogen life technologies TOPO TA Cloning kit (incorporating pCR 2.1-TOPO and TOP10F′ One Shot Chemically Competent cells) according to the manufacturer's instructions. If a heterozygote was suspected, five individual transformed One Shot colonies were sequenced using combinations of standard M13 forward and reverse primers to ensure that both alleles of the intron would be detected.

Preliminary Sequence Analyses and Alignments

Sequence chromatograms were read using Chromas (Technelysium Pty. Ltd, 1988–2000) and all mutations/indels were individually checked by eye at all sites across all taxa. Ambiguous peaks which may result from substitutional heterozygosity or errors in the sequencing process were recorded as Ns. In order to investigate a range of alignment strategies (Sanchis et al., 2001), five different alignment strategies were employed: (a) a "by eye" alignment and (b to e) four alignments performed by the distance-based program ClustalX 1.8 (Thompson et al., 1997), using a range of gap opening (GO: 20, 10, 5, and 2.5) and extension (GE: 10, 5, 1, 0.5) costs (Belshaw and Quicke, 2002). All ClustalX alignments were checked by eye and regions that were interpreted as being misaligned were amended manually.

LVHs

In order to include all the detected allelic size variants, all possible combinations of alleles from each specimen were included as independent terminals in the analyses (Sota and Vogler, 2003). Thus, if an individual was homozygotic for 7IβFIB and heterozygotic for 3ITBP, the heterozygotic individual would be represented by two identical 7IβFIB sequences and the two length variant 3ITBP sequences and vice versa. If, on the other hand, the individual was heterozygotic for both loci, all four combinations of the LVHs were included in the analyses.

Phylogenetic Analyses

Given that total evidence approaches to phylogenetic reconstruction have been shown to perform well in providing parsimonious resolutions of independent data sets (Flynn and Nedbal, 1998; Wenzel and Siddall, 1999;

Baker et al., 2001; Creer et al., 2003), the 7IβFIB and 3ITBP partitions were concatenated and analyzed simultaneously to maximize phylogenetic signal.

Contemporary molecular systematics often relies heavily on methods based on explicit models of sequence evolution, such as maximum likelihood (ML). However, readily applicable methods for incorporating indels have yet to be developed for model-based phylogenetic inference (Giribet and Wheeler, 1999; Kawakita et al., 2003), and ML and maximum parsimony (MP) searches and Bayesian approaches have always yielded congruent topologies within the *Trimeresurus* radiation (Malhotra and Thorpe, 2000, 2004). Thus, in the absence of a suitable likelihood-based search method, MP will be used here as the expedient phylogenetic approach. For completeness, however, Bayesian analyses were performed using MrBayes 3.0b4 (Huelsenbeck and Bollback, 2001; Huelsenbeck and Ronquist, 2001) on a number of alignments ("by eye" and edited ClustalX 20GO 10GE and 5GO 1GE) of the nucleotide portion of the intron data (indels treated as missing) to check for congruence between parsimony and MCMC approaches. In each analysis, a general time-reversible model was assumed (six substitution types, and a proportion of sites invariable with the rate for the remaining sites being drawn from a gamma distribution). Two independent runs each were performed of 500,000 generations. In order to ensure that searches had not become trapped in local optima, stationarity was determined for each run from plots of overall model likelihood scores against generation (Nylander et al., 2004). Consensus trees were then generated from the stationary phases of each run. The topologies of the consensus trees from each pair of runs were also compared to confirm the credibility of the resulting clades.

For the parsimony analyses, PAUP 4.0b10 (Swofford, 1998) was used as the core analytical program (with explicit assumptions replacing default parameters shown below). Six different approaches were used to analyze the substitutional and indel data generated by the "by eye" and the edited and unedited ClustalX 1.8 alignments:

1. Indels were treated as missing data.
2. As in 1, but the program GapCoder (Young and Healy, 2003) was used to generate a "simple indel coding" matrix as described by Simmons and Ochoterena (2000). Simple indel coding has the advantages of being conservative and easy to implement and has been shown to recover accurate phylogenies (Simmons and Ochoterena, 2000; Young and Healy, 2003).
   2a. Simple coding was assigned to the DNA data as in 2, but areas of ambiguous alignments that were deemed likely to violate positional homology (Lutzoni et al., 2000; Freudenstein and Chase, 2001) were excluded from the analyses.
3. This method was as in 2a, but the program INAASE 2.3b (INtegration of Ambiguously Aligned SEquences (Lutzoni et al., 2000) was used to generate multistate coded characters (step matrices, limited to 32 different character states) for the excluded areas of ambiguous alignments.
4. Indel positions were treated as a fifth character state.
5. As in 4, but areas of ambiguous alignment were excluded from the analyses.

For all PAUP analyses, MP searches were conducted using a random addition heuristic search (1000 replicates with tree-bisection-reconnection [TBR] branch swapping) with unlimited MaxTrees. For comparative purposes, strict consensus trees were subsequently generated for each analysis. A combination of bootstrap (100 replicates to reduce computational time) and Bremer (Bremer, 1994) support values (BS; performed using TreeRot v.2 [Sorenson, 1999]) were calculated to obtain relative measures of node support for the resulting trees (Felsenstein, 1985). Consistency indices (CI; Kluge and Farris, 1969) and retention indices (RI; Farris, 1989; Swofford, 1998) were generated to assess levels of homoplasy for all analyses.

Finally, the data were aligned and analyzed simultaneously using the program POY 3.0.11 (Wheeler and Gladstein, 2000) using the direct optimization method described in Wheeler (1996). The data were analyzed using a range of gap costs (1, 2, 4, 8) and the sequences were split into 1, 3, 5, and 10 unambiguously recognizable fragments in an attempt to reduce computational time and achieve better homology statements via successive data-partitioning steps (Giribet, 2001). In order to check the analytical process, the resulting alignments were checked by eye using the *–impliedalignment* function. As only one outgroup can be specified in POY, a *Triceratolepidophis sieversorum* (Ziegler et al., 2000) sequence, identified as an appropriate outgroup in the larger scale mitochondrial analyses in Malhotra and Thorpe (2004), was used to root all analyses.

In order to obtain a reference tree, the strict parsimony analysis presented in Malhotra and Thorpe (2004) was "pruned" to only include species that were used in the present study (Fig. 1). By pruning the tree, irrelevant OTUs were effectively removed from the data set, while leaving the actual robust phylogenetic hypothesis unmodified. Given that the 12S and 16S RNA data also contain indels, the pruned mtDNA data set was additionally reanalyzed with a selection of different alignment approaches (by eye and ClustalX 20GO, 10GE and 2.5GO, 0.5GE) and gap treatments (missing data and simple coded) to test for topological congruence with the phylogenetic hypothesis presented in Malhotra and Thorpe (2004). In the absence of uniform data sets (i.e., there were more taxa in the reference mtDNA tree), topological congruence (Giribet and Wheeler, 1999; Cognato and Vogler, 2001; Belshaw and Quicke, 2002) between the test tree and the reference tree was used to assess the performance of the diverse analytical methods on the different sequence alignments. The number of congruent nodes (with bootstrap support values $>50\%$ and $\geq 75\%$) out of a total of 23 were used to assess broad topological congruence and congruence between strongly supported

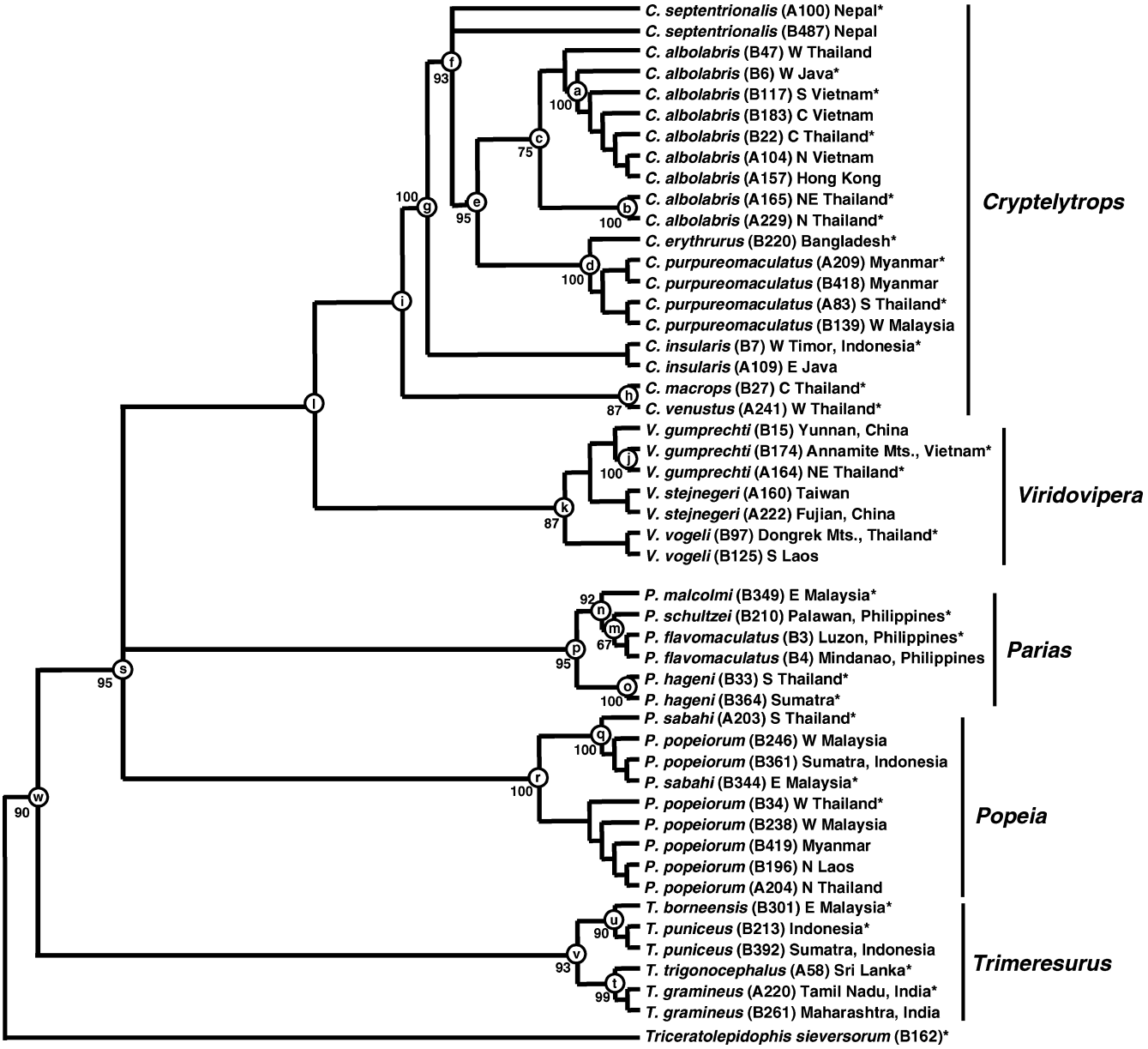

FIGURE 1. The strict consensus tree (based on a concatenated analysis of 2423 bp of 12S, 16S, ND4, and CYTB mtDNA gene partitions) presented in Malhotra and Thorpe (2004), which has been "pruned" to solely represent clades that include OTUs in the intron analyses. Bootstrap support values (>50%) are shown from 1000 replicates. Asterisks denote identical specimens (or conspecifics from geographically identical or close locations) that were included in the concatenated 7IβFIB and 3ITBP intron analyses. The lettered nodes correspond to potential monophyletic groupings that could be recovered from the intron analyses, which were used to assess topological congruence. For example, in order for node a to be congruent, the intron analyses would have to recover a monophyletic group comprising specimens B6, B22, and B117. Similarly, in order for node c to be congruent, the intron analyses would have to recover a monophyletic group comprising specimens B6, B117, B22, A165, and A229.

clades (Zharkikh and Li, 1992; Hillis and Bull, 1993), respectively. In order to fulfill the topological congruence criteria, both heterozygous alleles of the same OTU had to be included in the comparative monophyletic grouping.

Following the assignation of the most congruent tree, optimal sequence alignment, and analytical technique, base compositions were subjected to a chi-squared test of homogeneity across taxa and Fu and Li's *D* and *F* (Fu and Li, 1993) and Tajima's *D* (Tajima, 1989) test statistics were used to evaluate the possibility of non-neutral evolution. Saturation plots of pairwise differences due to (1) transitions and transversions, (2) transversions only, and (3) transitions only were plotted against the substitutional model of molecular evolution selected from distance analyses performed using the log likelihood function of ModelTest 3.06 (Posada and Crandall, 1998). Relative rates between the main lineages of the

ingroup were tested using a hierarchical approach as in Creer et al. (2003), implementing the two-cluster test of Takezaki et al. (1995) as implemented in PHYLTEST. In order to determine the level of phylogenetic signal related to each indel treatment approach, $g1$ statistics (Hillis and Huelsenbeck 1992) were calculated from the indel-only characters using the RandTrees function in PAUP. The level of indel-related homoplasy for each approach was also assessed by constraining parsimony searches (on the indel-only data), to retain trees only compatible with the intron nucleotide data.

Finally, the relevant alignment of the 7Iβ FIB and 3ITBP partition was concatenated with the 12S, 16S, ND4, and CYTB mtDNA data (using representative equivalent OTUs only) and analyzed using the identified optimal indel treatment approach. Minor indels in the mtDNA analyses of Malhotra and Thorpe (2004) were treated as fifth states (excluding 13 bp of areas of ambiguous alignment) and so replicate parsimony analyses were performed on the mtDNA data set (only containing OTUs represented by intron data) using both the optimal indel treatment approach and the treatment of indels as fifth states to check for topological congruence before final concatenation. For the total evidence approach, the heuristic search criteria were as above, but 1000 replicates were used to calculate bootstrap values and partitioned Bremer support (PBS; Baker and DeSalle, 1997) was performed using TreeRot v.2 (Sorenson, 1999) to assess the contribution of individual gene partitions to the overall support of the tree (Baker et al., 2001). For all analyses, two-tailed Wilcoxon signed-ranks tests (Templeton, 1983) were implemented in PAUP to test whether alternative constructed tree topologies were significantly different from the most parsimonious tree.

In order to critically appraise the phylogenetic utility of the indels themselves, each simple coded indel position (calculated by GapCoder) was assigned as either phylogenetically uninformative (i.e., present or absent in just one OTU) or phylogenetically informative. Indels were recorded that simply grouped monophyletic heterozygous alleles of the same OTU or appeared homoplasious due to obvious phylogenetic/taxonomic inconsistencies. Indels were also identified that could be unequivocally assigned as Hennigian according to the phylogenetic hypothesis generated by the total evidence analysis.

RESULTS

Sequence Analysis

In addition to the 7IβFIB (Creer et al., 2003), the 3ITBP conforms to GT (5′ donor, start) and AG (3′ acceptor, finish) consensus splice site rules and has a characteristic pyridimine-rich region preceding the 3′ splice site (Senapathy et al., 1990). Additionally, the concatenated intron partition displayed a distinctive A-T bias (30.3% A, 18.9% C, 17.1% G, and 33.7% T).

Of the OTUs that were cloned, 10 heterozygotes were discovered, with 9 showing heterogeneity in length (the Sumatran *P. hageni* alleles only exhibited substitutional differences). The LVHs differed by 1 to 15 bp at between 1 and 10 different sequence regions. Despite cloning PCR products that could not be sequenced directly, premature stops were still encountered. This resulted in eight taxa (*C. erythrurus* from Bangladesh and *C. septentrionalis* from Nepal, both *C. macrops* samples, *C. fasciatus*, *C. venustus*, *V. vogeli,* and *P. hageni* from Sumatra) with missing data approximately between positions 329 and 485 in the 7IβFIB partition. One region in each of the introns (23 bp after approximately 120 bp and 26 bp after approximately 192 bp in 7IβFIB and 3ITBP, respectively) could not be unequivocally aligned and so were excluded from the analyses and replaced with multistate coded characters according to the appropriate analytical methods. Unless otherwise stated, all sequence positions refer to the optimal alignment (see below).

Intron Phylogenetic Reconstructions

The Bayesian analyses resulted in tree topologies (not shown) that were congruent with parallel parsimony analyses, thus establishing confidence in the parsimony approach with these data. Table 1 provides summary data and tree statistics for a subset of the alignment strategies (by eye and edited and unedited ClustalX 20GO, 10GE and 5GO, 1GE) and all PAUP-based indel treatment methods. All alignments are additionally available at TreeBase www.treebase.org (study accession number S1345, matrix accession numbers M2367 to M2375). Due to the different alignments, concatenated 7IβFIB and 3ITBP intron sequence data partitions of between 1795 and 2165 characters were obtained for 31 OTUs containing between 178 and 655 parsimony informative sites. Numbers of equally parsimonious trees ranged from 1 to 2175, representing hypotheses with total lengths of between 472 and 1826. Within each alignment approach, the analysis of simple coded indels for whole sequences (method 2) plus the addition of multistate coded characters (method 4) produced trees that exhibited higher levels of homoplasy (as measured by the CI and RI) than other indel treatment methods. Conversely, the treatment of indels in whole sequences as missing data (method 1) and simple coded minus ambiguous regions (method 3) produced trees that exhibited the lowest levels of homoplasy. Trees resulting from the analysis of the different alignment and indel treatments of the pruned mtDNA data were all congruent with the topology presented in Malhotra and Thorpe (2004). Reference nodes (shown in Fig. 1) a to e, j to l, q, and r were not recovered in the analysis of any combination of alignment or indel treatment. Alternatively, nodes n and t were recovered from all analyses. The remaining nodes from the reference tree were represented to varying degrees throughout all analyses.

A wide range of variation in numbers of congruent nodes (with bootstrap support values $>50\%$ [between 3 and 10] and $\geq 75\%$ [between 3 and 8]) was observed throughout the strict consensus parsimony analyses. According to the topological congruence criteria, the

TABLE 1. Summary data for a selection of the different alignment methods ("by eye" and using ClustalX using a range of gap opening [20, 5] and extension [10, 1] costs for manually edited and unedited alignments, respectively) and indel analytical treatments (1 to 6: see Materials and Methods for details). The size (number of characters) refers to the length of the data matrix and PI shows the number of parsimony informative characters. Measures of homoplasy (CI and RI) are shown for equally parsimonious trees. N and tree length denote the number of equally parsimonious trees and respective tree lengths; Cong. nodes and nodes ≥75% refer to the number of congruent nodes with bootstrap support values >50% and ≥75%, respectively.

Alignment and analysis	Size	PI	CI	RI	N	Tree length	Cong. nodes	Nodes ≥75%
By eye—1	1866	178	0.672	0.827	15	472	7	6
By eye—2	1992	247	0.652	0.797	13	679	7	7
By eye—3	1888	215	0.670	0.821	6	579	7	7
By eye—4	1890	217	0.665	0.805	27	639	7	7
By eye—5	1866	597	0.551	0.803	6	1562	4	4
By eye—6	1812	556	0.585	0.818	516	1385	4	4
20_10—1	1832, 1843	204, 257	0.615, 0.652	0.800, 0.810	4, 2	553, 661	10, 8	7, 4
20_10—2	1937, 1961	268, 330	0.605, 0.621	0.778, 0.780	75, 58	735, 884	8, 7	7, 5
20_10—3	1870, 1894	229, 246	0.656, 0.661	0.802, 0.790	54, 58	654, 705	8, 7	7, 7
20_10—4	1872, 1896	230, 248	0.635, 0.658	0.784, 0.776	18, 15	677, 765	7, 8	6, 7
20_10—5	1832, 1843	560, 655	0.555, 0.564	0.806, 0.831	168, 1	1504, 1769	4, 4	3, 3
20_10—6	1795, 1807	548, 562	0.591, 0.597	0.826, 0.815	434, 76	1331, 1421	5, 5	4, 4
5_1—1	1850, 2031	184, 215	0.631, 0.699	0.811, 0.819	6, 167	501, 621	8, 7	6, 7
5_1—2	1963, 2165	256, 296	0.622, 0.661	0.790, 0.785	49, 14	690, 859	9, 9	8, 7
5_1—3	1886, 1967	216, 227	0.666, 0.672	0.819, 0.807	8, 4	574, 653	8, 9	7, 7
5_1—4	1888, 1969	218, 229	0.656, 0.666	0.801, 0.791	18, 331	636, 715	8, 8	8, 8
5_1—5	1850, 2031	559, 648	0.567, 0.609	0.813, 0.815	136, 90	1457, 1826	5, 6	5, 5
5_1—6	1805, 1864	529, 533	0.597, 0.638	0.830, 0.841	1311, 66	1302, 1337	5, 7	4, 5

"by eye" alignments produced trees that were notably poorer than those produced by the ClustalX-assisted alignments. Furthermore, the treatment of indels as fifth states (with and without the inclusion of the ambiguously aligned regions) produced significantly poorer trees compared to all other indel treatment methods. Of the remaining trees, minor inconsistencies were noted between the reference tree and the test trees for the following indel treatment strategies and alignments: indels as missing data (method 1) for the edited 20GO, 10GE; edited and unedited 5GO, 1GE; and the edited 2.5GO, 0.5GE ClustalX alignments and simple coded indels, including areas of ambiguous alignments (method 2) for the unedited 20GO, 10GE and 5GO, 1GE and all the edited ClustalX alignments. Of these, the most congruent tree (according to direct comparison to the reference tree) was recovered via analysis of the data set following simple coding of indels including areas of ambiguous alignments (method 2) for the 5GO, 1GE ClustalX-edited alignment (Fig. 2). The remaining trees all had major topological inconsistencies that made little or no taxonomic sense compared to the reference tree. Little (±1 node) or no difference was observed in 75% and 92% of the congruent nodes (with bootstrap support >50% and ≥ 75%, respectively) in the trees resulting from the analyses of the edited and unedited ClustalX alignments. Of the remaining nodes, 25% of the trees differed by between 1 and 3 nodes (more, and fewer, with bootstrap support >50%) and 8% of the unedited trees had up to 3 fewer congruent nodes only (with bootstrap support ≥75%), compared to the edited versions. Significant levels of phylogenetic signal were recorded from all treatments of the indel-only partitions ($g1$–simple coding −0.506, multistate coding −0.291, fifth state −0.262). Working from the ClustalX 5GO, 1GE alignment, the level of indel-related homoplasy (according to the constrained analyses) can be represented by fifth state (CI = 0.430, RI = 0.722) > simple coding (CI = 0.595, RI = 0.713) > multistate coding (CI = 0.631, RI = 0.688).

Working from the chosen optimal test tree (Fig. 2), it can be seen that alleles of 5 out of the 10 heterozygous OTUs formed monophyletic pairs. The alleles for the remaining five heterozygotes were distributed within or between the assigned generic groups. Incidental separate analyses of the individual introns (data not shown) showed that the monophyletic and paraphyletic groupings correspond to the actual LVH signal in the 7IβFIB and 3ITBP partitions, respectively. No significant differences were detected between the base frequency distributions; the Tajima and Fu and Li tests failed to reject the null hypothesis of neutral evolution and no saturation was detected in any of the plots of pairwise differences due to transitions, transversions, or both transitions and transversions against the transversional model (TVM) (Posada and Crandall, 1998) (PINVAR 0.558, gamma-shape parameter 0.735) selected by ModelTest. Furthermore, the relative rate tests failed to detect variation between any combination of the tested lineages.

POY produced biologically unrealistic trees and failed to align all sequences using gap costs of 1, 2, 4, and 8 on partitions demarcated into 1, 3, 5, and 10 unambiguously aligned regions (Giribet, 2001). Some misalignments may arise from the presence of small stretches of leading and trailing missing data commonly experienced in comparative phylogenetics. Therefore, further analyses were performed using the *–noleading* command and after all partitions were trimmed to remove all missing leading and trailing data, but problems still persisted. Finally, misalignments were still observed resulting from the analysis of a smaller data set only containing OTUs

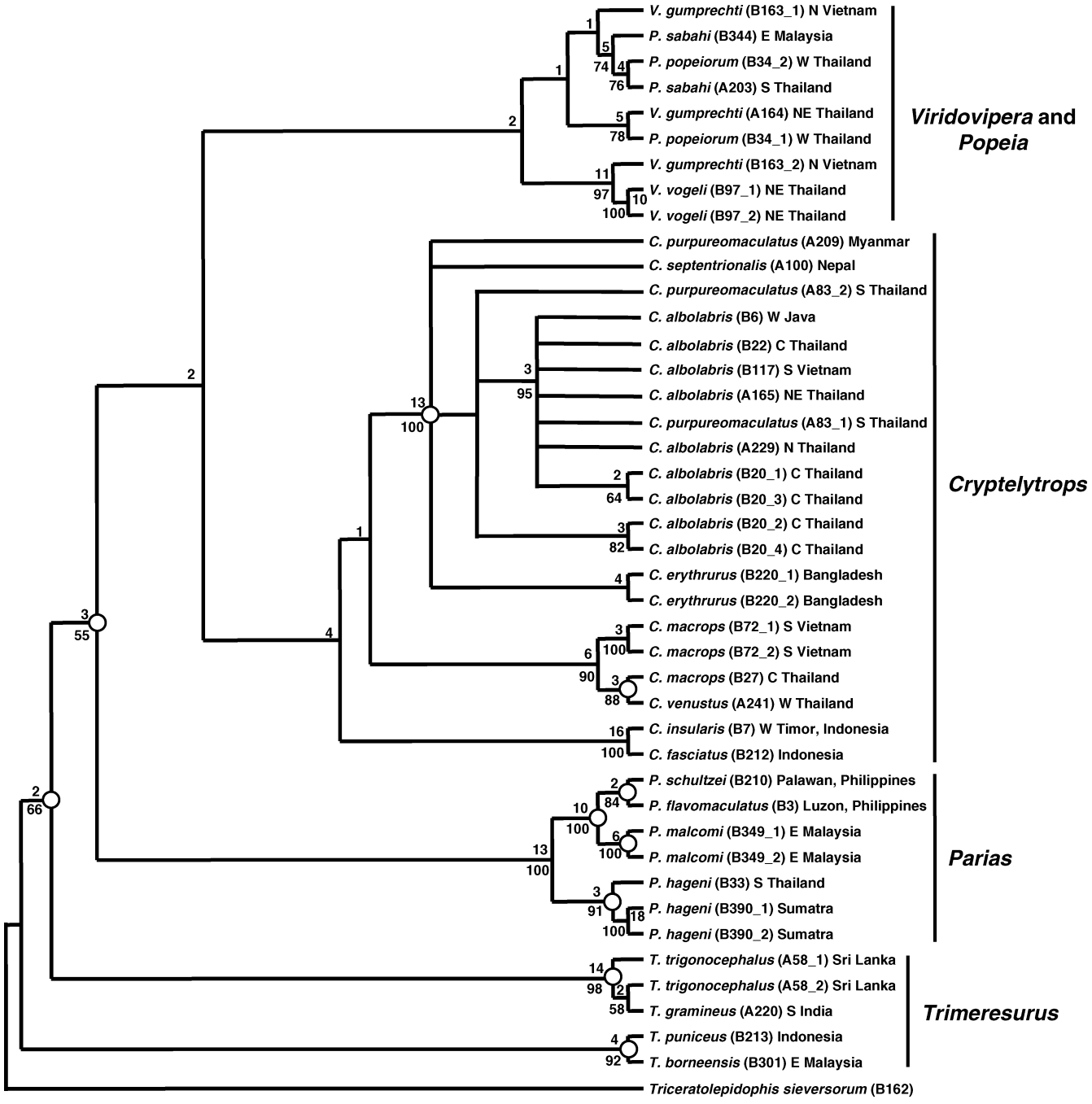

FIGURE 2. The "optimal" strict consensus tree based on the simple coded analysis of the 5GO, 1GE ClustalX alignment of the concatenated 7Iβ FIB and 3ITBP partitions (1963 characters). Bootstrap (>50%, from 1000 replicates) and Bremer support values are shown below and above nodes, respectively. Unfilled circles depict nodes that were congruent with the reference mtDNA tree in Figure 1. Suffixed specimen codes represent all possible combinations of heterozygotic alleles as described in Materials and Methods.

with uninterrupted sequence data and so no further analyses were performed using POY.

Total Evidence Phylogeny

Topologically identical trees were recovered from the analyses of the mtDNA data set (only containing OTUs represented by the intron data) treating indels as fifth states and via the addition of simple coded characters. The single most parsimonious total evidence tree (length 3053, CI 0.483, RI 0.709, RC 0.343, HI 0.517) resulting from the analysis of 894 PI sites from 4427 characters from four mtDNA partitions and the intron data can be seen in Figure 3. All of the recently assigned genera apart from *Trimeresurus s.s.* are represented as monophyletic groups in the total evidence tree.

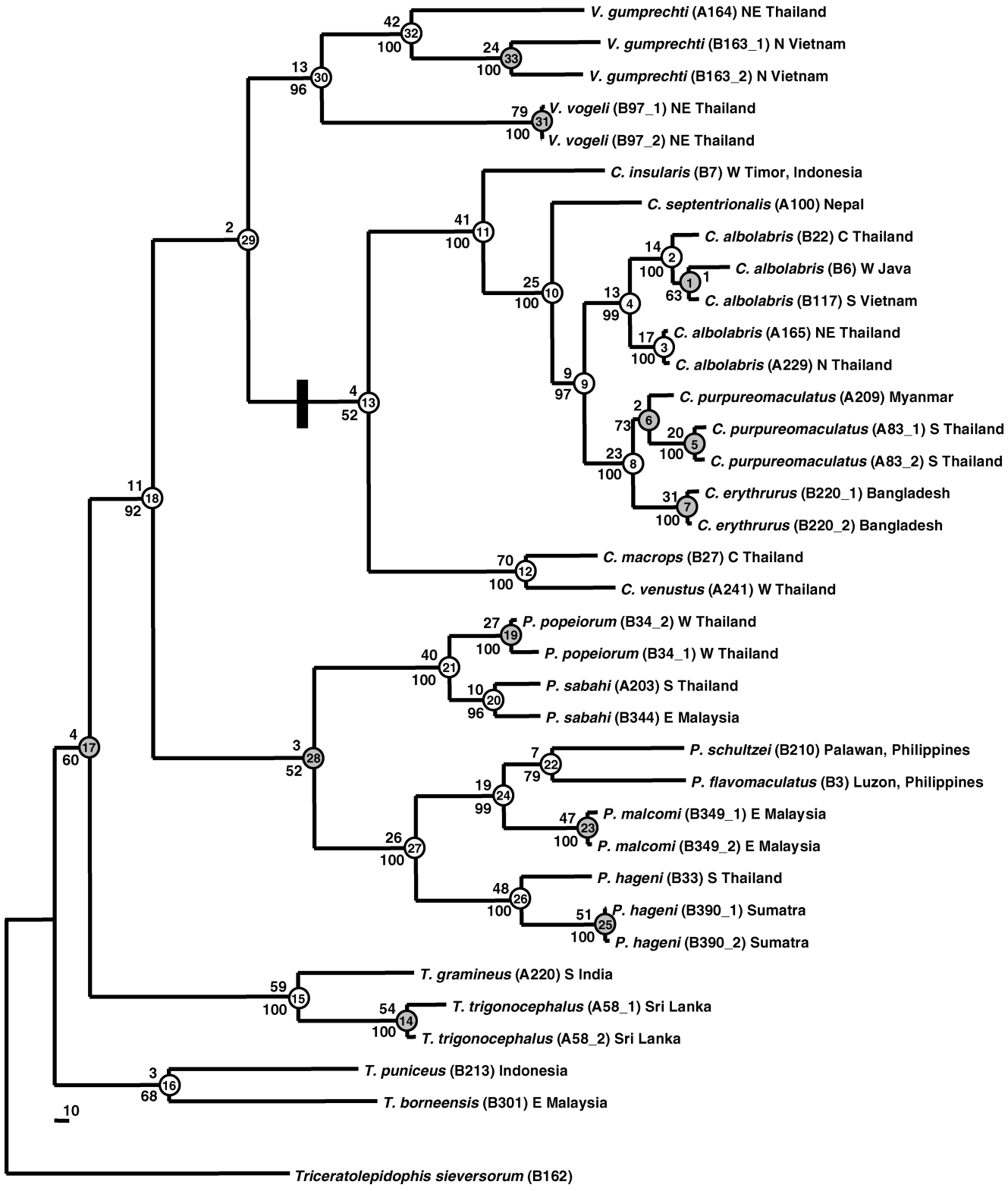

FIGURE 3. The single most parsimonious total evidence tree based on the simple coded, concatenated analysis of the 12S, 16S, ND4, and CYTB mtDNA and 7IβFIB and 3ITBP (ClustalX 5GO, 1GE) gene partitions (4427 characters). Bootstrap (>50%, from 1000 replicates) and Bremer support values are shown below and above nodes, respectively. The circled numbered nodes refer to nodes used in the calculation of the partitioned Bremer support values. The grey-filled circles depict nodes that were not congruent with the reference mtDNA tree in Figure 1. Suffixed specimen codes represent all possible combinations of heterozygotic alleles as described in Materials and Methods. The bold vertical line represents the position of the change from an unfused first supralabial and nasal scale to the fused condition. *Cryptelytrops* is the only genus to have a fused first supralabial and nasal scale and a long nonspiny hemipenis; *Viridovipera* species possess a type I spinose hemipenis; *Popeia* species possess a long calyculate hemipenis; *Parias* species possess a long papillose hemipenis; and *Trimeresurus* possess a type 2 spinose hemipenis (Malhotra and Thorpe, 2004).

TABLE 2. PBS for the total evidence parsimony tree. The PBS and total BS are shown for all gene partitions and associated simple coded characters for the 33 nodes shown in Figure 3. Where applicable, node numbers are suffixed by corresponding node letters represented in Figure 1. The suffix SC denotes the simple coded characters for the named gene partition.

Node no.	BS	7IβFIB	3ITBP	7IβFIB SC	3ITBP SC	12S	16S	ND4	CYTB	12S SC	16S SC
1	1	−3	1	0	0	0	0	0	3	0	0
2a	14	0	−0.5	0	0	0	2	2.5	9	1	0
3b	17	0	0	0	−1	0	2	10	6	0	0
4c	13	0	1	0	1	1	0	3	7	0	0
5	20	3	−1	1	−1	2	7	7	2	0	0
6	2	0	−1	0	0	0	0	2	1	0	0
7	31	4	1	2	0	0	0	8	14	1	1
8d	23	0	−4	0	−1	2	1	13	12	0	0
9e	9	0	0	1	−1	−1	0	3	7	0	0
10f	25	0	18	−1	2	0	0	1	4	0	1
11g	41	6	1	−2	1	5	5	10	13	1	1
12h	70	0	10	1	5	12	4	21	17	0	0
13I	4	6	0.7	0	1	2.3	−0.7	−4.7	−1.3	0	0.7
14	54	3	−9	4	0	1	9	21	23	0	2
15v	59	9	7	2	2	15	4	3	16	1	0
16u	3	0	1	1	1	0	−3	0	2	0	1
17t	4	2.5	0	1	0.5	4	0.5	−3.5	−1	0	0
18	11	5	−1	3	−1	4	0	2	−2	0	1
19	27	1	−6	0	1	3	4	10	13	1	0
20m	10	−4	0	−2	0	2	2	8	4	0	0
21n	40	6	0	1	1	10	4	10	8	0	0
22p	7	−1.5	0.5	1	1.5	−0.5	0.5	6	−0.5	0	0
23	47	2.5	1.5	−0.5	2.5	4	1.5	13.5	20	2	0
24q	19	11	2	−1	0	5	1	1	−2	0	2
25	51	10	2	1	5	0	2	12	16	2	1
26r	48	1.6	2.4	0.8	1.4	8.4	3.2	17.8	11.4	0	1
27s	26	−1	9	−1	4	2	6	0	7	0	0
28	3	−1	−0.7	−1	0.3	1.3	1.3	−1.7	4.3	0	0
29l	2	−1	1	0	1	1	1	−1	0	0	0
30k	13	9	−4	−1	−3	3	1	8	0	0	0
31	79	11	4	0	0	7	3	25	27	2	0
32j	42	3	−0.5	1	0	7	2	12.5	17	0	0
33	24	5.5	−3.5	−0.5	−0.5	3	0	4	14	2	0
Total	839	88	32	11	23	104	63	224	271	13	12

However, an analysis constrained to include *T. gramineus, T. trigonocephalus, T. puniceus*, and *T. borneensis* as a monophyletic group produced two trees of identical length that were not significantly longer than the most parsimonious tree (length 3057; $z = -0.555, -0.459$; $P = 0.579$, 0.646). The results from the PBS analysis are shown in Table 2 and show that the order of total support for the individual partitions (and the respective simple coded characters) in the simultaneous analysis can be represented by CYTB > ND4 > 12S > 7IβFIB > 16S > 3ITBP > 3ITBP—simple coded characters (SCC) > 12S SCC > 16S SCC > 7IβFIB SCC.

Phylogenetic Content of the Indel Data

Table 3 summarizes the putative phylogenetic content of all indels throughout the data partitions and Figure 4 shows the distribution of the indels throughout the 7IβFIB and 3ITBP. Between 56% and 64% of the indels were phylogenetically informative (PI) in the combined mtDNA and intron partitions, respectively. However, approximately 28% of the PI indels in the intron partitions only identified monophyletic alleles of the same OTU. Furthermore, only 33% and 19% of the PI indels in the combined mtDNA and intron partitions, respectively, could be assigned as Hennigian according to the total evidence phylogeny.

DISCUSSION

LVHs

The discovery that approximately one third of the OTUs exhibit a range of very diverse LVHs in the 7IβFIB and 3ITBP partitions is particularly relevant to molecular phylogenetic intron studies. These data and preliminary investigations with an independent intron locus (unpublished data) within the *Trimeresurus* radiation suggest that LVHs are common in intron loci. The reason for the discrepancy in the presence of LVHs

TABLE 3. A summary of the predicted phylogenetic signal of the indels throughout the mtDNA and the 7IβFIB and 3ITBP partitions. The total number and number of phylogenetically informative (PI) indels are shown in addition to indels that only identified monophyletic alleles of single OTUs and those that could be assigned as Hennigian.

	12S	16*S*	Total	7IβFIB	3ITBP	Total
No. indels	11	16	27	56	57	113
PI	6	6	12	25	16	41
Monophyletic alleles	—	—	—	12	8	20
Hennigian	2	3	5	5	9	14

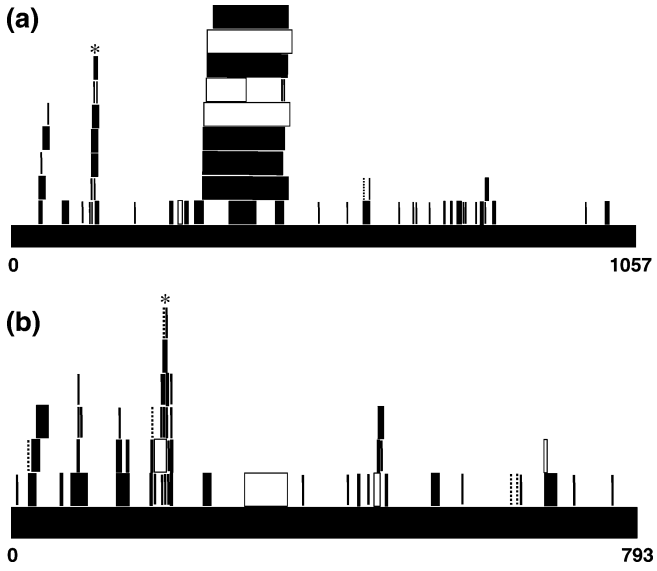

FIGURE 4. Schematic representation of the relative size and location of all indels (filled boxes) in (a) the 7IβFIB (1057 bp) and (b) the 3ITBP (793 bp). Indels of different sizes occurring in the same gene region are stacked vertically for representative purposes only. The areas of ambiguous alignment (regions characterized by T^nC^n repeats and the C/A region rich in 7IβFIB and 3ITBP, respectively) are denoted by asterisks. Putative Hennigian indels (according to comparison with the total evidence phylogenetic hypothesis [Figure 3]) are shown as broken lines (1 to 2 bp) and unfilled boxes (>3 bp).

between different intron data sets is not clear. LVHs may simply not be represented in some introns, or only homozygotic OTUs may have been incorporated in phylogenetic analyses in which LVHs are not included. It is clear, however, that LVHs should be incorporated into analytical frameworks in order to facilitate the detection of potential hybridization events and investigate the origin and mode of evolution of the relevant molecular marker (Palumbi and Baker, 1994; Sota and Vogler, 2003). The concatenation method, which combines different gene partitions to produce a composite OTU (Sota and Vogler, 2003), is an intuitive approach that accurately represents the biological nature of the data and reflects the phylogenetic contribution of the combined homozygous and heterozygous alleles. A concern may be that the duplication of the homozygous allele may inadvertently bias taxa towards monophyly, but parallel analyses showed that the concatenation approach reflected the monophyletic and paraphyletic nature of the individual intron partitions. With a limited number of alleles, the approach is manageable, but alternative strategies may be required if the number of composite OTUs become intractable.

In the present example, a possible explanation for the nonmonophyly of some of the LVHs is that coalescence has not yet occurred between alleles within ancestral taxa that are basal to derived paraphyletic clades. However, when the actual alleles are scrutinized, many LVHs indels appear homoplasious. This is illustrated by the non-monophyletic LVHs of the Vietnamese *V. gumprechti* (Figure 2). As the result of a trinucleotide GTT microsatellite repeat (positions 1509 to 1515 in the ClustalX 5GO, 1GE alignment), one allele of the Vietnamese *V. gumprechti* shares indels only with select OTUs from *Viridovipera*, *Cryptelytrops*, *Parias,* and the outgroup *T. sieversorum*. Similarly, the *P. popeiorum* LVHs shares a single-bp indel (position 1685 to 1685 in the above alignment) with many potentially closely related taxa, but also with the Sri Lankan *T. trigonocephalus* to the exclusion of all other *Trimeresurus* and *Parias*.

Indel Events

The homoplasious nature of a proportion of the indel data is further shown by the fact that only 33% and 19% (excluding indels which unite monophyletic LVHs) of the PI indels in the mtDNA and intron partitions respectively could be assigned to recognized taxonomic groupings. In the intron partitions, it is intuitive that microsatellite repeat regions (Beaumont and Bruford, 1999), or homopolymers (Beltrán et al., 2002), would be vulnerable to homoplasy, but there appears to be little consensus on the size of indels in either the ribosomal or intron partitions that unite biologically meaningful taxonomic groups in *Trimeresurus*. Indels that can be confidently assigned as homologous range from 1 to 5 bp in the ribosomal genes, 1 to 144 bp in the 7IβFIB, and between 1 and 52 bp in the 3ITBP. Doubt has been cast over the phylogenetic utility of smaller indels (between 1 and 30 bp), which may result from slipped-strand mispairing, compared to larger indels, which may have arisen through unequal crossing over or transposition (Giribet and Wheeler, 1999; Freudenstein and Chase, 2001). However, in the present example there appears to be no general rule which can be applied to identify the Hennigian or homoplasious nature of any of the indels. From the evidence provided, it appears that at least some indel positions are susceptible to reversible gain and loss events throughout evolutionary time.

The PBS values for the simple coded characters (Table 2) reflect the fact that there were more indels that could be assigned as Hennigian in the 3ITBP than the 7IβFIB (Table 3). However, the relative positions of indels within the introns give no indication of spatial phylogenetic utility as the putative Hennigian indels are randomly distributed throughout the data partitions (Figure 4). Some of the indel regions are obviously microsatellites, but both regions of ambiguous alignment close to the 5′ end of each intron (shown as stacked regions in Figure 4) are characterized by T^nC^n repeats and a C/A-rich region in 7IβFIB and 3ITBP, respectively. There is no immediate explanation for the T^nC^n repeats but it has recently been suggested that 5′ A/C rich regions are involved in the regulation of mRNA splicing (Hui et al., 2003). Interestingly, pitviper sequences that have entire or partial sequences within the very large indel in the 7IβFIB (Figure 4) also have two approximately 40-bp palindromic sequences immediately within the indel region (e.g., *T. borneensis* and the West Timor *C. insularis* between positions 333 and 366 and 431 and 468). It would seem intuitive that the latter region is involved

in the formation of secondary structure and errors in DNA replication (i.e., inadvertently cleaving the loop structure) and this may explain why the intervening region is missing in so many taxa. The latter also explains why sequencing reactions often terminated prematurely in exactly the same region as the large indels and the palindromes. Irrespective of the mode of molecular evolution that created the gaps in alignments, the indels represent an array of homoplasious and Hennigian characters and those which simply unite monophyletic alleles of LVHs. However, despite the high noise-to-signal ratio in the indel data set, analyses that omitted the indels were generally poorer (Table 1) than those which incorporated the data (including the optimal tree) as coded characters. The latter point obviously adds to the growing body of evidence that encourages the use of indel data, irrespective of noise (Wenzel and Siddall, 1999), in phylogenetic reconstruction (Giribet and Wheeler, 1999; Lutzoni et al., 2000; Simmons and Ochoterena, 2000; Freudenstein and Chase, 2001; Simmons et al., 2001; Kawakita et al., 2003; Miadlikowska et al., 2003).

Analytical Approaches

The considerable discrepancies between alignment lengths, number of parsimony informative sites, and number and topology of equally parsimonious trees demonstrates the importance of identifying appropriate alignment and analytical methods in indel-rich data sets. In general, the nodes that were not recovered from any of the alignment or analytical strategies either identify relatively derived clades (e.g., a–e/1–5 in Figs. 1 and 3 representing 0.87% mean uncorrected sequence divergence), clades that included nonmonophyletic LVHs (for the reasons given above), or are relationships that are associated that hard polytomies (e.g., *Viridovipera* and *Popeia*) (Creer et al., 2003; Malhotra and Thorpe, 2004). A comparison of the relative PBS values (Table 2) also highlights the discrepancy in support between the intron and mtDNA data partitions at the topologically incongruent nodes.

Although interspecific and even intergeneric gene flow has been documented for rattlesnakes (Klauber, 1972), there is no evidence to suggest that intergeneric hybridization occurs within *Trimeresurus* pitvipers. Thus, lineage sorting (Avise, 1994; Page and Holmes, 1998) is a more likely cause of absolute topological incongruence between the intron analyses and the reference mtDNA tree. Further to possessing different mutational mechanisms (Slade et al., 1994; Prychitko and Moore, 2000), it must be acknowledged that intron and mtDNA loci undergo very different coalescent processes as intron loci have a four times higher effective population size compared to mtDNA loci (Moore, 1995). Indeed, rates of evolution of at least the 7IβFIB have been shown to be between 0.18 and 0.26 times slower than comparable mtDNA partitions (Prychitko and Moore, 1997; Johnson and Clayton, 2000; Creer et al., 2003). Similarly, in the present analyses, the rate of evolution of the intron partition is 0.31 times that of the mtDNA partition, based on pairwise ingroup mean uncorrected sequence divergences of 3% and 9.6%, respectively. Consequently, unlike the mtDNA data (Moore, 1995), the intron data are less likely to have evolved at a rate that could track relatively rapid speciation events such as those occurring in the derived clades such as *Cryptelytrops*.

According to topological congruence with the reference mtDNA tree, a number of clear trends can be observed that may be relevant to data sets exhibiting a similar diverse array of indels (between 56 and 57 per intron, ranging from 1 and 151 bp in length). Trees resulting from the analysis of data aligned with the assistance of ClustalX 1.8 were generally more congruent than those produced by manual alignments, suggesting that manual alignments alone should be avoided. The analysis of manually edited and unedited ClustalX alignments had either little or no effect on the topologies of the large majority of trees. A limited number of trees did differ by more and less than up to three nodes (with bootstrap support $>50\%$), but only the analysis of unedited alignments resulted in trees with up to three fewer strongly supported congruent nodes (with bootstrap support $\geq 75\%$) compared to the edited alignments. The optimal tree also resulted from the analysis of an edited alignment and only analyses of unedited alignments failed to recover nodes f and o in Figure 1, which were recovered in all other analyses here and in all other analyses involving the *Trimeresurus* radiation (Malhotra and Thorpe, 2004). Although the manual editing of ClustalX alignments is routinely performed (Freudenstein and Chase, 2001; Sanchis et al., 2001; Kawakita et al., 2003), the practice remains a contentious issue regarding repeatability and subjectivity. The suitability of adjusting ClustalX alignments may be either more or less appropriate according to specific alignment parameters and individual data sets. However, in the current example, manual editing facilitated the resolution of the most strongly supported congruent topology.

The coding of indels as a fifth character state performed very poorly compared to all other indel treatments and displayed the highest levels of homoplasy in the indel-only constrained analyses. This is most probably due to the flawed assumption that indel events follow the same mutational mechanisms as nucleotide data. It is unlikely that 1–*n* bp length indel events represent 1–*n* independent mutations and it is more parsimonious to infer that multiposition gaps represent single events (Simmons and Ochoterena, 2000). Thus, by treating indel data as independent characters, larger indels will be significantly overweighted compared to smaller, equivalent indel events (Lutzoni et al., 2000; Freudenstein and Chase, 2001). The treatment of indels as fifth character states has been shown to produce identical tree topologies to trees generated by alternative indel treatments in the current mtDNA example and also in mixed data partitions (Hormiga et al., 2003). However, according to the above reasoning, it would appear inappropriate to apply fifth character state coding to intron data sets containing large and diverse indels.

Simple coding gaps (Simmons and Ochoterena, 2000; Simmons et al., 2001; Young and Healy, 2003) is a logical method of incorporating indels into phylogenetic analyses and the method performed well (including generating the most congruent tree) compared to other indel treatments. However, simple coding does not address the potential evolution of different sized indels resulting from nonindependent mutations. The natural progression from the latter point is to identify different character states within homologous regions and assign costs relative to putative mutational processes. Such an approach is implemented by INAASE 2.3b (Lutzoni et al., 2000), but simple coding areas of ambiguous alignments generally outperformed the assignation of multistate characters in the present analyses. The latter may be attributed to the fact that there was more Hennigian phylogenetic signal in the 31 simple coded characters (11 in the 7IβFIB and 20 in the 3ITBP) compared to the costs associated with the transformations between just two multistate characters. Indeed, the indel-only multistate coding constrained search yielded the lowest levels of homoplasy but resulted from the analysis of only 48 parsimony informative (PI) characters compared to 72 PI characters in the simple coding analysis.

Finally, there are no immediate explanations for the alignment difficulties experienced with POY. The direct optimization method implemented by POY has been widely used in molecular phylogenetics (Giribet, 2001; Sanchis et al., 2001; Braband et al., 2002; Hormiga et al., 2003; Sota and Vogler, 2003; Arnedo et al., 2004), but other authors have reported that POY has failed in alignments and recovering "test" clades (Belshaw and Quicke, 2002), or has produced alignments that appear gappy and biologically unrealistic (Cognato and Vogler, 2001). The size and diversity of the indels in any given data set may contribute to misalignments using the direct optimization method, and further usage and empirical testing on other data sets is recommended.

Phylogeny of Asian Pitvipers of the Trimeresurus Radiation

The five newly designated genera (*Cryptelytrops*, *Viridovipera*, *Popeia*, *Parias,* and *Trimeresurus*) were either recovered from strongly supported nodes (Figure 3) in the total evidence analysis (as in Malhotra and Thorpe, 2004) and/or recognized via combinations of diagnostic morphological characters (hemipenis type and fusion of the first upper labial scale, explained in Malhotra and Thorpe 2000, 2004). DeBry (2001) recently questioned if Bremer support values are comparable, even within a single tree, and so these support indices should be treated with a degree of caution. Nevertheless, combinations of bootstrap and Bremer support values are well-established parameters for estimating degrees of corroboration among competing topologies, and so trends in support values are still useful for estimating relative levels of support for phylogenetic hypotheses (Pons et al., 2004). Thus, node 11 (incorporating *C. insularis*, *C. septentrionalis*, *C. erythrurus*, *C. purpureomaculatus,* and *C. albolabris*), node 30 (*Viridovipera*), node 21 (*Popeia*), and node 27 (*Parias*) are all strongly supported. An analysis in which *Trimeresurus s.s.* (including *T. gramineus*, *T. trigonocephalus*, *T. puniceus,* and *T. borneensis*) was constrained as a monophyletic group produced two trees that were not significantly longer than the most parsimonious tree and so *Trimeresurus sensu* (Malhotra and Thorpe, 2004) can also be accepted (especially given the presence of a unique hemipenis type). Please note that "*Trimeresurus*"*gracilis* is actually no longer considered part of the *Trimeresurus* radiation (Malhotra and Thorpe, 2004) and is currently awaiting new generic allocation. There are, however, still poorly resolved regions of the phylogenetic hypothesis (e.g., nodes 13, 28, and 29 in Figure 3 that relate to the relationships between the genera). Thus, the analysis of 4427 characters from six mitochondrial and nuclear gene regions has failed to resolve these putative hard polytomies (Page and Holmes, 1998) that may have resulted from rapid speciation events. It is therefore unlikely that the addition of a limited number of base pairs from other sources would resolve the lack of strongly supported resolution. However, if more data were sought after, the PBS values strongly suggest that genes with a comparable rate of molecular evolution to the CYTB and ND4 partitions would have the greatest utility at the current taxonomic range (Creer et al., 2003). An alternative strategy may be the amplification and analysis of whole mitochondrial genomes (Kumazawa, 2004; Kumazawa and Endo, 2004) to further resolve the levels of support and phylogenetic affinities between genera and subclades (e.g., *C. macrops* and *C. venustus* within *Cryptelytrops*). For the time being, however, the current phylogenetic hypothesis and generic rearrangements associated with diagnostic morphological characters appears to be a realistic arrangement that at least provides a working benchmark in Asian pitviper systematics.

Intron Analysis Summary

The occurrence of LVHs (which can be easily included in multigene analytical frameworks) has been common to all independent introns tested to date and is the most common cause of the apparent failure of direct sequencing reactions (Palumbi and Baker, 1994). Critical analyses of the indels (in both LVHs and non-LVHs) suggest that they are a combination of homoplasious and homologous biological events. According to the reference mtDNA phylogeny (which in turn is supported by morphological characters), the analyses of ClustalX 1.8-assisted alignments were found to be more congruent than the analysis of "by eye" alignments. The treatment of indels as a fifth character state produced trees that were significantly poorer than all other treatment methods and direct optimisation using POY (Wheeler, 1996; Wheeler and Gladstein, 2000) appeared to be an inappropriate analytical tool for the present data. Furthermore, the analysis of simple coded indel characters (Simmons and Ochoterena, 2000; Simmons et al., 2001; Young and Healy, 2003) outperformed analyses where ambiguous regions of alignment had been

replaced with multistate coded characters (Lutzoni et al., 2000). Finally, the most congruent topology resulted from the analysis of simple coded indels generated from a manually edited ClustalX 1.8-assisted alignment (5GO, 1GE). Although not all characteristics of the present data will be universally relevant, a number of the above points will almost certainly be applicable to other intron data sets displaying similar frequencies (approximately every 19 and 14 bp in the 7IβFIB and 3ITBP, respectively), distributions and sizes (between 1 and 151 bp in length) of extensive and diverse indel events.

ACKNOWLEDGMENTS

This work was primarily funded by The Wellcome Trust (057257/Z/99/Z; 060384/Z/00/Z) and The Leverhulme Trust (F174/O) grants to RST and AM. We are grateful to the large numbers of people who have assisted us in the field, or supplied us with tissue samples for analysis. These include T. Chanard, J. Nabhitabhata (National Science Museum of Thailand), L. Chanhome (Queen Savoabha Memorial Institute, Thailand), K. Thirakhupt and P. P. Van Dijk (Chulalongkorn University, Thailand), M. J. Cox, J. Murray, T.-X. Kiem (Cho-Ray Hospital, Vietnam), B. L. Stuart (Field Museum of Natural History and Wildlife Conservation Society), M. Toriba (Japan Snake Institute), A. de Silva (Peridinaya University, Sri Lanka), S. Anuar (Universiti Sains Malaysia, Penang), U. Kuch (University of Frankfurt), M. Guillod and V. Morier (Ophiofarm), R. How (University of Western Australia), and K. L. Sanders (University of Wales, Bangor). We gratefully acknowledge the National Science Council of Thailand, Perhelitan, Malaysia, and the Ministry of Health, Vietnam, for permission to carry out fieldwork. Further thanks go to M. Lee, M. Simmons, T. Castoe, and an anonymous reviewer for valuable suggestions which greatly improved an earlier version of this manuscript.

REFERENCES

Alder, H., M. Yoshinouchi, M. B. Prystowsky, P. Appasamy, and R. Baserga. 1992. A conserved region in intron 1 negatively regulates the expression of the PCNA gene. Nucl. Acids Res. 20:1769–1775.

Arnedo, M. A., J. Coddington, I. Agnarsson, and R. G. Gillespie. 2004. From a comb to a tree: Phylogenetic relationships of the comb-footed spiders (Araneae, Theridiidae) inferred from nuclear and mitochondrial genes. Mol. Phyl. Evol. 31:225–245.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, New York. Pages 126–138.

Baker, R. H., G. S. Wilkinson, and R. DeSalle. 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). Syst. Biol. 50:87–105.

Beaumont, M. A., and M. W. Bruford. 1999. Microsatellites in conservation genetics. Pages 165–182 *in* Microsatellites. Evolution and applications (D. B. Goldstein and C. Schlotterer, eds.). Oxford University Press, Oxford, UK.

Belshaw, R., and D. L. J. Quicke. 2002. Robustness of ancestral state estimates: Evolution of life history strategy in ichneumonoid parasitoids. Syst. Biol. 51:450–477.

Beltrán, M., 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.

Birks, S. M., and S. V. Edwards. 2002. A phylogeny of the megapodes (Aves: Megapodiidae) based on nuclear and mitochondrial DNA sequences. Mol. Phyl. Evol. 23:408–421.

Braband, A., S. Richter, R. Hiesel, and G. Schlotte. 2002. Phylogenetic relationships within the Phyllopoda (Crustacea, Branchiopoda) based on mitochondrial and nuclear markers. Mol. Phyl. Evol. 25:229–244.

Bremer, K. 1994. Branch support and tree stability. Cladistics. 10:295–304.

Cognato, A. I., and A. P. Vogler. 2001. Exploring data interaction and nucleotide alignment in a multiple gene analysis of *Ips* (Coleoptera: Scolytinae). Syst. Biol. 50:758–780.

Creer, S., A. Malhotra, and R. S. Thorpe. 2003. Assessing the phylogenetic utility of four mitochondrial genes and a nuclear intron in the Asian pit viper genus *Trimeresurus*: Separate, simultaneous, and conditional data combination analyses. Mol. Biol. Evol. 20:1240–1251.

David, P., and I. Ineich. 1999. Les serpentes venimeux du monde: Systématique et répartition. Dumerilia (Paris) 3. Pages 3–499.

DeBry, R. W. 2001. Improving interpretation of the decay index for DNA sequence data. Syst. Biol. 50:744–752.

Dowling, T. E., C. Moritz, J. D. Palmer, and L. M. Rieseberg. 1996. Nucleic acids III: Analysis of fragments and restriction sites. Pages 249–320 *in* Molecular systematics, 2nd edition (D. M. Hillis, B. K. Mable, and C. Moritz, eds.). Sinauer Associates, Sunderland, Massachusetts.

Ericson, P. G. P., L. Christidis, M. Irestedt, and J. A. Norman. 2002. Systematic affinities of the lyrebirds (Passeriformes: Menura), with a novel classification of the major groups of passerine birds. Mol. Phyl. Evol. 25:53–62.

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

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

Flynn, J. J., and M. A. Nedbal. 1998. Phylogeny of the Carnivora (Mammalia): Congruence vs. incompatability among multiple data sets. Mol. Phyl. Evol. 9:414–426.

Freudenstein, J. V., and M. W. Chase. 2001. Analysis of mitochondrial *nad1* b-c intron sequences in Orchidaceae: Utility and coding of length-change characters. Syst. Bot. 26:643–657.

Friesen, V. 2000. Introns. Pages 274–294 *in* Molecular methods in ecology (A. J. Baker, ed.). Blackwell Science, Oxford, UK.

Friesen, V. L., B. C. Congdon, H. E. Walsh, and T. P. Birt. 1997. Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. Mol. Ecol. 6:1047–1058.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Gasch, A., U. Hinz, and R. Renkawitz-Pohl. 1989. Intron and upstream sequences regulate expression of the Drosophila ß3-tubulin gene in the visceral and somatic musculature, respectively. Proc. Natl. Acad. Sci. USA 86:3215–3218.

Giribet, G. 2001. Exploring the behavior of POY, a program for direct optimisation of molecular data. Cladistics. 17:S60–S70.

Giribet, G., and W. C. Wheeler. 1999. On gaps. Mol. Phyl. Evol. 13:132–143.

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.

Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Hered. 83:189–195.

Hormiga, G., M. Arnedo, and R. G. Gillespie. 2003. Speciation on a conveyor belt: Sequential colonisation of the Hawaiian islands by *Orsonwelles* spiders (Araneae, Linyphiidae). Syst. Biol. 52:70–88.

Huelsenbeck, J. P., and J. P. Bollback. 2001. Empirical and hierarchical Bayesian estimation of ancestral states. Syst. Biol. 50:351–366.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes—Bayesian inference of phylogeny, Version 3.0b4. Bioinformatics 17:754–755.

Hui, J., K. Stangl, W. S. Lane, and A. Bindereif. 2003. HnRNP L stimulates slicing of the *eNOS* gene by binding to variable-length CA repeats. Nat. Struct. Biol. 10:33–37.

Jenkins, T. M., R. E. Dean, R. Verkerk, and B. T. Forschler. 2001. Phylogenetic analyses of two mitochondrial genes and one nuclear intron region illuminate European subterranean termite (Isoptera: Rhinotermitidae) gene flow, taxonomy, and introduction dynamics. Mol. Phyl. Evol. 20:286–293.

Johnson, K. P., and D. H. Clayton. 2000. Nuclear and mitochondrial genes contain similar phylogenetic signal for pigeons and doves (Aves: Columbiformes). Mol. Phyl. Evol. 14:141–151.

Kawakita, A., T. Sota, J. S. Ascher, M. Ito, H. Tanaka, and M. Kato. 2003. Evolution and phylogenetic utility of alignment gaps within intron sequences of three nuclear genes in bumble bees (*Bombus*). Mol. Biol. Evol. 20:87–92.

Kirby, D. A., S. V. Muse, and W. Stephan. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA 92:9047–9051.

Klauber, L. M. 1972. Rattlesnakes, their habits, life histories and influence on mankind. University of California Press, Berkeley, California. Pages 239–248.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis for relationships among *Epicrates* (Boidae, Serpentes). Sys. Zool. 38:7–25.

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

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular evolutionary genetics analysis, version 1.01. Pennsylvania State University.

Kumazawa, Y. 2004. Mitochondrial DNA sequences of five squarnates: Phylogenetic affiliation of snakes. DNA Res. 11:137–144.

Kumazawa, Y., and H. Endo. 2004. Mitochondrial genome of the Komodo dragon: Efficient sequencing method with reptile-oriented primers and novel gene rearrangements. DNA Res. 11:115–125.

Leicht, B. G., S. V. Muse, M. Hanczyc, and A. G. Clark. 1995. Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila*. Genetics 139:299–308.

Lessa, E. P. 1992. Rapid surveying of DNA sequence variation in natural populations. Mol. Biol. Evol. 9:323–330.

Lutzoni, F., P. Wagner, V. Reeb, and S. Zoller. 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. 49:628–651.

Malhotra, A., and R. S. Thorpe. 2000. A phylogeny of the *Trimeresurus* group of pit vipers: New evidence from a mitochondrial gene tree. Mol. Phyl. Evol. 16:199–211.

Malhotra, A., and R. S. Thorpe. 2004. A phylogeny of four mitochondrial gene regions suggests a revised taxonomy for Asian pitvipers (*Trimeresurus* and *Ovophis*). Mol. Phyl. Evol. 32:83–100.

McDiarmid, R. W., J. A. Campbell, and T. S. A. Touré. 1999. Snake species of the world. A taxonomic and geographic reference. The Herpetologists League, Washington DC. Pages 344–345.

Miadlikowska, J., F. Lutzoni, T. Goward, S. Zoller, and D. Posada. 2003. New approach to an old problem: Incorporating signal from gap-rich regions of ITS and rDNA large subunit into phylogenetic analyses to resolve the *Peltigera canina* species complex. Mycologica 95:1181–1203.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Nakashima, K., I. Nobuhisa, M. Deshimaru, T. Ogawa, Y. Shimohigashi, Y. Fukumaki, M. Hattori, Y. Sakaki, S. Hattori, and M. Ohno. 1995. Structures of genes encoding TATA box-binding proteins from *Trimeresurus gramineus* and *T. flavoviridis* snakes. Gene 152:209–213.

Needleman, S. B., and C. D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443–453.

Nixon, K. C., and J. M. Carpenter. 1996. On simultaneous analysis. Cladistics 12:221–242.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Oppen, M. J. H. v., B. L. Willis, H. W. J. A. v. Vugt, and D. J. Miller. 2000. Examination of species boundaries in the *Acropora cervicornis* group (Scleractina, Cnidaria) using nuclear DNA sequence analyses. Mol. Ecol. 9:1363–1373.

Page, R. D. M., and E. C. Holmes. 1998. Molecular evolution. A phylogenetic approach. Blackwell Science, Oxford, UK. Pages 13–142.

Palumbi, S. R., and C. S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol. Biol. Evol. 11:426–435.

Phillips, A., D. Janies, and W. Wheeler. 2000. Multiple sequence alignment in phylogenetic analysis. Mol. Phyl. Evol. 16:317–330.

Pitra, C., D. Lieckfeldt, and J. C. Alonso. 2000. Population subdivision in Europe's great bustard inferred from mitochondrial and nuclear DNA sequence variation. Mol. Ecol. 9:1165–1170.

Pons, J., T. G. Barraclough, K. Theodorides, A. Cardoso, and A. P. Vogler. 2004. Using exon and intron sequences of the gene Mp20 to resolve basal relationships in *Cicindela* (Coleoptera: Cicindelidae). Syst. Biol. 53:554–570.

Posada, D., and K. A. Crandall. 1998. ModelTest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Prychitko, T. M., and W. S. Moore. 1997. The utility of DNA sequences of an intron from the ß-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves:Picidae). Mol. Phyl. Evol. 8:193–204.

Prychitko, T. M., and W. S. Moore. 2000. Comparative evolution of the mitochondrial cytochrome bgene and nuclear ß-fibrinogen intron 7 in woodpeckers. Mol. Biol. Evol. 17:1101–1111.

Prychitko, T. M., and W. S. Moore. 2003. Alignment and phylogenetic analysis of ß-fibrinogen intron 7 sequences among avian orders reveal conserved regions withing the intron. Mol. Biol. Evol. 20:762–771.

Rockman, M. V., D. M. Rowell, and N. N. Tait. 2001. Phylogenetics of *Planipapillus*, lawn-headed Onychophorans of the Australian Alps, based on nuclear and mitochondrial gene sequences. Mol. Phyl. Evol. 21:103–116.

Rosen, S., and H. J. Skaletsky. 1998. Primer3. Code available at http://www.genome.wi.mit.edu/genome software/other/primer3.html.

Rowe, D. L., and R. L. Honeycutt. 2002. Phylogenetic relationships, ecological correlates, and molecular evolution within the Cavioidea (Mammalia, Rodentia). Mol. Biol. Evol. 19:263–277.

Sanchis, A., J. M. Michelena, A. Latorre, D. L. J. Quicke, U. Gärdenfors, and R. Belshaw. 2001. The phylogenetic analysis of variable-length sequence data: Elongation factor-1α introns in European populations of the parasitoid wasp genus *Pauesia* (Hymenoptera: Braconidae: Aphidiinae). Mol. Biol. Evol. 18:1117–1131.

Senapathy, P., M. B. Shapiro, and N. L. Harris. 1990. Splice junctions, branch point sites, and exons: Sequence statistics, identification, and application to genome project. Methods Enzymal 183:252–278.

Simmons, M. P. 2004. Independence of alignment and tree search. Mol. Phyl. Evol. 31:874–879.

Simmons, M. P., and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49:369–381.

Simmons, M. P., H. Ochoterena, and T. G. Carr. 2001. Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. Syst. Biol. 50:454–462.

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.

Slade, R. W., C. Moritz, A. Heidemann, and P. T. Hale. 1993. Rapid assessment of single-copy nuclear DNA variation in diverse species. Mol. Ecol. 2:359–373.

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

Sota, T., and A. P. Vogler. 2003. Reconstructing species phylogeny of the carabid beetles *Ohomopterus* using multiple nuclear DNA sequences: Heterogenous information content and the performance of simultaneous analyses. Mol. Phyl. Evol. 26:139–154.

Swofford, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0. Sinauer Associates, Sunderland, Massachusests.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pages 407–446 *in* Molecular systematics, 2nd edition (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer Associates, Sunderland, Massachusetts.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearised trees. Mol. Biol. Evol. 12:823–833.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution 37:221–244.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res. 24:4876–4882.

Wenzel, J. W., and M. E. Siddall. 1999. Noise. Cladistics 15:51–64.

Wheeler, W. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? Cladistics 12:1–9.

Wheeler, W. C., and D. Gladstein. 2000. POY: The optimisation of alignment characters. American Museum of Ntaural History, New

York. Program and documentation available at ftp://ftp.amnh.org/pub/molecular/poy/.
Young, N. D., and J. Healy. 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinformatics 4:6. http://www.biomedcentral.com/1471-2105/4/6.
Zharkikh, A., and W. H. Li. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences I. Four taxa with a molecular clock. Mol. Biol. Evol. 9:1119–1147.
Ziegler, T., H.-W. Herrmann, P. David, N. L. Orlov, and O. S. G. Pauwels. 2000. *Triceratolepidophis sieversorum*, a new genus and species of pitviper (Reptilia: Serpentes: Viperidae: Crotalinae) from Vietnam. Russn. J. Herpetol. 7:199–214.

First submitted 15 February 2005; reviews returned 18 May 2005; final acceptance 30 August 2005
Associate Editor: Mike Lee

APPENDIX 1. Species, specimen code (AM refers to author's catalogue number), known locality, and GenBank accession numbers for the 7IβFIB and 3ITBP intron gene partitions. *Conspecifics from the same geographic locality/region that were nonidentical to the OTUs in the Malhotra and Thorpe (2004) study. Locality prefixes are N, north; S, south; E, east; W, west; and C, central. Two GenBank accession numbers for the intron partitions represent the two alleles of heterygotic individuals.

			GenBank accession no.	
Species	Specimen code	Locality	7IβFIB	3ITBP
Cryptelytrops albolabris	AM B6	W Java	AF517209	DQ117521
Cryptelytrops albolabris	AM B117	S Vietnam	AF517202	DQ117519
Cryptelytrops albolabris	AM B22	C Thailand	AF517205	DQ117520
Cryptelytrops albolabris	AM A165	NE Thailand	AF517196	DQ117522
Cryptelytrops albolabris	AM A229	N Thailand	DQ116985	DQ117507
Cryptelytrops albolabris	AM B20	S Thailand	DQ116965, DQ116966	DQ117487, DQ117488
Cryptelytrops erythrurus	AM B220	Bangladesh	DQ116984, DQ116986	DQ117506, DQ117508
Cryptelytrops purpureomaculatus	AM A209	S Myanmar	AF517210	DQ117518
Cryptelytrops purpureomaculatus	AM A83	S Thailand	DQ116967, DQ116968	DQ117489, DQ117490
Cryptelytrops septentrionalis	AM A100	Nepal	AF517194	DQ117517
Cryptelytrops insularis	AM B7	Indonesia	DQ116978	DQ117500
Cryptelytrops macrops	AM B27	C Thailand	AF517206	DQ117523
Cryptelytrops macrops	AM B72	S Vietnam	DQ116987, DQ116988	DQ117509, DQ117510
Cryptelytrops venustus	AM A241	S Thailand	DQ116994	DQ117516
Cryptelytrops fasciatus	AM B212	Tanadjampea Isl., Indonesia	DQ116993	DQ117515
Viridovipera gumprechti	AM B163*	N Vietnam	DQ116962, DQ116963	DQ117484, DQ117485
Viridovipera gumprechti	AM A164	NE Thailand	DQ116964	DQ117486
Viridovipera vogeli	AM B97	NE Thailand	DQ116989, DQ116990	DQ117511, DQ117512
Popeia sabahi	AM A203	S Thailand	DQ116975	DQ117497
Popeia sabahi	AM B344	E Malaysia	DQ116976	DQ117498
Popeia popeiorum	AM B34	W Thailand	DQ116977, DQ116974	DQ117499, DQ117496
Parias malcolmi	AM B349	Sabah, E Malaysia	DQ116982, DQ116983	DQ117504, DQ117505
Parias schultzei	AM B210	Palawan, Philippines	DQ116979	DQ117501
Parias flavomaculatus	AM B3	Luzon, Philippines	DQ116980	DQ117502
Parias hageni	AM B33	S Thailand	DQ116981	DQ117503
Parias hageni	AM B390*	Sumatra, Indonesia	DQ116991, DQ116992	DQ117513, DQ117514
Trimeresurus borneensis	AM B301	Sabah. E Malaysia	DQ116973	DQ117495
Trimeresurus puniceus	AM B213	Indonesia	AF517212	DQ117524
Trimeresurus trigonocephalus	AM A58	Sri Lanka	DQ116969, DQ116970	DQ117491, DQ117492
Trimeresurus gramineus	AM A220	S India	DQ116972	DQ117494
Triceratolepidophis sieversorum	AM B162	Laos	DQ116971	DQ117493