ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Molecular Phylogenetics and Evolution 31 (2004) 1031–1040

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Turtle phylogeny: insights from a novel nuclear intron

Matthew K. Fujita,[a,1] Tag N. Engstrom,[b,2] David E. Starkey,[a] and H. Bradley Shaffer[b,*]

[a] *Section of Evolution and Ecology, University of California, 1 Shields Avenue, Davis, CA 95616, USA*
[b] *Section of Evolution and Ecology, Center for Population Biology, University of California, 1 Shields Avenue, Davis, CA 95616, USA*

Received 28 May 2003; revised 9 September 2003

Abstract

Introns have gained considerable popularity as markers for molecular phylogenetics. However, no primers exist for a nuclear intron that amplifies across all turtles. Available data from morphology and mitochondrial DNA have not unambiguously resolved relationships within the superfamily Trionychoidea and the family Chelidae, which together form a large portion of extant turtle diversity. We tested the phylogenetic utility of a novel intron from the RNA fingerprint protein 35 (R35) as applied to these two areas of turtle systematics. We found the intron to be a single-copy locus that provides excellent resolving power for lineages among turtles, though problems with alignment made it impossible to infer deeper amniote relationships. Maximum parsimony and maximum likelihood both demonstrated the polyphyly of Trionychoidea and the reciprocal monophyly of Australian/New Guinea and South American chelid turtles. This is the first study to resolve such relationships with strong statistical support, and we suggest that R35 holds great promise for resolving additional persistent problems in the phylogeny of living turtles.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Nuclear phylogenetics; Intron; RNA fingerprint protein 35; Turtle phylogeny; Chelidae; Trionychoidea

1. Introduction

The majority of phylogenetic studies today rely on DNA sequence data to infer evolutionary history: it holds vast amounts of information and has a biologically universal presence (Nei and Kumar, 2000). In addition, the diverse nature of DNA allows investigators to choose from a suite of markers that best suit their study and taxa. The mitochondrial genome has several attributes, including a small effective population size with respect to the nuclear genome [and therefore on average a shorter coalescent time, albeit with a very high variance (Hudson and Turelli, 2003)], lack of significant recombination due to its uniparental inheritance (Avise, 1994; Moore, 1995; Sunnucks, 2000), and amenability to the polymerase chain reaction (PCR) as a consequence of the availability of universal primers (Kocher et al., 1989), that have made it an attractive phylogenetic tool for the past twenty years. It also offers a mix of fast-evolving genes, useful for studies of recently diverged lineages, and slow-evolving genes that reduce effects of homoplasy in studies of more ancient divergences. However, the inheritance of the mitochondrial genome as a single linkage group, the frequent demonstration of hybridization and natural selection, and lineage sorting often require additional nuclear data to test mitochondrial DNA (mtDNA)-based phylogenetic hypotheses. In addition, a strong base compositional bias and rapid rate of evolution may hinder mtDNA-based phylogenetic analyses as a result of homoplasy (Garcia-Machado et al., 1999; Naylor and Brown, 1998; Wiens and Hollingsworth, 2000). Thus, while mtDNA has provided and will continue to provide invaluable utility to phylogenetics, it is important to identify independent markers that can complement those of the mitochondrial genome.

The nuclear genome provides an incredible diversity of markers available to evolutionary biologists. There are coding and non-coding sequences that evolve at

* Corresponding author. Fax: 1-530-752-1449.
E-mail address: hbshaffer@ucdavis.edu (H.B. Shaffer).
[1] Present address: Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA.
[2] Present address: Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794-5245, USA.

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2003.09.016

different rates, allowing investigators to make a broad range of inferences from intrapopulation dynamics to the tree of life based on markers from independent linkage groups (Barns et al., 1996; Sunnucks, 2000). Despite these advantages of employing nuclear markers in phylogenetic studies, their use remains sparse when compared to mtDNA. Problems associated with nuclear gene families (i.e., the orthology problem; Sanderson and Shaffer, 2002), lack of universal PCR primers, and time-consuming optimization remain major roadblocks to the widespread use of nuclear sequences. Particularly for non-coding introns, markers designed for one taxon may or may not work in others, resulting in nuclear loci available for phylogenetic analyses with limited taxonomic applicability.

To date, the only nuclear markers available for turtle phylogenetics include coding sequences from only a few genes. The recombination activating gene RAG-1 consists of a single 3.1-kb exon that successfully amplifies via the PCR in many vertebrates including in birds, crocodilians, salamanders, and turtles (i.e., Groth and Barrowclough, 1999; Krenz et al., in preparation). While the phylogenetic utility of this gene has proven useful, it exhibits typically low levels of exon variation. Often, this translates to sequencing the entire gene for adequate phylogenetic signal for studies of recent to intermediate divergences. The oncogene C-*mos* is an approximately 1000 bp intronless gene that has been applied to higher-level reptile phylogenetics (Saint et al., 1998) including chelid turtles (Georges et al., 1998). However, the phylogenetic utility of C-*mos* is somewhat limited because the segment currently used in phylogenetics is less than 400 bp long (Georges et al., 1998; Saint et al., 1998). In addition, there is evidence that codon usage bias in C-*mos* may distort phylogenetic signal (Harris, 2003).

Nuclear introns provide an alternative to using coding sequence as genetic markers. As they are relatively free from many of the functional constraints of coding regions, introns tend to have an elevated rate of evolution when compared to coding sequence (Graur and Li, 2000). In addition, introns tend to have uniform and even base composition and unbiased substitution patterns, resulting in minimal among-site rate heterogeneity and a relatively low transition–transversion ratio (Armstrong et al., 2001; Birks and Edwards, 2002; Engstrom et al., submitted). The recent popularity of introns in phylogenetic studies has demonstrated their utility in a variety of phylogenetic analyses of plants and animals (Armstrong et al., 2001; Baker et al., 2001; Birks and Edwards, 2002; DeBry and Seshadri, 2001; Howarth and Baum, 2002; Johnson and Clayton, 2000; Prychitko and Moore, 1997; Weibel and Moore, 2002). However, few introns are available for non-avian reptilian phylogenetics, and none exist for turtles. Here we evaluate the utility of an intron in the RNA fingerprint protein 35 gene (R35) for use in turtle phylogenetics by examining its sequence properties, evolution, and ability to address important, current issues regarding turtle phylogeny. We have three goals in this study. First, we characterize attributes of the R35 intron based on sequences from several turtles. Second, we assess the phylogenetic utility of the R35 intron by attempting to clarify uncertain relationships within and between turtle families. Third, we briefly review prospects of nuclear gene phylogenetics in general.

1.1. The gene—RNA fingerprint protein 35

The RNA fingerprint protein 35 (R35) belongs to a very large, extremely diverse superfamily of 7-transmembrane proteins (Friedel et al., 2001). Analysis of the coding sequence indicates that R35 is a polypeptide consisting of 723 amino acids (Friedel et al., 2001). Currently, its function remains unknown, although sequence comparisons indicate that R35 may act as a G-protein-coupled receptor (GPCR) that is developmentally regulated and highly expressed in the cerebellum, spinal cord, and ganglion cells of chick embryos (Friedel et al., 2001). However, homology is too low to include R35 in any known gene subfamily, suggesting that it may act as a single-locus gene and/or that it belongs to a novel GPCR subfamily (Friedel et al., 2001; H. Schnürch, personal communication). Friedel et al. (2001) noted that an intron of approximately 1000–2000 bp resided in the transmembrane domain 6 in several birds (*Meleagris gallopavo*, *Anser anser*, *Struthio camelus*, and *Sitta europea*), a tortoise (*Geochelone denticulata*), and snake (*Elaphe guttata*) (Friedel et al., 2001).

1.2. Brief overview of the turtle phylogenetic questions

We assessed the phylogenetic utility of the R35 intron by examining two controversial aspects of relationships within turtles (order Testudines) for which different types of data (morphological and molecular) disagree. The first problem aims to clarify disagreement regarding the relationships within the family Chelidae, while the second tests the validity of the superfamily Trionychoidea (sensu Gaffney and Meylan, 1988).

Along with the families Pelomedusidae and Podocnemidae (sensu Shaffer et al., 1997), the Chelidae belong to the suborder Pleurodira, or side-necked turtles. Pleurodires are widespread in the southern hemisphere, with pelomedusids in Africa, podocnemids in Madagascar and South America, and chelids in Australia/New Guinea and South America. The phylogenetic relationships within Chelidae remain ambiguous despite extensive analyses using both morphological (Gaffney, 1977; Gaffney and Meylan, 1988) and molecular (mtDNA and nuclear DNA) data sets, either singly or in combination (Georges et al., 1998; Seddon et al., 1997; Shaffer et al.,

1997). Generally, morphology allies the long-necked chelids regardless of their continental origin (*Chelus* and *Chelodina* in Fig. 1A) (Gaffney, 1977; Gaffney and Meylan, 1988; Shaffer et al., 1997), while molecular data favor the monophyly of the South American and Australian/New Guinea groups regardless of neck length (Georges et al., 1998; Shaffer et al., 1997; Fig. 1B). However, this issue remains unresolved. Georges et al. (1998) and Krenz et al. (in preparation) found good bootstrap support for a South American clade but either could not resolve (Georges et al., 1998) or found low support for (Krenz et al., in preparation) an Australian/New Guinea group. The analysis presented in Shaffer et al. (1997) found moderately strong support for monophyly of Australian/New Guinea taxa and weak support for the monophyly of South American taxa.

The second issue we address involves the superfamily Trionychoidea, whose postulated extant members include the turtle families Trionychidae (soft-shelled turtles), Carettochelyidae (pig-nosed turtle), Dermatemydidae (Central American river turtle), and Kinosternidae (New World mud and musk turtles) (Gaffney and Meylan, 1988). Morphological analyses supports this group with moderately strong bootstrap proportions (77%; Shaffer et al., 1997). MtDNA data does not support this group, instead indicating that Trionychoidea is polyphyletic (i.e., Fig. 1B; Shaffer et al., 1997). As this problematic area involves several turtle families, any conclusive statement regarding the validity of Trionychoidea will affect a large portion of the turtle phylogeny.

2. Materials and methods

2.1. Taxonomic sampling

A comprehensive and complete treatment of the deep phylogeny of turtles is beyond the scope of this paper. Instead, we chose our study taxa with the goal of characterizing the phylogenetic utility of the R35 intron (Table 1). We included taxa from ten of the 14 commonly recognized turtle families, including all four from the superfamily Trionychoidea and all three from the suborder Pleurodira. Our sampling for the family Chelidae is the most extensive. We included two genera of South American turtles (the long-necked *Chelus* and short-necked *Phrynops*) and three Australian genera (the long-necked *Chelodina* and short-necked *Elseya* and *Emydura*). As outgroups for rooting the chelonian phylogeny, and because of our interest in reptile biology in general, we generated sequences for a bird (turkey, *M. gallopavo*) and a snake (gopher snake, *Pituophis catenifer*) in our initial assessment of the R35 intron.

2.2. Amplification and sequencing

A useful method to amplify introns of unknown sequence involves the development of primer sets that anneal to the flanking exons. Subsequent amplification via the PCR with these primers (exon-primed-intron-crossing, or EPIC, PCR) yields products that contain the intron and parts of the flanking exons (Palumbi and

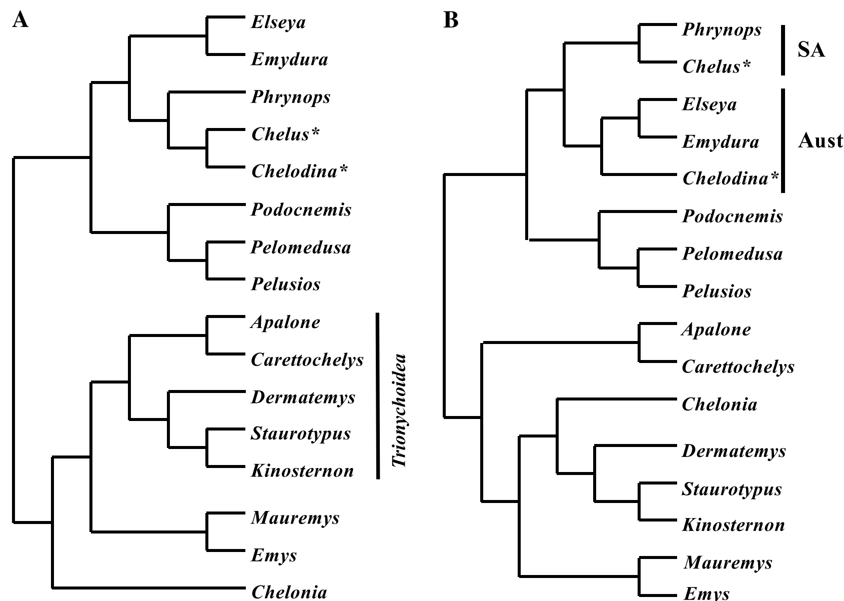

Fig. 1. Previous hypotheses of phylogenetic relationships of selected turtle lineages based on (A) morphological characters (Gaffney and Meylan, 1988) and (B) DNA sequence data (Georges et al., 1998; Shaffer et al., 1997). Stars indicate long-necked turtles of the family Chelidae. Based on morphological data (A) long-necked chelids are monophyletic regardless of continent and Trionychoidea forms a clade, and based on molecular data (B) chelids consist of reciprocally monophyletic South American and Australian/New Guinea groups and Trionychoidea is polyphyletic.

Table 1
Familial association, species, specimen, source, and sequence information for tissues used in this study

Family	Species	Specimen[a]	Source	GenBank Accession No.
Kinosternidae	*Kinosternon bauri*	HBS 16366	Field collected	AY339642
	Staurotypus triporcatus	HBS 16258	Field collected	AY339633
Dermatemydidae	*Dermatemys mawii*	HBS 38456	Zoo Specimen	AY339638
Cheloniidae	*Chelonia mydas*	TNE 1807	Field collected	AY339635
Geomydidae	*Mauremys caspica*	TNE 948	Field collected	AY339630
Emydidae	*Emys marmorata*[b]	HBS 14335	Field collected	AY339631
Trionychidae	*Apalone ferox*	TNE 710	Field collected	AY259580
Carettochelyidae	*Carettochelys insculpta*	TNE 931	Field collected	AY259571
Podocnemidae	*Podocnemis unifilis*	TNE 1361	Field collected	AY339634
Pelomedusidae	*Pelusios williamsi*	HBS 16408	Pet trade	AY339629
	Pelomedusa subrufa	HBS 16420	Pet trade	AY339639
Chelidae	*Chelus fimbriatus*	HBS 32256	Pet trade	AY339640
	Phrynops nasutus	HBS 27099	Pet trade	AY339637
	Elseya latisternum	HBS El Cairns 10	Field collected	AY339643
	Emydura subglobosa	HBS 32198	Field collected	AY339632
	Chelodina rugosa	HBS INVERELL#1	Field collected	AY339641
	Chelodina longicollis	HBS Walk B-1	Field collected	AY339636

[a] The numbers are specimen accessions from the tissue banks of H.B. Shaffer (HBS) or T.N. Engstrom (TNE).
[b] We follow Feldman and Parham (2002) in the recognition of *Emys*, rather than *Clemmys marmorata*.

Baker, 1994). The exon sequences should align with published sequences of the target gene, helping to verify that the primers annealed to the correct locus during PCR (Birks and Edwards, 2002; Prychitko and Moore, 1997). We used Primer3 (Rozen and Skaletsky, 2000; available online at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) to design primers to amplify the R35 intron based on published sequences of the exons 1 and 2 from the tortoise *G. denticulata* (GenBank Accession Nos. AJ293980 and AJ293981, respectively; Friedel et al., 2001). The forward primer anneals within exon 1 while the reverse primer anneals within exon 2. We repeated this process to develop squamate and bird primers based on sequences of the rat snake (*E. guttata*; GenBank Accession Nos. AJ293982 and AJ293983) and turkey (*M. gallopavo*; GenBank Accession Nos. AJ293972 and AJ293973). Table 2 lists the primer sequences for the turtle (R35Ex1, R35Ex2), snake (SQR35Ex1, SQR35Ex2), and bird (AR35F, AR35R).

We performed DNA extractions using standard phenol/chloroform or salt extraction protocols (Sambrook and Russell, 2001) to obtain whole genomic DNA from skeletal muscle, liver, or blood. Successful amplification utilized the following PCR conditions: 100 ng genomic DNA, 0.25 μM each primer, 0.0625 μM each dNTP, 3 mM $MgCl_2$, 2.5 μl 10× PCR buffer, and 1 U *Taq* DNA polymerase in a reaction volume of 25 μl. The cycling conditions were: 94 °C, 5 min; 94 °C for 30 s, 60 °C for 90 s, 72 °C for 120 s, repeated 35 times; 72 °C, 10 min. All PCRs were carried out in 0.6 ml PCR tubes in MJM Research PCT100 thermal cyclers (MJ Research). The 1.2 kb (turtle, turkey) or 2 kb (snake, *P. catenifer*) products were visualized with ethidium bromide by running 4 μl of the reaction mix on a 1% agarose gel using Tris–acetate–EDTA buffer. To prepare the PCR product for sequencing, 20 μl of the PCR mix was incubated first for 15 min with 20 U Exonuclease I and 2 U Shrimp alkaline phosphatase at 37 °C, followed by deactivation at 80 °C for 15 min. Sequencing reactions were performed in 10 μl volumes using ABI Prism BigDye Terminator chemistry (Applied Biosystems, Foster City, CA), and both strands of DNA were sequenced using an ABI Model 3100 automated sequencer at the UC Davis Division of Biological Sciences Automated Sequencing Facility. We deposited the sequences in GenBank (Accession Nos. AY339629–AY339643). Sequences for *Apalone ferox* and *Carettochelys insculpta* were deposited previously (Accession Nos. AY259580

Table 2
Forward (F) and reverse (R) sequences (5′ to 3′) of primers used in the PCR amplification and sequencing of the 1.2–2 kb R35 intron in the turtles, snake (*Pituophis catenifer*), and bird (*Meleagris gallopavo*) of this study

Taxon	Primer name	Primer sequences (5′ to 3′)	PCR product size (kb)
Turtles	R35Ex1	F: ACGATTCTCGCTGATTCTTGC	1.2
	R35Ex2	R: GCAGAAAACTGAATGTCTCAAAGG	
Snake	SQR35Ex1	F: CTTGCTCTGACCAAAGTCATTC	2
	SQR35Ex2	R: CCCAACAGTGTATTGAACAACC	
Bird	AR35F	F: TTTAGGAAAGGTGTGGCTGAGAG	1.2
	AR35R	R: GGAAGCTAAGTGTCTCAAACGAAAG	

and AY259571, respectively; Engstrom et al., submitted).

2.3. Southern blot

Using amino acid sequence comparisons, Friedel et al. (2001) hypothesized that R35 may belong to a novel GPCR subfamily, and may act as a single-locus gene (H. Schnürch, personal communication), though they did not specifically test this hypothesis. To determine the gene copy number of R35 in turtles, we performed a southern hybridization via a chemiluminescent detection system with the intron sequence acting as the probe [HybQuest Complete (DNP) System, Mirus]. We cloned the intron PCR product from *Elseya latisternum* using the pGEMT-Eazy Vector system (Promega) according to manufacturer's instructions. We randomly chose one clone to serve as a template for a PCR-produced probe. After amplification from this clone, we labeled the probe with dinitrophenyl according to manufacturer's instructions (Mirus).

We performed overnight digestions of genomic DNA (~20 μg) from the *E. latisternum* with the restriction endonuclease BstUI (New England BioLabs). We chose this enzyme because the intron lacked its recognition sequence, and thus a single band would indicate that the intron exists as a single-copy locus. After separating fragments on a 1% agarose gel, we transferred the DNA to a positively charged nylon membrane using standard alkaline transfer methods (Sambrook and Russell, 2001). Hybridization and detection were performed according to manufacturer's instructions (Mirus). The probe-genomic DNA hybridization step was carried out at a temperature of intermediate stringency (40 °C). Detection involved using an anti-DNP antibody with an alkaline phosphatase conjugate that reacts with the chemiluminescent substrate (Lumi Phos Plus).

2.4. Phylogenetic analyses

We aligned the sequences using Clustal X (Thompson et al., 1997), further refining the alignment by eye. Because several indels and repeating elements confounded assessments of homology, we removed these regions before performing any phylogenetic analyses. Base composition, transition:transversion ratio (ti:tv), sequence divergence, and other sequence properties were calculated using PAUP*4.0b10a (Swofford, 2000). We plotted uncorrected pairwise distances for transitions and transversions against maximum likelihood distances to assess levels of substitutional saturation. A graphical representation of variability along the length of the intron was constructed using SWAN 1.0 (Proutski and Holmes, 1998). We used Modeltest v. 3.06 (Posada and Crandall, 1998) to determine the best-fit model of sequence evolution.

We used maximum parsimony (MP) and maximum likelihood (ML) tree reconstruction methods as implemented in PAUP*4.0b10a (Swofford, 2000). In the MP analyses, we considered character changes as unordered and of equal weight, treated gaps as missing data, and used a random stepwise sequence addition (number of replicates = 10) algorithm with tree-bisection-reconnection (TBR) branch swapping. We also implemented a ti:tv weighting scheme to determine whether weighting character changes influenced the outcome of the tree searching. Bootstrap analyses (1000 replicates for parsimony, 100 replicates for ML) were conducted to estimate nodal support. As the likelihood ratio test and the Akaike information criterion (AIC) as implemented in Modeltest recommended different models of sequence evolution, we incorporated both into separate ML analyses. We were not able to root the turtle ingroup using outgroup sequences because bird and snake sequences did not align with turtle sequences. Instead we used the well-accepted monophyly of Pleurodira and Cryptodira (Gaffney and Meylan, 1988; Williams, 1950) to root the tree between these two lineages by constraining the three pleurodiran families Podocnemidae, Pelomedusidae, and Chelidae to be a monophyletic outgroup to the remaining taxa: this rooting has been confirmed with the fossil *Proganochelys* based on morphological evidence (Shaffer et al., 1997). To test for clock-like behavior, we performed a likelihood ratio test using likelihood scores from clock-enforced and clock-unenforced ML analyses.

3. Results

3.1. Sequence analysis

PCR amplifications resulted in a product of approximately 1.2 kb in turtles and turkey, and 2 kb in the snake. Visualization of the product on a gel and the sequencing of clones did not reveal any evidence for multi-locus amplification, and all PCR products produced sequences with uniform peaks and minimal background noise, implying we always amplified single products. The PCR product contained 55 bp (turtles), 38 bp (snake), or 147 bp (turkey) from the 3′ end of exon 1 and 66 bp (turtle), 30 bp (snake), or 64 bp (turkey) from the 5′ end of exon 2, allowing us to compare our exon sequences to available sequences in GenBank. The exon/intron boundary corresponded to predicted splice sites (program available at http://www.fruitfly.org/seq_tools/splice.html). Bird and snake sequences were too divergent to align with each other and to the turtle sequences, and we therefore deleted them from all further analyses.

The initial alignment was 1093 bp long and included indels ranging in size from 1 to 30 bp as well as several

repetitive elements. Assessment of homology was difficult for some indels, and two small repeats (a poly(T) at bp positions 445–456 and a GA-rich segment at bp positions 948–971) posed problems with alignment; we removed these from the data matrix for phylogenetic analyses. After excluding the troublesome regions, the pruned, aligned sequences consisted of 916 bp; 428 were variable and 211 were parsimony-informative among turtles.

Kimura-2-parameter estimates of pairwise sequence divergence ranged from 1.68% between *Chelus fimbriatus* and *Phrynops nasutus* to 24.1% between *Podocnemis unifilis* and *C. insculpta* (data not shown). Maximum-likelihood corrected base pair frequencies using the HKY + G model (see below; Hasegawa et al., 1985) did not indicate skewness (%A = 0.2890, %C = 0.1898, %G = 0.2076, and %T = 0.3136); a phylogenetically unconstrained χ^2 test of homogeneity could not reject homogeneity of base pair composition across the 17 taxa ($\chi^2 = 12.02$, df = 48, $P = 1$). Plots of maximum likelihood distances (HKY + G) versus the uncorrected *P* distances showed no evidence of the characteristic leveling off associated with saturation (Fig. 2). A plot of variability with respect to sequence position (Fig. 3) revealed an elevated level of variability from roughly nucleotide positions 500–900 relative to the first 500 bp. Interestingly, when we removed *Staurotypus* from the data set, the variability appeared more evenly distributed throughout the entire intron (data not shown).

3.2. Southern blot

The restriction enzyme BstUI is a 4-cutter (recognition sequence 5′-CGCG-3′, cutting on average every 256 bp. Because of the length of the intron and the absence of a BstUI recognition sequence within it, we predicted that the size of the fragment containing the intron should be approximately the size of the intron itself. A single band of approximately 1500 bp was detected in our Southern hybridization experiment, indicating that the intron is present only once in the genome of *E. latisternum* (data not shown).

3.3. Phylogenetic analyses

Both unweighted and weighted parsimony analysis produced single, completely resolved most parsimonious trees of identical topology (Fig. 4). For weighted parsimony, we used an ML-corrected ti:tv ratio of 1.7687:1 based on the HKY + G model of sequence evolution (see below). While most bootstrap values were 90% or more, the clade containing *Kinosternon*, *Staurotypus*, *Dermatemys*, and *Chelonia* was supported by bootstrap values of 89% with the unweighted parsimony analysis and 77% with the weighted analysis. *Apalone* and *Carettochelys* formed a clade that is sister to the remaining cryptodiran turtles used in this study. The chelids separated into two groups—a South American clade (100% bootstrap support) and an Australian/New Guinea clade (99% bootstrap support for both unweighted and weighted).

ModelTest found two optimal models of sequence evolution: the HKY + G model (ti:tv = 1.7687; $\alpha = 3.7892$; Hasegawa et al., 1985) based on the likelihood ratio test, and the K81uf + G model (*R*[A–T] = 0.7729, *R*[C–G] = 0.7729, *R*[C–T] = 3.2671, *R*[G–T] = 1, *R*[A–C] = 1, *R*[A–G] = 3.2671; $\alpha = 3.8008$; base frequencies: A = 0.2901, C = 0.1890, G = 0.2065, T = 0.3144;

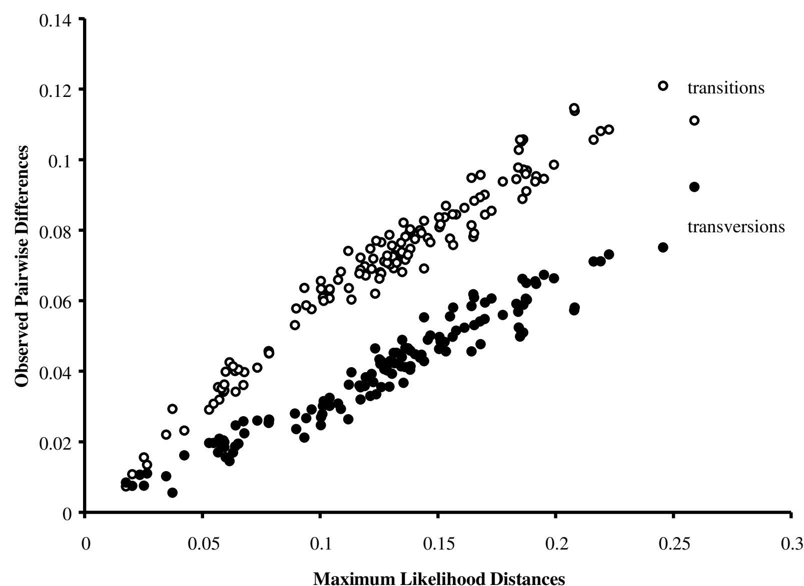

Fig. 2. Plots of maximum likelihood distances versus the uncorrected *P* distances among the 17 turtle R35 intron sequences.

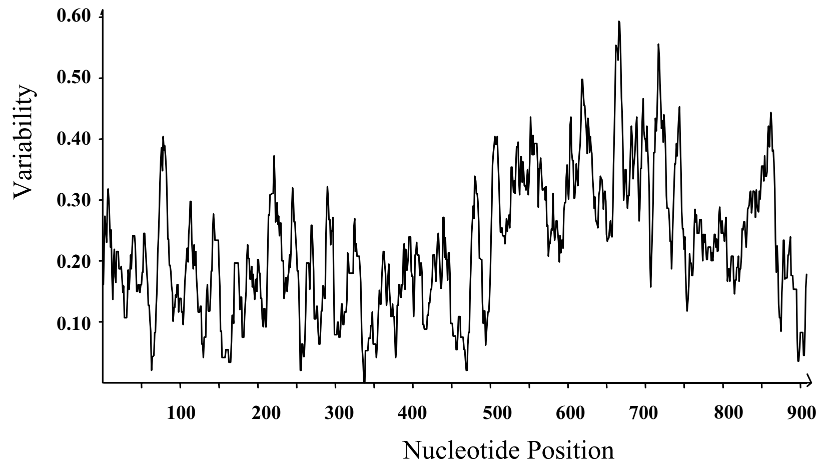

Fig. 3. Sliding window plot of variability across the length of the intron (interval length = 10 bp) as determined using the program SWAN (Proutski and Holmes, 1998), which calculates the nucleotide variation per column as an entropy function whose value range from 0 (no variability) to 1 (highest variability).

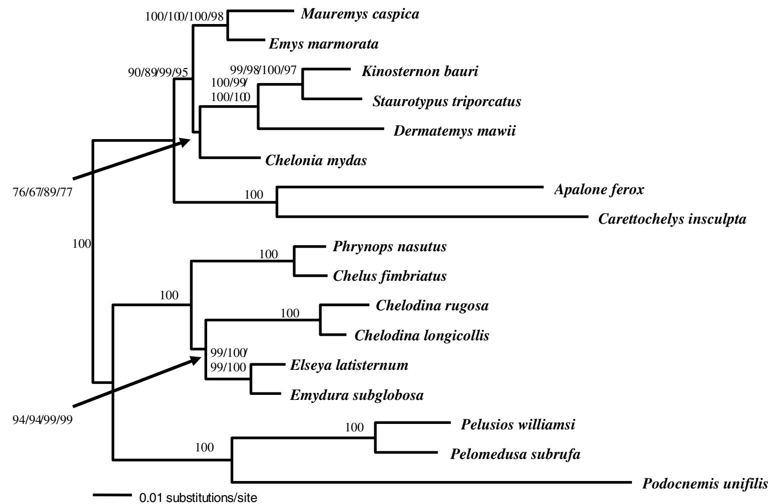

Fig. 4. Phylogram of turtles used in this study as inferred using maximum likelihood and the HKY + G model of sequence evolution. Numbers indicate bootstrap support based on the following analyses: ML (HKY + G)/ML (K81uf + G)/MP (unweighted)/MP (weighted); a single 100 indicates that the node received 100% bootstrap support for all analyses. See text for details.

Kimura, 1981) based on the AIC. Regardless of the model used, the ML phylogenetic analyses resolved trees topologically identical to those from the MP analyses (Fig. 4). The *Kinosternon–Staurotypus–Dermatemys–Chelonia* clade received moderate support with the HKY + G model (76%), and weaker support with the K81uf + G model (67%). In agreement with the parsimony analyses, South American and Australian/New Guinea chelids are reciprocally monophyletic groups. Of particular interest is the high bootstrap value (94% with both models) uniting the Australian chelids; this is the highest support found thus far for an Australian chelid clade. A likelihood ratio test indicated the intron did not evolve in a clocklike manner ($\chi^2 = 93.447$, df $= 12$, $P < .001$).

4. Discussion

4.1. Intron characterization

The sequence characteristics of the R35 intron are similar to those of other phylogenetically informative nuclear markers, including the introns of rhodopsin (Birks and Edwards, 2002), ovomucoid (Armstrong et al., 2001), and β-fibrinogen (Prychitko and Moore, 1997), as well as coding sequences of RAG-1 (Groth and Barrowclough, 1999), DMP-1 (Van Den Bussche et al., 2003), and the plant gene DMC-1 (Peterson and Seberg, 2002). In general, such characteristics include homogeneity in base pair composition, low ti:tv ratio, and low among-site rate heterogeneity. The taxonomic

applicability of these markers ranges widely—from the infraclass to the intergenus levels, and the utility of any particular marker must be investigated on a case-by-case basis. The variation in introns often reflects that of their flanking exons (Zang and Hewitt, 2003), implying that different introns can have drastically different rates of evolution. Thus far, as demonstrated here and in other studies, intron sequences tend to have their greatest phylogenetic utility for intermediately diverged lineages. More ancient divergences, such as that between turtles and birds, may require more conserved markers such as coding sequences (i.e., RAG-1; Krenz et al., in preparation). Too few studies of nuclear intraspecific phylogenetics exist to gauge intron utility at that level, though the prospects of nuclear phylogeography are both important and exciting (Hare, 2001).

Our results indicate that the R35 intron has the potential to be an excellent phylogenetic marker in turtles, and it may prove useful for other amniote taxa as well. As predicted by Friedel et al. (2001) and confirmed by our Southern blot experiment, the intron acts as a single-copy locus, eliminating the problems associated with paralogy and homology. A low ti:tv ratio, unbiased base pair composition, and lack of saturation (Fig. 2) indicate that the problems of long-branch attraction and statistical inconsistency associated with these issues (Sanderson and Shaffer, 2002) are not features of the R35 intron, at least across turtles. Although our sample of sequences failed the test of strict molecular clock, among-lineage and among-site rate heterogeneity (as demonstrated by a relatively low shape parameter in the model of molecular evolution, and sliding window plot of variability in Fig. 3) do not appear to interfere with R35's phylogenetic utility, based on general concordance with other markers for turtle phylogenetics. Methods exist that allow the dating of divergence times despite unclock-like evolution (Kishino et al., 2001; Sanderson, 1997, 2002; Thorne et al., 1998), and our laboratory is currently investigating molecular clock calibrations for turtles using the R35 intron.

While there are several characteristics that make the R35 intron an excellent candidate as a marker for phylogenetics, there are also drawbacks to its utility. We were unable to align turtle sequences with outgroup sequences, due to the large sequence divergences among snake, bird, and turtle. This made it impossible to test the rooting of the turtle phylogeny. This extreme divergence among distantly related taxa might be a general limitation on the application of intron sequences to phylogenetics. The crown group of turtles is at least 210 million years old based on the fossil *Proganochelys* (Gaffney, 1990), and R35 is both alignable and informative across this time span. However, our experience to date suggests that the R35 intron is not an appropriate marker for studies aimed at inferring deeper phylogenetic relationships among amniotes and beyond. The intron does harbor some variation at the intra/interspecific level in the snapping turtles (*Chelydra acutirostris*, *Chelydra rossignoni*, and *Chelydra serpentina*), North American softshelled turtles (*Apalone ferox*, *Apalone mutica*, and *Apalone spinifera*) and the painted turtle *Chrysemys picta* (unpublished data), and this level of variation may be sufficient for intraspecific phylogeographic studies (e.g., Hare, 2001). We are actively pursuing this possibility in our laboratory.

4.2. Turtle phylogeny

In general, phylogenetic hypotheses based on molecular evidence tend to agree with those based on morphological evidence; discordances between the two often result from weakly supported relationships or methodological differences (reviewed in Hillis and Wiens, 2000). This is the case for the turtle phylogeny. While there appear to be differences between the turtle phylogeny as inferred using molecular sequence data and morphological characters, the alternative relationships tend to have weak support in one or both hypotheses (see Georges et al., 1998; Seddon et al., 1997; Shaffer et al., 1997). Unfortunately, the differences involve relationships among several major lineages of turtles, resulting in an ambiguous phylogeny for many of the deep relationships among chelonian taxa. As a result, the turtle phylogeny contains important issues that require clarification. Despite the relatively sparse taxonomic sampling, our study is the first to resolve several of these troublesome relationships with very strong statistical support.

One of the most revolutionary aspects of the phylogeny presented by Gaffney and Meylan (1988) is the superfamily Trionychoidea. In their influential work, they demonstrated that the families (Kinosternidae + Dermatemydidae), and (Trionychidae + Carretochelyidae), were sister taxa, and that together these four families constituted a monophyletic superfamily Trionychoidea. They based this conclusion on six osteological and muscle characters (Gaffney and Meylan, 1988). Shaffer et al. (1997) conducted several analyses using morphological characters and fossil taxa, finding moderate to good statistical support for these groupings. Molecular data from the mitochondrial genome agreed with the sister pairings of (Kinosternidae + Dermatemydidae) and (Trionychidae + Carretochelyidae), but not on the monophyly of Trionychoidea. Rather, the mtDNA sequence data places *Carettochelys* plus Trionychidae (Trionychoidae sensu Shaffer et al., 1997) sister to all other Cryptodira, with no close relationship to (Kinosternidae + Dermatemydidae). However, bootstrap support for this placement was relatively weak, and the combined molecular and morphological data set agreed with the mtDNA in weakly supporting

Trionychoidae as sister to the remaining Cryptodira. In this study, we found strong support for the sister-group relationship of Trionychoidae to the remaining Cryptodira, and thus the first strong evidence rejecting the monophyly of Trionychoidea sensu Gaffney and Meylan (1988). This result has important implications for cryptodiran phylogenetics and systematics, although we feel more complete taxon sampling is necessary before making definitive conclusions. Our laboratory is currently investigating these sets of relationships with increased taxon sampling using both R35 and other nuclear markers.

Our data from the R35 intron also provide insights into the relationships of the pleurodiran family Chelidae. Our study supports the grouping of South American and Australian/New Guinea chelids into reciprocally monophyletic clades with very high bootstrap support (Fig. 4), in agreement with previous mitochondrial DNA studies (Georges et al., 1998; Seddon et al., 1997; Shaffer et al., 1997). As suggested in Shaffer et al. (1997), this result is consistent with the interpretation that the primary division of Australasian and South American chelids dates to the continental separation approximately 55 mya, rather than the more ancient diversification of chelid genera required under the morphology-based hypothesis.

We do not intend for these to be considered authoritative analyses of turtle phylogenetics; our taxon sampling is sparse, we have not examined the combined effects of all available data, and additional work is needed to confirm the placement of the root of the turtle tree. Instead, our goal has been to test the phylogenetic utility of the R35 across a broad spectrum of chelonian evolution, applying it to two long-standing and previously intractable problems in turtle systematics. The R35 intron has many desirable characteristics for a molecular marker, and our study shows strong support for relationships that have been difficult to resolve using other data. Our group is currently investigating whether such relationships hold upon further analyses including more loci and greater taxonomic sampling. We found the R35 intron to be an excellent molecular marker for phylogenetic studies of turtles, and believe its utility can extend to squamate and bird phylogenetics on similar divergence scales.

Acknowledgments

We would like to thank P. Meylan, A. Meylan, and P. Moler for tissue samples. The Shaffer laboratory group provided helpful comments on earlier drafts of the manuscript. This material is based upon work supported by the NSF under Grants 9727161/0213153, CalFed, and the UC Davis Agriculture Experimental Station.

References

Armstrong, M.H., Braun, E.L., Kimball, R.T, 2001. Phylogenetic utility of avian ovomucoid intron G: a comparison of nuclear and mitochondrial phylogenies in galliformes. Auk 118, 799–804.

Avise, J.C., 1994. Molecular Markers, Natural History, and Evolution. Kluwer Academic Publishers, Boston.

Baker, R.H., Wilkinson, G.S., DeSalle, R., 2001. Phylogenetic utility of different types of molecular data used to infer evolutionary relationships among stalk-eyed flies (Diopsidae). Syst. Biol. 51, 87–105.

Barns, S.M., Delwiche, C.P., Palmer, J.D., Pace, N.R., 1996. Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. 99, 9188–9193.

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

DeBry, R.W., Seshadri, S., 2001. Nuclear intron sequences for phylogenetics of closely related mammals: an example using the phylogeny of *Mus*. J. Mammal. 82, 280–288.

Feldman, C.R., Parham, J.F., 2002. Molecular phylogenetics of emydine turtles: taxonomic revision and the evolution of shell kinesis. Mol. Phylogenet. Evol. 22, 388–398.

Friedel, R.H., Stubbusch, J., Barde, Y.-A., Schnürch, H., 2001. A novel 7-transmembrane receptor expressed in nerve growth factor-dependent sensory neurons. Mol. Cell. Neurosci. 17, 31–40.

Gaffney, E.S, 1977. The side-necked turtle family Chelidae: a theory of relationship and using shared derived characters. Am. Mus. Novit. 2620, 1–28.

Gaffney, E.S., 1990. The comparative osteology of the Triassic turtle *Proganochelys*. Bull. Am. Mus. Nat. Hist. 194, 1–263.

Gaffney, E.S., Meylan, P.A., 1988. A phylogeny of turtles. In: Benton, M.J. (Ed.), The Phylogeny and Classification of Tetrapods. Clarendon Press, Oxford, England, pp. 157–219.

Garcia-Machado, E., Pempera, M., Dennebouy, N., Olivia-Suarez, M., Mounolou, J.-C., Monnerot, M., 1999. Mitochondrial genes collectively suggest the paraphyly of crustacea with respect to insecta. J. Mol. Evol. 49, 142–149.

Georges, A., Birrell, J., Saint, K.M., McCord, W., Donellan, S.C., 1998. A phylogeny of side-necked turtles (Chelonia: Pleurodira) based on mitochondrial and nuclear gene sequence variation. Biol. J. Linn. Soc. 67, 213–246.

Graur, D., Li, W.-H., 2000. Fundamentals of Molecular Evolution. Sinauer Associates, Sunderland, MA.

Groth, J.G., Barrowclough, G.F., 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12, 115–123.

Hare, M.P., 2001. Prospects for nuclear gene phylogeography. Trends Ecol. Evol. 16, 700–706.

Harris, D.J., 2003. Codon bias variation in C-*mos* between squamate families might distort phylogenetic inferences. Mol. Phylogenet. Evol. 27, 540–544.

Hasegawa, M., Kishino, H., Yano, T.-A., 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 2, 160–174.

Hillis, D.M., Wiens, J.J., 2000. Molecules versus morphology in systematics, conflicts, artifacts, and misconceptions. In: Wiens, J.J. (Ed.), Phylogenetic Analysis of Morphological Data. Smithsonian Institution Press, Washington.

Howarth, D.G., Baum, D.A., 2002. Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. Mol. Phylogenet. Evol. 23, 525–528.

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

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

Kimura, M., 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. 78, 454–458.

Kishino, H., Thorne, J.L., Bruno, W.J., 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18, 352–361.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. 86, 6196–6200.

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

Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47, 61–76.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

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

Peterson, G., Seberg, O., 2002. Molecular evolution and phylogenetic application of *DMC1*. Mol. Phylogenet. Evol. 22, 43–50.

Posada, D., Crandall, K.A.., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Proutski, V., Holmes, E.C., 1998. SWAN: sliding window analysis of nucleotide sequence variability. Bioinformatics 14, 467–468.

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

Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.

Saint, K.M., Austin, C.C., Donnellan, S.C., Hutchinson, M.N., 1998. *C-mos*, a nuclear marker useful for squamate phylogenetic analysis. Mol. Phylogenet. Evol. 10, 259–263.

Sambrook, J., Russell, D.W., 2001. Molecular Cloning: A Laboratory Manual, third ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.

Sanderson, M.J., Shaffer, H.B., 2002. Troubleshooting molecular phylogenetic analyses. Ann. Rev. Ecol. Syst. 33, 49–72.

Seddon, J.M., Georges, A., Braverstock, P.R., McCord, W., 1997. Phylogenetic relationships of chelid turtles (Pleurodira: Chelidae) based on mitochondrial 12S rRNA gene sequence variation. Mol. Phylogenet. Evol. 7, 55–61.

Shaffer, H.B., Meylan, P., McKnight, M.L., 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Syst. Biol. 46, 235–268.

Sunnucks, P., 2000. Efficient markers for population biology. Trends Ecol. Evol. 15, 199–203.

Swofford, D.L., 2000. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), version4. Sinauer Associates, Sunderland, MA.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15, 1647–1657.

Van Den Bussche, R.A., Reeder, S.A., Hansen, E.W., Hoover, S.R., 2003. Utility of the dentin matrix protein 1 (*DMP1*) gene for resolving mammalian intraordinal phylogenetic relationships. Mol. Phylogenet. Evol. 26, 89–101.

Weibel, A.C., Moore, W.S., 2002. A test of a mitochondrial gene-based phylogeny of woodpeckers (Genus *Picoides*) using an independent nuclear gene, β-fibrinogen intron 7. Mol. Phylogenet. Evol. 22, 247–257.

Wiens, J.J., Hollingsworth, B.D., 2000. War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. Syst. Biol. 49, 143–159.

Williams, E.E., 1950. Variation and selection in the cervical central articulations of living turtles. Bull. Am. Mus. Nat. Hist. 94, 505–562.

Zang, D.-X., Hewitt, G.M., 2003. Nuclear DNA analysis in genetics studies of populations: practice, problems, and prospects. Mol. Ecol. 12, 563–584.