MOLECULAR PHYLOGENETICS AND EVOLUTION
Vol. 10, No. 2, October, pp. 259–263, 1998
ARTICLE NO. FY980515

C-*mos*, A Nuclear Marker Useful for Squamate Phylogenetic Analysis

Kathleen M. Saint,[1] Christopher C. Austin,[2] Stephen C. Donnellan, and Mark N. Hutchinson

Evolutionary Biology Unit and Herpetology Section, South Australian Museum, Adelaide, 5000, Australia

Received September 12, 1997; revised December 16, 1997

Phylogenetic reconstruction in molecular systematics has largely been achieved using mitochondrial gene sequences and less frequently sequences of nuclear ribosomal RNA genes. At present few other nuclear genes have been identified that could be used to test these phylogenies. C-*mos*, a single-copy nuclear oncogene, has been identified as a candidate nuclear marker. Data are presented on the usefulness of c-*mos* sequences in the phylogenetic analysis of squamate reptile families. We obtained partial sequences of c-*mos* from 13 squamate reptile families and outgroup representatives from the orders Rhynchocephalia, Chelonia, and Crocodylia. Phylogenetic analysis reveals a high degree of phylogenetic information contained within the sequence for both the synonymous and nonsynonymous substitutions. Phylogenetic resolution was present at both the deepest and shallower divergences but relationships among the major squamate lineages were not resolved, possibly because rapid cladogenesis may have led to the diversification of these lineages.

INTRODUCTION

The usefulness of mitochondrial and nuclear ribosomal genes for recovering phylogenetic relationships among both recently and anciently diverged animal taxa is well known (reviewed in Hillis *et al.*, 1996). However, nuclear genes that are capable of resolving relationships among intermediately diverged taxa, i.e., 50–300 MYA (million years ago), are not commonly available. Single-copy nuclear genes are an increasingly important source of phylogenetic characters not only to resolve intermediate divergences but also for testing phylogenetic hypotheses based on mitochondrial and nuclear ribosomal RNA gene sequences. A number of genes have been identified that may prove to be useful for vertebrates (Graybeal, 1994) and invertebrates (Friedlander *et al.*, 1994; Mitchell *et al.*, 1997; Besansky and Fahey, 1997). At present there are no single-copy nuclear gene sequences that have been used for higher level phylogenetic analysis of reptiles. Our laboratory is interested in specific questions concerning the relationship of Australian reptiles to their overseas relatives (divergences up to 150 MYA) and nuclear genetic markers are required to test mitochondrial gene trees generated to date.

C-*mos*, a candidate nuclear gene proposed by Graybeal (1994), is a proto-oncogene that encodes a serine/threonine kinase expressed at high levels in germ cells. The protein is a regulator of meiotic maturation and is involved in the arrest of oocyte maturation at meiotic metaphase II (Yew *et al.*, 1993). The c-*mos* gene is single-copy, without introns, and is just over 1000 base pairs (bp). It has been found in the genomes of amphibians, birds, and mammals. The human c-*mos* open reading frame is 346 codons (Watson *et al.*, 1982). There are no repetitive elements in the sequence and few insertions or deletions that would complicate sequence alignment among vertebrates. These characteristics make it very amenable to PCR amplification from genomic DNA and direct sequencing of PCR products. Graybeal (1994) suggested that c-*mos* may provide phylogenetic information among taxa characterised by divergence times as old as 400 MYA. Encouragingly, Cooper and Penny (1997) have recently used c-*mos* sequences, obtained from genomic DNA by PCR, to recover evolutionary relationships among the avian orders. Here we describe primers that amplify a portion of the c-*mos* gene in a wide variety of reptilian families representative of the four reptilian orders. We determine the phylogenetic utility of the resulting sequence data by assessing concordance between the c-*mos* gene tree and recent summary cladograms of relationships based on morphological characters (Estes *et al.*, 1988; Macey *et al.*, 1997a).

MATERIALS AND METHODS

The 25 species examined and specimen museum identification numbers are: F. Agamidae: *Ctenophorus decresii* SAMA R42978, *Physignathus cocincinus* EBU 0188218; F. Amphisbaenidae: *Bipes biporus* CM22; F. Anguidae: *Elgaria multicarinata* CM199; F. Boidae:

[1] To whom correspondence should be addressed. Fax: 61 8 82077222, E-mail: sdonnell@genetics.adelaide.edu.au.
[2] Present address: Evolutionary Biology Unit, Australian Museum, Sydney, 2000, Australia.

1055-7903/98 $25.00

Candoia carinata USNM 512294; F. Chamaeleonidae: *Chamaeleo* sp. EBU D154; F. Gekkonidae: *Carphodactylus laevis* QM J8944, *Strophurus intermedius* SAMA R28963, *Eublepharis macularius* CM200, *Phyllodactylus marmoratus* SAMA R42098; F. Phrynosomatidae: *Sceloporus grammicus* CM331; F. Pygopodidae: *Delma fraseri* SAMA R32463; F. Pythonidae: *Python reticulatus* EBU D687; F. Scincidae: *Tiliqua scincoides* SAMA R30280, *Egernia luctuosa* AMS R98690, *Sphenomorphus leptofasciatus* THNC 51918, *Lipinia noctua* THNC 51595, *Ablepharus pannonicus* THNC 7676; F. Teiidae: *Cnemidophorus tigris* CM212, 213; F. Typhlopidae: *Ramphotyphlops australis* SAMA R36502; F. Varanidae: *Varanus rosenbergi* AMS R123331; O. Rhynchocephalia: F. Sphenodontidae: *Sphenodon punctatus* CM43; O. Chelonia: F. Chelidae: *Chelodina rugosa* NTM R13439, *Elseya dentata* NTM R13521; O. Crocodylia: F. Crocodylidae: *Crocodylus porosus* SAMA R34528. For a key to museum acronyms see Leviton *et al.* (1985). Additional acronyms are: CM, Craig Moritz Collection, University of Queensland; and EBU, Evolutionary Biology Unit, South Australian Museum. The species were selected to represent a wide range of evolutionary divergences across the four reptilian orders to test the utility of c-*mos* at a variety of levels of divergence.

Genomic DNA was extracted from small amounts of frozen liver or blood by the following method. Tissue was digested at 58°C for 1–3 h with 0.2 mg/ml proteinase K in 1% SDS in Tris buffer (pH 8) after homogenization. The resulting mixture was then extracted with phenol/chloroform (1:1) and the DNA was precipitated with 0.1 vol 3 M sodium acetate (pH 5.2) and 2 vol 100% ethanol. After washing in 70% ethanol, the genomic DNA was air dried and resuspended in 10 mM Tris-HCl (pH 8), 1 mM EDTA.

PCR primers were designed to anneal to regions conserved across the available vertebrate c-*mos* sequences in GenBank (human J00119, monkey X12449, rat X00422, chicken M19412, and frog M25366). The forward and reverse primers are G77 5′-TGG CYT GGT GCW NCA TNG ACT-3′ [409] and G78 5′-AGR GTG ATR WCA AAN GAR TAR ATG TC-3′ [1021], respectively. Numbers in brackets after the primer sequence refer to the 3′ position of the primer as localized on the human c-*mos* sequence. PCR products of about 660 bp in length were obtained from several skink species and sequenced. Using these sequence data, reptile-specific primers were designed to amplify a product of about 417 bp in the species listed above. The reptile-specific forward and reverse primers are G73 5′-GCG GTA AAG CAG GTG AAG AAA-3′ [513] and G74 5′-TGA GCA TCC AAA GTC TCC AAT C-3′ [888]. In a few species a product was only obtained by using one primer from each primer set (G73 and G78) which gave an intermediate size product. A second nested PCR using the reptile-specific primers G73 and G74 was carried out on this intermediate product which produced a fragment of the expected size. The PCR conditions were: 50–100 ng DNA, 12.5 pmol each primer, 0.2 mM each dNTP, 4 mM $MgCl_2$, 5 μl 10× *Taq* buffer, and 1 unit Promega *Taq* DNA polymerase in a total volume of 50 μl. PCR cycling conditions were: cycle 1 (94°C for 3 min, 48 or 55°C for 45 s, 72°C for 1 min); cycles 2 to 35 (94°C for 45 s, 48 or 55°C for 45 s, 72°C for 1 min); cycle 36 (72°C for 6 min, hold at room temperature) (FTS-320 Thermal Sequencer, Corbett Research). PCR products were purified using Bresa-Clean Nucleic Acid Purification Kit (Bresatec). Sequences of both strands were obtained with the PCR primers by direct sequencing of the double-stranded PCR product using the Perkin Elmer ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit. GenBank accession numbers are AF039462 to AF039486. The sequences were aligned using CLUSTAL W (Thompson *et al.*, 1994). To test whether the c-*mos* gene is present as a single copy in squamate reptiles, genomic DNA from two species of agamid, *Ctenophorus decresii* ($n = 1$) and *Ctenophorus nuchalis* ($n = 19$), was analyzed by Southern hybridization (Southern, 1975). Two micrograms of DNA was digested with restriction enzymes, separated by agarose gel electrophoresis, and transferred onto a positively charged nylon membrane (Hybond-N+, Amersham). Following the Boehringer Mannheim DIG High Prime Labelling and Detection Starter Kit I protocols, the c-*mos* sequence was localized by hybridizing, at 65°C overnight, a digoxigenin (dig)-labeled probe prepared from the c-*mos* PCR product of *C. decresii*, and the hybridized probe was immunodetected with anti-digoxigenin-alkaline phosphatase-conjugated antibody.

RESULTS

Southern analysis of *Eco*RI-digested DNA revealed a single band in 17 individuals and two bands in two individuals of *C. nuchalis*. Two bands were detected also in a single individual of *C. decresii*. Southern hybridization of *Hin*dIII- and *Pst*I-digested DNA from the individuals where two *Eco*RI bands were detected showed single bands when probed with the dig-labeled c-*mos* probe. The single *Pst*I band was less than 2.8 kb and the presence of more than one copy of the c-*mos* gene within this small fragment is unlikely. The results of the Southern analysis are consistent with the presence of a single copy of the c-*mos* gene in these species.

The aligned squamate c-*mos* sequences were 375 bp long (125 codons) and included 237 variable sites and 191 parsimony informative sites. The codon reading frame was inferred by comparison of translated sequences to the published amphibian, avian, and mammalian sequences (see accession numbers above). There was a high degree of amino acid homology between the reptilian sequences and the published c-*mos* sequences. Compared to the outgroups (Chelidae, Crocodylidae, and Sphenodontidae), six taxa displayed deletions that

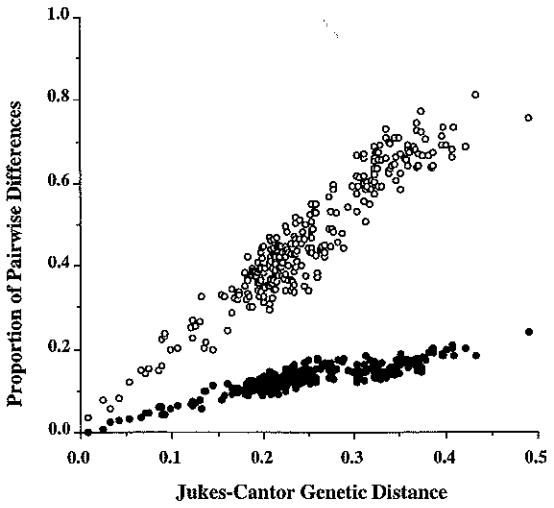

FIG. 1. Plots of frequencies of synonymous and nonsynonymous substitutions against Jukes–Cantor corrected genetic divergence among 25 reptilian c-*mos* sequences. Open circles, synonymous substitutions; solid circles, nonsynonymous substitutions.

did not disrupt the reading frame. Four of these deletions were one codon in length and of these, three were within the snake clade, families Boidae and Typhlopidae. All the deletions occurred within a 41-bp region (bp 727–768 of human c-*mos* sequence). Inclusion of the deletions in the phylogenetic analysis did not change the topology of the trees recovered. Figure 1 shows the relationship between Jukes–Cantor corrected genetic distance and the proportion of synonymous and nonsynonymous differences between taxa (Jukes and Cantor, 1969). The genetic distances and proportion of synonymous and nonsynonymous differences were calculated using the program MEGA (Kumar *et al.*, 1993). Polynomial regression lines were not significantly different from straight lines, indicating that saturation of third positions has not occurred and that there is phylogenetic signal at all codon positions.

The mean pair-wise comparisons for Jukes–Cantor corrected genetic distances (±SD) and the proportion of amino acid substitutions (±SD), respectively, are: interordinal (n = 89), 0.34 ± 0.05 and 0.26 ± 0.04; interfamilial (n = 282), 0.26 ± 0.07 and 0.24 ± 0.04; and intrafamilial (n = 18), 0.08 ± 0.04 and 0.10 ± 0.04. Of the 237 variable sites, 117 are third codon positions and 120 are at first and second codon positions.

The amount of phylogenetic signal was tested using the skewness test statistic, g_1 (Hillis and Huelsenbeck, 1992) with outgroups included. The g_1 value, using 10,000 randomly produced trees, was −1.05, which indicates that the data are significantly more structured than are random data (P = 0.01). The data were analyzed by maximum likelihood (ML) and maximum parsimony (MP) methods using test version 4.0d54 of PAUP*, written by D. L. Swofford. A transition/transversion ratio of 2.4, estimated using ML, was used in all analyses. The robustness of phylogenetic hypotheses was tested with bootstrapping using 100 pseudoreplicates for ML and 500 pseudoreplicates for MP. The outgroups are representatives from the orders Rhynchocephalia, Chelonia, and Crocodylia. Two ML trees were found with log likelihood values of −3769.37559. Three equally most parsimonious trees (length 703 steps) were found in the MP analysis. We present the consensus ML tree and bootstrap values for both ML and MP analyses in Fig. 2.

DISCUSSION

The Lepidosauria (the Sphenodontids and Squamates) diverged from the archosaurs (crocodiles) probably in the late Permian (Gow, 1975) and these two groups last shared a common ancestor with turtles in the late Carboniferous (Reisz, 1981). Monophyly of the lepidosaurs is supported in the c-*mos* phylogeny with strong bootstrap support from ML (89) and MP (94). The closest living relative of the squamates is *Sphenodon punctatus* with a time of divergence of 200 MYA (Evans, 1997). Squamate monophyly is strongly supported with bootstrap values of 100 in each analysis.

Within the squamate lineage, a sister relationship of *Cnemidophorus tigris* (F. Teiidae) with all the other squamates was recovered by both ML and MP analyses with bootstrap support of <50 and 63, respectively. Traditionally teiids are regarded as scincomorphans, i.e., allied to skinks, lacertids, and cordylids (Estes *et al.*, 1988; Macey *et al.*, 1997a). The unconventional relationship recovered in the present study may be due alternatively to the presence of a paralogous sequence in teiids or, less likely, to a mislabeled or contaminated sample as the sequence of c-*mos* from a second individual of *C. tigris* was identical.

The currently held idea (Estes *et al.*, 1988; Macey *et al.*, 1997b) that agamids are related to chamaeleons is strongly supported by the data with bootstrap values of 98 (ML) and 97 (MP). However, a sister-group relationship of the agamid/chamaeleon clade with the phrynosomatid, although recovered by the ML analysis, received poor support from bootstrapping (<50) and was not present in the MP phylogeny. The ML and MP analyses support a monophyletic clade for the two families Gekkonidae and Pygopodidae, with bootstrap values of 99 (ML) and 97 (MP), and monophyly of the Pygopodidae and Subfamily Diplodactylinae (*Carphodactylus* and *Strophurus*), with bootstrap values of 100 (ML) and 99 (MP).

Monophyly of the skinks is strongly supported by bootstrapping, 100 (ML and MP), and the two lygosomine lineages recognized by Greer (1979), the *Egernia/Mabuya* and *Sphenomorphus* groups, are also recovered by both analyses with strong bootstrap support, 98 (ML) and 99 (MP) for the *Egernia/Mabuya* lineage and 86 (ML) and 85 (MP) for the *Sphenomorphus*

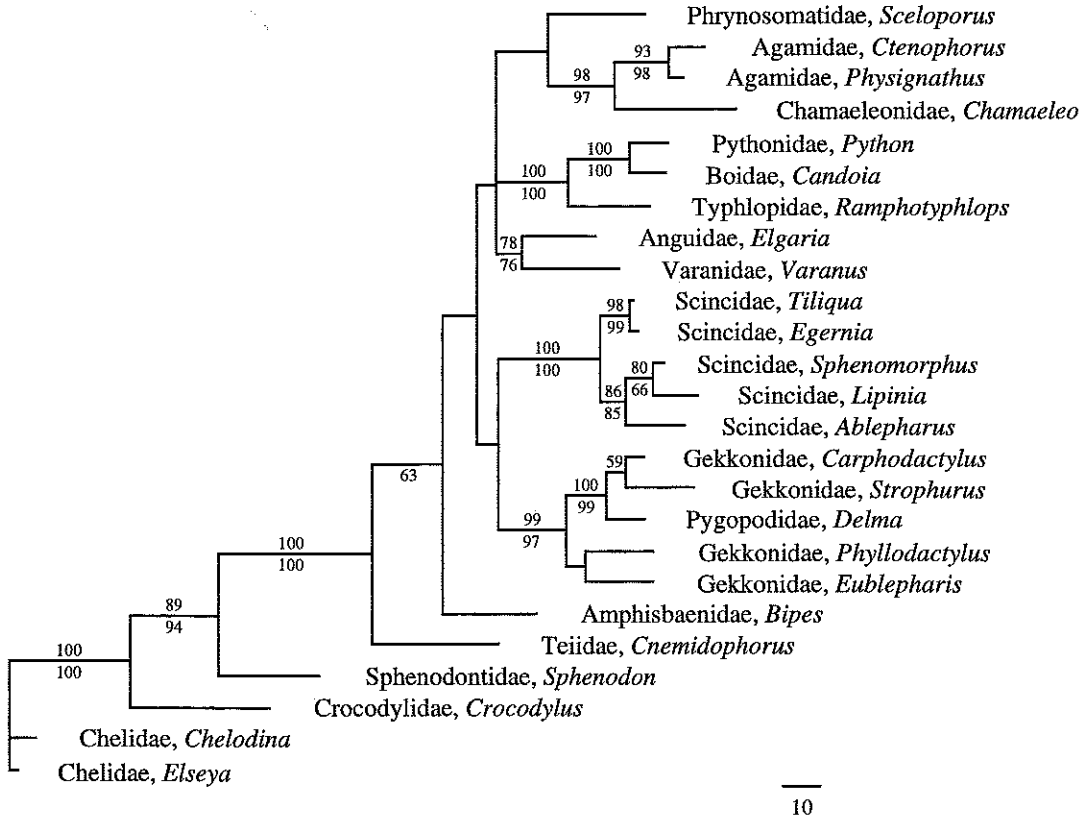

FIG. 2. Consensus of two equally most likely maximum likelihood trees of relationships among c-*mos* nucleotide sequences from 25 reptiles (Family and Genus names are given). Numbers at nodes represent node frequencies among 100 (ML—above) and 500 (MP—below) bootstrap pseudoreplicates. Scale indicates the expected number of substitutions along the branch.

lineage. Monophyly of the two snake lineages is strongly supported, with bootstrap values of 100 (ML and MP). Monophyly of the Superfamily Booidea (McDowell, 1987) is also strongly supported by both analyses with bootstrap values of 100. Monophyly of the two anguimorphan families, Anguidae and Varanidae, was recovered by both analyses with weaker bootstrap support of 78 (ML) and 76 (MP).

What is clear from Fig. 2 is that relationships among the major squamate lineages are not fully resolved by these data. This is most likely a result of the very short branch lengths of these divergences, i.e., rapid cladogensis, which is probably why historically there has been so much confusion and debate about higher-level squamate relationships. Whether additional sequence data from the remainder of the c-*mos* gene would resolve these relationships remains to be seen.

Phylogenetic analysis of c-*mos* sequences demonstrates the utility of this gene for recovering evolutionary relationships among and within squamate families. Phylogenetic analysis reveals a high degree of phylogenetic information contained within the sequence for both synonymous and nonsynonymous substitutions. The single copy, intronless nature of the gene and the presence of conserved regions within the gene makes it very amenable to PCR from genomic template and direct sequence analysis. The gene has been found in amphibians, birds, reptiles, and mammals and so could be a very useful "pilot" gene for higher-level systematics within the vertebrate lineages, in much the same way that Meyer (1994) recommends the use of the mitochondrial cytochrome *b* gene for "first pass" phylogenetic analyses. Furthermore, the regions of the gene that flank the central portion that we have amplified appear to be more variable judging from our alignment of the available vertebrate sequences and therefore may be very useful for groups that diverged more recently. We are attempting presently to clone the c-*mos* gene from several squamates to determine sequence of these flanking regions for this purpose.

ACKNOWLEDGMENTS

We thank R. Crombie, A. Georges, R. Macey, C. Moritz, O. Ryder, and T. Reeder for providing some tissues and DNA samples. D. Swofford kindly provided access to PAUP*. This study was supported by Australian Research Council Grant A19602789 to S.C.D. and M.N.H. and a National Science Foundation postdoctoral fellowship (INT-9505429) to C. C. A.

REFERENCES

Besansky, N. J., and Fahey, G. T. (1997). Utility of the *white* gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). *Mol. Biol. Evol.* **14:** 442–454.

Cooper, A., and Penny, D. (1997). Mass survival of birds across the Cretaceous-Tertiary boundary: Molecular evidence. *Science* **275:** 1109–1113.

Evans, S. E., and Barbadillo, J. (1997). Early Cretaceous lizards from Las Hoyas, Spain. *Zool. J. Linn. Soc.* **119:** 23–49.

Estes, R., de Queiroz, K., and Gauthier, J. (1988). Phylogenetic relationships within Squamata. *In* "Phylogenetic Relationships of the Lizard Families" (R. Estes and G. Pregill, Eds.), pp. 119–281. Stanford University Press, Stanford.

Friedlander, T. P., Regier, J. C., and Mitter, C. (1994). Phylogenetic information content of five nuclear gene sequences in animals: Initial assessment of character steps from concordance and divergence studies. *Syst. Biol.* **41:** 483–490.

Gow, C. E. (1975). The morphology and relationships of *Youngina capensis* Broom and *Prolacerta broomi* Parrington. *Paleontol. Afrikana* **18:** 89–131.

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

Greer, A. E. (1979). A phylogenetic subdivision of Australian skinks. *Rec. Aust. Mus.* **32:** 339–371.

Hillis, D. M., Mable, B., Larson, A., Davis, S., and Zimmer, E. (1996). Nucleic acids: Sequencing and cloning. *In* "Molecular Systematics" (D. Hillis, C. Moritz and B. Mable, Eds.), pp. 321–381. Sinauer, Sunderland, MA.

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

Jukes, T. H., and Cantor, C. R. (1969). Evolution of protein molecules. *In* "Mammalian Protein Metabolism" (H. N. Munro, Ed.), pp. 21–132. Academic Press, New York.

Kumar, S., Tamura, K., and Nei, M. (1993). "MEGA: Molecular Evolutionary Genetics Analysis," Version 1.0. The Pennsylvania State University, University Park, PA.

Leviton, A. E., Gibbs, R. H., Heal, E., and Dawson, C. E. (1985). Standards in herpetology and ichthyology: Part I. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. *Copeia* **1985:** 802–832.

Macey, J. R., Larson, A., Ananjeva, N. B., Fang, Z., and Papenfuss, T. J. (1997a). Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.* **14:** 91–104.

Macey, J. R., Larson, A., Ananjeva, N. B., and Papenfuss, T. J. (1997b). Evolutionary shifts in three major structural features of the mitochondrial genome among iguanian lizards. *J. Mol. Evol.* **44:** 660–674.

McDowell, S. B. (1987). Snake ecology and evolutionary biology. *In* "Systematics" (R. A. Seigel, J. T. Collins, and S. S. Novak, Eds.), pp. 3–50. McGraw-Hill, New York.

Meyer, A. (1994). Shortcomings of the cytochrome b gene as a molecular marker. *TREE* **9:** 278–280.

Mitchell, A., Cho, S., Regier, J. C., Mitter, C., Poole, R. W., and Matthews, M. (1997). Phylogenetic utility of elongation factor-1a in Noctuoidea (Insecta: Lepidoptera): The limits of synonymous substitutions. *Mol. Biol. Evol.* **14:** 381–390.

Reisz, R. (1981). A diapsid reptile from the Pennsylvanian of Kansas. *Univ. Kansas Mus. Nat. Hist. Special Pub.* **7:** 1–74.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Watson, R., Oskarson, M., and Vande Woude, G. F. (1982). Human DNA sequence homologous to the transforming gene (*mos*) of Moloney murine sarcoma virus. *Proc. Natl. Acad. Sci. USA* **79:** 4078–4082.

Yew, N., Strobel, M., and Vande Woude, G. F. (1993). Mos and the cell cycle: the molecular basis of the transformed phenotype. *Curr. Opin. Genet. Devel.* **3:** 19–25.