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

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:
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 G73 and G74 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 MgCl₂, 5 µl 10x 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+, Amer sham). 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 EcoRI-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 HindII- and PstI-digested DNA from the individuals where two EcoRI bands were detected showed single bands when probed with the dig-labeled c-mos probe. The single PstI 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...
that there is phylogenetic signal at all codon positions. That saturation of third positions has not occurred and significantly different from straight lines, indicating the skewness test statistic, \( g \) at first and second codon positions.

Of the 237 variable sites, 117 are third codon positions and 120 are amino acid substitutions. The mean pair-wise comparisons for Jukes–Cantor genetic distance and the proportion of synonymous and nonsynonymous differences between taxa (Jukes and Cantor, 1969). The genetic distances and the 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
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 anguimorph 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 cladogenesis, 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 nucleotide 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


