HERPETOLOGICAL JOURNAL, Vol. 9, pp. 147-151 (1999)

SQUAMATE RELATIONSHIPS BASED ON C-*MOS* NUCLEAR DNA SEQUENCES

D. J. HARRIS[1], E. A. SINCLAIR[1], N. L. MERCADER[2], J. C. MARSHALL[1] AND K. A. CRANDALL[3]

[1] *Department of Zoology, Brigham Young University, 574 Widtsoe Building, Provo, Utah, 84602, USA*

[2] *Departament de Biologia Animal, Universitat de Barcelona, Avinguda Diagonal 645, Barcelona 08028, Spain*

[3] *Department of Zoology and Monte L. Bean Museum, Brigham Young University, 574 Widtsoe Building, Provo, Utah, 84602, USA*

Relationships among squamate families have classically been difficult to establish, with morphological characters being interpreted to give many different topologies. Here we combine new C-*mos* nuclear DNA sequence data with those already published to assess relationships of 19 families within the Squamata. Monophyly of all the families examined is upheld. Many relationships between families are estimated, although it appears there may have been rapid cladogenesis associated with the origins of the Squamata.

Key words: Phylogeny, squamates, C-*mos*

INTRODUCTION

Squamate relationships have remained contentious since Camp's (1923) "Classification of the Lizards". Despite extensive analyses based on morphological characters many relationships remain unknown. Most widely accepted are the relationships suggested by Estes *et al.* (1988), although the analysis has been criticized (Kluge, 1989), and alternative suggestions for relationships have been made using different morphological characters (Presch, 1988). Surprisingly, the advent of DNA sequence data has had little impact on our understanding of squamate relationships. Although many studies have examined inter-familial relationships (e.g. Hedges *et al.*, 1991; Harris *et al.*, 1998), these have been limited due to their use of mitochondrial DNA sequences, which are typically saturated before the divergence times necessary to estimate relationships across squamates.

Recently Saint *et al.* (1998) used a fragment of the nuclear gene C-*mos* to investigate relationships of Australian reptiles relative to their overseas relatives. They showed that C-*mos* was likely to be a single copy gene in squamates, had no introns, and that a fragment of about 400 base pairs could be amplified across many squamate families. Graybeal (1994) had already shown that C-*mos* might be phylogenetically informative among taxa that had diverged up to 400 mya. To estimate relationships across squamates, we have extended the number of families included, and compared the estimates of phylogeny produced from Maximum Parsimony (MP) and Maximum Likelihood (ML) with those previously derived from morphological characters.

Correspondence: D. J. Harris, Campus Agrário de Vairão, R. Monte-Crasto, 4480 Vila do Conde, Portugal. *E-mail*: james@mail.icav.up.pt

METHODS

The additional species examined were: F. Cordylidae: *Cordylus cordylus*; F. Gekkonidae: *Bunopus tuberculatus*, *Stenodactylus doriae*; F. Iguanidae: *Dipsosaurus dorsalis*, *Iguana iguana*; F. Lacertidae: *Acanthodactylus scutellatus, Lacerta kulzeri, Podarcis hispanica*; F. Trogonophidae; *Diplometophon zarudnyi*; F. Xantusidae: *Lepidophyma gaigae, Xantusia vigilis*. These were selected to cover five families not included by Saint *et al.* (1998), and to extend the number of the family Gekkonidae examined from one to three.

Total genomic DNA was extracted from small (1 or 2 mm^3) pieces of tail tissue. The material was finely diced and digested with proteinase K (Kocher *et al.*, 1989). Purification was by phenol/chloroform extractions (Sambrook *et al.*, 1989), followed by centrifugal dialysis through a Centricon 30000 MW membrane (Amicon). Polymerase Chain Reaction (PCR) primers used in both the amplification and the sequencing were G73 and G74 (Saint *et al.*, 1998). PCR conditions were the same as those used by Saint *et al.* (1998). Successful PCR products were purified using a Qiaex II kit (Qiagen), and sequenced from both strands on an Applied Biosystems DNA Sequencing System.

SEQUENCE ANALYSIS

Genbank accession numbers are AF148702 to AF148712. The sequences were aligned by eye to the previously published sequences (Genbank AF039462 to AF039482) of Saint *et al.* (1998). The aligned sequences were 375bp long. The codon reading frame was infered by comparison with the published sequences. Of the new sequences, all the lacertids and the two geckos had a deletion of seven codons, and the *Diplometophon* had an eight codon deletion. These were in the same region (bp 727-768 of human C-*mos*

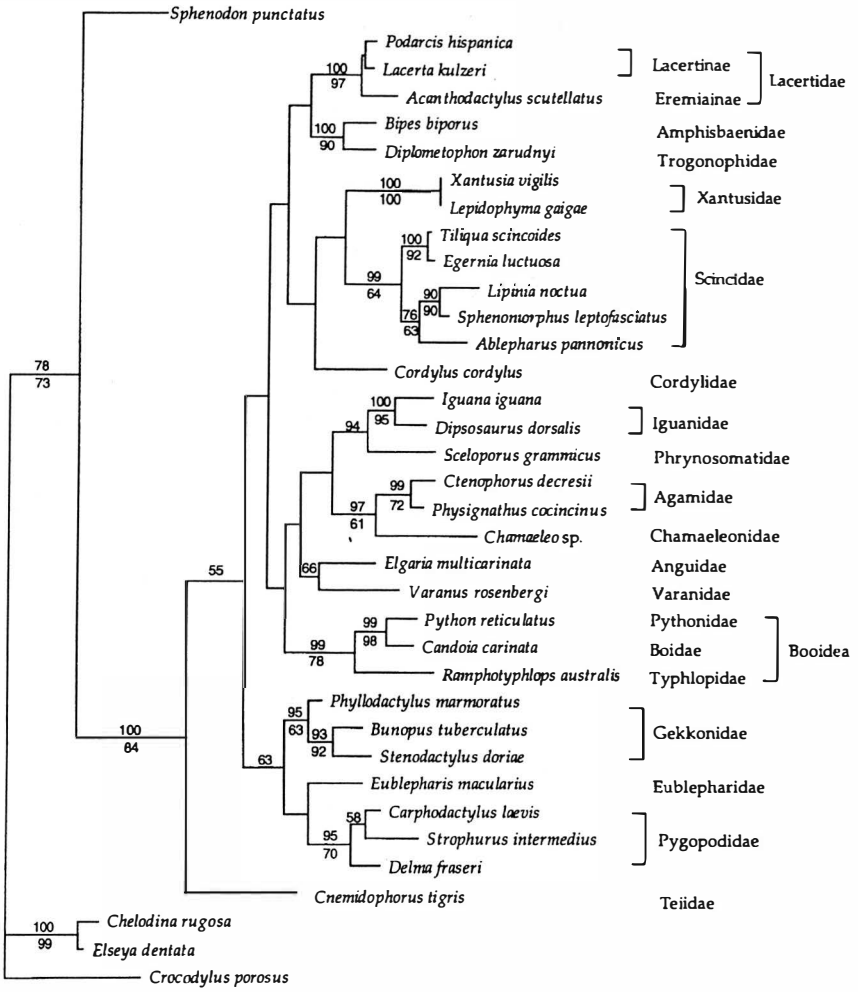

FIG. 1. Single most parsimonious tree derived from an analysis of C-*mos* nucleotide sequence. Numbers above branches indicate bootstrap support (1000 replicates). Numbers below branches indicate boostrap support from an MP analysis based on the amino acid sequence, with all changes weighted equally. See text for details. The tree was rooted using the *Crocodylus, Chelodina* and *Elseya* sequences.

sequence), and overlapped deletions also found in the skink *Lipinia noctua* and the teiid *Cnemidophorus tigris*. They were therefore treated as missing data in the analyses.

The data were analysed using PAUP* (Swofford, 1998). When estimating phylogenetic relationships among sequences, one assumes a model of evolution regardless of the optimality criteria employed. Determining which model to use given the data is a statistical problem (Goldman, 1993). We used the approach outlined by Huelsenbeck & Crandall (1997) to test alternative models of evolution, employing PAUP* and Modeltest (Posada & Crandall, 1998). A starting tree was obtained using neighbour-joining. With this tree, likelihood scores were calculated for various models of evolution and then compared statistically using a chi-square test with degrees of freedom equal to the difference in free parameters between the models being tested. The null hypotheses tested in this way included: (1) nucleotide frequencies are equal; (2) transition rates are equal to transversion rates; (3) transition rates are equal and transversion rates are equal; (4) rate homogeneity exists within the data set; and (5) there is no significant proportion of invariable sites. Once a model of evolution was chosen, it was used to estimate a tree using maximum likelihood (Felsenstein, 1981), using random sequence addition and a heuristic search with 10 replicates. Also an MP analysis was performed. Two hundred and nine of the 375 characters were parsimony-informative. A 10 replicate heuristic search was carried out, and support for nodes was estimated using the bootstrap (Felsenstein, 1985) technique, with 1000 replicates. A further MP analysis was carried out on the translated amino acid sequences. All changes were weighted equally.

RESULTS

Using MP, 209 of the 375 characters were parsimony-informative. A 10 replicate heuristic search found one MP tree with 892 steps. (CI= 0.46, HI= 0.54). Support for nodes was estimated using the bootstrap (Felsenstein, 1985) technique, with 1000 replicates (Fig. 1). In the translated amino acid sequence, 60 characters were informative. A ten replicate heuris-

TABLE 1. Tests of hypotheses relating to the model of evolution appropriate for phylogeny reconstruction (Huelsenbeck and Crandall, 1997). *P*-values were obtained using the computer program Modeltest (Posada & Crandall, 1998). Due to the performance of multiple tests, the significance level of rejection of the null hypothesis should be adjusted via the Bonferroni correction to $\alpha = 0.01$.

Null Hypothesis	Models Compared	$-\ln L_0$	$-\ln L_1$	df	*P*
Equal base frequencies	H_0: JC69, H_1: F81	5010.1	5005.5	3	0.012
Equal ti/tv rates	H_0: JC69, H_1: K80	5010.1	4763.8	1	<0.001
Equal ti and equal tv rates	H_0: K80, H_1: GTR	4763.8	4761.6	3	0.340
Equal rates among sites	H_0: K80, H_1: K80+G	4763.8	4582.3	1	<0.001
Proportion of invariable sites	H_0: K80+G, H_1: K80+G+invar	4582.3	4568.9	1	<0.001
Molecular clock	H_0: no rate heterogeneity, H_1: rate heterogeneity	4635.8	4568.9	34	<0.001

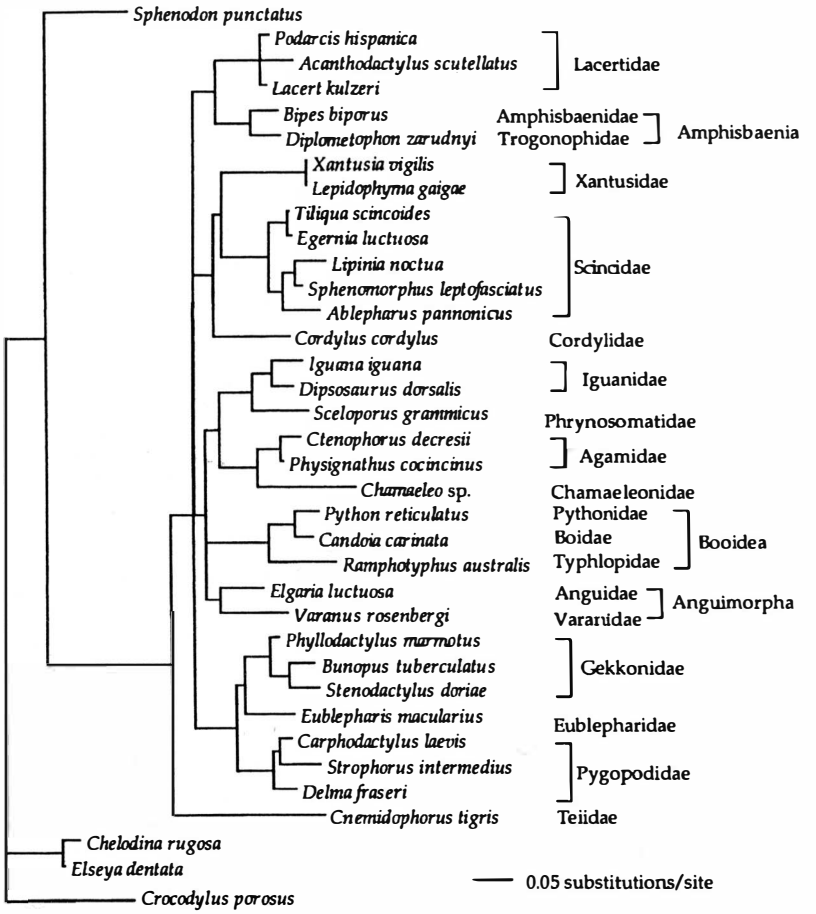

FIG. 2. Single maximum likelihood tree, derived using the K80 model with estimation of the proportion of invariant sites and a discreet approximation of the gamma distribution. See text for details.

tic search found three equally parsimonious trees of 376 steps (CI=0.61, HI=0.39). Support for nodes was again estimated using 1000 bootstrap replicates (Fig. 1). With ML, using Modeltest (Posada & Crandall, 1998) we concluded that the Kimura 80 model (transition/ transversion ratio = 2.6584), with a gamma distributed rate heterogeneity model (a= 3.0825), and an estimated proportion of invariable sites (0.2905) was the most appropriate model of evolution for these data. The data did not fit a molecular clock (Table 1). A ten replicate heuristic search using random sequence addition with this model produced a single maximum likelihood tree of score -ln 4568.9 (Fig. 2).

DISCUSSION

Analysis of our extended data set supports many of the conclusions drawn by Saint *et al.* (1998). The analyses based on C-*mos* sequences support the monophyly of the squamates, and that the closest living relative is *Sphenodon punctatus*. Within the squamates, all the superfamilies and families where multiple species were sampled came out as monophyletic groups - Agamidae (99% bootstrap support from MP tree), Amphisbaenia (100%), Booidea (99%), Gekkonidae (95%), Iguanidae (100%), Lacertidae (100%), Pygopodidae (95%), Scincidae (99%) and Xantusidae (100%). In the analy-

ses of the nucleotide sequences, the teiid *Cnemidophorus tigris* comes out basal to all other squamates, although this is not the case when the amino acid sequence is analysed. Based on morphological characters, teiids are usually regarded as the sister taxa to lacertids (e.g. Estes *et al.*, 1988). The basal position in this analysis could be due to the presence of a paralogous sequence in teiids, or it could be due to an artifact in the data such as long branch attractions (Felsenstein, 1978), or due to massive convergence in the morphological characters. Long branch attraction could be due to rate variation or inadequate sampling. Taxon sampling should not be a problem, as we have included C-*mos* sequences of lacertids, which are thought to be closely related to teiids (Estes *et al.*, 1988). Rate variation cannot be ruled out, as the data do not fit a molecular clock (Table 1), and it is clear from the ML analysis (Fig. 2) that *Cnemidophorus* has the longest external branch of all the squamates sampled. Since its position is only weakly supported (55% bootstrap in MP tree), and since the branches immediately above its position are extremely short, it cannot be placed with much confidence by this data set.

Based on morphological characters, the Scincomorpha is thought to include Scincidae, Cordylidae, Xantusidae, Lacertidae, Teiidae and Gymnophthalmidae (Estes *et al.*, 1988), with some authors suggesting that the amphisbaenians should be included (e.g. Schwenk, 1988). Excluding Gymnopthalmidae, which was not sampled, and *Cnemidophorus tigris*, these taxa are also associated by the MP analysis, with the Xantusidae being the sister taxon to the Scincidae, and with the next closest relative being the Cordylidae. These are the same relationships suggested by Presch (1988) based on morphological characters. The two amphisbaenians included, *Bipes biporus* and *Diplometophon zarudnyi*, are strongly grouped as monophyletic (100%), and appear to be the sister taxa to the Lacertidae. Evidence from amphisbaenian fossils also suggests they may be members of the Scincomorpha (Wu *et al.*, 1996). Within the Lacertidae, the monophyly of the subfamily Lacertinae - *Lacerta kulzeri* and *Podarcis hispanica* (Harris *et al.*, 1998) is weakly supported in the MP analysis.

The two Iguanids included, *Iguana iguana* and *Dipsosaurus dorsalis,* are strongly associated with the phrynosomatid *Sceloporus grammicus*, and this is also supported by morphology (Estes *et al.*, 1988). Most closely related to these is the clade made up of the Agamidae and Chamaeleonidae, again something found using morphological characters (Estes *et al.*, 1988). Monophyly of the two anguimorph families Anguidae and Varanidae was recoved by both ML and MP, but with low bootstrap support (66%).

Saint *et al.* (1998) labelled the subfamily Diplodactylinae (*Carphodactylus* and *Strophorus*) as members of the Gekkonidae. Kluge (1987) included the Diplodactylinae in the Pygopodidae because of a shared derived character of the muscle encircling the external ear opening. C-*mos* sequences support this, with the Diplodactylinae being sister group to the Pygopodidae (*Delma*) in both the ML and MP analysis (95% bootstrap support).

One difference between the MP and ML analyses was in the placement of *Eublepharis macularius*. ML analyses associate it with the Gekkonidae, while the MP analysis places it as sister taxon to the Pygopodidae. Morphological characters suggest that it is basal to a clade of these two groups (Grismer, 1988). While C-*mos* sequences clearly group Gekkonidae with Pygopodidae and Eublepharidae, the exact relationship between these three groups remains unresolved.

Most of the other intra-familial relationships are extremely weakly supported, as shown by very short internal branches in the ML analysis. As suggested by Saint *et al.* (1998), this could be the result of rapid cladogenesis, or simply a result of the limitations of using only one gene region to examine relationships. Only the inclusion of more sequence data will help to resolve this, although it is clear that C-*mos* is an extremely useful gene for examining many aspects of squamate relationships.

ACKNOWLEDGEMENTS

We thank E. N. Arnold, R. L. Bezy and J. W. Sites Jr. for supplying us with some of the tissues used. This work was supported by NSF IBN-97-02338 and the Alfred P. Sloan Foundation (to KAC). We also thank an anonymous reviewer who gave constructive criticisms of an earlier draft of this paper.

REFERENCES

Camp, C. L. (1923). Classification of the lizards. *Bulletin of the American Museum of Natural History* **48**, 289-481.

Estes, R., de Queiroz, K. & Gauthier, J. (1988). Phylogenetic Relationships within Squamata. In *Phylogenetic Relationships of the Lizard Families.* Estes, R. & Pregill, G. (Eds.). Stanford: Stanford University Press.

Felsenstein, J. (1978). Cases in which parsimony or compatability methods will be positively misleading. *Systematic Zoology* **27**, 401-410.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Goldman, N. (1993). Simple diagnostic statistical tests of models of DNA substitution. *Journal of Molecular Evolution* **37**, 650-651

Graybeal, A. (1994). Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Systematic Biology* **43**, 174-193.

Grismer, L. L. (1988). Phylogeny, taxonomy, classification and biogeography of Eublepharid geckos. In *Phylogenetic Relationships of the Lizard Families*. Estes, R. & Pregill, G. (Eds.). Stanford: Stanford University Press.

Harris, D. J., Arnold, E. N. & Thomas, R. H. (1998). Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. *Procedings of the Royal Society London B.* **265**, 1939-1948.

Hedges, S. B., Bezy, R. L. & Maxson, L. R. (1991). Phylogenetic relationships and biogeography of xantusiid lizards, inferred from mitochondrial DNA sequences. *Molecular Biology and Evolution* **8**, 767-780.

Huelsenbeck, J. P. & Crandall, K. A. (1997). Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**, 437-466.

Kluge, A. G. (1987). Cladistic relationships in the Gekkonidae (Squamata, Sauria). *Miscellaneous publications of the Museum of Zoology, University of Michigan* **173**, 1-54.

Kluge, A. G. (1989). Progress in squamate classification. *Herpetologica* **45**, 368-379.

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. *Proceedings of the National Academy of Sciences USA* **86**, 6196-6200.

Posada, D. & Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**(10).

Presch, W. (1988). Phylogenetic relationships of the Scincomorpha. In *Phylogenetic Relationships of the Lizard Families*. Estes, R. & Pregill, G. (Eds.). Stanford: Stanford University Press.

Saint, K. M., Austin, C. C., Donnellan, S. C. & Hutchinson, M. N. (1998). C-*mos*, a nuclear marker useful for squamate phylogenetic analysis. *Molecular Phylogenetics and Evolution* **10**, 259-263.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* (2nd ed.). New York: Cold Spring Harbour Press.

Schwenk, K. (1988). Comparative morphology of the Lepidosaur tongue and its relevance to squamate phylogeny. In *Phylogenetic Relationships of the Lizard Families*.Estes, R. & Pregill, G. (Eds.). Stanford: Stanford University Press.

Swofford, D.L. (1999). PAUP. Phylogenetic analysis using parsimony (and other methods). Sunderland, Mass: Sinauer Associates.

Wu, X.-C., Brinkman, D. B. & Russell, A. P. (1996). *Sineoamphisbaena hexatabularis*, an amphisbaenian (Diapsida: Squamata) from the Upper Cretaceous redbeds at Bayan Mandahu (Inner Mongolia, People's Republic of China), and comments on the phylogenetic relationships of the Amphisbaenia. *Canadian Journal of Earth Science* **33**, 541-577.

Accepted: 5.5.99