Short interspersed elements (SINEs) of squamate reptiles (Squam1 and Squam2): Structure and Phylogenetic Significance

VERNATA V. GRECHKO1, SERGEI A. KOSUSHKIN1, OLGA R. BORODULINA1, FATIMA G. BUTAEVA1, AND ILYA S. DAREVSKY2

1Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia
2Zoological Institute, Russian Academy of Sciences, Saint-Petersburg, Russia

Short interspersed elements (SINEs) are important nuclear molecular markers of the evolution of many eukaryotes. However, the SINEs of squamate reptile genomes have been little studied. We first identified two families of SINEs, termed Squam1 and Squam2, in the DNA of meadow lizard *Darevskia praticola* (Lacertidae) by performing DNA hybridization and PCR. Later, the same families of retrotransposons were found using the same methods in members of another 25 lizard families (from Iguania, Scincomorpha, Gekkota, Varanoidea, and Diploglossa infraorders) and two snake families, but their abundances in these taxa varied greatly. Both SINEs were Squamata-specific and were absent from mammals, birds, crocodiles, turtles, amphibians, and fish. Squam1 possessed some characteristics common to tRNA-related SINEs from fish and mammals, while Squam2 belonged to the tRNA*Ala* group of SINEs and had a more unusual and divergent structure. Squam2-related sequences were found in several unannotated GenBank sequences of squamate reptiles. Squam1 abundance in the Polychrotidae, Agamidae, Leiolepididae, Chamaeleonidae, Scincidae, Lacertidae, Gekkonidae, Varanidae, Helodermatidae, and two snake families were $10^2$–$10^4$ times higher than those in other taxa (Corytophanidae, Iguanidae, Anguidae, Cordylidae, Gerrhosauridae, Pygopodidae, and Eublepharidae). A less dramatic degree of copy number variation was observed for Squam2 in different taxa. Several Squam1 copies from Lacertidae, Chamaeleonidae, Gekkonidae, Varanidae, and Colubridae were sequenced and found to have evident orthologous features, as well as taxa-specific autapomorphies. Squam1 from Lacertidae and Chamaeleonidae could be divided into several subgroups based on sequence differences. Possible applications of these SINEs as Squamata phylogeny markers are discussed.

How to cite this article: Grechko VV, Kosushkin SA, Borodulina OR, Butaeva FG, Darevsky IS. 2011. Short interspersed elements (SINEs) of squamate reptiles (Squam1 and Squam2): structure and phylogenetic significance. J. Exp. Zool. (Mol. Dev. Evol.) 316:212–226.

Short interspersed elements (SINEs, short retroposons) are reliable and informative molecular markers of eukaryotic evolution (Shimamura et al., ’97; Borodulina and Kramerov, ’99; Deininger and Batzer, 2002; Salem et al., 2003; Shedlock et al., 2009). Their length in different genomes varies from ~100 to 500 bp, and they are repeated several to tens of thousands of times per haploid genome (Kramerov and Vassetzky, 2005, 2009). Their most significant advantage of using SINEs as a phylogenetic tool is that individual copies are created in evolution only once per
genome. They cannot be removed, as no precise excision mechanism for multiple copies is known (Serdobova and Kramerov, 1998; Kramerov and Vassetzky, 2005). Thus, the increases in the SINE copy number afforded by SINE dispersion into new loci and SINE divergence resulting from mutation are evolutionarily irreversible and seem to provide, as suggested by some authors, nearly homoplasy-free characteristics with low probability of reversal and convergence (Hillis, 1999; Shedlock and Okada, 2000; Salem et al., 2003). Therefore, taxa that lack SINEs or that have low copy number SINEs could be assumed to have appeared earlier in evolution than related taxa with higher copy number SINEs.

All copies of each SINE seem to have been amplified from one or several “master genes” (Shen et al., 1991; Okada et al., 2004) and to have evolved independently. Thus, the divergence of different copies reflects the relative age of taxa (Shedlock et al., 2009). Retroposon characteristics and their application to the phylogenetic study of fish and mammals (particularly rodents and primates) have been reviewed at length (Shimamura et al., 1997; Stoneking et al., 1997; Serdobova and Kramerov, 1998; Borodulina and Kramerov, 1999; Hills, 1999; Roos et al., 2004; Kramerov and Vassetzky, 2005). However, their application to other animal taxa, including reptiles, has attracted less attention. Only a turtle-specific SINE has been identified (Sasaki et al., 2004), and a family of short Squamata-specific retroposons was discovered independently by Piskurek et al. (2006, 2009) and by us (Kosushkin et al., 2006, 2008) from the DNA of the lacertid lizards Podarcis muralis and Darevskia praticola, respectively.

The application of retroposons to Squamata phylogeny could be used to verify controversial inferences from other molecular marker results (see review Grechko, 2002). During the past century, more than ten retroposon systems were constructed based on morphological (including paleontological) data (see reviews Alifanov, 2000; Evans, 2003; Conrad, 2008) or nuclear and mitochondrial markers (Harris et al., 2001; Albert et al., 2009; Vidal and Hedges, 2009). However, reptile taxonomy contains many controversial postulates for system construction, as well as many inter- and intra-taxa shuffles of genera and families (Alifanov, 2000; Harris et al., 2001; Whiting et al., 2003; Townsend et al., 2004). These uncertainties even include the general subdivision of Squamata into Iguania and Scleroglossa (Vidal and Hedges, 2009) and the generally accepted position of Acrodonta inside Iguania (Estes et al., 1988; Harris et al., 2001; Evans, 2003; Vitt et al., 2003; Böhme et al., 2007; Conrad, 2008; Albert et al., 2009). The Squamata order was recently revised to accommodate a large pool of current paleontological findings. Regardless, modern taxonomy does not reflect the real multitude and diversity of lizard and snake taxa, which are perpetually growing (Alifanov, 2000). New molecular markers such as SINEs would be useful to study the correlation of the genetic and morphological foundations of reptile relationships.

In this work, two families of tRNA-derived SINE retroposons, termed Squam1 and Squam2, were isolated and characterized. All currently available GeneBank Squam1-like unannotated sequences were also included in the analysis. Without endeavoring to solve all the difficulties in Squamata taxonomy, these markers were used to evaluate the possible evolutionary pathway of some taxa relative to others through the distribution and abundance of Squam1. Using this approach, searches for synapomorphies in Squam1 sequences could be used to identify closely related taxa within Squamata.

**MATERIALS AND METHODS**

**Samples and DNA Extraction**

DNA was extracted from blood or other tissue samples by standard proteinase K lysis followed by phenol/chloroform purification (Sambrook et al., 1989). DNA concentrations were determined with a NanoDrop® ND-1000 spectrophotometer. Blood samples of some lizards, snakes, and crocodiles were provided by Dr. D.B. Vasiliev (Moscow Zoo); those of other animals were collected in the field. Turtle DNA was provided by Dr. S.K. Semenova (Institute of Gene Biology, Moscow). Other nonreptilian DNA was provided by Dr. D.A. Kramerov (Engelhardt Institute of Molecular Biology, Moscow). A list of the species is shown in Table 1. In most experiments, DNA and blood from the same specimen were used.

**PCR Conditions**

PCR was used to amplify a hybridization probe and identify SINE sequences in the genomic DNA. The A-B PCR method was performed as described previously (Borodulina and Kramerov, 2005). Primers A (5'-TRGCTCAGTGG-3') and B (5'-GGRATYGA-ACY-3') were used, which corresponded to the A and B boxes of the tRNA PolIII promoter. The reaction was run mostly with 0.1–1.0 ng of D. praticola lizard genomic DNA for 27–33 cycles at 95°C for 1 min, 37°C for 1 min, and 72°C for 1 min.

For Squam1-specific primer pairs, Lac1 (5'-CAGAGCC-TAGGGCTTGC-3') and Lac2 (5'-AACCGCGTTTACCTTCCC-3') were used with PCR to prepare a hybridization probe. The additional primers Cor1 (5'-GCTCTGTGCAACCCTAGCA-3'), Cor2 (5'-GCTCATGTGGCCAGCATGA-3'), and Cor3 (5'-CGAGCAGGRT-GAGCTCCC-3') were used in various combinations along with Lac1 and Lac2 to detect Squam1 in genomic DNA. The Squam2 primers were described previously (Kosushkin et al., 2006). The PCR conditions were the same as those mentioned above, except that the annealing temperature was 53°C. Products were separated by electrophoresis on 3% agarose (Sigma) and visualized by ethidium bromide staining.

**Library Construction and Screening**

The A-B PCR products were cloned into the pGEM-T vector (Sambrook et al., 1989). The EcoRI and HindIII genomic DNA fragments of D. praticola were sized in a 3% WideRange agarose gel. Fragments of 400–1,000 bp were extracted and subsequently
Table 1. List of Squamata taxa studied by us (in bold), Piskurek et al. (2006) (in light), and taken from unannotated GeneBank sequences (enumerated in comments).

<table>
<thead>
<tr>
<th>Infraorder sensu Estes et al. ('88)</th>
<th>Family</th>
<th>Species</th>
<th>Abbr.</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iguania</td>
<td>Iguanida</td>
<td><em>Iguana iguana</em> Linnaeus 1758</td>
<td>Iig</td>
<td>Common green iguana</td>
</tr>
<tr>
<td></td>
<td>Cycluridae</td>
<td><em>Cyclura cornuta</em> Bonnaterre 1789</td>
<td>Cco</td>
<td>Rhinoceros iguana</td>
</tr>
<tr>
<td>Phrynosomatidae</td>
<td>Phrynosomatidae</td>
<td><em>Phrynosoma platyrhinos</em> Girard 1852</td>
<td>Ppl</td>
<td>Desert horned lizard</td>
</tr>
<tr>
<td>Corytophanidae</td>
<td>Corytophanidae</td>
<td><em>Anolis equestris</em> Merrem 1820</td>
<td>Aeq</td>
<td>Knight anole</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Anolis carolinensis</em> Linnaeus 1758</td>
<td>Aca</td>
<td>Green anole</td>
</tr>
<tr>
<td>Polychrotidae</td>
<td>Polychrotidae</td>
<td><em>Chamaeleolis barbatis</em> Garrido 1982</td>
<td>Cba</td>
<td>Bearded false chameleon</td>
</tr>
<tr>
<td>Agamida</td>
<td>Agamida</td>
<td><em>Phrynocephalus helioscopus</em> Pallas 1771</td>
<td>Phe</td>
<td>Sun-gazing agama</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Chlamydosaurus kingii</em> Gray 1827</td>
<td>Cki</td>
<td>Frillneck lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Hypsibius modestus</em> Meyer 1874</td>
<td>Hmo</td>
<td>Forest dragon</td>
</tr>
<tr>
<td>Leiolepididae</td>
<td>Leiolepididae</td>
<td><em>Leiolepis quentherpetesi</em> Daresky &amp; Kupriyanova 1993</td>
<td>Lgu</td>
<td>Peters’ butterfly lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Leiolepis reevesi</em> Gray 1831</td>
<td>Lre</td>
<td>Reeves’ butterfly lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Uromastyx acanthinura</em> Bell 1825</td>
<td>Uac</td>
<td>North African spiny-tailed lizard</td>
</tr>
<tr>
<td>Chamaeleonidae</td>
<td>Chamaeleonidae</td>
<td><em>Chamaeleo calyptratus</em> Dumeril 1851</td>
<td>Cca</td>
<td>Veiled chameleon</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Chamaeleo feae</em> Boulenger 1906</td>
<td>Cfe</td>
<td>Chameleon</td>
</tr>
<tr>
<td>Gekkota</td>
<td>Gekkonidae</td>
<td><em>Gekko gecko</em> Linnaeus 1758</td>
<td>Gge</td>
<td>Tokay gecko</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Gekko vittatus</em> Houttuyn 1782</td>
<td>Gvi</td>
<td>Lined gecko</td>
</tr>
<tr>
<td>Eublepharidae</td>
<td>Eublepharidae</td>
<td><em>Goniurosaurus sp.</em></td>
<td>Gos</td>
<td>Asian gecko</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Eublepharis macularius</em> Blyth 1854</td>
<td>Ema</td>
<td>Leopard gecko</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Hemithyconyx caudicinctus</em> Dumeril 1851</td>
<td>Hca</td>
<td>African fat tailed gecko</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Lialis burtonis</em> Gray 1835</td>
<td>Lbu</td>
<td>Burton's legless lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Lialis jicari</em> Boulenger 1903</td>
<td>Lji</td>
<td>Papua snake lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dibamus tiananensis</em> Diaz, Leong, Grismer, and Yaakob, 2004</td>
<td>Dta</td>
<td>Dibamus</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Dibamus tiananensis</em> Diaz, Leong, Grismer, and Yaakob, 2004</td>
<td>Dta</td>
<td>Dibamus</td>
</tr>
<tr>
<td>Lizards</td>
<td>Scincomorpha</td>
<td><em>Tiliqua sp.</em></td>
<td>Tis</td>
<td>Blue-tongued skink</td>
</tr>
<tr>
<td></td>
<td>Scincidae</td>
<td><em>Typhosaurus caucus</em> Cuvier 1817</td>
<td>Tca</td>
<td>Cuvier’s legless skink</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Gnypetoscinicus queenslandiae</em> De Vis 1890</td>
<td>Gqu</td>
<td>Prickly rainforest skink</td>
</tr>
<tr>
<td></td>
<td>Cordylidae</td>
<td><em>Cordylus giganteus</em> Smith 1844</td>
<td>Cgi</td>
<td>Giant sungazer lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Cordylus cataphractus</em> Boie 1828</td>
<td>Cct</td>
<td>Armadillo lizard</td>
</tr>
<tr>
<td></td>
<td>Gerrhosauridae</td>
<td><em>Gerrhosaurus sp.</em></td>
<td>Ges</td>
<td>Plated lizard</td>
</tr>
<tr>
<td></td>
<td>Lacertidae</td>
<td><em>Galatia sp.</em></td>
<td>Gas</td>
<td>Western Canaries lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Darevskia praticola</em> Eversmann 1834</td>
<td>Dpr</td>
<td>Meadow lizard</td>
</tr>
<tr>
<td></td>
<td>Teiidae (macro)</td>
<td><em>Podarcis taurica</em> Pallas 1814</td>
<td>Pta</td>
<td>Common wall lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Podarcis muralis</em> Laurenti 1768</td>
<td>Pmu</td>
<td>Common wall lizard</td>
</tr>
<tr>
<td></td>
<td>Lacertidae</td>
<td><em>Lacerta strigata</em> Eichwald 1831</td>
<td>Lst</td>
<td>Emerald lizard</td>
</tr>
<tr>
<td></td>
<td>Xantusiidae</td>
<td><em>Xantusia vigilis</em> Baird 1859</td>
<td>Xvi</td>
<td>Desert night lizard</td>
</tr>
<tr>
<td></td>
<td>Teiidae (micro)</td>
<td><em>Teindemis teguixin</em> Linnaeus 1758</td>
<td>Tei</td>
<td>Common tegu</td>
</tr>
<tr>
<td></td>
<td>Anguidae</td>
<td><em>Chormidophorus lemniscatus</em> Linnaeus 1758</td>
<td>Cle</td>
<td>Rainbow whiptail</td>
</tr>
<tr>
<td></td>
<td>Anguidae</td>
<td><em>Anguis fragilis</em> Linnaeus 1758</td>
<td>Afr</td>
<td>Slow-worm</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Ophisaurus apodus</em> Pallas 1775</td>
<td>Oap</td>
<td>European legless lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Ophisaurus attenuatus</em> Cope 1880</td>
<td>Oat</td>
<td>Slender glass lizard</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Gerrhonotus locophorus</em> Wiegmann 1828</td>
<td>Gli</td>
<td>Texas alligator lizard</td>
</tr>
<tr>
<td></td>
<td>Anniellidae</td>
<td><em>Anniella pulchra</em> Gray 1852</td>
<td>Apu</td>
<td>California legless lizard</td>
</tr>
<tr>
<td></td>
<td>Xenosauridae</td>
<td><em>Xenosaurus grandis</em> Gray 1856</td>
<td>Xgr</td>
<td>Knob-scaled lizard</td>
</tr>
</tbody>
</table>
cloned into the pGEM 3Z vector digested with the same restriction enzymes. Colony hybridization was performed at 60°C overnight in a mixture containing 4×SSC, 0.5% SDS, 5×Denhardt’s solution, 0.1 mg/mL boiled herring sperm DNA, and (α-32P)dATP-labeled probe. The probe was obtained by performing A-B PCR or PCR with Lac1 and Lac2 primers using D. praticola DNA as template.

**Dot-Blot DNA Hybridization**

One microgram of DNA was denatured in 10 μL of 0.5 M NaOH for 1 hr at 37°C. After incubation, 20 volumes of 6×SSC, 6% formaldehyde, and 0.025 M NaH2PO4 were added. The DNA mixture was applied to a Hybond-C Extra membrane (Amersham) using a dot-blot apparatus. The hybridization and washing conditions were the same as those used in colony hybridization. The probe was labeled by performing PCR with appropriate primers and (α-32P)dATP from Fosfor Center of Collective Use (Russian Academy of Sciences, Russia). Washing was performed at 42 and 52°C for 20 min. Results were visualized by either exposing the dot-blot membrane to an X-ray film or by scanning the membrane with a CyClone Phosphoimager. Quantification was performed with the OptiQuant software program.

**Sequencing and Sequence Analysis**

Cloned PCR products were sequenced either manually as described (Borodulina and Kramerov, 2005) or automatically using the ABI PRISM BigDye Terminator v. 3.1 kit on an ABI PRISM 3100-Avant. Sequences were manually aligned and edited with Genedoc (Nicholas and Nicholas, '97). A GenBank search for sequence similarity was performed with BLAST (Altschul et al., '90).

**NJ-Tree Construction**

The data were analyzed by the NJ method in the MEGA software package using 1,000 bootstrap replicas (Tamura et al., 2007). The maximum Composite Likelihood substitution model was chosen, and gaps were treated according to the Pairwise Deletion criterion.
RESULTS

Squam1 SINE

Several copies of a novel reptilian retroposon from the genomic library of the D. pratica lizards (fam. Lacertidae) were extracted by screening with a probe specific for PolIII promoter sites (A–B) that were homologous to SINEs from different taxa (Fig. 1A and Supplementary Fig. S1). The sequences were aligned, and primers (Lac1 and Lac2) corresponding to the most conserved central part of the consensus sequence were constructed and used to prepare a 188-bp DNA hybridization probe by PCR. Using this probe, we searched for and sequenced SINEs in genomic libraries of veiled chameleon (Chamaeleo calyptratus), emerald monitor (Varanus prasinus), tokay gecko (Gekko gecko), and dione ratsnake (Elaphe dione). About 4 to 12 complete copies were isolated from each species. All other SINE-like sequences were later found in GenBank and were included in our analysis (Figs. 1–3). Most of the full-length SINE copies were 330–380 bp long, and at least half of them were flanked by short direct repeats formed by retroposition insertion (target-site duplications, TSDs). The presence of the internal PolIII promoter boxes A and B, and of structural properties characteristic of other vertebrate SINES, suggested that the 5′-part of Squam1 was tRNA-derived. The downstream central parts were Squamata-specific and were more conserved than the tRNA-related 5′-end regions. At the very 3′-end, there were short arrays of one to four tandem repeats (underlined ACCTTT motif in Figs. 1–3), a characteristic of other eukaryotic SINEs (see Kramerov and Vassetzky, 2005).

The sequence alignments of the full-length Squam1 copies from several of the above-mentioned lizard families are shown in Figures 1–3 (in text and in Electronic Supplement). D. pratica lizards (Dpr, our data) and P. muralis (Pmu, 15 arbitrarily chosen sequences from the 54 available in GenBank; Piskurek et al., 2006) of fam. Lacertidae are represented in Figure 1A and Supplementary Figure S1. These two species are closely related morphologically, but were separated from Lacerta s. lato on the basis of certain morphological peculiarities. The alignment of Squam1 sequences from Dpr and Pmu showed a high level of similarity between these genera. Most of the lengths of these sequences contained long homologous regions (Fig. 1A, grey; yellow in Supplementary Fig. S1) with an identity of 95–98% between them. Unmarked sites were slightly less similar (80–85%). The use of several specific sites of 1–18 bp in length (Supplementary Fig. S1, red, blue, and green) permitted the isolation of at least three Squam1 subgroups (α, β, and γ).

Analysis of all available Pmu copies from GenBank showed that sequences from the α-subgroup were the most numerous (42 of 54 sequences available) (Fig. 1B). The NJ-tree shows that Squam1 of Dpr was clustered with either α- or β-subgroups of Pmu when the overall mean nucleotide sequence differences were compared. A more detailed analysis with handmade comparisons revealed that both genera shared autapomorphic Squam1 segments that belong to two subgroups (α and β), and in this analysis six and three Squam1 copies from Dpr clustered to α- and β-subgroups, respectively (Fig. 1A). The γ-subgroup was found only in Pmu and seemed to represent an apomorphy for this species (Fig. 1B), but this will need to be confirmed by searching further for Squam1 copies from the γ-subfamily in Dpr and other lacertid species.

The 12 Squam1 copies from the Chamaeleo calyptratus (fam. Chamaeleonidae) genome that were sequenced as part of this study are shown in Figure 2 and Supplementary Figure S2. Each of these copies has a much higher level of divergence than that of the lacertid genome, and all of them display homology, but only in the 5′-half of their sequences. There are at least three large homologous regions, as well as several shorter sites that have 92–97% similarity (Fig. 2, boxed; Supplementary Fig. S2, yellow). The remainder of the Cca Squam1 sequence was highly divergent (60–75% similarity, unmarked) compared to the nearly 90% homology between Squam1 sequences from lacertids. Most sequences (nine) formed a similar subgroup (α), while another three copies had very specific structures. The homology between two almost identical β-subgroup copies (Cca10 and Cca11; Supplementary Fig. S2, green) and all others was high only in the first half of the sequences. In the rest of the alignment, the homology ended after nucleotide position 213–215, and was followed by an 82-bp unrelated “tail” that was almost identical in both copies. The third copy (γ-subgroup) (Cca12; Supplementary Fig. S2, blue) was followed by an unrelated sequence that was completely different from all other sequences.

Despite having different tail sequences, β- and γ-subgroups are actually full-length SINEs: one sequence in the β-subgroup was flanked by a TSD (boxed), and the two other sequences (Cca11 and Cca12) contained a stretch of simple direct repeats that are characteristic of complete SINE copies. Since such Squam1 sequences were not found in the other taxa studied, they may represent a feature that is unique to Cca. The high divergence of Squam1 sequences from Cca might suggest a more ancient origin for these sequences than for the other taxa studied.

The analysis of 33 Squam1 copies of four species from fam. Varanidae is shown in Supplementary Figure S3. All copies contained specific ACCTTT motifs in their 3′-ends and two possessed TSDs. The intra- and interspecific differences among species from genus Varanus were much smaller than the intraspecific differences in Squam1 from Cca (Fig. 2), and were similar to the variability levels among Dpr+Pmu (Fig. 1A). No statistically meaningful clusters within NJ-tree topology (not shown) were found among the Varanus species copies, except for the three copies (1–3) of Vgr that clustered separately with bootstrap values of approximately 70%. However, it may be that these copies are members of rare Squam1 subgroups of Vgr.

The four available Squam1 copies from the tokay gecko genome (Supplementary Fig. S4, our data) had all the features of Squam1, and had the greatest number of long conserved
Figure 1. (A) Sequence alignments of individual Squam1 copies of fam. Lacertidae lizards D. praticola (Dpr, our data) and P. muralis (Pmu, 15 copies, GenBank). Squam1 subgroups a, b, and g are outlined. Positions of the PolIII A and B promoter sites are marked in bold rectangles above; short direct repeats (TSD) flanking Squam1 sequences of Dpr are boxed and short tandem repeats at the 3′-ends are underlined. Shared sequences with above 95% similarity are highlighted with light gray (Supplementary Fig. S1, yellow). The unmarked region had 85–95% similarity. Diagnostic nucleotides specific for sequence groups are given on a dark gray background (Supplementary Fig. S1, red, blue, and green). GenBank accession numbers for Dpr: DQ393692, DQ393693, DQ393695–DQ393699, DQ393700, DQ393702; and for Pmu: DQ023382–DQ023386, DQ023370, DQ023366, DQ023342, DQ023349, DQ023353, DQ023334, DQ023358, DQ023336, DQ023340, DQ023345. (B) Unrooted NJ tree for all available (54) Squam1 copies of Pmu (GenBank accession numbers: DQ023333–DQ023386) and Dpr. Squam1 subgroups are designated by Greek letters.
sequences (95–99%) among the taxa studied. They also displayed more diverged short sites of 80–85% identity, representing about 15% of the whole sequence. The variability was mainly attributable to random single mutations, and only one copy possessed 25–30 bp indel. The small number of available copies leaves open the possibility of the existence of minor subgroups.

The alignment of the consensus sequences of all the lizard Squam1 copies studied is shown in Supplementary Figure S5; the large number of synapomorphic regions among the sequences suggests monophyly. Three Lacertidae genera possessed several specific apomorphies (red).

Figure 2. Sequence alignments of individual Squam1 copies of Chamaeleo calyptratus (Cca, fam. Chamaeleonidae). Squam1 subgroups α, β, and γ are shown. Derived 3’-ends of copies 10–12 are dark gray (Supplementary Fig. S2, red and green). Shared sequences with 90–95% similarity are shown with light gray (Supplementary Fig. S2, yellow). Unmarked sites possess 60–85% similarity. For other details, refer to Figure 1A legend. Cca Squam1 GenBank accession numbers: DQ393676–DQ393681 and DQ393683–DQ393688. For abbreviations, see Table 1.

Figure 3. Sequence alignments of individual Squam1 copies of snake Thamnophis sirtalis (Tsi, fam. Colubridae, NCBI Trace Archive). α- and β-subgroups contain shared segments for both with ~95% of identities (light gray; Supplementary Fig. S6, yellow), and ~85% in unmarked segments. Dark gray (Supplementary Fig. S6, red) designates apomorphic sites for the β-subgroup. For other details, refer to Figure 1A legend.

Figure 1A legend.

Squam1 sequences of four snake genera were compared. The degree of divergence of Squam1 in a snake from fam. Colubridae, Thamnophis sirtalis (Tsi) is illustrated in Figure 3 and Supplementary Figure S6. The 10 most perfect copies of the 56 available in unannotated GenBank sequences were chosen. Among these, two subgroups were identified (α and β). Most of the sequence lengths were very similar (boxed), while the rest differed by several apomorphic segments (for details see Supplementary Fig. S6). The homology of the conserved regions reached 96–99%, whereas variable regions with about 85% identity comprised nearly 30% of the sequences. Some
apomorphic sites were seen among copies 8–10 (Supplementary Fig. S6, red), which leaves open the possibility of finding other subgroups in the remaining GenBank copies. To address this, more sequences need to be made available.

The individual copies from another snake of the same fam. Colubridae, *Elaphe dione* (Edi) had long sequence stretches of 97–99% similarity (Supplementary Fig. S7, yellow), and several shorter regions of 80–85% similarity (unmarked). However, the small number of copies does not allow any definitive conclusions to be made.

Squam1 copies of two species from the Viperidae snake family, available from GenBank (Piskurek et al., 2006), were analyzed. The level of similarity between intraspecific copies of *Echis ocellatus* (Eoc, 39 copies) was high, with an average similarity of 98% in the “yellow region” (Supplementary Fig. S8) and 85–90% in the unmarked region; the latter comprise about 30% of the molecules. One copy (N13) contained a long 45-bp insertion, while another (N21) contained a 5-bp insertion, indicating their more ancient origin.

The second Viperidae species, *Azemiops feae* (Afe), is represented by only five copies in GenBank (Supplementary Fig. S9). However, even this small number showed more evident divergence than in Eoc. Several long regions (47–61, 68–85, 232–268 bp) comprising more than two-thirds of the molecule had a low level of similarity (80–88%). The mean variability in the “yellow regions” was as significant as in Eoc (95–100%, Supplementary Fig. S8). The variability of Afe Squam1 was the largest among the snakes studied.

Many homologous regions were present in the alignment of Squam1 consensus sequences of snakes and lizards (Fig. 4 and Supplementary Fig. S10). Considering that the consensus sequences of each taxa represent the most frequent nucleotides at each position (and therefore the nucleotides that are more likely to be of a more ancient origin), the homology level of Squam1 copies in different taxa seems to be higher than even intrataxa homologies of different copies. All consensus sequences had several common homological fragments (8–33 bp) of 96–99% similarity. This implies that the Squam1-like SINEs of all snakes and lizards studied were homologous, reflecting their monophyly.

The only available copy of *Lacerta agilis* (Lacertidae) was also included in the analysis (Fig. 4 and Supplementary Fig. S5). Thus, among the families shown in these figures only fam. Lacertidae was represented by three genera, which clustered together at ~20 specific apomorphic positions (gray; Supplementary Fig. S10, red). Our subsequent DNA hybridization and PCR results supported the presence of Squam1-like sequences in other Squamata families.

**Squam1: Distribution Among Squamata**

Next, we searched for Squam1-like sequences in the DNAs of 24 lizard and 2 snake taxa by PCR with Dpr Squam1-specific primers and DNA hybridization with a Squam1-specific probe (Fig. 5). In addition to Dpr, Cca, Vpr, and Gge (described in a previous chapter), a PCR fragment of the expected size (with small variations) at a DNA concentration of 0.1–1.0 ng per sample was observed in 11 of the lizards and in the two snakes studied. This finding suggests the presence of Squam1 in their genomes (bold arrow), which was verified by DNA hybridization with radioactive probe (Fig. 5, dark circles).

Strong positive hybridization signals were revealed in the first group of species where a specific PCR product was abundant. The second group consisting of 13 taxa did not show specificity for Squam1 PCR bands at 0.1–1.0 ng/sample DNA concentration. However, in such cases, the hybridization of 1,000-fold greater amounts of DNA (1 μg) with a probe gave results that exceeded the blank radioactivity by 15–30% (Fig. 5, circles with a cross).

---

**Figure 4.** Consensus sequence alignment of Squam1 from lizards (L) and snakes (S) studied: Lacertidae (Dpr, Pmu, and Lsp, *Lacerta agilis*); Varanidae (Vpr: Vpr, Vin, Vgr, and Vti); Polychrotidae (Aca); Gekkonidae (Gge); Chamaeleonidae (Cca); Colubridae (Edi and Tsi); Viperidae (Eco and Afe) (see also Supplementary Fig. S10). Shared orthological regions (95–100% identity) for all the taxa covered approximately half of the molecules and are highlighted in gray (Supplementary Fig. S10, yellow). Unmarked sequences are the most divergent (60–75% of similarity). Segments of 227–242 bp contain an indel shared by snakes and nonlacertid lizards. There are about 20 identical sites (dark gray) which are apomorphic for Lacertidae (Dpr, Pmu, and Lsp) and 15 for three snakes (Fig. S10, red and green). For other details and designations, see Figure 1 and Table 1.
contrast, hybridization signals with DNAs of control outgroup taxa (xenopus, crocodile, turtle, birds, and mammals) did not exceed background (membrane without DNA) hybridization signal values (Fig. 5, empty circles). Therefore, we postulated that Squam1 copies may be present at low copy numbers in the DNA of these lizards. The PCR method is less sensitive under the specific conditions used, which can probably be ascribed to the relatively high degeneracy of Squam1 copies in some taxa and the resulting imperfect matching of PCR primers to the divergent SINE sequences.

To test the above possibility, we varied the PCR conditions to determine whether there was a relationship between PCR band intensity and DNA concentration ($10^{-5}$–100 ng per probe) in all taxa. The DNA of the first group gave bands at DNA concentrations of $10^{-5}$–1 ng (Fig. 6A). In contrast, the second taxa group yielded bands in the range of $10^{-1}$–100 ng (Fig. 6B–D). The Squam1 band intensities of the first group taxa did not differ significantly (Fig. 5; 1 ng DNA, and Fig. 6A). We approximated the Squam1 copy number in this group using previous estimations in Anolis carolinensis and P. muralis which have about $10^5$ copies per haploid genome (Piskurek et al., 2009). Since similar band intensities were observed in species from the same genera (Anolis equestris and Podarcis taurica, our data), we assumed that these species and other taxa of the first group would have approximately the same copy numbers. As regards the second group, the nearly equal intensities of the weak PCR bands were obtained at DNA concentrations that were $10^2$–$10^4$ times higher than in the first group. We concluded that the Squam1 copy number was proportionally less and would be ten to a few hundred copies per haploid genome. The first and the second groups were therefore designated as “H(highly)-copied” and “L(low)-copied,” respectively.

Figure 5. Distribution of Squam1 in Squamata taxa studied by PCR (thin arrow) and DNA-hybridization (empty arrow). Black circles, strong signals; empty circles, no signals; cross in circle, weak signals. M, ladder markers (1,000 bp). For abbreviations see Table 1.

Figure 6. PCR amplification of Squam1 material of some taxa at different DNA concentrations ($10^{-3}$–100 ng per probe, shown above the slots). (A) Dpr vs. Gge; (B) Gge vs. Bpl; (C) Gge vs. Afr; (D) Dpr vs. lig. K, no DNA; M, markers, shown by arrows. For abbreviations see Table 1.
The PCR results of some control taxa revealed the absence of specific Squam1-like amplified products in humans, mice, birds, crocodiles, and fish at the higher DNA concentrations tested (0.1–100 ng). Some trace amplified material was found at 100 ng of DNA in xenopus. This was a false-positive result, since Squam1-like sequences were not found by direct screening of the whole xenopus genome (http://genome.jgi-psf.org/Xentr4/Xentr4.download.ftp.html, search program FASTA). Turtle DNA had two very weak bands at the same DNA concentration, which migrated above and below the specific band. The probability that these bands are Squam1-specific is very low, but this will be tested by sequencing in the future. Squam1 of squamates is not related to the turtle-specific SINE reported by other authors (Sasaki et al., 2004).

Squam1-like sequences termed SAURIA SINE (Piskurek et al., 2006, 2009) were found by these authors also in the genera Ophisaurus, Basiliscus, and Leiolepis (L-copied), as well as in P. muralis, Heloderma suspectum, Chamaeleo feae, and Anolis carolinensis (H-copied). In addition, PCR indicated that Phrynocephalus helioscopus (Agamidae) and P. taurica (Lacertidae) belonged to the H-copied group (our data, not shown).

Squam2: Structure Peculiarities and Distribution in Squamata

A second, smaller (~200 bp) retroposon named Squam2 that showed no obvious sequence homology to Squam1 was also discovered in the genomic DNA library of D. praticola. Figure 7A shows the consensus Squam2 data of Dpr, of T73 copies of Squam2 copies in the draft genome of A. carolinensis (Aca, http://www.broad.mit.edu/models/anole/), and of several unannotated but Squam2-like sequences in GenBank. Aca and all these sequences possessed many orthologous regions of 80–90% identity (Fig. 7A, black; yellow in Supplementary Fig. S11) separated by more divergent, smaller sites with short indels. We could not find TSD-like or short tandem repeated sequences at the 5'- or 3'-ends. All these repeats contained shared trNAAs-like sequences that were similar to the 5'-parts of some mammalian SINEs (Fig. 7B). These data suggest a more ancient evolutionary origin for Squam2 than for Squam1, and the masking of some SINE-specific features by multiple mutations. Squam2-like sequences were present in all the Squamate taxa studied, and were absent from crocodiles, turtles, amphibians, birds, and mammals, as shown by PCR and DNA hybridization (Kosushkin et al., 2008). The abundance of Squam2 in different taxa was variable: Squam2 was revealed only by DNA hybridization in Tte (Tupinambis teguixin, Teiidae) (no PCR product available).
of expected length was obtained with this species DNA) and the band intensity of Squam2 in the veiled chameleon was very weak; however, Squam2 was present in all other Squamata taxa as judged by the results of both methods.

DISCUSSION AND CONCLUSION

The potencies of specific retroposons in phylogenetic studies of squamate reptiles were evaluated. Data from this and other studies, including those from unannotated sequences (GenBank) and from Piskurek et al. (2006), were summarized. This data pool covers only a small number of species, but all of them represent families of the infraorders after Estes et al. (88; see Table 1). This observation opens the perspective of determining Squam1 usage in the phylogeny of Squamata reptiles at the intra- and interfamilial systematic level.

First, we will outline some general principles for the study of molecular evolution using retroposons that were used in similar studies in other eukaryotes. There are three main general approaches to using SINEs as molecular markers (Serdobova and Kramerov, ’98; Veniaminova et al., 2007). The first involves the evaluation of the presence or absence of a SINE family in one or more taxa (“method of families”). As mentioned above, the monophyly of Squamata is shown by the presence of both specific SINEs in all taxa studied and their absence in crocodile, turtle, amphibians, and other orders.

The second approach involves the analysis of shared locus-specific copies and their levels of divergence (“method of copies”) (Murata et al., ’93; Roos et al., 2004; Wang and Kirkness, 2005; Ray et al., 2006). This latter approach may be used in intra- and intergenera relationship studies, as was shown by Piskurek et al. with varanid species (Piskurek et al., 2006). This approach was also profoundly explored in fish studies by the N. Okada group (see review Okada et al., 2004).

Besides these two approaches, one can use the phylogenetically important fact that a SINE family may have several discrete subfamilies that arise from a few individual “master” copies (Kim et al., ’94; Bashir et al., 2005; Kramerov and Vassetzky, 2005; Piskurek et al., 2006). The apomorphic features of such subfamilies between taxa are useful for studying phylogenetic relationships. Using the classification of SINE methods provided by Veniaminova et al. (2007), this approach could be termed the “method of subfamilies.” We and others have identified several subfamilies of Squam1 (Sauria SINE of Piskurek et al. (2006)) in Dpr+Pnu (Darevskia+Podarcis) (at least three, Fig. 1A), Cca, Chamaeleo calyptratus (three, Fig. 2A), Tsi, Thamnophis sirtalis (two, Fig. 3A), and A. carolinensis (five subfamilies, according to Piskurek et al., 2009).

According to the data obtained here, phylogenetic traces could potentially be found from the relative abundance of a specific SINE in related taxa. As soon as a new SINE family or subfamily emerges in evolution, it cannot be lost and can only increase in number. Thus, in principle, taxa containing few orthologs to Squam1 would be evolutionarily older than taxa with more (or many more) copies. This “copy number method” approach was mentioned previously in studies of phylogenetic relationships (Serdobova and Kramerov, ’98; Ray et al., 2006).

The first conclusion of our study is that all the Squamata families studied up until now seem to possess two kinds of SINEs—Squam1 and Squam2. One of them (Squam1) termed Sauria SINE by Piskurek and colleagues, and was found in some species studied by us and in several other Squamata species (Piskurek et al., 2006, 2009). This finding supports the monophyly of all the taxa studied, including those of serpents and tuatara (Piskurek et al., 2006). The alignments of all available Squam1 sequences showed that they shared more than half of their sequence lengths (consisting of 10–35 bp fragments), with similarities of close to 95% (90–100%) in homologous positions (Figs. 1–3).

Thus, a part (larger or smaller in different taxa) of the Squam1 sequence is extremely conserved (95–100% of identities). The rest of the sequence is much more variable and substrate for mutational process. Squam1 possesses apomorphic sites that can be used as a shared, derived (apomorphic) feature when comparing taxa at low taxonomic levels (species, genera, etc.). These apomorphies may be represented by shared clusters of several nucleotides or shared single nucleotides dispersed in a taxa-specific manner.

The same argument can be used for Squam2 which is also intrinsic to all the taxa studied. Slightly less similarity was seen among Squam2 sequences (Fig. 7), reflecting its more ancient origin. A more detailed analysis should be forthcoming when Squam1 and Squam2 from more taxa becomes available.

The second conclusion is that, in each taxon studied, Squam1 possessed some intrinsic features and different patterns of mutational distribution. These could be apomorphies which could be used in the future (after deciphering many more sequences) for inter- and intragenera phylogenetic studies (Figs. 1–3). The data presented here and those of Piskurek et al. (2006) suggest the very tight genetic relationships of the Darevskia and Podarcis genera as shown by the two shared Squam1 subgroups (z and β), and the data support (together with the single available copy in Lacerta agilis) the monophyly of Lacertidae among the other families studied (Figs. 1A and 4). Most of the P. muralis copies from GenBank belonged to the z-subgroup. At the same time, Podarcis is characterized by a minor apomorphic Squam1 subgroup (y). These findings stimulated investigation into the Squam1 subgroups of other Lacertidae genera, which has a very complex and poorly studied system. Notably, comparatively low levels of intraspecies copy divergence were seen in representatives of Varanidae and Gekkonidae Squam1 copies (Figs. 1A, 2B and C). High sequence variability was seen for Chamaeleo calyptratus (Chamaeleonidae) and for the snake Azemiops fœae (Viperidae), suggesting a more ancient
origin for Squam1 in these genera. Further investigations should allow for the study of evolution of the lizard and snake lineages.

The third conclusion is that, based on the very great differences in Squam1 contents in the different Squamata lines, all of them could be divided into two main groups: highly-copied (H) and low-copied (L) (Figs. 5 and 6). This study investigated the presence and contents of Squam1 (discovered initially in the Lacertidae genome) in representatives of the two main largest Squamata taxa (Scleroglossa and Iguania [http://www.reptile-database.org/db-info/taxa.html]) (Fig. 8). SINEs were present at high copy number in some specimens of these two lineages, and in low copy number in others. One possible explanation for this is that Squam1 in L-copied taxa became more divergent due to the longer evolutionary pathways of these taxa, which would have resulted in lower affinity of the primers designed for Squam1 from high copy number taxa (H-group). If this were true, then the H-copied taxa would be evolutionarily younger than L-copied taxa from the same taxonomic lineage. This preliminary conclusion should be verified in the future by comparing the Squam1 sequences of all taxa studied in this work to those of other taxa.

Another possible explanation for the above-mentioned differences would be that Squam1 emerged in few “master gene” copies in the ancestor of Squamata. According to Squam1 sequence divergence in Squamata, amplification of these not-numerous ancestral “master genes” would have taken place at different times. As a result, various subfamilies emerged with copy numbers that differed by several orders of magnitude between the lineages. More ancient taxa had more divergent Squam1 sequences in their genomes, due to a longer time of decay by chance mutations. Thus, taxa with smaller numbers of Squam1 copies or with higher sequence divergence could be considered more ancient, and the copy number could be considered an indirect indication of the relative evolutionary ages of the taxa. Moreover, more ancient taxa should possess more divergent copies of Squam1 that differ significantly from the “master genes” in the genome of their common ancestor.

To compare the available Squam1 data, we juxtaposed the Squam1 H- and L-copied taxa to a known system compiled by Yetz (2010), which was based on several publications by Zug et al. (2001) and other authors. This schema, if correct, does not exclude the possibility that several independent Squam1
amplification events occurred in the evolution of Squamata lineages. The L- and H-copied taxa appeared consecutively, as predicted by the intrinsic properties of SINE evolution. Therefore, L-copied precursors must have existed before Squam1 amplification events could occur.

Although we consider the Yetz's schema as quite preliminary, we demonstrated that retroposon abundance in different taxa can be used to verify taxonomic relations (see reviews Alifanov, 2000; Zug et al., 2001). The following examples illustrate this. There is a problem in the American macro- and microteiids identification which was developed on the basis of Boulangé's idea of the XIX century and which has resulted in considerable controversy among investigators. The opinion (Flesh, '83 cited in Alifanov, 2000) that macroteiids are more related to iguanids, and microteiids are more related to lacertids is supported by the fact that macroteiids (Tupinambis tequizin) and microteiids (Cnemidophorus lemniscatus; Piskurek et al., 2006) fall into different groups (L- and H-copied Squam1 correspondingly). This observation agrees with the observed differences in the reported morphological characteristics of macroteiids and microteiids.

The second example, namely, the division of Iguania into pleurodont iguanids and acrodont agamids, is supported by our data as the latter taxa (Agamidae, Chamaeleonidae, Leiolepididae, and Phrynocephalidae, which were placed by Alifanov into infraorder Pachyglossa), have H-copied Squam1 whereas Iguanidae+ Corythophanidae have L-copied Squam1. This leads to a question about the position of Phrynosomatidae and Polycrotidae which have a large number of Squam1 copies and which for this reason should be placed in the H-copied group.

As for snakes, it is well known that the “lizard origin of the snake may not be supported by any concrete evidence” (Alifanov, 2000). Our data have shown that several species of the Viperidae and Colubridae families have H-copied Squam1 retrotransposons. From a general point of view, it is possible that they branched from H-copied lizards; however, this is only an assumption that will need to be tested in the future.

Another example where the copy number of Squam1 may solve a number of controversial issues is the supposition that Sphenodontia (tuatara) and acrodont agamids, is supported by our data as the latter taxa (Agamidae, Chamaeleonidae, Leiolepididae, and Phrynocephalidae, which were placed by Alifanov into infraorder Pachyglossa), have H-copied Squam1 whereas Iguanidae+ Corythophanidae have L-copied Squam1. This leads to a question about the position of Phrynosomatidae and Polycrotidae which have a large number of Squam1 copies and which for this reason should be placed in the H-copied group.

Our data on the semi-quantitative evaluation of Squam1 copy numbers were correlated to those of Piskurek et al. (2006). The comparison revealed that the abundance of Squam1 (Sauria SINE of authors) in some taxa is less than that in others (see Fig. 5 in Piskurek et al., 2006), based on the relative intensities of the PCR bands. Thus, Sphenodontia (tuatara) and specimens of Dibamus tianmanensis (Dibamidae), Xanthusia vigilis (Xanthusidae), Phrynosoma platyrhinos (Phrynosomatidae), Amphisbaenia xera, and Blanus cinereus (Amphisbaenidae) (which were not studied in our analysis) have markedly less intensive SINE PCR bands. Bands of the representative genera of other families, including Lialis (Pygopodidae), Basiliscus (Corytophanidae), Ophisaurus (Anguidae), and Cnemidophorus (Teiidae), were also weaker than others, as was shown in this study (Figs. 5 and 6).

Molecular genetics investigations of higher level reptilian systems are now only beginning. However, it is already evident that it will be necessary to correlate and interpret the results obtained with different markers in different sets of taxa and species, although this will be complex. Independent evaluation of the relative ages of taxa using SINEs could be used to verify ambiguities. Thus, even the first “mosaic” (i.e., due to relatively low sampling of the taxa studied) results shed light on Squamata evolution using retroposons (SINEs). This approach, which uses the most homoplasies-free, as stated and supported by Hillis (1999), Shedlock and Okada (2000), Salem et al. (2003), and informative markers available (which reflect larger genome parts than any other single nuclear or mtDNA genes), should offer new perspectives for the determination of the evolutionary and systematic relationships among Squamata taxa.

ACKNOWLEDGMENTS

We thank D.B. Vassiliev (Moscow Zoo) and S.K. Semenova for providing us some animal specimens, D.A. Kramerov, A.A. Bannikova, and N.S. Vassetzky for discussion and help with computer data analysis.

LITERATURE CITED


Böhme MU, Fritzsch G, Tippman A, Schlegel M, Berendonk TU. 2007. The complete mitochondrial genome of the green lizard Lacerta...


Nicholas KB, Nicholas Jr HB. 1997. Genedoc: a tool for editing and annotating multiple sequence alignments, multiple sequence alignment editor and shading utility, distributed by the author per anonymous ftp (http://www.psc.edu/biomed/genedoc).


