Non-random accumulation of LINE1-like sequences on differentiated snake W chromosomes

M. Mezzasalma, V. Visone, A. Petraccioli, G. Odierna, T. Capriglione & F. M. Guarino

Department of Biology, University of Naples Federico II, Naples, Italy

Keywords
LINE1; nLTR; sex chromosome; serpentes; colubroids; FISH; genome; chromosome distribution.

Correspondence
Gaetano Odierna, Department of Biology, University of Naples Federico II, Via Cinthia, 80126, Naples.
Email: gaetano.odierna@unina.it

Editor: Jean-Nicolas Volff

Received 18 December 2015; revised 24 February 2016; accepted 1 April 2016
doi:10.1111/jzo.12355

Abstract

Due to their particular phylogenetic position and biological characteristics, squamate reptiles and, in particular, snakes are becoming an increasingly important model for fields such as evolutionary biology, molecular ecology and adaptation. Recently, during a study to analyze the evolutionary history of European whip snakes, we found a LINE1 (L1)-like sequence (GenBank accession no. LM644476), herein called TRL1L, and while there are data on the abundance of L1 in snakes, their genomic and chromosome localization is still largely unexplored. We therefore performed a study to obtain information on TRL1L abundance, distribution and conservation in snake species, belonging to the Colubridae, Lamprophiidae and Viperidae families, using quantitative dot-blot and fluorescence in situ hybridization (FISH). TRL1L showed a high identity with homologous segments of L1s of lizards the Anolis carolinensis and Lacerta agilis and the zebrafish Danio rerio. The discovered sequences are truncated L1 elements which occur with a low copy number, about 0.1% of the genome of the species studied. This evidence suggests that L1 retroposons have a similar landscape in lizard and snake genomes, probably because similar processes limited L1 distribution in their genomes. TRL1L showed a non-random chromosome distribution pattern. It was scarcely located on autosomes and on the euchromatic W chromosome of Cerastes vipera, while mostly found on the heterochromatic W chromosome of Hierophis carbonarius and Elaphe quatuorlineata. Our data support the hypothesis that a ‘purifying selection’ against the accumulation of L1 elements takes place in recombining regions and highlight the possible role of these elements in the differentiation processes of the snake heterochromatic W chromosome. Interestingly, the preferential distribution of TRL1L on the heterochromatic W chromosomes of the studied snakes appears to be similar to that observed in mammals for L1 accumulation on differentiated Y chromosomes. This finding suggests that a convergent process may have taken place in the differentiation of vertebrate heterochromatic sex chromosomes.

Introduction

Snakes are among the most successful vertebrate radiations, with about 3400 known species (Pincheira-Donoso et al., 2013). Due to their particular phylogenetic position and biological characteristics, squamate reptiles and, in particular, snakes are becoming an increasingly important model for different fields of study, such as evolutionary and developmental biology, molecular ecology and adaptation, and metabolic and biomedical venom-related research (see e.g. Secor & Diamond, 2000; Fry et al., 2006; Castoe et al., 2008; Vonk & Richardson, 2008; Ikeda et al., 2010; Delaguere, 2013; Di Pietro, Alcalde & Williams, 2014; Sheehy, 2015). That said, there are relatively few data available about the snake genome. Yet the information gathered so far highlights several distinctive characteristics in different taxa (Shedlock, 2006; Shedlock et al., 2007; Janes et al., 2010; Castoe et al., 2011). In general, the snake genome is relatively small with an average DNA content of about 2.2 pg/nucleus and has great variability in GC content (Hughes, Claya & Bernardi, 2002; Olmo, 2005, 2008; Gregory, 2015). In general, single-copy DNA sequences are almost constant in different taxonomic groups, while repetitive DNA sequences vary significantly (Olmo, 2008). Previously classified as ‘junk DNA’ (Ohno, 1972), repeated DNA sequences are now considered essential factors in genomic evolution and functionality (Biémont & Vieira, 2006; Volff, 2006).

The current knowledge about repeated DNA sequences of the snake genome is scarce and mostly limited to a number of microsatellites isolated in various species belonging to different
taxonomic groups (see e.g. Meister et al., 2009; Castoe et al., 2010; Lukoschek & Avise, 2011). However, various repeated sequences belonging to transposable elements (TEs) have been identified in snakes, namely LINEs, SINEs, CR1-like and Ty1-like elements (Flavell et al., 1995; Nobuhisa et al., 1998; Zupunski, Gubensek & Kordis, 2001; Piskurek, Austin & Okada, 2006; Piskurek, Nishihara & Okada, 2009; Castoe et al., 2011). Transposable elements represent a considerable fraction of eukaryotic genomes, and due to their mobility and amplification properties, may have profound effects on the structure of the host DNA, constituting a major source of genomic variability (see Kidwell & Lisch, 2001; Kazazian, 2004; Biémont & Vieira, 2006) and making it particularly important to understand the mechanisms involved in the evolution of genomes and sex chromosomes (see e.g. Kidwell & Lisch, 1997; Kordis & Gubensek, 1998; Kordis, 2010; Abbott et al., 2013; Chalopin et al., 2013, 2015).

Quantity, nature and level of activity of the various families of TEs may vary widely among vertebrates. Distinct evolutionary lineages are characterized by a different structure of their repeated element landscape (see Kordis, 2010).

Recently, during a study to analyze the evolutionary history of European whip snakes (Mezzasalma et al., 2015), we found a sequence (GenBank accession no. LM644476), herein called TRL1L, which showed a high identity with a transposable element of the LINE1 (L1) nLTR retrotransposon family of the lizard Anolis carolinensis Voigt, 1832. While there are data on the abundance of L1 in snakes (Kordis, 2010), their genomic and chromosome localization is still largely unexplored. We therefore performed a study to obtain information on TRL1L abundance, distribution and conservation in snake species, belonging to the Colubridae, Lamprophiidae and Viperidae families, using quantitative dot-blot and fluorescence in situ hybridization (FISH). The results of this study are presented here with a hypothesis about the non-random distribution of TRL1L in the genome of the snakes in question.

Materials and methods

Samples

We used ethanol-preserved tissues and/or chromosome suspensions of specimens from nine snake species belonging to the Colubridae, Lamprophiidae and Viperidae (see Table 1 for details). All the tissue or chromosomal samples used in this study were derived from specimens used in previous studies (see Gigantino et al., 2002; Aprea et al., 2003, 2006; Mezzasalma et al., 2014, 2015).

Table 1 Number, sex and origin of the studied snake species

<table>
<thead>
<tr>
<th>Family</th>
<th>Species</th>
<th>Origin</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colubridae</td>
<td>Hierophis viridiflavus (Lacépéde, 1789)</td>
<td>Italy</td>
<td>One female</td>
</tr>
<tr>
<td></td>
<td>Hierophis carbonarius (Bonaparte, 1833)</td>
<td>Italy</td>
<td>One female</td>
</tr>
<tr>
<td></td>
<td>Hierophis gemonensis (Laurenti, 1768)</td>
<td>Greece</td>
<td>One female</td>
</tr>
<tr>
<td></td>
<td>Coronella austriaca (Laurenti, 1768)</td>
<td>Italy</td>
<td>One female</td>
</tr>
<tr>
<td></td>
<td>Elaphe quatuorlineata (Bonnaterre, 1790)</td>
<td>Italy</td>
<td>One female</td>
</tr>
<tr>
<td></td>
<td>Zamenis longissimus (Laurenti, 1768)</td>
<td>Italy</td>
<td>One female</td>
</tr>
<tr>
<td>Lamprophiidae</td>
<td>Malpolon monspessulanus (Hermann, 1804)</td>
<td>Morocco</td>
<td>One female</td>
</tr>
<tr>
<td>Viperidae</td>
<td>Viper aspis (Linnaeus, 1758)</td>
<td>Italy</td>
<td>One male</td>
</tr>
<tr>
<td></td>
<td>Cerastes vipera (Linnaeus, 1758)</td>
<td>Morocco</td>
<td>One female</td>
</tr>
</tbody>
</table>

DNA extraction

Total genomic DNA was extracted from tissues of alcohol-preserved specimens and from chromosome suspensions stored at the Department of Biology of the University of Naples Federico II. Samples were washed twice in PBS 1x, and digested overnight at 55°C with proteinase K (0.2 mg·mL$^{-1}$) in the presence of 0.5–0.9% sodium dodecyl sulfate. After enzymatic digestion, the DNA was extracted twice, with phenol-chloroform (1:1), and chloroform–isoamyl alcohol (25:1) (Sambrook, Fritsch & Maniatis, 1989).

PCR amplifications, cloning and sequencing

To amplify TRL1L, we designed and used the following primer pair in standard PCR: TRL1L – Fw 5’CACAAAAACAC AACCAAAAAA- 3’ and TRL1L – Rv 5’-AAAACCAGTITT GATCGTTATGAA-3’.

PCR was performed for all species in 20 μL with the following parameters: initial denaturation at 95°C for 5 min, followed by 36 cycles at 94°C for 30 s, 50°C for 30 s and 72°C for 40 s, ending with a final elongation step at 72°C for 7 min. Amplicons were purified using Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). In addition, amplicons of H. viridiflavus and H. carbonarius were ligated in pGem-T easy vector (Promega) to test intraspecific variability of TRL1L. After transformation in DH5α cells, positive colonies were amplified by PCR using T7 and SP6 primer pair with the following conditions: 5 min at 94°C, 36 cycles at 94°C for 30 s, 50°C for 30 s and 70°C for 45 s, 5 min at 72°C. Sequencing of positive colonies and direct genomic amplicons were performed in both directions using the BigDye Terminator kit v1.1 and the automatic sequencer ABI Prism 310 (Applied Biosystems, Foster City, CA, USA). Chromatograms were checked by eye, edited and aligned using Chromas Lite© and BioEdit, version 7.0.5.3 (Hall, 1999). The sequences were blasted in GenBank (Altschul et al., 1990), RepeatMasker (Smit, Hubley & Green, 2015) and Censor
PCR products were purified, 52 as described below. PCR reactions were performed in 50 cycles under the following conditions: 94°C for 2 min, 94°C 30 s, 52°C 30 s, 65°C 1 min and 70°C 2 min. Biotinylated PCR products were purified with isopropanol (88%) + 0.3M K acetate and subsequently used in dot blot and FISH reactions as described below.

Quantitative dot-blot

Quantitative dot-blot analysis was performed on representatives of all the considered genera belonging to Colubridae, Lamprophiidae and Viperidae families. Genomic DNA was diluted to 1 µg/mL + 1 µg/mL of DNA of *E. coli*, serially diluted in denaturing buffer (0.4 M NaOH, 1 M NaCl). A 100 µL aliquot of each dilution was added per slot in a dot-blot apparatus (Bio-Rad, Hemel Hempstead, UK) and filtered on a nylon membrane (Bio Bond Sigma). Purified samples of directly genomic PCR amplicons of TRL1L of *H. viridiflavus* were used as standards at an initial concentration of 0.1 µg/mL + 1 µg/mL of *E. coli* DNA in denaturing buffer. Hybridization and staining procedures were performed as reported by Picariello et al. (2012). Copy numbers of TRL1L in the genome of the different snake species studied were calculated on the basis of the genome sizes reported by Gregory (2015): *H. viridiflavus* = 2.20 pg/N, *Coronella austriaca* = 1.89 pg/N, *Z. longissimus* = 1.82, *E. quatuorlineata* = 1.83 pg/N and *V. aspis* = 2.91 pg/N.

**FISH**

Metaphase plates of the studied species were obtained from fixed cell suspensions stored at the Department of Biology of the University of Naples Federico II and from blood cell cultures, following the methods described by Mezzasalma et al. (2013). FISH was performed as reported by Petraccioli, Maio & Odierna (2012), using mouse antibiotin antibody and FITC rabbit antimouse antibody. Chromosomes were counterstained with propidium iodide and images acquired through a Leica DM6000B epifluorescence microscope and LAS software (Leica Application Suite; Leica Microsystems Ltd., Heerbrugg, Switzerland).

**Results**

Isolation and nucleotide characterization of the repeated sequences

PCR amplification of TRL1L from DNA of *H. viridiflavus* and *H. carbonarius* resulted in a product of about 133 bp. After cloning and sequencing, we analyzed 12 positive clones of TRL1L (five of *H. viridiflavus* and seven of *H. carbonarius*), which showed an identity with their consensus sequence ranging from 87.5 to 97.8% and from 92.6 to 99.2%, respectively. Differences were mostly due to transitions and transversions (2.3:1) and rarely due to deletions (six out of 12 clones showed one deletion at nucleotide 65) (Fig. 1).

PCR amplifications from DNA of other snake species considered produced sequences from 130 to 144 bp, showing an identity with the TRL1L consensus sequence of *H. viridiflavus* ranging from 78.8 to 94.7% (Fig. 2a). Queries with the TRL1L consensus sequence of *H. viridiflavus* in Repbase with Censor (Kohany et al., 2006) or BLASTn evidenced a pairwise alignment, of about 86 bp, with an L1 nLTR retrotransposon of *Anolis carolinensis* (L1-81 Acar) and *Lacerta agilis* Linnaeus, 1758 (Accession No DQ117995), with an identity of 77.9 and 84.2%, respectively (Fig. 2b; Table 2).

Similarly, queries of RTL1L of the other snake species in Repbase with Censor (Kohany et al., 2006) resulted in identities ranging from 74.4 to 85.4% with various traits of L1 elements of *A. carolinensis*. The only exception was the TRL1L of *V. aspis* which showed an identity of 71.3% with an L1 element of the fish *Danio rerio* (Hamilton, 1822) (Table 2).

Search in BlastX with TRL1L of *H. viridiflavus, H. carbonarius, V. aspis* and *H. gemonensis* produced a pairwise alignment with various transcripts, including a region (from aa 181 to aa 206) of the L1-encoded reverse transcriptase-like protein of the crotaline snake, *Protobothrops flavoviridis* (Hallowell, 1861) (Accession No: dbj BAM62877) (Fig. 2c).

Quantitative dot-blot Analysis

In the genome of the snakes in question, TRL1L is present at about 0.15% in *H. viridiflavus, E. quatuorlineata, M. monspessulanae* and *V. aspis*, and less than 0.1% in *C. austriaca, Z. longissimus* and *C. vipera.* Therefore, considering the genome size of the studied species (see above, Materials and methods), TRL1L is present with about 25 000, 22 000 and 33 000 copies in the genome of *H. viridiflavus, E. quatuorlineata* and *V. aspis*, respectively, and less than 14 500 and 14 000 copies, respectively, in the genome of *C. austriaca* and *Z. longissimus* (Fig. 3).

**FISH analysis**

Metaphase plates suitable both in number and quality were obtained from samples of *H. carbonarius, E. quatuorlineata,* and *C. vipera* In *H. carbonarius* and *E. quatuorlineata,* discrete clusters of hybridization signals were mainly evident on interstitial and peritelomeric regions of various macrochromosomes. In addition, clear hybridization signals were more abundant on the W sex chromosome of both these species (Fig. 4a, b), while they were clearly fainter on autosomes and sex chromosomes of *C. vipera* (Fig. 4c).

**Discussion**

Our results show that TRL1L sequences occur in the genome of all the snake species studied, belonging to the Colubridae,
Lamprophiidae and Viperidae (Figs 2 and 3), which represent a large portion of Colubroid radiation (Pyron et al., 2011; Pyron, Burbrink & Wiens, 2013). A search with Censor retrieved that all TRL1L sequences have a high identity with homologous traits of autonomous non-long terminal repeat retrotransposons (nLTR-RTs) of the L1 family, found in the lizard A. carolinensis and in the zebra fish D. rerio (see Table 2).

Transposable elements of the L1 family are an abundant component (6%) of the mammalian genome (Deininger & Batzer, 2002). In the human genome they usually occur with more than 500,000 copies (McClure et al., 2005), mostly as L1 fossils, now unable to transpose (Sells, Provata & Almirantis, 2007), since they are byproducts of events of amplification, truncation and rearrangement occurring during mammalian evolution (Boissinot et al., 2006; Pontius et al., 2007). The nLTR-RT genomic landscape of mammals greatly differs from that of other vertebrates, where retroposons occur in a relatively low copy number and with few full-length (active) copies, probably as a consequence of a ‘purifying selection’ acting versus them (Charlesworth, Jarne & Assimacopoulos, 1994; Novick et al., 2009; Tollis & Boissinot, 2011, 2013).

Tollis & Boissinot (2013) showed that, in the genome of the lizard A. carolinensis, full-length L1 elements are rare and recent, and the accumulation of truncated inserts, due to rearrangements after transposition, is tolerated because it occurs under neutral selection. Our results seem to suggest a similar behavior for L1 retroposons in the genome of the studied snakes. In fact, TRL1L occurs in a small number of copies, most of which are likely to be represented by truncated inserts of full-length L1 elements. This hypothesis is also supported by the partial amino acid identity of TRL1L forced translation with the transposase gene of a full-length L1 element of P. flavoviridis (Tanaka et al., 2013) (see Fig. 2c).

Our attempts to extend TRL1L sequences at the 5′ or 3′ terminal ends in order to highlight the presence of full-length elements yielded negative results. Nevertheless, we cannot exclude their occurrence in the studied snakes, and hypothesize that the negative results could depend on the ineffectiveness of the primers used.

A recent origin of the snake TRL1L elements might be suggested by the high identity observed among them at both intraspecific and interspecific levels. In particular, the observed identity at

Figure 1 Pairwise alignments of 12 TRL1L clones (five of H. viridiflavus and seven of H. carborunius) against their respective consensus sequence. Lower cases indicate primer pairs.
intraspecific level (from 87 to 98%) is comparable to that found among recent L1 copies of A. carolinensis (Tollis & Boissinot, 2013).

In snakes, the genomic localization of different L1 elements is highly variable, ranging from coding to repetitive regions. In P. flavoviridis, a full-length L1 has been found in the intronic region of the Habu serum-like (HBL) protein, a member of the fetuin family, (Tanaka et al., 2013), while a truncated sequence showing about 60% identity with TRL1L sequences is present in a microsatellite region of Agkistrodon contortrix (Linnaeus, 1766) (GenBank A.N. GQ188304) (Castoe et al., 2010).

Our in situ hybridizations provide interesting results concerning TRL1L localization. Even though all the species analyzed show the plesiomorphic snake karyotype of 2n = 36 with the fourth pair being the sex chromosomes (Oguiura, Ferrarezi & Batistic, 2009; Vicoso et al., 2013; Mezzasalma et al., 2014), FISH analysis pointed out a differential trend in the occurrence of LINE retroposons on W chromosomes, depending on their stage of differentiation. TRL1L sequences are indeed present with a higher copy number on the heterochromatic W of H. carbonarius and E. quatuorlineata compared to C. vipera, whose W chromosome is still mostly

Figure 2 (a) Pairwise alignments of TRL1L sequences of the studied species (Hv, H. viridiflavus; Hc, H. carolinensis; Hg, H. gemonensis; Eq, E. quatuorlineata; Zl, Z. longissimus; Mm, M. monspessulanus; Va, V. aspis; Cv, C. viper. (b) Pairwise alignment of TRL1L of H. viridiflavus against homologous regions of nLTR L1 element of A. carolinensis (acro) and L. agilis (lag). (c) Alignments of putative transcripts (frame 2) of TRL1L of H. viridiflavus (Hv) and H. gemonensis (Hi) versus the homologous region of the L1-encoded reverse transcriptase-like protein (L1 ERT-LP) of the crotaline snake, Protobothrops flavoviridis (PI).

Table 2 Identity of TRL1L of the studied snakes with different L1 sequences of A. carolinensis (Acar or AC) and Danio rerio (DR)

<table>
<thead>
<tr>
<th>TRL1L</th>
<th>Identity with</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>H. viridiflavus (from 4 to 88 bp)</td>
<td>L1-81_Acar (dir., from 4032 to 4117 bp)</td>
<td>77.9</td>
</tr>
<tr>
<td>H. carolinensis (from 3 to 87 bp)</td>
<td>L1-81_Acar (dir., from 4032 to 4117 bp)</td>
<td>74.4</td>
</tr>
<tr>
<td>H. gemonensis (from 12 to 83 bp)</td>
<td>L1_AC_1 (dir., from 3968 to 4039 bp)</td>
<td>80.8</td>
</tr>
<tr>
<td>E. quatuorlineata (from 21 to 75 bp)</td>
<td>L1_AC_4 (dir., from 3816 to 3867 bp)</td>
<td>85.4</td>
</tr>
<tr>
<td>Z. longissimus (from 21 to 88 bp)</td>
<td>L1_AC_3B (dir., from 3823 to 3892 bp)</td>
<td>82.9</td>
</tr>
<tr>
<td>M. monspessulanus (from 23 to 95 bp)</td>
<td>L1-90_Acar (dir., from 2262 to 2328 bp)</td>
<td>82.9</td>
</tr>
<tr>
<td>V. aspis (from 24 to 138 bp)</td>
<td>L1-111_DR (dir., from 3333 to 3443 bp)</td>
<td>71.3</td>
</tr>
<tr>
<td>C. viper (from 18 to 128 bp)</td>
<td>L1_AC_15 (dir., from 3350 to 3368 bp)</td>
<td>74.1</td>
</tr>
</tbody>
</table>

dir, direct pairwise alignment.
euchromatic and homomorphic to the Z (Aprea et al., 2006; Mezzasalma et al., 2014, 2015). This evidence suggests that in heterochromatic chromosomes such as the W of *H. carbonarius* and *E. quatuorlineata*, a ‘purifying selection’ is not operative or very relaxed.

A similar preferential accumulation of transposable elements in non-recombining heterochromatic chromosomes or chromosome regions has been documented in various taxa, as on the Y of the Antarctic ice-fish, *Chionodraco hamatus* (Lönnberg, 1905) (Cocca et al., 2015) and in the sex-determining region of chromosome III of the midge *Chironomus riparius* (Meigen, 1804) (Kraemer & Schmidt, 1993).

In rodents of the genus *Microtus*, non-functional L1 elements are predominantly located on sex chromosomes, with distinct asymmetric distribution patterns on the Y and X elements of different species. In particular, L1 are always located on the Y heterochromatin and also, even if less abundant overall, on the X euchromatin (Marchal et al., 2006). These distinct distribution patterns are probably consistent with the action of different processes. The presence of L1 elements on mammalian X chromosomes has been hypothesized to have a functional significance due to their involvement in X gene inactivation (Bailey et al., 2000). Conversely, the preferential accumulation of L1 elements on almost completely heterochromatic Y/W might be tolerated due to the effect of neutral selection and contributes to the degenerative differentiation of heteromorphic sex chromosomes (Marchal et al., 2006; Melamed & Arnold, 2009; Chalopin et al., 2015). Our results on the TRL1L distribution on the heterochromatic W chromosome of *H. carbonarius* and *E. quatuorlineata* could suggest a convergent trend between squamates and mammals concerning the differentiation processes of heteromorphic sex chromosomes.

In conclusion, our data confirm the scarce presence and the recent origin of TRL1L sequences in squamate genomes, and evidence a non-random, preferential accumulation of these sequences in highly heterochromatic W sex chromosomes. These data also show a similar, convergent, distribution pattern of L1 retroposons on vertebrate sex chromosomes when their differentiation takes place through heterochromatinization events.

**Acknowledgements**

Thanks are due to many co-workers, especially Dora Baccaro and Elvira Gentile, for their technical support in quantitative dot-blot and FISH analysis. We thank Mark Walters for his language editing and advice, as well as two anonymous reviewers for their critical suggestions which improved this paper.

**Figure 3** Quantitative dot-blot of TRL1L on genomic DNA of the species considered. Numbers refer to DNA quantity in ng.

**Figure 4** Female metaphase plates of *H. carbonarius* (a), *E. quatuorlineata* (b) and *C. vipera* (c) probed with biotinylated TRL1L sequences of *H. viridiflavus*. Arrows point to the heteromorphic heterochromatic W chromosome of *H. carbonarius* and *E. quatuorlineata* (Mezzasalma et al., 2014, 2015), asterisks indicate the fourth homomorphic ZW chromosome pair (Aprea et al., 2006).
References


