*Doklady Biochemistry and Biophysics, Vol. 392, 2003, pp. 263–267. Translated from Doklady Akademii Nauk, Vol. 392, No. 3, 2003, pp. 410–414.*
*Original Russian Text Copyright © 2003 by Ciobanu, Grechko, Kramerov, Darevsky.*

**BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY**

# A New Subfamily of the Satellite DNA, CLsatIV, of the Lizard *Darevskia lindholmi* (Sauria, Laceridae): Structure and Evolution

**D. G. Ciobanu*, V. V. Grechko*, D. A. Kramerov*, and Corresponding Member of the RAS I. S. Darevsky****

Received April 4, 2003

A new subfamily (CLsatIV) of the tandem repeats CLsat, specific of the genus *Darevskia*, was discovered and sequenced only in the genomic DNA of the lizard *Darevskia lindholmi* (this endemic Crimean species until recently was regarded as the subspecies *Lacerta saxicola lindholmi* [1]). The CLsat repeated sequences were found in Caucasian representatives of the genus. They are characterized by a high level of variability and a sufficient rate of homogenization, which resulted in the formation of subfamilies [2–4]. The orthologs of the three subfamilies of CLsat of the Caucasian lizards were also found in the *D. lindholmi* DNA using the method of hybridization, and two of them (CLsatI and CLsatIII) were sequenced. The sequences and organization of CLsatIV, CLsatI, and CLsatIII of the Crimean species, as well as the two last sequences and similar satellites of the Caucasian species were subjected to comparative analysis. The CLsatIV subfamily is characterized by a 30 bp decrease in length and a higher degree of divergence from the general consensus of the family. It was concluded that the data on the presence of the specific CLsatIV repeats, together with some other molecular genetic data [5, 6], allows regarding the Crimean lizards as a separate species of *D. lindholmi*, differing from all studied populations of the species *D. saxicola*, with which it has been classified until recently.

An endemic Crimean lizard of the genus *Darevskia* (until recent time, the genus *Lacerta* [7, 8]) was attributed earlier to the *lindholmi* subspecies of the Caucasian rock lizard *Lacerta saxicola* [9]. On the basis of certain new findings (including our results [10]), this group, as well as other species attributed earlier to the so-called complex *L. saxicola*, are currently regarded as an individual genus *Darevskia* [7]. Within the framework of this genus, the form *lindholmi* is identified as a separate species [7]. However, the extent of molecular genetic similarity of *D. lindholmi* with the putative ancestor species *D. saxicola*, its subspecies, and other species of the genus *Darevskia*, living in Caucasus, remains obscure. For example, it is unclear whether *lindholmi* originated from the species *saxicola* or from earlier species precursors. The results obtained in this study, together with the data reported in our previous works [5, 6], can be regarded as evidence in favor of the assumption that the Crimean species and the Caucasian species (including the subspecies of *L. saxicola*) were formed independently.

The CLsat (Caucasian Lizards Satellite) family of the tandem DNA repeats of lizards *Darevskia* has been discovered and sequenced in the earlier works [3, 4, 10, 11] and found to be specific for this genus. At the initial stage of the study of the CLsat structure and organization, as well as the patterns of distribution of discriminate mutations and restriction sites, we described two subfamilies of monomers of these repeats [3, 4]. The third subfamily was identified later. In this study, we describe a new subfamily of the CLsat family, CLsatIV, discovered so far only in the *D. lindholmi* genome. The goal of this work was to perform a comparative analysis of this subfamily and three other subfamilies discovered by us earlier in Caucasian species of the genus *Darevskia.* In this work we also describe the intraspecific variability of individual CLsatIV monomers and two other CLsat monomers (I and III) of this species.

The genomic DNA of the lizard *D. lindholmi* was isolated from blood using the method described in [10]. The genomic DNA was subjected to exhaustive hydrolysis with *TaqI* endonuclease and separated in 2% agarose gel. The DNA from the zone 120–150 bp was loaded onto DEAE membrane, eluted with a solution containing 1 M NaCl, and cloned into a plasmid vector (pGEM3zf+). The *E. coli* colonies (strain XL-1) containing plasmids with inserts were selected based on their color (white) upon growing on selective LB-agar

* *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, ul. Vavilova 32, Moscow, 119991 Russia*
** *Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, St. Petersburg, 199034 Russia*

1607-6729/03/0910-0263$25.00 © 2003 MAIK "Nauka/Interperiodica"

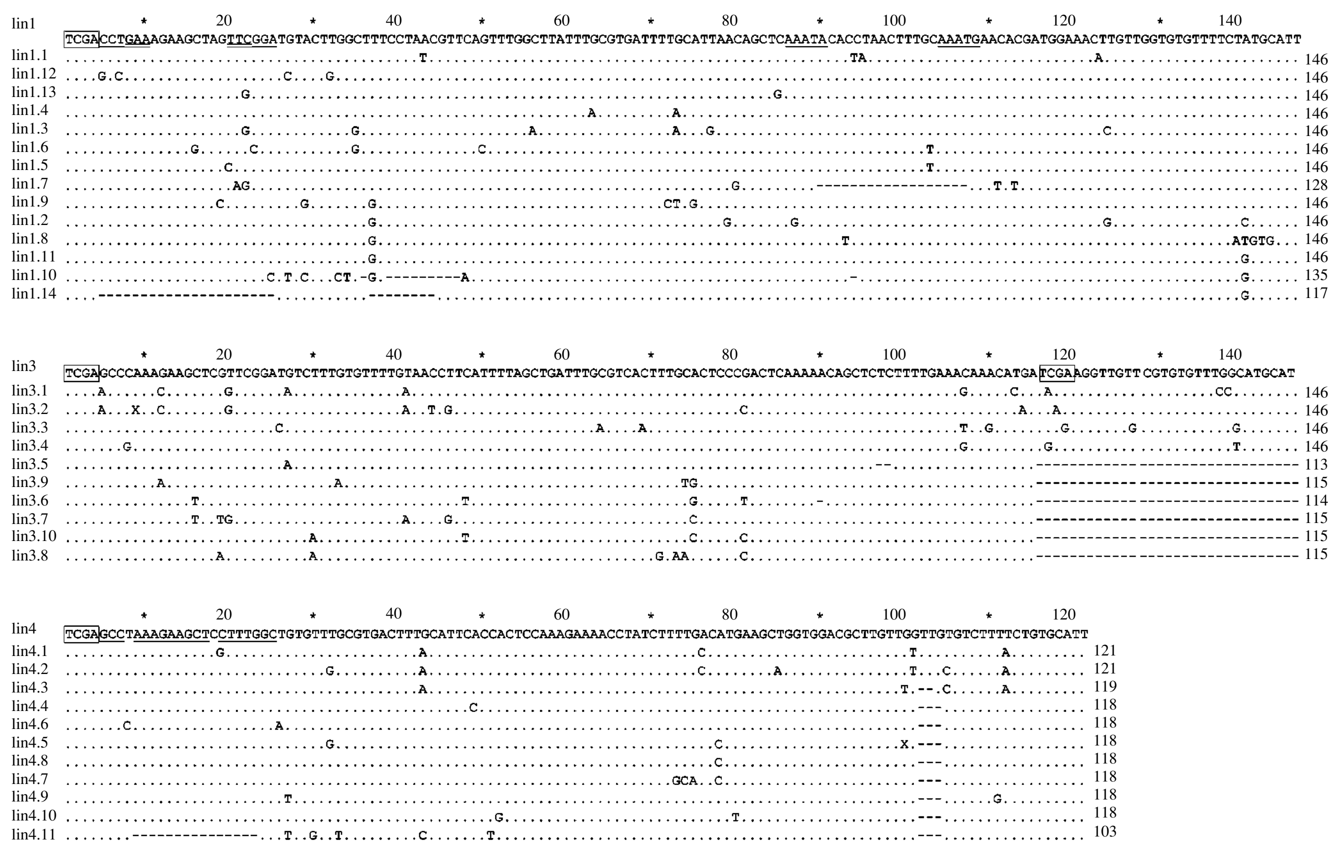

**Fig. 1.** Alignment of monomeric sequences in three CLsat subfamilies of *D. lindholmi* (lin1, 3, and 4). Accession numbers of the sequences used in the analysis are as follows: lin1.1, 14 AY256930-43; lin3.1, 10 AY256944-53; lin4.1, 11 AY256919-29. Dots indicate the nucleotides shared with the consensus sequence; letters, variations from the consensus sequence. Dashes indicate gaps. The length of each monomer in base pairs is shown on the right. The *Taq*I cloning site is boxed. Inverted, direct, and mirror repeats flanking long deletions are underlined.

with IPTG and X-gal. The sequences were determined using PCR-sequencing (Promega) by the Sanger method in accordance with the recommendations of manufacturer.

## STRUCTURE AND VARIABILITY OF THE CLsat (I, III, AND IV) TANDEM DNA REPEATS OF THE LIZARD *D. Lindholmi*

The three CLsat subfamilies, which were described in our earlier studies of the Caucasian species of the genus, were also found in the genome of the endemic Crimean species *D. lindholmi.* Subfamilies CLsatI and CLsatIII were dominant in the genome of the Crimean species (96% in total), whereas CLsatII accounted for only 4% of the total content of these CLsat in the genome. The consensus sequence of one of them was reported in our earlier works [3, 4].

The consensus sequence lin1 (CLsatI), typical for the *D. lindholmi*, was comprised of 14 monomers described here. The length of 11 monomers was 146 bp, similar to the length of the majority of CLsat monomers in the other species of the genus *Darevskia* (Fig. 1). Three monomers were 18, 10, and (21 + 8) bp shorter, which resulted in the formation of gaps in different sites of the nucleotide sequence during alignment. The deleted segments were flanked mainly by short repeats, which might enhance mispairing of bases and formation of loops with their subsequent excision as a result of slippage replication. One of the gaps, located at the 5'-end, could occur as a result of formation of the second site of restriction, which was used for cloning the monomer. Generally, the individual variability of the CLsatI repeats in *D. lindholmi* ranges from 3 to 10% (on average, 6%) and is caused by single base substitutions. The majority of substitutions are located randomly, and only a few of them are diagnostically significant, which allows the monomers to be divided into groups.

The subfamily CLsatIII of *D. lindholmi* was characterized on the basis of the sequences of ten monomers lin3 (Fig. 1). The length of four of them is typical of CLsat (146 bp). Six monomers are 25 bp shorter. Nevertheless, according to the structure of the common part of the sequence, the ten monomers are quite similar to each other and vary as a result of random single-nucle-

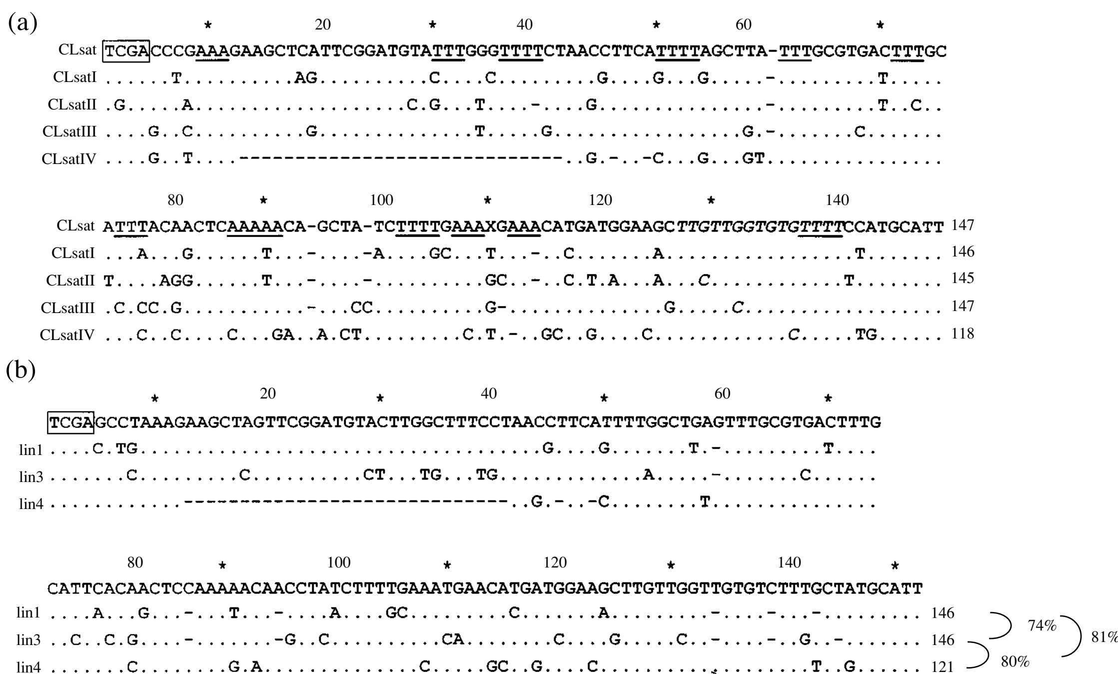


**Fig. 2.** Alignment of the CLsat family consensus sequence and the consensus sequences of each of the four subfamilies (CLsatI, II, III, and IV) and the consensus sequences of the lin1, 3, and 4 monomers. Dots indicate nucleotides shared with the consensus sequence; letters, variations from the consensus sequence. Dashes indicate gaps. The specific restriction site for each of the CLsatI, III, and IV consensus monomers is boxed. The sequence identities between the consensus sequences CLsat, lin1, 3, and 4 are shown at the bottom right.

otide substitutions within the range of 6–13%. The deleted segment in this case most probably appeared as a result of a single-nucleotide substitution at position 117, giving rise to the formation of an additional restriction site of cloning (*Taq*I–tcga).

Eleven discovered monomers, belonging to the new subfamily CLsatIV. However, representatives of this subfamily have not been found yet during random selection among many sequences of the clones of other *Darevskia* species. Their length ranged from 103 to 121 bp. The variability of the length is due to short (2-bp) indels. The degree of similarity between ten monomers ranged from 88 to 98% (in most cases, it was 95%). The CLsatIV species consensus sequence (lin4) was constructed on the basis of these monomers (Fig. 1). Generally, the variability of the monomers is determined by single-nucleotide substitutions and short indels. The extent of divergence of the last monomer (lin4.11) is 10% greater, mainly as a result of deletion of 15 bp. There are several group-marker sites (the first three monomers) and one deletion (2–3 bp) fixed at position 98–100 of the other eight monomers. This deletion appears in the region composed of $(T)_n(G)_n$ repeats (91-TTGTTGGTTGTGTCTTTT-108), suggesting that slippage replication is involved in the deletion occurrence. It should be noted that the base substitution T for G is typical at the position 97 in the first two monomers and the position 96 in the third monomer. Most probably, this substitution is implemented by the same mechanism. The segment, which was deleted in lin4.11, is represented in the other monomers by two inverted repeats (11 bp each) degenerate at two positions. The deleted segment in monomer lin 4.11 is flanked by GC, which could cause mispairing of bases, formation of loops, and excision of 15 bp in this monomer (or equiprobable amplification in another monomer).

## COMPARATIVE ANALYSIS OF THE SUBFAMILY CLsatIV AND THREE OTHER SUBFAMILIES CLsat OF THE SPECIES OF THE GENUS *Darevskia*

The CLsatIV sequences are significantly shorter than the sequences of the majority of CLsatI–III monomers. It might be suggested that this was caused by the deletion of 30 bp. This deletion could also occur as a result of slippage replication, which is indirectly evidenced by short flanking repeats. This mutation became a concomitant specific character of arising and fixation of the fourth subfamily, which could be explained by concerted evolution of this tandem DNA repeats [12].

Generally, CLsatIV, like the other three subfamilies, is characterized by the presence of a number of homopolymeric T- and A-runs (five $(T)_n$- and three $(A)_n$-runs in *D. lindholmi*) and the region containing $(T)_n(G)_n$ repeats, which is conserved in the whole CLsat family (Fig. 2). The total content of AT bases (55.4%) is slightly lower than in the other three subfamilies

Similarity level (%) between the consensus sequences of the CLsat subfamilies and the consensus sequence of the family CLsat

| Family of tandem repeats | CLsat | CLsatI | CLsatII | CLsatIII | CLsatIV |
|---|---|---|---|---|---|
| CLsat | | 86 | 82 | 86 | 78 |
| deletions | | 0 | 1 | 1 | 2 |
| CLsatI | | | 80 | 77 | 77 |
| deletions | | | 0 | 2 | 2 |
| CLsatII | | | | 78 | 73 |
| deletions | | | | 2 | 2 |
| CLsatIII | | | | | 79 |
| deletions | | | | | 2 |

Note: The level of similarity and deletions between the nucleotide sequences of CLsat consensuses leave out of account the 30-bp deletion in lin4.

CLsat (61.64, 62.75, and 60.76% in CLsatI, CLsatII, and CLsatIII, respectively), which suggests either another direction of nucleotide substitutions than that observed in ClsatI–III or a younger age of the CLsatIV repeats. The degree of similarity of the consensus sequence CLsatIV with any of the other three CLsat subfamilies is the same as the degree of similarity between these subfamilies themselves, and falls within the range from 75 to 81% (Fig. 2, table). It should be noted that the degree of similarity between the sequences CLsatIV and CLsatII is lower than that between the other pairs. This fact, as well as an extremely low content of CLsatII in the *D. lindholmi* genome (according to the results of hybridization), suggest that CLsatIV is much closer to CLsatIII and CLsatI. The level of divergence of lin4 from the general consensus sequence of the family is higher (20%) than the level of divergence of CLsatI, CLsatII, and CLsatIII (14, 18, and 13%, respectively). Let the general consensus of the subfamily CLsat be regarded as the hypothetical sequence, which is closest to the common ancestor sequence. Then, the most diverged and shortest sequence (ClsatIV) is thought to evolve later than at least one of three subfamilies CLsat described earlier.

## COMPARATIVE ANALYSIS OF EVOLUTION OF DIFFERENT SUBFAMILIES CLsat IN *D. Lindholmi*

The evaluation of correlation between the level of mutations and the homogenization rate of monomers lin1, lin3, and lin4 in *D. lindholmi* showed that the level of mutations was the highests in the lin3 monomers. The second highest mutation rate was observed in lin1, whereas the lin4 repeats were found to be the most homogeneous within the species. Nevertheless, the general level of variability within none of the *D. lindholmi* CLsat subfamilies exceeds 10%, whereas the level of variability between the monomers of the sub-

**Fig. 3.** Unrooted dendrogram of the monomer sequences of three CLsat subfamilies of *D. lindholmi* (lin1, lin3, and lin4) shown in Fig. 1. Accession numbers for all sequences are listed in the caption for Fig. 1. Tree was constructed using the Treecon software, the nearest-analysis neighbor, and the Jin–Nei [14] and Kimura two-parameter algorithms (to determine genetic distances) [15]. Confidence of the groupings was estimated using 1000 bootstrap replications. The bootstrap values are shown as the percentage above the nodes in which they are higher than 50%. The scale of genetic distance is shown at the top.

families is larger than 20% (unless the long deleted segments are taken into account). These facts suggest that the process of accumulation of mutations alternate with the processes of homogenization, which is consistent with the concept of concerted evolution of tandem repeats. This is also supported by the phylogenetic tree of divergence of the CLsat monomers of *D. lindholmi* (Fig. 3).

In conclusion, let's compare the results described above with the results of our earlier hybridization analysis of all species of the genus *Darevskia.* This analysis showed that the pattern of distribution of the content of each of the three subfamilies CLsat (I, II, and III) was species-specific [3, 4]. These findings seem to provide a new insight into some uncertain aspect of taxonomy of this genus and allow the level of relationship and status of certain species, subspecies, and populations to be elucidated. From this standpoint, the fact of discovery of the fourth subfamily (during random selection of clones) in the genome of the Crimean species *D. lindholmi* corroborates a significant difference between this species and Caucasian representatives of the genus *Darevskia.* This conclusion is also confirmed by the results of the taxonprint analysis [6]. The *Taq*I-taxonprints of the *D. lindholmi* genomic DNA were found to differ from the other rock lizards by the presence of an enriched fraction of repeats of about 120 bp, whose size coincides with the size of CLsatIV. The content of monomers with length of about 150 bp, which are mainly represented by the family CLsat (I, II, and III) of tandem repeats, in *D. lindholmi* is substantially smaller than the number of monomers in the other Caucasian species [6]. The results obtained by the RAPD method also showed that the difference between DNA of *D. lindholmi* and DNA of *D. saxicola* (as noted above, *D. lindholmi* was assumed to be a subspecies of *D. saxicola*) was significantly larger than the difference between four Caucasian subspecies of *D. saxicola* [5]. Our results are also consistent with the conclusions drawn by other authors on the basis of analysis of the allozyme composition of the enzymes [8].

We consider the results obtained in this work as evidence in favor of our earlier hypothesis that there is a certain correlation between the processes of speciation and evolution of tandem repeats. This correlation may be useful for reconstructing of low-level taxonomy [4, 10, 13]. Thus, satellite DNA is a valuable molecular marker in studies of taxonomy and phylogeny of biological species.

## ACKNOWLEDGMENTS

We are grateful to B.S. Tuniev for assistance in collecting of field material and A.A. Bannikova, N.L. Ryabinina, and S.A. Kosushkin for stimulating discussion.

This study was supported by the Russian Foundation for Basic Research, (project nos. 02-04-49548, 00-04-48245, and 02-04-06859).

## REFERENCES

1. Shcherbak, N.N., *Zemnovodnye i presmykayushchiesya Kryma* (Amphibians and Reptiles of the Crimea), Kiev: Akad. Nauk SSSR, 1966.
2. Chobanu, D., Rudykh, I.A., Grechko, V.V., and Kramerov, D.A., *Molekulyarnye mekhanizmy geneticheskikh protsessov i biotekhnologii* (Molecular Mechanisms of Genetic Processes and Biotechnology) (Proc. Int. Conference), Moscow, 2001, pp. 181–182.
3. Chobanu, D., Rudykh, I.A., Ryabinina, N.L., *et al.*, *Mol. Biol.* (Moscow), 2002, vol. 36, pp. 296–306.
4. Rudykh, I.A., Grechko, V.V., Chobanu, D.G., *et al.*, *Genetika* (Moscow), 2002, vol. 38, pp. 1110–1114.
5. Ryabinina, N.L., Grechko, V.V., and Darevskii, I.S., *Genetika* (Moscow), 1998, vol. 34, pp. 1661–1667.
6. Ryabinin, D.M., Grechko, V.V., Darevsky, I.S., *et al.*, *Russ. J. Herpetol.*, 1996, vol. 3, pp. 178–185.
7. Arribas, O.J., *Russ. J. Herpetol.*, 1999, vol. 6, pp. 1–22.
8. McCulloch, R.D., Fu, J., Darevsky, I.S., and Murphy, R.W., *Amphib. Reptil.*, 2000, vol. 21, pp. 169–176.
9. Darevskii, I.S., *Skal'nye yashcheritsy Kavkaza* (Rocky Lizards of the Caucasus), Leningrad: Nauka, 1967.
10. Grechko, V.V., Ryabinin, D.M., Fedorova, L.V., *et al.*, *Mol. Biol.* (Moscow), 1998, vol. 32, pp. 172–183.
11. Rudykh, I.A., Grechko, V.V., Kramerov, D.A., and Darevskii, I.S., *Dokl. Akad. Nauk*, 1999, vol. 367, no. 4, pp. 563–566.
12. Dover, G.A., Brown, S., Coen, E., *et al.*, *Genome Evolution*, London: Academic, 1982.
13. Grechko, V.V., *Genetika* (Moscow), 2002, vol. 38, pp. 1–20.
14. Jin, L. and Nei, M., *Mol. Biol. Evol.*, 1990, vol. 7, no. 1, pp. 82–102.
15. Kimura, M.J., *Mol. Evol.,* 1980, vol. 16, pp. 111–120.



View publication stats