JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 302B:505–516 (2004)

New Satellite DNA in *Lacerta* s. str. Lizards (Sauria: Lacertidae): Evolutionary Pathways and Phylogenetic Impact

DOINA CIOBANU[1,3*], VERNATA V. GRECHKO[1], ILYA S. DAREVSKY[2], AND DMITRI A. KRAMEROV[1]

[1]*Engelhardt Institute of Molecular Biology, Russian Academy of Science, 119991, Moscow, Russia*

[2]*Zoological Institute, Russian Academy of Sciences, 199034, St.Petersburg, Russia*

[3]*Institute of Genetics, Academy of Sciences of Moldova, 111120, Kishinev, Moldova*

ABSTRACT A new tandemly repeated (satellite) DNA family namely Agi160, from *Lacerta agilis* and *Lacerta strigata* (*Lacerta* sensu stricto (s. str.), Linnaeus 1758) have been cloned and sequenced. Agi160 is found in the above two species, as well as two other representatives of the same genus, *L. viridis* and *L. media.* DNA hybridization did not reveal it in *Darevskia, Podarcis, Zootoca, Eremias, Ophisops*, and *Gallotia* – the other genera of the family Lacertidae. The results suggest that Agi160 is a *Lacerta* s. str. specific family of tandem DNA repeats. However, a comparison between sequences of Agi160 and CLsat repeat units revealed 60 bp regions 62–74% identical. The latter is a satellite DNA family typical for *Darevskia* (syn. ''*L. saxicola* complex'') (Grechko et al., Molecular-genetic classification and phylogenetic relatedness of some species of Lacertidae lizards by taxonoprint data. Mol Biol 32:172–183, 1988.). Both Agi160 and CLsat tandem repeats share several common features (e.g., the same AT content and distribution of multiple short A-T runs, internal structure of repeated units, the presence of conservative regions). These data are indicative of their common origin and a possibly strong selective pressure upon conserving both satellites. A comparative analysis of structure, organization, and abundance of these two families of satDNA reveals evolutionary pathways that led to their formation and divergence. The data are consistent with the hypotheses of the concerted evolution of satellite DNA families. The possibility of use of Agi160 as a phylogenetic tool, defining relationships within *Lacerta* s. str., as well as within the whole family of Lacertidae is discussed. *J. Exp. Zool. (Mol. Dev. Evol.) 302B:505–516, 2004.*

1. INTRODUCTION

Tandem repeats of different monomer and array length form a major part of eukaryotic genomes, and most of them represent noncoding DNA families (Cox and Mirkin, '97). On the basis of buoyant density experiments they were called satellite DNA (reviewed in Beridze, '86). Repetitive DNA includes microsatellite, minisatellite, and proper satellite classes that differ by sequence complexity, chromosomal localization and peculiarities of evolutionary pathways (Charlesworth et al., '94). However, mechanisms involved in the homogenization and divergence of tandemly repeated coding DNA also affect satDNA arrays (Elder and Turner, '95). This evolutionary model, which allows for simultaneous action of several molecular mechanisms upon tandem repeats of nuclear DNA, is known as concerted evolution (reviewed by Elder and Turner, '95). The proper satDNA forms heterochromatin arrays, mainly in the regions of centromere and pericentromere, and therefore it is subjected to such processes as recombination to a lesser degree, which has an

Abbreviations: ms, microsatellite; bp, base pairs

Grant sponsor: Russian Foundation for Basic Research; Grant numbers: 02-04-49548, 00-04-48245 and 02-04-06859.

*Correspondence to: Doina Ciobanu, Engelhardt Institute of Molecular Biology, Russian Academy of Science, 119991, Moscow, Russia. E-mail: dia@id.ru

Received 21 April 2004; Accepted 30 June 2004

Published online 3 September 2004 in Wiley Interscience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21014

impact on its structure and evolutionary dynamics. Thus the peculiarities of satDNA evolutionary dynamics determine its species or taxon specificity (reviewed in Dover et al., '82; Grechko, 2002).

Information about the structure of reptile satDNA, and especially that of lizards, as well as about its evolution, is far from exhaustive. There are four families of satDNA described for Lacertidae (see detailed description and references in section 3.4. A comparative analysis of satDNA of Lacertidae). The concerted mode of evolutionary dynamics is a common feature of all of them, despite the different monomer structures. Three of the described satellite families are shared between species of the same Lacertidae genus and have species-specific variants (Capriglione, '95; Rudykh et al., '99). The satellite pGPS occurs in the genome of *Podarcis* and distantly related species *Algyroides moreoticus, Lacerta graeca, Lacerta viridis, Lacerta vivipara, Archaeolacerta bedrigae* (Capriglione, '95). These preliminary data do not allow making ultimate conclusions about structure and evolution of satDNA in reptiles.

We present here a new family of satDNA characteristic of the genus *Lacerta* s. str.. A comparison between Agi160 and four other satDNA families known within the Lacertidae reveals significant sequence similarity between only two of them (Agi160 and CLsat). Our results indicate a common evolutionary origin for the Agi160 and CLsat families characteristic of two genera (*Lacerta* s. str. and *Darevskia*), that diverged probably more than 25 Myr ago (Darevsky, '90; Mayer and Benyr, '94).

2. MATERIALS AND METHODS

2.1. Sampling

The sampling localities of all species studied are listed in Table 1.

2.2. Cloning and sequencing

The genomic DNA of lizard specimens was isolated from blood samples of *L. agilis exigua*, *L. agilis boemica* and *L. strigata*, following a previously described protocol (Grechko et al., '98). The total DNA from the specimens was digested with *Hind*III (Promega) restriction enzyme, fractionated by electrophoresis in 2% agarose and stained with ethidium bromide. The material from 120-200 bp region was electro eluted, ligated into the *Hind*III site of pGEM-3zf(+) plasmid vector (Promega) (Sambrook et al., '89) and transformed into the competent cells of *E. coli* XL1-Blue strain. Recombinant clones were selected on ampicillin LB plates, containing X-gal and IPTG. Twenty four positive (white) clones were randomly selected for each lizard species under analysis. Plasmid DNA preparation followed a previously described protocol (Lee and Rasheed, '90). The nucleotide sequences were determined by a dideoxy-chain-termination method using the PCR sequencing kits (Promega) according to the supplier's protocol.

2.3. DNA hybridization

Dot-blot hybridization analysis was performed with graded amounts of specific genomic DNA (0.01, 0.1, 1, 10, 100, 1000 ng), 1 ng of ae22 and Scz_11_1 satellite clones, and 1000 ng of the genomic DNA for each of the 32 investigated species. The concentration of DNA in each sample was calibrated against defined genomic DNA in agarose-EtBr gels. Satellite clones ae22 (Agi160) and Scz_11_1 (CLsat) were obtained from *Lacerta agilis exigua* and *Darevskia saxicola sczcerbaki* genomes respectively. DNA was spot-bloted on Hybond N^+ (Amersham) membrane following the manufacturer's protocol and subsequently hybridized with specific satellite monomers under low-stringency conditions (58°C).

Southern hybridization was performed after digestion of the genomic DNA (1000 ng for each probe) with *Hind*III (Promega) and blotting of DNA from agarose gels onto Hybond N^+ membrane following manufacturer instructions (Amersham). The hybridization with ae22 monomer was performed under high-stringency conditions (68°C). The [^{32}P] labelled hybridization probe was obtained by PCR amplification of monomer using sequence-specific primers (5'-TTGGTGTTTGTTTTGTG-forward and 5'-TACCATGCTTATTTCAC-reverse) and a recombinant plasmid pGEM-3zf(+) as template. 5°C step gradual washing was done in 0.1xSSC (0.15 M NaCl/0.015 M Na_3-citrate, pH 7.6), 0.1% SDS, with starting temperature of 42°C for both membranes.

2.4. Nucleotide sequence analysis

Searches for matches of nucleotide sequences in the databases were done using BLAST (Altschul et al., '90) and GeneBee (Brodsky et al., '95). The databases of GeneBank and EMBL have been used. Pairwise sequence comparisons and multiple

TABLE 1. Specimens used for satellite DNA cloning, dot and southern hybridizations and localities collected

N dots	N lane Southern	Species names[1]	Abbrev.	Localities
1	1, 2, 3	*L. viridis viridis*	*vir. vir*	Ukraine, Dnepropetrovsk
2	4	*L. media*	*med*	Armenia, Aranler
3	5	*L. strigata*	*str*	Armenia, Sevan Lake
3	6, 7	*L. strigata*	-//-	Russia, Dagestan, Mahach-Kala
4	8	*L. agilis chersonensis*	*a. che*	Byelorussia, Minsk region
4	9	*L. agilis chersonensis*	-//-	Russia, St-Petersburg region
5	10	*L. agilis agilis*	*a. agi*	Germany, Darmstadt
–	11	*L. agilis exigua*	*a. exi*	Russia, Lipetsk region
–	12	*L. agilis exigua*	-//-	Russia, Astrakhan, Solennoe village
Line A	13	*L. agilis exigua*	-//-	Russia, Kabardino-Balkaria, Nalchik
6	14	*L. agilis brevicaudata*	*a. bre*	Armenia, Aragatz, Kucbiak
7	15	*L. agilis boemica*	*a. boe*	Russia, Kabardino-Balkaria, Nalchik
7	16	*L. agilis boemica*	-//-	-//-
8	–	*L. agilis grusinica*		Russia, the Caucasus, Adler
9		*D. praticola*		Russia, Kabardino-Balkaria, Nalchik
10		*D. parvula*		Georgia, Ahaldaba
11		*P. taurica taurica*		Ukraine, Crimea, Yalta
12		*P. campestris*		Italy, Pescbicj
13		*P. tiliguerta*		Italy, Sardinia, prov. Nuoro, Monte-Albo
14		*P. muralis brugaeri*		Italy, Toscana, Rustici
15		*P. muralis nigriventris*		Italy, Volterra
16		*Z. vivipara*		Russia, St-Petersburg region; Byelorussia, Grodno
17		*D. nairensis*		Armenia, Razdan and Lchashen
18		*D. unisexualis*		Armenia, Ankavan
19		*D. armeniaca*		Armenia, Sevan
20		*D. valentini*		Armenia, Lchashen
21		*D. mixta*		Georgia, Ahaldaba
22		*D. dahli*		Armenia, Shagali
23		*D. lindholmi*		Ukraine, Crimea, Yalta
24		*D. rostombekovi*		Armenia, Dilijan
25		*D. caucasica*		Russia, Kabardino-Balkaria, Golubye ozera, ravine of river Cerek
26		*D. raddei*		Armenia, Hosrov and Gosh
27		*D. dryada*		Georgia, Adjaria, Girgami
28		*D. rudis*		Georgia, Ahaldaba
29		*D. clarcorum*		Turkey, Magden
30		*D. derjugini*		Georgia, Ahaldaba
31		*D. portschinskii*		Armenia, Stepanavan
32		*D. chlorogaster*		Azerbaijan, Dashdadjuk region
LineB		*D. sax. darevskii*		Russia, Sochi

L. – Lacerta; D. – Darevskia; P. – Podarcis; Z. – Zootoca.

sequence alignment were performed with GeneBee. GeneDoc redactor was used for the graphical editing of the aligned sequences. Phylogenetic trees were constructed using neighbour-joining (1000 bootstraps; Jin-Nei, '90; Kimura's two parameter) of TREECON (Van de Peer and De Wacher, '94) and NJ clustalW algorithm (MMfactor =8, (-0.001810) ;0.0327, 0.2617; (0.2617)) of GeneBee (Brodsky et al., '95).

3. RESULTS AND DISCUSSION

3.1. Detection, cloning, and general characterisation of a new family of tandemly repeated DNA in Lacerta agilis and L. strigata species

Both the *L. agilis* and *L. strigata* species belong to ''*L. agilis* complex''=*Lacerta* s. str. genus (Lacertidae, Sauria) (Arnold, '89), also known as

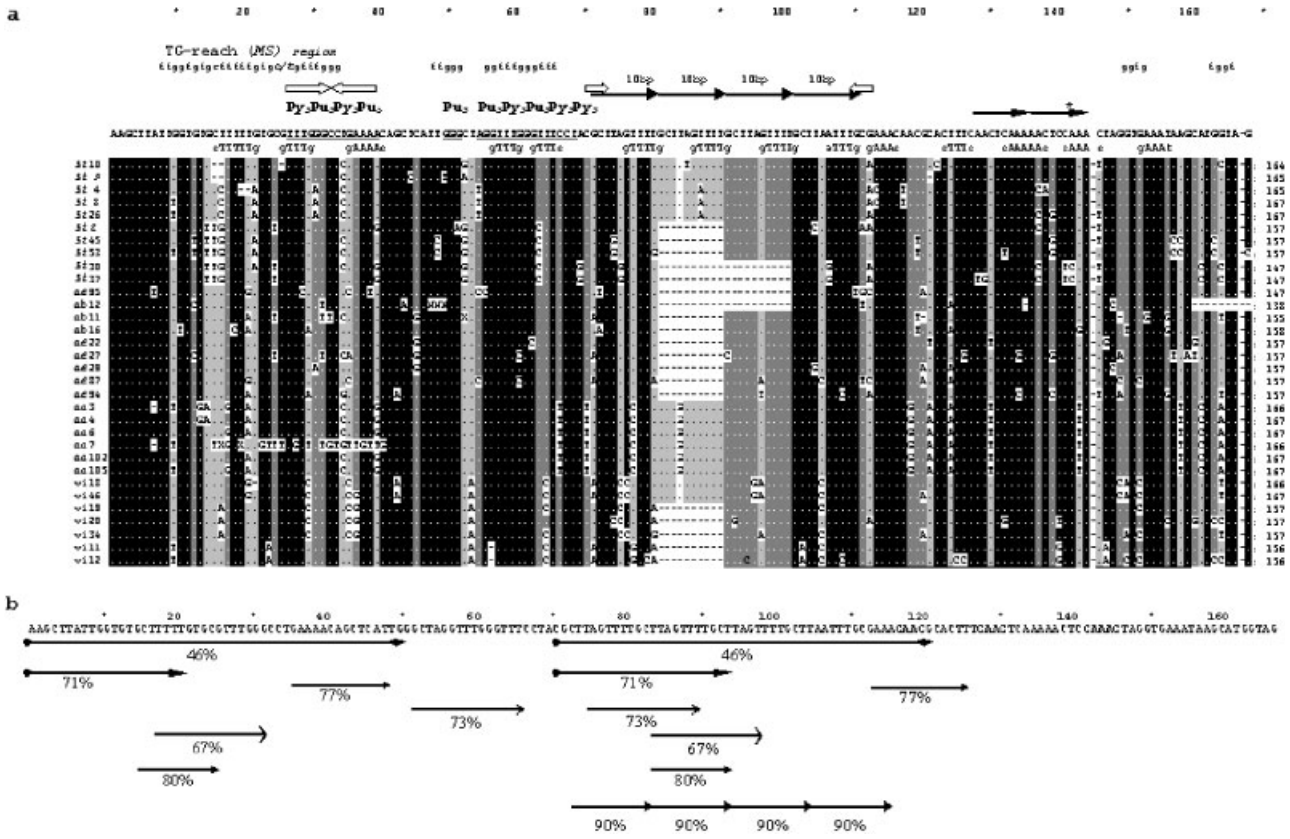

Fig. 1. a. Sequence alignment of Agi160 satDNA monomers of *Lacerta strigata* (st), *Lacerta agilis exigua* (ae) and *Lacerta agilis boemica* (ab) specimens (st4, AY184833; st9, AY184834; st10, AY184835; st2, AY184832; st8, AY652647; st26, AY652648; st30, AY652649; st37, AY652650; st45, AY652651; st53, AY652652; ae22, AY184827; ae27, AY184828; ae29, AY184829; ab11, AY184824; ab16, AY184826; ab12, AY184825; ae87, AY652638; ae94, AY652639; ae95, AY652640; aa3, AY652641; aa4, AY652642; aa6, AY652643; aa7, AY652644; aa102, AY652645; aa105 AY652646; vi1, AY652625; vi2, AY652626; vi3, AY652627; vi4, AY652628; vi5, AY652629; vi 6, AY652630; vi7, AY652631; vi8, AY652632; vi9, AY652633; vi10, AY652634; v11, AY652635; vi12, AY652636; vi13, AY652637). Consensus is shown above the aligned sequences. Dots indicate the nucleotides shared with consensus while letters show variations from the consensus sequence. Hyphens indicate the gaps. Conservative regions (one substitution per site as a maximum) are black-filled and variable ones (two or more substitutions per site), are filled with grey. The $(T)_{3-4}$ and $(A)_{3-5}$ stretches are depicted between consensus and aligned sequences. The lower-case italicized letters above the consensus indicate TG-rich, microsatellite-like regions (see text, section 3.3.); Pu_3Py_{3-5} indicates purine-pyrimidine tracts (the underlined sequence in the consensus). Several short direct repeats are shown as different arrows above; empty arrows indicate inverted repeats. b. Internal repeats of Agi160 monomer sequence. The consensus monomer of Agi160 sequences is shown as upper-case letters. The similar repeat stretches and their identity are shown as identical arrows and percentage under the arrows.

sand, or green lizards. ''*L. agilis* complex'' comprises at least five species (*L. agilis*, *L. strigata*, *L. viridis*, *L. media (trilineata)*, *L. shreiberi*) (Harris et al., '98). Four of them (excluding *L. shreiberi*) were studied in the present work.

Previous studies of Lacertidae by taxonoprint analysis showed that the members of ''*L. agilis* complex'' contained several characteristic repeated DNA fractions (Grechko et al., '98). PAAG electrophoresis of *Hind*III-digested DNA revealed a set of ''*L. agilis* complex''-specific bands in the 160–180 bp region. This material was chosen for further analysis because it is closer to a tandemly repeated DNA family CLsat (150 bp) in size. As we reported earlier, the CLsat family is characteristic of Caucasian rock lizards of ''*Lacerta saxicola* complex'' (Grechko et al., '98), currently recognised as a distinct genus *Darevskia* (Arribas, '99).

We isolated and characterised a new satDNA family called Agi160, from *L. strigata* and two subspecies of *L. agilis*, *L. a. exigua* and *L. a. boemica*. Sequencing of selected clones (see Materials and Methods) revealed ten 90–95%

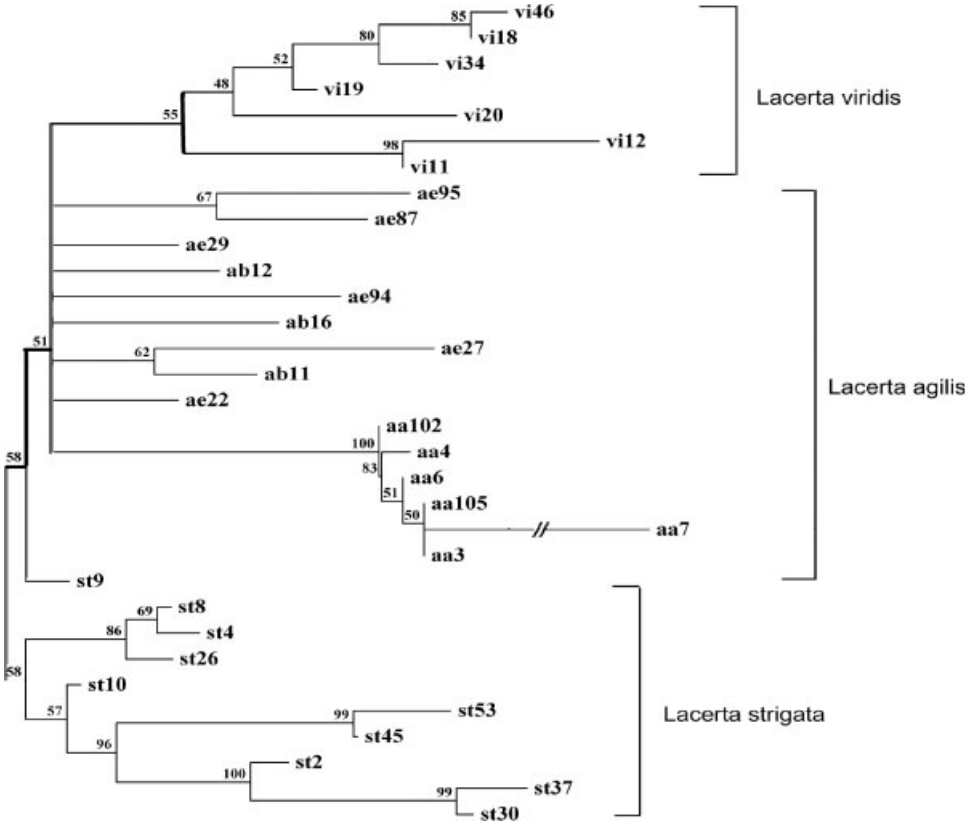

Fig. 2. Phylogenetic relationships between Agi160 satDNA repeat units. The tree shown was reconstructed using the neighbor- joining. The alignment taken in this analysis is the alignment of the monomer sequences shown on Fig. 1*. The numbers above each branch represent percentage values support (for details see section 2.4.).

identical sequences 138 and 165 bp in length. These sequences were designated as Agi160. The tandem organization of these repeats was shown by Southern hybridization analysis (see section 3.2.). The alignment of all of them revealed that the variability of the sequence length is determined by the number of 10 bp internal repeats (Fig. 1a). The similarity level of these 10 bp repeats is high enough (90%) to suggest their recent origin by amplification or insertion. The region of 10 bp repeats is flanked by two short inverted repeats (Fig. 1a). The occurrence of short 3-4 bp inverted repeats flanking the region of direct-repeat amplification has been documented for several satellites (Bonnewell et al., '83; Kato and Yoshida, '95).

Analysis of Agi160 monomers by several methods of phylogenetic reconstruction reveals several subfamilies of repeats (Fig. 2). Well-supported groupings of repeats are specific to both *L. strigata* and *L. agilis*, although some groupings contain monomers from different subspecies of *L. agilis*. The grouping of the repeat subfamilies was supported the both by analyses taking indels into account and ignoring them. Hence, the monomer variability, specifically in the region of 10 bp repeats, does not contain any phylogenetic signal. However, bootstrap values in the case of Agi160 monomers are indicative of the lower divergence rate compared to homogenization. The pattern of branching suggests that these DNA repeats remain conserved for million of years, the similarity of sequences suggests both a low evolutionary rate of this satDNA and close genetic relation of *Lacerta strigata* to *Lacerta agilis*, which is consistent with previous results (Darevsky et al., '76; Rykena, '91). Darevsky reported that *L. agilis* and *L. strigata* share common morphological characters in sympatric regions, which is thought to result from ongoing introgressive hybridization (Darevsky et al., '76). The possibility of hybridization between these two species (with formation of normal descendants) was also demonstrated in laboratory experiments (Bischoff, '69, '73; Nettmann and Rykena, '74). Nonetheless, our data demonstrate that Agi160

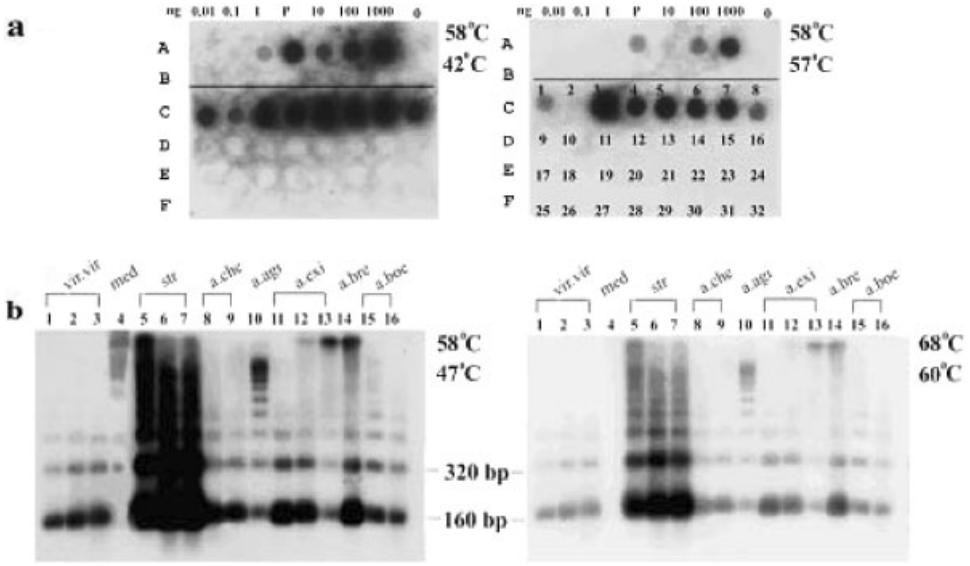

Fig. 3. Hybridization analysis of Agi160 satDNA within the Lacertidae. a. Dot-blot hybridization analysis of 32 species of lizards, representing *Lacerta* s. str., *Zootoca*, *Podarcis* and *Darevskia* genera. Hybridization was done at 58°C, followed by 5°C step gradual washes: 42, 47, 53, and 57.5°C (for details see text 2.4.). Here we present only the first – 42°C and the last – 57.5°C steps of washing. A-F letters at the left side of the picture indicate the dot lines. Graded concentrations of DNA are shown above (0.01–1000 ng). Line A: genomic DNA of *L. agilis exigua*. P – Plasmid carrying Agi160 DNA monomer sequence (ae22 satellite clone), 1ng. Line B: genomic DNA of *D. saxicola sczcerbaki*. P – Plasmid carrying CLsat DNA monomer sequence (Scz111 satellite clone), 1 ng. The lower four lines show hybridization with 10^3 ng of genomic DNA of each species studied. Line C: *Lacerta* s. str. species. Line D: *Darevskia*, *Podarcis* and *Zootoca* species. Lines E and F: *Darevskia* species. The numbers on dots indicate species as given in Table 1. b. Southern hybridization of Agi160 probe with *Hind*III completely digested genomic DNA of *Lacerta* s. str. species. Hybridization temperature was 58°C and 68°C. Each lane contains 10^3 ng of DNA. Washing was done at 47°C and 60°C respectively. Species names are given above the numbered lanes as abbreviations (see Table 1).

sequences may be used to discriminate these two species.

GenBank nucleotide sequence screening demonstrates that Agi160 tandem repeats represent a novel family of satDNA. Several homologies revealed in our paper will be further discussed in sections 3.2. and 3.4.

3.2. Agi160 hybridization analyses

Dot-blot hybridization analysis revealed the presence of Agi160 satellite in *Lacerta* s. str. species only (Fig. 3a). Even under mild hybridization conditions (58°C), the presence of Agi160 was not detected in other species of the Lacertidae family neither the closely related genera *Zootoca* (*Z. vivipara*), *Podarcis* and *Darevskia* (Fig. 3a), nor the more distant genera *Ophisops, Eremias,* and *Gallotia* (Southern hybridization under the same conditions, the data are not presented here).

The analysis of hybridization patterns revealed that the intensity of hybridization signal decreases with the increase of washing temperature. This fact reflects the level of divergence between Agi160 satellites of different *Lacerta* s. str. species and permits evaluation of the relative amount of satDNA similar to Agi160 in different species. Thus, a gradual wash shows that the intensity of hybridization signal with Agi160 satDNA drastically decreases in *L. media* with the increase of washing temperature and slightly in *L. viridis*, in contrast to *L. agilis* (dots 4–7) and *L. strigata* (Fig.3a). The estimated amounts of Agi160 repeats in *L. viridis, L. agilis,* and *L. strigata* are 7×10^2, $1.4–2 \times 10^4$ and more than 7×10^6 copies/haploid genome respectively. *L. media* shows much lower hybridization intensity even at the very beginning (at 42°C), and practically no detectable signal at higher temperature (Fig. 3a). This result suggests the presence of very few if any highly related Agi160 sequences. Therefore, the first two species seem less similar to *L. agilis* and *L. strigata,* and *L. media* may be considered the most distant species. The observed relationships agree with results of interspecific hybridization, morphology, zoogeography (Darevsky et al., '76; Arnold, '89;

Rykena, '91) and molecular data (Mayer and Benyr, '94) analysis.

Southern hybridization analysis of *L. viridis*, *L. media*, *L. strigata* and five subspecies of *L. agilis* was done under high-stringency conditions (68°C), following the complete *Hind*III digestion of the genomic DNA. As a result, we revealed a ladder-like pattern with intense monomer and dimer bands around 160 bp and 350 bp respectively, and higher-order bands resulting probably from the loss of a restriction site (Fig. 3b). This pattern proves the tandem nature of Agi160 repeats. The absence of hybridization between the DNA of *L. media* and Agi160 at 68°C is an exception. This fact, together with the detected very low content at 58°C on dot hybridization (Fig. 3a), confirms the high level of divergence between *L. media* satDNA and Agi160 inferred from dot-blot experiments. The Southern hybridization also reflects interspecific variation in quantity of Agi160 satDNA. Figure 3b shows that the amount of Agi160 repeat in *L. agilis* is lower than that in *L. strigata.* Mild washing reveals a similar content of Agi160 in *L. viridis* and in *L. agilis*, whereas strong washing shows that the amount of Agi160 in *L. viridis* is lower than that in *L. agilis* and *L. strigata* (Fig. 3b). The data are consistent with the results of dot-blot hybridization analysis and provide support for the existence of a significant amount of related but highly divergent Agi160 sequences in the genome of *L. viridis.*

It must be mentioned here that correct interpretation of the presence/absence of a satDNA is necessary to exclude homoplasy in terms of distribution of such families among extant species. It is thought, as I mentioned earlier, that satDNA evolution combine neutrally evolving monomer sequence and specific mode of evolution of tandem repeats, namely concerted evolution. Therefore a satDNA could be used as a phylogenetic marker only if the rate of concerted evolution is high enough and all copies in the repeat array may appear to be evolving almost as a single sequence. In this case the distinction between orthology and paralogy is blurred (Patterson, '88). So, we assume that a satDNA family could be used for phylogenetic assays/comparison between species only when the copy number of repeats is high enough to undergo a concerted evolution. This does mean that the presence of a satDNA family could be used as phylogenetic tool, while the absence involves much more ambiguity. However in the most cases the undetectable level of a satDNA family by quantitative hybridization points out to a fixed event of divergence of studied taxon. When the event of excision of a large array of repeated DNA becomes fixed in a population, this leads to the genetic divergence of this population.

3.3. Structure and variability of Agi160 repeats

The alignment of ten Agi160 sequences of cloned monomers was done to elucidate the monomer organization (Fig. 1a). The variability of the lengths of the satellite monomer unit (138–165 bp) primarily results from the different number of direct 10 bp repeats: some units contain four repeats, whereas others contain three or two only. One to three bp indels also contribute to the length variability in Agi160 (indicated with hyphens at the following positions: 16-17; 20-21; 26; 121; 122; 136; 146; 150; 165).

The analysis of nucleotide sequence variability revealed that hot-spot mutation segments alternate with conservative regions (Fig. 1a). Agi160 begins with the 10 bp region 100% identitical in all monomers. This region is followed by a variable TG-rich region 17 bp in length. Several copies of the latter region share 95–100% similarity with the ''GGTGTTTGTTTTGTGTGTTT'' microsatellite (ms) fragment from *Salmo salar* (AF257039.1), *Coffea arabica* (AJ308835.1; AJ308786.1; AJ308750.1; AJ308749.1), and genomic scaffold from *Drosophila melanogaster* (AE003669; AE003746). The 3'-end of the GT-''ms'' region forms a 12 bp palindrome with the 5'-end of the next highly variable part of Agi160. The remaining three regions shown in Figure 1a are: (i) the highly-conservative region containing another GT motif (GGTTTGGGTTT); (ii) the direct 10 bp repeats region with a conservative nucleotide sequence and a variable number of copies (2 to 4), flanked by short inverted repeats; (iii) A-rich variable region that includes four CAA(AA)C repeats, one GAAAC, and one GAAAT. Interestingly, the (T) runs of the DNA strand depicted in Fig. 1a are usually flanked by G-bases, making the complementary strand to be reach in CAA(AA)C repeats. The abundance of (GTTTG/CAAAC) stretches may be indicative of their importance for the structure and functioning of Agi160. It has been hypothesised and even proved in some cases that poly(A) stretches are associated with nuclear matrix (Lobov et al., 2001). Therefore, sequences abundant in $(A)_{3\text{-}4}$ motifs may be involved in the chromatin compactization and gene regulation by

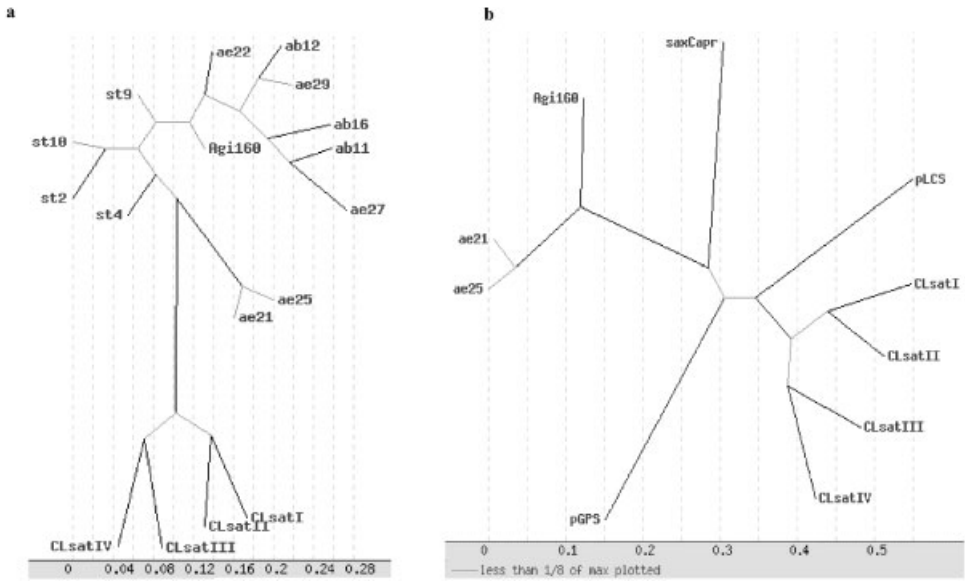

Fig. 4. Unrooted phylogenetic trees of sequences of Lacertidae satDNA. Trees shown were reconstructed using the neighbor joining method and the clustalW algorithm (for details see section 2.4.). a. Consensus sequences of CLsat subfamilies (cited in section 3.4.) and Agi160 consensus and monomer sequences shown on Fig. 1. b. Consensus sequences of Lacertidae satDNA. The alignment in this analysis is based on monomer sequences cited in section 3.4.

spatial positioning of some sequences (Plohl et al., '98; Trifonov, '99; Gottesfeld and Luger, 2001), and therefore may be exposed to natural selection.

Analysis of the AT-GC content and distribution reveals 60% AT base pairs and 11 to 13 short A-T runs, nine TTT(T), and one AAAA located in the first two thirds of the repeat unit. Four AAA(A) are located in the last third of the sequence. The A and T runs together form 30% of the Agi160 nucleotide sequence. The G/A (11) and C/T (7) transitions appeared to be the most frequent nucleotide changes in Agi160 family. We also found a considerable number of G/T transversions (7). It was found that G→A and C→T transitions occur much more frequently than the reverse ones. The same tendency of single-nucleotide changes was also documented for some other families of satellite DNA (Castagnone-Sereno et al., '98). It was supposed that high AT content was a characteristic feature of amphibian and reptilian genomes (Olmo, '91), although GC-rich fractions also occur (Olmo et al., '91).

Besides T-A clusters, there were also several GGG and CCC repeats leading, together with T and A runs, to two $(Py)_3$ $(Pu)_3$ $(Py)_3$ $(Pu)_3$ motifs. One of them follows the ''ms'' stretch, while the other precedes the region of direct 10-bp repeats. Searching for other repetitive motifs demonstrated occurrence of several short direct and inverted sequences (Fig. 1a).

Pairwise alignment of Agi160 monomer units also revealed longer repetitive stretches of a variable level of divergence (Fig. 1b). This structural feature, known as cryptic simplicity (Tautz et al., '86), was described for other families of satellite DNA (Plohl and Ugarkovic, '94) including one typical of lizards: CLsat (Ciobanu et al., 2001, 2003). The cryptic simplicity is thought to result from the joint action of replication slippage and point mutations, together with gene conversion-like and unequal crossing-over events (Tautz et al., '86). Dynamics of this interaction and its effect on the degree of degeneracy of internal repeats were discussed earlier (Stephan, '89). The structural features of Agi160 satDNA, analysed in this paper, point to the role of the short motif duplication and replication slippage events in the formation of this family of tandem repeats. Furthermore, replication slippage and single nucleotide changes are also involved in the process of divergence between Agi160 monomers. For instance, the length variability in Agi160, which might pass through the deletion/duplication of 10 bp, can be a result of replication slippage. The $(T)_2$ $(G)_2$ TGTG$(T)_n$ TGTGTCG$(T)_3$ $(G)_3$ ''ms'' motif, in its turn, may be a product of joint action of a number of duplications, replication slippage of TTGGTG sequence, and point mutations. Moreover, the $(T)_2$ $(G)_3$ N_3 $(G)_2$ $(T)_3$ $(G)_3$ $(T)_3$ region at 48–67 position in all monomers, the TTG

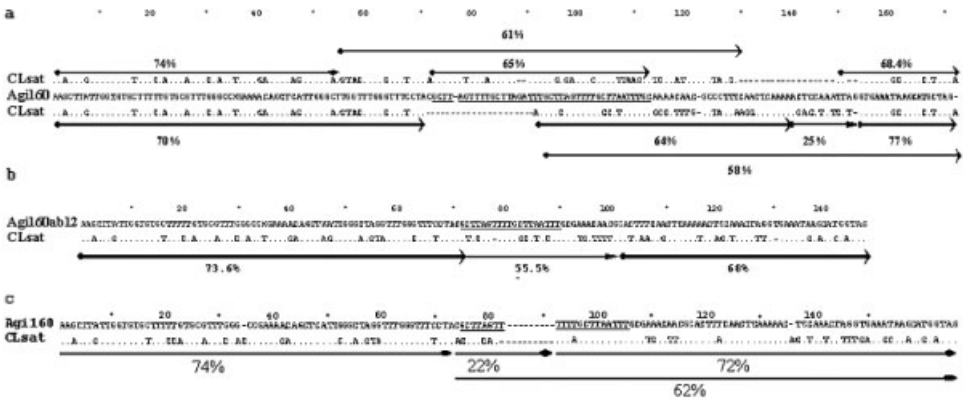

Fig. 5. Sequence similarity of Agi160 and CLsat monomers. a. Two alternative alignments of the Agi160ab12 monomer (AY184825) and the CLsat monomer (LCL308334). b. Alignment variants of the Agi160st4 monomer (AY184833) and the CLsat monomer (LMI252108) (see text, section 3.4.). c. The alignment of Agi160 and CLsat consensus monomers. Sequences of Agi160 monomers are shown as rows of upper-case letters. The region of 10 bp repeats (see text, section 3.3.) is underlined. Dots indicate the CLsat nucleotides shared with the Agi160 sequences. Hyphens indicate the gaps. Arrows show the distribution of more or less conserved regions. The sequence similarities are indicated as percentage numbers above and under arrows.

substitution at 15-17 position of monomer st2, as well as deletions of 2 bp at the same position in monomers st9 and st10, and at position 19–20 in monomer st4 result most likely from replication slippage (Fig. 1a).

3.4. Comparative analysis of satDNA within the family Lacertidae

Four satellite DNA families have been described for the genomes of lacertid lizards. The pLCS (190 bp) is specific for the genus *Podarcis* and two closely related species of *Algyroides* now included in *Podarcis* (Capriglione, '95); the pLHS (140 bp), specific for *Podarcis* only, is most probably a variant of pLCS (Capriglione, '95); the pGPS (185 bp) is present in the genome of *Podarcis* and a number of other taxa including *Algyroides moreoticus, Lacerta graeca, Lacerta viridis, Lacerta(-Zootoca) vivipara* and *Archaeolacerta bedriagae* (Capriglione, '95). This satellite family shows some similarity with the alphoid satDNA, and the authors suggest that it has an ancient origin (preceding at least the divergence within sub. fam. Lacertinae). The fourth type of lizard satDNA, named CLsat, was found to be specific for *Darevskia* (Grechko et al., '98; Rudykh et al., '99). This genus was distinguished from *Lacerta* s. lat. in 1999 only (Arribas, '99). Sequence comparison of Agi160 *Lacerta* s. str. satDNAs with the data on these reptilian satellites revealed several common features, such as the same range of monomer lengths $\approx$ 140–190 bp (see references above), AT content (tendency toward AT enrichment 50–65%) and homopolymeric (A_{3-4} and T_{3-4}) stretches (Capriglione et al., '91; Ciobanu et al., 2001). All these features were also found common for other non-reptilian satellites (King and Cummings, '97 and citations). Besides, the Agi160 sequences do not demonstrate significant similarity with the first three families of reptilian satDNA (pLCS, pLHS, pGPS). The separate placement of Agi160 DNA from all Lacertidae satDNA already described is also supported by a clustal analysis of consensus sequences (Fig. 4).

The Agi160 and CLsat monomers are characterised by the same distribution of A and T-runs and the existence of several internal repeats of a different level of divergence (absent in the other described reptilian satDNA), as well as by the common ''TG-ms'' region. Moreover, a comparison between CLsat and Agi160 revealed 60 bp regions with the 60–75% sequence identity (Fig. 4). Agi160 and CLsat monomers have an overall sequence similarity range from 50 to 69% and fail to form hybrid molecules even under mild stringency conditions (58°C). Since the Agi160 set of monomers significantly differs in length and variability level within the last third of the sequences (Fig. 1a), we compared all of them with the CLsat set of monomers available in the GeneBank database. The highest level of similarity was revealed for one group of the aligned pairs. It is formed by the pairs between the Agi160ab12 (AY184825) with two 10 bp repeats and CLsat monomer (LCL308334), as well as several other CLsat monomers (Fig. 4a). The monomer length of Agi160ab12 is the same as the length of CLsat monomers. However, the

pattern of similarity between the other Agi160 and CLsat proved to be similar for all of them. Thus, we can distinguish one ''conserved'' region for 70 bp preceding the zone of 10 bp-repeats of Agi160 monomers and another ''variable'' region constituting the remaining part of the Agi160 monomer. The conserved region shows 70–74% similarity and contains the TG- rich ''ms'' fragment (Figs. 1a and 4) that is highly conserved in CLsat monomers (Ciobanu et al., 2001, 2003). The variable region shows a more variable level of similarity (55–65%) with CLsat, as well as variable distribution of more or less similar regions. This region can be further described as consisting of two parts: (i) the zone of 10 bp direct repeats and (ii) A-rich 3'- part of Agi160 (see Fig.1a), which corresponds to randomly distributed $T_{3–5}$ and $A_{3–5}$ stretches in CLsat. To elucidate the origin and evolutionary scenario of the 10 bp-repeats region, we analysed the alignments between Agi160 monomers containing four 10 bp and CLsat monomers. The highest level of similarity was found for Agi160st4 (AY184833) and CLsat monomer (LMI252108) (Fig. 4b). Two variants of a gap position between Agi160st4 and CLsat can be proposed (Fig. 4b). The level of sequence identity is similar in both variants. The first variant of alignment illustrates the similarity between Agi160 and CLsat in the region of 10 bp direct repeats. It is 65%, while for the other Agi160 and CLsat monomers it is 55.5–60%. In the second variant (with a gap in the region of 10 bp repeats), the pattern of conservative regions is the same, but the variable regions are less similar (58%–61%). Thus, both variants of alignment suggest that the variable zone emerged after the divergence of *Lacerta* s. str. and *Darevskia*. As well as the higher similarity (90%) between all 10 bp repeats, suggests amplification of this region after the divergence of the rest part of the Agi160 monomer.

Our data suggest two possible evolutionary pathways of Agi160 and CLsat DNA. According to the first scenario, the Agi160 monomer originated from the 70 bp ''conserved'' region after the divergence of *Lacerta* s. str. and *Darevskia*. This could result from several slippage amplification events and is supported by the presence and different distribution of the shorter and less degenerated internal repeats in both Agi160 and CLsat monomers—the ''cryptic simplicity'' structure (Fig. 1b) specific also for other satellite DNA (Tautz et al., '86; Plohl and Ugarkovic, '94).

The second scenario suggests that duplication of the 70 bp conserved region occurred prior to the divergence of *Lacerta* s. str. and *Darevskia*. Note that ''conserved'' and ''variable'' regions correspond to the longest but highly degenerated internal repeat. In Agi160 similarity is 46% for the segment of 50 bp and 36% for the segment of 70 bp (Fig. 1b, section 3.2.). In CLsat similarity is 40% for the segment of 70 bp (Ciobanu et al., 2001, 2003). The low similarity between the longest internal repeats as compared to the significant similarity between Agi160 and CLsat in these regions favors the second interpretation.

The observed sequence similarity, as it is also seen from the consensus alignment (Fig. 4c), suggests the common origin of Agi160 and CLsat families of satDNA. Based on paleontological and immunological data on the divergence of *Lacerta* s. str. and *Darevskia* (Darevsky, '90; Mayer and Benyr, '94), we assume that the original sequence was present in ancestral genomes more than 25 Myr ago.

4. CONCLUSIONS

Regions with high and low sequence similarity between Agi160 and CLsat satellites suggest the possibility of selective maintenance of these sequences. Thus, the TG-rich region with approximately 75% similarity between two satellite families also appeared to be conserved in CLsat monomers. The conservative pattern (e.g., regional similarity, distribution of AT-tracts and AT-ratio, similar monomer length), shared by satellites that split more than 25 Myr ago suggests possible functional importance of these features. The observed similarity between Agi160 and CLsat families of satDNA is indicative of their common origin and has important implications for better understanding the interplay of molecular evolutionary mechanisms for forming various families of satDNA. Our data support monophyly of ''complex *Lacerta agilis*.''

ACKNOWLEDGEMENTS

We are grateful to our colleagues who had helped us with collecting lizards of ''*L. agilis* complex,'' to Dr. N. Vassetzky, for helpful discussions and participation in phylogenetic estimations of the data. We thank Dr. D. Frynta and V.F. Orlova for their valuable discussions on this subject and I.A. Rudykh for her valuable critical notes.

LITERATURE CITED

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215: 403–410.

Arnold EN. 1989. Towards a phylogeny and biogeography of the Lacertidae: relationships within an Old-World family of lizards derived from morphology. Bull Br Mus (Nat His) Zool 55:209–257.

Arribas OJ. 1999. Phylogeny and relationships of the mountain lizards of Europe and near East (Archeolacerta Mertens, 1921, sensu lato) and their relationships among the Eurasian Lacertid radiation. Russ J Herpetol 6:1–22.

Beridze TG. 1986. Satellite DNA. Springer, Berlin, Heidelberg, New York, Tokyo, 3–26.

Bischoff W. 1969. Lacerten-Bastarde. Lacerta strigata × Lacerta a. agilis und Lacerta strigata × Lacerta viridis meridionalis. Aquarien Terrarien, Leipzig 16:262–265.

Bischoff W. 1973. Lacertenbastarde II. Zoologischer Garten NF, Berlin 43:278–291.

Bonnewell V, Fowler RF, Skinner DM. 1983. An inverted repeat borders a fivefold amplification in satellite DNA. Science 221: 862–865.

Brodsky LI, Ivanov VV, Kalaidzidis YaL, Leontovich AM, Nikolaev VK, Feranchuk SI, Drachev VA. 1995. GeneBee-NET: Internet-based server for analyzing biopolymers structure. Biochemistry 60:923–928.

Capriglione T. 1995. Satellite DNA and phylogeny of Lacertid lizards. In: Llorente GA, Montori A, Santos X, Carretero MA, editors. Scientia Herpetologica p. 68–70.

Capriglione T, Cardone A, Odierna G, Olmo E. 1991. Evolution of a centromeric satellite DNA and phylogeny of lacertid lizards. Comp Biochem Physiol B 100: 641–645.

Castagnone-Sereno P, Semblat JP, Leroy F, Abad P. 1998. A new AluI satellite DNA in the root-node nematode *Meloidogyne fallax*: relationships with satellites from the sympatric species *M. halpa* and *M. chitwoodi*. Mol Biol Evol 15:1115–1122.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220.

Ciobanu D, Roudykh IA, Grechko VV, Kramerov DA. 2001. Molecular mechanisms of satellite DNA evolution in a group of closely related Caucasian rock lizards. In: Piruzian E, Volkova L, et al., editors. International Symposium Molecular Mechanisms of Genetic Processes and Biotechnology, Moskow. Thesis p 26.

Ciobanu D, Grechko VV., Darevsky IS. 2003. Molecular Evolution of Satellite DNA CLsat in Lizards from the Genus *Darevskia* (Sauria: Lacertidae): Correlation with Species Diversity. Russ J Genetics 39:1527–1541.

Cox R, Mirkin SM. 1997. Characteristic enrichment of DNA repeats in different genomes. Proc Natl Acad Sci USA 94:5237–5242.

Darevsky IS. 1990. Lizard of the genus *Lacerta* from the middle sarmatian lake deposits of the Northern Caucasus. Proc Zool Inst Leningrad 207:139–142.

Darevsky IS, Szczerbak NN, Peters G., et al. 1976. Origin and phylogeny of *Lacerta agilis* species. In: Yablokov AV, editor. Sand Lizard, Nauka, Moscow. p 91–95.

Elder JF, Turner BJ. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70: 297–320.

Dover GA, Brown S, Coen E, Dallas J, Strachan T, Trick M. 1982. The dynamics of genome evolution and species differentiation. In: Dover GA, Flavel RB, editors. Genome evolution. Academic Press, London.

Gottesfeld JM, Luger K. 2001. Energetics and affinity of the histone octamer for defined DNA sequences. Biochemistry 40:10927–10933.

Grechko VV. 2002. Molecular DNA markers in phylogeny and systematics. Russ J Genetics 38:851–868. Review.

Grechko VV, Ryabinin DM, Fedorova LV, Rudykh IA, Fedorov AN, Ryskov AP, Semenova SK, Darevsky IS. 1998. Molecular-genetic classification and phylogenetic relatedness of some species of Lacertidae lizards by taxonoprint data. Mol Biol 32:172–183.

Harris DJ, Arnold EN, Thomas RH. 1998. Relationships of lacertid lizards (Reptilia: Lacertidae) estimated from mitochondrial DNA sequences and morphology. Proc R Soc Lond B 265:1939–1948.

Jin L, Nei M. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7:82–102.

Kato M, Yoshida M. 1995. Nucleotide sequence of a major satellite DNA element isolated from a saltwater fish Sillago japonica. Mol Biol Rep 22:33–35.

King LM, Cummings MP. 1997. Satellite DNA repeat sequence variation is low in three species of burying beetles in the genus *Nicroporus* (Coleoptera: Silphidae). Mol Biol Evol 14:1088–1095.

Lee SY, Rasheed S. 1990. A simple procedure for maximum yield of high-quality plasmid DNA. Biotechniques 9: 676–679.

Lobov IB, Tsutsui K, Mitchell AR, Podgornaya OI. 2001. Specificity of SAF-A and lamin B binding *in vitro* correlates with the satellite DNA bending state. J Cell Biochem 83:218–229.

Mayer W, Benyr G. 1994. Albumin evolution and phylogenese in der familie Lacertidae. Ann Naturhist Mus Wien 96: 621–648.

Nettmann HK, Rykena S. 1974. Eine gelungene Kreuzung von Lacerta t. trilineata mit Lacerta v. viridis (Sauria, Lacertidae). Salamandra, Frankfurt/Main 10:83–85.

Olmo E. 1991. Genome variations in the transition from amphibians to reptiles. J Mol Evol 33:68–75.

Olmo E, Odierna G, Capriglione T. 1991. Genome evolution in reptiles. Symposium on the evolution of terrestrial vertebrates. Ghiara G et al., editors. Selected symposia and Monographs UZI 4. Mucchi, Modena. p 255–267.

Patterson C. 1988. Homology in Classical and Molecular Biology. Molecular Biology and Evolution 5: 603–625.

Plohl M, Ugarkovic D. 1994. Analysis of divergence of *Alphitobius diaperinus* satellite DNA – roles of recombination, replication slippage and gene conversion. Mol Gen Genet 242:297–304.

Plohl M, Mestrovic N, Bruvo B, Ugarkovic D. 1998. Similarity of structural features and evolution of satellite DNAs from *Palorus subdepresus* (Coleoptera) and related species. J Mol Evol 46:234–239.

Rudykh IA, Grechko VV, Kramerov DA, Darevsky IS. 1999. Distribution of *Hind*III-repeats in genomes of Caucasian lizards of the Lacerta species reflect their phylogenetic affiliation Dokl Akad Nauk 367:563–566.

Rykena S. 1991. Hybridization experiment as tests for species boundaries in the genus *Lacerta* sensu stricto. Mitt Zool Mus Berl 1:55–67.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: A laboratory manual, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Stephan W. 1989. Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6:198–212.

Tautz D, Trick M, Dover GA. 1986. Cryptic simplicity in DNA is a major source of genetic variation. Nature 322: 652–656.

Trifonov EN. 1999. Tuning function of tandemly repeating sequences: a molecular device for fast adaptation. Proc Intern Conf Evol Genomics. Costa Rica. Thesis. p 74

Van de Peer Y, De Wachter R. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Applic Biosci 10:569–570.